2013

Essays on information, inattention, and ambiguity

BOSTON UNIVERSITY

GRADUATE SCHOOL OF ARTS AND SCIENCES

Dissertation

ESSAYS ON INFORMATION, INATTENTION, AND AMBIGUITY

by

ANDREW ELLIS

B.A., University of North Carolina at Chapel Hill, 2007

Submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

2013

Approved by

First Reader

Larry G. Epstein
Professor of Economics

Second Reader

Barton Lipman
Professor of Economics

Third Reader

Jawwad Noor
Associate Professor of Economics

Fourth Reader

Laurent Bouton
Assistant Professor of Economics

Acknowledgments

First and foremost, I would like to thank my advisor Larry Epstein. My many long conversations with him have shaped my view of economics, and I cannot imagine an advisor more generous with his time and advice. This dissertation would not have been possible without his patience, support, advice, and guidance.

I am also particularly grateful to Laurent Bouton, Bart Lipman, and Jawwad Noor for their support and guidance throughout my time at Boston University. All three have always been willing to share their opinions and advice. Their input has greatly improved this work and me as an economist.

I would also like to thank the others who have shaped my views on economics, including (in alphabetical order) Tim Bond, Ralph Byrns, Andrea Canidio, Gabriele Gratton, Sambuddha Ghosh, Sean Horan, Asen Kochov, Dilip Mookherjee, Andy Newman, Bill Parke, Max Parrish, Dana Rotz, Jeremy Smith, and Boone Turchi. Moreover, I would like to thank my family and friends, especially Sunida Susantud, for their support. Financial support from the Boston University Economics Department, the Institute for Economic Development, and the Huse Foundation was essential in the completion of this work.

In writing the first chapter, I am especially grateful to Larry Epstein, Bart Lipman and Jawwad Noor for their advice, support and guidance. I would also like to thank Laurent Bouton, Sambuddha Ghosh, Sean Horan, Asen Kochov, Albert Ma, Andy Newman, Pietro Ortoleva, Dana Rotz, Todd Sarver, Jeremy Smith, Tomasz Strazlecki and various conference participants for helpful conversations and comments.

In writing the second chapter, I would especially like to thank Laurent Bouton, Larry Epstein, Bart Lipman, and Jawwad Noor. I would also like to thank Sandeep Baliga, Jacopo Bizzotto, Randy Ellis, Albert Ma, Michael Manove, Andy Newman, Juan Ortner, and Marc Rysman for helpful conversations and comments.

In writing the third chapter, I would especially like to thank Laurent Bouton, Larry Epstein and Bart Lipman for advice, encouragement and detailed comments. I would like to thank also Sophie Bade, Dino Gerardi, Sambuddha Ghosh, Gabriele Gratton, Sean Horan, Peter Klibanoff, Alessandro Lizzeri, Dilip Mookherjee, Jawwad Noor, Debraj Ray, Jeremy Smith, Tomasz Strzalecki, Tak-Yuen Wong and various conference participants for helpful conversations and comments.

ESSAYS ON INFORMATION, INATTENTION, AND AMBIGUITY

(Order No.)

ANDREW ELLIS

Boston University, Graduate School of Arts and Sciences, 2013
Major Professor: Larry G. Epstein, Professor of Economics

ABSTRACT

This dissertation consists of three essays studying economic agents with non-standard reactions to information. The standard model is often inadequate because it permits neither inattention nor ambiguity aversion. This dissertation provides both pure and applied theoretical analyses of these two phenomena.

The first essay models an agent who has a limited capacity to pay attention to information and thus conditions her actions on a coarsening of the available information. An optimally inattentive agent chooses both her coarsening and her actions by constrained maximization of an underlying subjective expected utility preference relation. The main result axiomatically characterizes the conditional choices of actions by an agent that are necessary and sufficient for her behavior to be seen as if it is the result of optimal inattention. The agent's utility index, cognitive constraint and prior are uniquely identified.

The second essay analyzes the implications of advertising in a model where consumers are optimally inattentive. Firms compete by choosing both prices and ad-

vertising levels. Consumers easily observe price but have a limited capacity to pay attention to information about quality. Advertisements increase consumer capacity for attention. An increase in capacity for attention results in more information processed by each consumer, which raises the likelihood that a high quality good is purchased but leads to an increased price. An exogenous decrease in the cost of advertising has a positive impact on equilibrium price but an ambiguous effect on equilibrium profit and surplus. Advertising generates some effects documented in the literature through a single mechanism.

The third essay studies strategic voting when voters have pure common values but exhibit Ellsberg-type behavior as modeled by maxmin expected utility preferences. The Condorcet Jury Theorem states that given subjective expected utility maximization and common values, the equilibrium probability that the correct candidate wins goes to one as the size of the electorate goes to infinity. In contrast, this essay provides sufficient conditions so that the equilibrium probability of the correct candidate winning the election is bounded above by one half in at least one state. As a consequence, there is no equilibrium in which information aggregates.

Contents

1 Foundations for Optimal Inattention **1**

1.1 Introduction 1

1.1.1 Objectives and Outline 1

1.1.2 Example 4

1.2 Setup and Model 7

1.2.1 Setup 7

1.2.2 Model 9

1.3 Foundations 11

1.3.1 Axioms 11

1.3.2 Characterization Result 18

1.3.3 Special cases 19

1.3.4 Counter-examples 20

1.4 Identification 22

1.5 Comparative Attention and the Value of Information 25

1.6 Application: Markets with Optimally Inattentive Consumers 28

1.6.1 Model 29

1.6.2 Equilibrium 30

1.7 Conclusion 34

2 Advertising and Inattention **36**

2.1 Introduction 36

2.2 The Model . . . 39
2.3 Consumer Behavior and Firm Demand . . . 43
2.3.1 Consumer Behavior . . . 43
2.3.2 Firm Demand . . . 46
2.3.3 Spillover . . . 50
2.4 Baseline: No Advertising . . . 51
2.4.1 Equilibrium . . . 52
2.4.2 Comparative Statics . . . 52
2.4.3 Welfare Analysis . . . 53
2.5 Equilibrium with Advertising . . . 55
2.5.1 Pricing when advertising is exogenous . . . 55
2.5.2 Equilibrium . . . 56
2.5.3 Comparative statics . . . 57
2.5.4 Welfare analysis . . . 58
2.6 Discussion . . . 60
3 Condorcet Meets Ellsberg **63**
3.1 Introduction . . . 63
3.2 Sincere Voting and Ambiguity . . . 66
3.3 Ambiguous Poisson Games . . . 69
3.4 The Condorcet Jury Theorem . . . 72
3.4.1 Ambiguous voting games . . . 72
3.4.2 Main result . . . 73
3.4.3 Information aggregation . . . 76
3.5 Strategic Abstention . . . 78
3.6 Conclusion . . . 81

A Appendix for Chapter 1 **83**

A.1 Preview of the Proof of Theorem 1 83

A.2 Proofs 85

A.2.1 Proofs from Section 1.3 85

A.2.2 Proof from Section 1.4 100

A.2.3 Proofs from Section 1.5 106

A.2.4 Proofs from Section 1.6 108

A.3 Counter-Examples 110

A.3.1 All but ACI 110

A.3.2 All but INRA 111

A.3.3 All but Monotonicity 112

A.3.4 All but Subjective Consequentialism 112

A.3.5 All but Continuity 113

A.3.6 Behavior compatible with optimal inattention but not inattention to alternatives 114

A.3.7 Behavior compatible with inattention to alternatives but not optimal inattention 115

B Appendix for Chapter 2 **116**

B.1 Proofs not in text 116

B.1.1 Proof of Lemma 2 116

B.1.2 Proof of Proposition 5 118

B.1.3 Proof of Proposition 6 119

B.1.4 Details for $p^* = v - t/2$ 120

B.1.5 Proof of Proposition 8 121

B.1.6 Proof of Proposition 9 123

B.1.7 Proof of Proposition 10 124

C Appendix for Chapter 3 **127**

C.1 Details for Section 3.2 127

C.2 Proof of Theorem 6 128

C.3 Preliminaries for the Remaining Proofs 130

C.4 Proofs from Section 3.4 131

C.5 Proofs from Section 3.5 147

References **165**

List of Tables

1.1 Conditional choices 5

2.1 Demand for firm 2 51

List of Figures

1.2.1 Timeline 8

1.6.1 Expected Consumer Surplus with $n = m = 5$ and $n_c(\phi) = 0$ 33

2.2.1 Setup with $n = 5$ firms 39

2.4.1 Consumer Surplus and Total Profits 54

List of Abbreviations

ACI	....	Attention Constrained Consequentialism
BR	....	Best Response
DM	....	Decision Maker
FIE	....	Full Information Equivalence
INRA	....	Independence of Never Relevant Acts
IS	....	Indirectly Selected over
LHS	....	Left Hand Side
MEU	....	Maxmin Expected Utility
RHS	....	Right Hand Side
SEU	....	Subjective Expected Utility
WARP	....	Weak Axiom of Revealed Preference

Chapter 1

Foundations for Optimal Inattention

1.1 Introduction

1.1.1 Objectives and Outline

Individuals often appear not to process all available information. This phenomenon, documented in both psychology and economics,[1] is usually attributed to agents' limited capacity to pay attention to information (Sims [2003]). When the available information exceeds their capacity, agents exhibit inattention, i.e. condition their choices on coarser information. This inattentiveness has significant economic consequences.[2]

This paper models agents who respond optimally to their limited attention. An *optimally inattentive* agent has a constraint that limits the information to which she can pay attention, and she chooses both her coarsening and her actions (or acts) conditional on that coarsening by maximizing a subjective expected utility preference relation. I axiomatically characterize the conditional choices of actions (acts) by a decision maker (DM) that are necessary and sufficient for her behavior to be seen *as if* it is the result of optimal inattention. These axioms clarify the model's implications

[1] See Pashler [1998] for a book-length treatment of attention in psychology. Economists have also empirically documented that agents, e.g. restaurant patrons (Luca [2011]), stock traders (DellaVigna and Pollet [2009]) and professional forecasters (Coibion and Gorodnichenko [2011a,b]), fail to process all available information.

[2] For instance, it can imply delayed response to shocks (Sims [1998, 2003]), sticky prices (Mackowiak and Wiederholt [2009]), under-diversification (Van Nieuweburgh and Veldkamp [2010]), sticky investment (Woodford [2008]), coordination failure (Hellwig and Veldkamp [2009]), specialization (Dow [1991]), exploitation (Rubinstein [1993]) and extreme price swings (Gul et al. [2011]).

for choice behavior and provide a choice-theoretic justification for it.

The modeler observes an objective state space and a partition describing the objective information. In contrast, the DM's tastes, her prior beliefs, her capacity for attention, and the information to which she pays attention (which I call her *subjective information* to distinguish it from the finer, objective information) are taken to be *unobservables* that must be inferred from choices. I assume a rich set of choice data, namely the DM's choices from each feasible set of acts and conditional on each state of the world.[3]

The rationale for assuming that the indicated range of behavior is observable is easily understood. First, with a narrower range of behavior, the model cannot be characterized. Choice out of a single feasible set cannot reveal much about underlying behavior for the reasons familiar from standard choice theory. Furthermore when the state space is unobservable and choice is observed conditional on a single state, earlier work by Van Zandt [1996] shows that optimal inattention has no testable implications.[4,5] Second, my setting allows analysis directly in terms of the economic object of interest – namely, the agent's chosen action, such as setting a price, selecting a bundle of goods, or deciding from which firm to purchase. Finally, this range of behavior permits unique identification of the unobservables, even though the DM's

[3]This data is an extension of that considered by the papers cited in Footnote 2, which study all conditional choices from a single feasible set. It is easily obtainable in a laboratory environment, and it could, in principle, be gathered from a real-world setting where both the realized state and the information received by the agent are independently and identically distributed across time.

[4]Van Zandt [1996] studies hidden information acquisition, which is readily reinterpreted as optimal inattention. Specifically, he takes as given any choice correspondence on a finite collection of alternatives. He shows that one can construct a state space and an information acquisition problem so that for every choice problem, the alternative selected in a fixed state matches the choice correspondence if the DM chooses information optimally.

[5]The model's implications when the state space and choice conditional on only one state are observable an open question. A partial answer is given by the axioms Monotonicity and ACI (below): conditional choices in a fixed state satisfy WARP when the problem contains only state-independent acts. However, they violate WARP (as well as many weaker properties implied by it) in general, and although identification of the utility index is possible, the attention constraint cannot be identified.

choices violate many of the well-understood properties that permit identification in other models, including the Weak Axiom of Revealed Preference (WARP).

The remainder of the paper proceeds as follows. In the next subsection, I use an example to illustrate my setting, the behavior of interest and how I achieve identification. Section 1.2 presents the model in detail. In Section 1.3, I formally describe the behavior of interest through five axioms. Theorems 1 and 2 show that these axioms characterize an optimally inattentive DM's choices. I also characterize two special cases: a DM who processes all information and a DM who processes the same information, regardless of the menu faced. In Section 1.4, Theorem 3 shows that the utility index, the attention constraint and, in many circumstances, the prior are uniquely identified by the agent's conditional choices.

In Section 1.5, Theorem 4 gives an intuitive, behavioral comparison equivalent to one optimally inattentive DM having a higher capacity for attention than another. I then argue that an optimally inattentive DM values information differently than a Bayesian DM because the former may not process all available information. Even if one information partition is *objectively* more valuable than another (in the sense of Blackwell [1953]), it may not be *subjectively* more valuable. That is, because of information overload a DM may reject an objectively more valuable information partition in favor of a coarser one. After generalizing my setting to allow the objective information to vary, Theorem 5 characterizes the subjective value of information to an optimally inattentive DM in terms of her choices.

In Section 1.6, I analyze a market where firms compete over optimally inattentive consumers. Intuition suggests that firms can exploit these consumers, and previous work (cf. Rubinstein [1993]) focuses on that aspect. Fixing prices, consumer surplus increases as capacity for attention increases. However, if strategic effects are taken into account, then *lower* consumer capacity for attention may lead to *higher* equilibrium consumer surplus. In fact, firms may *benefit* from facing *more attentive* consumers. The key difference from earlier work is that consumers perceive the price perfectly, but they are inattentive to information about the quality of the products. Intuitively, if consumers allocate their attention optimally, then a firm attracts attention only if it offers the consumer more surplus. This induces competition among firms who would not otherwise compete, which lowers prices and increases consumer surplus.

Section 1.7 concludes by discussing the relationship with other models of inattention. Proofs are collected in the remaining sections.

1.1.2 Example

Consider a benevolent doctor who treats patients suffering from a given disease. Glaxo, Merck, and Pfizer all produce pharmaceuticals that treat the disease, but the doctor knows that one of the three drugs will be strictly more effective than the other two. The one that works best for each patient is initially unknown, and the doctor can, in principle, determine it; for instance, by constructing a very detailed medical history. Uncertainty is modeled by the state space $\Omega = \{\gamma, \mu, \phi\}$, and the objective information by the partition

$$P = \{\{\gamma\}, \{\mu\}, \{\phi\}\}.$$

The state indicates whether the most effective drug is produced by Glaxo (γ), by Merck (μ) or by Pfizer (ϕ), and P indicates that the doctor can determine which state obtains.

Suppose there are two patients who are identical except that they have different

Table 1.1: Conditional choices

	γ	μ	ϕ
$c(\{g,m,f\}\|\cdot)$	$\{m\}$	$\{m\}$	$\{f\}$
$c(\{g,m\}\|\cdot)$	$\{g\}$	$\{m\}$	$\{m\}$

insurance plans: one's covers all three drugs, and the other's does not cover Pfizer's drug. Each patient is a choice problem, in which prescribing a drug corresponds to choosing an act (g, m and f represent prescribing the drugs produced by Glaxo, Merck and Pfizer respectively). The drug prescribed to each patient conditional on each state of the world is given by a *conditional choice correspondence*, a family of choice correspondences indexed by the state of the world. Table 1.1 lists the conditional choices of a doctor when facing $\{g, m, f\}$ (the problem associated with unrestricted insurance) and $\{g, m\}$ (the problem associated with restricted insurance).

Under the assumption that the doctor's choices result from optimal inattention, what can be inferred from them? One complication is that the doctor's choices violate WARP in state γ: she chooses g but not m from $\{g, m\}$ and chooses m but not g from $\{g, m, f\}$. Although WARP violations prevent identification of preference through the usual methods, the doctor's tastes, subjective information, and attention constraint can nevertheless be inferred from her choices. To begin with, the above choices reveal that the doctor cannot pay attention to all the objective information. If she did, then her choice when Glaxo's drug is effective from the first menu would reveal that she strictly prefers to prescribe it rather than to prescribe Merck's drug. Therefore, if Glaxo's drug is available in the larger problem and it is the most effective, then she should not prescribe Merck's. But because she chooses to prescribe the latter when facing $\{g, m, f\}$ in state γ, she does not pay attention to the objective information.

Since the doctor does not pay attention to the objective information, I turn to inferring her subjective information, i.e. the information to which she does pay attention. When facing $\{g, m\}$, she chooses differently conditional on γ than she does conditional on either μ or ϕ, so her subjective information must be at least as fine as $\{\{\gamma\}, \{\mu, \phi\}\}$. Moreover, it cannot be strictly finer because it would then be the objective information. Consequently, her subjective information is exactly $\{\{\gamma\}, \{\mu, \phi\}\}$ when facing $\{g, m\}$. Similarly, her subjective information must be $\{\{\phi\}, \{\gamma, \mu\}\}$ when facing $\{g, m, f\}$. Therefore, the doctor chooses as if she knows the answer to the question "Is Glaxo's drug the most effective?" when facing $\{g, m\}$ and "Is Pfizer's drug the most effective?" when facing $\{g, m, f\}$. With her subjective information known, her choices reveal her conditional preferences, which can then be aggregated to reveal her underlying unconditional preferences.

How can one tell if the doctor's choices have an optimal inattention representation? Theorems 1 and 2 show that a set of properties characterizes a doctor whose choices can be seen as if they result from optimal inattention. The doctor's choices do not violate any of these properties, so they are compatible with optimal inattention. However, consider a second doctor who chooses according to $c'(\cdot)$, where $c'(\cdot)$ is the same same as $c(\cdot)$ except $c'(\{g, m, f\}|\phi) = \{m\}$. Although both doctors select the same prescription from each choice problem in state γ, when the other conditional choices are considered, $c'(\cdot)$ cannot have an optimal inattention representation. To see this, note that the second doctor chooses Merck's drug when the patient has good insurance, regardless of the state of the world. As above, the modeler infers that the doctor knows whether Glaxo's drug is the most effective when facing $\{g, m\}$, so her choice of g in state γ reveals that she strictly prefers prescribing it to prescribing Merck's in that state. However, this implies that her choices from the smaller problem yield a better outcome in every state of the world than those from the larger problem, an impossibility if her subjective information is optimal when facing both problems.

1.2 Setup and Model

1.2.1 Setup

I adopt the following version of the classic Anscombe-Aumann setting. Uncertainty is captured by a set of states Ω and a set of events Σ, a $\sigma-$algebra of subsets of Ω. Consequences are elements of a separable metric space, Z. Let the set X consist of all finite-support probability measures on Z, endowed with the weak* topology. Objects of choice are acts, $\Sigma-$measurable simple (finite-ranged) functions $f : \Omega \to X$. Let $\mathcal{F}$ be the set of all acts, endowed with the topology of uniform convergence. Since $\mathcal{F}$ is metrizable, let $d(\cdot)$ be a compatible metric.[6]

The DM must choose an act from a compact set, i.e. her choice problem is a non-empty, compact subset of $\mathcal{F}$. Let $K(\mathcal{F})$ be the set of all choice problems, endowed with the Hausdorff topology generated by the metric $d(\cdot)$. She has access to objective information, represented by P, a finite partition of Ω.[7] I require every element of the partition be an element of Σ. Knowing the objective information allows the modeler to distinguish imperfect information from limited capacity for attention.[8]

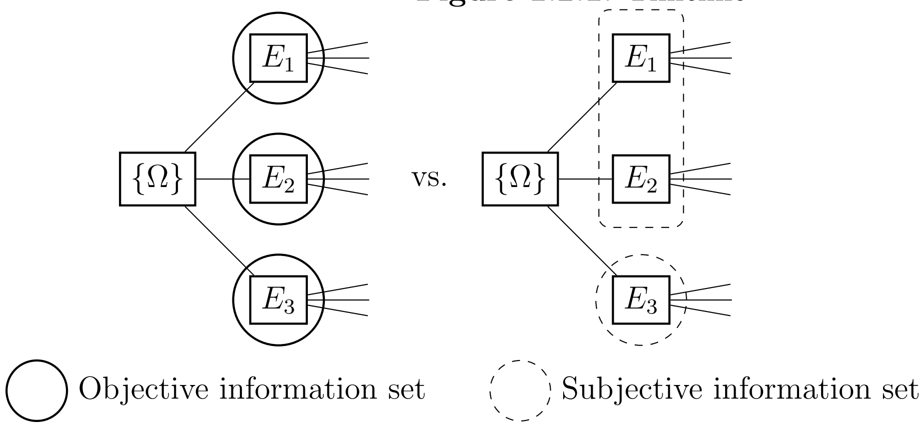
Choice results from a three-stage process. In stage 1, the state is realized but remains unknown to the DM. In stage 2, the DM chooses an act. Although in principle she observes the realized cell of P before making this choice, she instead acts as if she observes the realized cell of her subjective information. In stage 3, all uncertainty is

[6] X is metrizable by Theorem 15.12 of Aliprantis and Border [2006]. Let $\hat{d} : X \times X \to \mathbb{R}_+$ be a compatible metric. Since $\mathcal{F}$ has the topology of uniform convergence, $f_n \to f \iff \sup_\omega \hat{d}(f_n(\omega), f(\omega)) \to 0$. Therefore $d : \mathcal{F} \times \mathcal{F} \to \mathbb{R}_+$ given by $d(f, g) = \sup_\omega \hat{d}(f(\omega), g(\omega))$ is a compatible metric on $\mathcal{F}$; since $\mathcal{F}$ contains only simple acts, the supremum is attained and $d(f, g) < \infty$ for any $f, g \in \mathcal{F}$.

[7] With minor additional assumptions, P can be taken to be countable rather than finite.

[8] Because the modeler knows the objectively available information, inattention can be distinguished from imperfect information. For instance, if the modeler observes that the DM never distinguishes between states ω_1 and ω_2, then this is interpreted as inattention only if the objective information distinguishes ω_1 from ω_2.

Figure 1.2.1: Timeline



resolved and the DM gets the consequence specified by her chosen act and the realized state. The tree on the left in Figure 1.2.1 illustrates the timing. If the DM exhibits inattention, she acts as if facing a different tree than the objective one; for instance, the one on the right in Figure 1.2.1.

The modeler observes the DM's choices in stage 2 (or later) and the realization of the objective information, but does not observe the DM's subjective information. The choice data generates a conditional choice correspondence $c(\cdot)$, such that the DM is willing to choose the acts in $c(B|\omega)$ from the problem B when the state is ω. Formally, this is a set-valued, P-measurable function $c : K(\mathcal{F}) \times \Omega \to K(\mathcal{F})$ with $c(B|\omega) \subset B$ for all $B \in K(\mathcal{F})$ and all $\omega \in \Omega$.

I adopt the following notation throughout. Identify X with the subset of acts that do not depend on the state, i.e. $x \in X$ corresponds to the act $x \in \mathcal{F}$ such that $x(\omega) = x \ \forall \omega \in \Omega$, and let $K(X)$ be the set of compact, non-empty subsets of X, noting that $K(X) \subset K(\mathcal{F})$. For any partitions Q and Q', write $Q \gg Q'$ if Q is finer than Q'. For any acts $f, g \in \mathcal{F}$ and any event $E \in \Sigma$, define fEg to be the act that yields $f(\omega)$ if $\omega \in E$ and $g(\omega)$ if $\omega \notin E$. For any $\alpha \in [0,1]$ and any two $f, g \in \mathcal{F}$ let $\alpha f + (1-\alpha) g \in \mathcal{F}$ be the state-wise mixture of f and g, or the act taking the value in state ω of $\alpha f(\omega) + (1-\alpha)g(\omega)$, defined by the usual mixture operation on lotteries. For any $A, B \in K(\mathcal{F})$ and $\alpha \in [0,1]$, let $\alpha A + (1-\alpha)B \in K(\mathcal{F})$ be $\{\alpha a + (1-\alpha)b | a \in A, b \in B\}$. Any act in $\alpha A + (1-\alpha)B$ is an α mixture of an act in A with a $(1-\alpha)$ mixture of an act in B.

1.2.2 Model

An optimally inattentive agent is a tuple $(u(\cdot), \pi(\cdot), \mathbb{P}^*, \hat{P}(\cdot))$, where:

- $u : X \to \mathbb{R}$ is continuous and affine,
- $\pi : \Sigma \to [0,1]$ is finitely-additive and $\pi(E) > 0$ for every $E \in P$,
- $\mathbb{P}^* \subset \{Q : P \gg Q\}$ has the property that if $Q \in \mathbb{P}^*$ and $Q \gg Q'$, then $Q' \in \mathbb{P}^*$, and
- $\hat{P} : K(\mathcal{F}) \to \mathbb{P}^*$.

The *utility index* $u(\cdot)$ and *prior* $\pi(\cdot)$ have familiar interpretations. Neither varies with the problem, so an optimally inattentive DM has stable tastes and beliefs. I focus on interpreting the two new objects, the *attention constraint* $\mathbb{P}^*$ and the *attention rule* $\hat{P}(\cdot)$. The former describes to what the DM *can* pay attention, while the latter describes to what she *does* pay attention. The attention constraint $\mathbb{P}^*$ is a set of partitions, all of which are coarser than the objective information. If $Q \in \mathbb{P}^*$, then the DM has the capacity to pay attention to Q. I assume that $\mathbb{P}^*$ satisfies free disposal of information, in the sense that whenever she can pay attention Q, she can also pay attention to any Q' that is coarser than Q. Depending on the problem, the DM may have different subjective information, given by the attention rule $\hat{P}(\cdot)$. That is, $\hat{P}(B)$ is her subjective information when facing the problem B.

Definition 1. *A conditional choice correspondence* $c(\cdot)$ *has an* optimal inattention representation *if there exists an optimally inattentive agent so that for every problem* B*,*

$$\hat{P}(B) \in \arg\max_{Q\in\mathbb{P}^*} \sum_{E\in Q} \pi(E)[\max_{f\in B} \int u \circ f d\pi(\cdot|E)], \tag{1.2.1}$$

and for every problem B *and state* ω*,*

$$c(B|\omega) = \arg\max_{f\in B} \int u \circ f d\pi(\cdot|\hat{P}(B)(\omega)). \tag{1.2.2}$$

The DM's choices have an optimal inattention representation if they satisfy two properties. Equation (1.2.1) requires that her subjective information gives at least as high expected utility as any other partition in $\mathbb{P}^*$, i.e. it is chosen optimally. Equation (1.2.2) requires that the DM's choice from B in state ω maximizes expected utility conditional on the realized cell of her subjective information.

An optimally inattentive DM considers all available acts. In contrast, Masatlioglu et al. [2012] studies an agent who does not pay attention to the entire set of available actions. Although both models are motivated by the same underlying mechanism, neither nests the other: there are choices compatible with optimal inattention but not inattention to alternatives and vice versa.[9] While DMs conforming to either model may violate WARP, the reason for such violations is different.[10] In fact, an optimally inattentive DM may exhibit inattention yet satisfy WARP (Corollary 2).

One special case of optimal inattention is a DM who processes all available information. I call such a DM Bayesian, and say that $c(\cdot)$ *has a Bayesian representation* if

$$c(B|\omega) = \arg\max_{f\in B} \int u \circ f d\pi(\cdot|P(\omega))$$

[9]See Section A.3.6 and A.3.7.

[10]In Masatlioglu et al. [2012], removing unchosen alternatives may affect the options considered by the DM; in this paper, removing alternatives not chosen in a given state may alter the information processed.

for every B and ω. In the model, this corresponds to an optimally inattentive agent with $\mathbb{P}^* = \{Q : Q \ll P\}$ and $\hat{P}(B) = P$ for every problem B.

Another special case is a DM who always pay attention to the same information (which may differ from the objective information), regardless of the problem faced. I say that such a DM has fixed attention, and that $c(\cdot)$ has *fixed attention representation* if there is a partition Q satisfying $P \gg Q$ so that

$$c(B|\omega) = \arg\max_{f \in B} \int u \circ f d\pi(\cdot|Q(\omega))$$

for every B and ω. In the model, this corresponds to an optimally inattentive agent with $\mathbb{P}^* = \{Q' : Q' \ll Q\}$ and $\hat{P}(B) = Q$ for every problem B.

In addition to the above special cases, the model admits many others considered in the literature. For instance, $\mathbb{P}^*$ could equal the set of all partitions that are both coarser than P and have at most $\kappa \geq 1$ elements (Gul et al. [2011]). Alternatively, $\mathbb{P}^*$ could equal the set of all partitions that have mutual information with respect to P less than κ (similar to Sims [2003]).[11]

1.3 Foundations

1.3.1 Axioms

I impose the following axioms. The quantifier "for all $f, g \in \mathcal{F}$, $A, B \in K(\mathcal{F})$, $\omega \in \Omega$ and $\alpha \in (0, 1]$" is suppressed throughout.

A DM satisfies *WARP*, sometimes referred to as Independence of Irrelevant Acts, if for any $A \subset B$, whenever $c(B|\omega) \cap A \neq \emptyset$ it follows that $c(A|\omega) = c(B|\omega) \cap A$. If an inattentive DM's choices from problems A and B are conditioned on the same

[11] The mutual information is a measure of the information provided about the realization of one random variable by another. It corresponds to the reduction in entropy and is used by the rational inattention literature.

subjective information, then her choice in each state maximizes the same conditional preference relation, so these choices do not violate WARP. Therefore, if she violates it, then her choices from A and B must be conditioned on different subjective information. The first axiom, *Independence of Never Relevant Acts* or *INRA*, gives one situation where the DM should not violate WARP.

Axiom 1. *(INRA) If* $A \subset B$ *and* $c(B|\omega') \cap A \neq \emptyset$ *for every state* ω'*, then*

$$c(A|\omega) = c(B|\omega) \cap A.$$

Within the context of Section 1.1.2, INRA says that if two patients differ only in that one's plan drops the drug h but the doctor *never* prescribes h to the patient with better insurance, then she prescribes the same drug to both patients. To interpret the axiom, consider a problem B and a "never relevant" act f (i.e. $f \notin c(B|\omega')$ for all ω'), and let $A = B\backslash\{f\}$.[12] Suppose that her choices from B are conditioned on the subjective information Q. Because she never chooses f from B, the benefit of paying attention to Q when facing A is the same as it is when facing B. If Q is *optimal* when facing B, then Q is *still optimal* when facing A. Therefore, the DM should have the same subjective information when facing B as when facing A, so her choices from A and B should not violate WARP. More generally, the statement $c(B|\omega') \cap A \neq \emptyset$ for every state ω' implies that the entire set of acts that are in B but not in A is "never relevant" and removing them would not decrease the benefit of her subjective information when facing B. As above, if her subjective information is optimal when facing B, then it is still optimal when facing A. Consequently, the DM's choices should not violate WARP.

[12]Whenever B is finite, INRA is equivalent to "if $c(B|\omega') \neq \{f\}$ for all ω', then $c(B|\omega)\backslash\{f\} = c(B\backslash\{f\}|\omega)$."

In the present context, a DM satisfies *Independence* if

$$f \in c(A|\omega) \, and \, g \in c(B|\omega) \iff \alpha f + (1-\alpha)g \in c(\alpha A + (1-\alpha)B|\omega).$$

That is, if the DM chooses f over each h in A and g over each h' in B, then she chooses $\alpha f + (1-\alpha)g$ over each $\alpha h + (1-\alpha)h'$ in $\alpha A + (1-\alpha)B$.[13] If an optimally inattentive DM pays attention to the same information when facing the problems A, B and $\alpha A + (1-\alpha)B$, then her choice in each state maximizes the same conditional preference relation. Because her conditional preferences are expected utility, her choices do not violate Independence. This implies that whenever the DM violates this property for A, B and $\alpha A + (1-\alpha)B$, she must not pay attention to the same information when facing all three problems. The second axiom, *Attention Constrained Independence* or *ACI*, gives one situation where the DM should not violate Independence.

Axiom 2. *(ACI)* $f \in c(B|\omega)$ *if and only if* $\alpha f + (1-\alpha)g \in c(\alpha B + (1-\alpha)\{g\}|\omega)$.

In my example, this says that if there is a state-independent chance that the patient will take some drug h regardless of what the doctor actually prescribes, then her choice of prescription is unaffected by both the identity of h and the magnitude of that chance. To interpret the axiom, fix problems B and $\{g\}$. Because $\{g\}$ is a singleton, the DM makes the same choice from it no matter what her subjective information is. Therefore, the relationship between the benefits of any two subjective information partitions is the same for the problem B as it is for the problem $\alpha B + (1-\alpha)\{g\}$.[14] If paying attention to Q is *optimal* when facing B, then paying attention

[13] This follows from the standard formulation of Independence for a binary relation: $f \succeq g \iff \alpha f + (1-\alpha)h \succeq \alpha g + (1-\alpha)h$.

[14] One can think of $\alpha B + (1-\alpha)\{g\}$ as flipping a (possibly biased) coin, choosing from B if the coin comes up heads and otherwise choosing from $\{g\}$, where the DM must choose her subjective information before observing the outcome of the coin-flip. Since information only has value if the coin comes up heads, a partition is optimal when facing $\alpha B + (1-\alpha)\{g\}$ if and only if it would be optimal when facing B for sure.

to Q is *also optimal* when facing $\alpha B + (1-\alpha)\{g\}$. Consequently, an optimally inattentive DM conditions her choices on the same subjective information when facing $\alpha B + (1-\alpha)\{g\}$ as she does when facing B. Because her conditional preferences satisfy Independence, she chooses the mixture of her choices from B with g from $\alpha B + (1-\alpha)\{g\}$.

The next axiom adapts the standard Monotonicity axiom to the present setting. It also implies that tastes are state independent. For any lotteries x and y, say that x is *revealed* (resp. *strictly*) *preferred* to y if $x \in c(\{x,y\}|\omega)$ (resp. and $y \notin c(\{x,y\}|\omega)$) for some ω.[15]

Axiom 3. *(Monotonicity) (i) If $A \in K(X)$, then $c(A|\omega) = c(A|\omega')$ for all $\omega' \in \Omega$. (ii) If $f, g \in B$ and $f(\omega')$ is revealed preferred to $g(\omega')$ for every $\omega' \in \Omega$, then $g \in c(B|\omega) \implies f \in c(B|\omega)$. Moreover, if $f(\omega')$ is revealed strictly preferred to $g(\omega')$ for each $\omega' \in P(\omega)$, then $g \notin c(B|\omega)$.*

In my example, this says that the doctor cares only about the realized consequence of her choice, and if one drug gives a better consequence in every state than another, then she never prescribes the inferior drug. For interpretation, consider $B = \{x, y\}$ where x and y are lotteries. If the DM's tastes are state independent and she chooses x over y in state ω, then she also chooses x over y in state ω'. This reveals that the DM prefers x to y, i.e. considers x to be a better consequence than y. Now, consider acts f and g so that f yields a better consequence than g in every state of the world. Even if the DM received information revealing that the state on which g gives the best consequence would occur for sure, she would still be willing to choose f over g. Consequently, she never chooses only g when f is available. In addition, if f yields a strictly better consequence than g in every state in $P(\omega)$, then the DM does not

[15] Recall that $K(X)$ is the set of problems that contain only lotteries.

choose g. Thus, Monotonicity limits the scope of inattention; an inattentive DM will never pick a dominated act.

Another common property satisfied by most models of choice under uncertainty is *Consequentialism:* if $f, g \in B$ and $f(\omega') = g(\omega')$ for all $\omega' \in P(\omega)$, then

$$f \in c(B|\omega) \iff g \in c(B|\omega).$$

A DM who satisfies Consequentialism respects the objective information, in the sense that whenever two acts give the same outcome on every objectively possible state, then one of the acts is chosen if and only if the other is. A DM whose subjective information differs from the objective information will violate this property. The next axiom, Subjective Consequentialism, weakens Consequentialism to take this into account.

Axiom 4. *(Subjective Consequentialism) If $f, g \in B$ and $\forall \omega' [f(\omega') \neq g(\omega') \implies c(B|\omega') \neq c(B|\omega)]$, then $f \in c(B|\omega) \iff g \in c(B|\omega)$.*

Subjective Consequentialism implies that choice between any two acts is unaffected by their outcomes in states that the DM knows did not occur. To see this, fix B, f, and g as above, and suppose that the DM faces the problem B and that the realized state is ω. Whenever ω and ω' are in the same cell of her subjective information when facing B, the DM's choices in those states maximize the same conditional preference relation, so $c(B|\omega) = c(B|\omega')$. Consequently, if $c(B|\omega') \neq c(B|\omega)$, i.e. the DM makes different choices in states ω and ω', then these two states must be in different cells of her subjective information. By hypothesis, if f and g give a different consequence in state ω', then $c(B|\omega') \neq c(B|\omega)$, so the DM must know that ω' did not occur. Therefore, the DM knows that she receives the same consequenence from choosing either f or g, so she chooses f if and only if she chooses g.[16]

My final axiom is a technical condition ensuring the continuity of the underlying preference relation. Complicating its statement is that the DM's choices from different menus may be conditioned on different information. Consequently, her choices may appear discontinuous to the modeler, and the axiom must take into account that the underlying preference is revealed by choices that are not conditioned on the same subjective information. To state the axiom, I need two preliminary definitions. First, say that f *dominates* g if f is chosen from $\{f, g\}$ in every state of the world. If the DM has optimal inattention and f dominates g, then f must be (weakly) preferred to g conditional on every cell of every feasible subjective information partition. The second definition is:

Definition 2. The acts in A are indirectly selected over the acts in B, *written $A\ IS\ B$, if there are problems $B_1, ..., B_n \in K(\mathcal{F})$ so that $B_1 = A$ and $B_n = B$ and for each $i \in \{1, ..., n-1\}$ and every ω, $c(B_{i+1}|\omega) \cap B_i \neq \emptyset$.*

Suppose that the DM faces B and chooses an act in A regardless of the state of the world. Since her choices from B are available in A, her set of choices from A is selected over her choices from B. Moreover, if she chooses an act from B in every state of the world when facing C, then her set of choices from B is selected over any choices in C. Since the acts in A are selected over the acts in B that are in turn selected over the acts in C, the acts in A are indirectly selected over the acts in C. Write $A\ \overline{IS}\ B$ if there are sequences $(A_n)_{n=1}^{\infty}$ and $(B_n)_{n=1}^{\infty}$ so that $A_n \to A$, $B_n \to B$ and $A_n\ IS\ B_n$ for all n, i.e. $\overline{IS}$ is the sequential closure of IS.

The final axiom, Continuity, requires that each $c(\cdot|\omega)$ satisfies a weak continuity condition and that sequences of indirect selections do not contradict domination.

[16] Consequentialism implies Subjective Consequentialism: since $c(B|\cdot)$ is P-measurable, $P(\omega) \subset \{\omega'' : c(B|\omega'') = c(B|\omega)\}$ for every B and ω.

Axiom 5. *(Continuity) For any* $\{B_n\}_{n=1}^{\infty} \subset K(\mathcal{F})$*:*
(i) If $f_n \in c(B_n|\omega)$ *and*

$$\{\omega' : c(B|\omega') = c(B|\omega)\} = \{\omega' : c(B_n|\omega') = c(B_n|\omega)\}$$

for every $n \in \mathbb{N}$*, then* $B_n \to B$ *and* $f_n \to f$ *imply that* $f \in c(B|\omega)$*.*
(ii) If $\{f\}\ \overline{IS}\ \{g\}$ *and* g *dominates* f*, then* f *dominates* g *as well.*

The first condition of Continuity is a restriction of upper hemi-continuity.[17] It requires that this property holds only if the DM reveals that her choice is conditioned on the same information along the sequence and as it is at the limit. Both parts of Continuity are implied by combining WARP and upper hemi-continuity.

To interpret the second condition of the axiom, note that INRA suggests that the DM's set of choices from problem A is better than her set of choices from problem B whenever she chooses an act in A when facing B conditional on every state of the world. This direct ranking is incomplete but can be extended using finite sequences of choices to allow for indirect comparisons as well. These indirect comparisons are captured when the acts in A are indirectly selected over the acts in B. Because this ranks many more sets of choices, indirect selections are important for characterizing optimal inattention. Continuity insures a minimal consistency between these indirect comparisons and her direct comparisons. Specifically, suppose that f dominates g and g does not dominate f. Continuity implies that if B is sufficiently close to $\{g\}$ and A is sufficiently close to $\{f\}$, then the DM does not indirectly select the acts in B over the acts in A.

[17] This follows from Aliprantis and Border [2006, Cor 17.17].

1.3.2 Characterization Result

I can now state the main result: if the DM's choices satisfy the five axioms above, then she acts as if she has optimal inattention.

Theorem 1. *If $c(\cdot)$ satisfies INRA, ACI, Monotonicity, Subjective Consequentialism and Continuity, then $c(\cdot)$ has an optimal inattention representation.*

Theorem 1 shows that the above axioms are sufficient for the DM to have optimal inattention. For a discussion of the key ideas in its proof, see Secion A.1. Necessity is more complicated because I have not restricted attention to well-behaved tie-breaking rules. Consider two conditional choice correspondences, $c(\cdot)$ and $c'(\cdot)$, that both have optimal inattention representations with the same prior, utility index, and attention constraint. When their attention constraint is not a singleton, it is possible that the former has subjective information Q when facing the problem B, while the latter has subjective information Q' when facing B. This arises when a problem has multiple optimal subjective information partitions (i.e. the right hand side of (1.2.1) is not a singleton) and the DM must break ties between them. If she breaks these ties non-systematically, then the DM may violate ACI or INRA.[18] Though the axioms become necessary if I impose some conditions on tie-breaking when defining the model, Theorem 2 shows that the set of problems for which an optimally inattentive DM fails to satisfy either INRA or ACI is non-generic even without any such conditions.[19]

Theorem 2. *If $c(\cdot)$ has an optimal inattention representation, then $c(\cdot)$ satisfies Monotonicity, Subjective Consequentialism and Continuity. Moreover, there is a con-*

[18] A similar issue exists for random expected utility (Gul and Pesendorfer [2006]) with a finite state space. If ties are broken using a "regular" random expected utility function, then choices satisfy linearity, but if ties are broken differently, then linearity may fail.

[19] Say that an optimal inattention representation is *regular* if for any $A, B \in K(\mathcal{F})$ and $g \in \mathcal{F}$, $\hat{P}(B) = \hat{P}(\alpha B + (1-\alpha)\{g\})$ and $\arg\max_{f \in A} \mathbb{E}_\pi[u \circ f | \hat{P}(A)(\omega)] = \arg\max_{f \in A} \mathbb{E}_\pi[u \circ f | \hat{P}(B)(\omega)]$ for all ω whenever $c(B|\omega) \cap A \neq \emptyset$ for all ω. Given Theorems 1 and 2, one can verify that $c(\cdot)$ has regular optimal inattention if and only if $c(\cdot)$ satisfies all six axioms.

ditional choice correspondence $c'(\cdot)$ *satisfying INRA, ACI, Monotonicity, Subjective Consequentialism and Continuity as well as an open, dense* $K \subset K(\mathcal{F})$ *so that*

(i) $c(\cdot)$ *and* $c'(\cdot)$ *have optimal inattention representations parametrized by* $(u(\cdot), \pi(\cdot), \hat{P}(\cdot), \mathbb{P}^*)$ *and* $(u(\cdot), \pi(\cdot), \hat{Q}(\cdot), \mathbb{P}^*)$*, respectively, and*

(ii) $c(B|\omega) = c'(B|\omega)$ *for every* $\omega \in \Omega$ *and* $B \in K$.

Theorem 2 implies that INRA and ACI are generically necessary. This is because the set of problems for which ties can occur is "small." Consequently, INRA and ACI capture the economic content of optimal inattention. Though not strictly necessary, for any given prior, utility index and attention constraint, there are always attention rules that satisfy INRA and ACI.

1.3.3 Special cases

To understand the role of the axioms in the characterization, I characterize the two special cases of optimal inattention mentioned at the start of this section, the Bayesian model and the fixed attention model.

Corollary 1. $c(\cdot)$ *satisfies Consequentialism in addition to INRA, ACI, Monotonicity, and Continuity if and only if* $c(\cdot)$ *has a Bayesian representation.*

Intuitively, Consequentialism requires that the DM respects the objective information structure. For an optimally inattentive DM, this implies that she processes all information and chooses the act that maximizes expected utility. Since Consequentialism implies Subjective Consequentialism, $c(\cdot)$ has an optimal inattention representation and must have a Bayesian representation.

Corollary 2. *The following are equivalent:*

(i) $c(\cdot)$ *satisfies Independence in addition to INRA, Monotonicity, Subjective Consequentialism and Continuity.*

(ii) $c(\cdot)$ *satisfies WARP in addition to ACI, Monotonicity, Subjective Consequentialism and Continuity.*

(iii) $c(\cdot)$ has a fixed attention representation.

It immediately follows that WARP and Independence are equivalent for an optimally inattentive DM. The intuition behind Corollary 2 is that an optimally inattentive DM's choices from A and B violate Independence or WARP only if her subjective information differs at A, B or $\alpha A + (1-\alpha)B$. If her subjective information never changes, then she never violates either condition. Consequently, she has fixed attention if she satisfies either WARP or Independence.

1.3.4 Counter-examples

To help understand the role of the axioms, I provide a series of counterexamples showing what may go wrong if one or more of the axioms are not satisfied. An alternative model of particular interest is the *inattention* model. An inattentive DM maximizes expected utility conditional on her subjective information, but her subjective information is not necessarily optimal. Although she has stable tastes and beliefs, the information to which she pays attention varies with the problem in a general manner. Formally, $c(\cdot)$ has an *inattention representation* if Equation (1.2.2) holds for all problems B and states ω but the source of $\hat{P}(\cdot)$ is left unspecified.

Proposition 1. *If $c(\cdot)$ has an inattention representation, then $c(\cdot)$ satisfies Monotonicity, Subjective Consequentialism and Continuity (i).*

In particular, an inattentive DM's choices may violate INRA, ACI or Continuity (ii). Consequently, these three axioms reflect the optimality of her subjective information. They capture her reaction to her attention constraint but not that she exhibits inattention in the first place.[20]

[20]A characterization of the inattention model is available as supplementary material.

ACI reflects that the DM has an attention constraint. Consider the alternative model is *costly attention*. In this case, rather than being subject to a constraint on the information to which she can attend, the DM incurs a cost if she pays attention to a given partition. A function $\rho : \{Q : Q \ll P\} \to [0, \infty]$ is a *cost function* if $\rho(\{\Omega\}) = 0$ and $Q \gg Q'$ implies $\rho(Q) \geq \rho(Q')$. Formally, $c(\cdot)$ has a *costly attention representation* if there is an optimally inattentive agent and a cost function $\rho(\cdot)$ so that

$$\hat{P}(B) \in \arg\max_{Q} \sum_{E \in Q} \pi(E)[\max_{f \in B} \int u \circ f d\pi(\cdot|E)] - \rho(Q) \tag{1.3.1}$$

for every problem B and Equation (1.2.2) holds for every problem B and state ω. Given appropriate tie-breaking, this model satisfies all of the axioms except ACI. In fact, it satisfies the following weaker version of ACI:

$$\alpha f + (1-\alpha)g \in c(\alpha B + (1-\alpha)\{g\}|\omega) \iff \alpha f + (1-\alpha)h \in c(\alpha B + (1-\alpha)\{h\}|\omega)$$

for a fixed $\alpha \in (0, 1]$.[21]

INRA reflects that the DM's subjective information is optimal. If her subjective information cannot be represented as maximizing behavior, then the DM's choices violate INRA. For instance, suppose that Equation (1.2.1) holds when a minimum replaces each maxima and Equation (1.2.2) is satisfied. In this case, the DM's choices violate INRA but satisfy the remaining axioms.[22]

I now turn to Monotonicity, Subjective Consequentialism, and Continuity. If the utility index depends on the state, then the DM satisfies all axioms except Monotonicity. Fix a set of full support probability measures $\{\pi_\omega\}_{\omega \in \Omega}$ that containing at

[21] A full characterization of this model is work in progress.

[22] It is possible that Continuity (ii) is also violated. This is not surprising the interpretation of Continuity (ii) relies on INRA.

least two distinct measures. If

$$c(B|\omega) = \arg\max_{f\in B} \int u \circ f d\pi_\omega,$$

then the DM violates Subjective Consequentialism but satisfies the other axioms. The counter-example for Continuity involves lexicographic preferences. I defer it to Section A.3, which also contains details on the above counter-examples.

1.4 Identification

To interpret a model, it is important to understand how precisely the parameters are identified, i.e. what are the uniqueness properties of the representation. For instance, if certain parameters of the representation are not unique, then doing comparative statics is impossible. How much identification is possible within the current framework, given that the modeler does not directly observe ex ante preference, subjective information, or capacity for attention and that the DM's choices violate WARP? Of the four components that characterize an optimally inattentive agent, Theorem 3 shows that three are suitably unique, and in many cases of interest, all four are unique.

Before stating Theorem 3, one issue deserves elaboration. In general, many attention rules represent the same choice correspondence; for instance, if B contains only constant acts, then $\hat{P}(B)$ could be any partition. However, there is a unique *canonical* attention rule, given by the coarsest attention rule that represents choice. That is, $\hat{P}(\cdot)$ is canonical if $(u(\cdot), \pi(\cdot), \mathbb{P}^*, \hat{P}(\cdot))$ represents $c(\cdot)$ and for any $(u'(\cdot), \pi'(\cdot), \mathbb{P}'^*, \hat{Q}(\cdot))$ that also represents $c(\cdot)$, $\hat{Q}(B) \gg \hat{P}(B)$ for every B. The canonical attention rule is the partition

$$\hat{\hat{P}}(B) = \{\{\omega' : c(B|\omega') = c(B|\omega)\} : \omega \in \Omega\}. \tag{1.4.1}$$

To interpret this normalization, if paying attention to a finer partition has an arbitrarily small but positive cost, then the DM would always choose a canonical attention rule – she can make the same conditional choice in every state but avoid paying the cost. On the one hand, her subjective information may be finer than that given by her canonical attention rule. That is, she may pay attention to a partition strictly finer than it but make the same conditional choices on at least two cells. On the other hand, her subjective information must be at least as fine as it because otherwise, she does not pay attention to information that distinguishes two state on which she makes different conditional choices.

Theorem 3. *If $c(\cdot)$ is non-degenerate and represented by the optimally inattentive agents $(u_1(\cdot), \pi_1(\cdot), \mathbb{P}_1^*, \hat{P}_1(\cdot))$ and $(u_2(\cdot), \pi_2(\cdot), \mathbb{P}_2^*, \hat{P}_2(\cdot))$, then:*
(i) $u_1(\cdot)$ is a positive affine transformation of $u_2(\cdot)$,
(ii) there is a partition $\mathbb{Q} \ll P$ so that $\pi_1(\cdot|E) = \pi_2(\cdot|E)$ for any $E \in \mathbb{Q}$,
(iii) $\mathbb{P}_1^* = \mathbb{P}_2^*$, and
(iv) $\hat{P}_1(\cdot) = \hat{P}_2(\cdot)$ whenever $\hat{P}_1(\cdot)$ and $\hat{P}_2(\cdot)$ are both canonical.

Theorem 3 establishes that an optimally inattentive DM's utility index, attention rule and attention constraint are unique. However, her prior probability measure may not be. Because ex ante preference is unobserved, the DM's choices only reveal the likelihood of events that are relevant for choosing either her act or her subjective information. For this reason, the modeler can uniquely identify the DM's prior only up to conditioning on a partition $\mathbb{Q}$, which will be characterized below. Note that a coarser $\mathbb{Q}$ implies more precise identification, in the sense that fewer probability measures represent choices for given "true" prior beliefs. Since $P \gg \mathbb{Q}$, the prior of an optimally inattentive DM is identified at least as precisely as that of a Bayesian DM.

In many cases, $\mathbb{Q} = \{\Omega\}$, and the prior is uniquely identified. For instance,

$\mathbb{Q} = \{\Omega\}$ whenever $\mathbb{P}^*$ is all partitions coarser than P with at most κ elements and κ is less than the number of cells in P. One notable case where uniqueness does not obtain is when the DM is Bayesian; in this case, the coarsest $\mathbb{Q}$ is equal to P.

I now turn to characterizing the coarsest $\mathbb{Q}$. This partition is the set of "minimal isolatable events." Intuitively, E is an isolatable event if any choice problem can be partitioned into two distinct problems – one that depends on E and one that depends on E^c – so that either of the two can be varied without changing the DM's conditional choices of acts. The relative likelihood of events contained in different isolatable events is not relevant for her choices.

To define an isolatable event formally, first let

$$B_{E,x}B' = \{fEx : f \in B\} \cup \{gE^c x : g \in B'\}$$

for any problems B, B' and any consequence x. The problem $B_{E,x}B'$ is formed by combining the two problems B and B' into a single problem containing modifications of the acts in B so they differ from each other only on E and of the acts in B' so they differ from each other only on E^c.

Definition 3. *E is an* isolatable event *if for any B so that the right hand side of Equation (1.2.1) is a singleton and any B', whenever $z \notin c(\{g(\omega'), z\}|\omega')$ for any $g \in B \cup B'$ and ω', both*

$$f \in c(B|\omega) \implies fEz \in c(B_{E,z}B'|\omega)$$

for all $\omega \in E$ and

$$f \in c(B|\omega) \implies fE^c z \in c(B_{E^c,z}B'|\omega)$$

for all $\omega \in E^c$ hold.[23]

[23]One can define this condition from choices without referring to $\mathbb{P}^*$, but it is simpler to define it this way.

That is, whenever z is a bad enough outcome, the modifications of the DM's choices from B are still chosen from the problems $B_{E,z}B'$ and $B_{E^c,z}B'$, regardless of the contents of B'. Say that an isolatable event is *minimal* if it does not contain any other non-empty isolatable events. Note that Ω is always an isolatable event, but may not be minimal.

Lemma 1. *If $\mathbb{Q}$ is the coarsest partition satisfying (ii), then E is a minimal isolatable event if and only if $E \in \mathbb{Q}$.*

If the DM is Bayesian, then each $E \in P$ is a minimal isolatable event. However in the introductory example, the only isolatable event is Ω. When the DM is Bayesian, each element of the objective information is a minimal isolatable event, so it follows that her prior is only identified up to its conditional probabilities on every element of P.

1.5 Comparative Attention and the Value of Information

There are two comparatives of interest. The first is to compare two distinct DMs that have the same information. The second is to compare a single DM with two different objective information partitions.

Consider DM1 and DM2 with conditional choice correspondences given by $c(\cdot)$ and $c'(\cdot)$, respectively. Denote by $\hat{\hat{P}}_c(B)$ the canonical subjective information of $c(\cdot)$ when facing B and by $\hat{\hat{P}}_{c'}(B)$ the canonical subjective information of $c'(\cdot)$ when facing B using Equation (1.4.1). Note that these are defined from choices alone.

Definition 4. $c(\cdot)$ is more attentive than $c'(\cdot)$ *if for any B, there exists a B' so that*

$$\hat{\hat{P}}_c(B') = \hat{\hat{P}}_{c'}(B).$$

To understand this comparison, suppose that the modeler observes that $c'(\cdot)$ pays attention to Q when facing B, i.e. her canonical subjective information is Q. If $c(\cdot)$ is more attentive than $c'(\cdot)$, then there is a B' so that the modeler observes that $c(\cdot)$ pays attention to Q when facing B'. That is, whenever the modeler observes DM2 using information Q, the modeler also observes DM1 using Q, though possibly when facing a different choice problem. Theorem 4 shows that this comparison is equivalent to comparing their attention constraints when both DMs have optimal inattention.

Theorem 4. *If $c(\cdot)$ and $c'(\cdot)$ are non-degenerate and have optimal inattention representations, parametrized by $(u_c, \pi_c, \mathbb{P}^*_c, \hat{P}_c)$ and $(u_{c'}, \pi_{c'}, \mathbb{P}^*_{c'}, \hat{P}_{c'})$ respectively, then:*

$$c(\cdot)\, \textit{is more attentive than}\, c'(\cdot) \iff \mathbb{P}^*_{c'} \subset \mathbb{P}^*_c.$$

That is, $c(\cdot)$ is more attentive than $c'(\cdot)$ if and only if her attention constraint is larger. Note that their representations may have different priors and tastes. Therefore, $\mathbb{P}^*$ reflects the DM's capacity for attention: whenever $\mathbb{P}^*_{c'} \subset \mathbb{P}^*_c$, DM1 has a higher capacity for attention than DM2.

Another interesting comparative is how a DM reacts to changes in the available information. Up to now, I have considered a fixed information structure. I modify the primitives in order to allow the objective information to vary. This generalization allows me to consider the value of information to an optimally inattentive DM. Consider a set of finite partitions, $\mathcal{P}$, that represent the possible objective information. Suppose the DM's choices given objective information P are represented by a conditional choice correspondence indexed by $P \in \mathcal{P}$, i.e. $c_P : K(\mathcal{F}) \times \Omega \to K(\mathcal{F})$ where $c_P(B|\omega) \subset B$ and $c_P(B|\cdot)$ is P-measurable. I assume throughout that each $c_P(\cdot)$ has an optimal inattention representation parametrized by $(u, \pi, \mathbb{P}^*_P, \hat{P}(\cdot))$. In Section A.2.3, I provide a sufficient condition for this specification.

I now consider the value of inattention to an optimally inattentive DM. The typical formulation (for instance, Blackwell [1953]) says that Q_1 is objectively more valuable than Q_2 if the expected utility that a DM can obtain by choosing from any problem B is higher when she conditions her choices on Q_1 than when she conditions her choices on Q_2, regardless of her utility index and prior. To state this formally, define

$$V(Q, B, u, \pi) = \sum_{E \in Q} \pi(E) \max_{f \in B} \int u \circ f d\pi(\cdot|E)$$

for every $Q \in \mathcal{P}$, problem B, utility index u and probability measure π. Say that Q_1 *is objectively more valuable than* Q_2 if and only if

$$V(Q_1, B, u, \pi) \geq V(Q_2, B, u, \pi)$$

for every problem B, utility index u and prior π.

This definition only makes sense if the DM in the absence of inattention because an inattentive DM may not be able to condition her choices on the objective information. Instead, I propose the following alternative: Q_1 *is subjectively more valuable than* Q_2 if and only if

$$\max_{Q' \in \mathbb{P}^*_{Q_1}} V(Q', B, u, \pi) \geq \max_{Q' \in \mathbb{P}^*_{Q_2}} V(Q', B, u, \pi)$$

for every problem B, utility index u and prior π. A key difference between the two notions is that whether Q_1 is subjectively more valuable than Q_2 depends on the DM under consideration – DM1 may regard Q_1 as subjectively more valuable than Q_2 while DM2 reverses the ranking.

Theorem 5 relates this comparison to comparative capacity for attention. Note that "more attentive than" is defined for a fixed information structure, but can be easily adapted to our present context where the information structure varies. I omit a formal restatement.

Theorem 5. *For any $Q_1, Q_2 \in \mathcal{P}$, Q_1 is subjectively more valuable than Q_2 if and only if $c_{Q_1}(\cdot)$ is more attentive than $c_{Q_2}(\cdot)$.*

One case where objectively and subjectively more valuable agree is if the DM is Bayesian. Another case where the equivalence between (i) and (iii) holds is if

$$\mathbb{P}^*_P = \{Q' \ll P : Q' \, has \, at \, most \, \kappa \, elements\}$$

for every $P \in \mathcal{P}$. In general, however, the equivalence fails.[24] This accords with some real-life evidence on information overload. For instance, when choosing between health care plans, DMs may become overwhelmed by the sheer amount of information available and make decisions based on less information as more is provided. The contrast between objective and subjective valuation of information is one step towards analyzing information overload.[25]

1.6 Application: Markets with Optimally Inattentive Consumers

This section argues that inattention may increase competition among firms and benefit consumers. I use a simple model to show that, in equilibrium, more attentive consumers may have less expected consumer surplus. The key idea is that in order for firms that produce differentiated products to exploit their market power, consumers must pay attention to the differences between the products. If consumers are optimally inattentive, then firms must compete with each other for attention, even

[24] Suppose that $\Omega = \{a, b, c, d\}$, $P = \{\{a, b\}, \{c, d\}\}$, $Q = \{\{a\}, \{b\}, \{c\}, \{d\}\}$ and $\mathbb{P}^*_{Q'} = \{\{E, E^c\} : E \in Q\}$ for $Q' \in \{P, Q\}$. Then for $B = \{(100, 100, -100, -100), (-100, -100, 100, 100)\}$, facing P yields a higher ex ante expected utility than facing Q.

[25] Further work along this line is in progress.

if their products would not compete given that consumers were Bayesian. This result contrasts with several papers demonstrating that firms benefit when consumers exhibit inattention; for instance, Rubinstein [1993].

To illustrate the model, I return to the example in Section 1.2, but suppose now that the drugs are non-prescription, the patient must pay out of pocket, and the patient has access to the same information that the doctor did. The patient purchases at most one of the drugs and observes the price of all three drugs before deciding which to purchase. If she processes all available information and the three firms compete by setting prices, then each firm picks a price that extracts her entire surplus whenever its drug is most effective.[26] In contrast, if the patient cannot process all of her information and is optimally inattentive, then consumer surplus must be positive in equilibrium. To see this, suppose that one firm sets a price that would extract all surplus. The patient has no incentive to pay attention to information revealing if that firm's drug is effective. Consequently, to induce the patient to pay attention to information about its drug's effectiveness, each firm must set a price that gives positive consumer surplus.

1.6.1 Model

There are n risk-neutral firms. Each costlessly produce one of $m \geq 3$ distinct, non-divisible goods. A market ϕ is an element of $\{1, ..., m\}^n$ with the interpretation that firm i produces product ϕ_i. Let $n_\mu(\phi)$ be the number of firms who have a monopoly on producing a good of a given type, and $n_c(\phi)$ be the number of goods produced competitively, i.e. by at least two firms. All consumers and firms know the type of product that each firm produces.

[26] Similarly, if the consumer has fixed attention, then any firm whose information she distinguishes can extract all surplus.

A risk-neutral consumer purchases at most one unit of the good. The state space is $\Omega = \{1, ..., m\}$, and the consumer values a good at 1 if its type matches the state and otherwise values it at 0. She initially assigns equal probability to each state and has access to information that reveals the state of the world perfectly, i.e. her objective information is $P = \{\{\omega\} : \omega \in \Omega\}$. She has optimal inattention with an attention constraint parametrized by κ where

$$\mathbb{P}^* = \{Q \ll P : \#Q \leq \kappa\}.$$

The timing of the game is as follows. First, the state of the world is determined. Then, firms simultaneously choose a price without observing the state. Next, each consumer observes the price. Finally, each consumer chooses her subjective information, observes its realization, and purchases from one of the firms.

Let $\phi_i(q)$ be the act of buying from firm i in market ϕ at price q, so that

$$u(\phi_i(q))(\omega) = \begin{cases} 1 - q & if\, \omega = \phi_i \\ -q & otherwise \end{cases}.$$

A pair (ϕ, p), where ϕ is a market and $p \in \mathbb{R}^n_+$ is a price vector, corresponds to the problem $\{\phi_i(p_i) : i \leq n\} \cup \{\emptyset\}$, where $\emptyset$ is not buying from any firm. An equilibrium for a given market ϕ is a price vector $p \in \mathbb{R}^n_+$ so that each firm j in ϕ maximizes expected profit given p_{-j} and the choices of the consumer.

1.6.2 Equilibrium

In any equilibrium where there are more products available than the consumer has the capacity to differentiate between, i.e. attention is scarce, then the effective equilibrium price is zero.

Proposition 2. *For any market ϕ, if p is an equilibrium for ϕ and the consumer purchases from firm j, then $n_c(\phi)+n_\mu(\phi) > \kappa$ implies that $p_j = 0$, and $n_c(\phi)+n_\mu(\phi) \leq \kappa$ implies that for any j, either $p_j = 1$ or $p_j = 0$, where $p_j = 1$ if and only if j has a monopoly.*

To illustrate, consider first the case where there are $n = 6$ firms, $m = 3$ products and the market is $\phi = (1, 2, 3, 1, 2, 3)$. Two firms produce each type of product. If $\kappa = 3$, then the consumer is Bayesian and a firm makes a sale only if the product it produces matches the state of the world. Since consumers know the state of the world, in every state j, the two firms of type j play the Bertrand duopoly game with marginal cost equal to zero. Consequently, the only equilibrium is $p = (0, 0, 0, 0, 0, 0)$.

Consider the same market where $\kappa = 2$. The same price vector is an equilibrium, but the consumer behaves differently. The consumer pays attention to $\{\{j\}, \{j\}^c\}$ for some $j \in \{1, 2, 3\}$, and in state j, she is indifferent between purchasing from either of the two firms with type j; in any other state, she is indifferent between purchasing from any of the four remaining firms. WLOG, assume that the consumer's subjective information is $\{\{1\}, \{2, 3\}\}$.

Now, suppose firms 5 and 6 exit the market, so there are $n' = 4$ firms and the market is $\phi' = (1, 2, 3, 1)$. If $\kappa = 3$, then firms 2 and 3 have monopolies on producing goods of type 2 and 3, respectively, so both these firms charge 1. In contrast, firms 1 and 4 both produce good 1, so they compete as a Bertrand duopoly. The unique equilibrium price vector is $(0, 1, 1, 0)$, and consumers get expected consumer surplus equal to $\frac{1}{3}$.

But if $\kappa = 2$, then an equilibrium price vector is $(0, 0, 0, 0)$. Firms 1 and 4 compete as a duopoly; if either charged a positive price and made a sale, the other could undercut the price and make a larger profit. Since firm 1 sets a price equal to zero, the consumer must decide whether to pay attention to information that distinguishes either state 2 or state 3. To attract the customer, firms 2 and 3 must offer the consumer surplus conditional on paying attention to the information that reveals whether their product is optimal. Again, if either charged a positive price and made a sale, the other makes zero profit. The firm without a sale could undercut the other's price, causing the consumer to pay attention to different information and make a sale.[27]

Though the above equilibrium is not unique, in *any* equilibrium to the game, no firm that charges a positive price makes a sale with positive probability in equilibrium. Intuitively, if two firms share a type, and the first charges a positive price and makes a sale with positive probability, then the second can undercut its price to capture the whole market. Competition between the two firms drives the price to zero. If no other firm shares a type with a firm that charges a positive price, then the consumer does not pay attention to information about that firm's product. Consequently, if any firm charges a positive price, then no consumer purchases from it.

For a given market ϕ and price vector p, expected consumer surplus weakly increases with κ. However, *equilibrium* consumer surplus is non-monotonic in κ for the

[27] This can also be seen using Monotonicity and INRA. Define $B(p_2, p_3) = (\phi, (0, p_2, p_3, 0, 0, 0))$ and $B'(p_2, p_3) = (\phi', (0, p_2, p_3, 0))$ for $p_2, p_3 \geq 0$. Identify $B'(p_2, p_3)$ as the natural subset of $B(p_2, p_3)$, i.e. $\phi'_i(p)$ is $\phi_i(p)$ for $i \leq 4$. Suppose $p_3 = 0$ and that $p_2 > 0$. Because $\phi_5(0)$ dominates $\phi_2(p_2)$,

$$c(B(p_2, p_3)|\omega) = c(B(0, p_3)|\omega) \setminus \{\phi_2(p_2)\}$$

for every $\omega \in \Omega$ by Monotonicity and INRA. Since

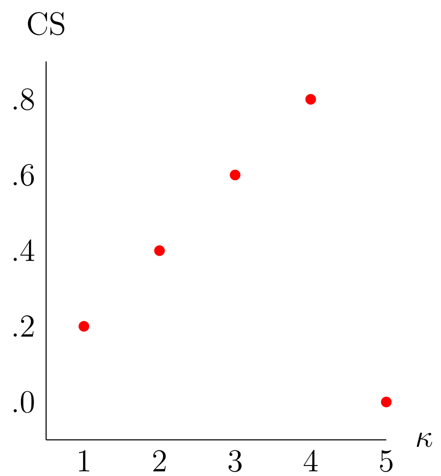
$$c(B(p_2, p_3)|\omega) \cap B'(p_2, p_3) \neq \emptyset$$

for every $\omega \in \Omega$, INRA implies that

$$\phi_2(p_2) \notin c(B'(p_2, p_3)|\omega)$$

for every ω. Consequently, firm 2 is indifferent between charging 0 and any other price. Repeating the above arguments but swapping p_3 with p_2 and firm 2 with firm 3 shows that the same holds for firm 3, so $(0, 0, 0, 0)$ is an equilibrium. In this equilibrium, the expected consumer surplus is $\frac{2}{3}$, larger than with $\kappa = 3$.

Figure 1.6.1: Expected Consumer Surplus with $n = m = 5$ and $n_c(\phi) = 0$



above market – it is maximized at $\kappa = 2$. Proposition 3 characterizes equilibrium consumer surplus for any market.

Proposition 3. *In any equilibrium for ϕ, expected total surplus is equal to*

$$\frac{1}{m}\min(n_\mu(\phi) + n_c(\phi), \kappa),$$

and expected consumer surplus equals $\frac{1}{m}\kappa$ if $\kappa < n_\mu(\phi) + n_c(\phi)$ or $\frac{1}{m}n_c(\phi)$ if $\kappa \geq n_\mu(\phi) + n_c(\phi)$.

Figure 1.6.1 illustrates Proposition 3 graphically. In a market with five firms that each produce different products, expected consumer surplus is maximized at $\kappa = 4$ and minimized at $\kappa = 5$. Similarly, expected profit is maximized when $\kappa = 5$ and equals 0 for any other value of κ. Proposition 3 shows that this non-monotonicity occurs whenever $n_c(\phi) < \kappa$.

1.7 Conclusion

In this paper, I have axiomatically characterized the properties of conditional choices that are necessary and sufficient for the DM to act as if she has optimal inattention. These axioms provide a choice-theoretic justification for the theory that agents respond to their limited attention optimally. The optimal inattention model is a versatile model with interesting implications: Dow [1991], Rubinstein [1993], and Gul et al. [2011] all consider consumers who conform exactly to the optimal inattention model.

Related papers by de Olivera [2012] and Mihm and Ozbek [2012] study rational inattention as revealed by a DM's ex ante preference over menus of acts.[28] Their representation of preference is similar to my own, but the primitives are very different. The DM chooses a menu in the anticipation that she will receive information and can choose what information to process at some cost.

Caplin and Martin [2012] study a related model, optimal framing. If frames are interpreted as states, then their analysis can be interpreted similarly to mine. Our papers are complementary, as their framework is designed for testing in the laboratory but does not achieve as precise identification. Moreover, Caplin and Martin relate choices to one another only through the existence of a utility function that solves a system of inequalities. In contrast to this paper, their primitive is stochastic choice, and their DM's prior is known to the modeler.

By way of conclusion, I compare the optimal inattention model with some other models of inattention that have been considered by the literature. The most prominent example is the rational inattention model, due to Sims [1998, 2003]. In this model, the constraint on attention takes the form of restricting the mutual infor-

[28]Ergin and Sarver [2010] can also be interpreted in this way, but it is not their focus.

mation, i.e. the reduction in entropy, between actions and the state of the world. This constraint implies that conditional choices are stochastic.[29] One interpretation is that the agent has access to arbitrarily precise, and arbitrarily imprecise, signals about the state of the world, but the modeler does not observe the realization of this information. Another, offered by Woodford [2012], is that the agent's perception of information is stochastic. Both these interpretations are outside the scope of my model: the objective information is known, and the agent's perceptions are deterministic.

Mankiw and Reis [2002] introduce the sticky information model. It postulates that agents update their information infrequently, and when they update, they obtain perfect information. The key difference between this model in a static setting and optimal inattention is that agents do not choose the information to which they pay attention.[30]

[29]Recently, Matejka and McKay [2012] have studied this model's implications in the context of discrete choices. Their focus is on solving the model in a discrete setting, and in the course of analysis, they provide testable implications in terms of choices from a suitably rich feasible set of actions. A full behavioral characterization of the model, even in this setting, remains an open question.

[30]The subjective learning literature, e.g. Dillenberger et al. [2012] and Natenzon [2012], studies an agent who has or anticipates receiving information (which is unobserved to the modeler) before making her choice. Though they do not focus on interpreting this behavior as inattention, these models have the same relationship to mine as does Mankiw and Reis [2002].

Chapter 2

Advertising and Inattention

2.1 Introduction

Introspection suggests, and psychologists have confirmed, that individuals often fail to pay attention to all available information. This paper explores how consumer inattention affects market outcomes and studies the effect of advertising that increases consumer capacity for attention. I analyze a market with consumers who easily observe prices but may fail to pay attention to match-specific quality information. Advertisements make it easier for consumers to pay attention to information about match quality. Each individual firm has an incentive to advertise because an increase in consumer capacity for attention has two effects, both of which increase profit. First, it increases the probability that a consumer knows that its good matches her preferences, increasing the chance of a sale. Second, it decreases the elasticity of substitution between the different products, so the firm can charge a higher price. Consequently, inattention leads to lower firm profits but may increase consumer surplus in the absence of advertising.[31]

Firms often benefit from providing information to consumers, e.g. Lewis and Sappington [1994]. However, if all firms provide information and consumers have a

[31] Most previous work focuses on cases where limitations on consumer information processing have the opposite effects. For instance, Rubinstein [1993], Piccione and Rubinstein [2003] and Spiegler [2006] provide situations where similar limitations either benefit firms or harm consumers.

limited capacity to pay attention, then the volume of available information can overwhelm their ability to process it, depriving firms of these benefits. Advertisements may increase consumer capacity for attention through several mechanisms. First, advertisements may change the presentation of the available information. Second, advertising may divert the consumer's capacity from other industries. Third, advertisements make the consumer aware of information available elsewhere. Note that the latter two mechanisms do not require that the ads themselves contain much, or any, information.

US advertising revenues reached $144 billion in 2011,[32] and the literature studying it has documented several effects that existing models struggle to explain with a single mechanism.[33] I show that the model of advertising studied herein generates the following effects holistically. First, advertising has a positive effect on price (e.g. Tremblay and Tremblay [1995] for US brewery industry). Second, advertising increases industry-level demand (e.g. Fischer and Albers [2010] for direct-to-consumer advertising in the US pharmaceutical industry). Third, advertising can have positive spillover effects (e.g. Kwoka [1993] for the US automobile industry). Fourth, if prices are sufficiently low, advertising can decrease the likelihood that a given consumer purchases from the advertiser (e.g. Anand and Shachar [2011] for US television programming).

Moreover, this theory of advertising has desirable features independent of the model. First, the theory is explicitly based on the idea that agents are boundedly rational. Bounded rationality is particularly appealing when considering advertising, which "is one of the topics in the study of industrial organization for which the tradi-

[32]Source: Kantar Media Reports (http://kantarmediana.com/intelligence/press/us-advertising-expenditures-increased-08-percent-2011)

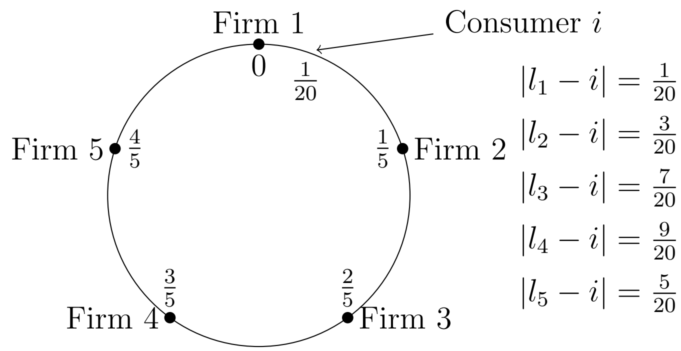
[33]It should be noted that advertising has different effects in different industries.

tional assumptions (especially those with regard to consumer behavior) are strained most" (Tirole [1988, p115]). Second, it applies to experience goods, i.e. goods whose value can be learned precisely only after purchase. According to the Nielsen Company, the automotive and pharmaceutical industries, which both produce experience goods, ranked first and third by advertising expenditure among all US industries in 2009.[34] Third, repeat purchase does not play any role in the model. Repeat purchase is unlikely for many experience goods, such as automobiles, pharmaceuticals, or computers, but is a key ingredient in models of advertising as a signal, e.g. Nelson [1974] or Milgrom and Roberts [1986]. Fourth, advertising need not contain much, or any, information to be effective. Abernethy and Franke [1996] provide a meta-analysis revealing that 29% of television ads contain no informational cues and that more than 65% contain one or fewer informational cues.

The remainder of the paper proceeds as follows. Section 2.2 introduces the model formally. To avoid Bertrand paradox-like results, I consider a variant of the spatial competition model introduced by Salop [1979]: firms and consumers are located on a circular product space. In addition to their different locations, each firm's product may be either a high- or a low-quality match for any given consumer. Consumers easily observe the price and location of each firm but must process information to learn match quality. Each is optimally inattentive (as in Rubinstein [1993] or Ellis [2012]), i.e. chooses the information that she processes to maximize ex ante expected utility. A consumer who sees an ad from a given firm can pay attention to information that reveals if that firm's product is valuable "for free." She still must process information about the quality of other firms, but she can devote the same capacity as before to a smaller number of firms. Consequently, advertising increases a consumer's capacity

[34]Source: http://smallbusiness.chron.com/industry-spends-advertising-22512.html

Figure 2.2.1: Setup with $n = 5$ firms



for attention.

Section 2.3 describes an inattentive consumer's purchasing behavior and derives the demand curve. I demonstrate that advertising in the model can have the effects on demand described above. Section 2.4 analyzes the equilibrium of the model in the absence of advertising as a baseline. Price, profit, and total surplus increase with the consumer's capacity for attention. Section 2.5 analyzes the full version of the model. Price increases with advertising. Both equilibrium price and equilibrium advertising increase as advertising costs decline, but equilibrium profit may either increase or decrease. Advertising is typically under-provided relative to the socially optimal level. Section 6 discusses alternate models of advertising and which effects and implications they deliver.

2.2 The Model

Firms, production and the product space

The product space is a circle of circumference 1. There are n firms located at equally spaced points along the circle. I assume $n \geq 3$ throughout. Let l_j be the location of firm j. Each firm produces a distinct product at constant marginal cost c.

Consumers

A unit mass of consumers is distributed uniformly along this circle. Identify consumer i with the consumer located at the point i on the circle. Each consumer wishes to purchase at most one unit of the good from some firm. To travel to the firm, she incurs a cost equal to $t > 0$ times the distance traveled.[35]

The gross value that a consumer gets from purchasing a good depends on the quality of the match between the good and the consumer, which may be either high or low. The state space is $\times_{i \in [0,1]} \Omega$ where $\Omega = \{H, L\}^n$ and $\omega_j^i = H$ if and only if firm j produces a high quality match for consumer i. For any i and j, the probability that $\omega_j^i = H$ is q, which is independent of the probability that $\omega_{j'}^i = H$ for all $j' \neq j$ and any i. If state $\omega_j^i = H$, then consumer i values the good produced by firm j at $v > c$ and otherwise values it at $b < c$. She may also opt not to purchase any goods. If she does so, then she receives utility 0. Denote the probability that $\omega^i = \omega$ by $\rho(\omega)$, noting this does not depend on i. If consumer i purchases from firm j at price p_j, then she gets utility

$$u_i = \begin{cases} v - t|l_j - i| - p_j & if\, \omega_j^i = H \\ b - t|l_j - i| - p_j & if\, \omega_j^i = L \end{cases}$$

where $|l_j - i|$ is the arc distance between l_j and i, i.e. it takes into account that distance "wraps" around the circle. Throughout, I write ω_j for ω_j^i when it will not

---

[35] I use the analogy of spatial competition throughout, but the product space need only reflect some easily observed characteristic with idiosyncratic valuations, such as name, color, or size. Many of my results require some form of easily observable horizontal differentiation because it smooths consumer responsiveness to changes in price. Without it, no non-trivial pure-strategy equilibrium exists. The choice of a circular product space with linear transportation costs is not crucial. It makes the model tractable and symmetric as well as allowing for relatively simple closed form solutions, but the results continue to hold (with different functional forms) if one assumes other forms of horizontal differentiation, e.g. Perloff and Salop [1985].

cause confusion.

Information and inattention

Each consumer observes the price each firm sets and knows the location of the firm. Moreover, each has access to information that reveals the state of the world perfectly. If she processes all available information, then she knows exactly which firms are a high quality match and which are a low quality match. However, she may not pay attention to all available information. I assume that each consumer's constraint allows her to pay attention to information revealing the quality of the match for at most κ firms and that each chooses to which she pays attention optimally. She observes all firms' prices before she observes quality information, so the information to which she pays attention varies with the prices. In other words, if the price vector is p, then she pays attention to the information that maximizes her ex ante expected consumer surplus given these prices. Following the terminology in Ellis [2012], each consumer has *optimal inattention* and the information that she processes is her *subjective information*.

Advertising technology

Receiving an ad from firm j allows the consumer to process information about the quality of firm j without affecting her ability to process information about the other firms. That is, if she sees an ad from only firm j, then she can pay attention to information that reveals the quality of any $\kappa + 1$ firms as long as at least one of these firms is j. More generally, if consumer i sees ads from the m firms in the set A_i, then she can pay attention to information revealing the quality of $\kappa + m$ firms, provided that at least m of these firms are in A_i. This increases her capacity for attention, necessarily increasing her consumer surplus.[36]

The advertising technology is similar to that in Grossman and Shapiro [1984].[37] Each firm selects a level of advertising, ϕ, between 0 and 1. If firm j selects advertising level ϕ_j, then a measure ϕ_j of the population of consumers receives an advertisement from firm j, and it incurs a cost $A(\phi_j)$. Ads cannot be targeted at a specific subset of consumers: all potential customers, regardless of their location on the circle, are equally likely to receive an ad from a given firm. The probability that consumer i sees an ad from firm j is independent of the probability that she sees an ad from firm $j' \neq j$. I assume that $A'(0) \geq 0$, $A'(\phi) > 0$ for all $\phi > 0$, $A''(\phi) > 0$ for all ϕ, and $A(0) = 0$. Examples of advertising cost functions include $A(\phi) = a\phi^2$ and $A(\phi) = -a\log(1-\phi)$.

Timing

The timing of the game is as follows.

1. Firms simultaneously choose an advertising level ϕ_j and a price p_j.

2. A fraction ϕ_j of consumers receive an ad from firm j.

3. Each consumer observes these prices and chooses her subjective information partition.

4. Each consumer observes the realized cell of her subjective information partition and purchases from at most one firm.

[36] Some readers may be troubled that the effect of advertising is too extreme. I do not mean for this to be taken as a literal description of the effect of advertising. Seeing an advertisement from firm j reduces the psychological cost of paying attention to information about j. The extreme nature of the reduction in cost simplifies the analysis and makes the model tractable.

[37] The examples of advertising technologies provided in that paper also work for this paper. Note that advertising has a very different effect on consumer behavior in this paper than in theirs.

Solution concept

I focus on a symmetric, pure strategy Nash equilibrium in prices and advertising. Each firm chooses its price and advertising level to maximize profits, given that every other firm prices at p^* and advertises at ϕ^*. If firm j sets price p, then net revenue equals the measure of consumers who buy from it multiplied by $p - c$. Profit equals net revenue minus advertising cost. Given a vector of prices and a set of firms from which a given consumer has seen advertisements, each firm knows from which of them she will purchase, conditional on the state of the world. Because there are a continuum of consumers and both match quality and advertisements are independently and identically distributed, profit is deterministic.

2.3 Consumer Behavior and Firm Demand

2.3.1 Consumer Behavior

For this section, I fix firm behavior and consider the purchasing behavior of a given consumer, labeled i. Consumer i faces two decisions: to what information should she pay attention and from which firm (if any) should she purchase? In what follows, it will be helpful to denote by

$$\hat{p}_j(i) = p_j + t|i - l_j|$$

the effective price that i pays if she buys from firm j. I will also assume that no firm plays an strictly dominated strategy, i.e. $p_j \geq c$ for all j and that each $\hat{p}_j(i) \leq v$ for every i.[38] For any price vector, the set of consumers for which at least two effective prices are equal is measure zero, so I focus only on the case where each effective price

[38] This holds for all consumers if and only if $p_j \leq v - \frac{t}{2}$ for every j. Consumer i never purchases from firm j if $\hat{p}_j(i) > v$. Therefore, if m firms have $\hat{p}_j(i) > v$, then all the analysis below goes through with $n^* = n - m$ replacing n and adding the qualifier "with effective price less than v" after each instance of "firm".

is distinct.

Conditional on the information to which she pays attention, the purchasing decision of consumer i is simple. She never purchases from a firm that she knows to be a low quality match. If there is at least one firm known to be a high equality match, then label the one with the lowest effective price as h, and if there is at least one firm of unknown match quality, then label the one with the lowest effective price as u. Purchasing from firm h (respectively, u) yields strictly higher utility than purchasing from any other firm known to be a high quality match (resp. firm with unknown match quality). If there are no firms of unknown quality and at least one firm known to be high quality, then the consumer purchases from firm h. If there are no firms known to be high quality, there is at least one firm of unknown quality, and $qv - (1-q)b > \hat{p}_u(i)$, then the consumer purchases from firm u. If there is at least one firm of unknown quality and at least one firm known to be high quality, then the consumer purchases from firm h if

$$v - \hat{p}_h(i) \geq qv + (1-q)b - \hat{p}_u(i)$$

and otherwise purchases from u.

The consumer always pays attention to information that reveals the quality of the firms from which she receives advertisements. The utility from purchasing from firm j only when she knows it is high quality is

$$q[v - \hat{p}_j(i)] + (1-q)0.$$

This is strictly greater than either the utility of purchasing a good of unknown quality,

$$qv + (1-q)b - \hat{p}_j(i),$$

or the utility not purchasing anything, 0. Consequently, consumer i gains by paying attention to information that reveals the match-quality of j whenever it does not consumer capacity to do so.

Consider her choice of to which of the remaining firms' quality she pays attention. If she sees more than $n-\kappa$ ads, then she pays attention to all of their qualities. If she sees less than $n-\kappa$ ads, then the optimal subjective information partition takes one of two forms. Either the information to which she pays attention reveals the quality of the κ remaining firms with the lowest effective prices, or it reveals the quality of the κ remaining firms whose effective prices rank between the second and $\kappa+1$ lowest from which she saw no ads.

It is typically optimal to choose the former subjective information.[39] If none of the firms are high quality (which occurs with probability $1-(1-q)^{\kappa+m}$ if she sees m ads), then she purchases from firm u if $qv+(1-q)b > \hat{p}_u(i)$ and makes no purchase otherwise. Assuming that she sees no ads and that the firms are labeled so that $\hat{p}_j(i) < \hat{p}_{j'}(i)$ if and only if $j < j'$, if the consumer pays attention to the information partition above, then the consumer's expected utility is

$$(1-(1-q)^{\kappa})v - \sum_{j=1}^{\kappa}(1-q)^{i-1}q\hat{p}_j(i) \tag{2.3.1}$$

if $qv+(1-q)b < \hat{p}_{\kappa+1}(i)$ and is

$$(1-(1-q)^{\kappa})v - \sum_{j=1}^{\kappa}(1-q)^{i-1}q\hat{p}_j(i) + (1-q)^{\kappa}[qv+(1-q)b-\hat{p}_{\kappa}] \tag{2.3.2}$$

otherwise.

With sufficiently low effective prices, a sufficiently low probability of a good match,

[39] In fact, I will impose an assumption from the following section onward that implies this will be the case.

and a sufficiently high benefit of purchasing a good match, the consumer may find the latter subjective information optimal. In this case, if none of the firms are high quality (which occurs with probability $1-(1-q)^{\kappa+m}$ if she sees m ads), then she purchases from the firm with the lowest effective price. Her ex ante expected utility is

$$(1-(1-q)^{\kappa})v - \sum_{j=2}^{\kappa+1}(1-q)^{i-1}q\hat{p}_j(i) + (1-q)^{\kappa}[qv + (1-q)b - \hat{p}_1]. \qquad (2.3.3)$$

Two necessary (but not sufficient) conditions for optimality of this subjective information are that $\hat{p}_{\kappa+1} < qv + (1-q)b$ and $(1-q)^{\kappa} > q$.

If the consumer finds the latter subjective information to be optimal, then the consumer is more likely to purchase from firm 1 if she has not seen an ad from it, (probability $(1-q)^{\kappa}$) than if she has seen one ad (probability q). Consequently, seeing an advertisement can actually decrease the probability that a consumer purchases from the advertising firm. Anand and Shachar [2011] document this effect for television program advertising. Conditional on viewing an advertisement for a given program, the probability of a given consumer watching it increases if it is a good match and decreases if it is a bad match.

2.3.2 Firm Demand

For simplicity, I make the following assumption for the remainder of the paper.

Assumption. $qv + (1-q)b < c$

That is, the expected gross benefit of purchasing from a firm of uncertain quality is less than the marginal cost of production. It implies that whenever a firm plays an undominated strategy, the consumer does not wish to purchase a good unless she paid attention to information that revealed it is a high quality match. This greatly simplifies the number of cases one must consider when deriving the firm's demand curve: all consumers pay attention to the information that reveals the quality of the κ firms with the lowest effective prices. The substance of the results does not change substantially without the assumption but it greatly simplifies the statement and interpretation of the results.

I now turn to deriving the firm's demand curve with fixed and symmetric levels of the advertising set by other firms. Let $f(r;\phi,n) = \binom{n}{r}\phi^r(1-\phi)^{n-r}$ and $F(r;p,n) = \sum_{i=0}^{r}\binom{n}{i}\phi^i(1-\phi)^{n-i}$ be the be the probability mass function and the cumulative distribution function, respectively, of the Binomial distribution with n trials at success probability ϕ.

Lemma 2. *If every other firm prices at $\bar{p} \leq v - \frac{t}{2}$ and selects advertising level $\bar{\phi}$, then the demand of firm j when it sets price p (which is suitably close to $\bar{p}$ and less than or equal to $v - \frac{t}{2}$) and advertises at ϕ is*

$$D(p,\phi;\bar{p},\bar{\phi}) = \phi\delta_I(\bar{\phi}) + (1-\phi)\delta_U(\bar{\phi}) + \frac{\bar{p}-p}{t}[\phi\Delta_I(\bar{\phi}) + (1-\phi)\Delta_U(\bar{\phi})], \qquad (2.3.4)$$

where:

$$\delta_I(\bar{\phi}) > \delta_U(\bar{\phi}) \text{ and } \Delta_I(\bar{\phi}) < \Delta_U(\bar{\phi}).$$

See Section B.1.1 for a precise statement of the demand function and derivation. To understand why Lemma 2 holds, it will be helpful to again denote the effective price that consumer i pays if she buys from firm j by $\hat{p}_j(i) = p_j + t|i - l_j|$. Fix a price vector p and a firm j. Divide the consumers up into n groups, where a consumer in group l ranks the effective price of firm j as the lth lowest. Let N_l be the measure of consumers in group l.

First, consider consumer i' in group l where $l \leq \kappa$. There are $l-1$ firms with effective prices lower than l, Firm j is a high quality match with probability q, and each of these $l-1$ firms is a low quality match with probability $(1-q)^{l-1}$. Consequently, firm j has the lowest effective price of all the firms she knows has a high quality match with probability

$$(1-q)^{l-1}q,$$

regardless of the ads she sees. Therefore, she buys from firm j with probability $(1-q)^{l-1}q$.

Next, consider consumer i' in group l where $\kappa < l$ who *does not see* an ad from firm j. She buys from firm j if she sees more than $n-\kappa$ total ads with probability $(1-q)^{l-1}q$. She also buys from firm j if she sees $z < n-\kappa$ total ads, $i \geq l-\kappa$ of which are from firms that have lower effective prices, with probability $(1-q)^{l-1}q$. Consequently, she sells to consumer i' with probability

$$\begin{aligned}&[1-F(n-\kappa-1;\bar{\phi},n-1)](1-q)^{l-1}q\\&\quad+\sum_{z=l-\kappa}^{n-\kappa-1}\sum_{i=l-\kappa}^{z}\binom{n-l}{z-i}\binom{l-1}{i}\bar{\phi}^{z}(1-\bar{\phi})^{n-1-z}(1-q)^{l-1}q.\end{aligned}$$

Finally, consider consumer i' in group l where $\kappa < l$ who *does see* an ad from firm j. She buys from firm j if she sees more than $n-\kappa$ total ads ($n-\kappa-1$ ads from its competitors) with probability $(1-q)^{l-1}q$. She also buys from firm j with probability $(1-q)^{\kappa+i}q$ if she sees $z < n-\kappa$ total ads, i of which are from firms that have lower effective prices. Finally, she buys from j with probability $(1-q)^{l-1}q$ if she sees $z < n-\kappa$ total ads, at least $l-\kappa$ of which are from firms that have lower effective prices. These events are mutually exclusive. Consequently, she sells to consumer i'

with probability

$$
\begin{aligned}
&[1 - F(n-\kappa-2; \bar{\phi}, n-1)](1-q)^{l-1}q \\
&\quad + \sum_{z=0}^{n-\kappa-2} \sum_{i=0}^{l-\kappa-1} \binom{n-l}{z-i}\binom{l-1}{i} \bar{\phi}^z (1-\bar{\phi})^{n-1-z}(1-q)^{\kappa+i} q \\
&\quad + \sum_{z=0}^{n-\kappa-2} \sum_{i=l-\kappa}^{z} \binom{n-l}{z-i}\binom{l-1}{i} \bar{\phi}^z (1-\bar{\phi})^{n-1-z}(1-q)^{l-1} q.
\end{aligned}
$$

If each firm sets the same price $\bar{p}$, then $N_1 = N_2 = ... = N_n = \frac{1}{n}$. The above cases imply that probability of a consumer in group N_l buying from j if she sees and ad from j is larger than if she does not. Consequently, $\delta_I(\bar{\phi}) > \delta_U(\bar{\phi})$.

Grossman and Shapiro [1984] show that when firm j increases (resp. decreases) its price by a small amount, N_1 decreases (increases) and N_n increases (decreases) by the same amount. A consumer in group n buys from j with probability

$$q(1-q)^{n-1}(1 - F(n-\kappa-1, \bar{\phi}, n-1)$$

if she does not see an ad from j and with probability

$$q(1-q)^{n-1}(1 - F(n-\kappa-2; \bar{\phi}, n-1)) + \sum_{i=0}^{n-\kappa-2} f(i; \bar{\phi}, n-1)(1-q)^{\kappa+i} q$$

if she does. Since the latter is larger than the former and firm j sells to a consumer in group 1 with probability q regardless of whether the consumer sees an ad from j, the change demand for good j if j changes its price is larger among those who see an ad from j, so $\Delta_I(\bar{\phi}) < \Delta_U(\bar{\phi})$.

2.3.3 Spillover

Empirical and theoretical work on advertising has shown that advertising may have a primary demand effect, i.e. increases industry-level demand, and may have positive spillover effects, i.e. increases rather than decreases the demand for competing products. This section shows that advertising for attention can have both these effects. An increase in advertising by firm j can increase the demand of firm j', holding prices and other firms' advertising levels constant, provided that initial advertising levels are not too high.

Many papers provide evidence of a positive effect of advertising on industry-level demand, e.g. Cowling et al. [1975] for cigarettes in the UK, Seldon and Doroodian [1989] for cigarettes in the US, or Nerlove and Waugh [1961] for oranges in the US. Moreover, several other theoretical models of advertising assume spillover, e.g. Examples 1, 2, and 4 in Ellison and Ellison [2011], without specifying why or how spillover occurs. Inattentive advertising provides micro-foundations for the spillover effect. Several papers document positive spillover empirically, e.g. Kwoka [1993] for automobiles in the US, Wosinska [2005] for direct-to-consumer advertising of US cholesterol drugs, and Kadiyali [1996] for the photographic film in the US.

I now demonstrate that these effect occur naturally in the model. Suppose that $n = 3$, $\kappa = 1$, every firm advertises with intensity ϕ, and all firms set the same price $p < v - \frac{t}{2}$. What happens to the demand of firm 2 when firm 1 increases its advertising level by ϵ?

Consider the demand for firm 2 when firms 2 and 3 advertise at ϕ and firm 1 advertises at $\phi + \epsilon$. The probability of seeing ads from a given subset of firms and demand for firm 2 in that instance is given by Table 2.1, where N_l is the measure of consumers who rank the effective price of firm 2 the lth lowest. When each firm sets

Table 2.1: Demand for firm 2

ads seen	probability	demand for firm 2
none	$(1-\phi-\epsilon)(1-\phi)^2$	$N_1 q$
only 1	$(\phi+\epsilon)(1-\phi)^2$	$N_1 q + \frac{1}{2} N_2 q(1-q)$
only 2	$(1-\phi-\epsilon)(1-\phi)\phi$	$N_1 q + N_2 q(1-q) + N_3 q(1-q)$
only 3	$(1-\phi-\epsilon)(1-\phi)\phi$	$N_1 q + \frac{1}{2} N_2 q(1-q)$
both 1 and 2	$(\phi+\epsilon)\phi(1-\phi)$	$N_1 q + N_2 q(1-q) + N_3 q(1-q)^2$
both 1 and 3	$(\phi+\epsilon)\phi(1-\phi)$	$N_1 q + N_2 q(1-q) + N_3 q(1-q)^2$
both 2 and 3	$(1-\phi-\epsilon)\phi^2$	$N_1 q + N_2 q(1-q) + N_3 q(1-q)^2$
all three	$(\phi+\epsilon)\phi^2$	$N_1 q + N_2 q(1-q) + N_3 q(1-q)^2$

the same price, $N_1 = N_2 = N_3 = \frac{1}{3}$. Consequently, the derivative of the demand of firm 2 with respect to ϵ evaluated and $\epsilon = 0$ is

$$(1-\phi)^2[\frac{1}{6}q(1-q)] + \phi(1-\phi)[\frac{1}{3}q(1-q)^2 - \frac{1}{6}q(1-q)].$$

This is positive if and only if $\frac{1-\phi}{\phi} > 2q-1$. Therefore, more advertising by firm 1 can lead to higher demand for firm 2.

Moreover, firms 2 and 3 are symmetric, so the demand for firm 3 increases. Advertising always increases the demand of firm 1 (given the assumption on v, b, and c). Therefore, *every firm in the market has increased demand.* Conclude that advertising has a primary demand effect.

2.4 Baseline: No Advertising

For this section, I assume that firms cannot advertise, i.e. $A(0) = 0$ and $A(\phi) = \infty$ for all $\phi > 0$. Therefore, in any equilibrium, $\bar{\phi}$ and ϕ must both equal 0.

2.4.1 Equilibrium

At $\bar{\phi}$ and ϕ equal to 0, the firm's demand function becomes

$$D(p,\bar{p}) = \begin{cases} \frac{(1-(1-q)^{\kappa})}{n} + q\frac{\bar{p}-p}{t} & if\, \kappa < n \\ \frac{(1-(1-q)^{n})}{n} + q(1-(1-q)^{n-1})\frac{\bar{p}-p}{t} & if\, \kappa \geq n \end{cases}. \tag{2.4.1}$$

The first order necessary condition for an optimal price is that

$$p = \begin{cases} \frac{\bar{p}+c}{2} + t\frac{(1-(1-q)^{\kappa})}{2nq} & if\, \kappa < n \\ \frac{\bar{p}+c}{2} + t\frac{(1-(1-q)^{n})}{2nq(1-(1-q)^{n-1})} & if\, \kappa \geq n \end{cases}.$$

This allows us to provide necessary conditions for an equilibrium.

Proposition 4. *If $\kappa \geq n$ and $p^* < v - \frac{t}{2}$ is an equilibrium, then*

$$p^* = t\frac{1-(1-q)^n}{nq(1-(1-q)^{n-1})} + c. \tag{2.4.2}$$

If $\kappa < n$ and $p^* < v - \frac{t}{2}$ is an equilibrium, then

$$p^* = t\frac{1-(1-q)^{\kappa}}{nq} + c. \tag{2.4.3}$$

Remark 1. This proposition, and many that follow, are stated as necessary conditions for equilibrium. For t large enough, such an equilibrium exists. Otherwise, only mixed strategy equilibria may exist. For instance if $t = 0$, then an equilibrium exists. This equilibrium involves pricing at cost when $\kappa < n$ and playing a mixed strategy when $\kappa \geq n$. Details available upon request.

2.4.2 Comparative Statics

I consider comparative statics on the endogenous variables, equilibrium price (p^*) and profit (π^*).

Proposition 5. *If p^* is an equilibrium with $p^* < v - \frac{t}{2}$ and $\kappa < n$, then:*

(i) As κ increases, π^* and p^* both increase.

(ii) As q increases, p^* decreases while π^* increases if $q < q^*$ but decreases if $q > q^*$, where

$$\frac{1-(1-q^*)^\kappa}{(1-q^*)^{\kappa-1}q^*} = 2\kappa.$$

(iii) As n increases, π^* and p^* both decrease.

(iv) As t increases, p^* and π^* both increase.

The first comparative static is the most important. Namely, both equilibrium price and profit increase with the consumer's capacity for attention. There is typically a very large increase from $\kappa = n-1$ to $\kappa = n$. The remaining comparatives are standard.

2.4.3 Welfare Analysis

An increase in capacity for attention increases both the probability that any given consumer purchases the good and the equilibrium price. Consequently, profit increases. However, the two effects move consumer surplus in opposite directions: the former increases it while the latter effect decreases it. Consequently, the effect on consumer surplus is ambiguous.

Proposition 6. *If $\kappa < n$ and $p^* \leq v - \frac{t}{2}$ is an equilibrium, then total surplus equals*

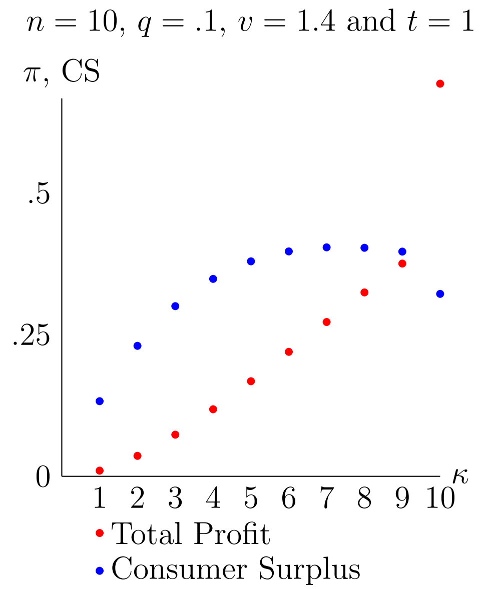
$$(v-c)(1-(1-q)^\kappa) - t\frac{2-q+(1-q)^\kappa(q-2+2q\kappa)}{4nq}, \tag{2.4.4}$$

and consumer surplus equals

$$(v-p^*)(1-(1-q)^\kappa) - t\frac{2-q+(1-q)^\kappa(q-2+2q\kappa)}{4nq}. \tag{2.4.5}$$

Total surplus monotonically increase as κ increases. All trades that occur benefit both the consumer and the firm. Since the probability of a trade occurring increases

Figure 2.4.1: Consumer Surplus and Total Profits



with κ, so does total surplus. Figure 2.4.1 shows equilibrium total profit and consumer surplus as a function of κ. Higher κ implies higher price and more demand for each firm, so profit increases with κ. However, consumer surplus may either increase or decrease as κ increases. In the example, total profit is maximized at $\kappa = n$ while consumer surplus is maximized at $\kappa = 7$. In fact, consumer surplus is increasing for $\kappa < 7$ and decreasing for $\kappa > 7$.

To see why consumer surplus is non-monotonic, consider its change as κ increases by one. With probability $(1 - q)^\kappa q$, the consumer's more precise information leads to an additional trade. However, Proposition 5 shows that p^* increases, so the consumer pays a higher price whenever she makes a trade. Consumer surplus increases if and only if the benefit resulting from the former is larger than the additional cost caused by the latter.

2.5 Equilibrium with Advertising

This section analyzes the equilibrium of the market when firms endogenously choose how much to advertise. When $\kappa \geq n$, advertising has no benefit, and since it is costly, no firm wishes to advertise. To avoid trivialities, I assume that $\kappa < n$ for the remainder of this paper. This guarantees that attention is scarce, so advertising may be valuable to firms.

2.5.1 Pricing when advertising is exogenous

For only this subsection, suppose that each firm advertises at the exogenously given level $\bar{\phi}$. Demand is given by Equation (2.3.4). This allows us to give a simple expression that the equilibrium price must satisfy when the whole market is served.

Proposition 7. *If $p^* < v - \frac{t}{2}$ is an equilibrium, then*

$$p^* = \bar{p}(\bar{\phi}) = t\frac{\bar{\phi}\delta_I(\bar{\phi}) + (1-\bar{\phi})\delta_U(\bar{\phi})}{\bar{\phi}\Delta_I(\bar{\phi}) + (1-\bar{\phi})\Delta_U(\bar{\phi})} + c. \tag{2.5.1}$$

Since $\delta_I(\bar{\phi}) > \delta_U(\bar{\phi})$ and $\Delta_I(\bar{\phi}) < \Delta_U(\bar{\phi})$, an increase in $\bar{\phi}$ increases equilibrium price and demand. To see where (2.5.1) comes from, fix $\bar{\phi}$ and $\bar{p} < v - \frac{t}{2}$. The first order necessary condition (from differentiating $(p-c)D(p, \bar{\phi}; \bar{p}, \bar{\phi})$ with respect to p) when $p < v - \frac{t}{2}$ is

$$p = \frac{\bar{p}}{2} + t\frac{D(p, \bar{\phi}; \bar{p}, \bar{\phi})}{2[\bar{\phi}\Delta_I(\bar{\phi}) + (1-\bar{\phi})\Delta_U(\bar{\phi})]} + \frac{c}{2} = \frac{\bar{p} + \bar{p}(\bar{\phi}) + c}{2}.$$

Solving for $p = \bar{p}$ yields Equation (2.5.1). Section B.1.4 considers the case where $p^* = v - \frac{t}{2}$. While p^* typically increases in ϕ, if $\bar{p}(\phi_h) = v - \frac{t}{2}$, then $p^* = v - \frac{t}{2}$ for $\phi \in [\phi_h, \phi_h + \epsilon)$ for some $\epsilon > 0$.

2.5.2 Equilibrium

If firms set a symmetric level of advertising $\bar{\phi}$, then Proposition 7 gives the price they set. It remains to determine the equilibrium level of advertising. If other firms advertise at level $\bar{\phi}$ and price at $\bar{p}$, then the demand curve of firm j when it sets advertising level ϕ and price $\bar{p}$ is

$$D(p, \bar{p}, \phi, \bar{\phi}) = \phi\delta_I(\bar{\phi}) + (1-\phi)\delta_U(\bar{\phi}).$$

In equilibrium, $\phi = \bar{\phi}$ and $p = \bar{p}$. Proposition 8 establishes equilibrium level of advertising using Proposition 7 and Envelope Theorem arguments.

Proposition 8. *If $0 < \phi^* < 1$ and $p^* \leq v - \frac{t}{2}$ are the equilibrium advertising level and price, then the implication given by Proposition 7 holds and*

$$A'(\phi^*) = (p^* - c)[\delta_I(\phi^*) - \delta_U(\phi^*)]. \tag{2.5.2}$$

To interpret (2.5.2), consider the problem of firm j when the other firms set price p^* and advertising level ϕ^*. If firm j also sets the equilibrium price and advertising level, then it sells to a $\delta_I(\phi^*)$ fraction of the consumers who see an ad from it and a $\delta_U(\phi^*)$ fraction of the consumers who do not see an ad from it. Because it was optimal to charge p^* when advertising at level ϕ^*, firm j still finds a price very close to p^* optimal. By deviating to advertising at level ϕ, the firm sells to approximately

$$(\phi - \phi^*)[\delta_I(\phi^*) - \delta_U(\phi^*)]$$

additional consumers, but it has to pay approximately

$$A'(\phi^*)(\phi - \phi^*)$$

more in advertising costs. Equation (2.5.2) implies that the change in cost is equal to the change in revenue.

In equilibrium, all firms charge the same price. Consequently, all n firms split demand evenly. Therefore, each firm's equilibrium demand equals

$$D(\phi^*) = \frac{1}{n}\{1 - \sum_{i=0}^{n-\kappa-1} f(i;\phi^*,n)(1-q)^{\kappa+i} - [1 - F(n-\kappa-1;\phi^*,n)](1-q)^n\}.$$

Equilibrium profit per firm π^* equals

$$\bar{p}(\phi^*)D(\phi^*) - A(\phi^*),$$

where ϕ^* solves Equation (2.5.2).

2.5.3 Comparative statics

The empirical literature documents a industry-level positive correlation of prices with advertising and of profits with advertising (see Bagwell [2007]). Because of the endogenous provision of advertising, one cannot distinguish the direction of the causality in the absence of a model. This section explores how equilibrium advertising intensity and profit change following a small exogenous shock to the cost in advertising.

To consider comparative statics in the cost of advertising, I assume that $A(\phi_j)$ can be written as $A(\phi_j; a)$, where $\frac{\partial A}{\partial a} > 0$ and $\frac{\partial^2 A}{\partial \phi \partial a} \geq 0$. That is, a is a shift parameter in the cost of advertising with a higher a corresponding to a higher absolute and marginal cost of advertising. To simplify the statement of Proposition 9, write $A_a \equiv \frac{\partial}{\partial a}A(\phi; a)$ and $A_\phi \equiv \frac{\partial}{\partial \phi}A(\phi; a)$, all evaluated at ϕ^*.

Proposition 9. *For every n and $\kappa > 0$,*

$$\frac{\partial \phi^*}{\partial a} < 0 \tag{2.5.3}$$

and

$$\frac{\partial \pi^*}{\partial a} = \frac{\partial \phi^*}{\partial a}[\bar{p}'(\phi^*)D(\phi^*) + \bar{p}(\phi^*)D'(\phi^*) - A_\phi] - A_a. \quad (2.5.4)$$

A small shock that raises the cost of advertising results in more advertising. Intuitively, the marginal cost of advertising increases without altering the marginal benefit to the firm from additional advertising. To restore equilibrium, the amount of advertising falls. Proposition 7 then implies that the equilibrium price decreases.

However, the same small shock may either increase or decrease equilibrium profit. Equation (2.5.4) can be rearranged to show that $\frac{\partial \pi^*}{\partial a} > 0$ if and only if

$$\frac{\partial \phi^*}{\partial a} A_\phi + A_a < \frac{\partial \phi^*}{\partial a}[\bar{p}'(\phi^*)D(\phi^*) + \bar{p}(\phi^*)D'(\phi^*)]. \quad (2.5.5)$$

On the one hand, revenue always decreases, as both price and demand decrease due to the fall in advertising. The RHS of (2.5.5) gives the value of this change, which is necessarily less than zero. On the other hand, advertising costs typically decrease as well, since the absolute level of advertising falls.[40] The LHS of (2.5.5) gives the decrease in costs, which may be either positive or negative. A sufficiently large decrease in the cost of advertising can more than offset the loss in revenues. In general, the difference between the change in revenue and the change in advertising costs may be either positive or negative.

2.5.4 Welfare analysis

As long as the whole market is served, price plays no role in determining social welfare. An increase in advertising increases the probability that a consumer purchases from a high-quality match. That is, it increases the gross social benefit obtained. The total

[40]For instance, advertising costs decrease for either $A(x, a) = ax^m$ with $m \geq 2$, $A(x, a) = -a \log(1 - x)$, or $A(x, a) = a(e^x - 1)$.

social benefit equals

$$v\{\sum_{i=0}^{n-\kappa-1} f(i,\phi,n)(1-(1-q)^{\kappa+i}) + [1-F(n-k,\phi,n)](1-(1-q)^n)\}.$$

It also affects the expected total transportation cost and consumes resources. Therefore, social welfare, as a function of ϕ, is given by

$$\begin{aligned} W(\phi) &= \sum_{i=0}^{n-\kappa-1} f(i,\phi,n)\{(1-(1-q)^{\kappa+i})(v-\mathbb{E}[TC|i])\} \\ &\quad +[1-F(n-k,\phi,n)]\{(1-(1-q)^n)(v-\mathbb{E}[TC|n])\} - nA(\phi) \end{aligned}$$

where $\mathbb{E}[TC|i]$ is the expected transportation costs given that the consumer sees i ads.

Proposition 10. *Suppose $\kappa = n-1$. If (ϕ^*, p^*) is an equilibrium, then*

$$\frac{\partial W}{\partial \phi}|_{\phi=\phi^*} > 0.$$

That is, advertising is under-provided from a total surplus perspective.

When $\kappa = n-1$, advertising increases the demand of all firms equally. Consequently, a given firm does not reap all of the gain in demand from its extra advertising, but each firm bears the full cost of producing advertising. Moreover, the price of the good is less than the marginal social benefit of selling to any additional customer, i.e. $p^* < v - (\mathbb{E}[TC|i+1] - \mathbb{E}[TC|i])$ for all i. Since each firm undervalues both the magnitude of change in demand and the benefit of of selling to an additional consumer, the private incentive to provides advertising is lower than the social incentive to provide advertising.

In general, advertising may be either over- or under-provided. The first order condition for the socially optimal level of advertising is

$$\sum_{i=0}^{n-\kappa-1} f(i;\bar{\phi},n-1)(1-q)^{\kappa+i}q[v-(\mathbb{E}[TC|i+1]-\mathbb{E}[TC|i])] = A'(\bar{\phi}). \qquad (2.5.6)$$

Comparing Equations (2.5.2) and (2.5.6), it is clear that if v is very large, then advertising is under-provided. For advertising to be over-provided, the inequality

$$\delta_I(\phi) - \delta_U(\phi) > \sum_{i=0}^{n-\kappa-1} f(i;\phi,n-1)(1-q)^{\kappa+i}q$$

must hold, which it does for many parameter values. For instance, if $n = 3$, $\kappa = 1$, then the above inequality holds whenever $1 < \phi(4q-2)$. Even if this difference is positive, advertising may be under-provided if the difference between p^* and $v - (\mathbb{E}[TC|i+1] - \mathbb{E}[TC|i])$ is large.

2.6 Discussion

I have analyzed a model in which advertising increases consumer capacity for attention. Advertising increases own demand, increases primary demand, and can have positive spillover. Additionally, seeing an ad from firm j can lower the probability that a consumer purchases from firm j (the "match effect"). In equilibrium, advertising increases price, and firms may either over- or under-provide advertising.

By way of conclusion, I discuss some related models and which effects they can or cannot deliver. Previous work typically classifies advertising as either persuasive, i.e. changes consumer's tastes, or informative, i.e. creates information for the consumers.[41] Advertising for attention has elements of both these views but does not fall into either cleanly. While it holds that advertising can alter consumer preferences,

[41] A third view of advertising, the complementary view, does not fit in with this theory at all. For an excellent survey of the advertising literature, see Bagwell [2007].

as in the persuasive view, it changes preferences by altering the information that consumers process, not altering their tastes. While it links advertising with information, as in the informative view, it does not create information, instead changing the information which consumers process.

Persuasive advertising models are typically reduced form and can thus accommodate many of the effects generated here. For instance, the model studied by Friedman [1983] can generate positive price effect, positive spillovers and a positive primary demand effect, though the latter two effects must be directly assumed. However, the match effect cannot occur in models of persuasive advertising. Persuasive advertising is typically over-provided in equilibrium, e.g. Dixit and Norman [1978], whereas it may be under-provided in this model. It applies to either search goods or experience goods, whereas advertising for attention does not make sense in the context of search goods.

There are three classes of models of informative advertising that I will consider: revealing alternatives Grossman and Shapiro [1984], signaling Milgrom and Roberts [1986], and comparative Meurer and Stahl [1994]. I begin with Grossman and Shapiro [1984]'s classic model of advertising that reveals the existence of search goods. The type of goods for which the models are appropriate differ: Grossman and Shapiro consider only search goods, whereas my model applies only to experience goods. Consequently, many findings are reversed: advertising lowers price, it can only have negative spillover effects and it typically has little or no primary demand effect. As in this model, firms may either over- or under-provide advertising in equilibrium.

An influential strand of the literature studies advertising as a signal of quality, e.g. Nelson [1974] or Milgrom and Roberts [1986]. As in this model, it applies to experience goods. Higher advertising signals higher quality and thus leads to a higher price. However, the positive primary demand effect, the positive spillover effect, and the match effect do not occur in Milgrom and Roberts [1986].[42] As noted earlier, signaling requires that repeat purchase occurs.

Lastly, Meurer and Stahl [1994] analyze comparative advertising of experience goods. They analyze a duopoly model that generates some of the same effects as advertising for attention, namely the match effect, the positive spillover effect, and that advertising increases price. However, positive spillover appears to be an artifact of the duopoly setting. In an oligopoly setting, advertising can only increase the demand of the firms that are compared by the ad. Moreover, advertising can have either a positive or a negative primary demand effect. Comparative advertising only applies to advertisements that are comparative in nature and contain a good deal of information. Advertising for attention applies not only to this situation, but also to ads that mention only the advertiser and contain little or no information. As in this model, firms may either over- or under-provide advertising in equilibrium.

[42] Positive primary demand and positive spillover effects could probably be added to the model by assuming that consumers perceive a correlation between the quality of two products.

Chapter 3

Condorcet Meets Ellsberg

3.1 Introduction

When deciding how to vote, each individual may have private information about which of the two candidates will be better. Both the information itself and how others react to it affect how a rational voter casts her ballot. If each voter maximizes subjective expected utility (henceforth, SEU) and voters have common values, then there exists an equilibrium to the voting game in which all private information is revealed for a large enough electorate.[43] This result, known as the Condorcet Jury Theorem, provides an important efficiency justification for democracy as a political system. It describes conditions under which democracy is superior to even a benevolent dictatorship, since the probability of selecting the better policy is higher when an election rather than a privately informed dictator picks the policy.

This paper shows that when voters are ambiguity averse and their private information is ambiguous, there may not exist an equilibrium in which information aggregates, regardless of the size of the electorate. In fact, Theorems 7 and 9 show that for many voting games, *no equilibrium of the game aggregates information.* A rational voter conditions her action on the probability that her vote changes the outcome of the election. Consequently, each voter's equilibrium strategy may differ from

[43] For instance, see Austen-Smith and Banks [1996], Feddersen and Pesendorfer [1997, 1999], Myerson [1998] or Wit [1998] .

the action that her private information would suggest is best if she disregarded others' strategies. Given SEU, each vote noisily reveals private information, and with enough voters, information aggregates. When voters are ambiguity averse, each picks a voting strategy as if to insure herself against altering the outcome in favor of the worse candidate. Theorem 9 relates this behavior to an extreme "swing voter's curse": if others play a strategy profile that would aggregate information, each voter best responds by minimizing the probability that she casts a pivotal vote. She either plays a mixed strategy (Theorem 7) or abstains (Theorem 9). In any equilibrium, no vote reveals information, precluding aggregation.

A large literature, initiated by Ellsberg [1961], criticizes SEU on both normative and descriptive grounds. When payoffs depend on ambiguous events – that is, events about which the decision maker has only vague information – SEU does not accurately describe preferences. Agents typically prefer betting on unambiguous events to ambiguous ones. For example, a bet on an event E, which is known to occur with probability 0.5, may be preferred both to a bet on the event F and a bet on its complement F^c when no information is provided about F. Ambiguity aversion explains evidence from asset markets that contradicts SEU (for instance, see Epstein and Schneider [2010]).

Many important policy decisions are made under ambiguity.[44] A policy to cap carbon emissions deals with poorly understood costs, base case emissions, and tails of the probability distribution of temperature changes. The recession of 2008-2009 resulted at least in part from an unprecedented event (systematic default in AAA rated bonds) in the credit market. The Federal Reserve decided whether or not to

[44]Papers that address political economy questions with ambiguity averse voters or candidates include Berliant and Konishi [2005], Ashworth [2005], Ghirardato and Katz [2006] and Bade [2011], though none consider strategic interaction between voters.

bail out banks and hedge funds based on their beliefs about the poorly understood connection between this default, these companies and the financial system as a whole. Many foreign policy decisions must be made despite possessing only poor quality information, such as that leading to the 2003 invasion of Iraq.

To accommodate ambiguity averse voters, this paper assumes that voter preference conforms to maxmin expected utility (henceforth, MEU; introduced and axiomatized in Gilboa and Schmeidler [1989]). Voters consider a set of probability measures and evaluate an act by taking its minimum expected utility with respect to every measure in that set. Formally, for some set of probability measures Π and a von Neumann-Morgenstern index $u(\cdot)$, the utility of an act f can be written as

$$\min_{p \in \Pi} \mathbb{E}_p[u \circ f].$$

SEU is the special case when Π is singleton. When Π is not singleton, the behavior in the Ellsberg paradox can be rationalized.

Section 3.2 gives an example that illustrates how ambiguity averse voters behave differently from their SEU counterparts. Section 3.3 introduces ambiguous Poisson games and proves existence of an equilibrium. Section 3.4 describes a common values voting game when voter preferences are MEU and presents the paper's main results. Theorem 7 shows that ambiguity aversion can preclude the existence of any equilibrium that aggregates information. Theorem 8 provides sufficient conditions for existence of an equilibrium that aggregates information. Section 3.5 modifies the setup by allowing voters to abstain strategically. Theorem 9 shows that information may fail to aggregate in this setting as well. Section 3.6 concludes by relating the main results to other works that show failure of information aggregation in voting games. Proofs are collected in the remaining sections.

3.2 Sincere Voting and Ambiguity

This section offers a brief example showing how ambiguity aversion alters the set of equilibria to voting games. Formal definitions of the game and equilibrium are deferred to Section C.1.

Consider an election with 101 voters who vote for one of two candidates, A and B. The candidate with the most votes wins. Suppose there are two states of the world, a and b, and all voters agree that A's policy is better in state a but B's policy is better in state b. Before voting, all voters observe a signal from the set $\{1, 2\}$. They believe that signal 1 occurs with probability 0.6 in state a, that signal 2 occurs with probability 0.6 in state b, and that signals are independently distributed conditional on the state of the world. After observing signal t, each voter considers the set of posteriors Π_t consisting of the Bayesian updates of the probability measures in some set Π. Because the state space is one dimensional, it is convenient to represent Π_t and Π as intervals, $[\underline{p}_t, \bar{p}_t]$ and $[\underline{p}, \bar{p}]$ respectively, corresponding to the probability each of the measures in the set assigns to a. For simplicity, suppose that the interval $[\underline{p}, \bar{p}]$ is symmetric about $\frac{1}{2}$. Voters get utility equal to 1 if the correct candidate is elected but 0 otherwise. After observing signal t, voter preference is represented by

$$\min_{p \in [\underline{p}_t, \bar{p}_t]} [p(Pr(A \text{ wins}|a)) + (1-p)Pr(B \text{ wins}|b)]. \tag{3.2.1}$$

Because of the noted symmetry, a voter (strictly) prefers to bet on a over b if she observes signal 1 and vice versa if she observes signal 2. If all voters who observe 1 vote for A and all those who observe 2 vote for B, information aggregates. If voters were SEU ($\underline{p} = \bar{p} = \frac{1}{2}$), then McLennan [1998, Thm. 1] would show that this sincere voting strategy is an equilibrium. In that equilibrium, information aggregates and each voter receives the same expected utility in each state, about 0.979. However, when $\underline{p} < 0.4$ and $0.6 < \overline{p}$ sincere voting is not an equilibrium because all players best respond by voting for both A and B with equal probability.

For instance, assume that $\underline{p} = .39$ and $\overline{p} = .61$. After updating, players who observe signal 1 use $\Pi_1 = [0.49, 0.7]$ and players who observe signal 2 use $\Pi_2 = [0.3, 0.51]$. Consider the problem of an arbitrary voter when all others vote sincerely. If this voter observes signal 1, then she picks her vote to maximize

$$\min_{p \in [0.49, 0.7]} [pPr(A \text{ wins}|a) + (1-p)Pr(B \text{ wins}|b)]. \tag{3.2.2}$$

She affects the outcome only when she is pivotal, or when exactly 50 of the others vote for A. Since all others vote sincerely,

$$Pr(A \text{ has 50 votes}|a) = Pr(B \text{ has 50 votes}|b) = \binom{100}{50} .6^{50}.4^{50} = \rho,$$

which is approximately 0.01, and

$$Pr(\text{ 51+ votes for } A|a) = Pr(\text{ 51+ votes for } B|b) = \sum_{j=51}^{100} \binom{100}{j} .6^{j}.4^{100-j} = \theta,$$

which is approximately 0.973. If she votes for A with probability α, then

$$Pr(A \text{ wins}|a) = \theta + \rho\alpha$$

and

$$Pr(B \text{ wins}|b) = \theta + \rho(1-\alpha).$$

Therefore, this voter's utility from voting for A with probability α is

$$\min_{p \in [0.49, 0.7]} p[\alpha\rho + \theta] + (1-p)[(1-\alpha)\rho + \theta]. \tag{3.2.3}$$

If she voted sincerely, then she would always vote for A ($\alpha = 1$) and her utility would be

$$\min_{p\in[0.49,0.7]} p[\rho+\theta]+(1-p)\theta = \theta + .49\rho,$$

about 0.9779. If she played her other pure strategy, voting for B ($\alpha = 0$), then she would get utility

$$\min_{p\in[0.49,0.7]} p\theta+(1-p)[\theta+\rho] = \theta + .3\rho,$$

about 0.976 which is less than if she voted for A.

When the voter picks her strategy, the state of the world is determined but unknown. By randomizing, she replaces subjective uncertainty with objective risk. While she prefers to follow her signal rather than vote against it, voting for A and B with equal probability insures her against ambiguity. By doing so, she receives utility equal to

$$\min_{p\in[0.49,0.7]} p[\theta+.5\rho]+(1-p)[\theta+.5\rho] = \theta + .5\rho,$$

about 0.978, so she prefers this mixture to sincere voting. A symmetric argument shows that the voter also prefers to mix in this way after observing signal 2. Hence, her best response is to randomize between voting for A and B regardless of the signal she observes.

As in the SEU case, each voter picks her strategy based on her "beliefs" about the state of the world if her vote decides the election. If all voters were SEU, then each vote would reveal something about the voter's private information, and as the number of voters approached infinity, the outcome of the election would reflect all private information. In contrast, in the example the voter minimizes the probability that she makes a mistake (conditional on her being pivotal) by randomizing between voting for A and B. She thinks that if she is pivotal, she will make a mistake with probability as high as 0.51 by voting for A or as high as 0.7 by voting for B. By mixing, she makes a mistake with precisely probability 0.5. Because the voter is ambiguity averse, she strictly prefers the latter strategy. Should the whole electorate play this strategy, information could not aggregate because no individual's vote reveals the underlying signal. Indeed, all voters randomizing as above is an equilibrium to this game.

That sincere voting fails to be an equilibrium is not in itself surprising; in fact, Austen-Smith and Banks [1996] show this is typically the case even with SEU voters. However, Theorem 7 below shows that there is *no* equilibrium to the above game in which information aggregates: if σ is an equilibrium where the expected winner in state a is A, then the expected winner in state b is *not* B.[45] Theorem 7 extends the logic above to any strategy profile. If information would aggregate should voters play strategy profile σ, then some voter prefers to insure herself rather than follow her prescribed strategy. Consequently, σ cannot be an equilibrium.

3.3 Ambiguous Poisson Games

This section introduces ambiguous Poisson games, a generalization of Myerson [1998]'s notion of *extended Poisson games*. Extended Poisson games simplify the analysis of large population games with some underlying uncertainty. Myerson proves that if an extended Poisson voting game has a common prior, common values and informative signals, then there exists an equilibrium in which information aggregates. The notation and definition of equilibrium are adapted from Myerson [1998]. Theorem 6 proves existence of an equilibrium.

[45] This paper's results are stated for games with Poisson population uncertainty, but only Theorem 9 relies on this assumption. Theorems 7 and 8 hold without population uncertainty. Details available from the author upon request.

For any finite set E, denote by ΔE the set of probability measures on E.

Definition. *An* ambiguous Poisson game Γ *is a collection* $(\Omega, C, T, U, (\Pi_t)_{t \in T}, r, n)$ *where:*

- Ω *is a finite set of states.*
- C *is a finite set of actions. Define* $Z(C) = \{x \in \mathbb{R}^C : x(c) \in \mathbb{N} \forall c \in C\}$*, the set of all possible realized action profiles (the number of players taking each action).*
- T *is a finite set of types.*
- $U : T \times C \times \Omega \times Z(C) \to \mathbb{R}$ *is a bounded function that represents preference.* $U(t, c, \omega, x)$ *is the utility for a voter of type* t *who takes action* c *when the realized state is* ω *and the realized action profile is* x*.*
- $\Pi_t \subset \Delta(\Omega)$ *is a closed, non-empty and convex set, representing the set of posteriors for each type. If* Π_t *is a singleton for every* t*, then all players are SEU, though they may have different priors.*
- $r : \Omega \to \Delta T$ *maps each state to a probability measure over types. Types are drawn independently according to* $r(\omega)$ *in state* ω*.*
- *The number of players is a random variable distributed Poisson with mean* $n \in \mathbb{R}_{++}$*.*

The timing of the game is as follows. Nature chooses the number of players according to the Poisson distribution with mean n and chooses the state of the world according to some unknown, unmodeled procedure. Each player learns her type and forms a set of posteriors.[46] Before learning the realized state, how many other players there are or what actions the other players have taken, she picks a strategy $s \in \Delta C$. When she picks this strategy, the state of the world is realized but unknown,

[46]Note that posterior beliefs rather than prior beliefs are taken as a primitive. One could specify a set of priors and an updating rule (in the example from Section 3.2, the updating rule is prior-by-prior Bayesian updating), which would constitute a special case of the above. However, there are no ex-ante actions so the set of priors only enters a voter's decision through her set of posteriors.

so ambiguity aversion leads each player to act as if Nature picked the distribution over states with the goal of minimizing her utility. A mixed strategy may equalize her expected utility across states, limiting her exposure to Nature's choice. For this reason, she may find a mixed strategy to be the only best response. For a more in depth discussion of this issue see Lo [1996] or Klibanoff [1996].

As in Myerson [1998], assuming a Poisson population yields convenient properties. Because types are conditionally independent and the population is distributed Poisson, the number of players that take each action c in state ω is also distributed Poisson and is independent of the number of players taking action $c' \neq c$ in state ω. Moreover, each player's conditional expectation does not depend on her private information. If $\lambda(\omega)(c)$ is the expected number of players in state ω that take action c, the probability of any given action profile x in state ω is $p(x|\lambda(\omega))$ where

$$p(x|\lambda) = \prod_{c \in C} \frac{e^{-\lambda(c)} \lambda(c)^{x(c)}}{x(c)!}. \tag{3.3.1}$$

These properties imply that the best response correspondence is the same for any two players with the same type. A *strategy profile* σ is a map from types to strategies, $\sigma : T \to \Delta(C)$. A player of type t picks a strategy $\sigma_t \in \Delta C$ to maximize

$$V_t(\sigma_t, \sigma) = \min_{q \in \Pi_t} \int_\Omega \int_{Z(C)} \sum_{c \in C} \sigma_t(c) U(t, c, \omega, x) dp(x|\lambda(\omega)) dq(\omega) \tag{3.3.2}$$

where

$$\lambda(\omega)(a) = n \sum_{t \in T} \sigma(t)(a) r(t|\omega). \tag{3.3.3}$$

Definition. *A strategy profile σ^* is an equilibrium for Γ if for each $t \in T$*

$$\sigma^*(t) \in \arg \max_{\hat{\sigma} \in \Delta C} V_t(\hat{\sigma}, \sigma^*). \tag{3.3.4}$$

If σ^* is an equilibrium, then every player picks her strategy to maximize the minimum expected utility over all measures in her set of posteriors, given she knows that the other players follow the strategy profile σ^*. When Π_t is singleton for all $t \in T$ this definition is equivalent to the definition in Myerson [1998]. Because each player maximizes her minimum expected utility given her beliefs and all player's beliefs agree, the behavior of each player is as in Lo [1996]'s "beliefs equilibrium with agreement." While he does not consider games with population uncertainty, this definition of equilibrium otherwise coincides with his.

Theorem 6. *For any ambiguous Poisson game Γ, there exists a strategy profile σ^* that is an equilibrium for Γ.*

3.4 The Condorcet Jury Theorem

This section describes common values voting games with MEU players and discusses the limiting equilibria. Theorem 7 establishes the existence of voting games for which no equilibrium aggregates information. Theorem 8 shows that information aggregates in equilibrium for some voting games where no voter is SEU. Neither of these two results depends on population uncertainty – very similar arguments work when the population is fixed at n. Details are available upon request.

3.4.1 Ambiguous voting games

Candidates A and B each commit to a distinct policy. Voters cast a vote for one of them, and the candidate with the most votes wins; in a tie, each candidate is selected with equal probability. Voters have common values and are instrumentally rational: they care only about the policy outcome and they have the same preference over policies given the state. Depending on the state of the world, the policy is either good or bad. There are two states, a and b, representing which policy is the good one.

Formally, an *ambiguous voting game* is an ambiguous Poisson game where the action set is $C = \{A, B\}$, the set of states is $\Omega = \{a, b\}$ and the utility function of all types takes value 1 if the candidate elected matches the state and 0 otherwise. The action A is interpreted as a vote for candidate A, B is interpreted as a vote for B and the set of types T is interpreted as a set of signals. Given that others play strategy profile σ, the payoff to a voter of type t using strategy $\hat{\sigma} \in \Delta\{A, B\}$ is

$$\begin{aligned} V_t(\hat{\sigma}, \sigma) \quad = \quad & \min_{\pi \in \Pi_t} \{\pi(a)[\hat{\sigma}(A) Pr(A \text{ wins}|a, v_A, \sigma) + \hat{\sigma}(B) Pr(A \text{ wins}|a, v_B, \sigma)] + \\ & + \pi(b)[\hat{\sigma}(A) Pr(B \text{ wins}|b, v_A, \sigma) + \hat{\sigma}(B) Pr(B \text{ wins}|b, v_B, \sigma)]\}, \end{aligned}$$

where $Pr(c \text{ wins}|\omega, v_d, \sigma)$ is the probability candidate c wins in state ω if she votes for candidate d and others play strategy profile σ.

As in Section 3.2, represent Π_t by the interval of probabilities that the measures within it assign to a. That is, $\Pi_t \equiv [p_t, q_t]$ where $p_t = \min_{\rho \in \Pi_t} \rho(a)$ and $q_t = \max_{\rho \in \Pi_t} \rho(a)$.

3.4.2 Main result

This subsection describes a set of ambiguous voting games for which no equilibrium aggregates information. Theorem 7 below shows that if voters lack confidence, then no equilibrium aggregates information. Voters lack confidence when the following condition on posteriors holds.

Definition. *An ambiguous voting game* has voters who lack confidence *if*

$$p_t < \frac{1}{2} < q_t$$

for all $t \in T$.

An outsider can detect when voters lack confidence through betting preferences.

If voters lack confidence, then all voters strictly prefer betting on the outcome of a fair coin toss over betting on either a or b. Even if the voter thinks that a is a better bet than b, she lacks confidence in this judgment and strictly prefers to hedge her bet on a by mixing it with a bet on b. This is impossible with SEU: if a is at least as likely as b when a and b are the only two states, then a bet on a is at least as good as a fifty-fifty lottery.

This translates into the voting setting as follows. Suppose a random voter were made a dictator – whichever policy she chooses will be implemented. If, irrespective of the signal she receives, she strictly prefers to pick the policy implemented by flipping a fair coin rather than implementing either policy for sure, then, and only then, voters lack confidence.

To give a better sense of the meaning of lacking confidence, suppose that voters form posterior beliefs by updating a common set of priors Π using prior-by-prior Bayesian updating. Precision of signals and the set of priors both contribute to posterior beliefs. Voters lack confidence when the signals do not provide enough information to offset the prior ambiguity. With very precise signals (there is some t so that $\frac{r(t|b)}{r(t|a)}$ is very high or very low), Π must be very close to $[0, 1]$ for voters to lack confidence; however, if signals are not very precise ($\frac{r(t|b)}{r(t|a)}$ close to one for all t), then Π can be a much smaller interval. For instance, with the signal structure described in Section 3.2, voters lack confidence whenever $[.4, .6] \subset \Pi$ without equality, but if $r(1|a) = r(2|b) = .51$, then voters lack confidence whenever $[.49, .51] \subset \Pi$ without equality. If $\Pi = [.45, .55]$, then voters lack confidence given the second signal structure but not the first.

Theorem 7. *Suppose that Γ is an ambiguous voting game with voters who lack confidence. If σ is an equilibrium for Γ in which the expected vote share for A in state a is greater than $\frac{1}{2}$, then the expected vote share for B in state b is less than $\frac{1}{2}$.*

Theorem 7 implies that the equilibrium probability of the correct candidate winning the election is bounded above by $\frac{1}{2}$ in at least one state, prohibiting information aggregation. The following outlines the proof, which is by contradiction.

Suppose, for the sake of contradiction, that there is an equilibrium strategy profile σ where the expected winner is correct in both states of the world. The key is to show that the worst case scenario (the state in which the wrong candidate is more likely to be elected) is not independent of the voter's strategy when the others play σ. If the worst case scenario were independent, then the voter acts as if maximizing SEU according to the posterior that maximizes the probability of the worst case scenario and familiar arguments (for instance, Myerson [1998, Thm. 2]) would imply that there is an equilibrium in which information aggregates when n is large.

To see why the worst case scenario depends on the voter's strategy, suppose that it doesn't, that n is "large" and that σ calls for all voters to play a pure strategy. Since the ratio of pivot probabilities must not go to either 0 or 1, Myerson [2000, Thm. 1] shows that the expected vote share for A in state a is the same as the expected vote share for B in state b. As a consequence, each voter thinks that if she abstains, then her conditional expected utility is equal across states. When a player votes for A for sure, her expected utility conditional on state a increases and her expected utility conditional on state b decreases (vice versa when voting for B). It follows that the worst case scenario depends on her vote. When voters lack confidence, this argument can be extended to any σ, regardless of n.

Because the worst case scenario depends on her vote, there is a mixed strategy that insures the voter against making a mistake and altering the election in favor of the wrong candidate when others play σ, as in Section 3.2. Since voters lack confidence, each voter weakly prefers to play this mixed strategy over any other strategy. If every voter insured herself, then information could not aggregate because this insurance strategy is independent of private information. Therefore, it must be that some voter is willing to play a different strategy. However, the only strategies that are at least as good as the insurance strategy assign higher probability to voting for the candidate that receives more votes from the insurance strategy. All voters expect to vote for the same candidate regardless of signal. This candidate is the expected winner in both states, a contradiction.

The following result characterizes one equilibrium to the game.

Proposition 11. *If an ambiguous voting game* Γ *has voters who lack confidence, then the strategy profile* σ *defined by* $\sigma(t)(A) = \frac{1}{2}$ *for all* $t \in T$ *is an equilibrium for* Γ.

In this equilibrium, both candidates are elected with equal probability regardless of the state. Therefore, knowing the winner of the election would not change the beliefs of a Bayesian agent. Neither Proposition 11 nor Theorem 7 show that this is the only equilibrium. However, Theorem 7 shows that if an equilibrium results in a higher probability of electing the correct candidate than this equilibrium in one state of the world, then it must result in a lower probability of electing the correct candidate in the other state of the world.

3.4.3 Information aggregation

This subsection provides a formal definition of information aggregation and proves that some ambiguous voting games have an equilibrium in which information aggregates. Because there is always some possibility of a mistake in a finite electorate, one cannot require full certainty that voters elect the proper candidate in a given game. Instead, the literature focuses on sequences of voting games where the probability of electing the wrong candidate vanishes along some sequence of equilibria. Below, a sequence of ambiguous voting games is indexed by the mean number of players, with all other primitives remaining the same.

Definition. *A sequence of ambiguous voting games* $(\Gamma_n)_{n=1}^{\infty}$ *satisfies* full information equivalence (FIE) *if there exists a sequence of strategy profiles* $(\sigma_n)_{n=1}^{\infty}$ *so that* σ_n *is an equilibrium for* Γ_n*, and for any* $\epsilon > 0$*, there exists* N *so that* $n > N$ *implies that the correct candidate is elected in each state with probability higher than* $1 - \epsilon$ *when* σ_n *is played.*[47]

An implication of Theorem 7 is that FIE fails for many sequences of ambiguous voting games. In contrast, as long as the signal structure is informative (the conditional distribution of signals varies with the state), any sequence of SEU voting games satisfies FIE. Since SEU is a special case of MEU, some ambiguous voting games satisfy FIE. However, SEU is not necessary for information to aggregate. In fact, Theorem 8 proves the existence of a sequence of equilibria that aggregates information whenever the game has disjoint* posteriors.

Definition. *An ambiguous Poisson game has* disjoint* posteriors *if for any distinct* t *and* t' *in* T *either* $p_{t'} = q_t$*,* $p_t = q_{t'}$ *or* $[p_t, q_t] \cap [p_{t'}, q_{t'}]$ *is empty.*

If all voters are SEU, then each Π_t is a singleton and the ambiguous Poisson game has disjoint* posteriors. More generally, one can distinguish between SEU, disjoint* posteriors and voters who lack confidence using Lemma 28. Consider an ambiguous voting game Γ. If Γ has singleton posteriors, then all voters act as SEU maximizers and none strictly prefer to randomize for any strategy profile. If Γ has disjoint* posteriors, then for any strategy profile at most one type of voter strictly prefers to randomize. If Γ has voters who lack confidence, then there exists a strategy profile such that all voters strictly prefer randomizing to playing a pure strategy.

[47] This definition is adapted from Feddersen and Pesendorfer [1997].

Theorem 8. *Suppose that* $(\Gamma_n)_{n=1}^{\infty}$ *is a sequence of ambiguous voting games that have disjoint* posteriors, that* $0 < p_t \leq q_t < 1$ *for all* t *and that for each* ω *and* t, $r(t|\omega) > 0$. *If there is some* $t \in T$ *s.t.* $r(t|a) \neq r(t|b)$, *then* $(\Gamma_n)_{n=1}^{\infty}$ *satisfies FIE.*

The proof generalizes the construction from Myerson [1998, Thm. 2]. As in that paper, the equilibrium consists of a "step strategy": at most one type of voter randomizes, and all others play a pure strategy, determined by how likely they view a relative to the randomizing voter. Because of disjoint* posteriors, at most one type of voter has a strict preference for randomization. The proof shows that along this sequence, even if some voter strictly prefers to randomize for every element of the sequence, the strategy will be the same as in the SEU game with the same signal structure at the limit. In fact, Myerson [1998, Thm. 2] is the special case where each Π_t is a singleton that results from Bayesian updating of a common prior.

3.5 Strategic Abstention

In SEU voting games, abstention typically improves the outcome of the election. This is due to the "swing voter's curse" (introduced in Feddersen and Pesendorfer [1996]): uninformed voters are more likely to abstain than informed voters. As a consequence, the expected number of votes for the correct candidate is larger than if voters could not abstain, so abstention improves the expected outcome of the election. The ambiguous voting games studied in Section 3.4 explicitly rule out the possibility of strategic abstention, leaving open the possibility that the conclusion of Theorem 7 fails when voters can choose to abstain.

This section will show that a version of Theorem 7 holds without mandatory voting. The analysis provides insight into the mechanism behind Theorem 7; namely, equilibrium behavior can be interpreted as an extreme swing voter's curse. Each voter prefers to minimize the chance that she casts a pivotal vote. If she abstained, then she would never be pivotal, which would be better than any available strategy. However, Theorem 7 assumes that she must vote. Among her available choices, her best option is to mimic abstention through a mixed strategy.

In order to allow for abstention, modify the ambiguous voting games from Section 3.4 by replacing the action set with $C = \{A, B, \emptyset\}$ and requiring that $T = \{1, 2\}$. The action $\emptyset$ corresponds to abstention. The payoffs for each voter are as in the previous section. The restriction to two types is for simplicity. Call such a game an *ambiguous voting game with abstention.*

Say that an ambiguous Poisson game has *symmetric signals* if $r(1|a) = r(2|b)$ and that players *have posteriors that respect likelihood ratios* if $\frac{r(t|a)}{r(t|b)} > \frac{r(t'|a)}{r(t'|b)}$ implies that $\min_{\pi\in\Pi_t} \pi(a) \geq \min_{\pi\in\Pi_{t'}} \pi(a)$ *and* $\max_{\pi\in\Pi_t} \pi(a) \geq \max_{\pi\in\Pi_{t'}} \pi(a)$ for every $t, t' \in T$. In ambiguous voting games with abstention satisfying these two assumptions, information does not aggregate along any sequence of equilibria.

Theorem 9. *Suppose that* $(\Gamma_n)_{n=1}^{\infty}$ *is a sequence of ambiguous voting games with abstention and symmetric signals. If voters lack confidence and have posteriors that respect likelihood ratios, then* $(\Gamma_n)_{n=1}^{\infty}$ *does not satisfy FIE.*

For a sequence of ambiguous voting games with abstention to satisfy FIE, it is necessary that there exists a sequence of equilibrium strategy profiles where the winner is correct in both states and the expected number of votes in each state goes to infinity. The proof of Theorem 9 adapts and extends the arguments from Theorem 7 to show that either the number of votes is bounded above in some state or the expected winner is incorrect in at least one state. As a consequence, FIE must fail.

With SEU voters, Feddersen and Pesendorfer [1999] and Bouton and Castanheira [2009] show the swing voter's curse persists in voting games similar to those considered here. Given two SEU voters who observe different signals, the voter whose signal conveys less information about the state of the world is more likely to abstain in equilibrium. As a consequence, for a fixed signal structure, the percentage of votes cast by more informed voters is higher in an election with abstention compared to one with mandatory voting. In contrast, Theorem 9 demonstrates that ambiguity aversion strengthens the swing voter's curse. An ambiguity averse swing voter perceives the probability of making a mistake with her vote to be larger than her SEU counterpart. Allowing abstention leads to fewer votes in expectation but, unlike SEU, may not change the composition of the votes when voters lack confidence.

Proposition 12. *If* Γ *is an ambiguous voting games with abstention that has voters who lack confidence, the strategy profile* σ^* *defined by* $\sigma^*(t)(\emptyset) = 1$ *for every* $t \in \{1, 2\}$ *is an equilibrium for* Γ*.*

Unlike the equilibrium shown to exist in Proposition 11, all voters play a pure strategy. When allowing for abstention, the equilibrium from Proposition 11 still exists. However, both equilibria lead to the same distribution of outcomes in both states, so the equilibrium payoffs are the same for each voter, as is the information that the outcome provides to an observer.

This result contrasts with Propositions 2 and 3 of Feddersen and Pesendorfer [1996] and Propositions 5 of Feddersen and Pesendorfer [1999]. In these papers, the fraction of voters who don't abstain remains bounded away from zero along any sequence of equilibria. This result is a consequence of SEU preferences: even a small difference in the expected benefits of voting for A instead of B induces a strict preference to vote for A.

3.6 Conclusion

Theorems 7 and 9 show that rational but ambiguity averse voters may find it optimal to insure themselves by minimizing the chance they cast a pivotal vote. This mechanism leads to a failure of information aggregation not documented by previous work. These papers show that the dimensionality of the uncertainty and the degree of commonality between voters are important in evaluating the efficiency of the election. In contrast, this paper suggests that how familiar the electorate is with the issues at stake also matters a good deal. By way of conclusion, this section reviews some of these results and contrasts them with Theorems 7 and 9.

Feddersen and Pesendorfer [1997] prove that if the distribution of preferences is unknown, then FIE fails generically. The problem is one of dimensionality; namely, each voter must infer both the distribution of signals and the distribution of preferences from these votes. Even if a voter knew which votes others cast and the electorate were large, she could not infer the state of the world. In contrast, this paper assumes common knowledge of the distribution of preferences. However, the distribution of votes may not vary with the state (see Proposition 11 or 12) because voters insure themselves against ambiguity by abstaining or randomizing.

Mandler [2011] shows that if the conditional distribution of signals is unknown, then FIE may fail. If all the signals were observed by each voter, then uncertainty would remain as to which state is correct even as the size of the electorate goes to infinity. In this paper, if all signals were observed, then the true state would be known with probability approaching 1 despite the prior ambiguity.

Bhattacharya [2008] drops the assumption of common values and characterizes the distributions of preferences for which FIE fails. For instance, FIE fails when any voter who receives information in favor of the Condorcet-winner with perfect information is very likely to strongly prefer the other candidate.[48] In contrast, this paper maintains pure common values.

Finally, the result in this paper relates to work that studies the effect of ambiguous information in other contexts. For instance, Condie and Ganguli [2011] demonstrates a failure of information transmission with ambiguity averse agents in general equilibrium. They show that a rational expectations equilibrium for an exchange economy may be partially revealing when agents are ambiguity averse; in contrast, fully revealing equilibria are generic with SEU agents. Two differences are worth pointing out. First, in their model agents do not act strategically – they are price takers. Second, they assume that only a subset of agents are ambiguity averse, while an ambiguous voting game has voters who lack confidence only if all voters are ambiguity averse.

[48] Additionally, the non-aggregation result in this paper is stronger because of his more demanding definition of FIE, which requires the definition of FIE from this paper to hold for *every* sequence of symmetric, Bayesian Nash equilibria in undominated strategies. Unlike Theorem 7, his conditions do not rule out the existence of a different equilibrium in which information would aggregate. For example, the game depicted by his Figure 1 fails his definition of FIE but satisfies the definition in this paper.

Appendix A

Appendix for Chapter 1

A.1 Preview of the Proof of Theorem 1

This Section discusses the key idea behind the proof of Theorem 1. The key idea of the proof is to map choices of acts onto a larger domain where it is suitably well-behaved. In particular, I consider the space of "plans." A plan is a mapping from each state to an act. Each set of conditional choices from a given menu defines one plan. In the example from Section 1.1.2, the modeler observes the doctor's conditional choices of acts. Instead of looking at her each of her conditional choices in isolation, one can think of them as choosing one plan. For instance, the doctor chooses the plan "pick g in state γ, otherwise pick m" from $\{g, m\}$ and chooses the plan "pick f in state ϕ, otherwise pick m" from $\{g, m, f\}$.

Although choice in a given state may violate WARP, INRA guarantees that her choice over plan maximizes a preference relation (whose domain is plans rather than acts). However, this preference relation may be discontinuous, incomplete and intransitive. Given the other axioms, one can extend it to a well-behaved preference relation so that the DM's choices are a maximal element of this preference relation.[49] I then show that this preference over plans can be represented as expected utility over a subset of "feasible" plans. I identify a candidate for $\mathbb{P}^*$ and show that any

[49]Although this is an extension, it is typically not a "compatible extension" in the sense that it may not preserve strict preference.

plan measurable with respect to some $Q \in \mathbb{P}^*$ is feasible. The coarsest partition that measures her chosen plan is identified as the DM's subjective information. This subjective information is optimal, in the sense that it maximizes expected utility according to the utility index and prior representing the preference over plans. The final step shows that her conditional choices can also be represented as maximizing this preference relation.

This suggests an alternative domain on which optimal inattention admits foundations: preference over plans. In supplementary material, I show that one can use preference over plans to derive both optimal inattention and costly attention representations. Observing choice of plan is more convenient because it requires observing a single ex ante choice rather than choices in each state of the world. It has also been used in applications (for instance, Gul et al. [2011]). However, this has some significant drawbacks. First, choice of plan is difficult to observe outside of a laboratory. Second, choice of plan typically reflects both constraints and true preference. Third, what a DM plans to choose may differ from what she actually chooses. Fourth, economic objects of interest are conditional choices, not ex ante choices. Therefore, I focus in the main paper on choice of acts. This data is closer to what economists typically work with, and reflects the DM's response to whatever constraints she faces rather than those she thinks that she will face. Moreover, if the DM follows through with her choice of plan, then her final conditional choices of acts satisfy the axioms from Section 1.3.

A.2 Proofs

A.2.1 Proofs from Section 1.3

Proof of Theorem 1:

Proof. If $c(B|\omega) = B$ for all $B \in K(X)$, then by Monotonicity, it follows that $c(B|\omega) = B \forall B \in K(\mathcal{F})$. Taking $\mathbb{P}^* = \{\Omega\}$ and $u(x) = 0 \forall x$ establishes the desired result. Therefore, assume that there are $x^*, x_* \in X$ and ω so that $x_* \notin c(\{x^*, x_*\}|\omega)$.

Lemma 3. *There exists an affine, continuous $u : X \to \mathbb{R}$ so that for any $B \in K(X)$, $x \in c(B|\omega) \iff u(x) \geq u(y)$ for all $y \in B$.*

Proof. Fix any $A \subset B \in K(X)$. By Monotonicity, if $x \in c(B|\omega)$, then $x \in c(B|\omega')$ for any ω'. By INRA, $c(B|\omega) \cap A \neq \emptyset$ implies that $c(B|\omega) \cap A = c(A|\omega)$, i.e. $c(\cdot|\omega)$ satisfies WARP when restricted to problems of lotteries. It is routine to verify that the resulting revealed preference relation satisfies the hypothesis of Grandmont [1972, Thm 2] and therefore an affine, continuous $u : X \to \mathbb{R}$ exists. □

Let $\mathcal{F}^\Omega$ be the set of functions from Ω to $\mathcal{F}$ that are $\sigma(P)$ measurable. I refer to elements of $\mathcal{F}^\Omega$ as "plans" with the interpretation that the DM chooses $F(\omega)$ in state ω. Since P is finite, any $F \in \mathcal{F}^\Omega$ is simple. I denote elements of X by $x, y, z, ...$, elements of $\mathcal{F}$ by $f, g, h, ...$ and elements of $\mathcal{F}^\Omega$ by $F, G, H, ...$ Identify X with the subset of $\mathcal{F}$ that does not vary with the state and $\mathcal{F}$ with the subset of $\mathcal{F}^\Omega$ that does not vary with the state, so $X \subset \mathcal{F} \subset \mathcal{F}^\Omega$.

Denote by $\mathcal{A}(\cdot)$ the partition defined as $\hat{\hat{P}}(\cdot)$ in Equation (1.4.1) and define $\hat{c} : K(\mathcal{F}) \twoheadrightarrow \mathcal{F}^\Omega$ by $F \in \hat{c}(B) \iff F(\omega) \in c(B|\omega)$ for every ω and $\sigma(F) \subset \sigma(\mathcal{A}(B))$. Since $\sigma(\mathcal{A}(B)) \subset \sigma(P)$, any $\sigma(\mathcal{A}(B))$-measurable selection from $c(B|\cdot)$ is in $\mathcal{F}^\Omega$. For any $F \in \mathcal{F}^\Omega$, define $\{F\} \in K(\mathcal{F})$ by $\{F\} = \{F(\omega) : \omega \in \Omega\}$. Since F is simple, $\{F\}$ is finite and compact. By INRA, if $F \in \hat{c}(B)$ then $F \in \hat{c}(\{F\})$.

Define $\mathcal{C} \subset \mathcal{F}^\Omega$ by $F \in \mathcal{C} \iff F \in \hat{c}(\{F\})$; $\mathcal{C}$ is the set of plans that the DM chose from some problem. Define a binary relation $\hat{\succeq}$ on $\mathcal{C}$ by

$$F \hat{\succeq} G \iff \{F\} \, \overline{IS} \, \{G\}.$$

Note that $F \in \hat{c}(B) \implies F \hat{\succeq} G$ for every $G \in \mathcal{C}$ so that $\{G\} \subset B$. For any $F \in \mathcal{F}^\Omega$, define $F^* \in \mathcal{F}$ to be $F^*(\omega) = F(\omega)(\omega)$.

Lemma 4. *If $F \in \hat{c}(\{F\})$, then $F^* \tilde{\sim} F$.*

Proof. Assume that $F = \hat{c}(\{F\})$ and define $\hat{F}$ so that

$$\hat{F}(\omega) = F(\omega)\mathcal{A}(\{F\})(\omega)\underline{x}$$

for some $\underline{x} \in X$ so that $u(\underline{x}) \leq u(F(\omega))\forall\omega$. It holds that $\hat{F} \in \hat{c}(\{F\} \cup \{\hat{F}\})$. To see this, note that

$$u \circ \hat{F}(\omega) \leq u \circ F(\omega)$$

for all ω so it must be that $F \in \hat{c}(\{F\} \cup \{\hat{F}\})$ by monotonicity and INRA. Therefore, $\mathcal{A}(\{F\} \cup \{\hat{F}\}) \gg \mathcal{A}(\{F\})$. For any ω, so that $\mathcal{A}(\{F\} \cup \{\hat{F}\})(\omega) \subset \mathcal{A}(\{F\})(\omega)$ so since

$$F(\omega)(\omega') = \hat{F}(\omega)(\omega')\forall\omega' \in \mathcal{A}(\{F\})(\omega),$$

it follows from Subjective Consequentialism and $F(\omega) \in c(\{F\} \cup \{\hat{F}\}|\omega)$ that $\hat{F}(\omega) \in c(\{F\} \cup \{\hat{F}\}|\omega)$. By INRA, $\hat{F} = \hat{c}(\{\hat{F}\})$. Further, if $F^* \tilde{\sim} \hat{F}$, it follows that $F^* \tilde{\sim} F$.

By construction, $\mathcal{A}(\{\hat{F}\})$ is finer than $\mathcal{A}(\{F^*\})$. By ACI,

$$\alpha\hat{F} + (1-\alpha)F^* \in \hat{c}(\alpha\{\hat{F}\} + (1-\alpha)\{F^*\})$$

for all $\alpha \in [0, 1]$. Further

$$u \circ \alpha\hat{F}(\omega) + (1-\alpha)F^* \geq u \circ \hat{F}(\omega)$$

by construction. Therefore, when $B_\alpha = (\alpha\{F^*\} + (1-\alpha)\{\hat{F}\}) \cup \{\hat{F}\}$

$$\alpha\hat{F} + (1-\alpha)F^* \in \hat{c}(B_\alpha)$$

by INRA. Since $\mathcal{A}(B_\alpha) \gg \mathcal{A}(\{\hat{F}\})$ for any $\alpha \in [0, 1)$ and for any ω,

$$[\alpha F^* + (1-\alpha)\hat{F}(\omega)](\omega') = \hat{F}(\omega)(\omega')$$

for all $\omega' \in \mathcal{A}(\{\hat{F}\})(\omega)$, by Subjective Consequentialism

$$\hat{F} \in \hat{c}(B_\alpha).$$

Therefore, for $n \in \{1, 2, ...\}$, $\{F\}$ IS $\frac{n-1}{n}\{F^*\} + \frac{1}{n}\{\hat{F}\}$, which goes to $\{F^*\}$, and by definition $F \hat{\succeq} F^*$. Since $F^* \in \hat{c}(\{F^*\} \cup \{\hat{F}\})$, it follows that $F^* \hat{\succeq} F$; combining yields $F \hat{\sim} F^*$. □

Lemma 5. *For any $h \in \mathcal{F}$, $\alpha \in [0,1]$ and $A, B \in K(\mathcal{F})$, if $A \, \overline{IS} \, B$, then $\alpha A + (1-\alpha)\{h\} \, \overline{IS} \, \alpha B + (1-\alpha)\{h\}$.*

Proof. There are sequences $(A_n)_{n=1}^{\infty}$ and $(C_n)_{n=1}^{\infty}$ that converge to A and B respectively where A_n IS C_n.

Consider an arbitrary n and the finite sequence $B_1, ..., B_m$ so that $A_n = B_1$ and $B_m = C_n$ and $c(B_i|\omega) \cap B_{i-1} \neq \emptyset$ for every ω. Since $\mathcal{A}(B_i) \gg \mathcal{A}(\{h\})$, $c(\alpha B_i + (1-\alpha)\{h\}|\omega) \cap [\alpha B_{i-1} + (1-\alpha)\{h\}]$ for every ω by ACI. Since $\alpha A_n + (1-\alpha)\{h\} = \alpha B_1 + (1-\alpha)\{h\}$ and $\alpha B_m + (1-\alpha)\{h\} = \alpha C_n + (1-\alpha)\{h\}$, $\alpha A_n + (1-\alpha)\{h\}$ IS $\alpha C_n + (1-\alpha)\{h\}$.

Since n was arbitrary, we can do this for all n. Note that $\alpha A_n + (1-\alpha)\{h\} \to \alpha A + (1-\alpha)\{h\}$ and $\alpha C_n + (1-\alpha)\{h\} \to \alpha B + (1-\alpha)\{h\}$, it follows that $\alpha A + (1-\alpha)\{h\} \, \overline{IS} \, \alpha B + (1-\alpha)\{h\}$. □

Lemma 6. *$\hat{\succeq}$ is transitive.*

Proof. Suppose $F \hat{\succeq} G$ and $G \hat{\succeq} H$. Set $A = \{F\}$, $B = \{G\}$ and $C = \{H\}$. Then $F, G, H \in \mathcal{C}$, $A \, \overline{IS} \, B$ and $B \, \overline{IS} \, C$.

Then there are sequence A_n, B_n, B'_n, C_n that converge to A, B, B, C respectively so that $A_n \, IS \, B_n$ and $B'_n \, IS \, C_n$. Pick $G_n \in \hat{c}(B'_n)$, noting that $\{G_n\} \, IS \, C_n$ by INRA. Let y be the worst outcome of any act in B. Let $z \in X$ be so that $u(y) - u(z) = k > 0$ (if such an outcome does not exist, replace each problem by mixing it with x^*). Pick $F_n \in \hat{c}(B'_n)$ for every n. Note that $\{F_n\}$ IS C_n for every n using INRA.

Because $u(\cdot)$ is continuous, for any $\epsilon > 0$ there is a $\delta(\epsilon)$ so that $d(x, x') < \delta(\epsilon)$ implies that $|u(x) - u(x')| < \epsilon$. Therefore, for any ϵ there is an $n(\epsilon)$ so that for any every $n > n(\epsilon)$ and any act $f' \in \{G_n\}$ there is an act $f \in B_n$ so that $u(f(\omega)) - u(f'(\omega)) < \epsilon$ (for every ω) and $u(f'(\omega)) > u(y) - \epsilon$.

Take a sub-sequence of B'_n and B_n so that $B_{n_i} = B_{n(\frac{1}{i})+1}$ and $B'_{n_i} = B'_{n(\frac{1}{i})+1}$. Pick $\bar{i}$ so that $\frac{1}{i} < k$. Set $\alpha_i = \frac{\frac{1}{n_i}}{(u(y) - \frac{1}{n_i} - u(z))}$ for every $i > \bar{i}$. Consider $f = F_{n_i}(\omega)$ for an arbitrary ω and $i > \bar{i}$. Pick $f' \in \{G_{n_i}\}$ so that $u(f(\omega')) - u(f'(\omega')) < \epsilon$ for every ω'.

Note that for every ω'

$$\begin{aligned}(1-\alpha)u(f(\omega'))+\alpha u(z)-u(f'(\omega')) &= u(f(\omega'))-u(f'(\omega')) \\ &\quad -\alpha(u(f(\omega'))-u(z)) \\ &< \frac{1}{n_i}-\alpha(u(f(\omega'))-u(z)) \\ &< \frac{1}{n_i}-\alpha(u(y)+\frac{1}{n_i}-u(z)).\end{aligned}$$

Therefore, for every ω,

$$u\circ((1-\alpha_i)F_{n_i}(\omega)+\alpha_i z)\leq u\circ f'$$

for some $f'\in\{G_{n_i}\}$.

By Monotonicity and INRA, $\hat{c}(B_{n_i})\subset\hat{c}(B_{n_i}\cup((1-\alpha_i)\{G_{n_i}\}+\alpha_i\{z\}))$ for $i<\bar{i}$. By Lemma 5, $\{G_{n_i}\}$ IS C_{n_i} implies that $(1-\alpha_i)\{G_{n_i}\}+\alpha_i\{z\}$ IS $(1-\alpha_i)C_{n_i}+\alpha_i\{z\}$. But then for $i>\bar{i}_k$ A_{n_i} IS $(1-\alpha_i)C_{n_i}+\alpha_i\{z\}$ and since $\alpha_i\to 0$, $A_{n_i}\to A$ and $C_{n_i}\to C$, we have $A\ \overline{IS}\ C$. Suppose we needed to mix all problems with x^* first. The amount of this mixture can be arbitrarily small. So using the same logic, we can find a sub-sequence $(n_k)_{k=1}^{\infty}$and a sequence $(\alpha_k)_{k=1}^{\infty}$ where $\alpha_k\to 0$ so that $(\frac{k-1}{k}A_{n_k}+\frac{1}{k}\{x^*\})$ IS $(1-\alpha_k)(\frac{k-1}{k}C_{n_k}+\frac{1}{k}\{x^*\})+\alpha_k\{z\}$ for every k. Again, this gives $A\,\bar{IS}\,C$. Conclude that $F\hat{\succeq}H$. □

Define $\mathbb{P}^*=\{\mathcal{A}(B):B\in K(\mathcal{F})\}$ and $\mathbb{P}^{**}=\{Q\in\mathbb{P}^*:\nexists Q'\in\mathbb{P}^*\,s.t.\,Q'\gg Q\}$. Identify $\mathcal{F}_Q^{\Omega}$ with the subset of $\mathcal{F}^{\Omega}$ that is $\sigma(Q)-$measurable.

Define the binary relation $\succeq$ on $\mathcal{F}^{\Omega}$ by

$$F\succeq G\iff \text{either } F\in\mathcal{H} \text{ and } F^*\hat{\succeq}G^* \text{ or } G\notin\mathcal{H}$$

$\succeq$ is a consistent extension of $\hat{\succeq}$.

Lemma 7. $\succeq$ *extends* $\hat{\succeq}$ *consistently.*

Proof. Suppose first that $F,G\in\mathcal{H}$ and that $G\in\mathcal{F}_Q^{\Omega}$.

[$F\hat{\succeq}G$ implies $F\succeq G$] Suppose $F\hat{\succeq}G$. Then $F,G\in\mathcal{C}$. By Lemma 4, $F^*\hat{\sim}F$ and $G^*\hat{\sim}G$. Since $\hat{\succeq}$ is transitive, $F^*\hat{\sim}F\hat{\succeq}G\implies F^*\hat{\succeq}G$. Since $G\hat{\sim}G^*$, we have that $F^*\hat{\succeq}G^*$ using transitivity of $\hat{\succeq}$.

[$F\hat{\succ}G$ implies $F \succ G$] There are two cases: $G \in \mathcal{H}$ and $G \notin \mathcal{H}$. In the latter case, it is impossible that $F\hat{\succ}G$. Therefore, it must be that both $F, G \in \mathcal{C}$. For contradiction, assume that $F\hat{\succ}G$ but $F \not\succ G$. Since $F\hat{\succeq}G$ implies $F \succeq G$, it must be that $G \succeq F$ (otherwise $F \succ G$) so $F^*\hat{\sim}G^*$. Since $F\hat{\sim}F^*$ and $G^*\hat{\sim}G$ and $\hat{\succeq}$ is transitive, it follows immediately that $G\hat{\sim}F$, a contradiction with $F\hat{\succ}G$. □

For any act $f \in \mathcal{F}$ and partition $Q \in \mathbb{P}^*$, define $\bar{f}_Q \in \mathcal{F}^\Omega$ as follows. If $\min_{x\in X} u(x)$ does not exist, then $\bar{f}_Q(\omega) = fQ(\omega)x$ where x is so that $u(x) = \min_{x\in\cup_{\omega\in\Omega} supp(G(\omega))} u(x) - 1$. If $\min_{x\in X} u(x)$ exists, then by $\bar{f}_Q(\omega) = fQ(\omega)x$ where x is so that $u(x) = \min_{x\in X} u(x)$.

Lemma 8. *$\succeq$ is a preorder.*

Proof. [Transitive] Suppose that $F \succeq G$ and $G \succeq H$. If $F \notin \mathcal{H}$ then $G \notin \mathcal{H}$ so $H \notin \mathcal{H}$ so $F \succeq H$. If $G \notin \mathcal{H}$ then $H \notin \mathcal{H}$ so $F \succeq H$. If $H \notin \mathcal{H}$, then $F \succeq H$. All that remains is the case where $F, G, H \in \mathcal{H}$ which would imply that

$$F^*\hat{\succeq}G^*\hat{\succeq}H^*$$

Since $\hat{\succeq}$ is transitive, $F^*\hat{\succeq}H^*$ and $F \succeq H$.

[Reflexive] Fix arbitrary F. If $F \notin \mathcal{H}$, then $F \succeq F$ by definition. If $F \in \mathcal{H}$, then noting that $F^*\hat{\sim}F^*$ (since $F^* \in \hat{c}(\{F^*\})$ immediately gives $F^*\hat{\succeq}F^*$ and $F \succeq F$. □

Lemma 9. *If $x, y \in X$ then either $x \succeq y$ or $y \succeq x$.*

Proof. This follows immediately from Lemma 3. □

Lemma 10. *For all $e, f, g, h \in \mathcal{F}$, the set $U = \{\lambda \in [0,1] : \lambda f + (1-\lambda)g \succeq \lambda h + (1-\lambda)e\}$ is closed in $[0,1]$.*

Proof. Suppose $\lambda_n \to \lambda$ and $\lambda_n \in U$ for all n. Then $\{\lambda_n f + (1-\lambda_n)g\} = B_n$, $\{\lambda_n h + (1-\lambda_n)e\} = C_n$ and $B_n \, \overline{IS} \, C_n$ by definition of $\succeq$. Therefore, for every n, there are sequences $(B^n_m)_{m=1}^\infty$ and $(C^n_m)_{m=1}^\infty$ so that B^n_m IS C^n_m and $d(B^n_m, B_n) + d(C^n_m, C_n) \to 0$. For every ϵ, there is an M^n_ϵ so that $m > M^n_\epsilon$ implies that $d(B^n_m, B_n) + d(C^n_m, C_n) < \epsilon$. Since $B_n \to \{\lambda f + (1-\lambda)g\} = B$ and $C_n \to \{\lambda h + (1-\lambda)e\} = C$, for every ϵ, there is an N_ϵ so that $n > N_\epsilon$ implies that $d(B_n, B) + d(C_n, C) < \epsilon$.

For every $n \in \{1, 2, ...\}$ define B'_n and C'_n by

$$B'_n = B_{N_{\frac{1}{n}}+1}^{M_{\frac{1}{n}}+1}$$

and

$$C'_n = C_{N_{\frac{1}{n}}+1}^{M_{\frac{1}{n}}+1}.$$

By the triangle inequality,

$$d(B'_n, B) \leq d(B, B_{N_{\frac{1}{n}}+1}) + d(B_{N_{\frac{1}{n}}+1}, B_{N_{\frac{1}{n}}+1}^{M_{\frac{1}{n}}+1})$$

and

$$d(C'_n, C) \leq d(C, C_{N_{\frac{1}{n}}+1}) + d(C_{N_{\frac{1}{n}}+1}, C_{N_{\frac{1}{n}}+1}^{M_{\frac{1}{n}}+1}).$$

Since

$$d(B, B_{N_{\frac{1}{n}}+1}) + d(B_{N_{\frac{1}{n}}+1}, B'_n) + d(C, C_{N_{\frac{1}{n}}+1}) + d(C_{N_{\frac{1}{n}}+1}, C'_n) \leq \frac{4}{n}$$

which goes to zero, $B'_n \to B$ and $C'_n \to C$; since B'_n IS C'_n for every n, $B\,\overline{IS}\,C$. It immediately follows from the definition of $\succeq$ and U that $\lambda f + (1-\lambda)g \succeq \lambda h + (1-\lambda)e$ so $\lambda \in U$. □

Lemma 11. *For any $f, g \in \mathcal{F}$, if $f(\omega) \succeq g(\omega)$ for all ω, then $f \succeq g$.*

Proof. For any $f, g \in \mathcal{F}$ so that $f(\omega) \succeq g(\omega)$ for every ω, it follows that $f(\omega) \in c(\{f(\omega), g(\omega)\}|\omega)$ from Lemma 3. From monotonicity, $f \in c(\{f, g\}|\omega)\forall\omega$ so $f \in \hat{c}(\{f, g\})$ and $f \succeq g$. □

Lemma 12. *For any $f, g, h \in \mathcal{F}$ and $\alpha \in (0, 1]$, $f \succeq g$ if and only if $\alpha f + (1-\alpha)h \succeq \alpha g + (1-\alpha)h$.*

Proof. Fix $f, g, h \in \mathcal{F}$ so that $f \succeq g$. Let $\alpha \in (0, 1]$ be arbitrary. Since $f \succeq g$, $\{f\}\,\bar{IS}\,\{g\}$. Apply Lemma 5 to get $\{\alpha f + (1-\alpha)h\}\,\bar{IS}\,\{\alpha g + (1-\alpha)h\}$, implying that $\alpha f + (1-\alpha)g \hat{\succeq} \alpha f + (1-\alpha)g$ so $\alpha f + (1-\alpha)g \succeq \alpha f + (1-\alpha)g$. □

Given the above and Lemma 10, Lemma 1.2 of Shapley and Baucells [1998] gives that $\alpha f + (1-\alpha)h \succeq \alpha g + (1-\alpha)h$ for $\alpha > 0$ implies $f \succeq g$.

Lemma 13. *There are $x, y \in X$ so that $x \succ y$.*

Proof. Recall that $x_* \notin \hat{c}(\{x^*, x_*\})$. Since $\mathcal{A}(\{x^*\}) = \mathcal{A}(\{x_*\}) = \{\Omega\}$ and $x^* \in \hat{c}(\{x_*, x^*\})$ by Lemma 3, x^* weakly dominates x_*. It follows that $x^* \succeq x_*$. Further, by Continuity, it is not the case that $\{x_*\}\, \bar{IS}\, \{x^*\}$ so $x^* \succ x_*$. □

Lemma 14. *If there is an $x \in X$ so that $u(f(\omega)) > u(x)$ for every ω, then $\bar{f}_Q \in \hat{c}(\{\bar{f}_Q\})$.*

Proof. Fix any such f and relabel $\bar{f}_Q = \bar{F}$. Let $y \in \arg\min_{x' \in \cup_\omega supp(\bar{F}(\omega))} u(x')$. Suppose not: $\bar{F} \notin \hat{c}(\{\bar{F}\})$ and $H \in \hat{c}(\{\bar{F}\})$.

It must be that $\mathcal{A}(\{\bar{F}\}) \not\gg Q$. If $\mathcal{A}(\{\bar{F}\}) \gg Q$, pick any ω so that $\bar{F}(\omega) \notin c(\{\bar{F}\}|\omega)$. Take $h = H(\omega)\mathcal{A}(\{\bar{F}\})x$. Since $u \circ H(\omega) \geq u \circ h$, $H \in \hat{c}(\{\bar{F}\} \cup \{h\})$ and $\mathcal{A}(\{\bar{F}\} \cup \{h\}) \gg \mathcal{A}(\{\bar{F}\})$. Further, by Subjective Consequentialism, $h \in c(\{\bar{F}\} \cup \{h\}|\omega)$. However, $u \circ \bar{F}(\omega) \geq u \circ h$ so monotonicity implies that $\bar{F}(\omega) \in c(\{\bar{F}\} \cup \{h\}|\omega)$. Since h is never strictly relevant, INRA implies that $\bar{F}(\omega) \in c(\{\bar{F}\}|\omega)$, a contradiction.

Since $\mathcal{A}(\{\bar{F}\}) \not\gg Q$, there is some $E \in P$ so that $u(H^*(\omega)) = u(y) \forall \omega \in E$. For every $\omega \notin E$, $u(F^*(\omega)) \geq u(H^*(\omega))$. Therefore, F^* weakly dominates H^* by monotonicity.

Define $\bar{H} \equiv \bar{H^*}_Q$ where $Q = \mathcal{A}(\{\bar{F}\})$. By definition, there is some J so that $J \in \hat{c}(\{J\})$ and $\mathcal{A}(\{J\}) = Q$. Since $\mathcal{A}(\{F^*\}) = \{\Omega\}$, $\alpha F^* + (1-\alpha)J \in \hat{c}(\alpha\{F^*\} + (1-\alpha)J)$

By INRA, monotonicity and Subjective Consequentialism, $\bar{H} \in \hat{c}(\{\bar{H}\} \cup \{\bar{F}\})$ and $\bar{H}^* \in \hat{c}(\{\bar{H}\} \cup \{\bar{H}^*\})$. Let $B_0 = \alpha\{\bar{H}^*\} + (1-\alpha)\{J^*\}$, $B_1 = \alpha\{\bar{H}^*\} \cup \{\bar{H}\} + (1-\alpha)\{J^*\}$ and $B_2 = \alpha\{\bar{H}\} \cup \{\bar{F}\} + (1-\alpha)\{J^*\}$. By ACI,

$$\alpha\bar{H}^* + (1-\alpha)J^* \in \hat{c}(B_1) \cap \hat{c}(B_0)$$

and

$$\alpha\bar{H} + (1-\alpha)J^* \in \hat{c}(B_2).$$

This implies that

$$\alpha\bar{H}^* + (1-\alpha)J^* \hat{\succeq} \alpha\bar{H} + (1-\alpha)J^* \hat{\succeq} \alpha\bar{F} + (1-\alpha)J^*. \tag{A.2.1}$$

Set $B_4 = \alpha\{F^*\} + (1-\alpha)\{J\}$. By ACI, $\alpha F^* + (1-\alpha)J(\omega) \in c(B_4|\omega)$. Let $B_5 = B_4 \cup \{\alpha\bar{F} + (1-\alpha)J\}$. By INRA, monotonicity and Subjective Consequentialism,

$\alpha F^* + (1-\alpha)J \in \hat{c}(B_5)$. By Subjective Consequentialism, $\alpha\bar{F}(\omega) + (1-\alpha)J(\omega) \in c(B_5|\omega)$ for all ω so by INRA, $\alpha\bar{F} + (1-\alpha)J \in \mathcal{C}$. Set $B_3 = B_2 \cup \{\alpha\bar{F} + (1-\alpha)J\}$. By Monotonicity and INRA,

$$\alpha\bar{H} + (1-\alpha)J^* \in \hat{c}(B_3)$$

so $\alpha\bar{H}^* + (1-\alpha)J^* \hat{\succeq} \alpha\bar{F} + (1-\alpha)J$.

By Lemma 4, $\alpha\bar{F} + (1-\alpha)J \hat{\sim} \alpha F^* + (1-\alpha)J^*$. By Lemma 6,

$$\alpha\bar{H}^* + (1-\alpha)J^* \hat{\succeq} \alpha F^* + (1-\alpha)J^*$$

which, by definition, is equivalent

$$\{\alpha\bar{H}^* + (1-\alpha)J^*\}\,\overline{IS}\,\{\alpha F^* + (1-\alpha)J^*\}.$$

Since F^* dominates $\bar{H}^*$ and $\bar{H}^*$ does not dominate F^* by Monotonicity, $\alpha F^* + (1-\alpha)J^*$ dominates $\alpha\bar{H}^* + (1-\alpha)J^*$ and $\alpha\bar{H}^* + (1-\alpha)J^*$ does not dominate $\alpha F^* + (1-\alpha)J^*$. This contradicts Continuity, so $\bar{f}_Q \in \hat{c}(\{\bar{f}_Q\})$. □

If $Q \in \mathbb{P}^*$ and $Q \gg Q'$, then Lemma 14 implies that $Q' \in \mathbb{P}^*$.

Lemma 15. *Suppose $F \in \hat{c}(B)$. If $\{G\} \subset B$, then $F \succeq G$.*

Proof. If $G \notin \mathcal{H}$, then $F \succeq G$. If $G \in \mathcal{H}$, then pick $Q \in \mathbb{P}^*$ so that $G \in \mathcal{F}_Q^\Omega$. There are two cases.

First, suppose $u(G^*(\omega)) > u(x)$ for some $x \in X$ and every ω. Set $\bar{G} = (\bar{G^*})_Q$. By monotonicity $F \in \hat{c}(B \cup \{\bar{G}\})$. By Lemma 14, $\bar{G} \in \hat{c}(\{\bar{G}\})$, which implies that $\{F\}\,\overline{IS}\,\{\bar{G}\}$. Since $G^* = \bar{G}^*$ by construction, it follows that $F \succeq G$.

Now, suppose $u(G^*(\omega')) = \min_{x\in X} u(x)$ for at least one ω'. Consider $F' = \frac{1}{2}F + \frac{1}{2}x^*$ and $G' = \frac{1}{2}G + \frac{1}{2}x^*$ and $\bar{G}' = (\bar{G'^*})_Q$. Now, $u(G'^*(\omega)) > \min_{x\in X} u(x)$. Apply the above argument to get $\{F'\}\,\overline{IS}\,\{G'\}$ and $(\frac{1}{2}F + \frac{1}{2}x^*)^* \hat{\succeq} (\frac{1}{2}G + \frac{1}{2}x^*)^*$. Lemma 12 gives that $F^* \succeq G^*$, so $F \succeq G$. □

Lemma 16. *There is a finitely additive probability measure on Σ, $\pi(\cdot)$, that assigns positive probability to every $E \in P$ so that for any $f, g \in \mathcal{F}$, $f \succ g$ implies $\int u \circ f d\pi > \int u \circ g d\pi$ and $f \sim g$ implies $\int u \circ f d\pi = \int u \circ g d\pi$.*

Proof. Let $\mathcal{F}' \subset \mathcal{F}$ be the acts that are $\sigma(P)$ measurable.

Claim 1. For any $f \in \mathcal{F}$, there is an $f' \in \mathcal{F}'$ so that $f' \sim f$.

Proof. First, I show that for any act f and any $E \in P$, there is an act g so that $g \sim f$ and g is constant on E and agrees with f on E^c. Pick any $f \in \mathcal{F}$ and any $E \in P$. Let $\bar{x} = \arg\max_{\omega \in E} u(f(\omega))$, $\underline{x} = \arg\min_{\omega \in E} u(f(\omega))$, $\bar{g} = \bar{x}Ef$ and $\underline{g} = \underline{x}Ef$. For every $\alpha \in [0,1]$, define

$$B_\alpha = \{f, \alpha\bar{g} + (1-\alpha)\underline{g}\}.$$

By Subjective Consequentialism and because $c(B|\cdot)$ must be P measurable, there is at least one h in every B_α so that $h \in c(B_\alpha|\omega)$ for all $\omega \in \Omega$.

Fix $\omega \in E$. By Monotonicity and INRA,

$$\alpha\bar{g} + (1-\alpha)\underline{g} \in c(B_\alpha|\omega) \,\&\, \beta > \alpha \implies \beta\bar{g} + (1-\beta)\underline{g} \in c(B_\beta|\omega)$$

and conversely

$$\alpha\bar{g} + (1-\alpha)\underline{g} \notin c(B_\alpha|\omega) \,\&\, \beta < \alpha \implies \beta\bar{g} + (1-\beta)\underline{g} \notin c(B_\beta|\omega).$$

Using the above and that $\bar{g} \in c(B_1|\omega)$, there is an $\bar{\alpha}$ so that $\alpha > \bar{\alpha}$ implies that $\{\alpha\bar{g} + (1-\alpha)\underline{g}\}\, IS\, \{f\}$ and $\alpha < \bar{\alpha}$ implies that $\{f\}\, IS\, \{\alpha\bar{g} + (1-\alpha)\underline{g}\}\, IS\, \{f\}$. Conclude that $\bar{\alpha}\bar{g} + (1-\bar{\alpha})\underline{g} \sim f$. Since f and E were arbitrary, this establishes the first step.

Now, label $P = \{E_1, ..., E_n\}$. Fix f. By the above, there is an f_1 so that $f \sim f_1$ and $f_1$is constant on E_1 and agrees with f on E_1^c. For $i = 2, ..., n$, the above shows that there is f_{i+1} so that $f_i \sim f_{i+1}$ and $f_{i+1}$is constant on E_{i+1} and agrees with f_i on E_{i+1}^c. By construction, f_n is $\sigma(P)$-measurable, and $f \sim f_1 \sim f_2 \sim ... \sim f_n \implies f \sim f_n$ by Lemma 8, so $f_n \in \mathcal{F}'$ and $f_n \sim f$, establishing the claim. □

Moreover, $\mathcal{F}'$ is finite dimensional. Restricted to $\mathcal{F}'$, $\succeq$ satisfies reflexivity, transitivity and independence by Lemmas 8 and 5. Lemma 10 implies that if $\lambda f + (1-\lambda)g \succeq g$ for every $\lambda \in (0,1)$, then it is not the case that $g \succ f$. Applying Aumann [1962, Thm. A] yields the existence of a mixture linear $U(\cdot)$ so that $f \succ g$ implies $U(f) > U(g)$ and $f \sim g$ implies $U(f) = U(g)$. By Monotonicity using choice from problems in the set $\{\{xEy, y\} : E \in P\}$ where $u(x) > u(y)$ and Lemma 3, there is a

$\pi(\cdot)$ with the desired properties, an $\alpha > 0$ and a $\beta \in \mathbb{R}$ so that $U(\cdot) = \int \alpha u \circ f d\pi(\cdot) + \beta$, the desired result. WLOG, take $\alpha = 1$ and $\beta = 0$. □

Lemma 17. *$c(B|\omega) = \arg\max_{f \in B} \int u \circ f d\pi(\cdot|\mathcal{A}(B)(\omega))$.*

Proof. Fix $B \in K(B)$ and set $E = \mathcal{A}(B)(\omega)$. First, suppose $f \in c(B|\omega)$ and set $F \in \hat{c}(B)$ so that $F(\omega) = f$ for all $\omega \in E$. Specifically, $F \succeq G$ whenever there is a $g \in B$ so that $G(\omega) = F(\omega)$ for every $\omega \notin E$ and $G(\omega) = g$ for every $\omega \in E$, so $f \in \arg\max_{g \in B} \int u \circ g d\pi(\cdot|E)$, implying that $c(B|\omega) \subset \arg\max_{f \in B} \int u \circ f d\pi(\cdot|E)$.

Now, suppose that $\int u \circ g d\pi(\cdot|E) \in \arg\max_{g \in B} \int u \circ g d\pi(\cdot|E)$. Set $x \in X$ so that $u(x) < \min_{\{f(\omega): f \in B \, and \, \omega \in \Omega\}} u(f(\omega))$. Define $\hat{F}(\omega) = F(\omega)\mathcal{A}(B)(\omega)x$ for all ω and $\hat{G}(\omega) = F(\omega)\mathcal{A}(B)(\omega)x$ for all $\omega \notin E$ and $\hat{G}(\omega) = gEx$ for $\omega \in E$ and $B' = B \cup \{\hat{F}\} \cup \{\hat{G}\}$ and note that Subjective Consequentialism and Monotonicity imply $\hat{F} \in \hat{c}(B')$. Then take $B'' = \{\hat{F}\} \cup \{\hat{G}\}$ and by INRA, $\hat{F} \in \hat{c}(B'')$. Now, take $y \in X$ so that $u(y) > \int u \circ f d\pi(\cdot|E)$ define $B_n = (B'' \backslash \{gEx\}) \cup \{\frac{1}{n}y + \frac{n-1}{n}gEx\}$. By Lemma 16, $G_n \succ F'$ for all F' so that $\{F'\} \subset B$ where $G_n(\omega) = \hat{F}(\omega)\mathcal{A}(B)(\omega)x$ for all $\omega \notin E$ and $G_n(\omega) = \frac{1}{n}y + \frac{n-1}{n}gEx$ for all $\omega \in E$. By Lemma 15, $G_n \in \hat{c}(B_n)$.[50]

By construction, $\mathcal{A}(B_n) = \mathcal{A}(B'')$ for all n. Since $\frac{1}{n}y + \frac{n-1}{n}g \in c(B_n|\omega)$ and $(\frac{1}{n}y + \frac{n-1}{n}g)Ex \to gEx$, it follows from Continuity that $gEx \in c(B''|\omega)$. By INRA, $c(B''|\omega) = c(B'|\omega) \cap B''$. Since $u \circ g \geq u \circ gEx$, $gEx \in c(B'|\omega) \implies g \in c(B'|\omega)$ by Monotonicity. By INRA, $c(B'|\omega) = c(B'|\omega) \cap B$, so $g \in c(B|\omega)$, completing the proof. □

Set $\hat{P}(B) = \mathcal{A}(B)$. Lemma 16 and 15 give that

$$\hat{P}(B) \in \arg\max_{Q \in \mathbb{P}^*} \sum_{E' \in Q} \pi(E') \max_{f \in B} \int u \circ f d\pi(\cdot|E')$$

because if $F \in \hat{c}(B)$ then $F \succeq G$ for all $G \in \mathcal{H}$ and $\{G\} \subset B$ implies that $\int u \circ F^* d\pi \geq \int u \circ G^* d\pi$, implying that

$$\sum_{E' \in \hat{P}(B)} \pi(E') \int u(F(\omega)(\omega)) d\pi(\cdot|E') \geq \sum_{E' \in Q} \pi(E') \max_{g \in B} \int u \circ g d\pi(\cdot|E')$$

[50] These inequalities for $u(x)$ and $u(y)$ can be taken to be strict even if $u(\cdot)$ is bounded because the remainder relies only on properties of $\tilde{R}$. Mixing with B with a constant $z \in int(u(X))$ ensures that there exists such $x, y \in X$.

for any $Q \in \mathbb{P}^*$. Using that $\hat{P}(B) = \mathcal{A}(B)$, Lemma 17 gives that

$$c(B|\omega) = \arg\max_{f \in B} \int u \circ f d\pi(\cdot|\hat{P}(B)(\omega)),$$

completing the proof. □

Proof of Theorem 2

Proof. Suppose $c(\cdot)$ is represented by $(u(\cdot), \pi(\cdot), \mathbb{P}^*, \hat{P}(\cdot))$. First, I show that $c(\cdot)$ satisfies Continuity.

Lemma 18. *$c(\cdot)$ satisfies Continuity.*

Proof. Suppose that both $\{f\}\,\overline{IS}\,\{g\}$ and g weakly dominates f. Let $B_n \to \{f\}$ and $C_n \to \{g\}$ so that $B_n\,IS\,C_n$. If $B_n\,IS\,C_n$, $F_n \in c(B_n)$ and $G_n \in c(C_n)$, then $\int u \circ F_n^* d\pi \geq \int u \circ G_n^* d\pi$. To see this, let $B_1, ..., B_n$ be the sequence from the definition of indirectly selected. Then $F_i \in \hat{c}(B_i)$ and $\{F_{i+1}\} \subset B_i$ implies that $\int u \circ F_i^* d\pi \geq \int u \circ F_{i+1}^* d\pi$ by construction of $c(\cdot)$ for all i. Therefore, $\int u \circ F_n^* d\pi \geq \int u \circ G_n^* d\pi$.

Now, note that $F_n^* \to f$ since all components of $B_n \to f$. Similarly, $G_n^* \to g$. Therefore, since $u(\cdot)$ is continuous, $\int u \circ f d\pi \geq \int u \circ g d\pi$. Since g weakly dominates f, $\int u \circ g d\pi(\cdot|\hat{P}(\{f,g\})(\omega)) \geq \int u \circ f d\pi(\cdot|\hat{P}(\{f,g\})(\omega))$, without equality for some ω'. This implies that $\int u \circ g d\pi(\cdot) > \int u \circ f d\pi$, a contradiction. □

For the second part, begin by defining a function $V : \{Q : P \gg Q\} \times K(\mathcal{F}) \to \mathbb{R}$ by

$$V(Q, B) = \sum_{E \in Q} \pi(E) \max_{f \in B} \int u \circ f d\pi(\cdot|E).$$

With this formulation, $\hat{P}(B) \in \arg\max_{Q \in \mathbb{P}^*} V(Q, B)$ for all B. By the maximum theorem, $V(Q, \cdot)$ is continuous and $\arg\max V(\cdot, B)$ is upper-hemi continuous.

If $u(\cdot)$ is constant, then set $K = K(\mathcal{F})$. Both INRA and ACI are satisfied because $c(B|\omega) = B$ for every B and ω. Clearly, K is open and dense in $K(\mathcal{F})$.

If not, then define K by

$$K = \{B \in K(\mathcal{F}) : \exists Q \in \mathbb{P}^* \, s.t. \, V(Q, B) > V(Q', B) \, \forall Q' \in \mathbb{P}^* \backslash \{Q\}\}.$$

I proceed by showing that $cl(K) = K(\mathcal{F})$ and then that K is open.

Lemma 19. *$cl(K) = K(\mathcal{F})$*

Proof. Pick any $B \in K(\mathcal{F})$ and any $\epsilon > 0$.

Fix $x \in X$ so that $u(x) \in int(u(X))$. Define $B' \in K(\mathcal{F})$ by $\alpha B + (1-\alpha)\{x\}$ for α close enough to 1 so that $d(B', B) < \frac{\epsilon}{3}$.

Pick a $Q \in \mathbb{P}^*$ so that $Q \gg \mathcal{A}(B')$ and $Q' \gg Q$ for $Q' \in \mathbb{P}^*$ implies that $Q = Q'$. Label $Q = \{E_1, ..., E_n\}$ and pick $f_1, ..., f_n$ so that $f_i \in c(B'|\omega)$ for some $\omega \in E_i$. Define f^* so that

$$f^*(\omega) = f_i(\omega)$$

whenever $\omega \in E_i$ and f^{**} so that $u(f^{**}) = u(f^*) + k$ for some $k > 0$. Since $u \circ f^* \in int(u(X)^\Omega)$ by construction of B', such a k exists.

Now, define f^i_α for every $\alpha \in [0,1]$ by $f^i_\alpha = (\alpha f_i + (1-\alpha) f^{**}) E_i f_i$ for every $i \in \{1, ..., n\}$. For α close enough to 1, $d(f^i_\alpha, f_i) < \frac{\epsilon}{3}$. Therefore, for α^* sufficiently high, note that $d(B'', B') < \frac{\epsilon}{3}$ where

$$B'' = B' \cup \{f^i_{\alpha^*}\}_{i=1}^n.$$

Conclude that $d(B'', B) \leq d(B'', B') + d(B', B) < \frac{2\epsilon}{3} < \epsilon$. Further, $V(Q, B'') > V(Q', B'')$ for all $Q' \in \mathbb{P}^*$ so that $Q' \neq Q$. Therefore, $B'' \in K$. Since B and ϵ are arbitrary, there is a $B'' \in K$ arbitrarily close to any $B \in K(\mathcal{F})$. Therefore, $cl(K) = K(\mathcal{F})$. □

Lemma 20. *K is open.*

Proof. Let $K^c = K(\mathcal{F}) \backslash K$. K is open if and only if K^c is closed. Because $K(\mathcal{F})$ is a metric space and thus first countable, it is sufficient to only show sequentially closed.

Pick $(B_n)_{n=1}^\infty \subset K^c$ and suppose that $B_n \to B$. Because $\mathbb{P}^*$ is finite, there are $Q \neq Q' \in \mathbb{P}^*$ and a sub-sequence $(B_{n_k})_{k=1}^\infty$ so that $Q, Q' \in \arg\max_{Q \in \mathbb{P}^*} V(Q, B_{n_k})$ for all $Q'' \in \mathbb{P}^*$. Because $\arg\max_{Q \in \mathbb{P}^*} V(Q, \cdot)$ is upper hemi-continuous and $B_{n_k} \to B$, $Q, Q' \in \arg\max_{Q \in \mathbb{P}^*} V(Q, B)$. Conclude that $B \in K^c$, so K^c is closed and K is open. □

Let $>$ be a linear order on $\mathbb{P}^{**}$ and set

$$\hat{Q}(B) = \max_{>} \arg\max_{Q \in \mathbb{P}^{**}} V(Q, B).$$

Define the conditional choice correspondence $c'(\cdot)$ by

$$c'(B|\omega) = \arg\max_{f \in B} \int u \circ f d\pi(\cdot|\hat{Q}(B)(\omega))$$

for every $B \in K(\mathcal{F})$. Clearly $c'(\cdot)$ has an optimal inattention representation and for every $B \in K$, $c'(B|\omega) = c(B|\omega)$ for every $\omega \in \Omega$.

Lemma 21. *$c'(\cdot)$ satisfies ACI.*

Proof. If $\mathcal{A}(B) \gg \mathcal{A}(C)$, then $\hat{Q}(B) \in \arg\max_{Q \in \mathbb{P}^{**}} V(Q, C)$. Therefore, $\hat{Q}(B) \in \arg\max_{Q \in \mathbb{P}^{**}} V(Q, \alpha B + (1-\alpha)C)$. Further, if $Q' \in \arg\max_{Q \in \mathbb{P}^{**}} V(Q, \alpha B + (1-\alpha)C)$, then $Q' \in \arg\max_{Q \in \mathbb{P}^{**}} V(Q, B) \cap \arg\max_{Q \in \mathbb{P}^{**}} V(Q, C)$. Since $\hat{Q}(B) > Q'$ for every $Q \in \arg\max_{Q \in \mathbb{P}^{**}} V(Q, B)$, it follows immediately that $\hat{Q}(\alpha B + (1-\alpha)C) = \hat{Q}(B)$. The conclusion follows immediately. □

Lemma 22. *$c'(\cdot)$ satisfies INRA.*

Proof. Suppose that $A \subset B$ and $c'(B|\omega) \cap A \neq \emptyset$ for all ω. Note that

$$\arg\max_{Q \in \mathbb{P}^{**}} V(Q, A) \subset \arg\max_{Q \in \mathbb{P}^{**}} V(Q, B)$$

and since $\hat{Q}(B) > Q'$ for all $Q' \in \arg\max_{Q \in \mathbb{P}^{**}} V(Q, B)$, $\hat{Q}(B) > Q'$ for all $Q \in \arg\max_{Q \in \mathbb{P}^{**}} V(Q, A)$ so $\hat{Q}(A) = \hat{Q}(B)$. Since

$$\begin{aligned} c'(B|\omega) \cap A &= [\arg\max_{f \in B} \int u \circ f d\pi(\cdot|\hat{Q}(B)(\omega))] \cap A \neq \emptyset \\ &= \arg\max_{f \in A} \int u \circ f d\pi(\cdot|\hat{Q}(B)(\omega)) \\ &= \arg\max_{f \in A} \int u \circ f d\pi(\cdot|\hat{Q}(A)(\omega)) \\ &= c'(A|\omega) \end{aligned}$$

it follows that $c'(B|\omega) \cap A = c'(A|\omega)$. □

Since K is open and $cl(K) = K(\mathcal{F})$, the Theorem follows immediately. □

Proof of Corollary 2:

Proof. (iii) clearly implies either (i) or (ii).

[(i) implies (iii)] Suppose $c(\cdot)$ satisfies Independence and has optimal inattention, with the canonical representation $\hat{P}(B) = \mathcal{A}(B)$. Let Q be coarsest common refinement of $\{\hat{P}(B)\}_{B\in K(\mathcal{F})}$. Claim that $c(B|\omega) = \arg\max_{f\in B} \int u \circ f d\pi(\cdot|Q(\omega))$ for every B. If not, there is a B' and an ω so that

$$c(B'|\omega) \neq \arg\max_{f\in B} \int u \circ f d\pi(\cdot|Q(\omega)).$$

Since $c(\cdot)$ has inattention,

$$c(B'|\omega) = \arg\max_{f\in B'} \int u \circ f d\pi(\cdot|\hat{P}(B')(\omega)).$$

There is a finite collection $\{B_1, ..., B_n\} \subset K(\mathcal{F})$ so that

$$[\cap_{i=1}^n \hat{P}(B_i)(\omega)] \cap \hat{P}(B')(\omega) = Q(\omega)$$

and $c(B_i|\omega) \neq c(B_j|\omega)$ for all $i \neq j$ (perhaps after mixing B_i with a singleton). Set $B^* = \prod_{i=1}^n \frac{1}{n} B_i$ and note that $\mathcal{A}(B^*)(\omega) = Q(\omega)$. Since we can take $\hat{P}(B^*) = \mathcal{A}(B^*)$, it follows that

$$c(B^*|\omega) = \arg\max_{f\in B^*} \int u \circ f d\pi(\cdot|Q(\omega)).$$

Now, since $c(\frac{1}{2}B^* + \frac{1}{2}B'|\omega) = \frac{1}{2}c(B^*|\omega) + \frac{1}{2}c(B'|\omega)$, $\mathcal{A}(\frac{1}{2}B^* + \frac{1}{2}B') \gg \mathcal{A}(B^*)$. By construction, $\mathcal{A}(\frac{1}{2}B^* + \frac{1}{2}B')(\omega) = Q(\omega)$, so

$$\begin{aligned} c(\frac{1}{2}B^* + \frac{1}{2}B'|\omega) &= \arg \max_{f \in \frac{1}{2}B^* + \frac{1}{2}B'} \int u \circ f d\pi(\cdot|Q(\omega)) \\ &= \frac{1}{2} \arg\max_{f \in B^*} \int u \circ f d\pi(\cdot|Q(\omega)) \\ &\quad + \frac{1}{2} \arg\max_{f \in B'} \int u \circ f d\pi(\cdot|Q(\omega)) \\ &\neq \frac{1}{2} \arg\max_{f \in B^*} \int u \circ f d\pi(\cdot|Q(\omega)) \\ &\quad + \frac{1}{2} \arg\max_{f \in B'} \int u \circ f d\pi(\cdot|\hat{P}(B')(\omega)) \\ &= \frac{1}{2}c(B^*|\omega) + \frac{1}{2}c(B'|\omega), \end{aligned}$$

which contradicts Independence.

[(ii) implies (iii)] Suppose that $c(\cdot|\omega)$ satisfies WARP and has optimal inattention. Since $K(\mathcal{F})$ includes all two and three element subsets, there is a complete and transitive binary relation $\succeq_\omega$ so that This binary relation is equal to the revealed preference relation. Let Q be any maximal element of $\mathbb{P}^*$ according to $\gg$. I show that $f \sim_\omega fQ(\omega)y$ for any f and an arbitrarily bad y. Therefore, $\hat{P}(B)(\omega) \subset Q(\omega)$ for every B and ω and consequently $\hat{P}(B)(\omega) = Q(\omega)$ represents choices.

Fix $f \in \mathcal{F}$, $\omega^* \in \Omega$ and $x, y \in X$ so that $u(x) > u(f(\omega)) > u(y)$ for every $\omega \in \Omega$. Define g_ω by $g_\omega(\omega') = \frac{1}{2}x + f(\omega)$ if $\omega' \in Q(\omega)$ and $g_\omega(\omega') = y$ otherwise. Consider the problem $B = \{g_\omega : \omega \notin Q(\omega^*)\} \cup \{f, fQ(\omega^*)y\}$. Clearly, $\hat{P}(B) = Q$ (otherwise, this is not optimal) and also $f, fQ(\omega^*)y \in c(B|\omega^*)$. Conclude that $f \sim_\omega fQ(\omega^*)y$. □

Proof of Corollary 1:

Proof. That (ii) implies (i) is trivial, so suppose $c(\cdot)$ has optimal inattention and satisfies Consequentialism.

Set $y, x \in X$ so that $u(x) > u(y)$ and consider $B = \{xEy : E \in P\} \cup \{x\}$. Clearly $x \in c(B|\omega)\forall\omega$. For any ω, note that $xP(\omega)y \in B$ and $xP(\omega)y(\omega') = x(\omega')$ for every $\omega' \in P(\omega)$. By Consequentialism, $xP(\omega)y \in c(B|\omega)$. However, if $\omega' \notin P(\omega)$, then monotonicity implies that $xP(\omega)y \notin c(B|\omega')$. Therefore, $\mathcal{A}(B) = P$, which implies that $P \in \mathbb{P}^*$. Since there is no $Q \in \mathbb{P}^* \backslash \{P\}$ finer than P or coarser than P, $\{P\} = \mathbb{P}^*$, implying that $c(\cdot)$ is Bayesian. □

A.2.2 Proof from Section 1.4

Proof of Theorem 3

Proof. [i,iv]Affine-uniqueness of $u(\cdot)$ is standard, and canonical uniqueness of $\hat{P}(\cdot)$ is trivial.

[iii] Suppose $\mathbb{P}_1^*, \mathbb{P}_2^*$ both represent $c(\cdot)$. Since $c(\cdot)$ is non-degenerate, there are $x, y \in X$ so that $u(x) > u(y)$. For any $Q \in \mathbb{P}_1^*$, define $B_Q = \{xEy : E \in Q\}$. Clearly, $xQ(\omega)y \in c(B_Q|\omega)$ for every ω, so $\hat{P}(B_Q) \gg Q$. Since $\mathbb{P}_2^*$ represents $c(\cdot)$, $Q \in \mathbb{P}_2^*$. Reversing the role of $\mathbb{P}_1^*$ and $\mathbb{P}_2^*$ give the converse, so they must be equal. T

[ii] Suppose that both π_1 and π_2 represent $c(\cdot)$. By (iii), let $\mathbb{P}^* = \mathbb{P}_1^* = \mathbb{P}_2^*$ and $\mathbb{P}^{**} = \{Q \in \mathbb{P}^* : Q' \gg Q \,\&\, Q' \in \mathbb{P}^* \implies Q = Q'\}$ be the set of the finest subjective information partitions in $\mathbb{P}^*$. Write

$$V(B) = \arg\max_{Q \in \mathbb{P}^*} \sum_{E \in Q} \pi(E)[\max_{f \in B} \int u \circ f d\pi(\cdot|Q(\omega))]$$

for any B. Let $\mathbb{Q}$ be the set of minimal isolatable events for $\mathbb{P}^*$.

Lemma 23. *E is a isolatable event for $\mathbb{P}^*$ if and only if any $Q_1, Q_2 \in \mathbb{P}^{**}$ are such that $Q_1 \gg \{E, E^c\}$, $Q_2 \gg \{E, E^c\}$, and there is a $Q_3 \in \mathbb{P}^{**}$ so that*

$$Q_3 \gg \{E' \in Q_1 : E' \subset E\} \cup \{E' \in Q_2 : E' \subset E^c\}.$$

Proof. [$\implies$] Suppose that E is an isolatable event for $\mathbb{P}^*$. Then pick any $Q_1, Q_2 \in \mathbb{P}^{**}$ and $x, y, z \in X$ so that $u(x) > u(y) > u(z)$; let $B_i = \{xE'y : E' \in Q_i\}$ for $i = 1, 2$.

Suppose $Q_i \not\gg \{E, E^c\}$. Then $\exists F \in Q_1$ s.t. $F \cap E \neq \emptyset$ and $F \cap E^c \neq \emptyset$. Fix $z \in X$ so that $u(x') = u(x) + \epsilon$ where $\epsilon > 0$ and $u(y) > u(z)$. Consider $B = B_i$ and $B' = B_i$, noting that $B_{E,z}B' = B_{E^c,z}B'$ by construction, and that Q_i is the only element of $V(B)$. If $\hat{P}(B_{E,z}B') \neq Q_i$, then E is not an isolatable event since there is no partition in $\mathbb{P}^*$ finer than Q_i except Q_i. However, if $\hat{P}(B_E B') = Q_i$, then either for any $\omega \in F \cap E$, $[xFy]Ex \notin c(B_{E,z}B'|\omega)$ or $[xFy]E^c z \notin c(B_{E^c,z}B'|\omega)$, implying that $xFy \notin c(B_E B'|\omega)$.[51] This contradicts that E is an isolatable event for $\mathbb{P}^*$.

Now, consider $B = B_1$ and $B' = B_2$. By construction,

$$B_{E,z}B' = B'_{E^c,z}B = B''.$$

Consequently,

$$[xQ_1(\omega)y]Ez \in c(B''|\omega)$$

for any $\omega \in E$ and

$$[xQ_2(\omega)y]E^c z \in c(B''|\omega)$$

for any $\omega \in E^c$, implying that

$$\hat{P}(B'') \gg \{E' \in Q_1 : E' \subset E\} \cup \{E' \in Q_2 : E' \subset E^c\}.$$

Therefore, there is some $Q_3 \in \mathbb{P}^{**}$ satisfying the desired property.

[$\Longleftarrow$] Suppose that any $Q_1, Q_2 \in \mathbb{P}^{**}$ are such that $Q_1 \gg \{E, E^c\}$, $Q_2 \gg \{E, E^c\}$, and there is a $Q_3 \in \mathbb{P}^{**}$ so that

$$Q_3 \gg \{E' \in Q_1 : E' \subset E\} \cup \{E' \in Q_2 : E' \subset E^c\}.$$

Consider any B so that $V(B)$ is singleton, and pick any B', labeling $\hat{P}(B) = Q_1$ and $\hat{P}(B') = Q$. If there is no z so that $z \notin c(\{f(\omega), z\}|\omega)$ for any $f \in B \cup B'$, then the condition is arbitrarily satisfied so suppose such a z exists and consider $B_{E,z}B'$.

I claim that

$$\{Q_1\} = \arg\max_{Q \in \mathbb{P}^*} \sum_{E' \in Q \& E' \subset E} \pi(E') \max_{f \in B} \int u \circ f d\pi(\cdot|E').$$

Suppose not, so $Q' \neq Q_1$ is in the argmax above. $Q'' = \{E' \cap E : E' \in Q'\} \cup \{E' \cap E^c : E' \in Q_1\}$ gives at least as high utility as Q_1 when facing, and by assumption, there is a $Q^* \in \mathbb{P}^{**}$ so that $Q^* \gg Q''$, contradicting that $V(B)$ is a singleton.

Consider now $B_{E,z}B'$. For the sake of contradiction, suppose that $\hat{P}(B_{E,z}B')$ cannot be written as $\{E' \in Q_1 : E' \subset E\} \cup \{E' \in Q_2 : E' \subset E^c\}$ for some $Q_2 \in \mathbb{P}^*$.

[51] This procedure must be modified slightly when $\pi_1(F \cap E) = \pi_1(F \cap E^c)$ so that the bet on F gives slightly higher utility on $F \cap E^c$, but arguments otherwise extend.

Let $\hat{P}(B_{E,z}B') = Q'$ and take any $Q \gg Q'$ so that $Q \in \mathbb{P}^{**}$. There is a $Q_3 \gg \{E' \in Q_1 : E' \subset E\} \cup \{E' \in Q : E' \subset E^c\}$ that is in $\mathbb{P}^{**}$. Further,

$$\sum_{E' \in Q_3 \& E' \subset E^c} \pi(E') \max_{f \in B_{E,z}B'} \int u \circ f d\pi(\cdot|E') \geq$$

$$\sum_{E' \in Q' \& E' \subset E^c} \pi(E') \max_{f \in B_{E,z}B'} \int u \circ f d\pi(\cdot|E')$$

because Q_3 is finer than Q' when restricted to E^c, and

$$\sum_{E' \in Q_3 \& E' \subset E} \pi(E') \max_{f \in B_{E,z}B'} \int u \circ f d\pi(\cdot|E') \geq$$

$$\sum_{E' \in Q' \& E' \subset E} \pi(E') \max_{f \in B_{E,z}B'} \int u \circ f d\pi(\cdot|E')$$

by the above claim because for any $E' \subset E$,

$$\max_{f \in B_{E,z}B'} \int u \circ f d\pi(\cdot|E') \leq \max_{f \in B} \int u \circ f d\pi(\cdot|E'),$$

where equality holds whenever $E' \in Q_1$. But this contradicts the assumption that $V(B)$ is a singleton: since $\{E' \in Q_3 : E' \subset E\} \neq \{E' \in Q_1 : E' \subset E\}$, there is a $Q_4 \gg \{E' \in Q_3 : E' \subset E\} \cup \{E' \in Q_1 : E' \subset E^c\}$, $Q_4 \neq Q_1$ and $Q_4 \in V(B)$ by construction. Conclude that $\hat{P}(B_{E,z}B') = \{E' \in Q_1 : E' \subset E\} \cup \{E' \in Q : E' \subset E^c\}$ for some $Q \in \mathbb{P}^*$.

Now, fix any $\omega \in E$ and suppose $f \in c(B|\omega)$. Since

$$c(B|\omega) = \arg\max_{g \in B} \int u \circ g d\pi(\cdot|Q_1(\omega))$$

and

$$c(B_{E,z}B'|\omega) = \arg \max_{f \in B_{E,z}B'} \int u \circ g d\pi(\cdot|Q_1(\omega))$$

and

$$\int u \circ f d\pi(\cdot|Q_1(\omega)) = \int u \circ fEz d\pi(\cdot|Q_1(\omega)),$$

it follows that $fEz \in c(B|\omega)$. Similar arguments show that the same property holds for E^c. Conclude that E is an isolatable event for $\mathbb{P}^*$. $\square$

This implies that $\mathbb{Q} \ll P$.

Lemma 24. *If E is an isolatable event for $\mathbb{P}^*$and F is a isolatable event $\mathbb{P}^*$, then $E \cap F$ is an isolatable event $\mathbb{P}^*$.*

Proof. Suppose E and F and isolatable events. Fix any $Q_1, Q_2 \in \mathbb{P}^{**}$. Since E is an isolatable event, there is a $Q_3 \gg \{E' \in Q_1 : E' \subset E\} \cup \{E' \in Q_2 : E' \subset E^c\} \in \mathbb{P}^{**}$. Since $Q_2, Q_3 \in \mathbb{P}^{**}$ and F is an isolatable event, there is a $Q_4 \gg \{E' \in Q_3 : E' \subset F\} \cup \{E' \in Q_2 : E' \subset F^c\} \in \mathbb{P}^{**}$. By construction, $Q_4 \gg \{Q_1 : E' \subset E \cap F\} \cup \{E' \in Q_2 : E' \subset (E \cap F)^c\}$. Since Q_1, Q_2 were arbitrary, this holds for any Q_1, Q_2. By Lemma 23, $E \cap F$ is a sub-problem. $\square$

Lemma 25. *If $Q \in \mathbb{P}^{**}$ and $E \in Q$, then $\pi_1(\cdot|E) = \pi_2(\cdot|E)$.*

Proof. Take any $Q \in \mathbb{P}^{**}$ and any $E \in Q$. Set $x, y \in int(u(X))$ so that $u(x) > u(y)$. WLOG, identify $x = u(x)$ and $y = u(y)$. Set ϵ so that

$$\pi_1(E)2\epsilon + (y - x)\pi_1(E') < 0$$

for all $E' \in P$. Take any simple $f, g \in \mathbb{R}^E$ so that $\int f d\pi_1(\cdot|E) \geq \int g d\pi_1(\cdot|E)$. There is an $\alpha \in (0, 1]$ so that $\alpha f + (1 - \alpha)x, \alpha g + (1 - \alpha)x \in [x - \epsilon, x + \epsilon]$. There are acts $f', g' \in \mathcal{F}$ so that $u(f')(\omega) = \alpha f(\omega) + (1 - \alpha)x$ for all $\omega \in E$ and $u(f')(\omega) = x$ otherwise and $u(g')(\omega) = \alpha g(\omega) + (1 - \alpha)x$ for all $\omega \in E$ and $u(g')(\omega) = x$ otherwise. Define $B = \{f'Q(\omega)y : \omega \in \Omega\} \cup \{g'Q(\omega)y : \omega \in \Omega\}$.

Claim that $\hat{P}(B) = Q$. If not, then there is a Q' so that $\hat{P}(B) = Q'$, so let $H \in \hat{c}(B)$ and consider H^*. It must be that $\int H^* d\pi_1 \geq \int f d\pi_1$. Since Q' is not finer than Q, there must be some $E' \in P$ so that $H^*(\omega) = y$ for all $\omega \in E'$. Let $E'' = \{\omega \in \Omega : u(H^*(\omega)) \geq u(f(\omega))\}$. By construction, $E'' \subset E$. Further, $2\epsilon \geq u(H^*(\omega)) - u(f(\omega))$. Therefore $\pi_1(E)2\epsilon + (y - x)\pi_1(E') + \int u \circ f' d\pi_1 \geq \int H^* d\pi_1$. However, $\int u \circ f' d\pi_1 > \pi_1(E)2\epsilon + (y - x)\pi_1(E') + \int u \circ f' d\pi_1$, contradicting that $\hat{P}(B) = Q'$.

Since $\int f d\pi_1(\cdot|E) \geq \int g d\pi_1(\cdot|E)$, $f'Q(\omega)x \in c(B|\omega)$. Since $\pi_2(\cdot)$ also represents $c(\cdot)$, $\int f d\pi_2(\cdot|E) \geq \int g d\pi_2(\cdot|E)$. Since f and g are arbitrary, $\int f d\pi_1(\cdot|E) \geq \int g d\pi_1(\cdot|E) \iff \int f d\pi_2(\cdot|E) \geq \int g d\pi_2(\cdot|E)$ for any f and g. Therefore, $\pi_1(\cdot|E) = \pi_2(\cdot|E)$ for every E. $\square$

Lemma 26. *If* $\pi_1(\cdot|E) = \pi_2(\cdot|E)$, $\pi_1(\cdot|E') = \pi_2(\cdot|E')$ *and* $E \cap E' \neq \emptyset$ *then* $\pi_1(\cdot|E \cup E') = \pi_2(\cdot|E \cup E')$.

Proof. Suppose that $\pi_1(\cdot|E) = \pi_2(\cdot|E)$, $\pi_1(\cdot|E') = \pi_2(\cdot|E')$ and $E \cap E' \neq \emptyset$. Note that $\pi_1(F|E) = \pi_2(F|E)$ and $\pi_1(F|E') = \pi_2(F|E')$. Using Bayes' rule on the event $E \cap E'$, it follows that

$$\frac{\pi_1(E|E \cup E')}{\pi_1(E'|E \cup E')} = \frac{\pi_2(E|E \cup E')}{\pi_2(E'|E \cup E')}.$$

For any $F \in \Sigma$, it holds that

$$\begin{aligned}\pi_1(F|E \cup E') &= \pi_1(E|E \cup E')(\pi_1(F|E) \\ &\quad -\pi_1(F \cap E'|E)) + \pi_1(E'|E \cup E')\pi_1(F|E')\end{aligned}$$

and that

$$\begin{aligned}\pi_2(F|E \cup E') &= \pi_2(E|E \cup E')(\pi_2(F|E) \\ &\quad -\pi_2(F \cap E'|E)) + \pi_2(E'|E \cup E')\pi_2(F|E').\end{aligned}$$

Because $\pi_1(E \cup E'|E \cup E') = \pi_2(E \cup E'|E \cup E') = 1$, conclude that $\pi_2(\cdot|E \cup E') = \pi_1(\cdot|E \cup E')$. □

Let Q^* be the finest common coarsening of $\mathbb{P}^{**}$. By successive applications of Lemma 26, we have that $\pi_1(\cdot|E') = \pi_2(\cdot|E')$ for all $E' \in Q^*$.

Lemma 27. *For any* $E \in \mathbb{Q}$, $\pi_1(\cdot|E) = \pi_2(\cdot|E)$.

Proof. To save on notation, write $\pi_1 = \pi_1(\cdot|E)$ and $\pi_2 = \pi_2(\cdot|E)$ and assume it is understood that each event E' is contained in E. Label the events in Q^* that are contained in E as $E_1, E_2, ..., E_n$. If $n = 1$, then we are done by the above, so assume $n \geq 2$.

Consider E_1 and E_2. By construction, there must be $Q_1, Q_2 \in \mathbb{P}^{**}$ so that $E' \subset E_1$ is in Q_1 but not Q_2, $E'' \subset E_2$ is in Q_2 but not in Q_1, and there is no $Q_3 \gg \{E' \in Q_1 : E' \subset E_1\} \cup \{E' \in Q_2 : E' \subset E_1^c\}$. Fix $x, y \in X$ so that $u(x) > u(y)$ and $u(x), u(y) \in int(u(X))$. Define $B_1 = \{xFy : F \in Q_1\}$ and $B_2 = \{xFy : F \in Q_2\}$. Let

$B_1^\epsilon = (B_1 \cup \{x'E'y\}) \backslash \{xE'y\}$ and $B_2^\epsilon = (B_1 \cup \{x'E''y\}) \backslash \{xE''y\}$ where $u(x') = u(x) + \epsilon$. For ϵ, ϵ' small enough but positive,

$$\hat{P}(B_1^\epsilon \cup B_2^{\epsilon'}) = \begin{cases} Q_1 & if\, \epsilon\pi_1(E') > \epsilon'\pi_1(E'') \\ Q_2 & if\, \epsilon\pi_1(E') < \epsilon'\pi_1(E''). \end{cases}$$

Therefore, there exists a $k = \frac{\pi_1(E'')}{\pi_1(E')}$ so that $\frac{\epsilon}{\epsilon'} > k$ implies $\hat{P}(B_1^\epsilon \cup B_2^{\epsilon'}) = Q_1$ and $\frac{\epsilon}{\epsilon'} < k$ implies $\hat{P}(B_1^\epsilon \cup B_2^{\epsilon'}) = Q_1$. Since π_2 also represents $c(\cdot)$, the same cutoff must hold for π_2. Therefore,

$$\frac{\pi_2(E'')}{\pi_2(E')} = \frac{\pi_1(E'')}{\pi_1(E')}.$$

By Lemma 26 and Bayes rule,

$$\frac{\pi_2(E_1)}{\pi_2(E_2)} = \frac{\pi_1(E_1)}{\pi_1(E_2)}.$$

By replacing E_1 with E_i and E_2 with E_{i+1}, we must have $\frac{\pi_2(E_i)}{\pi_2(E_{i+1})} = \frac{\pi_1(E_i)}{\pi_1(E_{i+1})}$. Since $\sum_{i=1}^n \pi_1(E_i) = \sum_{i=1}^n \pi_2(E_i) = 1$, we must have that $\pi_1(E_i) = \pi_2(E_i)$ for all i. □

Conclude that $\pi_1(\cdot|E) = \pi_2(\cdot|E)$ whenever E is a minimal isolatable event for $\mathbb{P}_1^*$, establishing the result. □

Proof of Lemma 1:

Proof. Fix a π_1 that represents $c(\cdot)$. It suffices to show that any π so that $\pi(\cdot|E) = \pi_1(\cdot|E)$ for every $E \in \mathbb{Q}$ also represents $c(\cdot)$. Fix any such π. It's clear that $\pi(\cdot|\hat{P}(B)(\omega)) = \pi_1(\cdot|\hat{P}(B)(\omega))$ for all ω because $\hat{P}(B) \gg \mathbb{Q}$. By Lemma 23,

$$\arg\max_{Q \in \mathbb{P}^*} \sum_{E \in Q} \pi(E)[\max_{f \in B} \int u \circ f d\pi(\cdot|Q(\omega))] =$$
$$\arg\max_{Q \in \mathbb{P}^*} \sum_{E \in Q} \pi_1(E)[\max_{f \in B} \int u \circ f d\pi_1(\cdot|Q(\omega))]$$

for every B. Conclude that $\pi(\cdot)$ also represents $c(\cdot)$. □

A.2.3 Proofs from Section 1.5

Assume $\mathcal{P}$ is finite and label $\mathcal{P} = \{P_1, ..., P_n\}$. Write $A\,\overline{IS}_P\,B$ if $A\,\overline{IS}\,B$ for $c_P(\cdot)$.

Axiom. (Agreement) If $\{f\}\ \overline{IS}_P\ \{g\}$ and not $\{g\}\ \overline{IS}_P\ \{f\}$, then for any $\alpha_1, ..., \alpha_n \in [0,1]$ and acts $f_1, ..., f_n, g_1, ..., g_n$ so that $\sum \alpha_i = 1$ and $\{g_i\}\ \overline{IS}_{P_i}\ \{f_i\}$, either $\sum \alpha_i g_i \neq g$ or $\sum \alpha_i f_i \neq f$.

Theorem 10. *If $\mathcal{P}$ is finite, each $c \in \{c_P(\cdot)\}_{P\in\mathcal{P}}$ has an optimal inattention representation and $\{c_P(\cdot)\}_{P\in\mathcal{P}}$ satisfies Agreement, then there is a $\pi(\cdot)$ and a $u(\cdot)$ so that each $c_Q(\cdot)$ is represented by $(u, \pi, \hat{Q}(\cdot), \mathbb{P}^*_Q)$.*

Proof. Define $\succeq_P$ on $\mathcal{F}$ by $f \succeq_P g$ if and only if $\{f\}\ \overline{IS}_P\ \{g\}$. $\succeq_P$ is a preorder that satisfies Gilboa et al. [2010]'s c-completeness, monotonicity and independence by Lemmas 8, 9, 11, and 12 respectively. Let B_0 be the set of simple, Σ-measurable, real functions. There exists a $u_P : X \to \mathbb{R}$ by Lemma 3. By agreement, it is WLOG to take the same $u(\cdot)$ for $u_P(\cdot)$ and assume that $0 \in int(u(X))$. For every $\succeq_P$, define a cone $K_P \subset B_0$ by

$$K_P = \{\lambda(u \circ f - u \circ g) : f \succeq_P g \, and \, \lambda \in \mathbb{R}_+\}$$

and let $K = co(\cup_{P\in\mathcal{P}} K_P)$. K is a cone: suppose $v \in K$. Then there are γ_i and $v_i \in K_i$ so that $\sum \gamma_i v_i = v$. But then $\lambda v_i \in K_i$ and consequently $\sum \gamma_i \lambda v_i = \lambda v$ and $\lambda v \in K$.

Note that Agreement implies that if there exist

$$f_1, ..., f_n, g_1, ..., g_n \in \mathcal{F}$$

and

$$\alpha_1, ..., \alpha_n \in [0,1]$$

so that $\sum \alpha_i = 1$, $f_i \succeq_{P_i} g_i$, $f = \sum \alpha_i f_i$ and $g = \sum \alpha_i g_i$, then $g \not\succ_P f$ for any $P \in \mathcal{P}$.

Identify $\hat{f}$ with $u \circ f \in B_0$. Define $\succeq$ by $f \succeq g$ if and only if $\hat{f} - \hat{g} \in K$. Claim that $\succeq$ extends each $\succeq_P$ compatibly. First, note that if $f \succeq_P g$, then $\hat{f} - \hat{g} \in K_P \subset K$, so $f \succeq g$.

Suppose $g \succ_{P_i} f$ but $f \succeq g$, so that $\hat{f} - \hat{g} = v \in K$. Then there are $v_1, ..., v_n$ so that $v_i \in K_{P_i}$ and $\gamma_1, ...\gamma_n \in \mathbb{R}_+$ so that

$$\sum \gamma_i v_i = v.$$

Consequently, there is a λ_i and two acts f_i, g_i so that $f_i \succeq_{P_i} g_i$ and $\lambda_i(\hat{f}_i - \hat{g}_i) = v_i$. Rewriting,

$$\hat{f} - \hat{g} = \sum \lambda_i \gamma_i \hat{f}_i - \sum \lambda_i \gamma_i \hat{g}_i$$

so defining $h(\omega) = \sum \lambda_i \gamma_i \hat{f}_i(\omega) - \hat{f}(\omega) = \sum \lambda_i \gamma_i \hat{g}_i(\omega) - \hat{g}(\omega)$ gives that

$$\hat{f} + h = \sum \lambda_i \gamma_i \hat{f}_i$$

and

$$\hat{g} + h = \sum \lambda_i \gamma_i \hat{g}_i.$$

Moreover, by mixing f, g, f_i, g_i with an act 0 so that $u(0(\omega)) = 0$ in every state at a given probability, we can take $h \in u(X)^\Omega$. Now, we have that $\frac{1}{2}f + \frac{1}{2}h = \sum \lambda_i \gamma_i (\frac{1}{2}f_i + \frac{1}{2}0)$ and $\frac{1}{2}g + \frac{1}{2}h = \sum \lambda_i \gamma_i (\frac{1}{2}g_i + \frac{1}{2}0)$. Conclude that there are acts $f'_i = \frac{1}{2}f_i + \frac{1}{2}0$ and $g'_i = \frac{1}{2}g_i + \frac{1}{2}0$ and $\alpha_i \in [0,1]$ so that $f'_i \succeq_{P_i} g'_i$ for every i and that $\frac{1}{2}f + \frac{1}{2}h = \sum \alpha_i f'_i$ and $\frac{1}{2}g + \frac{1}{2}h = \sum \alpha_i g'_i$. By Agreement, it is not the case that $\frac{1}{2}g + \frac{1}{2}h \succ_{P_i} \frac{1}{2}f + \frac{1}{2}h$. However $g \succ_{P_i} f \iff \frac{1}{2}g + \frac{1}{2}h \succ_{P_i} \frac{1}{2}f + \frac{1}{2}h$ because $\succeq_{P_i}$ satisfies Independence, a contradiction.

Now, $\succeq$ has an Aumann utility for the same reasons as above. Further, it also satisfies Independence, monotonicity, reflexivity, transitivity and continuity from Gilboa et al. [2010]. Conclude from their Theorem 1 that $\succeq$ has a unique set of priors Π so that $f \succeq g \iff \int u \circ f d\pi \geq \int u \circ g d\pi$ for all $\pi \in \Pi$. The prior from the Aumann utility must be in Π by routine arguments. □

Proof of Theorem 4:

Proof. First, suppose that $c(\cdot)$ is more attentive than $c'(\cdot)$. Fix an arbitrary $Q \in \mathbb{P}^*_{c'}$ and $x, y \in X$ so that $u_{c'}(x) > u_{c'}(y)$. Define the problem B by $\{xEy : E \in Q\}$. By construction, $\hat{P}_{c'}(B) \gg Q$ so $\hat{\hat{P}}_{c'}(B) = Q$. It follows from $c(\cdot)$ more attentive than $c'(\cdot)$ that there exists a B' so that $\hat{\hat{P}}_c(B') = Q$. Consequently, $\hat{P}_c(B') \in \mathbb{P}^*_c$ and $\hat{P}_c(B') \gg \hat{\hat{P}}_c(B') = Q$ so $Q \in \mathbb{P}^*_c$.

Now, suppose that $\mathbb{P}^*_{c'} \subset \mathbb{P}^*_c$. Fix an arbitrary B and suppose that $\hat{\hat{P}}_{c'}(B) = Q$. It follows immediately that $Q \in \mathbb{P}^*_{c'}$. Fix $x, y \in X$ so that $u_c(x) > u_c(y)$. Define the problem B' by $\{xEy : E \in Q\}$. By construction, $\hat{P}_c(B') \gg Q$, implying that $\hat{\hat{P}}_c(B') = Q$. Since B is arbitrary, there exists such a B' for every B. It follows that $c(\cdot)$ is more attentive than $c'(\cdot)$. □

Proof of Theorem 5:

Proof. It is clear that $\mathbb{P}^*_Q \subset \mathbb{P}^*_P$ implies that that P is more valuable than Q. From Theorem 4, this follows from $c_P(\cdot)$ has a higher capacity for attention than $c_Q(\cdot)$.

Suppose, for the sake of contradiction, that P is not more valuable than Q but that $c_P(\cdot)$ does not have a higher capacity for attention than $c_Q(\cdot)$. Then there must be some u, π, $B \in K(\mathcal{F})$ so that

$$\max_{Q' \in \mathbb{P}^*_Q} V(u, \pi, B, Q') > \max_{Q'' \in \mathbb{P}^*_P} V(u, \pi, B, Q'').$$

Let $Q^* \in \arg\max_{Q' \in \mathbb{P}^*_Q} V(u, \pi, B, Q')$. From Theorem 4, $\mathbb{P}^*_Q \subset \mathbb{P}^*_P$, so $Q^* \in \mathbb{P}^*_P$. Therefore,

$$\max_{Q'' \in \mathbb{P}^*_P} V(u, \pi, B, Q'') \geq V(u, \pi, B, Q^*) = \max_{Q' \in \mathbb{P}^*_Q} V(u, \pi, B, Q'),$$

a contradiction. □

A.2.4 Proofs from Section 1.6

Proof of Proposition 2

Proof. Let p be an equilibrium to the market ϕ where there is at least one firm of each type. For $i = 1, ..., m$, let $p_i = \min\{p_j : \phi_j = i\}$ where $\min(\emptyset) = \infty$. Relabel so that $p_1 \leq p_2 \leq ...$ and suppose that $p_i \leq 1$ for $i \leq \kappa$ (this is WLOG because no firm of type i sells anything if $p_i > 0$). I claim that $p_{\kappa-1} = 0$. By assumption, there is at least one firm, j^*, so that $\phi_{j^*} = \kappa + 1$.

Suppose $p_{\kappa-1} > 0$. If the consumer purchases a good of type i, she purchases from the firm that charges p_i by monotonicity. If she pays attention to the partition

$$\{\{1\}, \{2\}, ..., \{\kappa, \kappa + 1, ..., m\}\}, \tag{A.2.2}$$

then her expected utility is

$$\sum_{i=1}^{\kappa-1} \frac{1}{m}(1 - p_i)$$

which is a maximum, unless $\frac{1}{m}(1 - p_\kappa) \geq p_1(m - \kappa)$, in which case

$$\{\{1\}, \{2\}, \{3\}, ..., \{\kappa\}, \{j, \kappa + 1, ...m\}\} \backslash \{j\} \tag{A.2.3}$$

for some $1 \leq j \leq \kappa$ so that $p_j = p_1$ attains an expected utility of

$$\sum_{i=1}^{\kappa} \frac{1}{m}(1 - p_i) - (m - \kappa)p_1.$$

In either case, firm j^* makes zero profit.

Suppose j^* deviates to charging a price $\frac{p_{\kappa-1}}{2} > 0$. Now, the optimal partition is either

$$\{\{1\}, \{2\}, ..., \{\kappa + 1\}, \{\kappa - 1, \kappa, \kappa + 2, ..., m\}\}$$

if (A.2.2) was maximal in the first problem or

$$\{\{1\}, \{2\}, \{3\}, ..., \{\kappa - 1\}, \{\kappa + 1\}, \{j, \kappa, \kappa + 2, ..., m\}\} \backslash \{j\}$$

for j as above if (A.2.3) was maximal in the first problem. In state $\kappa+1$, the consumer purchases from j^* so it attains an expected profit of

$$\frac{p_{\kappa-1}}{2m} > 0$$

which is a profitable deviation and contradicts that p is an equilibrium where $p_{\kappa-1} > 0$. Therefore, $p_{\kappa-1} = 0$ in any equilibrium.

Since $p_1 = 0$, the second partition is always optimal. Suppose $1 \geq p_\kappa > 0$, so the consumer purchases from a firm of type κ and pays a positive price in state κ in the equilibrium. Firm j^*can charge $\frac{p_\kappa}{2}$ and attract customers to make positive profit. Therefore, $p_\kappa = 0$ so κ firms of different types charge price 0 in equilibrium and the consumer purchases from one of these firms no matter what the state is. □

Proof of Proposition 3

Proof. Expected total surplus is equal to the probability that the consumer purchases the good from a firm whose type matches the state. Clearly, this probability can be no larger than $\frac{n_m(\phi)+n_c(\phi)}{m}$. Further, the consumer makes at most κ different purchases, so it can also be no larger than $\frac{\kappa}{m}$. If $k \geq n_m(\phi) + n_c(\phi)$ but the consumer makes less than $n_m(\phi) + n_c(\phi)$ different purchases or $n_m(\phi) + n_c(\phi) > \kappa$ and the consumer makes less than κ, then a firm from which the consumer does not purchase can make positive profit by charging a price equal to $\epsilon > 0$ for ϵ small enough. Consequently, in equilibrium, the consumer purchases the good from a firm whose type matches the state with probability equal to $\max\{\frac{n_m(\phi)+n_c(\phi)}{m}, \frac{\kappa}{m}\}$.

For expected consumer surplus, if $\kappa < n_m(\phi) + n_c(\phi)$, the result follows from Proposition 2 and the above discussion. If $\kappa \geq n_m(\phi)+n_c(\phi)$, then it is easy to verify that in equilibrium, the monopolistic firms charge price 1 and at least one competitive firm of each type charges price 0. Consequently, expected consumer surplus equals $\frac{n_c(\phi)}{m}$. □

A.3 Counter-Examples

For the following, set $\Omega = \{\omega_1, \omega_2, \omega_3\}$, $P = \{\{\omega_1\}, \{\omega_2\}, \{\omega_3\}\}$ and $X = \Delta\mathbb{R}$. To economize on space, I write (a, b, c) for an act that gives a for sure in state ω_1, b for sure in state ω_2 and c for sure in state ω_3 and $c(B|\cdot) = (c(B|\omega_1), c(B|\omega_2), c(B|\omega_3))$.

A.3.1 All but ACI

Suppose $\mathbb{P}^* = \{Q \ll P : \#Q \leq 2\}$, $u(x) = x$ and $\pi(\omega) = \frac{1}{3}$ for every ω. Define a cost function ρ so that $\rho(Q) = \#Q - 1$ if $Q \in \mathbb{P}^*$ and $\rho(Q) = \infty$ otherwise. If

$$\hat{P}(B) = \max_{>} \arg\max_{Q \in \mathbb{P}^*} \sum_{E \in Q} \pi(E)[\max_{f \in B} \int u \circ f d\pi(\cdot|Q(\omega))] - \rho(Q)$$

where $\{\{\omega_1\},\{\omega_2,\omega_3\}\} > \{\{\omega_2\},\{\omega_1,\omega_3\}\} > \{\{\omega_3\},\{\omega_1,\omega_2\}\} > \{\Omega\}$ and Equation (1.2.2) holds, then $c(\cdot)$ violates ACI but satisfies the other 4 axioms. Define f, g, h by $(2,0,0)$, $(0,2,2)$ and $(0,0,0)$, respectively, and let $B_\alpha = \alpha\{f,g\} + (1-\alpha)\{h\}$. Note that $c(B_1|\cdot) = (\{f\},\{g\},\{g\})$. For $\alpha \geq \frac{1}{2}$, $\hat{P}(B_\alpha) = \{\{\omega_1\},\{\omega_2,\omega_3\}\}$, but for $\alpha < \frac{1}{2}$ $\hat{P}(B_\alpha) = \{\Omega\}$. Therefore,

$$c(B_{\frac{1}{4}}|\omega_1) = \{\frac{1}{4}g + \frac{3}{4}h\} \neq \{\frac{1}{4}f + \frac{3}{4}h\} = \frac{1}{4}c(B_1|\omega_1) + \frac{3}{4}\{h\},$$

contradicting ACI.[52] Monotonicity and Subjective Consequentialism are clearly satisfied. To see that INRA is satisfied, note that by Equation (1.2.2), if $A \subset B$ and $A \cap c(B|\omega) \neq \emptyset$ for all ω, then $\hat{P}(A) = \hat{P}(B)$. Again using equation (1.2.2), we have that $c(A|\omega) = c(B|\omega) \cap A$.

A.3.2 All but INRA

Keeping $\mathbb{P}^*$, $u(\cdot)$, $\pi(\cdot)$ and $>$ as above, suppose that

$$\hat{P}(B) = \max_{>} \arg\min_{Q \in \mathbb{P}^*} \sum_{E \in Q} \pi(E)[\min_{f \in B} \int u \circ f d\pi(\cdot|Q(\omega))]$$

for every B and that Equation (1.2.2) holds for each B and ω. This $c(\cdot)$ violates INRA. Take f, g, h, j so that $f = (3,1,2)$, $g = (1,3,1)$, $h = (1,0,0)$ and $j = (0,1,1)$. If $B = \{f,g,h,j\}$ and $A = \{f,g\}$, then $\hat{P}(B) = \{\{\omega_1\},\{\omega_2,\omega_3\}\}$ while $\hat{P}(A) = \{\{\omega_2\},\{\omega_1,\omega_3\}\}$. Consequently, $c(B|\cdot) = (\{f\},\{g\},\{g\})$ and $c(A|\cdot) = (\{f\},\{g\},\{f\})$, contradicting INRA.[53] Equation (1.2.2) implies that Subjective Con-

[52] Similar choices occur for any B' with $d(B',B) < \epsilon$ for ϵ suitably small.
[53] Similar choices occur for any B' with $d(B',B) < \epsilon$ for ϵ suitably small.

sequentialism and Monotonicity hold. To see why ACI holds, note that

$$\min_{f\in\alpha B+(1-\alpha)\{g\}}\int u\circ f d\pi(\cdot|E)=\alpha\min_{f\in B}\int u\circ f d\pi(\cdot|E)+(1-\alpha)\int u\circ g d\pi(\cdot|E)$$

for any B, g and E. This implies that $\hat{P}(\alpha B+(1-\alpha)\{g\})=\hat{P}(B)$, and Equation (1.2.2) gives that ACI holds.

A.3.3 All but Monotonicity

Let $v(x,\omega_1)=x$ and $v(x,\omega_2)=v(x,\omega_3)=-x$. Define

$$c(B|\omega)=\arg\max_{f\in B}\sum_{\omega\in\Omega}v(f(\omega),\omega)$$

and note that $0\in c(\{0,1\}|\omega)$ for all ω. Set $f=(1,0,0)$ and $B=\{f,0\}$. Since $\sum_{\omega\in\Omega}v(f(\omega),\omega)=1$ and $\sum_{\omega\in\Omega}v(0,\omega)=0$, $\{f\}=c(B|\omega)$. However,

$$0\in c(\{0,f(\omega)\}|\omega)$$

for all ω, so Monotonicity is contradicted. It is trivial to verify that the other axioms are satisfied.

A.3.4 All but Subjective Consequentialism

Return to the setup from the first two counter-examples. Set $\pi_1(\omega_1)=\pi_2(\omega_3)=\frac{1}{2}$, $\pi_1(\omega_2)=\pi_1(\omega_3)=\pi_2(\omega_1)=\pi_2(\omega_2)=\frac{1}{4}$, and $\pi_3=\pi_2$. Suppose that

$$c(B|\omega_i)=\arg\max_{f\in B}\int u\circ f d\pi_i$$

and consider $f=(4,2,2)$, $g=(4,2,0)$, $h=(0,4,5)$ and $B=\{f,g,h\}$. By construction $c(B|\cdot)=(\{f\},\{h\},\{h\})$. Note that $\{\omega_1\}=\{\omega'':c(B|\omega'')=c(B|\omega_1)\}$ and that $f(\omega_1)=g(\omega_1)$, a contradiction of subjective consequentialism. The other properties are trivial to verify.

A.3.5 All but Continuity

Take $\mathbb{P}^*$, $u(\cdot)$ and $\pi(\cdot)$ as in the first example. Write $P_i = \{\{\omega_i\}, \{\omega_i\}^c\}$. For every problem B, define an ordering $>_B$ by $P_i >_B P_j$ if and only if

$$\begin{aligned}
&\max_{f\in B} u(f(\omega_i))) > \max_{f\in B} u(f(\omega_j))\, OR \\
&\quad [\max_{f\in B} u(f(\omega_i))) = \max_{f\in B} u(f(\omega_j))\, AND \\
&\qquad (\max_{f\in B} \int u\circ f d\pi(\cdot|\{\omega_i\}^c) > \max_{f\in B} \int u\circ f d\pi(\cdot|\{\omega_j\}^c)] \\
&\quad OR\, [\max_{f\in B} u(f(\omega_i))) = \max_{f\in B} u(f(\omega_j))\, AND \\
&\qquad (\max_{f\in B} \int u\circ f d\pi(\cdot|\{\omega_i\}^c) = \max_{f\in B} \int u\circ f d\pi(\cdot|\{\omega_j\}^c) \\
&\qquad\quad AND\, i < j]
\end{aligned}$$

Also, set every $P_i >_B \{\Omega\}$. For every problem B, take $\hat{P}(B) = \max_{>_B} \mathbb{P}^*$ and suppose Equation (1.2.2) holds. I will show that Continuity fails. Set $f = (1, .9., 1)$. and $g = (1, .8, 1)$, noting that f dominates g but g does not dominate f. Set $h = (1, 0, 1)$, $j = (0, .9, 0)$ and $k = (0, 0, 0)$, and for every $n > 2$ define $B_{n,1} = \{g\}$,

$$B_{n,2} = \{g, \frac{n-1}{n}h + \frac{1}{n}k, \frac{n-1}{n}j + \frac{1}{n}k\},$$

$B_{n,3} = \{\frac{n-1}{n}h + \frac{1}{n}k, \frac{n-1}{n}j + \frac{1}{n}k, \frac{n-2}{n}f + \frac{2}{n}k\}$, $B_{n,4}\{, \frac{n-2}{n}f + \frac{2}{n}k\}$. Note that

$$(\{g\}, \{g\}, \{g\}) = c(B_{n,2}|\cdot),$$

that

$$(\{\frac{n-1}{n}h + \frac{1}{n}k\}, \{\frac{n-1}{n}j + \frac{1}{n}k\}, \{\frac{n-1}{n}h + \frac{1}{n}k\}) = c(B_{n,3}|\cdot),$$

and that

$$(\{\frac{n-2}{n}f+\frac{2}{n}k\},\{\frac{n-2}{n}f+\frac{2}{n}k\},\{\frac{n-2}{n}f+\frac{2}{n}k\}) = c(\{\frac{n-2}{n}f+\frac{2}{n}k\}|\cdot).$$

Therefore $\{g\}\, IS\, \{\frac{n-2}{n}f+\frac{2}{n}k\}$ for every $n>2$, and as $n\to\infty$,

$$\frac{n-2}{n}f+\frac{2}{n}k \to f.$$

Conclude that $\{g\}\, \overline{IS}\, \{f\}$ and f dominates g but g does not dominate f, contradicting Continuity. One can verify easily that the other four axioms are satisfied.

A.3.6 Behavior compatible with optimal inattention but not inattention to alternatives

Suppose $\Omega = \{a,b,c,d\}$ and $P = \{\{a\},\{b\},\{c\},\{d\}\}$. Consider π so that $\pi(\omega) = \frac{1}{4}$ for every ω and $\mathbb{P}^* = \{Q : Q \ll Q_1\} \cup \{Q : Q \ll Q_2\}$ where $Q_1 = \{\{a\},\{b,c\},\{d\}\}$ and $Q_2 = \{\{a,d\},\{b\},\{c\}\}$. Define acts x,y,z,w that give the utility values in the following table:

	a	b	c	d
$u\circ w$	6	6	6	4
$u\circ x$	8	9	0	0
$u\circ y$	0	0	0	16
$u\circ z$	2	0	9	0

One can verify that $\hat{P}(\{x,y,z,w\}) = Q_1$, $\hat{P}(\{x,z,w\}) = Q_2$, $\hat{P}(\{x,y,z\}) = Q_2$, and $\hat{P}(\{y,z\}) = Q_1$, so $c(\{x,y,z,w\}|a) = \{x\}$, $c(\{x,z,w\}|a) = \{w\}$, $c(\{x,y,z\}|a) = \{y\}$, and $c(\{y,z\}) = \{z\}$. But then by Lemma 1 and Theorem 3 of Masatlioglu et al. [2012], xPy and yPx so $c(\cdot|a)$ cannot be a choice with limited attention.

A.3.7 Behavior compatible with inattention to alternatives but not optimal inattention

Use the same setup as in C.6. Fix $x, y, z \in X$, i.e. all three are lotteries. Suppose $c(\cdot|a)$ is a choice with limited attention where $\Gamma(\{x, y, z\}) = \{y, z\}$, $\Gamma(\{x, y\}) = \{x, y\}$ and $x \succ y \succ z$. Then $c(\{x, y, z\}|a) = \{y\}$ and $c(\{x, y\}|a) = \{x\}$. If $c(\cdot)$ has an optimal inattention representation, then $c(\{x, y, z\}|a) = \{y\}$ implies $u(y) > u(x)$ but $c(\{x, y\}|a) = \{x\}$ implies $u(x) > u(y)$, a contradiction.

Appendix B

Appendix for Chapter 2

B.1 Proofs not in text

B.1.1 Proof of Lemma 2

I will show that

$$D(p,\phi;\bar{p},\bar{\phi}) = \phi D_I(p,\bar{p},\bar{\phi}) + (1-\phi)D_U(p,\bar{p},\bar{\phi}),$$

where:

$$\begin{aligned}
D_U(p,\bar{p},\bar{\phi}) &= \delta_U(\bar{\phi}) + \frac{\bar{p}-p}{t}\Delta_U(\bar{\phi}),\\
D_I(p,\bar{p},\bar{\phi}) &= \delta_I(\bar{\phi}) + \frac{\bar{p}-p}{t}\Delta_I(\bar{\phi}),\\
\delta_U(\bar{\phi}) &= \tfrac{(1-(1-q)^n)}{n} + \tfrac{1}{n}\textstyle\sum_{l=\kappa+1}^{n}[\sum_{z=l-\kappa}^{n-\kappa-1}\sum_{i=l-\kappa}^{z}\binom{n-l}{z-i}\binom{l-1}{i}\bar{\phi}^z(1-\bar{\phi})^{n-1-z}(1-q)^{\kappa+i}q\\
&\quad -F(n-\kappa-1;\bar{\phi},n-1)(1-q)^{l-1}],\\
\Delta_U(\bar{\phi}) &= q[1-(1-q)^{n-1}(1-F(n-\kappa-1;\bar{\phi},n-1))],\\
\delta_I(\bar{\phi}) &= \tfrac{(1-(1-q)^n)}{n} + \tfrac{1}{n}\textstyle\sum_{l=\kappa+1}^{n}[\sum_{z=0}^{n-\kappa-2}\sum_{i=0}^{z}\binom{n-l}{z-i}\binom{l-1}{i}\bar{\phi}^z(1-\bar{\phi})^{n-1-z}(1-q)^{\kappa+i}q\\
&\quad -F(n-\kappa-2;\bar{\phi},n-1)(1-q)^{l-1}q],\ and\\
\Delta_I(\bar{\phi}) &= q[1-(1-q)^{n-1}(1-F(n-\kappa-2;\bar{\phi},n-1))-\textstyle\sum_{i=0}^{n-\kappa-2}f(i;\bar{\phi},n-1)(1-q)^{\kappa+i}].
\end{aligned}$$

This follows from the discussion following Lemma 2 and noting that if firm j charges price p (which is "close" to $\bar{p}$), then

$$N_1 = \frac{\bar{p} - p}{t} + \frac{1}{n},$$

$N_2 = N_3 = ... = N_{n-1} = \frac{1}{n}$ and

$$N_n = \frac{p - \bar{p}}{t} + \frac{1}{n}.$$

See Grossman and Shapiro [1984] for explicit derivation of the measure of these groups.

Combining the above cases, the expected demand for firm j among the consumers who do not see her ad equals

$$\sum_{l=1}^{\kappa} N_l (1-q)^{l-1} q + \sum_{l=\kappa+1}^{n} N_l [1 - F(n-\kappa-1; \bar{\phi}, n-1)](1-q)^{l-1} q$$
$$+ \sum_{l=\kappa+1}^{n} N_l [\sum_{z=l-\kappa+1}^{n-\kappa-1} \sum_{i=l-\kappa+1}^{z} \binom{n-l}{z-i} \binom{l-1}{i} \bar{\phi}^z (1-\bar{\phi})^{n-1-z} (1-q)^{l-1} q].$$

Using the formulas above for N_l yields that

$$D_U(p, \bar{p}, \bar{\phi}) \quad = \quad \delta_U(\bar{\phi}) + \frac{\bar{p} - p}{t} \Delta_U(\bar{\phi}),$$

where δ_U and Δ_U are as above. Similarly, the expected demand for firm j among the consumers who see her ad equals

$$\sum_{l=1}^{\kappa+1} N_l (1-q)^{l-1} q + \sum_{l=\kappa+2}^{n} N_l [1 - F(n-\kappa-2; \bar{\phi}, n-1)](1-q)^{l-1} q$$
$$+ \sum_{l=\kappa+2}^{n} N_l \{ \sum_{z=0}^{n-\kappa-2} [\sum_{i=0}^{l-\kappa-1} \binom{n-l}{z-i} \binom{l-1}{i} \bar{\phi}^z (1-\bar{\phi})^{n-1-z} (1-q)^{\kappa+i} q$$
$$+ \sum_{i=l-\kappa}^{z} \binom{n-l}{z-i} \binom{l-1}{i} \bar{\phi}^z (1-\bar{\phi})^{n-1-z} (1-q)^{l-1} q] \}.$$

Using the formulas above for N_l yields that

$$D_I(p,\bar{p},\bar{\phi}) \quad = \quad \delta_I(\bar{\phi}) + \frac{\bar{p}-p}{t}\Delta_I(\bar{\phi}),$$

where δ_I and Δ_I are as above. Conclude that firm j has an expected demand given by (2.3.4). Note that $D_I(p,\bar{p},\bar{\phi}) \geq D_U(p,\bar{p},\bar{\phi})$, and that both are weakly increasing in $\bar{\phi}$. Therefore, the demand for firm j is also weakly increasing in $\bar{\phi}$.

B.1.2 Proof of Proposition 5

(iii) and (iv) are obvious from equations (2.4.1) and (2.4.2).

[(i)] Note that

$$\frac{\partial p^*}{\partial q} = -t\frac{1-(1-q)^\kappa + \kappa(1-q)^{\kappa-1}q}{nq^2}$$

using Equation (2.4.2). Since $1 > (1-q)^\kappa$ and $nq^2 > 0$ for every $q \in (0,1)$, it follows that $\frac{\partial p^*}{\partial q} < 0$.

Moving on to π^*, note that

$$\pi^* = t\frac{(1-(1-q)^\kappa)^2}{qn^2}.$$

Taking the first derivative,

$$\frac{\partial \pi^*}{\partial q} = -t(1-(1-q)^\kappa)\frac{1-(1-q)^\kappa + 2\kappa(1-q)^{\kappa-1}q}{n^2q^2}.$$

This is positive if and only if

$$1-(1-q)^\kappa + 2\kappa(1-q)^{\kappa-1}q < 0$$

which holds if and only if

$$\frac{1-(1-q)^\kappa}{(1-q)^{\kappa-1}q} < 2\kappa.$$

Note that q^* is the value making the first derivative equal to 0. I know show that the LHS is decreasing in q. Taking the first derivative yields that

$$\frac{\partial \frac{1-(1-q)^{\kappa}}{(1-q)^{\kappa-1}q}}{\partial q} = \frac{1-(1-q)^{\kappa}-\kappa q}{q^2(1-q)^k}.$$

Noting that $1-(1-q)^{\kappa}-\kappa q = 0$ when $q = 0$, since

$$\frac{\partial(1-(1-q)^{\kappa}-\kappa q)}{\partial q} = \kappa(1-q)^{\kappa-1} - \kappa \leq 0$$

the LHS is decreasing in q. Therefore, whenever $q < q^*$, π^* is decreasing in q, and whenever $q > q^*$, π^* is increasing in q.

[(ii)] To see that p^* increases as κ increases, note that $(1-q)^{\kappa+1} - (1-q)^{\kappa} = (1-q)^{\kappa}((1-q)-1) = -q(1-q)^{\kappa} < 0$. Therefore, $t\frac{1-(1-q)^{\kappa+1}}{q} > t\frac{1-(1-q)^{\kappa}}{q}$. Similarly, as κ increases, demand also increases. Therefore, π^* increases. □

B.1.3 Proof of Proposition 6

The key step is showing that total transportation costs in equilibrium are

$$t\frac{2-q+(1-q)^{\kappa}(q-2+2q\kappa)}{4nq}.$$

If I show this, then the rest of the proposition follows from the fact that a consumer purchases a good only if it is high quality and she process information about it. The probability of paying attention to information that reveals at least one firm is high quality equals $1-(1-q)^{\kappa}$.

In equilibrium, a consumer purchases from the closest firm with probability q, the second with probability $(1-q)q$ and so on. However, the probability of purchasing from the the jth closes firm when $j \geq \kappa+1$ is 0. Following Grossman and Shapiro

[1984], the aggregate cost of traveling to the jth closest firm if there are n firms is

$$t\frac{2j-1}{4n}$$

so transportation costs are given by

$$\sum_{j=1}^{\kappa} t\frac{2j-1}{4n}q(1-q)^{j-1} = t\frac{2-q+(1-q)^{\kappa}(q-2+2q\kappa)}{4nq},$$

which was to be verified. □

B.1.4 Details for $p^* = v - t/2$

If other firms charge $\bar{p} = v - \frac{t}{2}$ and firm j charges $p > \bar{p}$, then less than the entire market is served. In particular, when p is close to $\bar{p}$, firm j has a demand of

$$D_{I2}(p,\bar{p},\bar{\phi}) \quad = \quad \delta_I(\bar{\phi}) + \frac{\bar{p}-p}{t}\Delta_{I2}(\bar{\phi}),$$

where $\delta_I(\bar{\phi})$ is as above and

$$\Delta_{I2}(\bar{\phi}) \quad = \quad q[1+(1-q)^{n-1}(1-F(n-\kappa-2;\bar{\phi},n-1))+\textstyle\sum_{i=0}^{n-\kappa-2}\binom{n-1}{i}\bar{\phi}^i(1-\bar{\phi})^{n-1-i}(1-q)^{\kappa+i}].$$

from the consumers who see its ad, and a demand of

$$D_{U2}(p,\bar{p},\bar{\phi}) \quad = \quad \delta_U(\bar{\phi}) + \frac{\bar{p}-p}{t}\Delta_{U2}(\bar{\phi})$$

where $\delta_U(\bar{\phi})$ is as above and

$$\Delta_{U2}(\bar{\phi}) = q[1+(1-q)^{n-1}(1-F(n-\kappa-1;\bar{\phi},n-1))].$$

If price $p^* = v - \frac{t}{2}$ is an equilibrium, then

$$\bar{p}(\bar{\phi}) \geq v - \frac{t}{2} \geq \bar{\bar{p}}(\bar{\phi}) = t\frac{\bar{\phi}\delta_I(\bar{\phi}) + (1-\bar{\phi})\delta_U(\bar{\phi})}{2[\bar{\phi}\Delta_{I2}(\bar{\phi}) + (1-\bar{\phi})\Delta_{U2}(\bar{\phi})]}.$$

To see why, fix $\bar{\phi}$ and suppose that all other firms set $\bar{p} = v - \frac{t}{2}$. A price $p < \bar{p}$ is best response only if p satisfies Equation (2.5.1). A price $p > \bar{p}$ is a best response only if

$$p - c/2 = \frac{\bar{p}}{2} + t\frac{\bar{\phi}\delta_I(\bar{\phi}) + (1-\bar{\phi})\delta_U(\bar{\phi})}{2[\bar{\phi}\Delta_{I2}(\bar{\phi}) + (1-\bar{\phi})\Delta_{U2}(\bar{\phi})]} = \frac{\bar{p}}{2} + \frac{\bar{\bar{p}}(\bar{\phi})}{2}.$$

If $\bar{p}(\bar{\phi}) \geq v - \frac{t}{2} \geq \bar{\bar{p}}(\bar{\phi})$, then the only possible best response is $p = \bar{p}$.

B.1.5 Proof of Proposition 8

The Proposition claims only necessity. This can be established using the Envelope Theorem and Proposition 7. I provide more detailed arguments below because they provide sufficient conditions as well. The expected demand for firm j from setting price p and advertising ϕ is

$$\begin{aligned} D(p, \phi, \bar{p}, \bar{\phi}) &= \delta_U + \phi(\delta_I - \delta_U) + \frac{\bar{p} - p}{t}[\phi\Delta_I + (1-\phi)\Delta_U] \\ &= \delta_U + \phi(\delta_I - \delta_U) + \frac{\bar{p} - p}{t}C(\phi, \bar{\phi}). \end{aligned}$$

Fixing $\phi, \bar{\phi}$ and $\bar{p}$, $pD(p, \phi, \bar{p}, \bar{\phi})$ is a concave function of p. The first order condition for p is given by

$$0 = \delta_U + \phi(\delta_I - \delta_U) + \frac{\bar{p}}{t}C(\phi, \bar{\phi}) - \frac{2p}{t}C(\phi, \bar{\phi})$$

so

$$p = \frac{\bar{p}}{2} + \frac{t(\delta_U + \phi(\delta_I - \delta_U))}{2(C(\phi, \bar{\phi}))}.$$

In a symmetric equilibrium, $\bar{p} = \frac{t(\delta_U + \bar{\phi}(\delta_I - \delta_U))}{C(\bar{\phi},\bar{\phi})}$, so p is a function of only $\bar{\phi}$ and ϕ:

$$p(\phi,\bar{\phi}) = t[\frac{\delta_U + \bar{\phi}(\delta_I - \delta_U)}{2C(\bar{\phi},\bar{\phi})} + \frac{\delta_U + \phi(\delta_I - \delta_U)}{2C(\phi,\bar{\phi})}]$$

Assuming she prices at $p(\phi,\bar{\phi})$, expected demand as a function of ϕ and $\bar{\phi}$ can be written as

$$\begin{aligned} D(\phi,\bar{\phi}) &= \delta_U + \phi(\delta_I - \delta_U) + \frac{\bar{p} - p}{t} C(\phi,\bar{\phi}) \\ &= \delta_U + \phi(\delta_I - \delta_U) + C(\phi,\bar{\phi})[\frac{\delta_U + \bar{\phi}(\delta_I - \delta_U)}{2C(\bar{\phi},\bar{\phi})} + \frac{\delta_U + \phi(\delta_I - \delta_U)}{2C(\phi,\bar{\phi})}] \\ &= \frac{\delta_U + \phi(\delta_I - \delta_U)}{2} + (\delta_U + \bar{\phi}(\delta_I - \delta_U))\frac{C(\phi,\bar{\phi})}{2C(\bar{\phi},\bar{\phi})} \end{aligned}$$

so

$$\begin{aligned} D'(\phi,\bar{\phi}) &= \frac{\delta_I - \delta_U}{2} + \frac{C'(\phi,\bar{\phi})(\delta_U + \bar{\phi}(\delta_I - \delta_U))}{2C(\bar{\phi},\bar{\phi})} \\ &= \frac{\delta_I - \delta_U}{2} + \frac{(\Delta_I - \Delta_U)(\delta_U + \bar{\phi}(\delta_I - \delta_U))}{2C(\bar{\phi},\bar{\phi})} \end{aligned}$$

and $D''(\phi,\bar{\phi}) = 0$. Similarly,

$$\begin{aligned} p'(\phi,\bar{\phi}) &= t\frac{\delta_I - \delta_U}{2C(\phi,\bar{\phi})} - t\frac{C'(\phi,\bar{\phi})(\delta_U + \phi(\delta_I - \delta_U))}{2(C(\phi,\bar{\phi}))^2} \\ &= t\frac{\delta_I - \delta_U}{2C(\phi,\bar{\phi})} - t\frac{(\Delta_I - \Delta_U)(\delta_U + \phi(\delta_I - \delta_U))}{2C(\phi,\bar{\phi})^2} \\ &= t\frac{\Delta_U\delta_I - \Delta_I\delta_U}{2C(\phi,\bar{\phi})^2} \end{aligned}$$

and

$$p''(\phi,\bar{\phi}) = -t(\Delta_I - \Delta_U)\frac{\Delta_U\delta_I - \Delta_I\delta_U}{C(\phi,\bar{\phi})^3},$$

noting that both $p'(\phi,\bar{\phi}) \geq 0$ and $p''(\phi,\bar{\phi}) \geq 0$ since $\delta_I > \delta_U$ and $\Delta_U > \Delta_I$.

If there is a symmetric equilibrium in which price is less than $v - \frac{t}{2}$ and $\phi \in (0,1)$, then the first order condition is that

$$A'(\phi; a) = \frac{\partial R(\phi, \bar{\phi})}{\partial \phi}|_{\phi=\bar{\phi}}$$

Combining,

$$\begin{aligned}
\frac{\partial R(\phi, \bar{\phi})}{\partial \phi}|_{\phi=\bar{\phi}} &= p'D + pD' \\
&= t[\frac{\delta_I - \delta_U}{2C(\phi, \bar{\phi})} - \frac{(\Delta_I - \Delta_U)(\delta_U + \phi(\delta_I - \delta_U))}{2C(\phi, \bar{\phi})^2}]D(\phi, \bar{\phi}) + \\
&\quad t\frac{D(\bar{\phi}, \bar{\phi})}{C(\bar{\phi}, \bar{\phi})}[\frac{\delta_I - \delta_U}{2} + \frac{(\Delta_I - \Delta_U)(\delta_U + \bar{\phi}(\delta_I - \delta_U))}{2C(\bar{\phi}, \bar{\phi})}] \\
&= \bar{p}(\delta_I - \delta_U)
\end{aligned}$$

A local optimum requires that the second derivative of $\pi(\cdot) - A(\cdot)$ is negative. Since

$$\begin{aligned}
\frac{\partial^2 R(\phi, \bar{\phi})}{\partial \phi^2}|_{\phi=\bar{\phi}} &= p''D + 2p'D' + pD'' \\
&= 2(\delta_I - \delta_U)p' - t(\Delta_I - \Delta_U)\frac{\Delta_U \delta_I - \Delta_I \delta_U}{C(\phi, \bar{\phi})^3}D \\
&= 2(\delta_I - \delta_U)p' - \bar{p}(\Delta_I - \Delta_U)\frac{\Delta_U \delta_I - \Delta_I \delta_U}{C(\phi, \bar{\phi})^2} \\
&= [2(\delta_I - \delta_U)t + \bar{p}(\Delta_U - \Delta_I)]\frac{\Delta_U \delta_I - \Delta_I \delta_U}{C(\phi, \bar{\phi})^2}
\end{aligned}$$

because $D'' = 0$. A further necessary condition is then that $\frac{\partial^2 R(\phi, \bar{\phi})}{\partial \phi^2}|_{\phi=\bar{\phi}} < A''(\bar{\phi})$, so the second order sufficient condition for a local optimum is satisfied. □

B.1.6 Proof of Proposition 9

Again write $A_{\phi a} \equiv \frac{\partial^2 A(\phi, a)}{\partial a \partial \phi}$ and $A_{\phi\phi} \equiv \frac{\partial^2}{\partial^2 \phi}A(\phi; a)$, evaluated at $\phi = \phi^*$. If ϕ^* is an equilibrium, then the marginal benefit of advertising must be less than (resp. greater than) the marginal cost of advertising for a region above (resp. below) ϕ^*.

Consequently, $\frac{\partial \bar{p}(\phi)[\delta_I(\phi)-\delta_U(\phi)]}{\partial \phi}|_{\phi=\phi^*} = R' < A_{\phi\phi}$.

Totally differentiating equation (2.5.2) yields

$$d\phi^* R' = A_{\phi a} da + A_{\phi\phi} d\phi^*.$$

Therefore,

$$\frac{d\phi^*}{da} = \frac{A_{\phi a}(\phi^*)}{R' - A_{\phi\phi}(\phi^*)}$$

The assumption that $A_{\phi a}(\phi^*) > 0$ implies

$$\frac{d\phi^*}{da} < 0.$$

Since $\bar{p}'(\phi) > 0$, $\frac{d\bar{p}(\phi^*)}{da} < 0$ when $a < \bar{a}$. Equilibrium profit is given by

$$\pi(\phi^*) = \bar{p}(\phi^*)D(\phi^*) - A(\phi^*).$$

Therefore,

$$\begin{aligned} \frac{d\pi}{da} &= \frac{d}{d\phi^*}[\bar{p}(\phi^*)D(\phi^*)]\frac{d\phi^*}{da} - A_\phi \frac{d\phi^*}{da} - A_a \\ &= [\bar{p}'(\phi^*)D(\phi^*) + \bar{p}(\phi^*)D'(\phi^*) - A_\phi]\frac{d\phi^*}{da} - A_a, \end{aligned}$$

completing the claim. □

B.1.7 Proof of Proposition 10

Since $\kappa = n - 1$, the surplus from consumers who see at least one ad is

$$(1-(1-q)^n)v - t\frac{2 - q + (1-q)^n(q - 2 + 2qn)}{4nq},$$

and the surplus from consumers who see exactly zero ads is

$$(1-(1-q)^{n-1})v - t\frac{2-q+(1-q)^{n-1}(q-2+2q(n-1))}{4nq}.$$

This follows from Proposition 6. Since the proportion who see zero ads is $(1-\phi)^n$,

$$\begin{aligned} W(\phi) &= (1-(1-q)^n)v - t\frac{2-q+(1-q)^n(q-2+2qn)}{4nq} \\ &\quad -(1-\phi)^n[q(1-q)^{n-1}[v-t\frac{2n-1}{4n}]] - nA(\phi). \end{aligned}$$

Taking the first derivative yields that

$$W'(\phi) = n(1-\phi)^{n-1}[q(1-q)^{n-1}[v-t\frac{2n-1}{4n}]] - nA'(\phi).$$

Now,

$$D_I(\phi) = \frac{1-(1-q)^n}{n}$$

and

$$D_U(\phi) = \frac{1-(1-q)^n}{n} - \frac{(1-\phi)^{n-1}[q(1-q)^{n-1}]}{n}.$$

The first order necessary condition for an equilibrium is that

$$\bar{p}(\phi^*)(1-\phi^*)^{n-1}\frac{q(1-q)^{n-1}}{n} = A'(\phi^*).$$

Since $\bar{p}(\phi^*) \leq v - \frac{t}{2} < v - t\frac{2n-1}{4n} = v - \frac{t}{2} + \frac{1}{4n}$,

$$nA'(\phi^*) < (1-\phi^*)^{n-1}q(1-q)^{n-1}(v - t\frac{2n-1}{4n}).$$

Conclude that

$$\begin{aligned} W'(\phi^*) &= n(1-\phi^*)^{n-1}[q(1-q)^{n-1}[v - t\frac{2n-1}{4n}]] - nA'(\phi) \\ &> (n-1)(1-\phi)^{n-1}[q(1-q)^{n-1}[v - t\frac{2n-1}{4n}]] \\ &> 0 \end{aligned}$$

so if (ϕ^*, p^*) is an equilibrium, $W'(\phi^*) > 0$.

Appendix C

Appendix for Chapter 3

C.1 Details for Section 3.2

Consider $n = 101$, a set of players $I = \{1, ..., n\}$, a set of alternatives $\mathcal{A} = \{A, B\}$, set of types $T_i = \{1, 2\}$ for each $i \in I$ and $T_0 = \{a, b\}$. Set $T := T_0 \times T_1 \times ... \times T_n$. Each player has the same preference over state-alternative pairs given by a function $u : T \times \mathcal{A} \to \mathbb{R}$. Define $u(\cdot)$ by

$$u((t_0, ..., t_{101}), c) \equiv u(t_0, c) = \begin{cases} 1 & t_0 = c \\ 0 & t_0 \neq c \end{cases}$$

for all T. Player i's pure strategies are $S_i = \{A, B\}$; let $S = S_1 \times ... \times S_n$. An aggregation rule $f : S \to \mathcal{A}$ maps the profile of actions to an alternative. Set

$$f(s_0, ..., s_{101}) = \begin{cases} A & \text{if } \sum_{i=1}^{101} \chi_{s_i}(A) \geq 51 \\ B & otherwise \end{cases}$$

for all $(s_0, ..., s_{101}) \in S$, where $\chi_E(\cdot)$ is the indicator function of the set E. Fix a non-empty, closed and convex set of common priors $\Pi \in \Delta T$. Define Π by

$$\Pi = \{\pi \in \Delta T : \pi(\{a\} \times T_1 \times ... \times T_{101}) \in [\underline{p}, \bar{p}] \text{ and}$$

$$\frac{\pi(\{(a, t_1, ..., t_{101})\})}{\pi(\{a\} \times T_1 \times ... \times T_{101})} = \prod_{i=1}^{101} .6^{t_i}.4^{1-t_i} \text{ and}$$

$$\frac{\pi(\{(b, t_1, ..., t_{101})\})}{\pi(\{b\} \times T_1 \times ... \times T_{101})} = \prod_{i=1}^{101} .6^{1-t_i}.4^{t_i}\}$$

The game is defined by the collection $(\mathcal{A}, I, T, u, S, f, \Pi)$.

For each i, let $\hat{S}_i : T_i \to \Delta S_i$ be a strategy for player i and let $\Sigma := \hat{S}_0 \times ... \times \hat{S}_n$ be the set of strategy profiles. This requires that the player's strategy be measurable with respect to her type. As is convention, let σ_i denote player i's strategy and let σ_{-i} represents the vector of strategies chosen by players other than i. A strategy profile $\sigma^* \in \Sigma$ is an *equilibrium* if $\sigma_i^*(t_i)$ is in the set

$$\arg\max_{\sigma \in \Delta S_i} \min_{\pi \in \Pi} \mathbb{E}_\pi[\mathbb{E}_{\sigma^*_{-i}}[u((t_0, ..., t_n), f((s_0(t_0), ..., s_{i-1}(t_{i-1}), \sigma, s_{i+1}(t_{i+1}), ...)]|t_i]$$

for every $i \in I$.

C.2 Proof of Theorem 6

Proof. Define the set $\Lambda = \{l \in \mathbb{R}^{A \times \Omega} : \sum_{a \in A} l(a, \omega) = n\}$, noting that Λ is compact, and consider the correspondence $C_t : \Lambda \twoheadrightarrow \Delta A$ defined by

$$C_t(\lambda) = \arg\max_{\hat{\sigma} \in \Delta A} \min_{q \in \pi_t} \int_\omega \int_{Z(A)} \sum_{a \in A} \hat{\sigma}(a) U(x, t, a, \omega) dp(x|\lambda(\omega)) dq.$$

Define $C(\lambda) = \times_{t \in T} C_t(\lambda)$ and let $F : \Lambda \twoheadrightarrow \Lambda$ be defined by

$$F(\lambda) = \{n \sum_{t \in T} c_t(a) r(t|\omega) : c \in C(\lambda)\}.$$

If $\lambda \in F(\lambda)$ then (3.3.3) is satisfied, since then $\sigma^*(t) \in \arg\max_{\hat{\sigma}\in\Delta(A)} V_t(\hat{\sigma}, \sigma^*)$ for some σ^* that generates λ. Hence, existence of an equilibrium is equivalent to showing $F(\cdot)$ that has a fixed point. It remains to be shown that F is convex and closed. Since $F(\lambda)$ is an affine transformation of $C(\lambda)$, need to show that $C(\cdot)$ is convex and closed. Show first that all C_t are convex, compact and upper hemi continuous.

[Convex:] Define $\phi : C(\Omega) \to \mathbb{R}$ by $\phi(f) = \min_{q\in\pi_t} \int f dq$, where $C(\Omega)$ is the set of continuous functions from Ω to the real numbers). Then ϕ and $p(x|\lambda(\omega))U(x, t, \cdot, \omega)$ are both concave. So $g : \Delta A \to \mathbb{R}$ defined by

$$g(\hat{\sigma}) = \phi(p(x|\lambda(\omega)) \sum_{a\in A} \hat{\sigma}(a)U(x, t, a, \omega))$$

is also concave. Hence $g(x) = g(y) \implies g(\alpha x + (1-\alpha)y) \geq g(x)\forall\alpha \in [0, 1]$ and $x, y \in C_t(\lambda) \implies \alpha x + (1-\alpha)y \in C_t(\lambda)$. Therefore $C_t(\lambda)$ is convex, from which it follows that $C(\cdot)$ is convex since a product of convex sets is convex. Since $C(\cdot)$ is convex, $F(\cdot)$ is convex.

[Closed:] ϕ is continuous by the Maximum Theorem (Theorem 17.31 of Aliprantis and Border [2006]; henceforth, AB). $p(x|\cdot)$ is continuous since it is a product of continuous functions. $U(x, t, \cdot, \omega)$ is continuous by assumption. So

$$\min_{q\in\pi_t} \int_\omega p(x|\lambda(\omega))[\sum_{a\in A} \hat{\sigma}(a)U(x, t, a, \omega)]dq$$

is continuous. Hence $C_t(\lambda)$ is upper hemi continuous and compact by the Maximum Theorem as the set of solutions to a maximization problem.

$C(\lambda)$ is compact for all λ by the Tychonoff product theorem (AB Theorem 2.61) because $C(\lambda)$ is a product of compact sets. By AB Theorem 17.20, it suffices to show that if $\lambda_n \to \lambda$, $x_n \in C(\lambda_n)$, and $x_n \to x$ then $x \in C(\lambda)$. Given such sequences, let $x_{n,t}$ be the t-th component of x_n and x_t the t-th component of x for any $t \in T$. By definition of the product topology, $x_n \to x \iff x_{n,t} \to x_t$ for all $t \in T$. By definition of $C(\cdot)$, $x_{n,t} \in C_t(\lambda_n)$ for each n. Because $C_t(\cdot)$ is upper hemi continuous and compact, $x_t \in C_t(\lambda)$. Since t is arbitrary, $x_t \in C_t(\lambda)$ for all $t \in T$ and by definition of $C(\cdot)$, $x \in C(\lambda)$. Hence $C(\cdot)$ is upper hemi continuous and compact. AB Theorem 17.10 establishes that $C(\cdot)$ is closed and thus $F(\cdot)$ is closed.

Since Λ is compact and $F(\cdot)$ is closed and convex, applying Kakutani's Fixed point theorem (AB Corollary 17.55) yields a λ^* such that $\lambda^* \in F(\lambda^*)$. □

C.3 Preliminaries for the Remaining Proofs

Lemma 28 relies on two functions of the strategy profile.

Formally, if other votes unfold so that the realized action profile is in the event

$$Piv_A = \{x \in Z(C) : x(A) = x(B) \text{ or } x(A) = x(B) - 1\}, \tag{C.3.1}$$

then the voter is pivotal for candidate A; let Piv_B be the corresponding event for B. Each voter's best response depends on the relationship between her set of posteriors and the function $b : \Omega \times (\Delta C)^T \to [0, 1]$ given by $b(b, \sigma) =$

$$\frac{Pr(Piv_B|b,\sigma) + Pr(Piv_A|b,\sigma)}{Pr(Piv_B|b,\sigma) + Pr(Piv_B|a,\sigma) + Pr(Piv_A|b,\sigma) + Pr(Piv_A|a,\sigma)}. \tag{C.3.2}$$

and $b(a, \sigma) = 1 - b(b, \sigma)$. The probabilities in this function depend only on the strategy profile and not on an individual voter's type.

Another key equation is the *insurance strategy*, denoted $\hat{s}(\cdot, \sigma)$, is given by

$$\hat{s}(A,\sigma) = \frac{2(\mathbb{E}[U|b,\sigma] - \mathbb{E}[U|a,\sigma]) + Pr(Piv_B|b,\sigma) + Pr(Piv_B|a,\sigma)}{Pr(Piv_B|b,\sigma) + Pr(Piv_B|a,\sigma) + Pr(Piv_A|b,\sigma) + Pr(Piv_A|a,\sigma)}.$$

and $\hat{s}(B, \sigma) = 1 - \hat{s}(A, \sigma)$. This maps a strategy profile σ into the strategy a voter would play to ensure his expected utility is independent of the state if $\hat{s}(A, \sigma) \in [0, 1]$. Otherwise, no strategy equalizes a voters expected utilities between states.

Notice that expected utility in state ω if the voter abstained is given by

$$\mathbb{E}[U|a,\sigma] = \sum_{n=0}^{\infty} \frac{e^{-\lambda(a)(A)}\lambda(a)(A)^n}{n!}[\sum_{j=0}^{n-1} \frac{e^{-\lambda(a)(B)}\lambda(a)(B)^j}{j!} + \frac{1}{2}\frac{e^{-\lambda(a)(A)}\lambda(a)(A)^n}{n!}]$$

where $\lambda(\omega)(c) = \mathbb{E}[x(c)|\omega, \sigma]$ as in equation (3.3.3). Define $\mathbb{E}[U|b, \sigma]$ analogously.

This expression is precisely the probability that candidate ω wins in state ω. The expected utility of voting for candidate c in state ω when others play strategy profile σ is

$$\mathbb{E}[U|\omega, v_c, \sigma] = \mathbb{E}[U|\omega, \sigma] + [\chi_{\{\omega\}}(c) - \frac{1}{2}]Pr(Piv_c|\omega).$$

Additionally, let $\tau : C \times \Omega \times \Delta C^T \rightarrow [0, 1]$ be the expected vote share for a candidate in a state given a strategy profile. Formally,

$$\tau(c|\omega, \sigma) = \sum_{t \in T} r(t|\omega)\sigma(t)(c).$$

Note that this does not depend on the number of voters.

C.4 Proofs from Section 3.4

Lemma 28 establishes the form of a voter's best response correspondence. This will be used to prove both Theorem 7 and Theorem 8.

Lemma 28. *For any σ^*, σ^* is an equilibrium if $\sigma^*_t(A) \in BR_t(\sigma^*)(A)$ where*

$$BR_t(\sigma)(A) = \begin{cases} \{0\} & \text{if } \mathbb{E}[U|a, v_B, \sigma] \geq \mathbb{E}[U|b, v_B, \sigma] \;\&\; b(b, \sigma) > p_t \\ & \quad \text{or } \mathbb{E}[U|b, v_A, \sigma] \geq \mathbb{E}[U|a, v_A, \sigma] \;\&\; b(b, \sigma) > q_t \\ [0, 1] & \text{if } \mathbb{E}[U|a, v_B, \sigma] \geq \mathbb{E}[U|b, v_B, \sigma] \;\&\; b(b, \sigma) = p_t \\ & \quad \text{or } \mathbb{E}[U|b, v_A, \sigma] \geq \mathbb{E}[U|a, v_A, \sigma] \;\&\; b(b, \sigma) = q_t \\ \{1\} & \text{if } \mathbb{E}[U|a, v_B, \sigma] \geq \mathbb{E}[U|b, v_B, \sigma] \;\&\; b(b, \sigma) < p_t \\ & \quad \text{or } \mathbb{E}[U|b, v_A, \sigma] \geq \mathbb{E}[U|a, v_A, \sigma] \;\&\; b(b, \sigma) < q_t \\ \hat{BR}_t(\sigma)(A) & \text{otherwise} \end{cases}$$

and

$$\hat{BR}_t(\sigma)(A) = \begin{cases} \{0\} & \textit{if } b(b,\sigma) > q_t \\ [0, \hat{s}(A,\sigma)] & \textit{if } b(b,\sigma) = q_t \\ \{\hat{s}(A,\sigma)\} & \textit{if } q_t > b(b,\sigma) > p_t \\ [\hat{s}(A,\sigma), 1] & \textit{if } b(b,\sigma) = p_t \\ \{1\} & \textit{if } b(b,\sigma) < p_t \end{cases}$$

where $p_t = \min_{\rho\in\Pi_t} \rho(a)$ *and* $q_t = \max_{\rho\in\Pi_t} \rho(a)$. *If* $BR_t(\sigma) = \hat{BR}_t(\sigma)$ *then* $\hat{s}(A,\sigma) \in [0,1]$.

Proof. Throughout, a strategy is indexed solely by the probability of playing A. This is WLOG since ΔC is one dimensional. Let p_t and q_t be as in the statement of the Lemma.

A player of type t has a best response to σ of playing A with probability s if s maximizes

$$V_t(s,\sigma) = \min_{\rho\in\Pi_t} \mathbb{E}_\rho[\int [sU(t,A,\omega,x) + (1-s)U(t,B,\omega,x)]p(dx|\lambda(\omega))].$$

This function is not in general differentiable everywhere. Since $V_t(\cdot,\sigma)$ is concave as a minimum of a set of linear functions, the super-differential exists everywhere. By definition and adapted to this setting, the super-differential is given by

$$\partial V_t(s,\sigma) = \{x \in \mathbb{R}^\Omega : V_t(y,\sigma) \leq V_t(s,\sigma) + \sum_\omega [(y(\omega) - s(\omega))x(\omega)] \forall y \in \Delta A\}.$$

The best response correspondence is the set of all s s.t. $0 \in \partial V_t(s,\sigma)$ where $\partial V_t(s,\sigma)$ is the super-differential of $V_t(\cdot,\sigma)$ at s. This follows from the dual to Aliprantis and Border [2006, Lem 7.10], which states that s is a maximum of $V_t(\cdot,\sigma)$ if and only if $0 \in \partial V_t(s,\sigma)$.

Consider the case where $\mathbb{E}[U|a, v_B, \sigma] \geq \mathbb{E}[U|b, v_B, \sigma]$. Note that $V_t(s,\sigma)$ equals

$$
\begin{aligned}
& \min_{p\in\Pi_t(A)}\{p[s\frac{1}{2}Pr(Piv_A|a)-(1-s)\frac{1}{2}Pr(Piv_B|a)+\mathbb{E}[U|a,\sigma]]+\\
& \qquad +(1-p)[(1-s)\frac{1}{2}Pr(Piv_B|b)-s\frac{1}{2}Pr(Piv_A|b)+\mathbb{E}[U|b,\sigma]]\}\\
= & \; p_t[s\frac{1}{2}Pr(Piv_A|a)-(1-s)\frac{1}{2}Pr(Piv_B|a)+\mathbb{E}[U|a,\sigma]]+\\
& \qquad +(1-p_t)[(1-s)\frac{1}{2}Pr(Piv_B|b)-s\frac{1}{2}Pr(Piv_A|b)+\mathbb{E}[U|b,\sigma]]
\end{aligned}
$$

because for every s

$$
\begin{aligned}
& sPr(Piv_A|a)-(1-s)Pr(Piv_B|a)+2\mathbb{E}[U|a,\sigma]\geq\\
& \quad (1-s)Pr(Piv_B|b)-sPr(Piv_A|b)+2\mathbb{E}[U|b,\sigma]
\end{aligned}
$$

This occurs because the RHS reaches its minimum at $s=0$ and the LHS reaches its maximum at $s=0$. At $s=0$ the RHS equals $\mathbb{E}[U|a,v_B,\sigma]$ and the LHS equals $\mathbb{E}[U|b,v_B,\sigma]$. By hypothesis, $\mathbb{E}[U|a,v_B,\sigma]\geq\mathbb{E}[U|b,v_B,\sigma]$ so for every s the RHS is larger than the LHS. Thus, $V_t(s,\sigma)$ is differentiable in $s\in(0,1)$. By Aliprantis and Border [2006, Cor 7.17] $\partial V_t(s,\sigma)$ is singleton and coincides with the Gateaux derivative when it exists. Hence,

$$
\begin{aligned}
\partial V_t(s,\sigma) \;&=\; \{p_t[\frac{1}{2}Pr(Piv_A|a)+\frac{1}{2}Pr(Piv_B|a)]\\
&\qquad -(1-p_t)[\frac{1}{2}Pr(Piv_A|b)+\frac{1}{2}Pr(Piv_B|b)]\}
\end{aligned}
$$

and $0\in\partial V_t(s,\sigma)$ only if $b(b,\sigma)=p_t$. If $b(b,\sigma)<p_t$ this is positive and if $b(b,\sigma)<p_t$ this is negative and hence no $s\in(0,1)$ is a maxima.

If $s=1$ then the derivative is not defined since $V_t(1+\epsilon,\sigma)$ for any $\epsilon>0$ does not exist. The super-differential does exist:

$$\partial V_t(1,\sigma)=\{x\in\mathbb{R}:V_t(y,\sigma)-V_t(1,\sigma)\leq(y-1)x\forall y\in[0,1]\}.$$

Since

$$\begin{aligned} V_t(y,\sigma) - V_t(1,\sigma) &= (y-1)\frac{1}{2}(p_t[Pr(Piv_A|a) + Pr(Piv_B|a)] \\ &- (1-p_t)[Pr(Piv_A|b) + Pr(Piv_B|b)]) \end{aligned}$$

$0 \in \partial V_t(1,\sigma)$ if and only if

$$p_t[Pr(Piv_A|a) + Pr(Piv_B|a)] - (1-p_t)[Pr(Piv_A|b) + Pr(Piv_B|b) > 0.$$

As noted above, $b(b,\sigma) < p_t$ implies this is positive.

Additionally, if $s = 0$ the derivative is not defined since $V_t(0-\epsilon,\sigma)$ for any $\epsilon > 0$ does not exist. The super-differential does exist:

$$\partial V_t(0,\sigma) = \{x \in \mathbb{R} : V_t(y,\sigma) - V_t(0,\sigma) \leq y \cdot x \forall y \in [0,1]\}.$$

Since $V_t(y,\sigma) - V_t(0,\sigma) =$

$$y\frac{1}{2}(p_t[Pr(Piv_A|a) + Pr(Piv_B|a)] - (1-p_t)[Pr(Piv_A|n) + Pr(Piv_B|n)])$$

$0 \in \partial V_t(0,\sigma)$ if and only if

$$p_t[Pr(Piv_A|a) + Pr(Piv_B|a)] - (1-p_t)[Pr(Piv_A|b) + Pr(Piv_B|b)] < 0.$$

As noted above, $b(B,\sigma) > p_t$ implies this is negative.

The above observations show that if $\mathbb{E}[U|a, v_B, \sigma] \geq \mathbb{E}[U|b, v_B, \sigma]$, then the set of maximizers of $V_t(\cdot,\sigma)$ is

$$\arg\max_{s\in[0,1]} V_t(s,\sigma) = \begin{cases} \{1\} & \text{if } b(b,\sigma) > p_t \\ [0,1] & \text{if } b(b,\sigma) = p_t. \\ \{0\} & \text{if } b(b,\sigma) < p_t \end{cases}$$

If $\mathbb{E}[U|a, v_B, \sigma] \leq \mathbb{E}[U|b, v_B, \sigma]$, similar arguments show the same form of BR correspondence with the probability assigned to a equal to q_t instead of p_t.

Now, suppose that neither of the above inequalities hold. Then there exists an $\bar{s} \in (0,1)$ so that the conditional expected utilities in A and B are equal. Further, if $s > \bar{s}$ the conditional expected utility in state a is larger than that in state b and if $s < \bar{s}$ then the expected utility in state B is larger than that in state A. Algebra shows that

$$\bar{s} = \frac{2(\mathbb{E}[U|b,\sigma] - \mathbb{E}[U|a,\sigma]) + Pr(Piv_B|b) + Pr(Piv_B|a)}{Pr(Piv_B|a) + Pr(Piv_A|a) + Pr(Piv_B|b) + Pr(Piv_A|b)}$$

which is $\hat{s}(A,\sigma)$.

Since for all $s \in (0,\bar{s})$ and every $s \in (\bar{s},1)$ the minimizer is unique, the Gateaux derivative exists whenever $s \notin \{0,\bar{s},1\}$. If $s \in (\bar{s},1)$ then

$$\begin{aligned} \partial V_t(s,\sigma) &= \{\frac{\partial}{\partial s}V_t(s,\sigma)\} \\ &= \{p_t\frac{1}{2}[Pr(Piv_A|a) + Pr(Piv_B|a)] \\ &\qquad -(1-p_t)\frac{1}{2}[Pr(Piv_A|b) + Pr(Piv_B|b)]\}. \end{aligned}$$

If $s' \in (0,\bar{s})$ then

$$\begin{aligned} \partial V_t(s',\sigma) &= \{\frac{\partial}{\partial s}V_t(s,\sigma)\} \\ &= \{q_t\frac{1}{2}[Pr(Piv_A|a) + Pr(Piv_B|a)] \\ &\qquad -(1-q_t)\frac{1}{2}[Pr(Piv_A|b) + Pr(Piv_B|b)]\}. \end{aligned}$$

Thus any $s \in (\bar{s},1)$ is an optimum only if

$$p_t[Pr(Piv_A|a) + Pr(Piv_B|a)] - (1-p_t)[Pr(Piv_A|b) + Pr(Piv_B|b)] = 0,$$

which happens when $p_t = b(b,\sigma)$. Similarly, any $s \in (0,\bar{s})$ is an optimum when $q_t = b(b,\sigma)$. Otherwise there cannot be an optimum in $(0,1)\backslash\{\bar{s}\}$.

As above, when $s = 1$ then the derivative is not defined since $V_t(1+\epsilon,\sigma)$ for any $\epsilon > 0$ does not exist. The super-differential does exist:

$$\partial V_t(1,\sigma) = \{x \in \mathbb{R} : V_t(y,\sigma) - V_t(1,\sigma) \leq (y-1)x \forall y \leq 1\}.$$

Since $V_t(y,\sigma) - V_t(1,\sigma)$ is equal to

$$(y-1)(p_t\frac{1}{2}[Pr(Piv_A|a) + Pr(Piv_A|b)] - (1-p_t)\frac{1}{2}[Pr(Piv_B|a) + Pr(Piv_B|b)]),$$

$0 \in \partial V_t(1,\sigma)$ if and only if

$$V_t(y,\sigma) - V_t(1,\sigma) \leq 0 \iff b(b,\sigma) \leq p_t.$$

Hence $s = 1$ is optimal only if $b(b,\sigma) \geq p_t$. Similar arguments show then $0 \in \partial V_t(0,\sigma) \iff b(b,\sigma) \geq q_t$.

By the above, we have covered the cases where $b(b,\sigma) \geq q_t$ and $b(b,\sigma) \leq p_t$. Suppose $p_t < b(b,\sigma) < q_t$. In this case,

$$q_t[Pr(Piv_A|a) + Pr(Piv_B|a)] > (1-q_t)[Pr(Piv_A|b) + Pr(Piv_B|b)]$$

and

$$p_t[Pr(Piv_A|a) + Pr(Piv_B|a)] < (1-p_t)[Pr(Piv_A|b) + Pr(Piv_B|b)].$$

So for $s > \bar{s}$,

$$\begin{aligned}\partial V_t(s,\sigma) &= \{p_t\frac{1}{2}[Pr(Piv_A|a) + Pr(Piv_B|a)] \\ &\qquad -(1-p_t)\frac{1}{2}[Pr(Piv_A|b) + Pr(Piv_B|b)]\}\end{aligned}$$

is a singleton strictly smaller than zero. For $s' < \bar{s}$,

$$\begin{aligned}\partial V_t(s',\sigma) &= \{q_t\frac{1}{2}[Pr(Piv_A|a) + Pr(Piv_B|a)] \\ &\qquad -(1-q_t)\frac{1}{2}[Pr(Piv_A|b) + Pr(Piv_B|b)]\}\end{aligned}$$

is a singleton strictly larger than zero. However, for $s = \bar{s}$

$$\begin{aligned}\partial V_t(\bar{s},\sigma) &= \{p(a)\frac{1}{2}[Pr(Piv_A|a) + Pr(Piv_B|a)] \\ &\qquad -p(b)\frac{1}{2}[Pr(Piv_A|b) + Pr(Piv_B|b)] : p \in \Pi_t\}\end{aligned}$$

Since $q_t > \rho(A) > p_t$, $\exists \rho \in \Pi_t$ s.t.

$$\frac{\rho(a)}{1-\rho(a)} = \frac{Pr(Piv_A|b) + Pr(Piv_B|b)}{Pr(Piv_A|a) + Pr(Piv_B|a)}$$

implying that $0 \in \partial V_t(\bar{s}, \sigma)$ and $\bar{s}$ is the only maximizer when $q_t > b(B, \sigma) > p_t$.

Combining the above results yields the desired form of the best response function.

□

In order to prove Theorem 7, two more preliminary results are necessary. Lemma 29 and Lemma 30 allow characterization of the worst case scenario. The proof of Theorem 7 will use both these facts to show that no equilibrium exists where a voter thinks the worst case scenario is independent of her vote.

Lemma 29. *For any* $n \geq 1$*,*

$$\mathbb{E}[U|a, \sigma_n] \geq \mathbb{E}[U|B, \sigma_n] \iff \tau(A|a, \sigma_n) \geq \tau(A|b, \sigma_n).$$

Proof. Let $f(x, \lambda) = \frac{e^{-\lambda}\lambda^x}{x!}$, the probability mass function of the Poisson distribution with mean λ, and $F(x, \lambda)$ its CDF. The CDF of the Poisson distribution has the form $\frac{\Gamma([x+1], \lambda)}{[x]!}$ where $[z]$ is the greatest integer less than or equal to z and $\Gamma(z, y)$ is the generalized incomplete gamma function:

$$\Gamma(z, y) = \int_y^\infty e^{-t} t^{z-1} dt.$$

We can write

$$\mathbb{E}[U|a, \sigma_n] = Q(\tau(A|a, \sigma_n)n) + \frac{1}{2}\sum_{j=0}^{\infty} f(j, \tau(A|a, \sigma_n)n) f(j, \tau(B|a, \sigma_n)n)$$

where $Q(\cdot)$ is given by

$$Q(\lambda) = \sum_{j=0}^{\infty} f(j, \lambda) F(j-1, n-\lambda).$$

Observe that

$$\frac{\partial Q}{\partial \lambda} = \sum_{x=1}^{\infty}[\frac{\partial f(j,\lambda)}{\partial \lambda}F(j-1,n-\lambda) + f(j,\lambda)\frac{\partial F(j-1,n-\lambda)}{\partial \lambda}].$$

By the fundamental theorem of calculus,

$$\frac{\partial F(x,\lambda)}{\partial \lambda} = -\frac{e^{-\lambda}\lambda^x}{x!}$$

and

$$\frac{\partial f(x,\lambda)}{\partial \lambda} = \frac{e^{-\lambda}\lambda^{x-1}(x-\lambda)}{x!}$$

whenever x is an integer. Given this, the above sum can be written as

$$\frac{\partial Q}{\partial \lambda} = e^{-n}[\sum_{x=1}^{\infty}\frac{\lambda^x}{x!}\frac{(n-\lambda)^{x-1}}{x-1!} + \sum_{x=0}^{\infty}\frac{\lambda^x}{x!}\frac{(n-\lambda)^x}{x!}].$$

Now, we must deal with the second term.

$$\begin{aligned} \frac{\partial}{\partial\lambda}\sum_{j=0}^{\infty} f(j,\lambda)f(j,n-\lambda) &= \sum_{x=0}^{\infty}\frac{\partial}{\partial\lambda}e^{-n}\frac{\lambda^x(n-\lambda)^x}{x!x!} \\ &= \sum_{x=1}^{\infty} e^{-n}[\frac{x\lambda^{x-1}(n-\lambda)^x}{x!x!} - \frac{x\lambda^x(n-\lambda)^{x-1}}{x!x!}] \\ &= \sum_{x=1}^{\infty} e^{-n}[\frac{\lambda^{x-1}(n-\lambda)^x}{x!(x-1)!} - \frac{\lambda^x(n-\lambda)^{x-1}}{x!(x-1)!}] \end{aligned}$$

Adding together shows that $\frac{\partial}{\partial\lambda}\mathbb{E}[U|a,\sigma_n,n]$ is equal to

$$e^{-n}[\sum_{x=0}^{\infty}\frac{\lambda^x}{x!}\frac{(n-\lambda)^x}{x!} + \frac{1}{2}\sum_{x=1}^{\infty}(\frac{\lambda^x}{x!}\frac{(n-\lambda)^{x-1}}{x-1!} + \frac{\lambda^{x-1}(n-\lambda)^x}{x!(x-1)!})].$$

Clearly, this term is positive. Recall that $\lambda = \tau(A|a,\sigma_n)n$ so that

$$\frac{\partial \mathbb{E}[U|a,\sigma_n]}{\partial \tau(A|a,\sigma_n)} = \frac{\partial \mathbb{E}[U|a,\sigma_n]}{\partial \lambda}\frac{\partial \lambda}{\partial \tau(A|A,\sigma_n)} = n\frac{\partial \mathbb{E}[U|a,\sigma_n]}{\partial \lambda}.$$

Since $\frac{\partial \mathbb{E}[U|a,\sigma_n]}{\partial \lambda} \geq 0$, so is $\frac{\partial \mathbb{E}[U|a,\sigma_n]}{\partial \tau(A|a,\sigma_n)}$.

Since $\frac{\partial \mathbb{E}[U|a,\sigma_n]}{\partial \tau(A|a,\sigma_n)} \geq 0$, as $\tau(A|a,\sigma_n)$ increases, $\mathbb{E}[U|a,\sigma_n]$ increases. Similarly for $\tau(A|b,\sigma_n)$ and $\mathbb{E}[U|b,\sigma_n]$. Since the expected number of voters in each state is equal, the terms $\mathbb{E}[U|a,\sigma_n]$ and $\mathbb{E}[U|b,\sigma_n]$ are equal whenever $\tau(A|a,\sigma_n)$ and $\tau(B|b,\sigma_n)$ are equal. This establishes the claim. □

Lemma 30. *If $\frac{1}{2} < \tau(B|b,\sigma_n) < \tau(A|a,\sigma_n)$, then $\hat{s}(A,\sigma_n) < \frac{1}{2}$. In particular, when* *the expected winner in each state is correct,* $\hat{s}(A, \sigma_n) < \frac{1}{2} \iff b(b, \sigma_n) > \frac{1}{2}$.

Proof. Suppose $\frac{1}{2} < \tau(B|b, \sigma_n) < \tau(A|a, \sigma_n)$. By Lemma 29,

$$\mathbb{E}[U|b, \sigma_n] < \mathbb{E}[U|a, \sigma_n].$$

Consider the numerator of $\hat{s}(A, \sigma_n)$. Recall that it is

$$\begin{aligned}\phi(\sigma_n) &= 2(\mathbb{E}[u|b, \sigma_n] - \mathbb{E}[u|a, \sigma_n]) \\ &\quad + Pr(Piv_B|b, \sigma_n) + Pr(Piv_B|a, \sigma_n).\end{aligned}$$

The fraction is less than $\frac{1}{2}$ if and only if

$$\begin{aligned}2\phi(\sigma_n) &< Pr(Piv_B|b, \sigma_n) + Pr(Piv_B|a, \sigma_n) \\ &\quad + Pr(Piv_A|b, \sigma_n) + Pr(Piv_A|a, \sigma_n).\end{aligned}$$

Equivalently, this holds if and only if

$$\begin{aligned}0 &> 4(\mathbb{E}[u|b, \sigma_n] - \mathbb{E}[u|a, \sigma_n]) + Pr(Piv_B|b, \sigma_n) \\ &\quad + Pr(Piv_B|a, \sigma_n) - Pr(Piv_A|b, \sigma_n) - Pr(Piv_A|a, \sigma_n)\end{aligned}$$

We can rewrite

$$\gamma = Pr(Piv_B|b, \sigma_n) + Pr(Piv_B|a, \sigma_n) - Pr(Piv_A|b, \sigma_n) - Pr(Piv_A|a, \sigma_n)$$

as a function only of $\tau(A|a, \sigma_n)$ and $\tau(B|b, \sigma_n)$. Set $t = \tau(B|b, \sigma_n)$ and $s = \tau(A|a, \sigma_n)$ for convenience. Expanding and writing in terms of t and s,

$$\begin{aligned}
\gamma &= e^{-n}\sum_{j=0}^{\infty} n^{2j}[\frac{t^j(1-t)^j}{j!j!} + n\frac{t^j(1-t)^{j+1}}{j!j+1!} + \frac{s^j(1-s)^j}{j!j!} + n\frac{s^{j+1}(1-s)^j}{j!j+1!}] \\
&\quad -e^{-n}\sum_{j=0}^{\infty} n^{2j}[\frac{t^j(1-t)^j}{j!j!} + n\frac{t^{j+1}(1-t)^j}{j!j+1!} + \\
&\quad \frac{s^j(1-s)^j}{j!j!} + n\frac{s^j(1-s)^{j+1}}{j!j+1!}] \\
&= \sum_{j=0}^{\infty} e^{-n}n^{2j+1}[\frac{t^j(1-t)^{j+1}}{j!j+1!} + \frac{s^{j+1}(1-s)^j}{j!j+1!} \\
&\quad -\frac{t^{j+1}(1-t)^j}{j!j+1!} - \frac{s^j(1-s)^{j+1}}{j!j+1!}].
\end{aligned}$$

Recall that

$$\begin{aligned}
\mathbb{E}[U|b,\sigma_n] &= \sum_{j=0}^{\infty} f(j;tn)F(j-1;(1-t)n) + \frac{1}{2}\sum_{j=0}^{\infty} f(j;tn)f(j;(1-t)n) \\
&:= \hat{\psi}_n(\tau(B|b,\sigma_n))
\end{aligned}$$

and similarly $\mathbb{E}[U|a,\sigma_n] = \hat{\psi}_n(\tau(A|a,\sigma_n))$. Setting

$$\theta_n(t) = \sum_{j=0}^{\infty} e^{-n}n^{2j+1}[\frac{t^j(1-t)^{j+1}}{j!j+1!} - \frac{t^{j+1}(1-t)^j}{j!j+1!}]$$

and

$$\psi_n(x) = 4\hat{\psi}_n(x) + \theta_n(x).$$

gives that

$$\hat{s}(A,\sigma_n) < \frac{1}{2} \iff \psi_n(t) - \psi_n(s) < 0.$$

From Lemma 29 and writing $\lambda = nt$, we have that

$$\frac{\partial\hat{\psi}_n}{\partial\lambda} = e^{-n}[\sum_{x=0}^{\infty} \frac{\lambda^x}{x!}\frac{(n-\lambda)^x}{x!} + \frac{1}{2}\sum_{x=1}^{\infty}(\frac{\lambda^x}{x!}\frac{(n-\lambda)^{x-1}}{x-1!} + \frac{\lambda^{x-1}(n-\lambda)^x}{x!(x-1)!})].$$

which is positive. Now,

$$\theta_n(\lambda) = \sum_{j=0}^{\infty} e^{-n} \frac{\lambda^j (n-\lambda)^{j+1} - \lambda^{j+1}(n-\lambda)^j}{j!(j+1)!}$$

so that

$$\begin{aligned}
\frac{\partial \theta_n}{\partial \lambda} &= \sum_{j=0}^{\infty} e^{-n} \frac{\partial}{\partial \lambda} \frac{\lambda^j (n-\lambda)^{j+1} - \lambda^{j+1}(n-\lambda)^j}{j!(j+1)!} \\
&= \sum_{j=0}^{\infty} e^{-n} \frac{\lambda^{j-1}(n-\lambda)^{j+1}}{(j-1)!(j+1)!} - \frac{\lambda^j (n-\lambda)^j}{j!j!} - \frac{\lambda^j (n-\lambda)^j}{j!j!} + \frac{\lambda^{j+1}(n-\lambda)^{j-1}}{(j-1)!(j+1)!} \\
&= \sum_{j=1}^{\infty} e^{-n} \frac{\lambda^{j-1}(n-\lambda)^{j+1} + \lambda^{j+1}(n-\lambda)^{j-1}}{(j-1)!(j+1)!} - 2\sum_{j=0}^{\infty} e^{-n} \frac{\lambda^j (n-\lambda)^j}{j!j!}
\end{aligned}$$

Combining,

$$\begin{aligned}
\frac{\partial \psi_n}{\partial t} &= [4\frac{\partial \hat{\psi}_n}{\partial \lambda} + \frac{\partial \theta_n}{\partial \lambda}]\frac{\partial \lambda}{\partial t} \\
&= n[2\sum_{x=0}^{\infty} e^{-n} \frac{\lambda^x (n-\lambda)^x}{x!^2} + 3\sum_{j=1}^{\infty} e^{-n} \frac{\lambda^{j-1}(n-\lambda)^{j+1} + \lambda^{j+1}(n-\lambda)^{j-1}}{(j-1)!(j+1)!}]
\end{aligned}$$

which is clearly greater than 0.

To show that $\psi_n(\tau(B|b,\sigma_n)) - \psi_n(\tau(A|a,\sigma_n)) < 0$, recall that we can write this as $\int_s^t \frac{\partial \psi_n(x)}{\partial x} dx$ which is negative because the integrand is positive but $\tau(A|a,\sigma_n) > \tau(B|b,\sigma_n)$. Therefore, whenever $\tau(A|a,\sigma_n) > \tau(B|b,\sigma_n)$ it must be that $\hat{s}(A,\sigma_n) < \frac{1}{2}$.

To complete the second part of the Lemma, note the following.

Claim 2. $b(b,\sigma_n) > \frac{1}{2} \iff |\tau(A|a,\sigma_n) - \frac{1}{2}| > |\tau(B|b,\sigma_n) - \frac{1}{2}|$.

Proof. Note

$$b(b,\sigma_n) > \frac{1}{2} \iff Pr(Piv_A|b) + Pr(Piv_B|b) > Pr(Piv_A|a) + Pr(Piv_B|a).$$

Let $t = \tau(A|a,\sigma_n)$ so that $Pr(Piv_A|a) + Pr(Piv_B|a)$ equals

$$2\sum_{j=0}^{\infty} p(2j)\left(\begin{array}{c}2j\\j\end{array}\right)t^j(1-t)^j$$
$$+\sum_{j=0}^{\infty} p(2j+1)\left(\begin{array}{c}2j+1\\j+1\end{array}\right)[t^j(1-t)^{j+1}+t^{j+1}(1-t)^j]$$

where $p(x)=\frac{e^{-n}n^x}{x!}$. Take the derivative with respect to t to get

$$(1-2t)[2\sum_{j=0}^{\infty} j\{p(2j)\left(\begin{array}{c}2j\\j\end{array}\right)t^{j-1}(1-t)^{j-1}$$
$$+\sum_{j=0}^{\infty} p(2j+1)\left(\begin{array}{c}2j+1\\j+1\end{array}\right)t^{j-1}(1-t)^{j-1}\}]$$

which is positive whenever $t<.5$ and negative whenever $t>.5$. Similarly for $Pr(Piv_A|b)+Pr(Piv_B|b)$. Given the symmetry of $Pr(Piv_A|a)+Pr(Piv_B|b)$ with respect to $\tau(A|a,\sigma_n)$ and $Pr(Piv_A|b)+Pr(Piv_B|b)$ with respect to $\tau(B|b,\sigma_n)$, the claim follows immediately. □

From Claim 2, whenever $b(b,\sigma_n)>\frac{1}{2}$, $|\tau(A|a,\sigma_n)-\frac{1}{2}|>|\tau(B|a,\sigma_n)-\frac{1}{2}|$. Further, if the expected winners are correct, it must be that both $\tau(A|a,\sigma_n)>\frac{1}{2}$ and $\tau(B|b,\sigma_n)>\frac{1}{2}$. It follows that $\tau(A|a,\sigma_n)>\tau(B|b,\sigma_n)$, so $\hat{s}(A,\sigma_n)<\frac{1}{2}$. Similarly, suppose that $\hat{s}(A,\sigma_n)<\frac{1}{2}$ and the expected winners are correct. From the above, $\tau(A|a,\sigma_n)>\tau(B|b,\sigma_n)>\frac{1}{2}$, so by Claim 2 $b(b,\sigma_n)>\frac{1}{2}$. □

Proof of Theorem 7:

Proof. First, note that if there is no t so that $r(t|a)\neq r(t|b)$, vote shares must be equal across states, completing the proof. Therefore, assume that for some t, $r(t|a)\neq r(t|b)$.

Suppose, for the sake of contradiction, that σ_n is an equilibrium for Γ_n where $\tau(A|a,\sigma_n)>\frac{1}{2}$ and $\tau(B|b,\sigma_n)>\frac{1}{2}$.

Claim 3. $BR_t(\sigma_n)=\hat{BR}_t(\sigma_n)$ for all t.

Proof. If $BR_t(\sigma_n) \neq \hat{BR}_t(\sigma_n)$ then either

$$\mathbb{E}[U|a,\sigma_n] \geq \mathbb{E}[U|b,\sigma_n] + \frac{1}{2}(Pr(Piv_B|b) + Pr(Piv_B|a)) \tag{C.4.1}$$

or

$$\mathbb{E}[U|b,\sigma_n] \geq \mathbb{E}[U|a,\sigma_n] + \frac{1}{2}(Pr(Piv_A|b) + Pr(Piv_A|a)) \tag{C.4.2}$$

by Lemma 28.

In the first case, some type $\hat{t}$ plays a mixed strategy in σ_n. First, suppose equation (C.4.1) holds. In this case, because σ_n is an equilibrium, Lemma 28 implies that $b(b,\sigma_n) = p_{\hat{t}}$. Because voters lack confidence, $p_{\hat{t}} < \frac{1}{2}$ which implies that

$$|\tau(A|a,\sigma_n) - \frac{1}{2}| < |\tau(B|b,\sigma_n) - \frac{1}{2}|$$

by Claim 2. Since $\tau(A|a,\sigma_n), \tau(B|b,\sigma_n) > \frac{1}{2}$ it follows that $\tau(A|a,\sigma_n) < \tau(B|b,\sigma_n)$ and Lemma 29 implies that $\mathbb{E}[U|b,\sigma_n] > \mathbb{E}[U|a,\sigma_n]$, a contradiction.

Similarly, if instead equation (C.4.2), it must be that $b(n,\sigma_n) = q_{\hat{t}}$. Because voters lack confidence, $q_{\hat{t}} > \frac{1}{2}$ which implies that

$$|\tau(A|a,\sigma_n) - \frac{1}{2}| > |\tau(B|b,\sigma_n) - \frac{1}{2}$$

by Claim 2. Since $\tau(A|a,\sigma_n), \tau(B|b,\sigma_n) > \frac{1}{2}$ it follows that $\tau(A|a,\sigma_n) > \tau(B|b,\sigma_n)$, Lemma 29 implies that $\mathbb{E}[u|a,\sigma_n] > \mathbb{E}[U|b,\sigma_n]$, a contradiction.

In the second case, all types play pure strategies in σ_n. Further, at least one type (WLOG, 1) votes for A for sure and another type (WLOG, 2) votes for B for sure. By Lemma 28, if equation (C.4.1) holds then $p_2 \leq b(b,\sigma_n) \leq p_1 < \frac{1}{2}$. By Lemma 29, $\mathbb{E}[U|b,\sigma_n] > \mathbb{E}[U|a,\sigma_n]$, contradicting that equation (C.4.1) holds. Suppose instead that equation (C.4.2) holds. By Lemma 28, $\frac{1}{2} < q_2 \leq b(b,\sigma_n) \leq q_1$. By Lemma 29, $\mathbb{E}[U|a,\sigma_n] > \mathbb{E}[U|b,\sigma_n]$, contradicting that equation (C.4.2) holds.

Hence, $BR_t(\sigma_n) = \hat{BR}_t(\sigma_n)$ for all t. □

I now show that no type plays a pure strategy.

Claim 4. $\sigma_n(t) \in (0,1)$ for all t.

Proof. Suppose $\sigma_n(t)(A) \in \{0,1\}$ for some t. WLOG, assume that either $\sigma_n(2)(A) =$

1 or $\sigma_n(2)(B) = 1$.

Assume the former. Then it must be that $1 \in BR_2(\sigma_n)(A)$ so

$$b(b, \sigma_n) \leq p_2 < \frac{1}{2}$$

by Lemma 28. By Lemma 30 it must be that $\hat{s}(A, \sigma_n) > \frac{1}{2}$. By assumption, some type of voter must vote for A with probability smaller than $\frac{1}{2}$. WLOG, assume this type is 1, so that $\sigma_n(1)(A) \leq \frac{1}{2} < \hat{s}(A, \sigma_n)$ for n high enough. Hence, it must be that $b(b, \sigma_n) \geq q_1$. Combining with $p_2 \geq b(b, \sigma_n)$ gives that $p_2 \geq q_1$, which is a contradiction of $p_2 < \frac{1}{2} < q_1$.

Now, assume the latter. It must be that $0 \in BR_2(\sigma_n)(A)$ so $b(B, \sigma_n) \geq q_2$ by Lemma 28. By Lemma 30 it must be that $\hat{s}(A, \sigma_n) < \frac{1}{2}$. By assumption, some type of voter must vote for A with probability larger than $\frac{1}{2}$. WLOG, assume this type is 1, so that $\sigma_n(1)(A) \geq \frac{1}{2} > \hat{s}(A, \sigma_n)$ for n high enough. Lemma 28 implies that $b(B, \sigma_n) \leq p_1$. Combining with $b(B, \sigma_n) \geq q_2$ gives that $p_1 \geq q_2$, which is a contradiction of $p_1 < \frac{1}{2} < q_2$. □

This claim shows that all types of voters must play a mixed strategy. Setting $[\underline{p}, \bar{p}] = \cap_{t \in T} [p_t, q_t]$, $b(b, \sigma_n) \in [\underline{p}, \bar{p}]$, since otherwise at least one type of voter plays a pure strategy by Lemma 28. Further, if $b(b, \sigma_n) \in (\underline{p}, \bar{p})$, the best response of all voters is to play $\sigma_n(t)(A) = \hat{s}(A, \sigma_n)$. Because of this, vote shares in each state are the same, a contradiction.

Claim 5. Suppose that $b(b, \sigma_n) = \underline{p}$. Then the expected winner in state b is not B.

Proof. WLOG, assume that $\underline{p} = p_1$; in fact, $p_1 = \max_{t \in T} p_t$ so

$$q_t > b(b, \sigma_n) > p_t \forall t \neq 1.$$

By Lemma 28, $\sigma_n(1)(A) \geq \hat{s}(A, \sigma_n)$ and $\sigma_n(t)(A) = \hat{s}(A, \sigma_n)$ for all $t \neq 1$. Because $b(b, \sigma_n) = p_1 < \frac{1}{2}$, by Lemma 30 it must be that $\hat{s}(A, \sigma_n) > \frac{1}{2}$. Therefore $\sigma_n(t)(A) > \frac{1}{2}$ for all t. Therefore, $\tau(B|b, \sigma_n) < \frac{1}{2}$ and thus B is not the expected winner in state b. □

Claim 6. Suppose that $b(B, \sigma_n) = \bar{p}$. Then the expected winner in state a is not A.

Proof. WLOG, assume that $\bar{p} = q_1$. By Lemma 28, $\sigma_n(1)(A) \leq \hat{s}(A, \sigma_n)$ and $\sigma_n(t)(A) = \hat{s}(A, \sigma_n)$ for all $t \neq 1$. By Lemma 30 it must be that $\hat{s}(A, \sigma_n) < \frac{1}{2}$. Therefore $\sigma_n(t)(A) < \frac{1}{2}$ for all t, so $\tau(A|a, \sigma_n) < \frac{1}{2}$ and A is not the expected winner in state a. □

Therefore, there is no equilibrium where both $\tau(A|a, \sigma_n) > \frac{1}{2}$ and $\tau(B|b, \sigma_n) > \frac{1}{2}$. □

Proof of Proposition 11:

Proof. Suppose σ is played. Clearly, $\mathbb{E}[U|a, \sigma] = \mathbb{E}[U|b, \sigma] = \frac{1}{2}$. This implies that $BR_t(\sigma) = \hat{BR}_t(\sigma)$ for all t by Lemma 28. Further, note that $b(b, \sigma) = \frac{1}{2}$ since $Pr(Piv_c|a) = Pr(Piv_c|b)$ for $c \in \{A, B\}$ since the vote shares are equal in both states. Since $b(b, \sigma) \in [p_t, q_t]$, voters of type t are willing to play $\sigma(t)(A) = \hat{s}(A, \sigma) = \frac{1}{2}$. Therefore, σ is an equilibrium. □

Proof of Theorem 8:

Proof. This proof adapts the arguments of Myerson [1998] Theorem 2.

Relabel $T = \{1, 2, ..., T\}$ so that $\min_{p \in \Pi_i} p(a) < \min_{p \in \Pi_{i+1}} p(a)$ for every $i \in \{1, 2, ..., T-1\}$. Denote $[h] = \max_{z \in \mathbb{Z}} z \leq h$ and $\sigma(h)$ for some $h \in [1, T]$ the strategy profile such that if h is an integer then $\sigma(t)(A) = 0$ if $t \leq h$ and $\sigma(t)(A) = 1$ if $t > h$. If h is not an integer then $\sigma(h)$ is such that $\sigma(t)(A) = 0$ if $t < [h]$ and $\sigma(t)(A) = 1$ if $t > h$ and $\sigma([h])(A) = h - [h]$. The proof will show that for all n high enough, there is an $h(n)$ so that $\sigma(h(n))$ is an equilibrium and that the expected winner in a is A and the expected winner in b is B.

Define functions $z : [1, T] \times \mathbb{N} \to [0, 1]$ and $\beta : [1, T] \times \mathbb{N} \to [0, 1]$ by the formulas

$$z(h, n) := \begin{cases} \hat{s}(A, \sigma(h), n) & \hat{s}(A, \sigma(h), n) \in [0, 1] \\ 1 & \hat{s}(A, \sigma(h), n) > 1 \\ 0 & \hat{s}(A, \sigma(h), n) < 0 \end{cases}$$

where $\hat{s}(c, \sigma, n)$ is $\hat{s}(c, \sigma)$ when there are n expected players. Further, define $\beta(h, n)$ to be $b(b, \sigma(h))$ when there are n expected players.

Let $q_t = \max_{p\in\Pi_t} p(a)$ and $p_t = \min_{p\in\Pi_t} p(a)$. If $\hat{s}(A,\sigma,n) < 0$ then

$$\mathbb{E}[U|b,\sigma] - \mathbb{E}[U|a,\sigma] + \frac{1}{2}(Pr(Piv_B|b) + Pr(Piv_B|a)) < 0$$

so that $\mathbb{E}[U|a,v_B,\sigma] > \mathbb{E}[U|b,v_B,\sigma]$. Hence

$$BR_t(\sigma) = \begin{cases} 1 & b(b,\sigma) > p_t \\ [0,1] & b(b,\sigma) = p_t \\ 0 & b(b,\sigma) < p_t \end{cases}$$

by Lemma 28. Similarly, if $\hat{s}(A,\sigma) > 1$ then $1 - \hat{s}(A,\sigma) < 0$ which implies

$$\mathbb{E}[U|a,\sigma] - \mathbb{E}[U|b,\sigma] + \frac{1}{2}((Pr(Piv_A|a) + Pr(Piv_A|b))) < 0$$

and thus $\mathbb{E}[U|a,v_B,\sigma] > \mathbb{E}[U|b,v_B,\sigma]$. Hence

$$BR_t(\sigma) = \begin{cases} 1 & b(b,\sigma) > q_t \\ [0,1] & b(b,\sigma) = q_t \\ 0 & b(b,\sigma) < q_t \end{cases}$$

by Lemma 28. Otherwise, $BR_t(\sigma)(A) = \hat{BR}_t(\sigma)(A)$.

Given the above notes, Lemma 28 shows that σ_h is an equilibrium if $\beta(h,n) \in \eta(h,n)$

$$\eta(h,n) = \begin{cases} [q_h, p_{h+1}] & h \in \mathbb{Z} \\ q_{[h]} & h \in ([h] + z(h,n), [h]+1) \\ [p_{[h]}, q_{[h]}] & h = [h] + z(h,n) \\ p_{[h]} & h \in ([h], [h] + z(h,n)) \end{cases}.$$

It's clear that $\hat{s}(\cdot,\sigma_n)$ is continuous by construction. It follows that $z(\cdot,n)$ is continuous since it can be written as the minimum of two continuous functions. Therefore $\eta(\cdot,n)$ is upper hemi continuous, compact and convex.

There exists numbers $I(a) \neq I(b)$ so that $\tau(A|\omega,\sigma_{I(\omega)}) = \tau(B|\omega,\sigma_{I(\omega)})$ for each $\omega \in \{a,b\}$ and for every $h \in (I(a), I(b))$ (or $(I(b), I(a))$ if $I(b) < I(a)$) $\tau(A|\omega,\sigma_h) \neq$

$\tau(B|\omega,\sigma_h)$ for each ω because $r(\cdot|a) \neq r(\cdot|b)$ and $h \mapsto \tau(c|\omega,\sigma_h)$ is a continuous function with range equal to $[0,1]$. Assume WLOG that $I(a) < I(b)$. For n high enough, $\exists h(n)$ so that $\beta(h(n),n) \in \eta(h(n),n)$ and $h(n) \in (I(a),I(b))$. This follows from $\beta(I(a),n) \to 0$, $\beta(I(b),n) \to 1$, $\beta(\cdot,n)$ is continuous and $\eta(\cdot,n)$ is convex and upper hemi continuous. Since $h(n) \in (I(a),I(b))$, $\tau(A|a,\sigma_{h(n)}) > \tau(B|a,\sigma_{h(n)})$ and $\tau(B|b,\sigma_{h(n)}) > \tau(A|b,\sigma_{h(n)})$. Define $\sigma_n^* = \sigma_{h(n)}$ when n is large enough; otherwise, σ_n^* let σ_n^* be an arbitrary equilibrium. Using the arguments of Myerson [1998] Theorem 2, the sequence of equilibrium vote shares from $(\sigma_n^*)_{n=1}^{\infty}$ must converge. Applying the law of large numbers gives that the correct candidate is elected with arbitrarily high probability in both states. Therefore, $(\Gamma_n)_{n=1}^{\infty}$ satisfies FIE. □

C.5 Proofs from Section 3.5

Theorem 9 follows from a special case of Theorems 11 and 12.

Theorem 11. *Suppose Γ_n is an ambiguous voting game with abstention that has voters who lack confidence and posteriors that respect likelihood ratios. If σ_n is an equilibrium for Γ_n where the worst case scenario for all voters is not independent of their vote and the expected vote share for A in state a is greater than $\frac{1}{2}$, then the expected vote share for B in state b is less than $\frac{1}{2}$.*

Proof. The proof will be by contradiction. Suppose σ_n is an equilibrium for Γ_n where the worst case scenario for all voters is not independent of their vote and the expected vote share for A in state a is greater than $\frac{1}{2}$ and the expected vote share for B in state b is also greater than $\frac{1}{2}$.

Begin by deriving the best response correspondence for voters when the worst case scenario varies with the strategy played. For any strategy $s \in \Delta C$, represent s by the ordered pair $(\frac{s(A)}{1-s(\emptyset)}, s(\emptyset))$ if $s(\emptyset) < 1$ and $(0,1)$ otherwise. Note that there is a bijection between these ordered pairs corresponds and each strategy profile. Now, define a function $\hat{s} : \Omega \times [0,1) \times (\Delta C)^T \to \mathbb{R}$ by $\hat{s}(A; s, \sigma)$ equals

$$\frac{2(\mathbb{E}[U|b,\sigma] - \mathbb{E}[U|a,\sigma]) + (1-s)[Pr(Piv_B|a,\sigma) + Pr(Piv_B|b,\sigma)]}{(1-s)[Pr(Piv_B|a,\sigma) + Pr(Piv_B|b,\sigma) + Pr(Piv_A|a,\sigma) + Pr(Piv_A|b,\sigma)]}$$

and $\bar{s} : \sigma \to [0,1]$ implicitly by

$$\hat{s}(A;\bar{s}(\sigma),\sigma)=\begin{cases}1 & if\ \mathbb{E}[U|b,\sigma]>\mathbb{E}[U|a,\sigma]\\ 0 & if\ \mathbb{E}[U|b,\sigma]<\mathbb{E}[U|a,\sigma]\end{cases}$$

and $\bar{s}(\sigma)=1$ if $\mathbb{E}[U|b,\sigma]=\mathbb{E}[U|a,\sigma]$. Note that if $\sigma(t)(\emptyset)<\bar{s}(\sigma)$, the voter's worst case scenario still changes with her vote. In this case, playing the strategy defined by $\sigma(t)(A)=\hat{s}(A;\sigma(t)(\emptyset),\sigma)$ equalizes the voter's expected utilities across states. On the other hand, if $\sigma(t)(\emptyset)\geq\bar{s}(\sigma)$, the voter is abstaining enough that her vote will no longer affect the worst case scenario.

Lemma 31. *Suppose that the worst case scenario is not independent of the strategy picked given σ and that the expected winner is correct in each state. If σ is an equilibrium and $b(b,\sigma)\in(p_t,q_t)$, then $\sigma(t)\in BR_t(\sigma)$ where*

$$BR_t(\sigma)=\begin{cases}\{(0,1)\} & if\ \mathbb{E}[U|b,\sigma]=\mathbb{E}[U|a,\sigma]\\ =\begin{cases}\{(0,1)\} & if\ \frac{q_t}{1-q_t}<\frac{Pr(Piv_A|b,\sigma)}{Pr(Piv_A|a,\sigma)}\\ \{1\}\times[\bar{s}(\sigma),1] & if\ \frac{q_t}{1-q_t}=\frac{Pr(Piv_A|b,\sigma)}{Pr(Piv_A|a,\sigma)}\\ \{(1,\bar{s}(\sigma))\} & if\ \frac{q_t}{1-q_t}>\frac{Pr(Piv_A|b,\sigma)}{Pr(Piv_A|a,\sigma)}\end{cases} & if\ \mathbb{E}[U|b,\sigma]>\mathbb{E}[U|a,\sigma]\\ =\begin{cases}\{(0,1)\} & if\ \frac{p_t}{1-p_t}>\frac{Pr(Piv_B|b,\sigma)}{Pr(Piv_B|a,\sigma)}\\ \{0\}\times[\bar{s}(\sigma),1] & if\ \frac{p_t}{1-p_t}=\frac{Pr(Piv_B|b,\sigma)}{Pr(Piv_B|a,\sigma)}\\ \{(0,\bar{s}(\sigma))\} & if\ \frac{p_t}{1-p_t}<\frac{Pr(Piv_B|b,\sigma)}{Pr(Piv_B|a,\sigma)}\end{cases} & if\ \mathbb{E}[U|b,\sigma]<\mathbb{E}[U|a,\sigma]\end{cases}.$$

Proof. (I drop the subscript t for convenience).

Suppose $p<b(b,\sigma)<q$. If the voter plays strategy (s,θ), she gets

$$\begin{aligned}V_t(s,\theta;\sigma) &= \min_{\pi\in[p,q]}\pi\{\mathbb{E}[U|a,\sigma]\\ &\quad +(1-\theta)[sPr(Piv_A|a,\sigma)-(1-s)Pr(Piv_B|a,\sigma)]\}+\\ &\quad +(1-\pi)\{\mathbb{E}[U|b,\sigma]\\ &\quad +(1-\theta)[(1-s)Pr(Piv_B|b,\sigma)-sPr(Piv_A|b,\sigma)]\}.\end{aligned}$$

Given a fixed $\theta<\bar{s}(\sigma)$, consider $v_{\theta\sigma}:[0,1]\to\mathbb{R}$ define by $v_{a\sigma}(s)=V_t(s,\theta;\sigma)$.

Note that $\partial v_{a\sigma}(s)$ is equal to

$$\begin{cases} \{(1-\theta)[p[Pr(Piv_A|a,\sigma)+Pr(Piv_B|a,\sigma)]- & if\ s>\hat{s}(A;\theta,\sigma) \\ \quad -(1-p)[Pr(Piv_B|b,\sigma)+Pr(Piv_A|b,\sigma)]]\} & \\ \{(1-a\theta)[\pi[Pr(Piv_A|a,\sigma)+Pr(Piv_B|a,\sigma)]- & if\ s=\hat{s}(A,\theta,\sigma) \\ \quad -(1-\pi)[(Pr(Piv_B|b,\sigma)+Pr(Piv_A|b,\sigma)]] : \pi\in[p,q]\} & \\ \{(1-\theta)[q[Pr(Piv_A|a,\sigma)+Pr(Piv_B|a,\sigma)]+ & if\ s<\hat{s}(A,\theta,\sigma) \\ \quad +(1-q)[Pr(Piv_B|b,\sigma)+Pr(Piv_A|b,\sigma)]]\} & \end{cases}.$$

As in Lemma 1, given $p<b(b,\sigma)<q$, $0\in\partial v_{\theta\sigma}(s)$ only if $s=\hat{s}(A,\theta,\sigma)$. Given this, consider $v_\sigma:[0,1]\to\mathbb{R}$ defined by $v_\sigma(\theta)=V_t(\hat{s}(A,\theta,\sigma),\theta,\sigma)$. Write $pc\omega = Pr(Piv_c|\omega,\sigma)$. By construction

$$\mathbb{E}[U|a,\sigma]+(1-\theta)[\hat{s}Pr(Piv_A|a,\sigma)-(1-\hat{s})Pr(Piv_B|a,\sigma)]] = \\ [\mathbb{E}[U|b,\sigma]+(1-\theta)[(1-\hat{s})Pr(Piv_B|b,\sigma)-\hat{s}Pr(Piv_A|b,\sigma)]]$$

when $\hat{s}=\hat{s}(A,\theta,\sigma)$. So if $\theta<\bar{s}(\sigma)$

$$\begin{aligned} v_\sigma(\theta) &= \mathbb{E}[U|a,\sigma]+(1-\theta)[\hat{s}Pr(Piv_A|a,\sigma)-(1-\hat{s})Pr(Piv_B|a,\sigma)]] \\ \partial v_\sigma(\theta) &= \{\tfrac{\partial}{\partial\theta}[(1-\theta)\tfrac{2(\mathbb{E}[U|b,\sigma]-\mathbb{E}[U|a,\sigma])+(1-\theta)[pBa+pBb]}{(1-\theta)[pAa+pBb+pBa+pAb]}pAa- \\ &\qquad -(1-\theta)\tfrac{2(\mathbb{E}[U|a,\sigma]-\mathbb{E}[U|b,\sigma])+(1-\theta)[pAa+pAb]}{(1-\theta)[pAa+pBb+pBa+pAb]}pBa]\} \\ &= \{\frac{\partial}{\partial\theta}[\frac{2(\mathbb{E}[U|b,\sigma]-\mathbb{E}[U|a,\sigma])-2(\mathbb{E}[U|a,\sigma]-\mathbb{E}[U|b,\sigma])}{pAa+pBb+pBa+pAb}+ \\ &\qquad +\frac{(1-\theta)[pBa+pBb]pAa-pBa[pAa+pAb]}{pAa+pBb+pBa+pAb}]\} \\ &= \{\frac{pBa[pAa+pAb]-[pBa+pBb]pAa}{pAa+pBb+pBa+pAb}\} \\ &= \{\frac{pBa(pAb)-pBb(pAa)}{pAa+pBb+pBa+pAb}\} \end{aligned}$$

Since FIE implies that $\frac{Pr(Piv_A|A,\sigma)}{Pr(Piv_B|A,\sigma)}<\frac{Pr(Piv_A|B,\sigma)}{Pr(Piv_B|B,\sigma)}$, no $\sigma(t)(\emptyset)<\bar{s}(\sigma)$ is optimal. Therefore, the voter abstains enough that worst case scenario is independent of whether she votes for A or B when she votes.

We can think of her as a SEU voter that assigns either probability p to a (if $\mathbb{E}[U|b,\sigma] < \mathbb{E}[U|a,\sigma]$) or q to a (if $\mathbb{E}[U|b,\sigma] > \mathbb{E}[U|a,\sigma]$). In this case, because $p < b(b,\sigma) < q$, the voter votes for B (in the first case) or A (in the second case) for sure conditional on voting. In the first case, she abstains for sure if $\frac{p}{1-p} > \frac{Pr(Piv_B|b,\sigma)}{Pr(Piv_B|a,\sigma)}$, and abstains with probability $\bar{s}(\sigma)$ if $\frac{p}{1-p} < \frac{Pr(Piv_B|b,\sigma)}{Pr(Piv_B|a,\sigma)}$. She is willing to abstain with any probability between $[\bar{s}(\sigma), 1]$ if $\frac{p}{1-p} = \frac{Pr(Piv_B|b,\sigma)}{Pr(Piv_B|a,\sigma)}$. In the second case, she abstains for sure if $\frac{q}{1-q} < \frac{Pr(Piv_A|b,\sigma)}{Pr(Piv_A|a,\sigma)}$, and abstains with probability $\bar{a}(\sigma)$ if $\frac{q}{1-q} > \frac{Pr(Piv_A|b,\sigma)}{Pr(Piv_A|a,\sigma)}$. She is willing to abstain with any probability between $[\bar{a}(\sigma), 1]$ if $\frac{q}{1-q} = \frac{Pr(Piv_A|b,\sigma)}{Pr(Piv_A|a,\sigma)}$. This establishes the best response correspondence when $p < b(b,\sigma) < q$. □

Lemma 32. *Suppose that the worst case scenario is not independent of the strategy picked given σ and that the expected winner is correct in each state. If σ is an equilibrium and $b(b,\sigma) \leq p_t$, then $\sigma(t) \in BR_t(\sigma)$ where*

$$BR_t(\sigma) = \begin{cases} \{(1,0)\} & if \ \frac{p_t}{1-p_t} > \frac{Pr(Piv_A|b,\sigma)}{Pr(Piv_A|a,\sigma)} \\ \{1\} \times [0, \bar{s}(\sigma)] & if \ \frac{p_t}{1-p_t} = \frac{Pr(Piv_A|b,\sigma)}{Pr(Piv_A|a,\sigma)} \\ \widetilde{BR_{A,t}}(\sigma) & if \ \frac{p_t}{1-p_t} < \frac{Pr(Piv_A|b,\sigma)}{Pr(Piv_A|a,\sigma)} \end{cases}$$

and $\widetilde{BR_{A,t}}(\sigma)$ is equal to

$$\begin{cases} = \begin{cases} \{(0,1)\} & if \ \frac{q_t}{1-q_t} < \frac{Pr(Piv_A|b,\sigma)}{Pr(Piv_A|a,\sigma)} \\ \{1\} \times [\bar{s}(\sigma), 1] & if \ \frac{q_t}{1-q_t} = \frac{Pr(Piv_A|b,\sigma)}{Pr(Piv_A|a,\sigma)} \\ \{(1, \bar{s}(\sigma))\} & if \ \frac{q_t}{1-q_t} > \frac{Pr(Piv_A|b,\sigma)}{Pr(Piv_A|a,\sigma)} \end{cases} & if \ \mathbb{E}[U|b,\sigma] > \mathbb{E}[U|a,\sigma] \\ \{(0,1)\} & otherwise \end{cases}$$

Proof. She votes for A conditional on voting because $b(b,\sigma)$ is low enough relative to her priors. She never abstains if $\frac{p_t}{1-p_t} > \frac{Pr(Piv_A|b,\sigma)}{Pr(Piv_A|a,\sigma)}$. If $\frac{p_t}{1-p_t} = \frac{Pr(Piv_A|b,\sigma)}{Pr(Piv_A|a,\sigma)}$, she's indifferent between abstaining and voting for B and so is willing to play any mixture between voting and abstaining. She abstains at least enough that she can't affect the outcome with her vote if $\frac{p_t}{1-p_t} < \frac{Pr(Piv_A|b,\sigma)}{Pr(Piv_A|a,\sigma)}$. If she abstains more than $\bar{s}(\sigma)$, she acts as if she's an SEU voter who assigns probability p_t to a if $\mathbb{E}[U|a,\sigma] > \mathbb{E}[U|b,\sigma]$ and q_t to a if $\mathbb{E}[U|b,\sigma] > \mathbb{E}[U|a,\sigma]$. Her best response correspondence is exactly as in

Bouton and Castanheira [2009], establishing the result. □

Lemma 33. *Suppose that the worst case scenario is not independent of the strategy picked given σ and that the expected winner is correct in each state. If σ is an equilibrium and $b(B,\sigma) \geq q_t$, then $\sigma(t) \in BR_t(\sigma)$ where*

$$BR_t(\sigma) = \begin{cases} \{(0,0)\} & if\ \frac{q_t}{1-q_t} < \frac{Pr(Piv_B|b,\sigma)}{Pr(Piv_B|a,\sigma)} \\ \{0\} \times [0, \bar{s}(\sigma)] & if\ \frac{q_t}{1-q_t} = \frac{Pr(Piv_B|b,\sigma)}{Pr(Piv_B|a,\sigma)} \\ \widetilde{BR_{B,t}}(\sigma) & if\ \frac{q_t}{1-q_t} > \frac{Pr(Piv_B|b,\sigma)}{Pr(Piv_B|a,\sigma)} \end{cases}$$

and

$$\widetilde{BR_{B,t}}(\sigma) = \begin{cases} = \begin{cases} \{(0,1)\} & if\ \frac{p_t}{1-p_t} > \frac{Pr(Piv_B|b,\sigma)}{Pr(Piv_B|a,\sigma)} \\ \{0\} \times [\bar{s}(\sigma), 1] & if\ \frac{p_t}{1-p_t} = \frac{Pr(Piv_B|b,\sigma)}{Pr(Piv_B|a,\sigma)} \\ \{(0, \bar{s}(\sigma))\} & if\ \frac{p_t}{1-p_t} < \frac{Pr(Piv_B|b,\sigma)}{Pr(Piv_B|a,\sigma)} \end{cases} & if\ \mathbb{E}[U|a,\sigma] > \mathbb{E}[U|b,\sigma] \\ \{(0,1)\} & otherwise \end{cases}$$

Proof. She votes for B conditional on voting because $b(b,\sigma)$ is high enough. She never abstains if $\frac{q_t}{1-q_t} < \frac{Pr(Piv_B|b,\sigma)}{Pr(Piv_B|a,\sigma)}$. If $\frac{q_t}{1-q_t} = \frac{Pr(Piv_B|b,\sigma)}{Pr(Piv_B|a,\sigma)}$, she's indifferent between abstaining and voting for B and so is willing to play any mixture between voting and abstaining. She abstains at least enough that she can't affect the outcome with her vote if $\frac{q_t}{1-q_t} > \frac{Pr(Piv_B|b,\sigma)}{Pr(Piv_B|a,\sigma)}$. If she abstains more than $\bar{s}(\sigma)$, she acts as if she's an SEU voter who assigns probability p_t to a if $\mathbb{E}[U|a,\sigma] > \mathbb{E}[U|b,\sigma]$ and q_t to a if $\mathbb{E}[U|b,\sigma] > \mathbb{E}[U|a,\sigma]$. Her best response correspondence is exactly as in Bouton and Castanheira [2009], establishing the result. □

Now, focus on the specific conditions at equilibrium. Because posteriors that respect likelihood ratios and voters lack of confidence, $p_2 \leq p_1 < \frac{1}{2} < q_2 \leq q_1$ (perhaps after relabeling). These values partition $[0,1]$ into regions where the best response correspondence of the voters has similar properties when $b(\cdot)$ is within that region. Proceed by analyzing these regions separately.

Suppose now that $b(b,\sigma_n) \in (p_2, p_1]$. By assumption, σ_n is so that $\sigma_n(1)(\emptyset) < 1$. Since $b(b,\sigma_n) < p_1$, Lemma 32 gives that $\sigma_n(1)(B) = 0$.

First, consider the case where $\sigma_n(1)(\emptyset) < \bar{s}(\sigma_n)$ so $\frac{p_1}{1-p_1} \geq \frac{Pr(Piv_A|b,\sigma)}{Pr(Piv_A|a,\sigma)}$. Since $\tau(B|b,\sigma_n) > \tau(A|b,\sigma_n)$, it must be that $\sigma_n(2)(A) = 0$ and $\sigma_n(1)(\emptyset) < 1$. By Lemma

31, $\mathbb{E}[U|b, \sigma_n] > \mathbb{E}[U|a, \sigma_n]$ and

$$\frac{q_2}{1-q_2} \leq \frac{Pr(Piv_B|b,\sigma)}{Pr(Piv_B|a,\sigma)}$$

. Because $\tau(B|b, \sigma_n) > \tau(A|b, \sigma_n)$ and $\tau(A|a, \sigma_n) > \tau(B|a, \sigma_n)$, it follows that

$$Pr(Piv_A|b, \sigma_n) > Pr(Piv_B|b, \sigma_n)$$

and

$$Pr(Piv_A|a, \sigma) < Pr(Piv_B|a, \sigma)$$

and

$$\frac{Pr(Piv_A|b,\sigma)}{Pr(Piv_A|a,\sigma)} > \frac{Pr(Piv_B|b,\sigma)}{Pr(Piv_B|a,\sigma)}.$$

However, $p_1 < q_2$ so these are mutually impossible.

Now, consider the case where $1 > \sigma_n(1)(\emptyset) \geq \bar{s}(\sigma_n)$. Since $b(b, \sigma_n) \leq p_1$, Lemma 32 gives that $\sigma_n(1)(B) = 0$ and $\sigma_n(1)(A) > 0$ implies that $\mathbb{E}[U|b, \sigma_n] > \mathbb{E}[U|a, \sigma_n]$. Since $\sigma_n(1)(B) = 0$, for the expected winner in state b to be correct it must hold that $\sigma_n(2)(B) > 0$. But because $\mathbb{E}[U|b, \sigma_n] > \mathbb{E}[U|a, \sigma_n]$ and $p_2 < b(b, \sigma_n) < q_2$, Lemma 31 gives that $\sigma_n(1)(B) = 0$, a contradiction.

Now suppose that $b(b, \sigma_n) \in [q_2, q_1)$. By assumption, $\sigma_n(2)(\emptyset) < 1$. Since $b(B, \sigma_n) > q_2$, Lemma 33 gives that $\sigma_n(2)(A) = 0$.

First, consider the case where $\sigma_n(2)(\emptyset) < \bar{s}(\sigma_n)$. From Lemma 33, $\frac{q_2}{1-q_2} \leq \frac{Pr(Piv_B|b,\sigma)}{Pr(Piv_B|a,\sigma)}$. By assumption, it must be that $\sigma_n(1)(A) > 0$ so $\sigma_n(1)(\emptyset) < 1$. Because $b(b, \sigma_n) \in (p_1, q_1)$, Lemma 31 requires that $\mathbb{E}[U|a, \sigma_n] > \mathbb{E}[U|b, \sigma_n]$ and $\frac{p_1}{1-p_1} \geq \frac{Pr(Piv_A|b,\sigma_n)}{Pr(Piv_A|a,\sigma_n)}$. Because $\tau(B|b, \sigma_n) > \tau(A|b, \sigma_n)$ and $\tau(A|a, \sigma_n) > \tau(B|a, \sigma_n)$, it follows that $Pr(Piv_A|b, \sigma_n) > Pr(Piv_B|b, \sigma_n)$ and

$$\frac{Pr(Piv_B|b,\sigma)}{Pr(Piv_B|a,\sigma)} < \frac{Pr(Piv_A|b,\sigma_n)}{Pr(Piv_A|a,\sigma_n)},$$

which is impossible since $p_1 < q_2$.

Now, consider the case where $1 > \sigma_n(2)(\emptyset) \geq \bar{s}(\sigma_n)$. By assumption and Lemma 33, $\sigma_n(2)(B) > 0$. From Lemma 32, $\mathbb{E}[U|a, \sigma_n] > \mathbb{E}[U|b, \sigma_n]$. But since $p_1 < b(b, \sigma_n) < q_1$, Lemma 31 yields that $\sigma_n(1)(A) = 0$, a contradiction.

If $b(b, \sigma_n) \in [0, p_1] \cup [p_2, q_1] \cup [q_2, 1]$ it follows from Lemmas 31-33 that all voters will vote for the same candidate whenever they do not abstain, a contradiction. $\square$

Assume WLOG that $r(1|a) + r(1|b) \geq 1$ and that $r(1|a) \geq r(1|b)$ (otherwise, relabel candidates and types).

Define $\hat{\tau} : \{A, B\} \times \Omega \to [0, 1]$ by

$$\begin{aligned}
\hat{\tau}(A|a) &= \left(\frac{\sqrt{r(2|a)} + \sqrt{r(2|b)}}{\sqrt{r(1|a)} + \sqrt{r(1|b)}}\right)^2 r(1|a) \\
\hat{\tau}(B|a) &= r(2|a) \\
\hat{\tau}(A|b) &= \left(\frac{\sqrt{r(2|a)} + \sqrt{r(2|b)}}{\sqrt{r(1|a)} + \sqrt{r(1|b)}}\right)^2 r(1|b) \\
\hat{\tau}(B|b) &= r(2|b)
\end{aligned}$$

which would be the limiting vote shares for each candidate in each state if voters were expected utility.

Theorem 12. *Fix any sequence $(\Gamma_n)_{n=1}^{\infty}$ of AVGAs with voters who lack confidence and posteriors that respect the likelihood ratio. If the inequalities*

$$2 + \sqrt{\frac{\hat{\tau}(B|b)}{\hat{\tau}(A|b)}} + \sqrt{\frac{\hat{\tau}(B|a)}{\hat{\tau}(A|a)}} > 2\left(\left(\frac{\hat{\tau}(A|b)\hat{\tau}(B|b)}{\hat{\tau}(A|a)\hat{\tau}(B|a)}\right)^{\frac{1}{4}} - 1\right)\frac{\sqrt{\frac{\hat{\tau}(A|b)}{\hat{\tau}(B|b)}}}{1 - \sqrt{\frac{\hat{\tau}(A|b)}{\hat{\tau}(B|b)}}} \tag{C.5.1}$$

and

$$2 + \sqrt{\frac{\hat{\tau}(B|b)}{\hat{\tau}(A|b)}} + \sqrt{\frac{\hat{\tau}(B|a)}{\hat{\tau}(A|a)}} > \frac{\sqrt{\frac{\hat{\tau}(A|b)}{\hat{\tau}(B|b)}}}{1 - \sqrt{\frac{\hat{\tau}(A|b)}{\hat{\tau}(B|b)}}} - \frac{\frac{\hat{\tau}(B|a)}{\sqrt{\hat{\tau}(A|b)\hat{\tau}(B|b)}}}{1 - \frac{\hat{\tau}(B|a)}{\sqrt{\hat{\tau}(A|b)\hat{\tau}(B|b)}}} \tag{C.5.2}$$

both hold and σ_n is an equilibrium for Γ_n where the expected votes in each state goes to infinity and the expected winners are correct given σ_n, then for n sufficiently high, the worst case scenario for all voters is independent of their vote in σ_n.

Proof. The proof will be by contradiction.

Without loss of generality, suppose that $r(1|a)+r(1|b) \geq r(2|a)+r(2|b)$ and that $\frac{r(1|a)}{r(1|b)} > \frac{r(2|a)}{r(2|b)}$ (which implies that $r(1|a) > r(1|b)$), so that $p_2 \leq p_1 < \frac{1}{2} < q_2 \leq q_1$. Define $\sigma(1)(\emptyset) = \bar{a} = 1 - (\frac{\sqrt{1-r(1|a)}+\sqrt{1-r(1|b)}}{\sqrt{r(1|a)}+\sqrt{r(1|b)}})^2$, $\sigma(1)(A) = 1-\bar{a}$ and $\sigma(2)(B) = 1$.

Lemma 34. *Suppose that $(\sigma^*_{n_k})$ is a convergent sub-sequence of equilibrium strategy profiles to Γ_{n_k} so that the worst case scenario for every voter is independent of her strategy for every $\sigma^*_{n_k}$. Then $\sigma_{n_k} \to \sigma$. Moreover, $\tau(A|a,\sigma) \leq \tau(B|b,\sigma)$ and $\frac{\tau(A|b,\sigma)}{\tau(B|b,\sigma)} \geq \frac{\tau(B|a,\sigma)}{\tau(A|a,\sigma)}$*, with equality only if $r(1|a)+r(1|b) = r(2|a)+r(2|b)$.

Proof. This follows follows from Bouton and Castanheira [2009] Lemma 1 and Theorem 1, noting that when the strategy profile is played all voters act as if SEU with posterior p_t or q_t. At the limit, it must be that

$$\mu(a) = \mu(b) \iff (\sqrt{\tau(A|a)} - \sqrt{\tau(B|a)})^2 = (\sqrt{\tau(B|b)} - \sqrt{\tau(A|b)})^2.$$

Rewriting,

$$\sqrt{(1-\bar{a})r(1|a)} - \sqrt{r(2|a)} = \sqrt{r(2|b)} - \sqrt{(1-\bar{a})r(1|b)}$$

where $\bar{a} = \sigma(1)(\emptyset)$. Solving for $\bar{a}$ yields

$$1 - (\frac{\sqrt{1-r(1|a)}+\sqrt{1-r(1|b)}}{\sqrt{r(1|a)}+\sqrt{r(1|b)}})^2.$$

The remaining results follow from algebra. □

Note that $\hat{\tau}(c|\omega) = \tau(c|\omega,\sigma)$.

The worst case scenarios is independent of the strategy chosen given σ is played if and only if either

$$\mathbb{E}[U|b,\sigma] - \frac{1}{2}Pr(Piv_A|b,\sigma) \geq \mathbb{E}[U|a,\sigma] + \frac{1}{2}Pr(Piv_A|a,\sigma) \quad \text{(C.5.3)}$$

or

$$\mathbb{E}[U|a,\sigma] - \frac{1}{2}Pr(Piv_B|a,\sigma) \geq \mathbb{E}[U|b,\sigma] + \frac{1}{2}Pr(Piv_B|b,\sigma) \quad \text{(C.5.4)}$$

as in Lemma 28. If $r(1|a) = r(2|b)$, then it's clear that neither of the equalities are satisfied because $\mathbb{E}[U|b,\sigma,n] = \mathbb{E}[Ua,\sigma,n]$. Therefore, suppose $r(1|a) \neq r(2|b)$ an

Consider the limiting equilibrium strategy profile. At this strategy profile, neither of these equations holds for n large enough.

Lemma 35. *If*

$$2+\sqrt{\frac{\tau(B|b,\sigma)}{\tau(A|b,\sigma)}}+\sqrt{\frac{\tau(B|a,\sigma)}{\tau(A|a,\sigma)}}>2(\frac{\tau(A|b,\sigma)\tau(B|b,\sigma)}{\tau(A|a,\sigma)\tau(B|a,\sigma)})^{\frac{1}{4}}-1)\frac{\sqrt{\frac{\tau(A|b,\sigma)}{\tau(B|b,\sigma)}}}{1-\sqrt{\frac{\tau(A|b,\sigma)}{\tau(B|b,\sigma)}}}$$

then $\mathbb{E}[U|b,\sigma]-\mathbb{E}[U|a,\sigma]<\frac{1}{2}(Pr(Piv_A|a,\sigma,n)+Pr(Piv_A|b,\sigma,n))$ *for n large enough.*

Proof. For notational purposes, drop the dependence on σ. Lemma 34 shows that

$$\tau(B|b)\tau(A|b)>\tau(A|a)\tau(B|a) \tag{C.5.5}$$

and

$$\frac{\tau(A|b)}{\tau(B|b)}>\frac{\tau(B|a)}{\tau(A|a)} \tag{C.5.6}$$

whenever the above conditions are satisfied. Set

$$\mu(\omega)=-\tau(A|\omega)-\tau(B|\omega)+2\sqrt{\tau(B|\omega)\tau(A|\omega)}$$

noting that $\mu(A)=\mu(B)=\mu\in[-1,0]$.

Since

$$\begin{aligned}\mathbb{E}[U|a,\sigma] &= 1-e^{-(\tau(A|a)+\tau(B|a))n}\{\sum_{k=1}^{\infty}(\frac{\tau(B|a)}{\tau(A|a)})^{\frac{k}{2}}I_k(2n\sqrt{\tau(A|a)\tau(B|a)}) \\ &\quad -\frac{1}{2}I_0(2n\sqrt{\tau(A|a)\tau(B|a)})\}\end{aligned}$$

and

$$\begin{aligned}\mathbb{E}[U|b,\sigma] &= 1-e^{-(\tau(A|b)+\tau(B|b))n}\{\sum_{k=1}^{\infty}\sqrt{\frac{\tau(A|b)}{\tau(B|b)}}^{k}I_k(2n\sqrt{\tau(A|b)\tau(B|b)}) \\ &\quad -\frac{1}{2}I_0(2n\sqrt{\tau(A|b)\tau(B|b)}\}\end{aligned}$$

(where $I_k(\cdot)$ is a modified Bessel function of the first kind (see Myerson [2000], p.

27)), the conclusion is equivalent to

$$e^{-(\tau(A|a)+\tau(B|a))n}\sum_{k=1}^{\infty}\sqrt{\frac{\tau(B|a)}{\tau(A|a)}}^{k} I_k(2n\sqrt{\tau(A|a)\tau(B|a)}) \qquad \text{(C.5.7)}$$
$$-e^{-(\tau(A|b)+\tau(B|b))n}\sum_{k=1}^{\infty}\sqrt{\frac{\tau(A|b)}{\tau(B|b)}}^{k} I_k(2n\sqrt{\tau(A|b)\tau(B|b)})$$

is less than

$$\frac{1}{2}(Pr(Piv_A|b)+Pr(Piv_A|a)+I_0(2n\sqrt{\tau(A|b)\tau(B|b)})-I_0(2n\sqrt{\tau(A|a)\tau(B|a)})).$$

Let $\phi(n)$ be the value of (C.5.7).

By Baricz [2010] equation (2.6) we have that if $y > x > 0$ and $k > 0$ is an integer then

$$I_k(x) < e^{x-y}(\frac{y}{x})^{\frac{1}{2}}I_k(y). \qquad \text{(C.5.8)}$$

Using equations (C.5.5) and (C.5.8), we have that

$$e^{-(\tau(A|a)+\tau(B|a))n}(\sum_{k=1}^{\infty}\sqrt{\frac{\tau(B|a)}{\tau(A|a)}}^{k} I_k(2n\sqrt{\tau(A|a)\tau(B|a)})$$
$$< e^{(\mu-2\sqrt{\tau(B|b)\tau(B|b)})n}(\frac{\tau(A|b)\tau(B|b)}{\tau(A|a)\tau(B|a)})^{\frac{1}{4}}\sum_{k=1}^{\infty}\sqrt{\frac{\tau(B|a)}{\tau(A|a)}}^{k} I_k(2n\sqrt{\tau(A|b)\tau(B|b)})$$

so we find that

$$\begin{aligned}
\frac{\phi(n)}{e^{-2\sqrt{\tau(B|A)\tau(B|B)})n}} &< e^{\mu n}(\frac{\tau(A|b)\tau(B|b)}{\tau(A|a)\tau(B|a)})^{\frac{1}{4}} \sum_{k=1}^{\infty} \sqrt{\frac{\tau(B|a)}{\tau(A|a)}}^{k} I_k(2n\sqrt{\tau(A|b)\tau(B|b)}) \\
&\quad -e^{\mu n} \sum_{k=1}^{\infty} \sqrt{\frac{\tau(A|b)}{\tau(B|b)}}^{k} I_k(2n\sqrt{\tau(A|b)\tau(B|b)}) \\
&< (\frac{\tau(A|b)\tau(B|b)}{\tau(A|a)\tau(B|a)})^{\frac{1}{4}} \sum_{k=1}^{\infty} \sqrt{\frac{\tau(B|b)}{\tau(A|b)}}^{k} I_k(2n\sqrt{\tau(A|b)\tau(B|b)}) \\
&\quad -e^{\mu n} \sum_{k=1}^{\infty} \sqrt{\frac{\tau(A|b)}{\tau(B|b)}}^{k} I_k(2n\sqrt{\tau(A|b)\tau(B|b)})
\end{aligned}$$

since $\frac{\tau(B|a)}{\tau(A|a)} < \frac{\tau(A|b)}{\tau(B|b)}$. Setting

$$\bar{K}(n) = \sum_{k=1}^{\infty} \sqrt{\frac{\tau(A|b)}{\tau(B|b)}}^{k} I_k(2n\sqrt{\tau(A|b)\tau(B|b)}) > 0$$

and

$$\theta = (\frac{\tau(A|b)\tau(B|b)}{\tau(A|a)\tau(B|a)})^{\frac{1}{4}} - 1 > 0$$

yields that

$$\mathbb{E}[U|b,\sigma] - \mathbb{E}[U|a,\sigma] < \theta e^{\mu n}\bar{K}(n)e^{-2n\sqrt{\tau(A|b)\tau(B|b)}}$$

Note that

$$\begin{aligned}
\bar{K}(n) &= \sum_{k=1}^{\infty} \sqrt{\frac{\tau(A|b)}{\tau(B|b)}}^{k} I_k(2n\sqrt{\tau(A|b)\tau(B|b)}) \\
&< \sum_{k=1}^{\infty} \sqrt{\frac{\tau(A|b)}{\tau(B|b)}}^{k} I_0(2n\sqrt{\tau(A|b)\tau(B|b)}) \\
&\approx \sum_{k=1}^{\infty} \sqrt{\frac{\tau(A|b)}{\tau(B|b)}}^{k} \frac{e^{\sqrt{(2n\sqrt{\tau(A|b)\tau(B|b)})^2}}}{\sqrt{2\pi\sqrt{(2n\sqrt{\tau(A|b)\tau(B|b)})^2}}} \\
&= \frac{\sqrt{\frac{\tau(A|b)}{\tau(B|b)}}}{1-\sqrt{\frac{\tau(A|b)}{\tau(B|b)}}} \frac{e^{2n\sqrt{\tau(A|b)\tau(B|b)}}}{2\sqrt{\pi n\sqrt{\tau(A|b)\tau(B|b)}}}
\end{aligned}$$

by Abramowitz and Stegun [1972] equations (9.7.1) and (9.7.7) and that when $k \geq 0$ it follows that $I_k(x) > I_{k+1}(x)$. Therefore, for n large enough

$$\phi(n) < \frac{\sqrt{\frac{\tau(A|b)}{\tau(B|b)}}}{1-\sqrt{\frac{\tau(A|b)}{\tau(B|b)}}} \theta e^{\mu n} I_0(2n\sqrt{\tau(A|b)\tau(B|b)}) e^{-2n\sqrt{\tau(A|B)\tau(B|B)}}$$

Since

$$\begin{aligned}
& \frac{1}{2}[Pr(Piv_A|b) + Pr(Piv_A|a) + e^{-(\tau(A|b)+\tau(B|b))n} I_0(2n\sqrt{\tau(A|b)\tau(B|b)}) \\
& \quad - e^{-(\tau(A|a)+\tau(B|a))n} I_0(2n\sqrt{\tau(A|a)\tau(B|a)})] \\
= & \frac{1}{2}[e^{-(\tau(A|b)+\tau(B|b))n}(2I_0(2n\sqrt{\tau(A|b)\tau(B|b)}) \\
& \quad + I_1(2n\sqrt{\tau(A|b)\tau(B|b)})\sqrt{\frac{\tau(B|b)}{\tau(A|b)}}) \\
& \quad + e^{-(\tau(A|a)+\tau(B|a))n} I_1(2n\sqrt{\tau(A|a)\tau(B|a)})\sqrt{\frac{\tau(B|a)}{\tau(A|a)}}] \\
\approx & \frac{e^{\mu n}}{4\sqrt{\pi n}}\left(\frac{\sqrt{\frac{\tau(B|a)}{\tau(A|a)}}}{(\tau(A|a)\tau(B|a))^{\frac{1}{4}}} + \frac{2+\sqrt{\frac{\tau(B|b)}{\tau(A|b)}}}{(\tau(A|b)\tau(B|b))^{\frac{1}{4}}}\right),
\end{aligned}$$

it suffices to show that

$$[\frac{\sqrt{\frac{\tau(B|a)}{\tau(A|a)}}}{(\tau(A|a)\tau(B|a))^{\frac{1}{4}}} + \frac{2+\sqrt{\frac{\tau(B|b)}{\tau(A|b)}}}{(\tau(A|b)\tau(B|b))^{\frac{1}{4}}}] > \frac{\sqrt{\frac{\tau(A|b)}{\tau(B|b)}}}{1-\sqrt{\frac{\tau(A|b)}{\tau(B|b)}}} \frac{2(\frac{\tau(A|b)\tau(B|b)}{\tau(A|a)\tau(B|a)})^{\frac{1}{4}}-1)}{(\tau(A|b)\tau(B|b))^{\frac{1}{4}}}.$$

Note that

$$\frac{\sqrt{\frac{\tau(B|a)}{\tau(A|a)}}}{(\tau(A|a)\tau(B|a))^{\frac{1}{4}}} + \frac{2+\sqrt{\frac{\tau(B|b)}{\tau(A|b)}}}{(\tau(A|b)\tau(B|b))^{\frac{1}{4}}} > \frac{2+\sqrt{\frac{\tau(B|b)}{\tau(A|b)}}+\sqrt{\frac{\tau(B|a)}{\tau(A|a)}}}{(\tau(A|b)\tau(B|b))^{\frac{1}{4}}}$$

because $\tau(B|b)\tau(A|b) > \tau(B|a)\tau(A|a)$. Therefore, if

$$2+\sqrt{\frac{\tau(B|b)}{\tau(A|b)}}+\sqrt{\frac{\tau(B|a)}{\tau(A|a)}} > 2(\frac{\tau(A|b)\tau(B|b)}{\tau(A|a)\tau(B|a)})^{\frac{1}{4}}-1)\frac{\sqrt{\frac{\tau(A|b)}{\tau(B|b)}}}{1-\sqrt{\frac{\tau(A|b)}{\tau(B|b)}}}$$

then the claim holds. □

Lemma 36. *If*

$$2+\sqrt{\frac{\tau(B|b,\sigma)}{\tau(A|b,\sigma)}}+\sqrt{\frac{\tau(B|a,\sigma)}{\tau(A|a,\sigma)}} > \frac{\sqrt{\frac{\tau(A|b,\sigma)}{\tau(B|b,\sigma)}}}{1-\sqrt{\frac{\tau(A|b,\sigma)}{\tau(B|b,\sigma)}}} - \frac{\frac{\tau(B|a,\sigma)}{\sqrt{\tau(A|b,\sigma)\tau(B|b,\sigma)}}}{1-\frac{\tau(B|a,\sigma)}{\sqrt{\tau(A|b,\sigma)\tau(B|b,\sigma)}}}$$

then $\mathbb{E}[U|a,\sigma] - \mathbb{E}[U|b,\sigma] < \frac{1}{2}Pr(Piv_B|a,\sigma) + \frac{1}{2}Pr(Piv_B|b,\sigma)$ *for* n *large enough.*

Proof. For notational purposes, drop the dependence on σ. As in Lemma 35, we can write the claim as

$$e^{-(\tau(A|b)+\tau(B|b))n}\sum_{k=1}^{\infty}\sqrt{\frac{\tau(A|b)}{\tau(B|b)}}^{k} I_k(2n\sqrt{\tau(A|b)\tau(B|b)}) - \tag{C.5.9}$$
$$- e^{-(\tau(A|a)+\tau(B|a))n}\sum_{k=1}^{\infty}\sqrt{\frac{\tau(B|a)}{\tau(A|a)}}^{k} I_k(2n\sqrt{\tau(A|a)\tau(B|a)})$$

is less than

$$\frac{1}{2}(Pr(Piv_A|b) + Pr(Piv_A|a) - I_0(2n\sqrt{\tau(A|b)\tau(B|b)}) + I_0(2n\sqrt{\tau(A|a)\tau(B|a)})).$$

Write $\phi(n)$ to be the value of (C.5.9). Note that

$$\begin{aligned}
&e^{-(\tau(A|a)+\tau(B|a))n}\sum_{k=1}^{\infty}\sqrt{\frac{\tau(B|a)}{\tau(A|a)}}^k I_k(2n\sqrt{\tau(A|a)\tau(B|a)}) \\
&> e^{-(\tau(A|a)+\tau(B|a))n}\sum_{k=1}^{\infty}\sqrt{\frac{\tau(B|a)}{\tau(A|a)}}^k \sqrt{\frac{\tau(A|a)\tau(B|a)}{\tau(A|b)\tau(B|b)}}^k \times \\
&\quad I_k(2n\sqrt{\tau(A|b)\tau(B|b)})e^{2n\sqrt{\tau(A|a)\tau(B|a)}-2n\sqrt{\tau(A|b)\tau(B|b)}} \\
&= e^{\mu n-2n\sqrt{\tau(A|b)\tau(B|b)}}\sum_{k=1}^{\infty}\frac{\tau(B|a)}{\sqrt{\tau(A|b)\tau(B|b)}}^k I_k(2n\sqrt{\tau(A|b)\tau(B|b)})
\end{aligned}$$

since whenever $k > \frac{1}{2}$ and $y > x$ we have

$$I_k(x) > (\frac{x}{y})^k e^{x-y} I_k(y) \tag{C.5.10}$$

by equation (2.2) of Baricz [2010].

We have that

$$\begin{aligned}
\phi(n) &< e^{-(\tau(A|b)+\tau(B|b))n}\sum_{k=1}^{\infty}\sqrt{\frac{\tau(A|b)}{\tau(B|b)}}^{k} I_k(2n\sqrt{\tau(A|b)\tau(B|b)}) - \\
&\quad - e^{\mu n-2n\sqrt{\tau(A|b)\tau(B|b)}}\sum_{k=1}^{\infty}\frac{\tau(B|a)}{\sqrt{\tau(A|b)\tau(B|b)}}^{k} I_k(2n\sqrt{\tau(A|b)\tau(B|b)}) \\
&= \frac{e^{\mu n}\sum_{k=1}^{\infty}(\sqrt{\frac{\tau(A|b)}{\tau(B|b)}}^{k} - \frac{\tau(B|a)}{\sqrt{\tau(A|b)\tau(B|b)}}^{k})I_k(2n\sqrt{\tau(A|b)\tau(B|b)})}{e^{2n\sqrt{\tau(A|b)\tau(B|b)}}} \\
&< \frac{e^{\mu n}I_0(2n\sqrt{\tau(A|b)\tau(B|b)}\sum_{k=1}^{\infty}(\sqrt{\frac{\tau(A|b)}{\tau(B|b)}}^{k} - \frac{\tau(B|a)}{\sqrt{\tau(A|b)\tau(B|b)}}^{k})}{e^{2n\sqrt{\tau(A|b)\tau(B|b)}}} \\
&\approx \frac{e^{\mu n}\sum_{k=1}^{\infty}(\sqrt{\frac{\tau(A|b)}{\tau(B|b)}}^{k} - \frac{\tau(B|a)}{\sqrt{\tau(A|b)\tau(B|b)}}^{k})}{2\sqrt{\pi n\sqrt{\tau(A|b)\tau(B|b)}}} \\
&= \frac{e^{\mu n}(\frac{\sqrt{\frac{\tau(A|b)}{\tau(B|b)}}}{1-\sqrt{\frac{\tau(A|b)}{\tau(B|b)}}} - \frac{\frac{\tau(B|a)}{\sqrt{\tau(A|b)\tau(B|b)}}}{1-\frac{\tau(B|a)}{\sqrt{\tau(A|b)\tau(B|b)}}})}{2\sqrt{\pi n\sqrt{\tau(A|b)\tau(B|b)}}}
\end{aligned}$$

so it suffices to show that

$$2+\sqrt{\frac{\tau(B|b)}{\tau(A|b)}}+\sqrt{\frac{\tau(B|a)}{\tau(A|a)}} > \frac{\sqrt{\frac{\tau(A|b)}{\tau(B|b)}}}{1-\sqrt{\frac{\tau(A|b)}{\tau(B|b)}}} - \frac{\frac{\tau(B|a)}{\sqrt{\tau(A|b)\tau(B|b)}}}{1-\frac{\tau(B|a)}{\sqrt{\tau(A|b)\tau(B|b)}}}$$

which is the hypothesis. □

Now, consider the specific conditions at equilibrium. Suppose that σ is an equilibrium. If the election is not close, then it must be that either

$$\mathbb{E}[U|a,\sigma]-\mathbb{E}[U|b,\sigma] > \frac{1}{2}(Pr(Piv_B|b,\sigma)+Pr(Piv_B|a,\sigma))$$

or

$$\mathbb{E}[U|b,\sigma]-\mathbb{E}[U|a,\sigma] > \frac{1}{2}(Pr(Piv_A|b,\sigma)+Pr(Piv_A|a,\sigma)).$$

By Bouton and Castanheira [2009] Lemma 1, restrict attention to profiles indexed by $\theta \in [0,1]$ defined by $\sigma_\theta(1)(\emptyset) = \theta$, $\sigma_\theta(1)(A) = 1-\theta$ and $\sigma_\theta(2)(B) = 1$. Let $\bar{a}$ be defined as in Lemma 1.

For n high enough, if σ_θ is an equilibrium then $\theta \in (0,1)$. Therefore, it must be the case that either

$$p_1(Pr(Piv_A|a,\sigma_\theta)) = (1-p_1)Pr(Piv_A|b,\sigma_\theta), \tag{C.5.11}$$

$$\frac{p_2}{1-p_2} < \frac{Pr(Piv_A|b,\sigma_\theta) + Pr(Piv_B|b,\sigma_\theta)}{Pr(Piv_A|a,\sigma_\theta) + Pr(Piv_B|a,\sigma_\theta)} < \frac{p_1}{1-p_1}, \tag{C.5.12}$$

and (C.5.3) all hold or

$$q_1(Pr(Piv_A|a,\sigma_\theta)) = (1-q_1)Pr(Piv_A|b,\sigma_\theta), \tag{C.5.13}$$

$$\frac{q_2}{1-q_2} < \frac{Pr(Piv_A|b,\sigma_\theta) + Pr(Piv_B|b,\sigma_\theta)}{Pr(Piv_A|a,\sigma_\theta) + Pr(Piv_B|a,\sigma_\theta)} < \frac{q_1}{1-q_1}, \tag{C.5.14}$$

and (C.5.4) all hold.

By Lemmas 35 and 36 above neither (C.5.4) nor (C.5.3) holds at $\sigma_{\bar{\theta}}$. The following inequalities hold given the signal structure, as long as θ is so that $\tau(A|a,\sigma_\theta) > \frac{1}{2}$ and $\tau(B|b,\sigma_\theta) > \frac{1}{2}$.

- $\frac{\partial \frac{Pr(Piv_A|b,\sigma_\theta)+Pr(Piv_B|b,\sigma_\theta)}{Pr(Piv_A|a,\sigma_\theta)+Pr(Piv_B|a,\sigma_\theta)}}{\partial\theta} < 0$
- $\frac{\partial \frac{Pr(Piv_A|b,\sigma_\theta)}{Pr(Piv_A|a,\sigma_\theta)}}{\partial\theta} < 0$
- $\frac{\partial}{\partial\theta}(\mathbb{E}[U|b,\sigma_\theta] - \frac{1}{2}Pr(Piv_A|b,\sigma_\theta)) > 0$
- $\frac{\partial}{\partial\theta}(\mathbb{E}[U|a,\sigma_\theta] + \frac{1}{2}Pr(Piv_A|a,\sigma_\theta)) < 0$
- $\frac{\partial}{\partial\theta}(\mathbb{E}[U|b,\sigma_\theta] + \frac{1}{2}Pr(Piv_B|b,\sigma_\theta)) > 0$
- $\frac{\partial}{\partial\theta}(\mathbb{E}[U|a,\sigma_\theta] - \frac{1}{2}Pr(Piv_B|a,\sigma_\theta)) < 0$

Suppose that equations (C.5.11), (C.5.12) and (C.5.4) all hold for some σ_θ. It is the case that

$$\frac{Pr(Piv_A|b,\sigma_{\bar{\theta}})}{Pr(Piv_A|a,\sigma_{\bar{\theta}})} > 1$$

for n large enough (using standard formulas for pivot probabilities). Since (C.5.11) holds and $\frac{p_1}{1-p_2} < 1$, it must be that $\theta > \bar{\theta}$ because $\frac{\partial \frac{Pr(Piv_A|B,\sigma_{\bar{\theta}})}{Pr(Piv_A|A,\sigma_{\bar{\theta}})}}{\partial \theta} < 0$. However, this implies that

$$\mathbb{E}[U|a,\sigma_\theta] - \frac{1}{2}Pr(Piv_A|a,\sigma_\theta) < \mathbb{E}[U|a,\sigma_{\bar{\theta}}] - \frac{1}{2}Pr(Piv_A|a,\sigma_{\bar{\theta}}))$$

and

$$\mathbb{E}[U|b,\sigma_\theta] + \frac{1}{2}Pr(Piv_B|b,\sigma_\theta) > \mathbb{E}[U|b,\sigma_{\bar{\theta}}] + \frac{1}{2}Pr(Piv_B|b,\sigma_{\bar{\theta}}).$$

Note therefore that

$$\mathbb{E}[U|b,\sigma_\theta] + \frac{1}{2}Pr(Piv_B|b,\sigma_\theta) > \mathbb{E}[U|a,\sigma_\theta] - \frac{1}{2}Pr(Piv_A|a,\sigma_\theta))$$

which means that (C.5.4) cannot hold.

Now, suppose that equations (C.5.13), (C.5.14) and (C.5.3) all hold for some σ_a. It can be verified that

$$\frac{Pr(Piv_A|b,\sigma_{\bar{\theta}}) + Pr(Piv_B|b,\sigma_{\bar{\theta}})}{Pr(Piv_A|a,\sigma_{\bar{\theta}}) + Pr(Piv_B|a,\sigma_{\bar{\theta}})} \leq 1$$

for n large enough using Myerson [2000] Equation 5.5, with equality holding only if $r(1|a) = r(2|b)$. Since (C.5.14) holds and $\frac{q_2}{1-q_2} > 1$, since

$$\frac{\partial \frac{Pr(Piv_A|B,\sigma_\theta)+Pr(Piv_B|B,\sigma_\theta)}{Pr(Piv_A|A,\sigma_\theta)+Pr(Piv_B|A,\sigma_\theta)}}{\partial \theta} < 0$$

it must be that $\theta < \bar{\theta}$ for n large enough. However, this implies that

$$\mathbb{E}[U|a,\sigma_\theta] + \frac{1}{2}Pr(Piv_A|a,\sigma_\theta)) > \mathbb{E}[U|a,\sigma_{\bar{\theta}}] + \frac{1}{2}Pr(Piv_A|a,\sigma_{\bar{\theta}}))$$

and

$$\mathbb{E}[U|b,\sigma_\theta] - \frac{1}{2}Pr(Piv_A|b,\sigma_\theta) < \mathbb{E}[U|b,\sigma_{\bar{\theta}}] - \frac{1}{2}Pr(Piv_A|b,\sigma_{\bar{\theta}}).$$

Note therefore that

$$\mathbb{E}[U|b,\sigma_\theta] - \frac{1}{2}Pr(Piv_A|b,\sigma_\theta) < \mathbb{E}[U|a,\sigma_\theta] + \frac{1}{2}Pr(Piv_A|a,\sigma_\theta))$$

which means that (C.5.3) cannot hold. Therefore, σ_θ is not an equilibrium, which is a contradiction and completes the proof. □

Proof of Proposition 11:

Proof. Note that $Pr(Piv_c|\omega, \sigma^*) = 1$ for all ω, c and $\mathbb{E}[U|a, \sigma^*] = \mathbb{E}[U|b, \sigma^*]$ since no one votes. From there, the logic in Proposition 11 shows that a fixed voter would prefer to randomize with equal probability between A and B rather than play any other strategy that mixes between voting for A and B. On the other hand, since the tie breaking rule is a coin-flip, she can induce the same distribution over outcomes by abstaining. Hence, she weakly prefers to abstain rather than flip a coin and is thus willing to play her strategy profile. □

References

Avery M. Abernethy and George R. Franke. The information content of advertising: A meta-analysis. *Journal of Advertising*, 25(2):pp. 1–17, 1996.

Milton Abramowitz and Irene A. Stegun. *Handbook of Mathematical Tables.* National Bureau of Standards, Applied Mathematics Series. Tenth edition, 1972. Original version 1964.

Charalambos Aliprantis and Kim Border. *Infinite Dimensional Analysis.* Springer, 3rd edition, 2006.

Bharat N. Anand and Ron Shachar. Advertising, the matchmaker. *The RAND Journal of Economics*, 42(2):205–245, 2011.

Scott Ashworth. Camapaign affects with ambiguity-averse voters. *mimeo*, 2005.

Robert J. Aumann. Utility theory without the completeness axiom. *Econometrica*, 30(3):445–462, 1962.

David Austen-Smith and J. Banks. Information aggregation, rationality and the Condorcet jury theorem. *American Political Science Review*, 90(1), 1996.

Sophie Bade. Electoral competition with uncertainty averse parties. *Games and Economic Behavior*, 72:12–29, 2011.

Kyle Bagwell. The economic analysis of advertising. In Mark Armstrong and Rob Porter, editors, *Handbook of Industrial Organization*, volume 3, pages 1701–1844. North-Holland, 2007.

Arpad Baricz. Bounds for modified Bessel functions of the first and second kinds. *Proceedings of the Edinburgh Mathematical Society*, 53(3), 2010.

Marcus Berliant and Hideo Konishi. Salience: Agenda choices by competing candidates. *Public Choice*, 125, 2005.

Sourav Bhattacharya. Preference monotonicity and information aggregation in elections. *mimeo*, 2008.

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

Laurent Bouton and Micael Castanheira. The Condorcet-Duverger trade-off: Swing voters and voting equilibria. In E. Aragones, C. Bevia, H. Llavador, and N. Schofield, editors, *The Political Economy of Democracy*. Fundacion BBVA, 2009.

Andrew Caplin and Daniel Martin. Framing effects and optimization. *Mimeo*, 2012.

Olivier Coibion and Yuriy Gorodnichenko. What can survey forecasts tell us about informational rigidities? *Journal of Political Economy*, Forthcoming, 2011a.

Olivier Coibion and Yuriy Gorodnichenko. Information rigidity and the expectations formation process: A simple framework and new facts. *mimeo*, 2011b.

Scott Condie and Jayant Ganguli. Ambiguity and rational expectations equilibria. *Review of Economic Studies*, Forthcoming, 2011.

Keith Cowling, John Cable, Michael Kelly, and Tony McGuinness. *Advertising and Economic Behavior*. Holmes and Meier Publishers, 1975.

Henrique de Olivera. Axiomatic foundations for rational inattention. *Mimeo*, 2012.

Stefano DellaVigna and Joshua Pollet. Investor inattention and friday earnings announcements. *Journal of Finance*, 64, 2009.

David Dillenberger, Juan Sebastian Lleras, Philipp Sadowski, and Norio Takeoka. A theory of subjective learning. *Mimeo*, 2012.

Avinash Dixit and Victor Norman. Advertising and welfare. *The Bell Journal of Economics*, 9(1):pp. 1–17, 1978.

James Dow. Search decisions with limited memory. *Review of Economic Studies*, 58 (1):1–14, 1991.

Andrew Ellis. Foundations for optimal inattention. *mimeo*, 2012.

Glenn Ellison and Sara Fisher Ellison. Strategic entry deterrence and the behavior of pharmaceutical incumbents prior to patent expiration. *American Economic Journal: Microeconomics*, 3(1):1–36, 2011.

Daniel Ellsberg. Risk ambiguity and the Savage axioms. *Quarterly Journal of Economics*, 75(4), 1961.

Larry Epstein and Martin Schneider. Ambiguity and asset markets. *Annual Review of Financial Economics*, 2, 2010.

Haluk Ergin and Todd Sarver. A unique costly contemplation representation. *Econometrica*, 78(4), 2010.

Timothy Feddersen and Wolfgang Pesendorfer. The swing voter's curse. *American Economic Review*, 86(3), 1996.

Timothy Feddersen and Wolfgang Pesendorfer. Voting behavior and information aggregation in elections with private information. *Econometrica*, 65(5), 1997.

Timothy Feddersen and Wolfgang Pesendorfer. Abstention in elections with asymmetric information and diverse preferences. *American Political Science Review*, 93 (2), 1999.

Marc Fischer and Sonke Albers. Patient- or physician-oriented marketing: What drives primary demand for prescription drugs? *Journal of Marketing Research*, 47 (1), 2010.

James W. Friedman. Advertising and oligopolistic equilibrium. *The Bell Journal of Economics*, 14(2):pp. 464–473, 1983.

Paolo Ghirardato and Jonathan Katz. Indecision theory: Quality of information and voting behavior. *Journal of Public Economic Theory*, 8, 2006.

Itzhak Gilboa and David Schmeidler. Maxmin expected utility with a unique set of priors. *Journal of Mathematical Economics*, 18(1), 1989.

Itzhak Gilboa, Fabio Maccheroni, Massimo Marinacci, and David Schmeidler. Objective and subjective rationality in a multiple prior model. *Econometrica*, 78(2), 2010.

Jean-Michel Grandmont. Continuity properties of von Neumann-Morgenstern utility. *Journal of Economic Theory*, 4(1):45–57, 1972.

Gene M. Grossman and Carl Shapiro. Informative advertising with differentiated products. *The Review of Economic Studies*, 51(1):pp. 63–81, 1984.

Faruk Gul and Wolfgang Pesendorfer. Random expected utility. *Econometrica*, 74 (1):pp. 121–146, 2006.

Faruk Gul, Wolfgang Pesendorfer, and Tomasz Strzalecki. Behavioral competitive equilibrium and extreme prices. *mimeo*, 2011.

Christian Hellwig and Laura Veldkamp. Knowing what others know: Coordination motives in information acquisition. *Review of Economic Studies*, 76(1):223–251, 2009.

Vrinda Kadiyali. Entry, its deterrence, and its accommodation: A study of the u. s. photographic film industry. *The RAND Journal of Economics*, 27(3):pp. 452–478, 1996.

Peter Klibanoff. Uncertainty, decision and normal form games. *mimeo*, 1996.

John E. Kwoka. The sales and competitive effects of styling and advertising practices in the u.s. auto industry. *The Review of Economics and Statistics*, 75(4), 1993.

Tracy R Lewis and David E M Sappington. Supplying information to facilitate price discrimination. *International Economic Review*, 35(2):309–27, 1994.

Kin Chung Lo. Equilibrium in beliefs under uncertainty. *Journal of Economic Theory*, 70(1), 1996.

Michael Luca. Reviews, reputation, and revenue: The case of yelp.com. *mimeo*, 2011.

Bartosz Mackowiak and Mirko Wiederholt. Optimal sticky prices under rational inattention. *American Economic Review*, 99:769–803, 2009.

Michael Mandler. The fragility of information aggregation in large elections. *Games and Economic Behavior*, http://dx.doi.org/10.1016/j.bbr.2011.03.031, 2011.

N. Gregory Mankiw and Ricardo Reis. Sticky information versus sticky prices: A proposal to replace the new keynesian phillips curve. *The Quarterly Journal of Economics*, 117(4):1295–1328, 2002.

Yusufcan Masatlioglu, Daisuke Nakajima, and Erkut Y. Ozbay. Revealed attention. *American Economic Review*, 105:2183–2205, 2012.

Filip Matejka and Alisdair McKay. Rational inattention to discrete choices: A new foundation for the multinomial logit model. *mimeo*, 2012.

Andrew McLennan. Consequences of the Condorcet jury theorem for beneficial infromation aggregation by rational agents. *American Political Science Review*, 92 (2), 1998.

Michael Meurer and Dale Stahl. Informative advertising and product match. *International Journal of Industrial Organization*, 12(1):1–19, March 1994.

Maximilian Mihm and M. Kemal Ozbek. Decision making with rational inattention. *Mimeo*, 2012.

Paul Milgrom and John Roberts. Price and advertising signals of product quality. *Journal of Political Economy*, 94(4):pp. 796–821, 1986.

Roger Myerson. Extended Poisson games. *Games and Economic Behavior*, 25(1), 1998.

Roger Myerson. Large Poisson games. *Journal of Economic Theory*, 94(1), 2000.

Paulo Natenzon. Random choice and learning. *Mimeo*, 2012.

Phillip Nelson. Advertising as information. *Journal of Political Economy*, 82(4):pp. 729–754, 1974.

Marc Nerlove and Frederick V. Waugh. Advertising without supply control: Some implications of a study of the advertising of oranges. *Journal of Farm Economics*, 43(4):pp. 813–837, 1961.

Harold E. Pashler. *The Psychology of Attention*. MIT Press, Cambridge, MA, USA, 1998.

Jeffrey M. Perloff and Steven C. Salop. Equilibrium with product differentiation. *The Review of Economic Studies*, 52(1):pp. 107–120, 1985.

Michele Piccione and Ariel Rubinstein. Modeling the economic interaction of agents with diverse abilities to recognize equilibrium patterns. *Journal of the European Economic Association*, 1(1):212–223, 2003.

Ariel Rubinstein. On price recognition and computational complexity in a monopolistic model. *Journal of Political Economy*, 101(3):473–484, 1993.

Steven C. Salop. Monopolistic competition with outside goods. *Bell Journal of Economics*, 10(1):141–156, Spring 1979.

Barry J. Seldon and Khosrow Doroodian. A simultaneous model of cigarette advertising: Effects on demand and industry response to public policy. *The Review of Economics and Statistics*, 71(4):pp. 673–677, 1989.

Lloyd S. Shapley and Manel Baucells. A theory of multiperson utility. *mimeo*, 1998.

Christopher A. Sims. Stickiness. In *Carnegie-Rochester Conference Series on Public Policy*, volume 49, pages 317–356. Elsevier, 1998.

Christopher A. Sims. Implications of rational inattention. *Journal of Monetary Economics*, 50:665–690, 2003.

Ran Spiegler. Competition over agents with boundedly rational expectations. *Theoretical Economics*, 1(2):207–231, June 2006.

Jean Tirole. *The Theory Of Industrial Organization*. MIT Press, 1988.

Carol Horton Tremblay and Victor J. Tremblay. Advertising, price, and welfare: Evidence from the u. s. brewing industry. *Southern Economic Journal*, 62(2):pp. 367–381, 1995.

Stijn Van Nieuweburgh and Laura Veldkamp. Information acquisition and underdiversification. *Review of Economic Studies*, 77(2):779–805, 2010.

Timothy Van Zandt. Hidden information acquisition and static choice. *Theory and Decision*, 40:235–247, 1996.

Jurgen Wit. Rational choice and the Condorcet jury theorem. *Games and Economic Behavior*, 22:364–376, 1998.

Michael Woodford. Inattention as a source of randomized discrete adjustment. *Mimeo*, 2008.

Michael Woodford. Inattentive valuation and reference-dependent choice. *Mimeo*, 2012.

Marta Wosinska. Direct-to-consumer advertising and drug therapy compliance. *Journal of Marketing Research*, 42(3):pp. 323–332, 2005.

Curriculum Vitae

[redacted]