NORTHWESTERN UNIVERSITY

Complexity in Economic Theory

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY

Field of Economics

By

Modibo Khane Camara

EVANSTON, ILLINOIS

June 2022

Abstract

This dissertation leverages methodologies from computer science to understand sources of complexity in economic theory. Chapter 1 considers time complexity: how much time is needed to make a decision. Chapters 2 and 3 consider informational complexity: how much and what kind of data is needed to make a decision. I argue that this line of work is critical for our understanding of economic behavior. The fact that economic models often assume agents are capable of very complex behaviors can lead to predictions and policy recommendations that are themselves quite complex. Through applications in decision theory and mechanism design, this work aims to help distinguish predictions and recommendations that are unrealistically complex from those that are at least plausible.

Acknowledgements

This dissertation would not have been possible without the guidance and support of my committee. This includes my committee chairs, Eddie Dekel and Jason Hartline, as well as committee members, Marciano Siniscalchi and Jeffrey Ely. I am deeply grateful to them for their patience and open-mindedness as they taught me to transform vague intuitions into concrete theorems, and broad pronouncements into well-substantiated arguments.

My committee has been unfailingly generous with both their time and their advocacy, even when my research interests drifted outside of the typical topics of inquiry. Their breadth of knowledge and sharp insights have proven critical to the substance of my research as well as my ability to communicate it effectively. Jason has exemplified extroverted, forward-thinking research that is not limited by disciplinary boundaries. He has been my conduit to the broader community of theoretical computer science, but also shaped my vision for economics and made me incredibly optimistic about its future. Eddie has taught me how to think rigorously and clearly, and dramatically reshaped the way that I communicate ideas. He has consistently pushed me to improve and refine my work, and has shown a remarkable willingness to engage deeply from the earliest stages of a research idea to the final stages of polishing a draft. Marciano has shown me, through his extraordinary enthusiasm for research, that there is a path for unconventional approaches in the field of economics. His optimism has made it possible for me to pursue risky projects, while his willingness to engage deeply has allowed me to execute those projects effectively. Finally, I benefited from Jeff's unreasonably good judgement in both the substance of research and in its presentation. In particular, I pursued Chapter 1 of this dissertation in part because of his resounding encouragement even in its very early stages.

I am also grateful to many of the faculty members and fellow students at Northwestern, with whom I have had the pleasure to spend the past six years, as well as my friends at other universities that are or have undergone the same process. The econometrics community at Northwestern, including Joel Horowitz and especially Chuck Manski, shaped my perspective on research tremendously. Beyond his perspectives on credibility and ‘unlearning’, Chuck introduced me to some key ideas that led to Chapter 1 of this dissertation. Brendon Andrews is a kindred spirit and dear friend whose intellectual integrity has made me hopeful about academia. In addition to Brendon, Martin Eftimoski, Michael Gmeiner, and Justin Gardiner made it possible for me to survive the first year of my PhD. Aleck Johnsen has been a dear friend and co-author who always brings good humor and wisdom into the research process. Xiaoyu Cheng and Lorenzo Stanca, two of the three caballeros, have been superb intellectual partners that helped me refine my own understanding of decision theory.

My parents, Elisabeth and Modibo, instilled in me a value for education and independence of thought, inspiring me to pursue a PhD. They have provided unconditional love and support, which I have often leaned on in the ten years since I left for college. My brother, Lamine, has continued to inspire me with his energy and unbridled curiosity as he pursues his own PhD. My partner, Samantha, has brought a special joy into the final years of my PhD. I cannot imagine going through the job market without her support and encouragement, as well as her perspective as a fellow intellectual. I love them all dearly.

Preface

This dissertation leverages methodologies from computer science to understand sources of complexity in economic theory. Chapter 1 considers time complexity: how much time is needed to make a decision. Chapters 2 and 3 consider informational complexity: how much and what kind of data is needed to make a decision. I consider the implications of these different sources of complexity for different economic settings. Chapter 1 focuses on applications in decision theory and behavioral economics, whereas Chapters 2 and 3 focus on applications in mechanism design.

In Chapter 1 on "Computationally Tractable Choice", I develop theoretical foundations for behavioral heuristics by bringing powerful models of computation to decision theory. Specifically, I propose a new kind of axiom – computational tractability – and apply it to a model of choice under risk. This axiom is quite weak: it only rules out behaviors that are thought to be implausible for *any* algorithm to exhibit in a reasonable amount of time. If a decisionmaker could make intractable choices, we could convert those choices into efficient solutions to problems of significant importance to science and industry, for which no efficient solutions are known.

First, I use this framework to show that, under standard rationality assumptions, computational constraints necessarily lead to forms of choice bracketing (a common behavioral heuristic). If a decisionmaker's choices are rational (i.e. maximize expected utility) and tractable, I show that her choices are observationally equivalent to forms of choice bracketing. Equivalently, I show that expected utility maximization is intractable unless the utility function satisfies a strong separability property. This demonstrates that even mild computational constraints can substantially sharpen our predictions about the decisionmaker's behavior relative to rationality alone.

Second, I use these results to give a formal justification for behavior that violates the expected utility axioms. Consider a decisionmaker who wants to maximize the expected value of an objective function that happens to be intractable. For many such objective functions, I show that a computationally-constrained decisionmaker *cannot* simultaneously (i) guarantee any non-zero fraction of her optimal payoff and (ii) have revealed preferences that satisfy the expected utility axioms. Then I show that the decisionmaker *can* guarantee a reasonable payoff, but only by using heuristics that an outside observer would not recognize as rational.

The remaining chapters lay the foundation for a theory of data-driven mechanism design. Both consider incomplete information games between a policymaker who sets a policy, an agent who responds, and a hidden state of nature. In line with the longstanding Wilson critique, these papers take the position that neither policymakers nor agents are likely to have a perfect understanding of the world. But, in contrast to more conservative approaches in robust mechanism design, they do not insist on mechanisms that are entirely detail-free. Instead, they observe that any available data can be used by the policymaker to (i) learn about the world and (ii) make more credible assumptions about the agent's beliefs. Formalizing this requires new ideas from statistical learning.

In Chapter 2 on "Mechanisms for a No-Regret Agent" (with Jason Hartline and Aleck Johnsen), we replace common knowledge with common history: data accumulated over time through repeated interaction between the policymaker and agent. The state is revealed after every period, but we make no other assumptions on the state-generating process. We develop calibrated policies that adapt to historical data over time, assuming the agent does the same, even if the data is highly non-stationary. This requires new behavioral assumptions that build on prior work on learning in games and capture ideas like "rationality" and "unpredictability" in a fully ex post sense.

In Chapter 3 on "Mechanism Design with a Common Dataset", I replace common knowledge with a common dataset: a random sample of states. The high-level idea is straightforward: if the data convincingly demonstrates some fact about the world, the agent should believe that fact. To formalize this, I rely on tools from statistical learning theory that characterize the complexity of learning an optimal response to a given policy. I propose a penalized policy that performs well under weak assumptions on how the agent learns from data, where policies that are too complex are penalized if they lead to unpredictable responses by the agent. This leads to new insights in models of vaccine distribution, prescription drug approval, performance pay, and product bundling.

To Eva, Isidro, and Oumou.

Contents

Chapter 1

Computationally Tractable Choice

1.1 Introduction

Any decisionmaker has a limited amount of time to make decisions, whether that means seconds or a lifetime. Yet, making good decisions can be time-intensive. This paper explores the implications of these two facts by integrating computational constraints into decision theory.

For context, observe that making good decisions can be especially time-intensive when considering many related decisions at once. For example, consider a consumer choosing from hundreds of products in a grocery store or an investor trading in dozens of assets. To ensure that they make decisions in a reasonable amount of time, people tend to narrowly frame their choices; they rely on heuristics like choice bracketing or mental accounting to break down complicated decisions into many simpler ones (see e.g. Tversky and Kahneman 1981; Rabin and Weizsäcker 2009). This can have a meaningful economic impact (see e.g. Choi et al. 2009; Hastings and Shapiro 2018).

To better understand these heuristics – and the broader implications of computational constraints for behavior – I propose a model of computationally tractable choice. Specifically, I impose an axiom of *computational tractability* in a model of choice under risk. This axiom is quite weak: it only rules out behaviors that are thought to be implausible for *any* algorithm to exhibit in a reasonable amount of time. If a decisionmaker could make intractable choices, we could convert those choices into efficient solutions to problems of significant importance to science and industry. Despite great effort, there are no known efficient solutions to these problems.

I use this framework of computationally tractable choice to obtain two kinds of results. First, I show that, under standard rationality assumptions, computational constraints necessarily lead to forms of choice bracketing. If a decisionmaker's choices are rational (i.e. maximize expected utility) and tractable, I show that her choices are observationally equivalent to forms of choice bracketing. Equivalently, I show that expected utility maximization is intractable unless the utility function satisfies a strong separability property. This demonstrates that even mild computational constraints, like tractability, can substantially sharpen our predictions about the decisionmaker's behavior relative to rationality alone.

Second, I use these results to give a formal justification for behavior that violates the expected utility axioms. Suppose a decisionmaker wants to maximize the expected value of a given objective function. If her objective function is not separable, my earlier results imply that exact optimization is intractable. What are the implications for her behavior? For many objective functions, I show that a computationally-constrained decisionmaker *cannot* simultaneously (i) guarantee any non-zero fraction of her optimal payoff and (ii) have revealed preferences that satisfy the expected utility axioms. The decisionmaker *can* guarantee a reasonable payoff, but only by using heuristics that an outside observer would not recognize as rational.

I now discuss the model and results in more detail.

Model. I consider a model of choice under risk. A decisionmaker cares about high-dimensional random vectors $X = (X_1, \ldots, X_n)$. A choice correspondence maps a menu of feasible options to the decisionmaker's choices X from that menu. This correspondence must be defined over *at least* all binary menus, as well as *product menus* in which it is feasible (but not necessarily optimal) to choose X_i independently of X_j. I call choices *rational* if they maximize expected utility for some continuous utility function (von Neumann and O. Morgenstern 1944).

I refer to two running examples: one represents utility functions that satisfy a symmetry property; another represents utility functions that do not. This distinction will soon be useful. In the first example, an investor has preferences over income X_i from assets $i = 1, \ldots, n$. In the second example, a consumer has preferences over consumption bundles, where X_i represents the quantity consumed of good i. In general, the decisionmaker's choices are *symmetric* if she is indifferent between the vectors (X_i, X_j) and (X_j, X_i). Symmetry may be plausible for investors: income from one asset i is interchangeable with income from another asset j. It is not plausible for consumers: commodities are not usually interchangeable.

Next, I introduce computation. The decisionmaker's choices are generated by a Turing machine, a powerful model of computation used in computational complexity theory to study what algorithms can and cannot do. Given an appropriate description of a menu, the Turing machine outputs a choice from that menu within a certain amount of time. This represents the state of the art: up to variations, the Turing machine is the most powerful model of computation to date. The Church-Turing thesis captures the sense in which the Turing machine is thought to be universal.[1]

A choice correspondence is *tractable* if it can be generated by a Turing machine within a reasonable amount of time. I use a definition of "reasonable" that has proven itself useful in computer science. The decisionmaker is allowed to take longer when facing a menu that is more complicated to describe, but the time taken must grow at most polynomially in the length of the description.This definition is used in computational complexity theory to

[1] Strictly speaking, it is not necessary for choices to be generated by a Turing machine for the results in this paper to hold. All that is necessary is that people with access to a computer are unable to efficiently solve problems that are thought to be fundamentally hard. On the other hand, suppose some person *can* make choices that I label intractable. Then, using the results in this paper, we could leverage that person's choices to efficiently solve problems that are thought to be fundamentally hard. If anything, that would *increase* the significance of this paper.

distinguish problems that computers can plausibly solve from ones that they cannot. I pair this definition with computational hardness conjectures, like P $\neq$ NP. Conjectures like these are commonly used in computational complexity theory to argue that computational problems are intractable.

Next, I consider what this framework can tell us about behavior.

Narrow Choice Bracketing. First, I ask: what are the implications of rationality, tractability, and symmetry for behavior? This case is a useful warmup for the next result, which drops the symmetry assumption. It is also important in its own right, as the investor example illustrates.

The answer to my question is "narrow choice bracketing". A decisionmaker *narrowly choice brackets* if her choice X_i does not depend on her choice X_j. This procedure is well-defined (but not necessarily optimal) on product menus. Theorem 1 shows that rational, tractable, and symmetric choice correspondences are observationally equivalent to narrow choice bracketing. Figure 1.1 illustrates. Equivalently, this result shows that expected utility maximization is intractable unless the utility function is *additively separable*, i.e. $u(x) = u_1(x_1) + \ldots + u_n(x_n)$.

It is worth emphasizing that this result is quite strong, despite the fact that (as I argued earlier) the tractability assumption is remarkably weak. Without tractability, the agent is restricted to any continuous and symmetric utility function. Additive separability is *much* stronger than that. For example, if the investor cares about total income $X_1 + \ldots + X_n$, additive separability implies risk neutrality.[2] The strength of this result illustrates two things: (i) tractability can significantly sharpen our predictions about behavior, and (ii) rationality appears to be a strong assumption, in the presence of computational constraints that are

[2] Keep in mind that this is a model of choice under risk. The von Neumann-Morgenstern utility function is cardinal, not ordinal. Monotone transformations of additively separable utility functions are generally not additively separable.

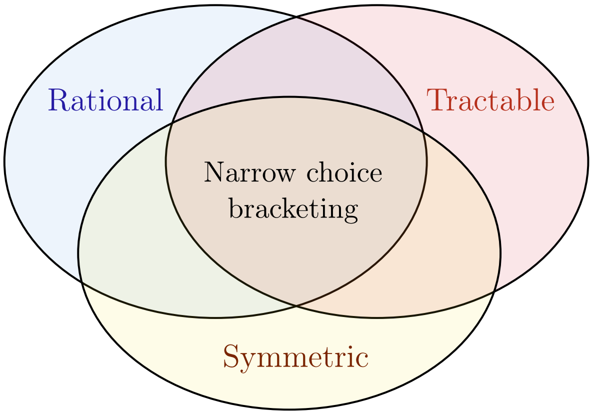


Figure 1.1: This diagram depicts the space of choice correspondences. The blue region consists of rational choice correspondences, the red region consists of tractable choice correspondences, and the yellow region consists of symmetric choice correspondences. The intersection of these three regions corresponds to narrow choice bracketing.

often missing from our models but likely to bind in practice. I will substantiate the second point after completing the analysis of rational and tractable choice.

Next, I generalize Theorem 1 by dropping the symmetry assumption.

Dynamic Choice Bracketing. I generalize choice bracketing to *dynamic choice bracketing*. In the spirit of dynamic programming, the decisionmaker considers choices X_i sequentially. Her choice of X_i only depends on a small number of choices X_j that she has not yet considered, even if it depends on more choices X_k that she has already considered. These heuristics preserve the computational advantages of choice bracketing but allows for richer patterns of behavior.

I illustrate dynamic choice bracketing in a simple example. I cannot use the investor example, because investor choices are likely to be symmetric, in which case dynamic choice bracketing is equivalent to narrow choice bracketing.Instead, consider the consumer example. Let the consumer have preferences over (i) location, (ii) sunscreen, and (iii) winter coats. Even if her preferences over sunscreen and coats are separable, the availability of sunscreen may influence her need for a coat by affecting where she wants to live. She can dynamically

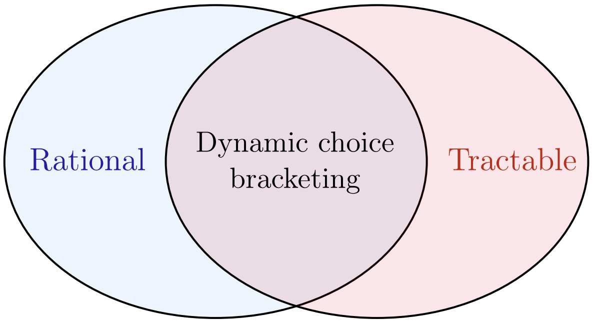


Figure 1.2: This diagram depicts the space of choice correspondences. The blue region consists of rational choice correspondences and the red region consists of tractable choice correspondences. The intersection of these two regions corresponds to dynamic choice bracketing.

bracket her choices by making a consumption plan conditional on her location, consistent with narrow choice bracketing, and only then deciding where to live.

Theorems 2 and 3 show that any rational and tractable choice correspondence is observationally equivalent to dynamic choice bracketing.[3] Figure 1.2 illustrates. As in Theorem 1, it is useful to restate this result in terms of a separability property. Theorem 2 shows that expected utility maximization is intractable unless the utility function is *Hadwiger separable*. This property is a novel relaxation of additive separability that allows for some complementarity and substitutibility across decisions, but limits their frequency. Like additive separability, this is a strong restriction relative to rationality, which only requires that the utility function be continuous.

Together, Theorems 1, 2, and 3 form the first main contribution of this paper. They articulate the implications of computational constraints for behavior under standard rationality assumptions. In doing so, they demonstrate that a behavioral heuristic – choice bracketing – is not only consistent with but *predicted by* an essentially standard model of choice with mild computational constraints.

Next, I turn to the second main contribution, which builds on these results.

[3] More precisely, it corresponds to *relatively narrow* dynamic choice bracketing. Dynamic choice bracketing can be broad or narrow, just as choice bracketing can be broad or narrow. This will be clear later in the paper.

Choice Trilemma. Having explored the implications of tractability for rational choice, I can now revisit a normative question: should a decisionmaker satisfy the expected utility axioms?

To formalize this, suppose a decisionmaker intrinsically wants to maximize the expected value of a given objective function. If her objective happens to be Hadwiger separable, Theorem 3 implies that her optimal choices are tractable. If not, Theorem 2 implies that optimization is intractable. In that case, the decisionmaker could still make choices that appear rational to an outside observer, insofar as they can be rationalized by preferences that satisfy the expected utility axioms. However, her revealed preferences would not match her true preferences. Alternatively, she could turn to *approximation algorithms* that guarantee her a positive fraction of her optimal payoff. However, her choices may not appear rational to an outside observer.

Theorem 4 shows that, for many objective functions, no rational and tractable choice correspondence can guarantee a non-zero fraction of the decisionmaker's optimal payoff. However, there do exist tractable approximation algorithms that can guarantee a meaningful fraction of her optimal payoff. These approximation algorithms violate the expected utility axioms: they do not maximize expected utility for *any* continuous utility function. They range from greedy algorithms that violate the continuity axiom to randomized algorithms that involve stochastic choice.

Altogether, my results imply an impossibility result, or *choice trilemma*, that relates three properties of choice. Figure 1.3 illustrates.[4] For many objectives, a choice correspondence can be tractable and approximately optimal, in that it guarantees a meaningful fraction of the optimal payoff (e.g. one-half or two-thirds). It can be tractable and rational if the agent is willing to dynamically choice bracket, in which case her revealed preferences are

[4] This figure was inspired by a similar figure in Akbarpour and Li (2020). The same is true for the term "trilemma".

Hadwiger separable. But it cannot satisfy all three properties at once. Given tractability, choice correspondences that perform well according to the decisionmaker's true objective cannot be rationalized by preferences that satisfy the expected utility axioms.

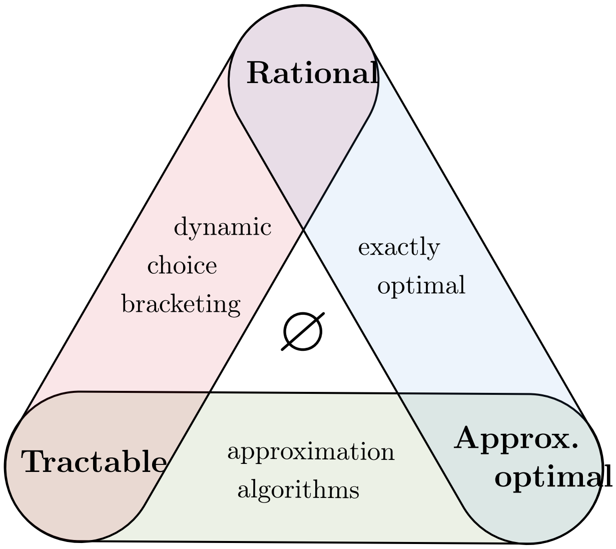


Figure 1.3: This diagram depicts the choice trilemma. The blue region connecting rationality and approximate optimality includes the traditional assumption of exact optimization. The green region connecting tractability and approximate optimality corresponds to approximation algorithms studied in computer science. The red region connecting rationality and tractability corresponds to dynamic choice bracketing. The ∅ symbol says that the intersection of all three regions is empty.

Related Literature. This paper contributes to three research efforts within economics. I briefly discuss the related literature now, and provide a more detailed discussion in section 1.6.

First, I contribute to the literature on bounded rationality. Previous work has introduced computational models of behavior in specific economic settings, such as repeated games, learning, and contracting (see e.g. Rubinstein 1986, A. Wilson 2014, Jakobsen 2020, respectively). Many papers rely on specialized models of computation, like finite automata, that rule out behaviors that anyone with access to a computer should be capable of. In contrast, I apply a very general model of computation to a very general model of choice, and still manage to obtain strong results.

The most related paper on bounded rationality is Echenique et al. (2011). Their revealed preference approach to computational complexity anticipates the tractability axiom in this paper. In a model of consumer choice, they show that, if a finite dataset of choices can be rationalized at all, then it can be rationalized by tractable preferences. In contrast, I consider a model of choice under risk and find that tractability rules out preferences that are not Hadwiger separable.

Second, I contribute to the subfield of economics and computation, which uses models from computer science to gain insight into economic phenomena. Computational complexity theory has been applied to classic problems like mechanism design, Nash equilibrium, and learning (see e.g. Nisan and Ronen 2001, C. Daskalakis et al. 2009, Aragones et al. 2005, respectively). In the same spirit, I apply similar methods to another classic problem: choice under risk.

Third, I contribute to the literature on choice bracketing and related phenomena. There is a sizable experimental and non-experimental literature that finds empirical evidence of narrow framing, including choice bracketing, mental accounting, and myopic loss aversion. There is also a small but growing theoretical literature that includes an axiomatic foundation without computational constraints (Zhang 2021) and models of rational inattention (Kőszegi and Matějka 2020; Lian 2020).

Overview for Computer Scientists. The computational results in this paper appear to be new and may be of independent interest to theoretical computer scientists.

I prove dichotomy theorems, in the sense of Schaefer (1978). I consider a large class of computational problems that correspond to expected utility maximization on product menus. Let u-EUM denote expected utility maximization with symmetric utility function

u. Restricting attention to symmetric utility functions u, Theorem 1 shows that if u is not additively separable then u-EUM is NP-hard. Proposition 2 shows that if u is additively separable then u-EUM is tractable.

I can generalize to asymmetric utility functions if I model the decisionmaker as a Turing machine with advice, following the literature on non-uniform complexity. If the non-uniform exponential-time hypothesis holds, Theorem 2 shows that if u is not Hadwiger separable then u-EUM is not tractable. Theorem 3 shows that if u is Hadwiger separable then u-EUM is tractable.

Theorem 3 is especially noteworthy as a fixed-parameter tractability result. I propose a graph-theoretic measure of how separable a utility function is, based on the Hadwiger number. Holding that measure fixed, I show that a suitable dynamic programming algorithm can efficiently maximize expected utility. This appears to be distinct from existing graph-based dynamic programming algorithms that rely on stronger sparsity assumptions based on treewidth.

Finally, Theorem 4 establishes a gap like the revelation gap of Feng and Hartline (2018), except even larger. I identify objective functions u where there exist constant factor approximation algorithms to u-EUM. However, unless NP $\subset$ P/poly, there exists no constant factor approximation algorithm that satisfies a standard rationality property.

Organization. The paper is organized as follows. Section 1.2 introduces the model of choice under risk and specializes it to high-dimensional settings. Subsections 1.2.2 through 1.2.5 introduce the computational model of choice, along with the necessary background for readers new to computational complexity. Section 1.3 relates rationality, tractability, and symmetry to narrow choice bracketing, and additive separability. Section 1.4 relates rationality and tractability to dynamic choice bracketing, and Hadwiger separability. Section 1.5 establishes the choice trilemma. Section 1.6 surveys the related literature. Section 1.7

concludes. Omitted proofs can be found in Appendix **??**.

1.2 Model

I introduce a standard model of choice under risk and specialize it to focus on high-dimensional problems where a decisionmaker has many decisions to make. Then I formalize choice as a computational problem, introducing the necessary definitions and concepts along the way.

A decisionmaker is tasked with choosing a lottery X from a *finite* menu M of feasible lotteries. The lottery X is a random variable that takes on values in a compact space of outcomes $\mathcal{X}$. Formally, let $(\Omega, \mathcal{F}, P)$ be a probability space where the sample space $\Omega = [0,1]$ is the unit interval, $\mathcal{F}$ is the Borel σ-algebra, and P is the Lebesgue measure. A lottery X is a map from the sample space Ω to the outcome space $\mathcal{X}$. I restrict attention to lotteries that can be described using a finite number of characters, in the following sense.

Assumption 1. *I restrict attention to lotteries X whose support is finite and, for every outcome x in the support, $X^{-1}(x)$ is a finite union of intervals (open or closed).*

A choice correspondence c describes the agent's behavior. If the agent is presented with menu M, then $c(M)$ describes her choices from that menu. Formally, a collection of menus $\mathcal{M}$ describes the universe of possible menus an agent may be presented with. A choice correspondence c maps menus $M \in \mathcal{M}$ to lotteries $X \in M$, where $c(M) \subseteq M$ and $c(M) \neq \emptyset$ for every $M \in \mathcal{M}$. That is, the agent's choices $X \in c(M)$ must be belong to menu M, and the agent always chooses at least one lottery $X \in M$ from every menu $M \in \mathcal{M}$. If $c(M)$ contains two or more lotteries, this is interpreted as the agent being indifferent between those lotteries,

The collection $\mathcal{M}$ is interpreted as a collection of menus that the decisionmaker could *potentially* be faced with. This is a potential outcomes interpretation, where $c(M)$ is the decisionmaker's choice in the hypothetical where she is presented with menu M.

Definition 1. *A choice correspondence c is* rational *if there exists a continuous, cardinal utility function $u : \mathcal{X} \to \mathbb{R}$ such that*

$$\forall M \in \mathcal{M}, \quad c(M) = \arg\max_{X \in M} \mathrm{E}[u(X)]$$

This is the notion of rationality that was axiomatized by von Neumann and O. Morgenstern (1944) (see Mas-Collell et al. 1995, chapter 6 for a standard textbook treatment). As usual, this does not mean that the decisionmaker explicitly performs any calculations, or that the decisionmaker has an intrinsic objective function that she wants to maximize. It only says that the agent's behavior can be rationalized by preferences that satisfy the expected utility axioms. In that case, they can be represented *as if* they maximize expected utility for some continuous utility function u that is u is revealed from the agent's choices.

Many behavioral heuristics are rational under this definition, include satisficing, consideration sets, and choice bracketing (see Proposition 2). However, the revealed utility function u may appear odd or detached from the agent's economic incentives.[5] That is not a problem for this paper: the less restrictive the definition of rationality, the stronger my results.

Assumption 2. *The collection $\mathcal{M}$ includes all binary menus (i.e. those with at most two lotteries).*

This assumption ensures that a rational choice correspondence c uniquely identifies its revealed utility function u, up to affine transformation.

1.2.1 High-Dimensional Choice

I specialize this model of choice under risk to focus on high-dimensional choices. This is intended to capture settings in which a decisionmaker is tasked with making many different

[5] For example, the utility function u that rationalizes choice bracketing may not respect the fungibility of money. If the outcome $X \in \mathbb{R}^n$ is a vector of incomes from n assets, u may depend on more than total income $X_1 + \ldots + X_n$.

decisions. Many settings have this flavor. Consider a consumer that decides how much to purchase of many different goods, or an investor that decides how much to invest in many different assets.

Outcomes x are arbitrarily high-dimensional vectors. I restrict attention to rational-valued vectors, i.e. $x_i \in \mathbb{Q}$, because they can always be described by a finite number of characters. Formally, the set $\mathcal{X}^n$ of n-dimensional outcomes is

$$\mathcal{X}^n = \{x \in \mathbb{Q} \cap [0,1]^\infty \mid x_i = 0, \forall i > n\}$$

The outcome space $\mathcal{X}$ is the union of n-dimensional outcomes for all $n > \infty$. Formally,

$$\mathcal{X} = \bigcup_{n=1}^{\infty} \mathcal{X}^n$$

There is an implicit assumption being made here. Consider a consumer with preferences over bundles X. I assume that her preferences over goods $i < n$ do not depend on whether there are n or N goods available, if she consumes none of goods $n+1, \ldots, N$ either way (that is, if $x_i = 0$ for $i = n+1, \ldots, N$). In other words, if she prefers apples to oranges in a local corner store, she should not change her mind when purchasing those two items from a large grocery store.

A lottery over n-dimensional outcomes is effectively an n-dimensional random vector

$$X = (X_1, \ldots, X_n, 0, 0, \ldots)$$

where the *partial lotteries* $X_i : \Omega \to [0,1]$ are univariate random variables. These partial lotteries X_i may be correlated, since they are defined on the same sample space Ω.

A *partial menu* M_i is a finite set of partial lotteries X_i.

Definition 2. *A* product menu M *is the Cartesian product of* n *partial menus* M_i*, i.e.*

$$M = M_1 \times \ldots \times M_n \times \{0\} \times \{0\} \ldots$$

In a sense, product menus are the simplest kind of high-dimensional menu. The fact that it is possible to choose $X_i \in M_i$ independently of $X_j \in M_j$ means that a decisionmaker can narrowly frame her choices without violating feasibility constraints. For that reason, product menus are typically used in lab experiments that study choice bracketing (see e.g. Tversky and Kahneman (1981), Rabin and Weizsäcker (2009)).

Assumption 3. *The collection* $\mathcal{M}$ *of menus includes all product menus.*[6]

This assumption does *not* restrict the decisionmaker to product menus. The collection $\mathcal{M}$ must include binary menus and product menus (assumptions 2 and 3). But it can also include menus of other kinds, like menus with budget constraints. Enlarging the collection $\mathcal{M}$ can only shrink the set of utility functions u for which expected utility maximization is tractable.

1.2.2 Choice as Computation

From a computational perspective, a choice correspondence c describes a *computational problem.*[7] A menu is a particular *instance* of that problem. Choice is a process by which the decisionmaker takes in a description of the menu M and outputs a chosen lottery $X \in c(M)$.

I model the decisionmaker as a Turing machine TM whose choice correspondence c_{TM} reflects the output of TM. A Turing machine is an abstract model of computation that takes in a string of characters and outputs another string. As depicted in figure 1.2.2, a

[6] This can be weakened slightly. I only require the collection $\mathcal{M}$ to include all product menus M consisting of binary partial menus $M_i = \{X_i, X_i'\}$.

[7] In general, a choice correspondence may or may not be an optimization problem, but any rational choice correspondence is an optimization problem since it is equivalent to expected utility maximization.

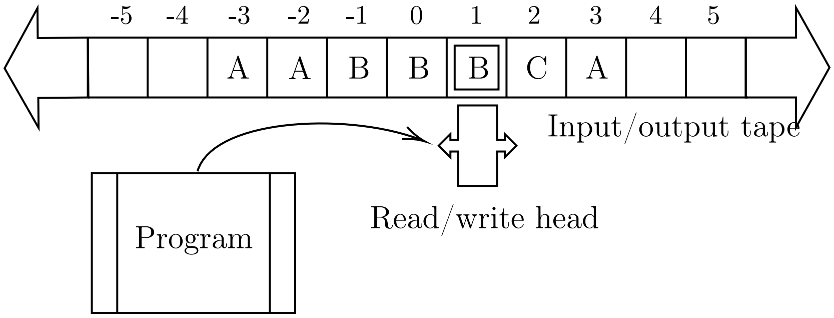


Figure 1.4: A depiction of a Turing machine, in the process of reading entry 1 on its tape.

Turing machine consists of a program, a read/write head, and an input/output tape. The tape is infinite and represents memory. The head can either modify a given entry of the tape, move to the next entry of the tape, or move to the previous entry of the tape. The program maintains a finite set of states and specifies a transition function. The transition function maps the current state and the symbol on the current entry of the tape to a new state and instructions for the head (shift left, shift right, or overwrite the current entry). The initial contents of the tape represent the input and the program ends when a terminal state is reached. The output is whatever is left on the tape.[8]

The Turing machine is a mathematically precise way to describe an algorithm, making it possible prove results about what algorithms can and cannot do. As such, the reader is welcome to think of the Turing machine as an algorithm written in their favorite all-purpose programming language, like Python or Java. This is typically how Turing machines are thought about in computer science. After all, most programming languages are Turing-complete, which means that they can simulate any Turing machine. Conversely, the Church-Turing thesis asserts that any physically-realizable computer can be simulated by a Turing machine.

In modeling the decisionmaker as a Turing machine, I rely on a hypothesis that the cog-

[8] For a more formal definition of the Turing machine, please refer to any textbook on computational complexity (e.g. S. Arora and Barak 2009, ch.1). Note that there are many variations on this model, but most are formally equivalent.

nitive process underlying human choice can be efficiently simulated with a Turing machine. The analogy between the human brain and computation is not new to this work. Researchers in computational neuroscience and elsewhere have long found value in taking an algorithmic perspective on the nervous system (see e.g. Papadimitriou et al. 2020). It is beyond the scope of this paper to evaluate whether that analogy is apt. However, it seems clear that computational constraints are binding on the human brain, and the Turing machine is the most compelling model of computation we have to date.

1.2.3 Representing Menus

Having modeled the decisionmaker as a Turing machine, I need to represent menus in a form that is legible to her. I describe a menu M with a string $s(M)$ of length $\ell(M)$, written in a standard alphabet. In principle, this could also be used to represent visual input from scanning a restaurant menu or a shelf on the grocery store, or audio input from hearing a list of options described.

The description $s(\cdot)$ is an essential primitive of this model. The same menu M described in two different ways may have different computational properties. The following example illustrates.

Example 1. This example conveys how the complexity of choice may depend on how the menu M is described. Suppose an investor is offered a share in a large holding company that consists of n subsidiaries. If she accepts, she receives payments X^A, where X_i^A denotes the share of profits from subsidiary i. If she rejects, she receives payments of 0.

The investor's choice is not especially complex if the holding company describes the earnings potential of each of its subsidiaries in a natural way. For example, for each subsidiary i, the holding company could describe each partial lottery X_i^A in order from $i = 1$ to $i = n$. It can describe partial lotteries as a list of claims like "in the event that $\omega \in [a, b]$, income is

$X_i(\omega) = x_i$."

The investor's choice is more complex if the holding company tries to obfuscate. For example, it could say "profits are high ($x_i = 1$) if a particular instance of the traveling salesman problem can be solved with a route of length k; otherwise profits are low ($x_i = 0$)." This would be sensible if, for example, the subsidiaries are trucking and shipping companies where the ability find quick routes that visit multiple locations will directly affect profitability. However, if the investor needs to solve the traveling salesman problem in order to decide whether to invest, she is unlikely to invest optimally, because that problem is thought to be fundamentally hard.[9]

I resolve this challenge by assuming, wherever possible, that menus are described in a simple and systematic way. This biases my results towards being more conservative. After all, it would be easy to argue that a choice correspondence is intractable if the menus are presented in complicated or obfuscatory ways. Instead, I show that a choice correspondences are intractable despite the fact that menus are presented in straightforward ways.

First, I specify the description $s(M)$ of binary menus M.

Assumption 4. *Let M be a binary menu.*

1. *Describe n-dimensional outcomes x as a list of values $x_1, \dots, x_n$ in decimal notation.*
2. *Describe partial lotteries X_i as a list of triples $[x_i, a, b]$ where $[a, b] \in \Omega$ is an interval of the sample space $\Omega = [0, 1]$ where $X_i(\omega) = x_i$.*[10]
3. *Describe n-dimensional lotteries X as an ordered list of partial lotteries $X_1, \dots, X_n$.*
4. *The description $s(M)$ is an ordered list of lotteries $X \in M$.*

[9] Thanks to Ehud Kalai for providing this example. Lipman (1999) studies a related problem where a decisionmaker does not know all of the logical implications of the information she is presented with.

[10] This list is always finite, due to assumption 1.

Next, I specify the description $s(M)$ of product menus M. This description is efficient since it takes advantage of the simple structure of product menus. In contrast, it would be very inefficient to describe a product menu M in the way that I describe binary menus: list every lottery $X \in M$.[11]

Assumption 5. *Let M be an n-dimensional product menu.*

1. *Describe partial lotteries as in assumption 4.*

2. *Describe partial menus M_i as a list of partial lotteries X_i.*

3. *The description $s(M)$ is an ordered list of partial menus $M_1, \ldots, M_n$.*

My results hold for any function $s(\cdot)$ that is consistent with these two assumptions. To be clear, if the collection $\mathcal{M}$ includes menus M' that are neither binary nor product menus, the set of tractable choice correspondences will generally depend on how $s(M')$ is defined. But my results only depend on how binary and product menus are described.

The description length $\ell(M)$ is bounded by a function of three parameters. Let lotteries $X \in M$ be n-dimensional. Let partial lotteries X_i be measurable with respect to m intervals $[a_i, b_i] \in \Omega$ in the sample space. Finally, let partial menus M_i consist of k lotteries. Then

$$\ell(M) = O(nmk)$$

In contrast, the size of a product menu M is $O(k^n)$. This difference is what makes high-dimensional optimization hard: product menus that can be described in only $O(n)$ characters require the agent to choose from as many as k^n lotteries

[11] For example, suppose that each partial menu $M_i = \{X_i^A, X_i^B\}$ is binary. The first entry in the list would be $X_1^A, X_2^A, X_3^A, \ldots$, the second entry would be $X_1^B, X_2^A, X_3^A, \ldots$, the third entry would be $X_1^A, X_2^B, X_3^A, \ldots$, and so on. This is incredibly inefficient. For example, there would be 2^{n-1} redundant descriptions of the partial lottery X_i^A.

1.2.4 Computationally Tractable Choice

A choice correspondence c is *computationally tractable* if there exists an algorithm that computes the agent's choice $c(L)$ from any given menu L within a reasonable amount of time.

Formally, the time it takes for the agent to make a choice $c_{\mathrm{TM}}(M)$ from menu M is the number of steps taken by TM before it arrives at its output. Let $\mathrm{runtime}_{\mathrm{TM}}(M)$ denote that number of steps. It is natural that an agent should take more time to make a decision on menus that have more lotteries or are otherwise more complicated. For this reason, time constraints restrict how quickly the runtime increases as the menu becomes more complicated.

Definition 3. *A time constraint T is a function $T : \mathbb{N} \to \mathbb{R}_+$ that maps a menu M's description length $\ell(M)$ to a maximum allowable runtime, $T(\ell(M))$.*

A Turing machine TM satisfies a time constraint T in a strong sense if

$$\mathrm{runtime}_{\mathrm{TM}}(M) \leq T(\ell(M)) \quad \forall M \in \mathcal{M} \tag{1.1}$$

In a moment, I will use this to define a strong axiom of computational tractability.

It is also possible to satisfy a time constraint T in a weaker sense. This reflects the notion that a decisionmaker may be adapted to a world in which menus M never exceed a certain description length $\ell(M)$. For example, one could hypothesize that the human brain has evolved over time to choose over bundles with $n \leq N$ goods, where N is the maximum number of goods that the consumer will ever encounter. As we will see in section 1.4, it can sometimes help to know N before commiting to an algorithm for making choices, irrespective of how large N is.

Formally, a Turing machine satisfies the time constraint in a weak sense if it requires additional input, called *advice*, to meet that constraint. An advice string A_j is associated with a menu M with description length $\ell(M) = j$. This could reflect the output of any pre-processing the decisionmaker does after learning the description length $\ell(M)$ but before learning the menu M. The Turing machine receives a menu-advice pair $\left(M, A_{\ell(M)}\right)$ as its initial input, and satisfies time constraint T if

$$\text{runtime}_{\text{TM}}\left(M, A_{\ell(M)}\right) \leq T(\ell(M)) \quad \forall M \in \mathcal{M} \tag{1.2}$$

I will use this to define a weak axiom of computational tractability.[12]

In order to define computational tractability, I need to specify a time constraint. This involves taking a stand on what constitutes "a reasonable amount of time." In doing so, I try to adhere to two guiding principles. First, I want to err on the side of being conservative. I prefer to label implausible behavior as tractable in order to avoid ruling out plausible behavior as intractable. Second, I want to defer whenever possible to the current state of the art in computer science.

Definition 4. *The choice correspondence c_{TM} is* strongly tractable *if the Turing machine* TM *satisfies* (1.1) *for some time constraint $T(k)$ that grows at most polynomially in k.*

Definition 5. *The choice correspondence c_{TM} is* weakly tractable *if the Turing machine* TM *satisfies* (1.2) *for some time constraint $T(k)$ and advice A_k that grow at most polynomially in k.*

The notion that "polynomial time" defines the boundary between tractable and intractable is common in computational complexity theory. This reflects a belief that any algorithm whose runtime is exponential in k will take an unreasonable amount of time

[12] In computational complexity theory, Turing machines with advice are studied in the literature on non-uniform time complexity. They are formally related to boolean circuits, another general model of computation used in computer science (S. Arora and Barak 2009, ch.6).

unless k is quite small. Clearly the converse is not true: an algorithm whose runtime is polynomial in k need not be quick. For example, an algorithm that requires $O(k^{100})$ steps runs in polynomial time but is unlikely to be feasible in practice. Furthermore, even $O(k)$ problems, like adding two numbers, can be challenging for human beings if k is large. In that sense, both definitions of tractability rule out only the very hardest problems.

It is also worth emphasizing that this is an asymptotic notion of computational tractability. The time constraint bounds the rate at which the runtime increases as the description length increases. There are good reasons for taking an asymptotic perspective. First, it does not force us to make a precise assessment of how quickly the decisionmaker can process information. From an asymptotic perspective, an intractable problem is intractable regardless of whether the decisionmaker is a child or an expert aided by a supercomputer. Second, it does not force us to specify exactly how complicated the decisionmaker's menu M is. Suppose M is an n-dimensional product menu, reflecting n individual decisions. How many decisions does a person face in her lifetime? Clearly n is large; even a consumer entering a grocery store faces hundreds if not thousands of products. Specifying exactly how large n is seems both hopeless and unnecessary.

Finally, I stress that computational tractability – like other axioms – is a restriction on the choice correspondence c rather than on the menu M or the choice $c(M)$. Tractability constraints how the decisionmaker's choices vary as the menu changes. This is very different from other constraints, like a budget constraint, which restrict the lotteries that the agent can choose from. One implication of this is that there is no unambiguous sense in which an agent can maximize expected utility "subject to" a time constraint.[13] The following example clarifies.

[13] The exception is if we know that the agent is running a particular search algorithm. If it hasn't stopped after T iterations, we can insist that it return the best option identified so far. It may be possible to formulate interesting models along these lines, but it would require going beyond computational constraints. We would need to hypothesize that decisionmakers use a *particular* algorithm to make choices.

Example 2. I claim that tractability has no implications for choice $c(M)$ in a fixed menu M. To see this, suppose that lottery $X \in M$ is optimal in M according to some objective. There exists a tractable choice correspondence c that chooses X from M. The algorithm simply memorizes the answer: if the input menu M' is M, output X, otherwise output the entire menu M'. This choice correspondence chooses optimally in M, but may not optimize in other menus.

This observation can be strengthened. Given a tractable choice correspondence c that fails to maximize expected utility on some finite collection $\mathcal{M}'$ of menus, it is always possible to create a new, tractable choice correspondence c' that outputs the optimal choice for menus $M \in \mathcal{M}'$ and outputs $c(M)$ for menus $M \notin \mathcal{M}'$. The algorithm for c' takes the algorithm for c and carves out an exception for every menu $M \in \mathcal{M}'$.

Clearly, these algorithms do not scale. But they underscore a key point: tractability axioms constrain how choices vary across the entire collection $\mathcal{M}$ of potential menus. They do not constraint choice within a given menu.

1.2.5 Computational Hardness Conjectures

Most results in computational complexity theory rely on computational hardness conjectures, and this paper is no exception.[14] The most well-known of these conjectures is P $\neq$ NP. In this subsection, I will state this conjecture, as well as two refinements.

These conjectures relate to an important class of computational problems that arise in mathematical logic. I introduce these problems not only because they are necessary to state the conjectures, but because they will come up again when proving results in sections 1.3 and 1.4.

[14] This is also true for many applications of computer science in economics. For example, the celebrated result that finding Nash equilibria is computationally-hard relies on the conjecture that PPAD $\neq$ FP (C. Daskalakis et al. 2009).

The satisfiability problem (SAT) asks whether a logical expression is possibly true, or necessarily false. To define it, I need to introduce a few objects. A *boolean variable* $v \in \{\text{true}, \text{false}\}$ can be either true or false. A *literal* is an assertion that v is true (v) or false ($\neg v$). A *clause* CL is a sequence of literals combined by "or" statements. For example,

$$CL = (v_1 \vee \neg v_2 \vee v_3)$$

A *boolean formula* BF in *conjunctive normal form* (CNF) is a sequence of clauses combined by "and" statements. For example,

$$BF = CL_1 \wedge CL_2$$

Finally, BF is *satisfiable* if there exists an *assignment* of values to $v_1, \ldots, v_n$ such that $BF = \text{true}$.

Definition 6. *The computational problem* SAT *asks whether a given formula is satisfiable.*

There are many variants of SAT. An especially important one is 3-SAT, which restricts attention to formulas where each clause has exactly three literals.

Definition 7. *The computational problem 3-SAT asks whether a given formula*

$$BF = CL_1 \wedge \ldots \wedge CL_m$$

is satisfiable, where each clause CL_j has exactly three literals.

The famous P $\neq$ NP conjecture has many equivalent formulations, including the following.

Conjecture 1 (P $\neq$ NP)**.** *There is no Turing machine that solves 3-SAT in polynomial time.*

Cook (1971) showed that a Turing machine that could solve 3-SAT in polynomial time could solve any problem in the complexity class NP in polynomial time. Roughly, NP consists of all computational problems whose solutions can be *verified* in polynomial time. In contrast, the class P consists of all problems whose solutions can be *obtained* in polynomial time. In other words, P $=$ NP would mean that if it's easy to verify a solution, then it is easy to obtain a solution.

Beginning with Karp (1972), computer scientists have shown that P $\neq$ NP is equivalent to the non-existence of a polynomial-time algorithm for hundreds of other notoriously hard problems. That is, if there exists a polynomial-time algorithm for *any* of these problems, then P $=$ NP. The fact that efficient algorithms have not been found for any of these problems, despite their scientific and industrial importance, has led to a widespread belief that P $\neq$ NP. For example, in a 2018 poll of theoretical computer scientists, 88% of respondents believed P $\neq$ NP (Gasarch 2019).

There are many refinements of P $\neq$ NP, two of which will be useful in this paper. These are stronger conjectures (they imply P $\neq$ NP), but they can be motivated in similar ways.

Conjecture 2 (NP $\not\subset$ P/poly)**.** *There is no Turing machine that solves 3-SAT in polynomial time with at most polynomial-size advice.*

Karp and Lipton (1980) showed that if this conjecture were false, it would imply a partial collapse of the so-called polynomial hierarchy (also see S. Arora and Barak (2009), section 6.4).

Conjecture 3 (Nonuniform Exponential Time Hypothesis, NU-ETH)**.** *There is no Turing machine that solves 3-SAT in subexponential time with at most polynomial-size advice.*

This is a refinement of the better-known exponential time hypothesis. Please note that there are different variants of the NU-ETH used in the literature.

I rely on these conjectures to prove my results. I use the weakest conjecture, P $\neq$ NP, to motivate narrow choice bracketing. I use the strongest conjecture, the NU-ETH, to motivate dynamic choice bracketing. I use the intermediate conjecture, NP $\not\subset$ P/poly, to establish the choice trilemma. These conjectures are sufficient but it is not clear whether the latter two are necessary.

1.3 Narrow Choice Bracketing

This section relates narrow choice bracketing to rational, strongly tractable, and symmetric choice correspondences. I begin with a formal definition of narrow choice bracketing and an informal explanation of the role it plays in reducing the computational complexity of choice.

First, let $c_i(M) \subseteq M_i$ denote the decisionmaker's partial choices from a product menu M.[15]

Definition 8. *A choice correspondence c is* narrowly bracketed *on product menus M if the partial choices $c_i(M)$ only depend on the partial menu M_i.*

This definition of narrow bracketing does not imply that the agent is optimizing in any sense. Typically, we associate narrow choice bracketing with a decisionmaker that is optimizing within each bracket. This corresponds to choice correspondences that are both rational and narrowly bracketed, and is indistinguishable from expected utility maximization with an additively separable utility function.

Definition 9. *A utility function u is* additively separable *if there exist univariate functions $u_i : [0, 1] \to \mathbb{R}$ such that, for any outcome $x \in \mathcal{X}$,*

$$u(x) = \sum_{i=1}^{\infty} u_i(x_i)$$

[15] Formally, if $X \in c(M)$ is a lottery chosen from menu M, then $X_i \in c_i(M)$.

Proposition 1. *A choice correspondence c is rational and narrowly bracketed if and only if it reveals a continuous and additively separable utility function.*

Narrow choice bracketing reduces the effective dimension of a high-dimensional optimization problem. To see why dimensionality drives computational hardness, consider *brute-force search*, a simple algorithm that optimizes within a menu M by searching over every lottery $X \in M$ and evaluating its expected utility $\mathrm{E}[u(X)]$. The number of lotteries $X \in M$ that need to be evaluated is k^n, where lotteries $X \in M$ are n-dimensional and partial menus M_i consist of k partial lotteries. If partial lotteries X_i are measurable with respect to the same m intervals in the sample space, the runtime is on the order of $O(mk^n)$. However, recall that the description length $\ell(M)$ of a product menu M is on the order of $O(nmk)$. Clearly, mk^n is not a polynomial function of nmk. Moreover, it is the dimension n, rather than quantities k or m, that is the key bottleneck.

A decisionmaker that narrowly choice brackets avoids this bottleneck, by transforming one n-dimensional optimization problem into n 1-dimensional optimization problems. Brute-force search on each partial menu M_i only needs to evaluate k partial lotteries. Since there are n partial menus, the total runtime is on the order of $O(nmk)$. This is polynomial in the description length.

Proposition 1 shows that narrow choice bracketing is without loss of optimality when the utility function u is additively separable. In that case, it is not necessary to evaluate every lottery $\ell \in M$ to be confident that one has made the optimal choice. Likewise, if the utility function u is increasing, then narrow choice bracketing is optimal on deterministic product menus. By deterministic, I mean that they consist of sure outcomes $x \in M$ rather than lotteries X. In that case, optimization is straightforward because there is no trade-off: simply choose the highest $x_i \in M_i$ in each partial menu. Of course, this is true only because M is a product menu.

However, narrow choice bracketing is suboptimal in general. The following example illustrates.

Example 3. A decisionmaker cares about bundles of fruit, where x_i denotes the quantity consumed of fruit i. She faces partial menus with two partial lotteries each. Their outcomes depend on a coin that can turn up heads ($\omega \leq 0.5$) or tails ($\omega > 0.5$) with equal probability. For each fruit i, the decisionmaker can choose between a partial lottery X_i^H that returns one unit of fruit i if the coin turns up heads, and X_i^T that returns one unit of fruit i if the coin turns up tails. Formally,

$$X_i^H(\omega) = \begin{cases} 1 & \omega \leq 0.5 \\ 0 & \omega > 0.5 \end{cases} \qquad X_i^T(\omega) = \begin{cases} 0 & \omega \leq 0.5 \\ 1 & \omega > 0.5 \end{cases}$$

Suppose the decisionmaker can only consume one fruit before it spoils. She is indifferent between fruits, so her utility function is

$$u(x) = \max_i x_i$$

It is optimal to hedge, by choosing partial lotteries that are negatively correlated. If $n = 2$, this can be achieved by choosing (X_1^H, X_2^T) or (X_1^T, X_2^H). This guarantees the decisionmaker a payoff of 1, whereas any other feasible lottery give the decisionmaker a payoff of 0.5.

However, a decisionmaker that narrowly brackets her choices will evaluate partial lotteries only by their marginal distributions.[16] For each fruit i, she will be indifferent between X_i^H and X_i^T. Her choices $c(M)$ include lotteries that obtain only half the optimal payoff.[17]

In section 1.5, I show a much stronger result: even if we allow for dynamic choice bracketing, we can always find a product menu where the decisionmaker strictly prefers a lottery

[16] Indeed, narrow choice bracketing is closely related to correlation neglect (see e.g. Zhang 2021). However, narrow choice bracketing may be suboptimal even if partial lotteries are independent (see e.g. Rabin and Weizsäcker 2009).

[17] Of course, one could easily perturb the lotteries to break indifference in favor of the suboptimal lotteries.

that obtains a negligible fraction of the optimal payoff. This holds for a much larger class of utility functions.

1.3.1 Representation Theorem

My first theorem shows that brute-force search, although naive and impractical, is effectively the best we can do unless u is additively separable. That is, there is no clever way to avoid the bottleneck associated with the dimension n, unless P = NP.

To state my theorem, I need to define two more properties: symmetry, and efficient computability of the utility function. Symmetry is an assumption of theorem 1, whereas efficient computability is an implication.

Symmetry says that relabeling coordinates does not affect choice. More formally, for any n and permutation $k_1, \ldots, k_n$ of $1, \ldots, n$, an outcome x' is a *permutation* of lottery x if

$$x' = (x_{k_1}, \ldots, x_{k_n}, x_{n+1}, \ldots)$$

A menu M' is a *permutation* of menu M if

$$M' = \{(X_{k_1}, \ldots, X_{k_n}, X_{n+1}, \ldots) \mid X \in M\}$$

Definition 10. *A choice correspondence c is* symmetric *if $c(M) = c(M')$ for any permutation M' of M. A utility function u is* symmetric *if $u(x) = u(x')$ for any permutation x' of x.*

For example, symmetry is plausible when each coordinate x_i of the outcome corresponds to income from some source i. If the decisionmaker only cares about a function of total income, i.e.

$$u(x) = f\left(x_1 + x_2 + \ldots\right)$$

then her utility function satisfies symmetry.

Next, a utility function u is efficiently computable if there exists a reasonably quick algorithm that computes $u(x)$ with at most ϵ imprecision. The caveat is that utility functions are only identified up to affine transformations, so at best we can compute a normalized utility function. Given utility function u over n-dimensional outcomes x, the normalized utility function u^n is:

$$u^n(x) = \frac{u(x) - \min_x u\left(x_1, \ldots, x_n, 0, 0, \ldots\right)}{\max_x u\left(x_1, \ldots, x_n, 0, 0 \ldots, \right) - \min_x u\left(x_1, \ldots, x_n, 0, 0, \ldots\right)}$$

where the vector $1, \ldots, 1$ is of length n. For any n-dimensional menu M, this is observationally equivalent to u because cardinal utility functions are only defined up to affine transformations. Effectively, this renormalizes the utility function separately for each n.

Definition 11. *A utility function u is* efficiently computable *if there exists a Turing machine that takes in a constant $\epsilon \in [0,1]$ and n-dimensional outcome $x \in \mathcal{X}$, and then outputs a real number y such that the normalized utility function u^n satisfies*

$$y - \epsilon \leq u^n(x) \leq y + \epsilon$$

with runtime $O(\text{poly}(n, 1/\epsilon))$.

I stress that efficient computability of the utility function is much weaker than tractability of the choice correspondence, and unrelated to separability. It is essentially a regularity condition: typical utility functions will satisfy it, irrespective of whether expected utility maximization is tractable. Intuitively, being able to calculate utility for a given outcome is very different from being able to choose the best lottery amongst a large set of lotteries.

Theorem 1 gives the main direction of my representation theorem. It associates rational, tractable, and symmetric choice correspondences with additively separable utility functions.

As I showed in proposition 1, this is indistinguishable from narrow choice bracketing.

Theorem 1. *Let choice correspondence c be rational, strongly tractable, and symmetric. If $P \neq NP$ then c reveals an additively separable, symmetric, and efficiently computable utility function.*

This theorem accomplishes two things. First, it provides a foundation for narrow choice bracketing as observed in lab experiments (symmetry is plausible in experiments where outcomes are monetary). Second, it provides a very strong restriction on the utility function based on relatively weak assumptions. Consider again the decisionmaker who only cares about total income, i.e.

$$u(x) = f\left(x_1 + x_2 + \ldots\right)$$

Theorem 1 suggests that expected utility maximization is tractable only if f is linear. That is, either the decisionmaker fails to maximize expected utility, or she is risk-neutral. Risk neutrality is often seen as a strong assumption, but in this case it is a straightforward implication of theorem 1. In fact, it would take special justification to argue that this decisionmaker is *not* risk-neutral, and yet somehow still capable of maximizing expected utility.[18]

Next, I provide a partial converse: when the utility function is additively separable, expected utility maximization is tractable on product menus.

[18] For example, one justification could be that expected utility maximization is tractable within a restricted collection of menus $\mathcal{M}'$, and only those menus are relevant to a given model. Verifying that expected utility maximization is tractable within a proposed model strikes me as a good practice for authors, in the same spirit as robustness checks.

With that said, this justification is not bullet-proof from a normative perspective. If we know that the decisionmaker is failing to optimize *somewhere* then why is it safe to assume that the decisionmaker is optimizing *anywhere*? Recall example 2. It is always possible to specialize an algorithm so that it optimizes in a particular menu or finite set of menus. But the cost of this is a slower runtime. Maximizing expected utility within collection $\mathcal{M}'$ may involve trading quick and optimal choices in menus $M' \in \mathcal{M}'$ for slower and potentially suboptimal choices in some menu $M \in \mathcal{M}$. How does one argue that menu M' should be prioritized over menu M?

Proposition 2. *Let the utility function u be additively separable and efficiently computable. Then expected utility maximization is strongly tractable on the collection of product menus.*

Proposition 2 follows from the fact that narrow choice bracketing is without loss of optimality for additively separable utility functions, and can be implemented in polynomial time. This argument does not generalize because narrow choice bracketing is only defined on product menus.[19]

1.3.2 Proof Outline of Theorem 1

I now outline the proof of Theorem 1, which relies on two key lemmas and two minor ones. In the next subsection, I illustrate how the key lemmas are proven in two special cases.

Recall the satisfiability problems introduced in section 1.2.5. I will make use of two variants.

Definition 12. *The computational problems MAX 2-SAT (MIN 2-SAT) takes a boolean formula*

$$BF = CL_1 \wedge \ldots \wedge CL_m$$

with variables $v_1, \ldots, v_n$, where each clause CL_j has exactly two literals, representing distinct variables. It outputs the maximum (minimum) number of clauses that can be simultaneously satisfied, i.e.

$$\max_{v_1,\ldots,v_n} \sum_{j=1}^{m} 1(CL_j = \text{true})$$

Garey et al. (1976) showed that there does not exist a polynomial-time algorithm for MAX 2-SAT unless P = NP. Later, Kohli et al. (1994) proved the analogous result for MIN 2-SAT.

[19] On ternary menus M, expected utility maximization is strongly tractable even if u is not additively separable. A simple brute-force search algorithm can evaluate each of the three lotteries $X \in M$ and choose the best one. This evaluation can be done quickly as long as u is efficiently computable.

The high-level structure of the proof is an *algorithmic reduction*, which is a particular kind of proof by contradiction. I show that solving MAX 2-SAT (or MIN 2-SAT) can be reduced to solving expected utility maximization for utility function u. More precisely, a polynomial-time algorithm for the latter can be used as a subroutine to construct a polynomial-time algorithm for the former. Of course, this contradicts P $\neq$ NP.

Although each reduction is unique, algorithmic reductions are the prototypical proof technique in computational complexity theory. What distinguishes this result is that it is not enough to establish just one reduction, for *some* utility function u. I need to show that polynomial-time reductions exist for *every* symmetric utility function u, armed only with the knowledge that u is symmetric and not additively separable. In that sense, I need to prove a *dichotomy theorem* (c.f. Schaefer 1978), which characterizes the time complexity of a large class of computational problems.

The first lemma establishes a useful fact about additively separable utility functions.

Lemma 1. *Let u be a symmetric utility function. Then u is additively separable iff there do not exist constants $a, b \in \mathbb{Q}$ and an outcome $x \in \mathcal{X}$ such that*

$$u(a, a, x_3, x_4, \ldots) + u(b, b, x_3, x_4, \ldots) \neq u(a, b, x_3, x_4, \ldots) + u(b, a, x_3, x_4, \ldots)$$

The next two lemmas establish polynomial-time reductions for two different cases. It follows from Lemma 1 that these cases are collectively exhaustive.

Lemma 2. *Suppose a tractable choice correspondence c maximizes expected utility, where there exist constants $a, b \in \mathbb{Q}$ and an outcome $x \in \mathcal{X}$ such that*

$$u(a, a, x_3, x_4, \ldots) + u(b, b, x_3, x_4, \ldots) > u(a, b, x_3, x_4, \ldots) + u(b, a, x_3, x_4, \ldots)$$

Then there exists a polynomial-time algorithm for MAX 2-SAT.

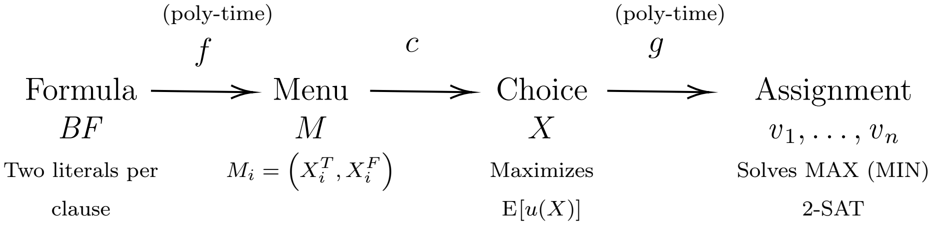


Figure 1.5: The high-level structure of the reduction algorithm used in Lemma 2 (Lemma 3). The function f maps formulas into product menus M. The choice correspondence c maps menus M into lotteries X^M that maximize expected utility. The function g maps lotteries X into true/false assignments to $v_1, \ldots, v_n$ that solve MAX (MIN) 2-SAT. Since f and g can be computed in polynomial-time, this algorithm has polynomial runtime if and only if c is tractable.

Lemma 3. *Suppose a tractable choice correspondence c maximizes expected utility, where there exist constants $a, b \in \mathbb{Q}$ and an outcome $x \in \mathcal{X}$ such that*

$$u(a, a, x_3, x_4, \ldots) + u(b, b, x_3, x_4, \ldots) < u(a, b, x_3, x_4, \ldots) + u(b, a, x_3, x_4, \ldots)$$

Then there exists a polynomial-time algorithm for MIN 2-SAT.

The high-level structure of the proof of Lemma 2 (Lemma 3) is illustrated in figure 1.3.2. The goal is to construct a reduction algorithm that solves MAX (MIN) 2-SAT, using an algorithm that maximizes expected utility as a subroutine.

This algorithm is tied to a specific utility function u, and is well-defined whenever u is symmetric but not additively separable. First, I define a function f that maps a given formula BF into a product menus M. The partial menus M_i are binary, and consist of two partial lotteries: X_i^T that will represent a "true" value for v_i and X_i^F that will represent a "false" value for v_i These partial lotteries are constructed in the proof, and depend on BF through the function f. Second, I compute the agent's choice $X = c(M)$ from menu M. Third, I define a function g that maps lotteries X to true/false assignments. This is straightforward: $v_i = 1$ if and only if $X_i = X_i^T$. Finally, I verify that this algorithm satisfies a special property: assignment $g(c(f(BF)))$ solves MAX (MIN) 2-SAT if the choice correspondence c maximizes expected utility.

The rest of the argument is proof by contradiction. If maximizing expected utility were tractable for *any* utility function that is symmetric but not additively separable, the reduction algorithm runs in polynomial time. The reason is that f and g can be computed in polynomial time, and polynomial functions are closed under composition. Of course, that leads to a contradiction. The reduction algorithm is a polynomial-time algorithm for MAX (MIN) 2-SAT. But this contradicts P $\neq$ NP, as discussed above.

The final step in proving Theorem 1 is to verify efficient computability.

Lemma 4. *A choice correspondence that is rational and strongly tractable reveals an efficiently computable utility function.*

The proof of Lemma 4 transforms the choice-generating algorithm into a utility-computing algorithm. I associate a utility level $y \in [0, 1]$ with lottery X^y that outputs the least desirable outcome with probability y and the most desirable outcome with probability $1 - y$. Then I assign outcome x a utility $u(x) = y$ if the agent chooses x when offered $\{x, X^{y-\epsilon}\}$, but chooses $X^{y+\epsilon}$ when offered $\{x, X^{y+\epsilon}\}$. This argument relies on assumption 2 which ensures that binary menus are represented in the collection $\mathcal{M}$.

1.3.3 Proof of Special Cases

I consider two special cases that illustrate how I prove lemma 2, which is similar to how I prove lemma 3. I leave the full proofs to appendix **??**.

Maximum Utility. First, I show that maximizing expected utility with

$$u(x) = \max_i x_i$$

is intractable, assuming P $\neq$ NP. This turns out to be straightforward, so it is a useful warmup.

Let BF be a boolean formula with n variables $v_1, \ldots, v_n$ and m clauses $CL_1, \ldots, CL_m$. Each clause has at most two literals, which I can write as

$$CL_j = v_{j_1} \vee v_{j_2}$$

The auxilliary variables v_{j_k} are meant to represent literals v_i or $\neg v_i$ for the original n variables. Given this formula, MAX 2-SAT solves

$$\begin{aligned} \max_{v_i \in \{\text{true, false}\}} \sum_{j=1}^{m} \mathbf{1}(CL_j) &= \max_{v_i \in \{\text{true, false}\}} \sum_{j=1}^{m} \mathbf{1}\left(v_{j_1} \vee v_{j_2}\right) \\ &= \max_{v_i \in \{\text{true, false}\}} \sum_{j=1}^{m} \max\left\{\mathbf{1}\left(v_{j_1}\right), \mathbf{1}\left(v_{j_2}\right)\right\} \qquad (1.3) \end{aligned}$$

where the indicator satisfies $\mathbf{1}(\text{true}) = 1$ and $1(\text{false}) = 0$. Compare this with expected utility maximization with the n-dimensional product menu described in the previous subsection, i.e.

$$\max_{X_i \in \{X_i^T, X_i^F\}} \mathrm{E}[\max\{X_1, \ldots, X_n\}] \qquad (1.4)$$

These optimization problems are already quite similar. It only remains to define the partial lotteries X_i^T, X_i^F in a way that makes them equivalent.

The high-level idea behind the partial lottery X_i^T is that it chooses a random clause j to evaluate. It returns $x_i = 1$ if setting $v_i = \text{true}$ makes CL_j true, and otherwise returns $x_i = 0$. Similarly, X_i^F returns $x_i = 1$ if setting $v_i = \text{false}$ makes CL_j true, and otherwise returns $x_i = 0$. The decisionmaker will end up choosing the lottery X that maximizes the probability that a randomly-chosen clause j is true under assignment $f(X)$. But this is equivalent to maximizing the number of clauses j that are true under $f(X)$, i.e. solving MAX 2-SAT.

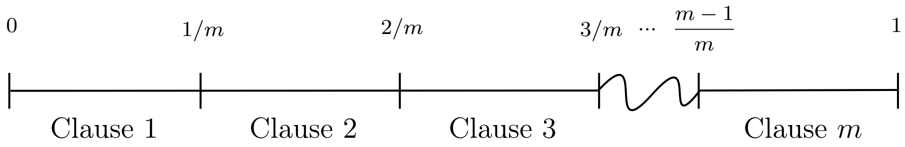


Figure 1.6: This diagram depicts the sample space $\Omega = [0, 1]$, broken up into m intervals of equal size. Interval j is associated with clause j.

Formally, I divide the sample space Ω into m equally-sized intervals. Figure 1.3.3 illustrates. When ω falls in the j^{th} interval, i.e. $\omega \in [(j-1)/m, j/m)$, define

$$X_i^T(\omega) = \begin{cases} 1 & v_{j_1} = v_i \\ 1 & v_{j_2} = v_i \\ 0 & \text{otherwise} \end{cases} \qquad X_i^F(\omega) = \begin{cases} 1 & v_{j_1} = \neg v_i \\ 1 & v_{j_2} = \neg v_i \\ 0 & \text{otherwise} \end{cases}$$

It follows that

$$\begin{aligned} \mathrm{E}[\max\{X_1, \ldots, X_n\}] &= \frac{1}{m} \sum_{j=1}^{m} \mathrm{E}\left[\max\{X_1, \ldots, X_n\} \mid \omega \in \left[\frac{j-1}{m}, \frac{j}{m}\right)\right] \\ &= \frac{1}{m} \sum_{j=1}^{m} \mathrm{E}\left[\max\left\{\mathbf{1}\left(v_{j_1} \mid f(X)\right), \mathbf{1}\left(v_{j_2} \mid f(X)\right)\right\} \mid \omega \in \left[\frac{j-1}{m}, \frac{j}{m}\right)\right] \\ &= \frac{1}{m} \sum_{j=1}^{m} \max\left\{\mathbf{1}\left(v_{j_1} \mid f(X)\right), \mathbf{1}\left(v_{j_2} \mid f(X)\right)\right\} \end{aligned}$$

where $(v_{j_k} \mid f(X))$ refers to the value of auxilliary variable v_{j_k} given the assignment

$$v_1, \ldots, v_n = f(X)$$

This is proportional to the objective (1.3) of MAX 2-SAT. Therefore, the lottery X that maximizes expected utility (1.4) also leads to an assignment $f(X)$ that solves MAX 2-SAT.

Quadratic Utility. Next, I show that maximizing expected utility with

$$u(x) = (x_1 + x_2 + \ldots)^2$$

is intractable, assuming P $\neq$ NP. This builds on the same structure as the previous case, but is somewhat more involved. Essentially, the goal is to manipulate this utility function into something that looks like maximum utility.

The previous construction no longer works, but it is instructive to see why. Consider a clause $CL_j = x_1 \vee x_2$. If I choose partial lotteries X_1^T, X_2^T representing true assignments to both variables x_1 and x_2, then expected utility conditional on the interval associated with clause j is

$$(1 + 1)^2 = 4$$

If I instead choose partial lotteries X_1^T, X_2^F where variable x_2 is now false, that conditional expected utility becomes

$$(1 + 0)^2 = 1$$

In both cases, clause j would be true under assignment $f(X)$. However, expected utility assigns a higher payoff when both literals in clause j are true, compared to when only one is true. Essentially, expected utility corresponds to a multi-valued or fuzzy logic where clause j is merely "somewhat true" if only one literal is true.

To address this, I need to refine the construction. First, I rewrite each clause CL_j in its disjunctive normal form, i.e.

$$CL_j = (v_{j_1} \wedge v_{j_2}) \vee (\neg v_{j_1} \wedge v_{j_2}) \vee (v_{j_1} \wedge \neg v_{j_2})$$

Next, I take each interval in the sample space Ω and break it down into three subintervals. Each subinterval corresponds to one term in the disjunctive normal form of CL_j. Figure

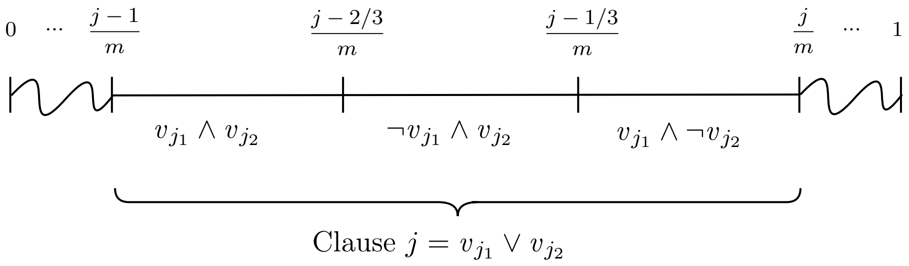


Figure 1.7: This diagram depicts the sample space $\Omega = [0, 1]$, broken up into $3m$ intervals of equal size. Each triple of intervals is associated with a clause. Within each triple associated with clause j, the three intervals correspond to the three terms in the disjunctive normal form of clause j.

1.3.3 illustrates. This enriched state space will offer additional degrees of freedom to turn the expected quadratic utility function into something that mimics the maximum utility function.

The high-level idea behind the partial lottery X_i^T is similar to before. It evaluates a random entry in the disjunctive normal form of clause j. If setting $v_i =$ true does not falsify that entry, it returns a positive value (the precise value depends on the entry selected). Otherwise, it returns zero. The lottery X_i^F is defined analogously. I claim that, as before, the decisionmaker will end up choosing the lottery X that maximizes the probability that a randomly-chosen clause j is true under assignment $f(X)$.

To define these partial lotteries, fix a constant $\gamma \in (0, 1)$. When ω falls into the first subinterval associated with clause j, set

$$X_i^T(\omega) = \begin{cases} \gamma & v_{j_1} = v_i \\ \gamma & v_{j_2} = v_i \\ 0 & \text{otherwise} \end{cases} \qquad X_i^F(\omega) = \begin{cases} \gamma & v_{j_1} = \neg v_i \\ \gamma & v_{j_2} = \neg v_i \\ 0 & \text{otherwise} \end{cases}$$

When ω falls into the second subinterval associated with clause j, set

$$X_i^T(\omega) = \begin{cases} 1 & \neg v_{j_1} = v_i \\ 1 & v_{j_2} = v_i \\ 0 & \text{otherwise} \end{cases} \qquad X_i^F(\omega) = \begin{cases} 1 & \neg v_{j_1} = \neg v_i \\ 1 & v_{j_2} = \neg v_i \\ 0 & \text{otherwise} \end{cases}$$

When ω falls into the third subinterval associated with clause j, set

$$X_i^T(\omega) = \begin{cases} 1 & v_{j_1} = v_i \\ 1 & \neg v_{j_2} = v_i \\ 0 & \text{otherwise} \end{cases} \qquad X_i^F(\omega) = \begin{cases} 1 & v_{j_1} = \neg v_i \\ 1 & \neg v_{j_2} = \neg v_i \\ 0 & \text{otherwise} \end{cases}$$

Now, consider the decisionmaker's expected utility from lottery X, conditioned on the interval associated with clause j. That is,

$$\mathrm{E}\left[(X_1 + \ldots + X_n)^2 \mid \omega \in \left[\frac{j-1}{m}, \frac{j}{m}\right)\right] \tag{1.5}$$

When the assignment $f(X)$ makes both variables in clause j true, expected utility (1.5) equals

$$A := \frac{1}{3}\left((\gamma+\gamma)^2 + (0+1)^2 + (1+0)^2\right) = \frac{4\gamma^2}{3} + \frac{2}{3} \tag{1.6}$$

When the assignment $f(X)$ makes v_{j_1} true but v_{j_2} false, expected utility (1.5) equals

$$B := \frac{1}{3}\left((\gamma+0)^2 + (1+1)^2 + (0+0)^2\right) = \frac{\gamma^2}{3} + \frac{4}{3} \tag{1.7}$$

Since u is symmetric, this is also true if $f(X)$ makes v_{j_2} true but v_{j_1} false. Finally, when

$f(X)$ makes neither entry in clause j true, expected utility (1.5) equals

$$C := \frac{1}{3}\left((0+0)^2 + (1+0)^2 + (0+1)^2\right) = \frac{2}{3} \tag{1.8}$$

Since $\gamma > 0$, it is clear that the last case (1.8) (where clause j is false) delivers lower expected utility than the three other cases (where clause j is true). To finish the construction, we need to ensure that the expected utility is the same in any case where the clause j is true. Setting (1.6) equal to (1.7) and solving for γ yields

$$\gamma = \sqrt{\frac{2}{3}}$$

Given this value of γ, the previous argument goes through.[20] The lottery X that maximizes expected quadratic utility in this menu also yields an assignment $f(X)$ that solves MAX 2-SAT.

This section only proved lemma 2 for two special cases. The full proof in appendix **??** has a similar structure. As long as the utility function is symmetric and not additively separable, it is possible to make an argument along these lines.

1.4 Dynamic Choice Bracketing

This section relaxes the symmetry assumption of section 1.3, in order to complete the characterization of rational and tractable choice. In particular, I show that rational and tractable choice corresponds to *dynamic choice bracketing*, a novel generalization of choice bracketing. Equivalently, it corresponds to expected utility maximization with a *Hadwiger separable*

[20] One concern is that γ is an irrational number, and therefore lacks a finite decimal representation. But it is not necessary to set γ to be exactly this number. All that is needed is for the discrepancy between (1.6) and (1.7) to be less than $1/m$ times the difference between (1.6) and (1.8), as well as the difference between (1.7) and (1.8). This ensures that the discrepancy will not affect the optimal choice. It can be achieved with a γ that has $O(\log m)$ digits.

utility function, where Hadwiger separability is a novel relaxation of additive separability.

I begin with two examples, which demonstrate that the conclusion of theorem 1 no longer holds when the symmetry assumption is relaxed. More precisely, a utility function need not be additively separable in order for expected utility maximization to be tractable.

Example 4. Suppose the decisonmaker is choice bracketing, but less narrowly than in section 1.3. She partitions dimensions $i = 1, \dots, n$ into mutually exclusive brackets $B_1, \dots, B_m$. For each bracket B_j, she maximizes expected utility according to a utility function u_j that is defined over the coordinates $i \in B_j$. Let $k = \max_j |B_j|$ be the size of the largest bracket.

For concreteness, consider a consumer choosing from eight available products: cereal, napkins, milk, ground beef, chicken, jam, apples, oranges. The consumer has four brackets. The first bracket consists of breakfast foods (cereal, milk, jam). The second bracket is just napkins. The third bracket consists of raw meat (ground beef, chicken). The fourth bracket consists of fruits (apples, orange). The consumer's revealed utility function is

$$u(x) = u_1(x_1, x_3, x_6) + u_2(x_2) + u_3(x_4, x_5) + u_4(x_7, x_8)$$

where x_i denotes the amount of product i she consumes. Each bracket includes natural complements (e.g. cereal and milk) or substitutes (e.g. apples and oranges). But across brackets, the consumer ignores any complementarity or substitutability.

Although u is not additively separable in example 4, expected utility maximization is tractable as long as the bracket size does not grow too quickly with the number of products n. Formally, it is tractable as long as $k = O(\log n)$.[21] I call this *relatively narrow* choice bracketing.

[21] It is always possible to optimize within each bracket by brute-force search. The runtime of the algorithm will be exponential in k, where k is the size of the largest bracket. When $k = O(\log n)$, a runtime that is exponential in k is only polynomial in n. Therefore, brute-force search can maximize expected utility in polynomial time.

Example 5. Suppose the decisionmaker is willing to narrowly bracket decisions $i = 2, \ldots, n$, but only after conditioning on decision 1.

For concreteness, consider an individual whose first decision is where she wants to live. Then she must decide how much of several different products to acquire: gasoline, snow boots, swimsuits, gardening tools, hammocks, etc. These products lack obvious complementarities or substitutabilities, so the consumer is willing to evaluate each product without considering the others. However, her preferences over all of these products depend on where she lives. For example, she may value gasoline more in Los Angeles than in Chicago, but snow boots more in Chicago than Los Angeles. The decisionmaker's revealed utility function is

$$u(x) = u_2(x_1, x_2) + u_3(x_1, x_3) + u_4(x_1, x_4) + u_5(x_1, x_5) + u_6(x_1, x_6) + \ldots$$

This decisionmaker cannot evaluate one product separately from another. For example, she cannot fully separate gasoline from snow boots. If gasoline were unavailable, then she probably would not move to Los Angeles, which might make her value snow boots more.

Although u is not additively separable in example 5, expected utility maximization is tractable. It is straightforward to maximize expected utility in polynomial time using backwards induction. There are two steps to this algorithm:

1. Conditional on her choice X_1, compute her optimal choices $X_2^*(X_1), \ldots, X_n^*(X_1)$, i.e.

$$X_i^*(X_1) \in \arg\max_{X_i \in M_i} \mathrm{E}[u_i(X_1, X_i)]$$

2. Choose X_1 to maximize expected utility, given her planned choices $X_2^*, \ldots, X_n^*$, i.e.

$$\mathrm{E}[u\left(X_1, X_2^*(X_1), \ldots, X_n^*(X_1)\right)]$$

This is not choice bracketing, but it has a similar flavor. For each product $i = 2, \ldots, n$, the decisionmaker brackets together her consumption decision X_i with her location decision X_1. Then, when the time comes to choose X_1, she does not need to reconsider her consumption decisions. After all, she has already determined her choices $X_2^*, \ldots, X_n^*$ as a function of X_1.

1.4.1 Dynamic Choice Bracketing

These examples are both special cases of what I call dynamic choice bracketing.

Dynamic choice bracketing is a family of algorithms that combine principles of dynamic programming with principles of choice bracketing. Like choice bracketing, it may selectively ignore links between decisions i and j. Unlike choice bracketing, the relevant brackets may change in the process of making the choice. For instance, this is what happened in example 5.

The formal definition of dynamic choice bracketing is given below (see algorithm 1). It represents a family of algorithms mapping product menus M to lotteries $X \in M$, which have a particular form. These algorithms visit coordinates $1, \ldots, n$ in a prespecified order. The goal for each coordinate i is to define a function $X_i^*(\cdot)$ that maps choices X_j to a choice X_i. As more coordinates are visited, the algorithms redefines $X_i^*(\cdot)$ so that it remains a function of unvisited coordinates. Eventually, the algorithms visit all coordinates, and the functions $X_i^*(\cdot)$ have no remaining arguments. The output is simply $X_1^*, \ldots, X_n^*$.

The brackets in dynamic choice bracketing are not as neatly defined as in choice bracketing. In particular, they are dynamic. I say that coordinate i belongs to bracket B_i, where

$$B_i = \{i\} \cup S_i \cup I_i$$

consists of i's successors and indirect influencers, as defined in algorithm 1. Although coordinate i's predecessors $j \in P_i$ also enter into value function V_i, the decisionmaker does not

Input: product menu M.

Process: visit coordinates $i \in \{1, \dots, n\}$ in a prespecified order. At each i:

1. Specify the *successors* S_i of i.

   This is some subset of the unvisited coordinates j.

2. Identify the *predecessors* P_i of i.

   This is the subset of visited coordinates j where the choice $X_j^*(\cdot)$ depends on X_i.

3. Specify value function V_i that depends on i, successors S_i, and predecessors P_i, i.e.

   $$V_i\left(X_i, X_{S_i}, X_{P_i}\right) \in \mathbb{R}$$

4. Identify the indirect influencers I_i of i.

   This is the subset of unvisited coordinates j where choice $X_k^*(\cdot)$ depends on X_j for predecessors $k \in P_i$.

5. Define the choice $X_i^*(\cdot)$ as a function of successors and indirect influencers.

   This is done by optimizing the value function as follows:

   $$X_i^*\left(X_{S_i}, X_{I_i}\right) \in \arg\max_{X_i \in M_i} V_i\left(X_i, X_{S_i}, X_{P_i}^*\left(X_i, X_{I_i}\right)\right)$$

6. Redefine the choices X_j^* for predecessors $j \in P_i$ by replacing X_i with $X_i^*(\cdot)$, i.e.

   $$X_j^*\left(X_{S_i}, X_{I_i}\right) := X_j^*\left(X_i^*\left(X_{S_i}, X_{I_i}\right), X_{I_i}\right) \quad \forall j \in P_i$$

Output: $(X_1^*, \dots, X_n^*) \in \mathcal{X}$. This is well-defined because, once all coordinates have been visited, choices $X_i^*(\cdot)$ have no remaining arguments.

Algorithm 1: A prototypical dynamic choice bracketing algorithm.

need to reconsider those choices: they are given by $X_j^*(X_i, X_{P_i})$ as a function of i and its predecessors.

As with choice bracketing, dynamic choice bracketing is only a meaningful restriction when the brackets are small.[22] In this case, the size of the largest bracket (or *bracket size*) is

$$k = \max_i |B_i|$$

Definition 13. *A choice correspondence c is consistent with* relatively narrow *dynamic choice bracketing if it can be generated by some specification of algorithm 1 with bracket size*

$$O(\log n)$$

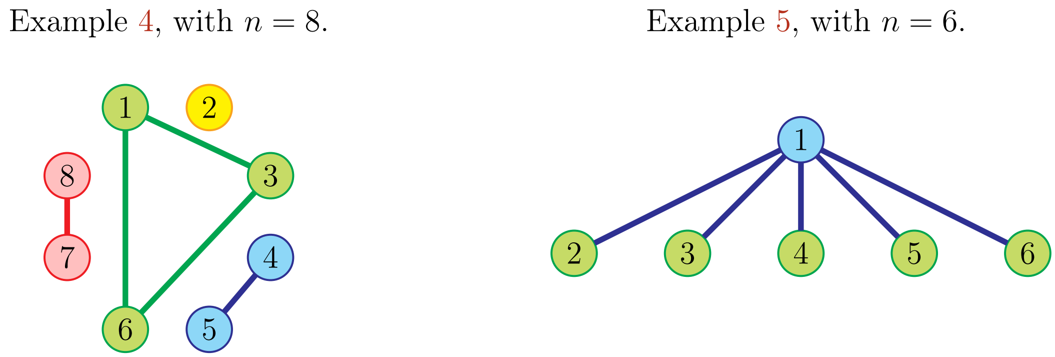
where n is the dimension of the menu M.

1.4.2 Hadwiger Separability

Previously, I related narrow choice bracketing to additive separability, and used theorem 1 to motivate additive separability. In that same spirit, I will relate dynamic choice bracketing to Hadwiger separability. In the next subsection, I use theorem 2 to motivate Hadwiger separability.

Hadwiger separability is a relaxation of additive separability. It captures a sense in which most pairs (x_i, x_j) are evaluated separately from each other, but not necessarily all. Its advantage relative to additive separability is that it preserves computational tractability while being capable of modeling a much richer class of phenomena. Nonetheless, it is still quite restrictive, especially in combination with other assumptions like symmetry.

[22] When $k = n$, dynamic choice bracketing includes backwards induction, which can be used to maximize expected utility for any utility function u. Of course, backwards induction will not necessarily run in polynomial time.


Example 4, with $n = 8$.

Example 5, with $n = 6$.

Figure 1.8: The inseparability graphs $G_n(u)$ associated with utility functions in Example 4 and 5.

The first step to defining Hadwiger separability is to define a pairwise notion of separability.

Definition 14. *A utility function u is (i, j, n)-separable if there exist functions u_i, u_j such that, for all n-dimensional outcomes x,*

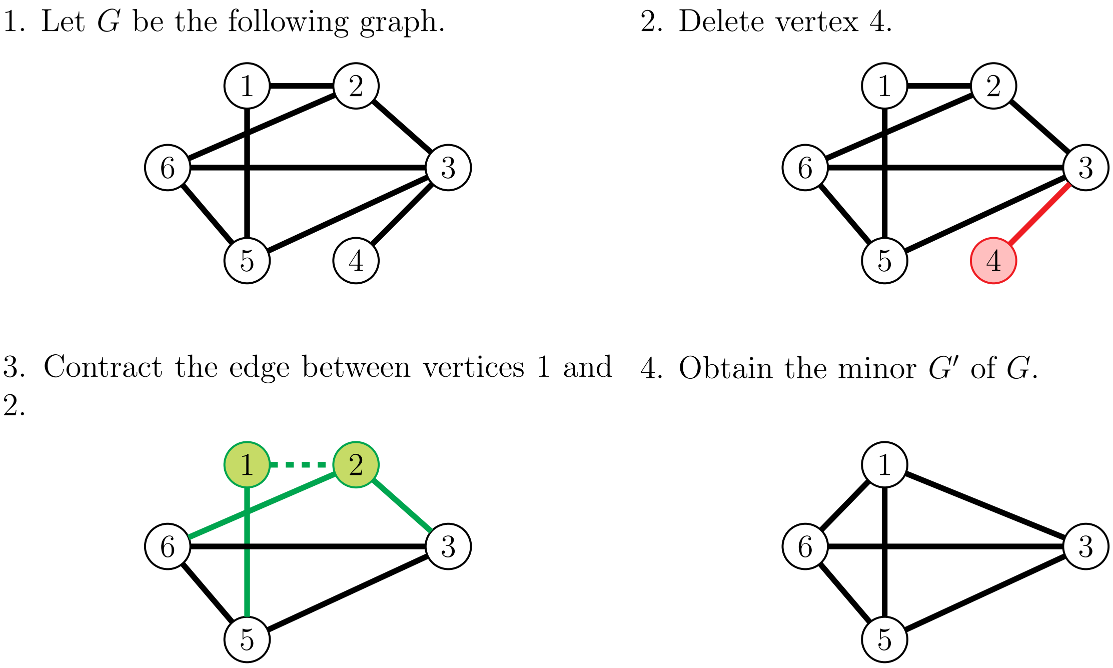
$$u(z) = u_i(x_i, x_{-ij}) + u_j(x_j, x_{-ij})$$

The second step is introduce a graph that identifies which pairs (i, j, n) are not separable.

Definition 15. *The* inseparability graph *$G_n(u)$ of utility function u is an undirected graph with n nodes. There is an edge between nodes i and j if and only if u is* not *(i, j, n)-separable.*

Figure 1.8 depicts the inseparability graphs associated with Example 4 and 5. As we will see, these are also examples of sparse graphs.

The utility function u is Hadwiger separable if its inseparability graph $G_n(u)$ quickly becomes sparse as n grows large. To formalize this, I need a measure of graph sparsity. It turns out that the right measure was formulated by Hadwiger (1943) to state his longstanding conjecture about the chromatic number of graphs. It refers to a concept called graph minors.



1. Let G be the following graph.

2. Delete vertex 4.

3. Contract the edge between vertices 1 and 2.

4. Obtain the minor G' of G.

Figure 1.9: In this example, I find the Hadwiger number of the graph G. The minor G' is complete and has four vertices. In fact, this is the largest complete minor, so $\text{Had}(G) = 4$.

Definition 16. *Let G' be a subgraph of the undirected graph G. Then G' is a* minor *if it can be formed from G by some sequence of the following two operations:*

1. Delete *a node i and all of its incident edges (i, j).*

2. Contract *an edge (i, j). This deletes nodes i, j and replaces them with a new vertex k. It also replaces any edges (i, l) and (j, l) with a new edge (k, l).*

Definition 17. *Let G be an undirected graph. The* Hadwiger number $\text{Had}(G)$ *of G is the number of nodes in its largest complete minor.*[23]

Figure 1.9 illustrates these definitions, through an example.

Definition 18. *The function u is* Hadwiger separable *if*

$$\text{Had}(G_n(u)) = O(\log n)$$

[23] A complete graph (or minor) is one in which all nodes share an edge.

Hadwiger separability is an asymptotic property, like computational tractability. However, it is often easy to verify whether a utility function is Hadwiger separable or not.[24] Take Example 4 and 5, whose inseparability graphs are depicted in Figure 1.8. In Example 4, the Hadwiger number is the size k of the largest bracket. This is consistent with Hadwiger separability if and only if choice bracketing is relatively narrow, i.e. $k = O(\log n)$. In Example 5, the Hadwiger number is 1, independently of n. Clearly this is consistent with Hadwiger separability.

Hadwiger separability may appear quite general, but it is not. Compared to additive separability, it is capable of modeling a richer set of preferences, such as preferences involving a limited number of complementarities and substitutions between goods. However, it remains restrictive in the sense that "most" utility functions are not Hadwiger separable. In fact, if we restrict attention to symmetric utility functions, Hadwiger separability is equivalent to additive separability.

Proposition 3. *A symmetric utility function is Hadwiger separable iff it is additively separable.*

Proof. See Figure 1.10. □

Finally, I relate dynamic choice bracketing with Hadwiger separability.

Proposition 4. *Let c be a rational choice correspondence. If c reveals a Hadwiger separable utility function, then it can be generated by relatively narrow dynamic choice bracketing. If the NU-ETH holds then the converse is also true.*

This relationship is not obvious, in contrast to the relationship between narrow choice bracketing and additive separability. It will become clearer in the proof outline of theorem 3.

[24] In general, computing the Hadwiger number is NP-hard (Eppstein 2009). However, for any inseparability graph $G_n(u)$ and constant C, it is possible to determine whether $\mathrm{Had}(G_n(u)) \leq C \log n$ within $O(\mathrm{poly}(n, C))$ time. This follows from a fixed parameter tractability result of Alon et al. (2007).

An empty graph G, with $\mathrm{Had}(G) = 0$.

A complete graph G, with $\mathrm{Had}(G) = 8$.

1
2
3
4
5
6
7
8
1
2
3
4
5
6
7
8

Figure 1.10: When the function u is symmetric, the inseparability graph $G_n(u)$ is either empty or complete. If $G_n(u)$ is empty (left), then u is additively separable. If $G_n(u)$ is complete (right), then $\mathrm{Had}(G_n(u)) = n$ and therefore u is not Hadwiger separable. It follows that Hadwiger separability is equivalent to additive separability when u is symmetric.

1.4.3 Representation Theorem

Theorem 2 shows that rational and weakly tractable choice implies expected utility maximization with a Hadwiger separable utility function. This result relies on the non-uniform exponential time hypothesis (NU-ETH). Theorem 3 gives a partial converse: expected utility maximization is weakly tractable on product menus when the utility function is Hadwiger separable.

These results refer to weak tractability, whereas my earlier results (theorem 1, proposition 2) referred to strong tractability. This has three implications. First, it makes the hardness result (theorem 2) stronger and the partial converse (theorem 3) weaker. Second, I rely on stronger computational hardness conjectures. Third, I use a relaxed notion of efficient computability.

Definition 19. *A utility function u is* efficiently computable with advice *if it satisfies definition 11 with a Turing machine that has access to $O(\mathrm{poly}(n, 1/\epsilon))$-size advice.*

Theorem 2. *Let choice correspondence c be rational and weakly tractable. If the NU-ETH holds, then c reveals a Hadwiger separable utility function, which is efficiently computable* *with advice.*

This result essentially implies theorem 1 as a corollary. Suppose that the choice correspondence c is symmetric, as well as rational and weakly tractable. By theorem 2, u is Hadwiger separable. Since u is also symmetric, proposition 3 implies that u is additively separable. That is the conclusion of theorem 1. The only remaining differences are that theorem 1 made a stronger tractability assumption and relied on a weaker computational hardness conjecture.

I argue that theorem 2 is tight by providing a partial converse.

Theorem 3. *Let the utility function u be Hadwiger separable and efficiently computable with advice. Then expected utility maximization is weakly tractable on the collection of product menus.*

Next, I outline the proofs of theorems 2 and 3. The full proofs are left to appendix **??**.

1.4.4 Proof Outline of Theorem 2

The argument is similar to the reduction argument in Theorem 1, only more involved. Most of the work is done by the following lemma, which plays the same role here that Lemmas 2 and 3 played in the proof of Theorem 1.

Lemma 5. *Suppose a weakly tractable choice correspondence maximizes expected utility, where*

$$d_n := \mathrm{Had}(G_n(u))$$

Then there exists an $O(\mathrm{poly}(n))$-time algorithm to solve MAX 2-SAT for any boolean formula with at most d_n variables. This algorithm uses at most $O(\mathrm{poly}(n))$-size advice.

For the purposes of solving MAX 2-SAT, the utility function u is effectively d_n-dimensional. In other words, even if u is separable across some dimensions, it is effectively high-dimensional as long as its inseparability graph has a large complete minor.

The advice in Lemma 5 provides two kinds of information. First, it describes the largest complete minor of the inseparability graph $G_n(u)$. This is the same minor that is used to define the Hadwiger number, and it has exactly d_n nodes. Second, it identifies points where the utility function u is not (i,j,n)-separable. To define this precisely, I need additional notation.

Definition 20. *An n-dimensional outcome x and quadruple $a_1, a_2, b_1, b_2 \in [0,1]$ constitute a* violation *of (i,j,n)-separability if*

$$\begin{aligned} &u\left(\ldots, x_{i-1}, a_1, x_{i+1}, \ldots, x_{j-1}, a_2, x_{j+1}, \ldots\right) + u\left(\ldots, x_{i-1}, b_1, x_{i+1} \ldots, x_{j-1}, b_2, x_{j+1}, \ldots\right) \\ &\neq u\left(\ldots, x_{i-1}, a_1, x_{i+1}, \ldots, x_{j-1}, b_2, x_{j+1}, \ldots\right) + u\left(\ldots, x_{i-1}, b_1, x_{i+1}, \ldots, x_{j-1}, a_2, x_{j+1}, \ldots\right) \end{aligned} \tag{1.9}$$

Lemma 6. *A utility function u is (i,j,n)-separable iff there exists no violation of (i,j,n)-separability.*

The algorithm in Lemma 5 takes a violation of (i,j,n)-separability as advice, for every pair of dimensions $i, j \leq n$ where u is not (i,j,n)-separable.

Lemma 5 has two immediate corollaries.

Corollary 1. *Suppose a weakly tractable choice correspondence maximizes expected utility, where*

$$\text{Had}(G_n(u)) = \Omega(\text{poly}(n)) \tag{1.10}$$

Then there exists a $O(\text{poly}(n))$-time algorithm for MAX 2-SAT with n variables. This algorithm uses at most $O(\text{poly}(n))$-size advice.

This contradicts NP $\not\subset$ P/poly and will be useful in the proof of Theorem 4.

Corollary 2. *Suppose a weakly choice correspondence c maximizes expected utility, where*

$$\mathrm{Had}(G_n(u)) = \omega(\log n) \tag{1.11}$$

Then the there exists a $O(2^{o(n)}$-time algorithm for 3-SAT with n variables. This algorithm uses at most $O(\mathrm{poly}(n))$-size advice.

This completes the proof of Theorem 2: the conclusion of Corollary 2 contradicts the NU-ETH, and the only utility functions that do not satisfy condition (1.11) are Hadwiger separable.

1.4.5 Proof Outline of Theorem 3

To prove Theorem 3, I construct a dynamic choice bracketing algorithm that maximizes expected utility in polynomial time as long as the utility function u is Hadwiger separable.

In order to define the algorithm, I need to review another measure of graph sparsity called *contraction degeneracy*, which turns out to be closely related to the Hadwiger number.

Definition 21. *Let G be an undirected graph.*

1. *The* degree *of node i is the number of nodes $j \neq i$ with which i shares an edge.*

2. *The contraction degeneracy $\mathrm{cdgn}(G)$ is the smallest number d such that every minor G' of G has a vertex with degree less than or equal to d.*

Lemma 7. *The utility function u is Hadwiger separable if and only if*

$$\mathrm{cdgn}(G_n(u)) = O(\log n)$$

I define algorithm 2 on the next page. This algorithm takes in a product menu M and outputs a lottery $X^* \in M$. It is parameterized by a utility function u and the contraction degeneracy d of the inseparability graph $G_n(u)$. The algorithm requires a description of the inseparability graph $G_n(u)$ as advice, as well as any advice needed to efficiently compute the utility function u.

After reading the description of algorithm 2, it may be useful to refer to Figure 1.11 for a more concrete example. The figure depicts nine iterations of the algorithm on a nine-dimensional product menu. It shows how the predecessors, successors, and indirect influencers are defined in terms of the inseparability graph $G_n(u)$, and how these sets change as the algorithm iterates.

Algorithm 2 is at least superficially consistent with dynamic choice bracketing. However, it is not immediately clear that it is well-defined. There are two steps that require clarification. First, I show that step 1 is always possible, and can be done in polynomial time.

Lemma 8. *Let G be an undirected graph with contraction degeneracy d. There exists a polynomial-time algorithm that converts G into an directed acyclic graph $\vec{G}$ by assigning a direction to each edge in G, where each node i has at most d outgoing edges.*

Next, I show that step 5d will never return an error.

Lemma 9. *Let G be an undirected graph with contraction degeneracy d. Let $\vec{G}$ be the directed acyclic graph from Lemma 8. There exists a node i in $\vec{G}$ that has at most d indirect influencers.*

I show that this algorithm is consistent with dynamic choice bracketing.

Lemma 10. *Algorithm 2 is a special case of Algorithm 1.*

I show that this algorithm is optimal.

Lemma 11. *Algorithm 2 maximizes expected utility. That is, it outputs a lottery $X^* \in c(M)$ that maximizes expected utility.*

Input: product menu M.

Advice:

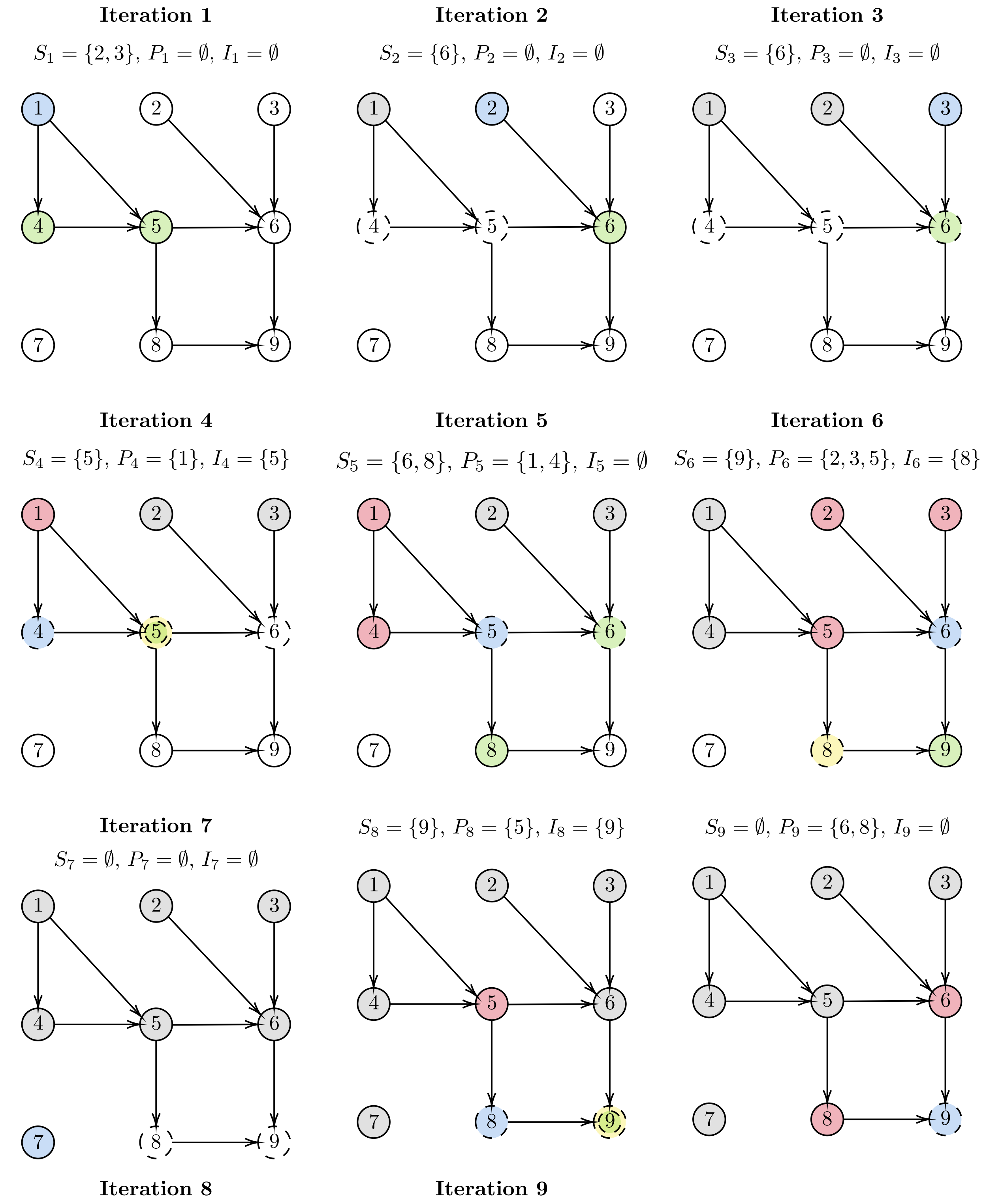
1. Inseparability graph $G := G_n(u)$ with contraction degeneracy d.
2. Any advice needed to efficiently compute the utility function u.

Process:

1. Convert the undirected graph G into a directed acyclic graph $\vec{G}$ by assigning a direction to each edges in G. Each node in $\vec{G}$ has at most d outgoing edges.
2. Do a topological sort of $\vec{G}$. Without loss of generality, assume that the coordinates $i = 1, \ldots, n$ are already sorted correctly.
3. Define a *frontier* $F \subseteq \{1, \ldots, n\}$ that is initially empty. Later, this will keep track of unvisited nodes i that are successors to some visited node j.
4. Let i be the smallest unvisited node in $\vec{G}$.
5. (a) The successors S_i are unvisited nodes where G contains an edge between i and j.
   (b) The predecessors P_i are visited nodes j where $X_j^*(\cdot)$ depends on X_i.
   (c) The indirect influencers I_i are frontier nodes $j \in F$ where G contains a path between i and j that does not pass through F.
   (d) If there are more than d indirect influencers, i.e. $|I_i| > d$, repeat step 5 with the smallest unvisited node $j > i$.
   (e) Define
   $$V_i(X_i, X_{S_i}, X_{P_i}) = \mathrm{E}[u\,(X_i, X_{S_i}, X_{P_i}, 0, 0, \ldots)]$$
   That is, the value function equals expected utility under the (potentially false) assumption that $X_j = 0$ for all coordinates $j \notin \{i\} \cup S_i \cup P_i$.
6. Run steps 5 and 6 of the dynamic choice bracketing algorithm (1).
7. Label node i as visited. Update the frontier F by adding S_i and deleting i, i.e.
   $$F := (F \cup S_i) \setminus \{i\}$$
   Return to step 4 if any unvisited nodes remain in $\vec{G}$.

Output: $(X_1^*, \ldots, X_n^*) \in \mathcal{X}$. This is well-defined because, once all coordinates have been visited, choices $X_i^*(\cdot)$ have no remaining arguments.

Algorithm 2: A dynamic choice bracketing algorithm that maximizes expected utility.



Iteration 1

$S_1 = \{2,3\}$, $P_1 = \emptyset$, $I_1 = \emptyset$

Iteration 2

$S_2 = \{6\}$, $P_2 = \emptyset$, $I_2 = \emptyset$

Iteration 3

$S_3 = \{6\}$, $P_3 = \emptyset$, $I_3 = \emptyset$

Iteration 4

$S_4 = \{5\}$, $P_4 = \{1\}$, $I_4 = \{5\}$

Iteration 5

$S_5 = \{6,8\}$, $P_5 = \{1,4\}$, $I_5 = \emptyset$

Iteration 6

$S_6 = \{9\}$, $P_6 = \{2,3,5\}$, $I_6 = \{8\}$

Iteration 7

$S_7 = \emptyset$, $P_7 = \emptyset$, $I_7 = \emptyset$

$S_8 = \{9\}$, $P_8 = \{5\}$, $I_8 = \{9\}$

$S_9 = \emptyset$, $P_9 = \{6,8\}$, $I_9 = \emptyset$

Iteration 8

Iteration 9

Figure 1.11: Each diagram depicts the directed graph $\vec{G}$ for some iteration of algorithm 2. The node i that is currently being visited is blue. The frontier nodes F have a dashed outline. The predecessors P_i are red, and all other visited nodes are grey. The successors S_i are green. The indirect influencers I_i are yellow. A node $j \in S_i \cap I_i$ is green on the interior and yellow on the exterior, like node 5 in iteration 4. The bracket size is three because there are never more than three nodes in $\{i\} \cup S_i \cup I_i$.

It only remains to show that Algorithm 2 runs in polynomial time. The key step is to show that the algorithm's runtime depends exponentially on d, but polynomially on all other relevant parameters. This is known as a fixed-parameter tractability result, because the algorithm is efficient as long as the parameter d is held fixed.

As before, let M be an n-dimensional product menu where partial menus M_i consist of k partial lotteries X_i, and each X_i is measurable with respect to the same m intervals in the sample space.

Lemma 12. *Algorithm 2 has a runtime of*

$$O(\text{poly}(n, m, k) \cdot \text{poly}(k)^d) \tag{1.12}$$

Finally, recall that u is Hadwiger separable. Therefore, $d = O(\log n)$ by Lemma 7. Plugging this into expression (1.12) yields a runtime of

$$O(\text{poly}(n, m, k))$$

which completes the proof.

1.5 Choice Trilemma

In this section, I establish the choice trilemma. The choice trilemma suggests that the decisionmaker may actually be better off if she is willing to violate the rationality axioms.

To motivate the exercise, consider a thought experiment. A decisionmaker intrinsically cares about expected utility for some utility function $\bar{u}$, but is computationally constrained. I refer to $\bar{u}$ as her objective function, to distinguish it from the *revealed* utility function u. The objective function is what the decisionmaker intrinsically cares about, like profits or pleasure, whereas revealed utility is any utility function that rationalizes the decisionmaker's choices.

In the presence of computational constraints, maximizing expected utility according to $\bar{u}$ may be intractable. Theorem 2 implies that it will be intractable whenever $\bar{u}$ is not Hadwiger separable. In that case, the decisionmaker has one of two options. First, she can make choices that are both tractable and rational, in that they satisfy the expected utility axioms, or equivalently, they maximize expected utility for *some* utility function u. Since these choices are tractable, it must be that $u \neq \bar{u}$. Second, she can make choices that are tractable but violate the expected utility axioms.

When optimal choice according to $\bar{u}$ is intractable, the decisionmaker may settle for tractable choices that are only approximately optimal. For example, she may prefer a choice correspondence that guarantees her at least half of the optimal payoff in any given menu, relative to one that may perform even worse. Theorem 4 shows that it is possible for the decisionmaker to make tractable and approximately optimal choices, but only if she is willing to violate the expected utility axioms. In other words, she will appear irrational to an outside observer.

To reason about approximate optimality, I need to quantify "approximately". For that purpose, I turn to the approximation ratio. This measure of approximate optimality is widely used in computer science to evaluate approximation algorithms for intractable problems. Within economics, it has been applied to the literature on mechanism design (e.g. Hartline and Lucier 2015, Feng and Hartline 2018, Akbarpour, Kominers, et al. 2021).

I make the following assumption in order to simplify the definition of the approximation ratio.

Assumption 6. *Let $\bar{u}$ be the objective function. Then $\bar{u}(0, 0, \dots) = 0$ and $\bar{u}(x) \geq 0$ for all $x \in \mathcal{X}$.*

Definition 22. *Let $\bar{u}$ be a objective function. Then* $\mathrm{APX}^{\bar{u}}_n(c)$ *denotes the* approximation ratio *achieved by a choice correspondence c, where*

$$\mathrm{APX}^{\bar{u}}_n(c) \leq \frac{\mathrm{E}[\bar{u}(c(M))]}{\max_{X\in M} \mathrm{E}[\bar{u}(X)]}$$

for any n-dimensional product menu M.

Theorem 4 has two parts. The first part shows that it is not possible for a choice correspondence to simultaneously be rational, tractable, and approximately optimal. The second part shows that it is possible to be tractable and approximately optimal if one is willing to drop rationality.

Theorem 4. *If NP* $\not\subset$ P/poly*, there exists a objective function $\bar{u}$ where the following are true.*

1. *Let the choice correspondence c be rational and weakly tractable. Then c fails to achieve any constant approximation ratio, i.e.*

$$\lim_{n\to\infty} \mathrm{APX}^{\bar{u}}_n(c) = 0$$

2. *There exists a strongly tractable (but not rational) choice correspondence c' where*

$$\mathrm{APX}^{\bar{u}}_n(c') \geq {}^1\!/_2$$

In fact, the result is stronger than the theorem statement implies. The proof is constructive and identifies a large class of utility functions where the result applies. For example, this class includes

$$\bar{u}(x) = \sqrt{x_1 + x_2 + \dots}$$

I outline the proof in the next subsection.

The choice trilemma refers to the fact that the decisionmaker may care about three properties: rationality, tractability, and approximate optimality. She can satisfy rationality and approximate optimality by maximizing expected utility with respect to her objective function. She can satisfy rationality and tractability by dynamically choice bracketing, as established by Theorem 3. This is only optimal when the objective function is Hadwiger separable, as established by Theorem 2. She can satisfy tractability and approximate optimality, as I show in Theorem 4. But she cannot satisfy all three properties at once, as I also show in Theorem 4.

The choice trilemma implicitly assumes that the approximation ratio is a *reasonable* way to measure approximate optimality. However, I do not claim that it is the *only* way. In particular, we might be concerned if different ways of measuring approximate optimality led to radically different conclusions. Fortunately, I can strengthen theorem 4 to partially address this concern. I begin with a property that any measure of approximate optimality should satisfy: respect for weak dominance.

Definition 23. *Let c, c' be choice correspondences. Then c'* weakly dominates *c if*

$$\mathrm{E}[\bar{u}(c'(M))] \geq \mathrm{E}[\bar{u}(c(M))]$$

for all product menus M, where the inequality is strict for at least one menu.

The next corollary strengthens theorem 4. Rather than compare a rational and tractable choice correspondence c with a tractable and approximately optimal choice correspondence c', it compares c with a tractable and approximately optimal choice correspondence c'' that weakly dominates c. In that sense, any reasonable measure of approximate optimality should agree that c'' is weakly better than c. The approximation ratio is only used to break ties; it gives a sense in which c'' is strictly better than c.

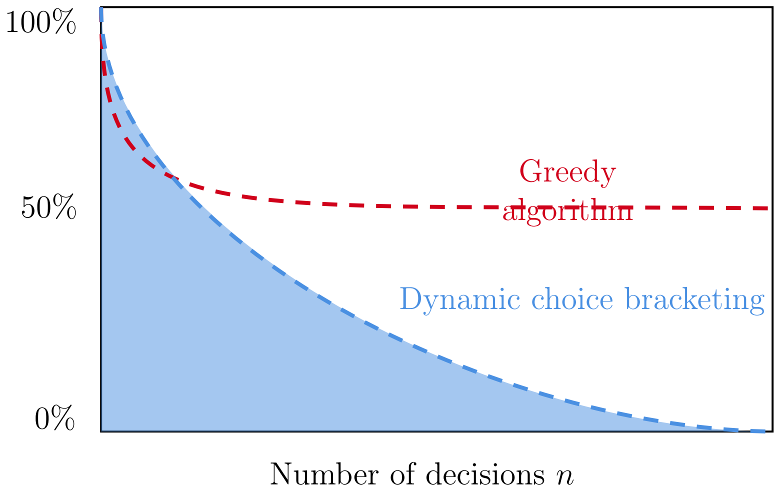


Figure 1.12: This figure plots the approximation ratio (y-axis) against the dimension n (x-axis). It depicts the so-called greedy algorithm in red, and the class of dynamic choice bracketing algorithms in blue, for a particular hedonic utility function. The greedy algorithm violates the expected utility axioms but guarantees a $1/2$ approximation, irrespective of n. However, with dynamic choice bracketing, the approximation ratio vanishes as n grows. The same is true of any rational and tractable choice correspondence, which can be represented as dynamic choice bracketing, by theorem 2.

Corollary 3. *Let $\bar{u}$ be an efficiently computable objective function where theorem 4 holds. Let the choice correspondence c be rational and strongly tractable. If NP $\not\subset$ P/poly, there exists a strongly tractable (but not rational) choice correspondence c'' that weakly dominates c, where*

$$\mathrm{APX}_n^{\bar{u}}(c'') \geq 1/2$$

Proof. Let c' be defined as in the statement of theorem 4. Let c'' be generated by the following algorithm. First, given a product menu M, compute $X \in c(M)$ and $X' \in c'(M)$. Second, evaluate $\mathrm{E}[\bar{u}(X)]$ and $\mathrm{E}[\bar{u}(X')]$, and choose the better of the two lotteries $\{X, X'\}$. □

1.5.1 Proof Outline of Theorem 4

Here, I give a high-level outline of how to prove theorem 4. Figure 1.12 illustrates.

First, I show that rational and weakly tractable choice correspondences may not even be approximately optimal, when n is large. Clearly, this is not always true. If the objective function $\bar{u}$ is Hadwiger separable, then theorem 3 implies that expected objective maximization is weakly tractable. However, this is true whenever $\bar{u}$ satisfies the following property.

Definition 24. *The objective function $\bar{u}$ is ϵ-*sublinear *for some constant $\epsilon > 0$ if*

$$\bar{u}(\underbrace{1,\ldots,1}_{n \textit{ times}},0,\ldots) = O\left(n^{1-\epsilon}\right)$$

Lemma 13. *Let the objective function $\bar{u}$ be symmetric, ϵ-sublinear, and strictly increasing.*[25] *Let the choice correspondence c be rational and weakly tractable. If NP $\not\subset$ P/poly then*

$$\mathrm{APX}_n^{\bar{u}}(c) = \tilde{O}(n^{-\epsilon})$$

That is, no rational and weakly tractable choice correspondence guarantees a constant approximation. As n grows, the approximation ratio converges to zero. The rate of convergence is determined by ϵ. Intuitively, for any given n, objective functions $\bar{u}$ that are more sublinear will be harder to approximate with a rational and weakly tractable choice correspondence.

Next, I turn to the second part of theorem 4. I present a greedy algorithm (3) that generalizes Johnson's (1974) approximation algorithm for MAX SAT.

The greedy algorithm has a lexicographic flavor. In the first iteration, the decisionmaker chooses the partial lottery X_1. Rather than anticipate her remaining choices, she incorrectly assumes that eventual outcome x will be zero-valued in all other dimensions, i.e. $x_i = 0$ for $i \geq 2$. She then maximizes expected objective under that assumption. In the i^{th} iteration, the decisionmaker chooses the partial lottery X_i. Now, she takes into account her choices

[25] By increasing, I mean that $\bar{u}(x) > \bar{u}(x')$ whenever $x_i > x'_i$ for some i.

```
Parameters: objective function $\bar{u}$ that is efficiently computable.
Input: product menu $M$.
Process: iterate over $i = 1, \ldots, n$. For each $i$, define
    $$X_i^* \in \arg\max_{X_i \in M_i} \mathrm{E}\big[\bar{u}\left(X_1^*, \ldots, X_{i-1}^*, X_i, 0, 0, \ldots\right)\big]$$
Output: $(X_1^*, \ldots, X_n^*) \in M$
```

Algorithm 3: A greedy approximation algorithm.

$X_1^*, \ldots, X_{i-1}^*$, but she incorrectly assumes that her eventual outcome x will be zero-valued in all dimensions $j > i$.

Despite appearing naive, the greedy algorithm guarantees a $1/2$-approximation when the objective function $\bar{u}$ satisfies a diminishing returns property. Roughly, an decisionmaker that prefers outcome x to x' should not prefer $x + x''$ to $x' + x''$ *even more* after she is given a lump sum of x''.

Definition 25. *The objective function $\bar{u}$ features* diminishing returns *if*

$$\bar{u}(x) - \bar{u}(x') \geq \bar{u}(x + x'') - \bar{u}(x' + x'') \quad \forall x, x', x'' \in \mathcal{X}$$

Johnson (1974) showed that a similar greedy algorithm guarantees a $1/2$-approximation for MAX 2-SAT. His proof applies almost immediately to this setting when $\bar{u}$ is the maximum objective function $\bar{u}(x) = \max_i x_i$. It turns out that this result holds more generally. I show that this result requires only two properties of the objective function: that $\bar{u}$ is non-decreasing and has diminishing returns.

Lemma 14. *Let objective function $\bar{u}$ be non-decreasing with diminishing returns, and efficiently computable. Then the greedy algorithm (3) guarantees a $1/2$-approximation.*

Theorem 1 follows immediately from lemmas 14 and 13. Furthermore, these results identify a large class of objective functions $\bar{u}$ in which theorem 1 holds. These are objective functions $\bar{u}$ that are symmetric, strictly increasing, and ϵ-sublinear, with diminishing returns.

There are many natural objectivey functions that all of these assumptions. For example, consider objective functions of the form

$$\bar{u}(x) = f\left(\sum_{i=1}^{n} x_i\right)$$

where f is strictly increasing. These satisfy diminishing returns when f is concave, and are ϵ-sublinear as long as

$$f(z) = O(z^{1-\epsilon})$$

This requirement is not much stronger than strict concavity. For example,

$$f(z) = \sqrt{z}$$

is $1/2$-sublinear.

I conclude with a remark on the greedy algorithm and its interpretation. In principle, the greedy algorithm can be understood as maximizing expected "utility" with respect to the limit of a sequence of utility functions, i.e.

$$\lim_{\epsilon \to 0^+} \left[\bar{u}(x_1, 0, 0, \ldots) + \epsilon \cdot \bar{u}(x_1, x_2, 0, 0, \ldots) + \epsilon^2 \cdot \bar{u}(x_1, x_2, x_3, 0, 0, \ldots) + \ldots\right]$$

This limiting behavior reflects lexicographic revealed preferences and violates the expected utility axioms because it violates the continuity axiom.[26]

This raises a natural question: is theorem 4 driven by approximation algorithms that *could* be rationalized, if only we dropped the continuity axiom? The answer is no; the

[26] For a survey of lexicographic choice under uncertainty, see Blume et al. 1989.

greedy algorithm is not the only approximation algorithm, and it does not appear to be the best one. For example, the algorithm in Corollary 3 weakly dominates the greedy algorithm and violates the weak axiom of revealed preference (since the utility function varies based on the menu).

Likewise, in appendix **??** gives a randomized algorithm that guarantees a $(1 - {}^{1}\!/_{e})$ approximation in the special case of the maximum objective function. This is better than the ${}^{1}\!/_{2}$ approximation established in lemma 14 for the greedy algorithm. Moreover, because the algorithm is randomized, the decisionmaker makes stochastic choices. This is an even further departure from standard rationality assumptions than the greedy algorithm or the algorithm in Corollary 3.[27]

1.5.2 Proof of Lemma 13

In this subsection, I prove lemma 13, which shows that rational and weakly tractable choice correspondences can be poor approximations. I leave the proof of lemma 14 to the appendix.

Let c be a rational and weakly tractable choice correspondence with revealed utility function u. Let $\bar{u}$ be a payoff function that is symmetric, ϵ-sublinear, and strictly increasing. I want to construct a menu M where $c(M)$ performs poorly relative to the optimal choice, according to $\bar{u}$.

The first step is to simplify the agent's behavior by focusing on coordinates i over which the revealed utility function u is additively separable. To do this, I use the concept of graph coloring.

Definition 26. *Let G be an undirected graph. Let C be a set of colors.*

1. *A C-*coloring *of G is map f from nodes $i \in \{1, \ldots, n\}$ to colors in C where adjacent nodes have different colors: if nodes i and j share an edge then $f(i) \neq f(j)$.*

[27] Technically, my model assumes that the algorithm generating the decisionmaker's choices is deterministic. But I show in appendix **??** that my results do not change if the decisionmaker is allowed to randomize.

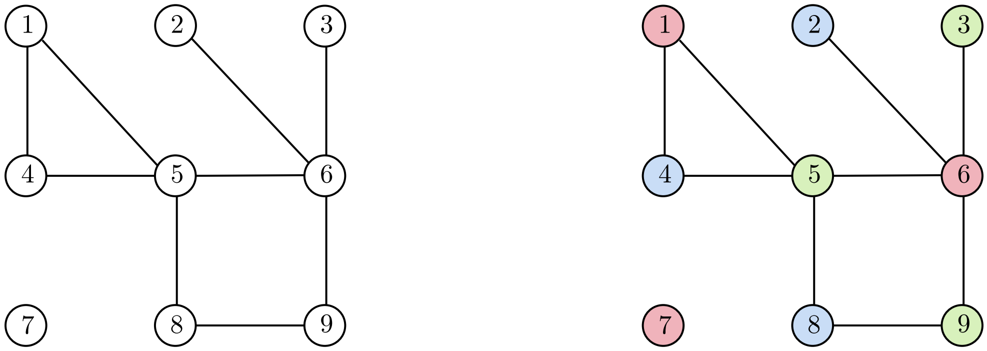


Figure 1.13: This diagram depicts a graph with a chromatic number of 3. On the left is the original graph. On the right is the graph where each node is assigned a color from the set $\{\text{red}, \text{blue}, \text{green}\}$.

2. *The* chromatic number *of G is the size of the smallest set C such that a C-coloring exists.*

Figure 1.13 illustrates. Let N be the set of nodes i that are assigned the most common color in a minimal coloring of the inseparability graph $G_n(u)$. For example, in figure 1.13, we could define $N = \{1, 6, 7\}$ or $N = \{2, 4, 8\}$ or $N = \{3, 5, 9\}$ because all three colors have the same number of nodes. Let S be the set of nodes i where $i \in N$ and

$$u(\underbrace{0, \ldots, 0}_{i-1 \text{ times}}, 1, 0, 0, \ldots) \geq u(0, 0, \ldots)$$

If u is nondecreasing, then $S = N$. In general, S may be a strict subset of N. For now, assume that at least half the elements of N are in S, i.e. $|S| \geq |N|/2$. The other case is even easier, and I return to it at the end of the proof.

Let d be the number of nodes in S. For convenience, let $S = \{1, \ldots, d\}$ be the first d nodes. This is without loss of generality because $\bar{u}$ is symmetric.

Restricting attention to nodes i with the same color is useful because those nodes can be separated from one another. Formally, I claim that there exist functions $u_1, \ldots, u_d$ such that, for any d-dimensional consequence x,

$$u(x) = \sum_{i=1}^{d} u_i(x_i)$$

This follows from the fact that any two nodes $i, j \in S$ cannot share an edge in the inseparability graph $G_n(u)$. Otherwise, they could not have the same color according to a valid C-coloring. Since none of the first d nodes share an edge, the inseparability graph $G_d(u)$ is empty, and therefore u is additively separable over d-dimensional consequences x.

Next, I construct a product menu M where the choice correspondence c will perform poorly. For all dimensions $i > d$, let $M_i = \{0\}$ consist of a single partial lottery that always returns $x_i = 0$. For all dimensions $i \leq d$, let $M_i = \{X_i^G, X_i^B\}$ consist of two partial lotteries. For all $i \leq d$, let

$$X_i^G(\omega) = \begin{cases} 1 & \omega \in \left[\frac{i-1}{d+2}, \frac{i}{d+2}\right) \\ 0 & \text{otherwise} \end{cases} \qquad X_i^B(\omega) = \begin{cases} 1 & \omega \geq \frac{d}{d+2} \\ 0 & \text{otherwise} \end{cases}$$

Note that X_i^B delivers a higher expected value than X_i^G, but the partial lotteries X_i^G are negatively correlated with each other while the partial lotteries X_i^B are positively correlated.

I claim that the choice correspondence c will choose the partial lottery X_i^B over X_i^G for each i. Compare the expected utility of X_i^B, i.e.

$$\mathrm{E}\left[u_i\left(X_i^B\right)\right] = \left(\frac{2}{d+2}\right) \cdot u_i(1) + \left(\frac{d}{d+2}\right) \cdot u_i(0)$$

with expected utility of X_i^G, i.e.

$$\mathrm{E}\left[u_i\left(X_i^G\right)\right] = \left(\frac{1}{d+2}\right) \cdot u_i(1) + \left(\frac{d+1}{d+2}\right) \cdot u_i(0)$$

Since $u_i(1) \geq u_i(0)$ for all $i \in S$, it is clear that X_i^B is better according to u. Because u is additively separable across $i \in S$, the decisionmaker ignores the correlation across dimensions i.

Unfortunately, always choosing X_i^B is suboptimal from the perspective of the payoff function $\bar{u}$. The expected payoff will be

$$\begin{aligned}\left(\frac{2}{d+2}\right) \cdot \bar{u}(\underbrace{1, \dots, 1}_{d \text{ times}}, 0, 0) + \left(\frac{d}{d+2}\right) \cdot \bar{u}(0, 0, \dots) &= \left(\frac{2}{d+2}\right) \cdot O(d^{1-\epsilon}) \\ &= O(d^{-\epsilon}) \qquad (1.13)\end{aligned}$$

where the first equality follows from sublinearity and the fact that $\bar{u}(0, 0, \dots) = 0$. However, the expected payoff from always choosing X_i^G is

$$\begin{aligned}\left(\frac{2}{d+2}\right) \cdot \bar{u}(0, 0, \dots) + \sum_{i=1}^{d} \left(\frac{1}{d+2}\right) \cdot \bar{u}(\underbrace{0, \dots, 0}_{i-1 \text{ times}}, 1, 0, 0, \dots) &= \left(\frac{d}{d+2}\right) \cdot \bar{u}(1, 0, 0, \dots) \\ &= \Theta(1) \qquad (1.14)\end{aligned}$$

The first equality follows from symmetry and the fact that $\bar{u}(0, 0, \dots) = 0$. The second equality follows from the fact that $\bar{u}$ is strictly increasing, and therefore $\bar{u}(1, 0, 0, \dots) > \bar{u}(0, 0, \dots)$.

Divide (1.13) by (1.14) to show that the approximation ratio is at most

$$\text{APX}_n^{\bar{u}}(c) = O(d^{-\epsilon})$$

However, the lemma claimed that the approximation ratio is at most $O(n^{-\epsilon})$. Therefore, I still need to show that $n = \tilde{O}(d)$. To do this, I need to introduce one more property of graphs, closely related to contraction degeneracy (21).

Definition 27. *Let G be an undirected graph. The degeneracy* $\mathrm{dgn}(G)$ *is the smallest number d such that every subgraph of G has a node with degree less than or equal to d.*

Szekeres and Wilf (1968) show the chromatic number of a graph G is at most $1+\mathrm{dgn}(G)$. By the pigeonhole principle, the number of nodes $i \in N$ must be at least

$$\frac{n}{1+\mathrm{dgn}(G)}$$

By assumption, at least half of these nodes i satisfy $u_i(1) \geq u_i(0)$. Therefore, the number of nodes $i \in S$ is at least

$$d \geq \frac{n}{2+2\mathrm{dgn}(G)} \tag{1.15}$$

Finally, I can bound the degeneracy as follows.

$$\begin{aligned}\mathrm{dgn}(G_n(u)) &\leq \mathrm{cdgn}(G_n(u)) \\ &= \tilde{O}\left(\mathrm{Had}(G_n(u))\right) \\ &= O(\log n) \end{aligned} \tag{1.16}$$

The first line follows immediately from the definitions of degeneracy and contraction degeneracy (21). The second line follows from lemma 7. The third line follows from theorem 2. Note that this third line, which is absolutely essential, is the only time I use the fact that c is weakly tractable.

Combining (1.15) and (1.16) gives

$$d \geq \frac{n}{O(\log n)} = \tilde{O}(n)$$

which is what I sought to show.

All that remains is to consider the case where $u_i(0) > u_i(1)$ for more than half of the nodes $i \in N$. I claimed this case was even easier. Redefine S as the set of all nodes i where $i \in N$ and $u_i(0) > u_i(1)$. Redefine $X_i^B(\omega) = 0$ for all $\omega \in [0,1]$, and let X_i^G be defined as above. The choice correspondence c will still choose X_i^B over X_i^G. The decisionmaker's expected payoff

$$\mathrm{E}\left[\bar{u}\left(X_1^B, \ldots, X_n^B\right)\right]$$

can only decrease relative to my original construction, but the expected payoff

$$\mathrm{E}\left[\bar{u}\left(X_1^G, \ldots, X_n^G\right)\right]$$

will stay the same. The rest of my argument goes through verbatim.

1.6 Related Literature

This work contributes to three research efforts. First, it contributes to the literature on bounded rationality, which incorporates cognitive limitations into economic models. Second, it contributes to the subfield of economics and computation, which uses computational complexity to gain insight into economic phenomena. Third, it contributes to the sizable literature in behavioral economics on choice bracketing and related phenomena.

Bounded Rationality. There is a longstanding effort to incorporate bounded rationality in economic modeling. Here, I emphasize work that is methodologically similar to mine.

Echenique et al. (2011) also consider a model of computationally-constrained consumer choice. Thet develop a "revealed preference approach to computational complexity" in response to alternative approaches used in prior work. They propose only ruling out utility functions for which maximization is computationally hard and evaluate the implications for observed behavior. Their definition of tractability is similar to mine: a utility function over bundles is tractable if there exists a polynomial-time algorithm that maximizes the consumer's utility subject to budget constraints. The main difference in our definitions that Echenique et al. (2011) interpret choices as a dataset, whereas I work with a choice correspondence.[28]

Directionally, my paper is very much aligned with Echenique et al. (2011). The results are quite different, however. Echenique et al. (2011) consider a classic model of budget-constrained consumer choice. They show that tractability does not meaningfully constrain behavior: any rationalizable dataset is consistent with a tractable utility function. In contrast, I consider a model of choice under risk and show that tractability has significant implications for behavior.

Gilboa, Postlewaite, et al. (2021) also study computationally-constrained consumer choice, but take a different approach. In their model, the utility derived from consumption is a property of the menu that the agent faces, in contrast to the revealed preference approach that Echenique et al. (2011) and I take. In this formulation, the authors show that the consumer's problem is NP-hard. The authors argue that, as a result, consumers may turn to heuristics like mental accounting. My results show that these kinds of arguments for behavioral heuristics can actually be formalized, by imposing computational tractability as an axiom.

Other researchers have used Turing machines to study the computability of choice (Richter and Wong 1999a), equilibria (Richter and Wong 1999b), and repeated game strategies (Anderlini and Sabourian 1995). Computability is a much weaker property than computational tractability. It asks whether behavior can be generated by a Turing machine, whereas tractability asks whether behavior can be generated by a Turing machine with a reasonable runtime.

[28] In their model, the dataset consists of observed choices $c(M_1), \ldots, c(M_k)$ for k distinct menus. In my model, the choice correspondence c describes the choices $c(M)$ that the agent would make, *if* she were presented with the menu M. The interpretation is similar to the potential outcomes framework in econometrics.

Beyond Turing machines, researchers have used specialized models that impose more structure on how the agent generates her choices. Some of these models are borrowed from computer science, like perceptrons (Rubinstein 1993) and finite automata (e.g. Rubinstein 1986, A. Wilson 2014). Other models are original, and capture forms of procedural reasoning (e.g. Mandler et al. 2012, Mandler 2015), costly reasoning (e.g. Gabaix et al. 2006, Ergin and Sarver 2010), incomplete reasoning (e.g. Lipman 1999, Jakobsen 2020), or limited attention (e.g. Gabaix 2014).

It is tempting to use more specialized models of computation in order to obtain stronger results (although the representations in this paper are already quite strong). After all, the Turing machine is a general model of computation, and computations that are easy for a computer may be challenging for a human. However, most individuals and firms have access to computers and are free to use them to support their decisionmaking. It would be odd if a theory of computationally-constrained choice could not account for decisionmakers who take advantage of modern computing power.

Economics and Computation. The notion that computational constraints bind on economic phenomena is widely accepted in the interdisciplinary subfield of economics and computation. Most famously, computational complexity theory has had a big impact on mechanism design, where optimal mechanisms are often intractable (e.g. Nisan and Ronen 2001). However, both economists and computer scientists have applied this framework to many other topics, like equilibrium (e.g. Gilboa and Zemel 1989, C. Daskalakis et al. 2009), learning (Aragones et al. 2005), social learning (Hązła et al. 2021), testing (Fortnow and Vohra 2009), and rationalizing choices (Apesteguia and Ballester 2010). Most of these papers, like mine, rely on an expansive interpretation of the Church-Turing thesis that uses the Turing machine to model behavioral or social processes.

This subfield also inspired the choice trilemma, which takes the perspective of approximation algorithms in order to critique the expected utility axioms. Feng and Hartline (2018) take the same perspective to critique the revelation principle in mechanism design. In prior-independent settings, they show that designers may obtain a better approximation to their objective if they are willing to use non-revelation mechanisms. Specifically, they find that the revelation gap is between 1.013 and e in the setting they study, whereas a value of 1 means no gap. In the settings identified in Theorem 4, the approximation gap grows to ∞ as $n \to \infty$.

Choice Bracketing and Related Phenomena. There is considerable empirical support for choice bracketing and other forms of narrow framing, like mental accounting (Thaler 1985) and myopic loss aversion (Benartzi and Thaler 1995). Read et al. (1999) coined the term "choice bracketing" as a way to explain behavior observed in prior experiments (e.g. Tversky and Kahneman 1981). Since then, behavioral experiments have highlighted potential factors that influence choice bracketing, including choice complexity (Stracke et al. 2017), cognitive ability (Abeler and Marklein 2016), framing (Brown et al. 2021), and the desire for self control (Koch and Nafziger 2019).

Choice bracketing and other forms of narrow framing seem to be economically meaningful. Observational studies have found evidence for narrow framing in taxi services (e.g. Camerer et al. 1997; Martin 2017), eBay bidding (Hossain and Morgan 2006), savings behavior Choi et al. (2009), food stamp expenditures (Hastings and Shapiro 2018), and MBA admissions (Simonsohn and Gino 2013). Others have proposed narrow framing as an explanation for stock market non-participation (Barberis et al. 2006) and the equity premium puzzle (Benartzi and Thaler 1995).

Moreover, choice bracketing can lead to surprising behavior. For example, Rabin and Weizsäcker (2009) consider a decisionmaker choosing from a product menu. Their model specializes mine by assuming that partial lotteries X_i, X_j are independent of each other and that the decisionmaker cares about total income, i.e. $X_1 + \ldots + X_n$. Unless the decisionmaker's preferences satisfy constant absolute risk aversion, the authors show that she will violate first order stochastic dominance in some menu. Then they provide experimental evidence that many decisionmakers narrowly bracket their choices to the point where they choose dominated lotteries.

In light of this empirical evidence, researchers have proposed various theories of choice bracketing and mental accounting. Zhang (2021) provides an axiomatic foundation for narrow choice bracketing, by relaxing the independence axiom and introducing an axiom of correlation neglect. Lian (2020) conceptualizes a decisionmaker as a narrow thinker if she uses different information to make different decisions, and formulates a model of rational inattention where the decisionmaker chooses what information to use for each decision. Similarly, Kőszegi and Matějka (2020) develop a model of rational inattention to understand mental accounting. Finally, Koch and Nafziger (2016) takes a different approach, justifying choice bracketing as a commitment device.

1.7 Conclusion

In this paper, I propose a new theoretical framework for studying computationally-tractable choice. Specifically, I apply a remarkably powerful model of computation, the Turing machine, to a quite general model of choice under risk. With these ingredients, I address two problems. First, I provide a formal justification for the claim that computational constraints lead to forms of choice bracketing (Theorems 1, 2, 3). Second, I provide a formal justification for behavior that violates the expected utility axioms (Theorem 4). I summarize these results as a choice trilemma (Figure 1.3).

These results show the potential value of computational tractability to economic theory. First, by recognizing computational constraints as binding on the world around us, we can make sharper predictions about economic behavior. The first step to realizing this potential is to formulate computational constraints correctly, as an axiom that restricts how choices vary across counterfactual menus. Then we can impose tractability on top of other assumptions, like rationality, to obtain useful representations and sharper predictions. Second, computational tractability clarifies the meaning of *other* assumptions in our models. For example, it seems natural to assume that investors only care about total income, but not if this implies risk neutrality. More generally, it seems natural to assume that choices reveal preferences that satisfy the expected utility axioms, but not if this means that she is making choices that are objectively worse than they need to be (Theorem 4).

Since P. A. Samuelson (1938), economic theory has typically associated rationality with "exact maximization of revealed preferences". The choice trilemma suggests that, instead, the decisionmaker should prioritize "approximate maximization of hedonic preferences" where hedonic preferences reflect the decisionmaker's intrinsic objective function (if she has one). In the the presence of computational constraints, revealed preferences cannot match hedonic preferences unless the objective function is Hadwiger separable. For that reason, it is generally not clear why revealed preferences should exist at all. Presumably, the decisionmaker's priority is to perform well according to her hedonic preferences (if they exist), irrespective of whether an outside observer would be able to make sense of her choices (see e.g. Manski 2011). Theorem 4 sharpens this argument by showing that, in fact, it is in the decisionmaker's best interest to make choices that do not reveal preferences that satisfy the expected utility axioms.

To develop alternative definitions of rationality that are more compatible with computational constraints, it may be useful to learn from the "beyond worst-case analysis" literature in computer science (see e.g. Roughgarden 2021). It is common in computer science to evaluate algorithms on their runtime in the worst-case instance. Consider an algorithm A that takes one minute to solve 99% of inputs and one year for 1% of inputs (assuming a measure over inputs). The worst-case runtime is one year. But a decisionmaker that does not have a year to deliberate might use another algorithm A': see whether A returns an answer within a minute, otherwise choose something suboptimal. This is optimal 99% of the time, suboptimal 1% of the time, and takes about a minute. Perhaps A' should be regarded as rational, even though strictly speaking it cannot be rationalized.

Chapter 2

Mechanisms for No-Regret Agents[1]

2.1 Introduction

Mechanism design is a branch of economic theory concerned with the design of social institutions. It encompasses a wide range of phenomena that have historically been of interest to economists, including, but not limited to, auctions (Myerson 1981; Vickrey 1961), matching markets (Gale and Shapley 1962; A. E. Roth 1982), taxation (Mirrlees 1971), contracts (Ross 1973; Spence and Zeckhauser 1971), and persuasion (Kamenica and Gentzkow 2011).

Despite this field's potential, it is often unclear whether and how mechanisms derived from economic theory can be implemented in practice. In particular, one modeling practice stands out as a barrier to implementation: the *common prior* assumption. Many mechanism design problems are only interesting in the presence of uncertainty, and this uncertainty is typically modeled as stochasticity. The *state* of the world is drawn according to some distribution and, importantly, the distribution is commonly known by the designer and all participants in the mechanism.[2]

This paper will dispense with the common prior assumption. In its place, we consider a model of adversarial online learning where the principal and a single agent are learning about the state, over time, using data. The static mechanism design problem is a Stackelberg game of incomplete information. The principal chooses a policy p, the agent chooses a response

[1] Joint work with Jason Hartline and Aleck Johnsen. See Camara et al. (2020) for a published version.

[2] This assumption is limiting in two ways. First, mechanisms based on a common prior may not be practicable, because they rely on knowledge that a real-world designer is unlikely to possess. Second, even if the designer knows the distribution (resp. has beliefs), the participants may not arrive with the same knowledge (resp. share those beliefs).

r, nature chooses a state y, and payoffs are realized. In the online problem, this game is repeated T times, where state y_t is revealed at the end of period t. The sequence of states is arbitrary and the principal's mechanism should perform well without prior knowledge of the sequence. The principal's present choices can affect the agent's future behavior; this makes mechanism design a reinforcement learning problem in our model.

In the absence of distributional assumptions, standard restrictions on the agent's behavior, like Bayesian rationality, become toothless. In its place, we define *counterfactual internal regret* (CIR) and assume that the agent obtains low CIR. This is an ex post definition of rationality that includes Bayesian rationality (with a well-calibrated prior) as a special case. We develop data-driven mechanisms that are guaranteed to perform well under our behavioral assumptions. Specifically, we prove upper bounds on the principal's regret from following our mechanism, relative to the single fixed policy that performs best in hindsight. Our results take the form of reductions from the principal's problem to robust versions of static mechanism design with a common prior.

Running Example. Bayesian persuasion is a model of strategic communication, due to Kamenica and Gentzkow (2011). It has received considerable attention from economists and, more recently, algorithmic game theorists (e.g. Dughmi and Xu 2016, Cummings et al. 2020). It is a useful test case for our framework because (a) it is interesting even with only one agent, (b) the optimal solution varies with the agent's beliefs, and (c) it has the potential to be widely applicable.[3]

Our running example is adapted from Kamenica and Gentzkow (2011). A drug company (the principal) seeks approval from a regulator (the agent) for a newly-developed drug. The state $y \in \{\text{High}, \text{Low}\}$ describes the drug's quality. Neither the regulator nor the

[3] Bayesian persuasion has been used to study a wide range of topics, including recommendation systems (Mansour et al. 2016), traffic congestion (Das et al. 2017), congested social services (Anunrojwong et al. 2020), financial stress-testing (Goldstein and Leitner 2018), and worker motivation (Ely and Szydlowski 2020).

company know the quality in advance. The company needs to design a clinical trial that will generate (possibly noisy) information about the drug's quality. Roughly, a trial p specifies the probability $p(m, y)$ of sending a message m to the regulator, conditional on the drug quality y. Informally, the message describes the outcome of the trial. After hearing the message, the regulator decides whether to approve the drug. The regulator receives a payoff if it approves a high-quality drug or rejects a low-quality drug. The company receives a payoff if the regulator approves, regardless of quality. Its challenge is to design a clinical trial that convinces the regulator to approve as many drugs as possible.

To predict behavior in incomplete-information games, we need to make assumptions about how the agents deal with uncertainty. The common prior is one such assumption. In our running example, the common prior would specify a probability $q \in [0, 1]$ that the drug is high quality. Consider the case $q = 1/3$. If the company does not run a trial – e.g. it recommends "approve" in every state – the regulator would never approve, as the drug is more likely to be low quality than high quality ex ante. If the company runs the most thorough trial possible – e.g. it recommends "approve' if and only if the drug is high quality – the regulator would approve with probability 1/3. Finally, consider the optimal trial. The optimal trial always recommends "approve" if the drug is high quality. If the drug is low quality, it recommends "approve" and "reject" with equal probability. After hearing "approve", the regulator's posterior puts equal weight on both states, and so it might as well approve. Here, the regulator approves with probability 2/3.

Online Mechanism Design. In our model, both the company and the regulator would be learning about drug quality over time. New drugs arrive sequentially. For each drug, the company designs a clinical trial and generates a message. The regulator hears the message and decides whether to approve. Regardless of whether the drug is approved, both parties eventually learn the drug's true quality, and the next drug arrives. The company's strategy, called a *mechanism*, maps the drug (i.e. state) history and the approval decision (i.e. response) history to a trial for the current drug. The regulator's strategy, called a learning algorithm or *learner*, maps the drug quality history and the trial (i.e. policy) history to an approval decision for the current drug. This model is *online* because the company and regulator must make decisions while the drugs are still arriving. It is *adversarial* in the sense that we impose no assumptions on the sequence of drugs, and so any results (e.g. claiming that a mechanism performs well) must hold for all such sequences.

The company's problem is to develop a mechanism that performs as well as the best-in-hindsight trial. That is, the company should not regret following its mechanism relative to any simple alternative where it picks the same trial p in every period. To evaluate what would have happened under an alternative sequence of trials, the company must take into account how the regulator's behavior would have changed. Therefore, the company faces a reinforcement learning problem and its benchmark corresponds to the notion of *policy regret* in the literature on bandit learning with adaptive adversaries (e.g. R. Arora, O. Dekel, et al. 2012). In that setting, R. Arora, O. Dekel, et al. (2012) show that guaranteeing sublinear (policy) regret is generally impossible.[4] This fact precludes a simple solution to the company's problem; we must constrain the regulator's behavior.[5]

The standard way to constrain the regulator/agent's behavior – i.e. to capture "self-interest" in the absence of a meaningful notion of ex ante optimality – is to impose upper bounds on the agent's regret. This will be our approach as well. We build on existing

[4] R. Arora, O. Dekel, et al. (2012) obtain positive results when the adversary satisfies a bounded memory assumption. Ryabko and Hutter (2008) obtain positive results under a different kind of assumption, that the environment is sufficiently "forgiving" of mistakes. These papers reflect two prominent approaches in reinforcement learning: (a) restricting attention to Markov decision processes, and (b) assuming an ability to "reset" the problem (Kearns et al. 1999).

[5] R. Arora, Dinitz, et al. (2018) consider policy regret in a repeated game and use the self-interest of the adaptive adversary to motivate behavioral restrictions. This is reminiscent of a literature on multi-agent reinforcement learning when the state is Markovian (Buşoniu et al. 2010; Hu and Wellman 1998; Littman 1994; Uther and Veloso 2003). Unlike these papers, we do not have the ability in our model to advise all participants simultaneously.

no-regret assumptions, in ways that are intended to refine and better motivate those assumptions.

No-Regret Agents. Two notions of regret have been used historically: external and internal (or swap) regret (ER and IR). For example, Nekipelov et al. (2015) show how ER bounds combined with bidding data can be used to partially identify bidder valuations in a dynamic auction. Braverman et al. (2018) consider a dynamic pricing problem against no-ER agents.[6] Their analysis is generalized by Deng et al. (2019), who study repeated Stackelberg games of complete information. Furthermore, the literature on no-regret learning in games has established that if agents satisfy a no-ER (resp. no-IR) property in a repeated game, the empirical distribution of their actions will converge to a coarse correlated equilibrium (resp. correlated equilibrium) (Blum, Hajiaghayi, et al. 2008; Foster and Vohra 1997; Hart and Mas-Colell 2001; Hartline, Syrgkanis, et al. 2015).

Both ER and IR can be thought of as "non-policy" regret, because they do not take into account how the agent's behavior affects the behavior of others. The justification for these regret bounds is that (a) they are satisfied by well-known learning algorithms (see e.g. Littlestone and Warmuth 1994 for ER), and (b) they generalize optimality conditions associated with a stationary equilibrium. Nonetheless, these regret bounds can be problematic. Effectively, they assume that agents are (a) sophisticated enough to obtain low non-policy regret, but (b) not aware that their true objective is policy regret. Keep in mind that an agent who minimizes policy regret can easily obtain high non-policy regret, and thereby violate the regret bounds.

To avoid this problem, the principal in our model can commit to a mechanism that is

[6] In their model, the agent is "learning" an appropriate response to the principal's pricing strategy. If the agents use naive mean-based learners, Braverman et al. (2018) provide a mechanism that extracts the full surplus. In particular, the agent fails to anticipate the mechanism that the principal is using. As they point out, this leads to odd behavior: the agent may purchase goods at a price exceeding her valuation. In our setting, the agent does not face uncertainty with respect to the mechanism; instead, she faces uncertainty with respect to the state sequence.

nonresponsive to the agent's behavior: the policy p_t depends on the state history but not on the agent's response history. When mechanisms are nonresponsive, non-policy regret and policy regret coincide for the agent. Then, bounds on the agent's regret are permissive assumptions that allow a wide range of sophisticated and self-interested behavior, including Bayesian rationality. Keep in mind, there is no need to resort to responsiveness if nonresponsive mechanisms tightly bound the principal's regret.[7]

Counterfactual Internal Regret. Without constraints on the agent's behavior, an early mistake by the principal can result in a permanent, undesirable shift in the agent's behavior. As we will see, this can occur when the agent behaves as if she has additional information about the state of the world that is not accounted for in our description of the model. The agent can make the principal's problem infeasible if she is willing to exploit her information selectively, i.e. based on the principal's choice of policies. Unfortunately, neither no-ER nor no-IR assumptions can rule out selective use of information.

Our notion of rationality requires the agent to fully and consistently exploit her information, regardless of the principal's chosen policies. Existing benchmarks like external and internal regret cannot capture this requirement. To see why, it helps to consider the fable of the tortoise and the hare. Both animals have an hour to traverse a one-mile track. For the tortoise, this requirement is feasible and binding: finishing in time means hustling, without substantial breaks or detours. For the hare, however, the requirement is hardly restrictive: it may stop for a break, walk rather than run, or even run around in circles while still finishing the race in time. Benchmarks like external or internal regret imply reasonable behavior for an uninformed agent (i.e. the tortoise). But for an informed agent (i.e. the hare), these

[7] This approach seems spiritually similar to that of Immorlica, Mao, et al. (2020), who develop mechanisms that incentivize efficient social learning. By restricting attention to simple disclosures (i.e. unbiased subhistories), they significantly simplify the agents' inferential problem and can motivate a permissive notion of frequentist rationality. Having restricted disclosure in this manner, they nonetheless design mechanisms with optimal rates of convergence.

benchmarks are easy enough to satisfy that it may engage in all kinds of frivolous behavior – possibly to the detriment of the principal.

The solution to our analogy is to strengthen the hare's benchmark. If the hare has to traverse the track in three minutes, it needs to hustle, like the tortoise. Similarly, if the agent has to obtain no-regret with her information as additional context, this would preclude the kind of frivolous behavior that makes the principal's problem infeasible. Of course, setting this benchmark requires us to know the nature and quality of the agent's information, just as we needed to know the top speed of the hare. The idea behind counterfactual internal regret is that we can identify the agent's information with her past behavior under counterfactual mechanisms. Intuitively, any information that is useful should eventually reveal itself through variation in behavior.

Main Results. This paper considers three variations on our model: one where the principal knows the agent's information, one where the agent has no private information, and one where the agent may have private information. In each case, we propose a mechanism and bound on the principal's regret in terms of the agent's counterfactual internal regret (CIR).

Our first mechanism is intended as a warmup. It requires oracle access to the agent's information and has poor performance in finite samples, but avoids some complications associated with information asymmetry between the principal and agent. First, the mechanism produces a calibrated forecast of the state in the current period using off-the-shelf algorithms, using the oracle as additional context for the forecast. Then, it chooses the worst-case optimal policy in a (hypothetical) ϵ-robust version of the common prior game. In that game, the agent's response only needs to be ϵ-approximately optimal, and the mechanism substitutes its forecast for the prior.

Theorem 5 bounds the principal's regret under this mechanism, under some restrictions on the stage game. Suppose there are $n_{\mathcal{Y}}$ states, $n_{\mathcal{P}}$ policies, and $n_{\mathcal{R}}$ responses. Fix a parameter $\epsilon > 0$ (controlling robustness) and $\delta > 0$ (controlling the fineness of a grid). Our bound is

$$\underbrace{O(\epsilon)}_{\text{cost of }\epsilon\text{-robustness}} + \frac{1}{\epsilon}\left(\underbrace{O(\text{CIR})}_{\text{agent's regret}} + \underbrace{\tilde{O}\left(\frac{\delta^{1-n_{\mathcal{Y}}} n_{\mathcal{Y}} n_{\mathcal{R}}^{2n_{\mathcal{P}}}}{T^{1/4}}\right)}_{\text{forecast miscalibration}} + \underbrace{O\left(\delta^{1/2}\right)}_{\text{approximation error}} \right) \tag{2.1}$$

If the agent satisfies no-CIR, i.e. CIR → 0 as $T \to \infty$, then the principal's regret vanishes in T as long as $\epsilon, \delta \to 0$ at the appropriate rates. Moreover, the principal's average payoffs converge to a natural benchmark: what he would have obtained in a stationary equilibrium of the repeated game with a common prior (the empirical distribution conditioned on agent's information).

Our second mechanism applies when the agent is as uninformed as the principal. This mechanism is identical to the first, except its forecast does not use information revealed by the learner. We formalize "uninformedness" by assuming that the agent's external regret is non-negative (in conjunction with no-CIR). Theorem 6 bounds the principal's regret under this mechanism, under some additional restrictions on the stage game. Our bound is

$$\underbrace{O(\epsilon)}_{\text{cost of }\epsilon\text{-robustness}} + \frac{1}{\epsilon}\left(\underbrace{O(\text{CIR})}_{\text{agent's regret}} + \underbrace{\tilde{O}\left(\frac{\delta^{1-n_{\mathcal{Y}}} n_{\mathcal{Y}}}{T^{1/4}}\right)}_{\text{forecast miscalibration}} + \underbrace{O\left(\delta^{1/2}\right)}_{\text{approximation error}} \right) \tag{2.2}$$

Compared to (2.1), this drops the exponential dependence on the number $n_{\mathcal{P}}$ of policies. This is because the principal's forecast does not need to take into account the agent's information, which significantly reduces the forecast miscalibration in finite samples.

Our third mechanism applies even when the agent is more informed than the principal. Here, we consider an "informationally robust" version of the stage game, due to Bergemann and Morris (2013), where the agent receives a private signal from an unknown information structure. Like before, we formulate an ϵ-robust version of this game, where the agent's response need only be ϵ-approximately optimal. Our mechanism is identical to the second mechanism, except that it chooses the worst-case optimal policy in the ϵ-informationally-robust game instead of the ϵ-robust game.

Theorem 7 bounds the principal's regret under this mechanism, under some restrictions on the stage game. Let $\hat{\pi}_T$ denote the empirical distribution of states $y_{1:T}$. Given a common prior π, let $\nabla(\pi)$ be the difference between the principal's maxmin payoff and his maxmax payoff across all possible information structures. Roughly, our bound is

$$\underbrace{\nabla(\hat{\pi}_T)}_{\text{cost of informational robustness}} + \underbrace{O(\epsilon)}_{\text{cost of }\epsilon\text{-robustness}} + \frac{1}{\epsilon}\left(\underbrace{O(\text{CIR})}_{\text{agent's regret}} + \underbrace{\tilde{O}\left(\frac{\delta^{1-n_{\mathcal{Y}}} n_{\mathcal{Y}}}{T^{1/4}}\right)}_{\text{forecast miscalibration}} + \underbrace{O\left(\delta^{1/2}\right)}_{\text{approximation error}}\right) \tag{2.3}$$

Unlike (2.1) and (2.2), the principal's regret does not vanish as $T \to \infty$. However, it is vanishing up to the cost of informational robustness $\nabla(\hat{\pi}_T)$ that would also be present under a common prior, if the agent were more informed than the principal.

Finally, although our focus is not on computational complexity, the reader should note that the computational tractability of our mechanisms will depend critically on our ability to solve robust mechanism design problems under a common prior. So, while our bounds on the principal's regret apply to a large class of games, evaluating tractability may require a case-by-case analysis.

Additional Related Work. Within computer science, many researchers share our goal of replacing prior knowledge in mechanism design with data. The literature on sample complexity in mechanism design allows the principal to learn the state distribution from i.i.d. samples (Balcan, Blum, Hartline, et al. 2008; Cole and Roughgarden 2014; J. Morgenstern and Roughgarden 2015; Syrgkanis 2017). Here, the data arrives as a batch rather than online, there is no repeated interaction and the question of responsiveness does not arise. However, there has also been work that applies online learning to auction design (e.g. C. Daskalakis and Syrgkanis 2016; Dudík et al. 2017) and Stackelberg security games (e.g. Balcan, Blum, Haghtalab, et al. 2015). Here, agents are either short-lived or myopic, whereas our agent is long-lived and potentially forward-looking.

These papers can avoid the agent's learning problem because they emphasize applications where the agent does not face uncertainty, or where truthfulness is a dominant strategy. In contrast, Cummings et al. (2020) and Immorlica, Mao, et al. (2020) study problems that are closer to our own, insofar as both the principal and the agent must learn from data. They impose behavioral assumptions that are suited for i.i.d. data, whereas our model generalizes to adversarial data.

Within economics, research has focused on relaxing prior knowledge, rather than replacing it entirely. Part of the literature on robust mechanism design relaxes the common prior to some kind of approximate agreement on the distribution (Artemov et al. 2013; Jehiel, Meyer-ter-Vehn, and Moldovanu 2012; Meyer-ter-Vehn and Morris 2011; Ollár and Penta 2017; Oury and Tercieux 2012). Our approach will suggest ϵ-robustness and ϵ-informational-robustness as alternatives to "approximate agreement".

Organization. Section 2.2 introduces the stage game and ϵ-robustness. Section 2.3 introduces the repeated game. Section 2.4 defines external, internal, and counterfactual internal regret. Section 2.4 presents our mechanism and regret bounds when the agent's learner is known. As preparation for the remaining results, section 2.5 introduces the stage game with private signals. Section 2.4 presents our mechanism and regret bounds when the agent is uninformed. Section 2.4 presents our mechanism and regret bounds when the agent may be more informed than the principal. Section 2.9 concludes with a discussion of open problems.

Appendix B.1 applies these results to two special cases: our running example, and a principal-agent problem. Appendix B.2 considers the complexity of the agent's learning problem. Appendix B.3 describes our forecasting algorithms in more detail. Appendix B.4 relaxes some of the restrictions on the stage game and generalizes our results. Appendix B.5 collects proofs.

2.2 Stage Game

Our model features three participants: a male principal, a female agent, and nature. As advertised, we are interested in a repeated interaction between these participants. To begin with, however, we describe the stage game, which will constitute a single-round of the repeated game. In the stage game, the principal moves first and commits to a policy $p \in \mathcal{P}$. Next, the agent observes the policy p and then chooses a response $r \in \mathcal{R}$. Utility functions depend on the response r, the policy p, and an unknown state of the world $y \in \mathcal{Y}$, chosen by nature. Formally, the agent's utility function is $U : \mathcal{R} \times \mathcal{P} \times \mathcal{Y} \to [0, 1]$ while the principal's utility function is $V : \mathcal{R} \times \mathcal{P} \times \mathcal{Y} \to [0, 1]$.

Assumption 7 (Regularity)**.** *We impose the following regularity conditions.*

1. *The state space $\mathcal{Y}$ is finite.*
2. *The response space $\mathcal{R}$ is a compact space with metric $d_{\mathcal{R}}$.*
3. *The policy space $\mathcal{P}$ is a compact space with metric $d_{\mathcal{P}}$.*
4. *The utility U is equi-Lipschitz continuous in (r, p) for Lipschitz constants $K^U_{\mathcal{R}}$ and $K^U_{\mathcal{P}}$, i.e.*

$$\forall y \in \mathcal{Y} : \quad |U(r, p, y) - U(\tilde{r}, \tilde{p}, y)| \leq K^U_{\mathcal{R}} d_{\mathcal{R}}(r, \tilde{r}) + K^U_{\mathcal{P}} d_{\mathcal{P}}(p, \tilde{p})$$

5. *The utility V is equi-Lipschitz continuous in (r, p) for Lipschitz constants $K^V_{\mathcal{R}}$ and $K^V_{\mathcal{P}}$, i.e.*

$$\forall y \in \mathcal{Y} : \quad |V(r, p, y) - V(\tilde{r}, \tilde{p}, y)| \leq K^V_{\mathcal{R}} d_{\mathcal{R}}(r, \tilde{r}) + K^V_{\mathcal{P}} d_{\mathcal{P}}(p, \tilde{p})$$

Later on, we will use covers to convert infinite action spaces into discrete approximations. For example, our running example involved an infinite policy space.

Definition 28 (Covers)**.** *Let $\mathcal{X}$ be a metric space with metric $d_{\mathcal{X}}$. Generally, lower case letters x denote elements of $\mathcal{X}$ while upper case letters X denote subsets.*

1. *Fix $\delta_{\mathcal{X}} > 0$. Let the partition $\mathcal{C}_{\mathcal{X}}$ be a $\delta_{\mathcal{X}}$* cover *of $\mathcal{X}$. That is, for every set $X \in \mathcal{C}_X$, any two elements $x, \tilde{x} \in X$ must be within distance $\delta_{\mathcal{X}}$ of one another, i.e. $d_{\mathcal{X}}(x, \tilde{x}) < \delta_{\mathcal{X}}$.*

2. *To reduce notation, we also let $\mathcal{C}_{\mathcal{X}}$ denote a discretized subset of $\mathcal{X}$. That is, for each set $X \in \mathcal{C}_{\mathcal{X}}$, choose a unique $x \in X$ to represent X. In that case, we say $x \in \mathcal{C}_{\mathcal{X}}$.*

3. *Let $x \in \mathcal{X}$ and $\tilde{x} \in \mathcal{C}_{\mathcal{X}}$. We say that $\tilde{x}$ is the* discretization *of x if $x, \tilde{x}$ belong to the same subset $X \in \mathcal{C}_{\mathcal{X}}$.*

We will refer to covers $\mathcal{C}_{\mathcal{P}}$ of the policy space (with metric $d_{\mathcal{P}}$), $\mathcal{C}_{\mathcal{R}}$ of the response space (with metric $d_{\mathcal{R}}$), and $\mathcal{C}_{\Delta(\mathcal{Y})}$ of the state distributions $\Delta(\mathcal{Y})$ (with the l_1 metric).[8] Of course, if the underlying set $\mathcal{X}$ is finite to begin with, we can simply set $\delta_{\mathcal{X}} = 0$ and let $\mathcal{C}_{\mathcal{X}} = \mathcal{X}$.

The stage game plays an important role in our analysis. Two of our results (theorems 5 and 6) are best understood as reducing the online mechanism design problem to the simpler task of finding a "locally-robust" policy in the stage game. In the locally-robust problem, we maintain the traditional common prior assumption: that is, the state y is drawn from a commonly known distribution π. However, we relax the assumption that the agent maximizes her expected utility $\mathrm{E}_{y \sim \pi}[U(r, p, y)]$. Instead, she chooses a response (or a distribution μ over responses) that guarantees her an expected utility that is within an additive constant ϵ of the optimum. Since this assumption only partially identifies the agent's behavior, the principal's utility can take on a range of values. The principal's worst-case utility from following policy

[8] Note that $\mathcal{P}$, $\mathcal{R}$, $\Delta(\mathcal{Y})$ are all compact. Therefore, we can always construct a finite cover.

p is described by the function

$$\alpha_p(\pi,\epsilon) = \min_{\mu \in \Delta(\mathcal{R})} \mathrm{E}_{y\sim\pi}[\mathrm{E}_{r\sim\mu}[V(r,p,y)]] \quad \text{s.t.} \quad \max_{\tilde{r}\in\mathcal{R}} \mathrm{E}_{y\sim\pi}[U(\tilde{r},p,y)] - \mathrm{E}_{y\sim\pi}[\mathrm{E}_{r\sim\mu}[U(r,p,y)]] \leq \epsilon$$

and his best-case utility is described by

$$\beta_p(\pi,\epsilon) = \max_{\mu \in \Delta(\mathcal{R})} \mathrm{E}_{y\sim\pi}[\mathrm{E}_{r\sim\mu}[V(r,p,y)]] \quad \text{s.t.} \quad \max_{\tilde{r}\in\mathcal{R}} \mathrm{E}_{y\sim\pi}[U(\tilde{r},p,y)] - \mathrm{E}_{y\sim\pi}[\mathrm{E}_{r\sim\mu}[U(r,p,y)]] \leq \epsilon$$

The worst-case optimal (or ϵ-robust) policy, defined below, is one of two main ingredients in our proposed mechanisms (the other is a calibrated forecasting algorithm).

Definition 29 (ϵ-Robustness)**.** *The* ϵ-robust *policy is worst-case optimal over all response distributions μ that achieve at least the agent's optimal expected utility minus ϵ. Formally, policy is*

$$p^*(\pi,\epsilon) \in \arg\max_{p\in\mathcal{P}} \alpha_p(\pi,\epsilon)$$

Definition 30 (Cost of ϵ-Robustness)**.** *Fix a distribution π and parameter $\epsilon > 0$. The cost of ϵ-robustness is the distance between the principal's best-case utility (under the best-case optimal policy) and worst-case utility (under the worst-case optimal policy). Formally,*

$$\Delta(\pi,\epsilon) = \max_{p\in\mathcal{P}} \beta_p(\pi,\epsilon) - \alpha_{p^*(\pi,\epsilon)}(\pi,\epsilon)$$

The cost of ϵ-robustness will be a key variable in our upper bounds on the principal's regret in the repeated game. It will be convenient to assume that this cost is growing at most linearly in ϵ, although this assumption is not really necessary (see appendix B.4).

Assumption 8. *For any distribution π, $\Delta(\pi,\epsilon) = O(\epsilon)$.*

Finally, the following lemma will be important to our results. Suppose that the principal misjudges the agent. Instead of choosing a response that achieves at least her optimal expected utility minus ϵ, the agent only achieves her optimal expected utility minus $\epsilon + \tilde{\epsilon}$, for $\tilde{\epsilon} > 0$. Nonetheless, if the principal uses the ϵ-robust policy, his utility degrades smoothly in the residual $\tilde{\epsilon}$.

Lemma 15. *Assume regularity (assumption 7). For any distribution π, policy p, and constants $\epsilon, \tilde{\epsilon} > 0$, the principal's worst-case and best-case utilities satisfy*

$$\alpha_p(\pi, \epsilon + \tilde{\epsilon}) \geq \alpha_p(\pi, \epsilon) - \frac{\tilde{\epsilon}}{\epsilon} \quad \text{and} \quad \beta_p(\pi, \epsilon + \tilde{\epsilon}) \leq \beta_p(\pi, \epsilon) + \frac{\tilde{\epsilon}}{\epsilon}$$

Appendix B.1 describes two well-known special cases of our model: Bayesian persuasion and the principal-agent problem. For each case, we provide a simple example, check that the example satisfies all relevant assumptions, and evaluate our results. In that sense, these examples serve as sanity checks for the rest of the paper, which involves assumptions and solutions that are sometimes rather abstract.

2.3 Repeated Game

In the repeated game, the stage game is repeated T times. In period t, the principal chooses policy p_t, the agent chooses response r_t, and nature chooses the state y_t. At the end of period t, the state y_t is revealed to both the principal and the agent.

The agent's repeated game strategy (henceforth, *learner* L) maps the state history $y_{1:t-1}$, the response history $r_{1:t-1}$, the policy history $p_{1:t-1}$, and the current policy p_t to a distribution μ_t over responses. Formally, the response distribution in the t^{th} period is given by[9]

$$L_t : \mathcal{Y}^{t-1} \times \mathcal{R}^{t-1} \times \mathcal{P}^t \to \Delta(\mathcal{R})$$

[9] The fact that the response distribution μ_t may depend on realized response history $r_{1:t-1}$ allows the learner to introduce correlation between responses across time, if desired.

The principal's repeated game strategy (henceforth, *mechanism* σ) maps the state history $y_{1:t-1}$, the response history $r_{1:t-1}$, and the policy history $p_{1:t-1}$ to a distribution ν_t over policies. Formally, the policy distribution in the t^{th} period is given by

$$\sigma_t : \mathcal{Y}^{t-1} \times \mathcal{R}^{t-1} \times \mathcal{P}^{t-1} \to \Delta(\mathcal{P})$$

Our goal is to design a mechanism σ^* that the principal would not regret using, relative to a finite set of alternative mechanisms. Regret – which we define momentarily – measures the gap in performance between σ^* and the alternative mechanism σ that performed best in hindsight, given the realized sequence of states $y_{1:T}$. We consider a simple set of alternative mechanisms, corresponding to some finite set of fixed policies $\mathcal{P}_0 \subseteq \mathcal{P}$ that the principal wishes to consider.[10] Formally, by a fixed policy p, we mean a *constant mechanism* σ^p that selects the same policy

$$\sigma^p_t(y_{1:t-1}, r_{1:t-1}, p_{1:t-1}) = p$$

in all periods t and for all histories.

To define the principal's regret, we need notation for the agent's behavior under the proposed mechanism σ^*, as well as under the counterfactual mechanisms σ^p. Fix the state sequence $y_{1:T}$. Let μ^*_t describe the agent's behavior under σ^*, i.e.

$$\mu^*_t = L_t\left(y_{1:t-1}, r^*_{1:t-1}, p^*_{1:t}\right)$$

given the realized history of responses $r^*_{1:t-1}$ and policies $p^*_{1:t}$ under σ^*. Let μ^p_t describes the agent's behavior under σ^p, i.e.

$$\mu^p_t = L_t(y_{1:t-1}, r^p_{1:t-1}, (\underbrace{p, \ldots, p}_{\text{t times}}))$$

[10] Many of our results and definitions can be adapted to any finite set of nonresponsive mechanisms.

given the realized history of responses $r^p_{1:t-1}$ under σ^p.

Definition 31 (Principal's Regret)**.** *The* principal's regret *relative to the best-in-hindsight fixed policy $p \in \mathcal{P}_0$ is*

$$\text{Regret}^P_n(L, y_{1:T}) = \sup_{p \in \mathcal{P}_0} \frac{1}{T} \sum_{t=1}^{T} \left(\mathrm{E}_{r \sim \mu^p_t}[V(r, p, y_t)] - \mathrm{E}_{r \sim \mu^*_t}\left[V(r, \sigma^*(y_{1:t-1}, r^*_{1:t-1}, p_{1:t-1}), y_t)\right]\right)$$

The mechanism σ^* satisfies no-regret if the principal's regret is $o(1)$, i.e. it vanishes as $T \to \infty$. Recall that the no-regret mechanism design problem is infeasible without further assumptions on the learner L. The following proposition formalizes this simple observation.

Proposition 5 (Impossibility Result for Unrestricted Learners)**.** *In our running example, for every mechanism σ^*, there exists a learner L along with a state sequence $y_{1:\infty}$ such that the principal's regret does not vanish, i.e.*

$$\lim_{T \to \infty} \text{Regret}^P_n(L, y_{1:T}) > 0$$

2.4 Behavioral Assumptions

In this section, we develop a restriction on the learner L that captures "rational" behavior by the agent, without requiring assumptions on the state sequence $y_{1:T}$. In particular, we build on no-regret assumptions pioneered in the literature on learning in games.

In online learning, regret measures how much better or worse off the agent would have been had she followed the best-in-hindsight "simple" strategy instead of her learner. Different notions of regret correspond to different definitions of simplicity. All of the regret notions used in this paper will be special cases of *contextual regret*, defined as follows. Given a sequence $z_{1:T}$ of variables in some arbitrary set $\mathcal{Z}$, contextual regret considers a strategy "simple" if, for any two periods t and τ, sharing the same context $z_t = z_\tau$ implies taking the same response $r_t \neq r_\tau$.

Definition 32. *Given a sequence $z_{1:T}$ of covariates, the agent's* contextual regret *relative to a best-in-hindsight modification rule $h : \mathcal{Z} \to \mathcal{R}$ is*

$$\mathrm{CR}(p_{1:T}, y_{1:T}) = \max_h \frac{1}{T} \sum_{t=1}^{T} \left(U(h(z_t), p_t, y_t) - U(r_t, p_t, y_t)\right)$$

Note that, unlike our definition of the principal's regret, the agent's contextual regret does not take into account how changes in her past behavior would have also affected the principal's behavior. This omission is justified when the mechanism is nonresponsive.

Definition 33 (Responsiveness)**.** *A mechanism σ is* nonresponsive *if*

$$\sigma_t(y_{1:t-1}, r_{1:t-1}, p_{1:t-1}) = \sigma_t(y_{1:t-1}, \tilde{r}_{1:t-1}, p_{1:t-1})$$

for any period t, state history $y_{1:t-1}$, policy history $p_{1:t-1}$, and response histories $r_{1:t-1}, \tilde{r}_{1:t-1}$.

Our mechanisms will be nonresponsive. This is a design choice, not an assumption. In restricting attention to nonresponsive mechanisms, we simplify the agent's problem and make our behavioral assumptions more credible. If our mechanisms were responsive, non-myopic agents would not necessarily satisfy no-regret as defined above. For example, an agent might decide to forgo an otherwise-optimal response if she believes said response would trigger an undesirable policy by the principal going forward.[11] This behavior would be perfectly reasonable but could cause the agent to accumulate regret. Finally, as it turns out, even nonresponsive mechanisms can guarantee vanishing principal's regret in two of the scenarios we study (sections 2.5 and 2.7). In these scenarios, there is limited room for responsive mechanisms to improve our guarantees.

[11] For instance, in models of repeated sales, a buyer may refuse to purchase a good at a reasonable price if she believes that holding out will cause the seller to reduce prices in the future (Devanur et al. 2019; Immorlica, Lucier, et al. 2017).

In the remainder of this section, we define three special cases of contextual regret: *external regret* (ER), *internal regret* (IR), and *counterfactual internal regret* (CIR).

2.4.1 External Regret

In our model, external regret is contextual regret where the policy p_t is the context in period t. That is, no-ER requires the agent to perform as well as the best-in-hindsight mapping from policies p_t to responses r_t. Now, why should external regret include the policy as context? Basically, because our stage game is an extensive form. Technically-speaking, a *strategy* in the stage game is not simply a response; it is a function from the observed policy to a response. Our definition of external regret compares the agent's performance to the best-in-hindsight strategy in the stage game.[12]

An immediate difficulty with defining ER is that the set $\mathcal{P}$ may be continuous.[13] For instance, this is true in our running example. To ensure that the agent's learning problem is feasible in that case, we allow the agent to group together nearby policies according to the cover $\mathcal{C}_\mathcal{P}$ (defined in section 2.2), and consider regret with respect to this coarser context. Of course, when the policy space $\mathcal{P}$ is finite, there is no need for this, and we can set $\mathcal{C}_\mathcal{P} = \mathcal{P}$.

Definition 34 (External Regret)**.** *The agent's* external regret (ER) *relative to the best-in-*

[12] Suppose that, instead, we compared the agent's performance to the best-in-hindsight response $r \in \mathcal{R}$. Defining external regret in this way would confound variation in policies with variation in the state, and could lead to odd behavior. For example, consider the"mean-based" learner in Braverman et al. (2018), which never deviates far from the response that maximizes the agent's empirical utility. In that paper, the learner engages in odd behavior, like spending more than the agent's valuation.

Our definition is more similar to that of Hartline, Johnsen, et al. (2019), where agents following a dashboard provided by the mechanism will best respond to an allocation rule given the empirical value distribution, rather than best respond to the empirical bid distribution. This way, the agent adapts sensibly to changes in the principal's policy.

[13] If $\mathcal{P}$ is continuous, the policy p_t may be unique in every period $t = 1, \ldots, \infty$. In that case, requiring no-ER would be equivalent to requiring ex post optimality. That is unreasonably strong.

hindsight modification rule $h : \mathcal{C}_{\mathcal{P}} \to \mathcal{R}$ *is*

$$\mathrm{ER}(p_{1:T}, y_{1:T}) = \max_h \frac{1}{T} \sum_{t=1}^{T} \left(U(h(p_t), p_t, y_t) - U(r_t, p_t, y_t)\right)$$

Note the slight abuse of notation. By $h(p_t)$*, we mean* $h(P_t)$ *where* P_t *is the unique set in the partition* $\mathcal{C}_{\mathcal{P}}$ *that contains* p_t*.*

Although common in the literature (e.g. Nekipelov et al. 2015, Braverman et al. 2018), no-ER assumptions are insufficient for our problem. They do not circumvent the impossibility result (proposition **??**) that motivated us to restrict the agent's behavior in the first place. In particular, this is because they fail to rule out certain pathological behaviors. Because these pathological behaviors are clearly not in the agent's best interest, we also conclude that no-ER fails to rule out "irrational" behavior and is therefore not a good definition of "rationality". The following proposition (and its proof) clarifies the issue.

Proposition 6 (Impossibility Result for No-ER Learners)**.** *In our running example, for every mechanism* σ^**, there exists a learner* L *that guarantees no-ER on all state/policy sequences, i.e.*

$$\lim_{T\to\infty} \sup_{\tilde{p}_{1:T}, \tilde{y}_{1:T}} \mathrm{E}_L[\mathrm{ER}(\tilde{p}_{1:T}, \tilde{y}_{1:T})] = 0$$

along with a state sequence $y_{1:\infty}$ *such that the principal's regret does not vanish, i.e.*

$$\lim_{T\to\infty} \mathrm{Regret}_n^P(L, y_{1:T}) > 0$$

2.4.2 Internal and Counterfactual Internal Regret

Before defining CIR, we provide a brief intuition: what went wrong with external regret? Recall the tortoise and hare analogy in the introduction. For a behavioral assumption to rule out pathological behaviors, it may have to adapt to the information of the agent (or the speed of the animal).

What do we mean by information? Implicit in most stochastic models is the idea that the state is fundamentally unpredictable. But there is no ex ante sense in which the deterministic sequence $y_{1:T}$ is predictable or not. In particular, the agent may behave as if she possesses "private information" about the sequence of states that goes beyond the "public information" inherent in the description of the model. In practice, the agent may have access to data that the principal lacks, notice a pattern that did not occur to the principal, or succeed through dumb luck. Formally, this reflects an adversary who simultaneously chooses the state sequence $y_{1:T}$ and the learner L to cause the mechanism σ^* to underperform. In particular, even though the agent may not observe y_t when choosing a response r_t, this cannot prevent the adversary from "correlating" r_t and y_t.[14]

No-CIR requires the agent to consistently and fully exploit her private information. In the spirit of revealed preference, private information is identified with her behavior across counterfactual mechanisms. Intuitively, if the agent is able to distinguish between periods t, τ and finds it useful to do so, then her behavior should also differ between those two periods. If her behavior under one mechanism reveals private information, this information should also be accessible to her under a different mechanism. This logic allows us to define a purely ex post notion of rationality that does not refer to the agent's beliefs or to a distribution over state sequences.

No-CIR refines no-IR, a weaker condition that was developed in the literature on calibration (e.g. Foster and Vohra 1997). Internal regret is contextual regret where the context is the agent's own behavior $r_{1:T}$. To ensure that the agent's learning problem is feasible when

[14] To be clear, this "correlation" is non-causal. For example, the adversary might choose a state sequence such that $y_t = 1$ on even periods and $y_t = 0$ on odd periods, and a learner L such that $r_t = 1$ on even periods and $r_t = 0$ on odd periods. Empirically-speaking, there would be a correlation between the states and the responses. However, if we subsequently changed the value of state y_t in some period t, this would not affect the response r_t, because the state is not observed and cannot affect the output of the learner L. That is, there is no causal relationship between r_t and y_t.

the response space $\mathcal{R}$ is continuous, we allow the agent to group together nearby responses according to the cover $\mathcal{C}_{\mathcal{R}}$, and consider regret with respect to this coarser context. Of course, when the response space $\mathcal{R}$ is finite, as in our running example, there is no need for this, and we can set $\mathcal{C}_{\mathcal{R}} = \mathcal{R}$.

Definition 35 (Internal Regret)**.** *The agent's* internal regret (IR) *relative to the best-in-hindsight modification rule $h : S_{\mathcal{P}} \times S_{\mathcal{R}} \to \mathcal{R}$ is*

$$\mathrm{IR}(p_{1:T}, y_{1:T}) = \max_h \frac{1}{T} \sum_{t=1}^{T} \left(U(h(p_t, r_t), p_t, y_t) - U(r_t, p_t, y_t)\right)$$

Like earlier, note the slight abuse of notation. By $h(p_t, r_t)$, we mean $h(P_t, R_t)$ where (P_t, R_t) is the unique set in the collection $\mathcal{C}_{\mathcal{P}} \times \mathcal{C}_{\mathcal{R}}$ that contains (p_t, r_t).

Counterfactual internal regret is contextual regret where the context is the concatenation of: the policy p_t^* under the proposed mechanism σ^*; the agent's behavior $r_{1:T}^*$ under σ^*; and her counterfactual behavior $r_{1:T}^p$ under the fixed policies $p \in \mathcal{P}_0$. The following definitions formalize this.

Definition 36 (Information)**.** *Let the* information partition *be*

$$\mathcal{I} = \underbrace{S_{\mathcal{P}}}_{\textit{policy } p_t^*} \times \underbrace{S_{\mathcal{R}}}_{\textit{response } r_t^*} \times \underbrace{(S_{\mathcal{R}})^{|\mathcal{P}_0|}}_{\textit{responses } r_t^p \textit{ for } p \in \mathcal{P}_0}$$

and let the information I_t *in period t be the unique set in $\mathcal{I}$ that satisfies*

$$I_t \ni (p_t^*, r_t^*, (r_t^p)_{p \in \mathcal{P}_0})$$

Note that, by definition, the same information I_t is available to the agent regardless of whether the principal follows our mechanism σ^* or deviates to some fixed policy $p \in \mathcal{P}_0$.

Intuitively, the principal's choice of mechanism should not affect what information the agent has available.

Definition 37 (Counterfactual Internal Regret)**.** *The agent's* counterfactual internal regret (CIR) *relative to the best-in-hindsight modification rule $h : \mathcal{I} \to \mathcal{R}$ is*

$$\mathrm{CIR}(p_{1:T}, y_{1:T}) = \max_h \frac{1}{T} \sum_{t=1}^{T} (U(h(I_t), p_t, y_t) - U(r_t, p_t, y_t))$$

The discussion in the proof of proposition **??** clarifies how no-CIR rules out the kinds of pathological or irrational behavior that no-ER fails to rule out. In the next section, we will see the crucial role that no-CIR plays in our proving our bounds on the principal's regret. The essential property is that, conditional on information I_t, the agent chooses a roughly constant response that is approximately best-in-hindsight for whichever mechanism the principal is considering.

2.5 Mechanism for an Informed Principal

Our first result should be viewed as pedagogical. It bounds the principal's regret under a mechanism that requires oracle access to the agent's learner. This requirement is unrealistic and will be removed in sections 2.5 and 2.6. Likewise, the bound itself will feature an exponential dependence on the size of the policy space. This dependence will also be removed in later sections.

Definition 38 (Information Oracle)**.** *The* information oracle $\Omega_t : \mathcal{P} \to \mathcal{I}$ *specifies the information I_t that the learner L would generate in period t given any policy $p_t \in \mathcal{P}$ and the realized history.*

This case is a convenient starting point because it avoids the bulk of the information asymmetries between the principal and the agent that our later results need to address. That follows from the fact that any private information generated by the learner can be anticipated by the principal with access to the information oracle. This case is also a convenient point of departure from the common prior assumption because it permits a wider range of agent behavior without relaxing the principal's knowledge of said behavior. To be clear, under a common prior, the fact that the principal knows the agent's prior means that he also has precise knowledge of the agent's learner. In addition, since the agent is Bayesian, the agent does not find it beneficial to randomize and her learner will typically be deterministic. Essentially, the common prior provides an information oracle for free.

Mechanism 1. *Let the distribution π_t be a forecast of the state y_t generated by a calibrated forecasting algorithm that uses the agent's information as context.*

- *Our forecasting algorithm applies a generic no-internal-regret algorithm due to Blum and Mansour (2007) in an auxilliary learning problem where the action space consists of discretized forecasts $\pi \in \mathcal{C}_{\Delta(\mathcal{Y})}$ and the loss function is the negated quadratic scoring rule S. In each period, the algorithm makes a prediction π_t and incurs loss $-S(\pi_t, y_t)$. Further details as well as rates of convergence are in appendix B.3.*
- *The context is the vector of outputs $\Omega(p)$ of the information oracle under discretized policies $p \in \mathcal{C}_{\mathcal{P}}$. The forecasting algorithm is run separately for each context.*

Fix a parameter $\bar{\epsilon} > 0$. In period t, the informed-principal mechanism *σ^* chooses the discretization of the $\bar{\epsilon}$-robust policy $p^*(\pi_t, \bar{\epsilon})$ that treats the forecast π_t as a common prior.*

Before stating the theorem in full, we present the reasoning behind the result and clarify the components of the regret bound, as well as the assumptions required. First, we require some additional notation. Let "$t \in I$" indicate that information I is present in period t, i.e. $I_t = I$. Let $n_I = \sum_{t=1}^{T} \mathbf{1}(t \in I)$ indicate the number of periods with information I. Let $\hat{\pi}_I$

be the empirical distribution conditioned on the agent having information I, i.e.

$$\hat{\pi}_I(y) = \frac{1}{n_I} \sum_{t \in I} \mathbf{1}(y_t = y)$$

We begin with a straightforward but important observation: across all periods $t \in I$, the agent's response r_t^* is roughly constant, as are her counterfactual responses r_t^p under fixed policies $p \in \mathcal{P}_0$. By regularity (7), slight variations in responses have correspondingly slight impacts on the agent's and principal's utility. Suppose that these responses are exactly constant, i.e. $r_t = r_I$. Note that $p_t = p_I$ is exactly constant as well, across these time periods, for all constant mechanisms σ^p as well as the proposed mechanism σ^*, which uses discretized policies. With everything constant, the principal's average utility across context I takes on a familiar form:

$$\frac{1}{n_I} \sum_{t \in I} V(r_I, p_I, y_t) = \mathrm{E}_{y \sim \hat{\pi}_I}[V(r_I, p_I, y)]$$

Similarly, the agent's average utility is

$$\frac{1}{n_I} \sum_{t \in I} U(r_I, p_I, y_t) = \mathrm{E}_{y \sim \hat{\pi}_I}[U(r_I, p_I, y)]$$

Essentially, within each context I, we have recreated the stage game with common prior $\hat{\pi}_I$. The agent accumulates regret

$$\epsilon_I = \max_{\tilde{r}} \mathrm{E}_{y \sim \hat{\pi}_I}[U(\tilde{r}, p, y)] - \mathrm{E}_{y \sim \hat{\pi}_I}[U(r_I, p, y)]$$

Under mechanism 1, the principal chooses (roughly) the $\bar{\epsilon}$-robust policy for the forecast π_t. Suppose for the moment that the forecasts are also roughly constant for all periods $t \in I$, i.e. $\pi_t = \pi_I$. Since the forecast is calibrated and uses information I_t as context, π_I cannot be too far in the l_1 distance from $\hat{\pi}_I$ (this is essentially the definition of calibration, and follows from results in appendix B.3). It follows from regularity that the $\bar{\epsilon}$-robust policy for π_I is nearly $\bar{\epsilon}$-robust for $\hat{\pi}_I$.

At this point, the principal has (roughly) applied the $\bar{\epsilon}$-robust policy for the empirical distribution $\hat{\pi}_I$, to an agent that obtains regret ϵ_I. In that sense, the principal has misjudged the agent's capacity to make mistakes. However, recall lemma 15: this affects the principal's best-case and worst-case utilities by at most $\epsilon_I/\bar{\epsilon}$. It follows that, roughly-speaking, the principal's utility is not much worse than the worst-case optimal utility. At the same time, it cannot be much better than the best-case optimal utility. More precisely,

$$\max_{\tilde{p}} \beta_{\tilde{p}}(\hat{\pi}_I, \bar{\epsilon}) + \frac{\epsilon_I}{\bar{\epsilon}} \geq \mathrm{E}_{y\sim\hat{\pi}_I}[V(r_I, p_I, y)] \geq \max_{\tilde{p}} \alpha_{\tilde{p}}(\hat{\pi}_I, \bar{\epsilon}) - \frac{\epsilon_I}{\bar{\epsilon}} \tag{2.4}$$

By assumption 8, the difference between the upper bound and the lower bound is

$$O(\bar{\epsilon}) + O\left(\frac{\epsilon_I}{\bar{\epsilon}}\right) \tag{2.5}$$

This pins down the principal's utility under mechanism 1. Moreover, the upper bound in (2.4) also applies to any constant mechanism σ^p for $p \in \mathcal{P}_0$. Therefore, (2.5) also bounds the regret accumulated by the principal in context I.

This brings us to our key assumption: the agent's CIR is at most some constant ϵ.

Assumption 9 (Bounded CIR)**.** *Let $y_{1:T}$ be the realized state sequence and let $p^*_{1:T}$ be the policy sequence generated by the proposed mechanism σ^*. There exists a constant $\epsilon \geq 0$ such that*

$$\epsilon \geq \mathrm{CIR}(y_{1:T}, p^*_{1:T}) \quad \text{and} \quad \epsilon \geq \mathrm{CIR}(y_{1:T}, \underbrace{p, \dots, p}_{t \text{ times}}),\ \forall p \in \mathcal{P}_0$$

Remark 1. It is worth emphasizing that this bound applies only to the realized state sequence $y_{1:T}$. That is, the agent does not need to perform well over all state sequences, and her objective need not be worst-case regret minimization. If the agent is Bayesian, for example, she will obtain low CIR as long as her beliefs are well-calibrated.

Since CIR is contextual regret with information I_t as context, bounded CIR ensures that

$$\epsilon \geq \frac{1}{T} \sum_{I \in \mathcal{I}} n_I \epsilon_I$$

Combine this with our bound (2.5) on the agent's regret ϵ_I in the context of information I, and it follows that the principal's regret is bounded above by

$$O(\bar{\epsilon}) + O\left(\frac{\epsilon}{\bar{\epsilon}}\right)$$

To transform this intuition into a result, we need to address an assumption made along the way: that the forecast π_t is roughly constant across all periods $t \in I$. This is not necessarily true. The adversary can choose a sequence of states $y_{1:T}$ that makes the principal appear more informed than the agent. Indeed, variation in forecasts can be interpreted as private information of the principal, even if it is spurious. On the other hand, any variation in π_t that affects the policy p_t will also be included in the agent's information I_t. What remains is variation in π_t that does not affect the policy – useless information from the principal's perspective, but not necessarily useless to the agent. If the principal expects the agent to exploit this information and the agent does not, this can lead to a suboptimal policy choice.

The following assumption restricts attention to stage games where this problem does not arise; that is, the agent's failure to exploit information that is useless to the principal does not affect the principal's utility. In appendix B.4, we avoid this restriction by instead assuming that the principal – using our publicly-announced mechanism – is not more informed than the agent.

Assumption 10. *Let $\epsilon > 0$. Let π and $\tilde{\pi}$ be distributions in the stage game. If the ϵ-robust policies under π and under $\tilde{\pi}$ are close to one another, then they are also close to the ϵ-robust* *policy under any convex combination of these distributions. Formally, for any $\lambda \in [0,1]$,*

$$d_{\mathcal{P}}\left(p^*(\pi,\epsilon), p^*\left(\lambda\pi + (1-\lambda)\tilde{\pi}, \epsilon\right)\right) = O\left(d_{\mathcal{P}}\left(p^*(\pi,\epsilon), p^*(\tilde{\pi},\epsilon)\right)\right)$$

The following theorem formalizes the preceding discussion and bounds the principal's regret under mechanism 1.

Theorem 5. *Assume regularity (assumption 7), restrictions on the stage game (assumptions 8, 10), and ϵ-bounded CIR (assumption 9). Let σ^* be the mechanism 1. Given access to the information oracle, for any constant $\bar{\epsilon} > 0$, the principal's expected regret* $\mathrm{E}_{\sigma^*}\left[\mathrm{Regret}_n^P(L, y_{1:T})\right]$ *is at most*

$$\underbrace{O(\bar{\epsilon})}_{\textit{cost of } \bar{\epsilon}\textit{-robustness}} + \frac{1}{\bar{\epsilon}} \cdot \Bigg(\underbrace{O(\epsilon)}_{\textit{agent's regret}} + \underbrace{\tilde{O}\left(T^{-1/4}\sqrt{|\mathcal{Y}||\mathcal{C}_{\Delta(\mathcal{Y})}||\mathcal{C}_R|^{(|\mathcal{P}_0|+|\mathcal{C}_{\mathcal{P}}|)/2}}\right)}_{\textit{forecast miscalibration}}$$
$$+ \underbrace{O\left(\delta_{\Delta(\mathcal{Y})}^{1/2}\right) + O(\delta_{\mathcal{R}}) + O(\delta_{\mathcal{P}})}_{\textit{discretization error}} \Bigg)$$

Remark 2. Here are a few comments on this result.

1. The bound depends on the size of the partitions $\mathcal{C}_{\mathcal{P}}$, $\mathcal{C}_{\mathcal{R}}$, and $\mathcal{C}_{\Delta\mathcal{Y}}$. However, if we define these partitions to be as small as possible, we can replace these terms with the covering numbers of $\mathcal{P}$, $\mathcal{R}$, and $\Delta(\mathcal{Y})$, respectively. In that sense, our finite sample bounds will deteriorate as one increases the complexity of the action and state spaces.

2. Furthermore, if we define these partitions to be the smallest possible, then theorem 5 implies that the principal's regret vanishes if $T \to \infty$ and $\epsilon, \bar{\epsilon}, \delta_{\Delta(\mathcal{Y})}, \delta_{\mathcal{P}}, \delta_{\mathcal{R}} \to 0$ at the appropriate rates. It also follows from the proof that the principal's payoffs converge to a natural benchmark: what he would have obtained in a stationary equilibrium of the repeated game where it is common knowledge that y_t is drawn independently from

the empirical distribution $\hat{\pi}_{I_t}$. Formally,

$$\frac{1}{T}\sum_{t=1}^{T} V(r_t, p_t, y_t) - \frac{1}{T}\sum_{I\in\mathcal{I}} n_I \max_{p\in\mathcal{P}} \beta_p(\hat{\pi}_I, 0) \to 0$$

3. Finally, note the exponential dependence on the number of alternative mechanisms $|\mathcal{P}_0|$ and the size of the policy space cover $|\mathcal{C}_\mathcal{P}|$. This dependence, which is not present in theorems 6 and 7, reflects the fact that the mechanism 1 uses the agent's information I_t as context for its forecast π_t. Since our bound is uniform across all learners that satisfy ϵ-bounded CIR on the realized state sequence $y_{1:T}$, it must accommodate learners that generate a lot of information, regardless of whether that information is useful. As mentioned at the beginning of this section, this is another reason why the "informed principal" setting seems less compelling than the settings studied in sections 2.7 and 2.8.

2.6 Stage Game with Private Signals

In general, we cannot expect the principal to have access to an information oracle. Fortunately, we can still construct mechanisms σ^* that obtain vanishing or bounded principal's regret without any knowledge of the learner. However, in order to state the relevant assumptions (sections 2.7 and 2.8) and describe the mechanism (section 2.8), we need to consider scenarios where the agent has private information that the principal lacks. This requires a brief detour. In this section, we revisit the stage game in order to introduce terminology that reflects agent's private information.

Suppose that the state y is drawn from a known distribution π, but the agent has access to a private signal $I \in \mathcal{I}$ generated by the *information structure* γ.

Definition 39 (Information Structure). *An* information structure *is a function* $\gamma : \mathcal{I} \times \mathcal{Y} \to$

$[0,1]$ *where* $\gamma(\cdot, y)$ *is a probability distribution over* $\mathcal{I}$.

The game proceeds as follows. First, nature chooses a hidden state $y \sim \pi$. Second, the principal chooses a policy p. Third, the agent observes a signal $I \sim \gamma(\cdot, y)$ and chooses a response r_I. For instance, if the agent maximizes her expected utility, her responses after signals I would be

$$r_I \in \arg\max_{\tilde{r}_I \in \mathcal{R}} \mathrm{E}_{y\sim\pi}\left[\mathrm{E}_{I\sim\gamma(\cdot,y)}[U(\tilde{r}_I, p, y)]\right]$$

Finally, the state y is revealed and payoffs are determined.

As in section 2.2, suppose the agent does not necessarily maximize her expected utility. Instead, she chooses responses r_I (or distributions μ_I over responses) that guarantees her an expected utility that is within an additive constant ϵ of the optimum. For a given information structure γ, the principal's worst-case utility from following policy p is described by

$$\begin{aligned} \alpha_p(\pi, \gamma, \epsilon) = & \min_{\mu_I \in \Delta(\mathcal{R})} \mathrm{E}_{y\sim\pi}\left[\mathrm{E}_{I\sim\gamma(\cdot,y)}[\mathrm{E}_{r\sim\mu_I}[V(r,p,y)]]\right] \\ \text{subject to} \quad & \max_{\tilde{r}_I \in \mathcal{R}} \mathrm{E}_{y\sim\pi}\left[\mathrm{E}_{I\sim\gamma(\cdot,y)}[U(\tilde{r}_I, p, y)]\right] - \mathrm{E}_{y\sim\pi}\left[\mathrm{E}_{I\sim\gamma(\cdot,y)}[\mathrm{E}_{r\sim\mu_I}[U(r,p,y)]]\right] \leq \epsilon \end{aligned}$$

and his best-case utility is described by

$$\begin{aligned} \beta_p(\pi, \gamma, \epsilon) = & \max_{\mu_I \in \Delta(\mathcal{R})} \mathrm{E}_{y\sim\pi}\left[\mathrm{E}_{I\sim\gamma(\cdot,y)}[\mathrm{E}_{r\sim\mu_I}[V(r,p,y)]]\right] \\ \text{subject to} \quad & \max_{\tilde{r}_I \in \mathcal{R}} \mathrm{E}_{y\sim\pi}\left[\mathrm{E}_{I\sim\gamma(\cdot,y)}[U(\tilde{r}_I, p, y)]\right] - \mathrm{E}_{y\sim\pi}\left[\mathrm{E}_{I\sim\gamma(\cdot,y)}[\mathrm{E}_{r\sim\mu_I}[U(r,p,y)]]\right] \leq \epsilon \end{aligned}$$

Note that $\alpha(\pi, \epsilon)$, the worst-case utility in the stage game without a private signal, is equivalent to $\alpha(\pi, \gamma, \epsilon)$ when γ is uninformative. The same applies to β.

Recall that our theorem 5 could be interpreted as reducing the online mechanism design problem to the simpler task of finding a ϵ-robust policy in the stage game without a private signal. The same is true of our next result, theorem 6. In contrast, theorem 7 reduces the online problem to solving for a robust policy when the agent has a private signal generated by an unknown information structure. This corresponds to notion of informational robustness introduced by Bergemann and Morris (2013) and applied by Bergemann, Brooks, et al. (2017), applied to our single-agent setting.

Definition 40 (ϵ-Informational-Robustness)**.** *The worst-case optimal (or* ϵ-informationally-robust*) policy for an unknown information structure γ is*

$$p^\dagger(\pi, \epsilon) \in \arg\max_{p \in \mathcal{P}} \inf_{\gamma} \alpha_p(\pi, \gamma, \epsilon)$$

Definition 41 (Cost of ϵ-Informational-Robustness)**.** *Fix a distribution π and parameter $\epsilon > 0$. The cost of* ϵ-informational-robustness *is the distance between the principal's best-case utility (under the best-case optimal policy (for the best-case information structure) and worst-case utility (under the worst-case optimal policy for the worst-case information structure). Formally,*[15]

$$\nabla(\pi, \epsilon) = \max_{p \in \mathcal{P}} \sup_{\gamma} \beta_p(\pi, \gamma, \epsilon) - \max_{p \in \mathcal{P}} \inf_{\gamma} \alpha_p(\pi, \gamma, \epsilon)$$

Let $\nabla(\pi) = \nabla(\pi, 0)$ denote the cost of informational robustness in the traditional setting where the agent is optimizing exactly ($\epsilon = 0$). It will be convenient to assume that the cost is growing at most linearly in ϵ, although this assumption is not really necessary (see appendix B.4).

Assumption 11. *For any distribution π, $\nabla(\pi, \epsilon) = \nabla(\pi) + O(\epsilon)$.*

Finally, we verify that lemma 15 still applies in the presence of private signals.

[15] Why do we evaluate the cost of informational robustness under the worst-case information structure? Because the regret guarantee that we obtain in theorem 7 applies uniformly across all learners L. As we will see, different learners will induce different empirical information structures γ. Our cost of informational robustness must accommodate the worst-case information structure, which loosely corresponds to the worst-case learner.

Lemma 16. *Assume regularity (assumption 7). For any distribution π, information structure γ, policy p, and constants $\epsilon, \tilde{\epsilon} > 0$, the principal's worst-case and best-case utilities satisfy*

$$\alpha_p(\pi, \gamma, \epsilon + \tilde{\epsilon}) \geq \alpha_p(\pi, \gamma, \epsilon) - \frac{\tilde{\epsilon}}{\epsilon} \quad \text{and} \quad \beta_p(\pi, \gamma, \epsilon + \tilde{\epsilon}) \leq \beta_p(\pi, \gamma, \epsilon) + \frac{\tilde{\epsilon}}{\epsilon}$$

2.7 Mechanism for an Uninformed Agent

Our second result bounds the principal's regret under a mechanism that does not require detailed knowledge of the learner L. Instead, this result assumes that the agent is not more informed than the principal. To begin, the mechanism is as follows.

Mechanism 2. *Let the distribution π_t be a forecast of the state y_t.*

- *Our forecasting algorithm applies a generic no-internal-regret algorithm due to Blum and Mansour (2007) in an auxilliary learning problem where the action space consists of discretized forecasts $\pi \in \mathcal{C}_{\Delta(\mathcal{Y})}$ and the loss function is the negated quadratic scoring rule.*

Fix a parameter $\bar{\epsilon} > 0$. In period t, the uninformed-agent mechanism *σ^* chooses the discretization of the $\bar{\epsilon}$-robust policy $p^*(\pi_t, \bar{\epsilon})$ that treats the forecast π_t as a common prior.*

What does it mean for an agent to be uninformed? Following the intuition developed in section 2.4, the agent's behavior cannot reveal an understanding of the state sequence that goes far beyond the principal's forecast. This can be formalized by adding a lower bound on the agent's ER to our upper bound on the agent's (counterfactual) IR.[16]

[16] Although they study a different problem, Blum, Gunasekar, et al. (2018) also use lower bounds on ER to prove results, exploiting the fact that exponential weights guarantees non-negative expected ER (Gofer and Mansour 2016).

Assumption 12 (Lower-Bounded ER)**.** *Let $y_{1:T}$ be the realized state sequence and let $p^*_{1:T}$ be the policy sequence generated by the proposed mechanism σ^*. There exists a constant $\tilde{\epsilon} \geq 0$ such that*

$$\mathrm{ER}(y_{1:T}, p^*_{1:T}) \geq -\tilde{\epsilon} \quad \text{and} \quad \mathrm{ER}(y_{1:T}, \underbrace{p, \ldots, p}_{t\ times}) \geq -\tilde{\epsilon},\ \forall p \in \mathcal{P}_0$$

While there is no a priori sense in which the deterministic sequence $y_{1:T}$ is predictable or not, this combination of bounds can be seen as an ex post definition of unpredictability. Intuitively, if an agent fully exploits the information she reveals under the proposed mechanism σ^* (no-IR) without outperforming the best use of public information (non-negative ER), her private information cannot be particularly useful. Fully exploiting useless information generally means ignoring it.

To see this, suppose the policy p is fixed and that the learner obtains non-positive IR and non-negative ER. It is trivial to show that IR is non-negative and bounded below by ER, so it follows that the learner's IR and ER both equal zero. In turn, IR and ER can only be equal when the best-in-hindsight responses conditional on the context (i.e. the learner's response) are the same in every context. That is, the context is useless. To achieve zero IR, the learner's response must equal some best-in-hindsight response conditional on the context. If the best-in-hindsight response is unique, this means that the learner's response is the same in every period.

What this amounts to, essentially, is that our reasoning for theorem 5 largely applies to theorem 6. Let us recall the first steps of that argument. Previously, we considered all periods $t \in I$ with information I as context. It followed immediately from the definition of information that the agent's responses r_t were roughly some constant r_I. Furthermore, since the principal's forecasts used I_t as context, the constant policy p_I was calibrated to the empirical distribution $\hat{\pi}_I$.

Now, our mechanism does not have access to I_t and is not calibrated to $\hat{\pi}_I$. Instead, for every policy context $P \in \Sigma_{\mathcal{P}}$, it is calibrated to the empirical distribution $\hat{\pi}_P$ conditioned on $p_t \in P$. Formally,

$$\hat{\pi}_P(y) = \frac{1}{n_P} \sum_{t \in P} \mathbf{1}(y_t = y)$$

where $t \in P$ indicates $p_t \in P$ and n_P is the number of periods $t \in P$. The policy context P is coarser than information I, by definition of the latter. So, the principal behaves as if the agent shares his prior $\hat{\pi}_P$, while the agent behaves as if she receives I as a private signal.

This is where non-negative ER comes in. The agent's information I is useless to her. If there is a unique best-in-hindsight response within policy context P, then the agent will choose roughly the same response $r_t = r_P$ in every period $t \in P$. In other words, the policy context P coincides with the agent's information I, and the principal is correct in assuming that the agent (roughly) optimizes against the empirical distribution $\hat{\pi}_P$. Our previous argument goes through.

Again, we just assumed that there is a unique best-in-hindsight response within policy context P. What if this is not the case, i.e. the best-in-hindsight response is not unique? In general, the argument breaks down. The agent can condition her action on her private information I, which no longer necessarily coincides with P. To be clear, this private signal I remains useless to the agent. Moreover, the $\bar{\epsilon}$-robust policy is by definition robust to multiplicity of best responses. However, if the agent's best response is correlated with the state, this can undermine the principal's utility even if it does not affect the agent's.[17]

The following assumption restricts attention to stage games where this issue does not arise. Informally, it asserts that if a private signal is useless to the agent, then it has limited relevance to the principal, assuming that the principal is following (nearly) optimal policies.

[17] For example, consider a stage game with a binary response $r \in \{0, 1\}$, a binary state $y \in \{0, 1\}$, and a binary policy $p \in \{\text{Risky}, \text{Safe}\}$. The agent's utility is always zero. The principal's utility under the risky policy is 1 if $r = y$ and -1 otherwise. It is slightly negative under the safe policy. If y is drawn from the uniform distribution, and the agent optimizes without a signal, then the principal prefers the risky policy. If the agent receives a signal that is perfectly correlated with the state, and sets $r = 1 - y$, then the principal prefers the safe policy.

Formally, the value of information structure γ to the agent in the stage game with common prior π and policy p is

$$\phi_p(\pi, \gamma) = \max_{r, r_I \in \mathcal{R}} \mathrm{E}_{y \sim \pi}\left[\mathrm{E}_{I \sim \gamma(\cdot, y)}[U(r_I, p, y)] - U(r, p, y)\right]$$

This is the expected utility of the agent that optimizes given information structure γ minus the expected utility of the agent if she does not receive a private signal.

Assumption 13. *Let π be a distribution, $\epsilon > 0$ be a constant, and γ be an information structure (intuitively, one that is not useful to the agent).*

1. *If the principal uses ϵ-robust policy $p^*(\pi, \epsilon)$, his maxmin payoff without γ, i.e. $\alpha_{p^*(\pi,\epsilon)}(\pi, \epsilon)$, is not much larger than his maxmin payoff with γ, i.e. $\alpha_{p^*(\pi,\epsilon)}(\pi, \gamma, \epsilon)$. That is,*

$$\alpha_{p^*(\pi,\epsilon)}(\pi, \epsilon) - \alpha_{p^*(\pi,\epsilon)}(\pi, \gamma, \epsilon) = O\left(\phi_{p^*(\pi,\epsilon)}(\pi, \gamma)\right) + O(\epsilon)$$

2. *The principal's maxmax payoff with γ under any policy $p \in \mathcal{P}$, i.e. $\beta_p(\pi, \gamma, \epsilon)$, is not much larger than his maxmax payoff without γ under the best-case policy, i.e. $\max_{\tilde{p} \in \mathcal{P}} \beta_{\tilde{p}}(\pi, \epsilon)$. That is,*

$$\beta_p(\pi, \gamma, \epsilon) - \max_{\tilde{p} \in \mathcal{P}} \beta_{\tilde{p}}(\pi, \epsilon) = O\left(\phi_p(\pi, \gamma)\right) + O(\epsilon)$$

Both parts of assumption 13 would be immediate if the information structure γ were uninformative, because the left-hand sides would be non-positive. Basically, we require useless (to the agent) private signals to be similar to uninformative private signals in these two respects.

Finally, we are ready to bound the principal's regret under mechanism 2.

Theorem 6. *Assume regularity (assumption 7), restrictions on the stage game (assumptions 8, 10, 13), ϵ-bounded CIR (assumption 9), and $\tilde{\epsilon}$-lower-bounded ER (assumption 12). Let σ^* be the uninformed-agent mechanism 2. For any constant $\bar{\epsilon} > 0$, the principal's expected regret $\mathrm{E}_{\sigma^*}\left[\mathrm{Regret}_n^P(L, y_{1:T})\right]$ is at most*

$$\begin{aligned}\underbrace{O(\bar{\epsilon})}_{\textit{cost of } \bar{\epsilon}\textit{-robustness}} + \underbrace{O(\tilde{\epsilon})}_{\textit{agent's information}} + \frac{1}{\bar{\epsilon}} \cdot \Bigg(\underbrace{O(\epsilon)}_{\textit{agent's regret}} + \underbrace{\tilde{O}\left(T^{-1/4}\sqrt{|\mathcal{Y}||\mathcal{C}_{\Delta(\mathcal{Y})}|}\right)}_{\textit{forecast miscalibration}} \\ + \underbrace{O\left(\delta_{\Delta(\mathcal{Y})}^{1/2}\right) + O(\delta_{\mathcal{R}}) + O(\delta_{\mathcal{P}})}_{\textit{discretization error}} \Bigg)\end{aligned}$$

Remark 3. If we define the partition $\mathcal{C}_{\Delta\mathcal{Y}}$ to be the smallest possible, then theorem 6 implies that the principal's regret vanishes if $T \to \infty$ and $\epsilon, \bar{\epsilon}, \tilde{\epsilon}, \delta_{\Delta(\mathcal{Y})}, \delta_{\mathcal{P}}, \delta_{\mathcal{R}} \to 0$ at the appropriate rates. It also follows from the proof that the principal's payoffs converge to a natural benchmark: what he would have obtained in a stationary equilibrium of the repeated game where it is common knowledge that y_t is drawn independently from the empirical distribution $\hat{\pi}_{P_t}$. Formally,

$$\frac{1}{T}\sum_{t=1}^{T} V(r_t, p_t, y_t) - \frac{1}{T}\sum_{P \in \mathcal{C}_{\mathcal{P}}} n_P \max_{p \in \mathcal{P}} \beta_p(\hat{\pi}_P, 0) \to 0$$

2.8 Mechanism for an Informed Agent

In section 2.4, we assumed that the principal knows the agent's learner L. The implication of this assumption is that the principal is as informed as the agent. In section 2.5, we assumed that the agent is as uninformed as the principal. In this section, we allow the agent to be more informed than the principal. This generality comes at a cost: we no longer ensure vanishing principal's regret. Instead, we show that, in the limit, the following mechanism guarantees regret that is no greater than the cost of informational robustness.

Mechanism 3. *Let the distribution π_t be a forecast of the state y_t.*

- *Our forecasting algorithm applies a generic no-internal-regret algorithm due to Blum and Mansour (2007) in an auxilliary learning problem where the action space consists of the discretized forecasts $\pi \in \mathcal{C}_{\Delta(\mathcal{Y})}$ and the loss function is the negated quadratic scoring rule.*

Fix a parameter $\bar{\epsilon} > 0$. In period t, the informed-agent mechanism *σ^* chooses the discretization of the $\bar{\epsilon}$-informationally-robust policy $p^\dagger(\pi_t, \bar{\epsilon})$ that treats the forecast π_t as a common prior.*

Theorem 7 builds on the same reasoning as theorems 5 and 6. First, we need to adapt assumption 10 to the case with private signals.

Assumption 14. *Let $\epsilon > 0$. Let π and $\tilde{\pi}$ be distributions in the stage game. If the ϵ-informationally-robust policies under π and under $\tilde{\pi}$ are close to one another, then they are also close to the ϵ-informationally-robust policy under any convex combination of these distributions. Formally, for any $\lambda \in [0, 1]$,*

$$d_{\mathcal{P}}\left(p^\dagger(\pi, \epsilon), p^\dagger\left(\lambda\pi + (1-\lambda)\tilde{\pi}, \epsilon\right)\right) = O\left(d_{\mathcal{P}}\left(p^\dagger(\pi, \epsilon), p^\dagger(\tilde{\pi}, \epsilon)\right)\right)$$

Next, recall how, in the previous section, we were concerned that the principal's policy p_t in period t was calibrated to the empirical distribution $\hat{\pi}_P$ given policy context $P \in \mathcal{C}_{\mathcal{P}}$ (where $t \in P$) rather than the empirical distribution $\hat{\pi}_I$ given information $I = I_t$. There, we resolved that problem by assuming the agent was uninformed (non-negative ER). Here, our solution is even simpler: choose a policy p_t that is robust to the agent's private information I, whatever that may be.

To be more precise, recall that the policy context P is coarser than information I. We can interpret periods $t \in I$ as those periods in which the agent received a private signal

I. By looking at the frequency of information I within policy context P, we can define an empirical information structure $\hat{\gamma}_P$ using Bayes' rule, i.e.

$$\hat{\gamma}_P(I, y) = \frac{n_I \hat{\pi}_I(y)}{n_P \hat{\pi}_P(y)} \cdot \mathbf{1}(I \subseteq P)$$

where $I \subseteq P$ is shorthand for $t \in I \implies t \in P$. Before, we could roughly approximate principal's and agent's utility as their expected utility in the stage game where the state y was drawn from the empirical distribution $\hat{\pi}_I$. Now, the approximation is the expected utility in the stage game where $y \sim \hat{\pi}_P$ and the agent receives private signal I from the empirical information structure $\hat{\gamma}_P$. Of course, the principal's policy p_t is robust to all information structures γ, including $\hat{\gamma}_P$.

Next, we formalize this discussion and bound the principal's regret under mechanism 3.

Theorem 7. *Assume regularity (assumption 7), restrictions on the stage game (assumptions 11, 14), and ϵ-bounded CIR (assumption 9). Let σ^* be the informed-agent mechanism 3. For any constant $\bar{\epsilon} > 0$, the principal's expected regret $\mathrm{E}_{\sigma^*}\left[\mathrm{Regret}_n^P(L, y_{1:T})\right]$ is at most*

$$\underbrace{\frac{1}{T}\sum_{P \in \mathcal{C}_\mathcal{P}} n_P \nabla(\hat{\pi}_P) + O(\bar{\epsilon})}_{\textit{cost of } \bar{\epsilon}\textit{-informational-robustness}} + \frac{1}{\bar{\epsilon}} \cdot \Bigg(\underbrace{O(\epsilon)}_{\textit{agent's regret}} + \underbrace{\tilde{O}\left(T^{-1/4}\sqrt{|\mathcal{Y}||\mathcal{C}_{\Delta(\mathcal{Y})}|}\right)}_{\textit{forecast miscalibration}}$$
$$+ \underbrace{O\left(\delta_{\Delta(\mathcal{Y})}^{1/2}\right) + O(\delta_\mathcal{R}) + O(\delta_\mathcal{P})}_{\textit{discretization error}} \Bigg)$$

Remark 4. In contrast to our previous results, this regret bound does not vanish. However, if we define the partition $\mathcal{C}_{\Delta\mathcal{Y}}$ to be the smallest possible, the bound does converge to

$$\frac{1}{T}\sum_{P \in \mathcal{C}_\mathcal{P}} n_P \nabla(\hat{\pi}_P)$$

as $T \to \infty$ and $\epsilon, \bar{\epsilon}, \delta_{\Delta(\mathcal{Y})}, \delta_\mathcal{P}, \delta_\mathcal{R} \to 0$ at the appropriate rates. This is the best possible guarantee in a stationary equilibrium of the repeated game where (a) it is common knowledge that y_t is drawn independently from the empirical distribution $\hat{\pi}_{P_t}$ and (b) the agent has access to an unknown information structure γ.

2.9 Conclusion

We studied single-agent mechanism design where the common prior assumption is replaced with repeated interaction and frequent feedback about the world. Our primary motivation was to remove a barrier (the common prior) that makes it difficult to implement mechanisms in practice. However, we also want to emphasize that this work can be viewed as a learning foundation for (robust) mechanism design. Indeed, our results show that policies similar to those predicted by a common prior can perform well even without making any assumptions about the data-generating process. This lends credibility to researchers who invoke the common prior for tractability, but do not expect it to be taken literally. However, there are two caveats.

1. Our policies are robust to agents that behave suboptimally by up to some $\epsilon > 0$. In contrast, most papers on local robustness involve an optimizing agent with misspecified beliefs (e.g. Artemov et al. 2013; Meyer-ter-Vehn and Morris 2011; Oury and Tercieux 2012). These notions coincide sometimes but not always. In addition, our policies sometimes require informational robustness (Bergemann and Morris 2013).

2. The number of interactions T required for our mechanisms to approximate the static common prior game may be large. In particular, our bounds depend on features of the stage game, like the size of the policy and response spaces, and the number of states. These features may also affect the agent's learning rate, which in turn affects our bounds. In that sense, the common prior assumption may be less appealing in more games that are more complex.

Further Work. There are several directions in which to generalize and improve this work. To begin, it is not clear whether our finite sample bounds have a tight dependence on the number of periods T and various other parameters. For example, is it possible to remove the exponential dependence in theorem 5 on the size of the policy space?[18] In addition, there may be opportunities for tightening our results in less abstract settings where the stage game has more structure.

Our analysis was restricted to single-agent problems. Suppose there are multiple agents. From the perspective of any one agent, her opponents correspond to adaptive adversaries (c.f. R. Arora, Dinitz, et al. 2018) whose future behavior is influenced by the agent's present response. However, if the number of participants is large and the mechanism's outcome preserves the differential privacy of each agent's response history (c.f. Manski and Tetenov 2007), the behavioral assumptions developed here may also be suitable for the multi-agent setting.

We assumed that the principal and agent observe the state after every interaction, but this may be unrealistic in many applications. For instance, in contract theory the state is a function from the agent's actions to outcomes. Let us briefly refer to the principal-agent problem in appendix B.1.2. There, if the agent chooses to work, we do observe whether the project succeeds or not. However, we may not learn whether the project would have succeeded had the agent shirked. To mitigate this issue, we could consider the case with bandit feedback, where participants observe their own payoffs but not the state. The challenge with bandit feeback is that it requires responsive mechanisms, as the mechanism must depend on the principal's payoffs, which in turn depend on the agent's response.[19]

[18] One approach the principal might take is to attempt to discern the agent's beliefs from the description of her learner L, and substitute those beliefs for his own forecast. If successful, this would tie the principal's forecast miscalibration to the agent's counterfactual internal regret.

[19] Relatedly, Balcan, Blum, Haghtalab, et al. (2015) consider a repeated Stackelberg game where the state is the agent's private type. The principal receives bandit feedback: he never observes the type directly but can infer it from the agent's behavior. The issues associated with responsiveness do not arise in this model as the agent is myopic (or more precisely, there is a sequence of short-lived agents).

In section 2.8, where the agent may be more informed than the principal, the principal's regret did not vanish but rather converged to the cost of informational robustness under a common prior. There is reason to believe that this result is not tight. Although the principal will never have access to the private signal I of the agent, he may attempt to learn (via the agent's past behavior) about the information structure γ that generates it. In turn, the agent may anticipate this and attempt to manipulate the principal's policy by feigning (partial) ignorance of her private signal. This suggests a less conservative definition of informational robustness, where the principal learns the quality of any information that the agent decides to exploit. However, in the repeated game, this approach would require responsive mechanisms.

As the last two paragraphs illustrate, we need a theory of behavior for responsive mechanisms. The no-regret conditions used here and elsewhere are not as well-motivated when the mechanism (or adversary) is responsive, insofar as they do not generalize traditional rationality assumptions. Extending the logic of no-regret conditions to a larger set of mechanisms – but not necessarily all mechanisms – is a clear priority for further work.

Chapter 3

Mechanism Design with a Common Dataset

3.1 Introduction

It is a truism in economics that beliefs are important determinants of behavior. In any number of settings, ranging from vaccine distribution to compensation policy, understanding agent's beliefs can make the difference between a successful policy and one that fails dramatically. Unfortunately, in many instances, the rich behavioral or survey data needed to identify agents' beliefs may not be available. This can limit our ability to provide clear policy recommendations.

While existing modeling techniques can circumvent our ignorance of agents' beliefs, they are widely recognized as imperfect. On one extreme, it is common to assume that the agent shares the policymaker's beliefs or that she knows the data-generating process. These assumptions have been criticized as unrealistic, both in economic theory (e.g. R. Wilson 1987, Bergemann and Morris 2005) and in econometrics (e.g. Manski 1993, Manski 2004a). On the other extreme lie robust solution concepts that make no assumptions about the agents' beliefs. But these solution concepts lead to unreasonably conservative recommendations in many applications.

This paper proposes a new, less extreme approach: rather than assume a common prior belief, assume access to a common dataset. The high-level idea is straightforward. If the data convincingly demonstrates some fact about the world, the agent should believe that fact. But if there is insufficient data to reach a particular conclusion, the agent's beliefs are undetermined.

I formalize this approach by integrating a model of statistical learning into mechanism design. To develop this approach, I restrict attention to single-agent problems where a policymaker commits to a policy, an agent responds, and payoffs are determined by a state of nature. Both the policymaker and agent have access to an i.i.d. dataset that they can use to learn about the distribution of the state. *Regret bounds* limit how suboptimal the agent's action can be with respect to the true distribution. A regret bound is *feasible* if even an ignorant agent can satisfy it using off-the-shelf learning rules. Typical regret bounds tighten as the sample size grows.

I derive feasible regret bounds and propose *penalized policies* that highlight new tradeoffs for the policymaker. Policy choices can influence the complexity of the agent's learning problem, which in turn affects her regret bound. Policies that are too complex, in a precise sense, can increase the likelihood that the agent makes a mistake, as well as the severity of that mistake. Penalized policies implicitly penalize complexity, by guarding against the worst-case mistake by the agent.

I propose a *data-driven penalized policy* and present the key technical results of the paper. Since the optimal penalized policy depends on the true data-generating process, it is generally not feasible. However, the policymaker can learn from the common dataset, just like the agent. The fact that the policymaker is learning at the same time as the agent poses theoretical challenges that appear to be new to both statistics and economics. Nonetheless, I develop a data-driven penalized policy and characterize its rate of convergence, which is approximately optimal. In the limit as the sample size grows, this policy coincides with the optimal policy under the common prior.

Finally, I show that penalization can lead to new insights through illustrative examples. Specifically, I consider models of vaccine distribution, prescription drug approval, performance pay, and product bundling. This framework captures important dimensions of policy complexity and highlights trade-offs that are obscured by both the common prior and robust solution concepts.

Model. I consider a rich class of incomplete information games where a policymaker commits to a policy and a single agent responds. Payoffs are determined by the policy, the response, and a hidden state of nature. The state is drawn from some unknown distribution. This setup captures a number of classic design problems in economics, like monopoly regulation, Bayesian persuasion, and contract design. Traditionally, these models are solved by assuming the state distribution is common knowledge, or that the policymaker and the agent have common prior beliefs.

Rather than assume a common prior belief, I assume that both the policymaker and the agent have access to a common dataset. This dataset consists of n i.i.d. draws from the state distribution. Both participants have strategies that map the data to actions. If there were only one participant, this would be a standard statistical decision problem (see e.g. Wald 1950, Manski 2004b).

Agent's Behavior. The goal of this paper is to produce a strategy for the policymaker that performs well under reasonable assumptions on the agent's behavior. To formalize these assumptions, I adapt ideas from statistical learning theory.

I impose bounds on the agent's *regret* that vary with sample size and the policymaker's strategy. Regret measures how suboptimal the agent's strategy is, in expectation, according to the true distribution. An agent that knows the true distribution can guarantee zero regret. An agent whose beliefs are only somewhat misspecified will obtain low regret. An agent who fails to learn from the available data, or has deeply misspecified beliefs, will obtain high regret.

I restrict attention to *feasible* regret bounds. These are regret bounds that the agent can satisfy using off-the-shelf heuristics like empirical utility maximization, even if she has no prior knowledge of the true distribution. I do not assume that the agent uses these heuristics. Instead, I assume that she does not underperform these heuristics. She can be Bayesian or non-Bayesian, well-informed or poorly-informed, and still satisfy a feasible regret bound.

I provide sufficient conditions for a regret bound to be feasible by borrowing a central concept from statistical learning theory: *Rademacher complexity*. This measures the complexity of a statistical learning problem. Naturally, an agent facing a more complex learning problem should be more likely to accumulate regret. Since the agent's learning problem is linked to the policy choice, Rademacher complexity captures a form of policy complexity from the agent's perspective.

To complete the specification of the sufficient condition, I introduce a new concept called *sample privacy*. This concept is closely related to differential privacy (Dwork, McSherry, et al. 2006). It measures how aggressively the policymaker makes use of the realized sample. Sample privacy is a bridge between traditional statistical learning problems and economic problems where multiple participants learn from the same data. The agent's learning problem looks more traditional if the policymaker limits how aggressively he uses the available data, and standard tools like Rademacher complexity remain useful.

Penalized Policy. This model formalizes a sense in which complex policies are undesirable. Policies that are more complex or more sensitive to the data lead to looser regret bounds. As a result, the agent's behavior becomes less predictable. For a policymaker that is concerned with worst-case guarantees, less predictable behavior can only make him worse off.

I propose *penalized policies* as a way to handle this new trade-off. To make the agent's behavior more predictable, a policymaker can set policies that are simpler and less sensitive to the data. But these changes can be costly. A penalized policy balances the advantages of complexity with the disadvantages of unpredictability. More precisely, it evaluates policies according to the policymaker's payoff under the worst-case agent response that respects the regret bound. This implicitly penalizes policies that are too complex for the agent, given the amount of data available.[1]

Data-Driven Penalized Policy The optimal penalized policy can generate useful insights, but it still depends on the true distribution. Fortunately, this paper shows that it is possible to approximate the optimal penalized policy by using the available data.

Specifically, I construct a *data-driven penalized policy.* There are three steps. First, it evaluates the policymaker's expected payoff with respect to the empirical distribution. Second, it adds white noise to the policymaker's payoff, following the exponential mechanism of Manski and Tetenov (2007). Third, it estimates the set of behaviors that satisfy the agent's regret bound, and optimizes against the worst case behavior in this estimated set.

Theorem 8 shows that the data-driven penalized policy satisfies sample privacy. They key challenge here is that sample privacy is a property of the policymaker's strategy, but I cannot define the strategy without specifying sample privacy parameters for the agent's regret bound.

Theorem 9 shows that the rate of convergence of the data-driven penalized policy is approximately optimal. First, I show that the policymaker's expected payoff converges as the sample size grows. More precisely, it converges to his optimal payoff in a hypothetical model where both participants know the true distribution. Next, I evaluate the rate of convergence. The optimal rate is that of the optimal penalized policy, where the policymaker knows the true distribution but the agent is still learning. If the optimal rate is $n^{-\gamma}$, I show that the data-driven penalized policy converges at a $n^{-\frac{\gamma}{1+2\gamma}}$ rate. In typical applications where $\gamma = 1/2$, the rate of convergence is $n^{-1/4}$.

[1] This is an important point of contrast with the large literature on penalization and regularization in statistics. Here, the goal is not to penalize policies that are too complex for the policymaker to learn effectively. Instead, the policymaker wants to penalize policies that complicated *the agent's* learning problem.

Illustrative Examples. I show that penalized policies can provide new insights in four examples.

First, I consider a model of vaccine distribution. Here, the penalized policy waits for statistically-significant clinical trial results before distributing vaccines. While this is common practice, it has been criticized (Wasserstein and Lazar 2016) and conflicts with the recommendations of the treatment choice literature (Manski 2019). My model offers a *strategic* reason to insist on statistical significance: the population needs to be convinced of vaccine quality. People may not take up the vaccine if it has not been proven effective beyond a reasonable doubt. If there are fixed costs to vaccine distribution, this outcome is worse than not distributing the vaccine at all.

Second, I consider a model of prescription drug approval by a regulator. Here, the penalized policy restricts doctors' ability to prescribe drugs that haven't been proven effective in clinical trials. Limiting the number of drugs approved reduces the risk that doctors fall prey to false positives, in which an ineffective drug appears to be effective by random chance. The penalized policy sets a standard for approval that increases as more drugs are approved, as in stepwise methods for multiple hypothesis testing (e.g. Holm 1979, Romano and Wolf 2005). In contrast, alternative approaches lead to extreme recommendations. If the doctors knew the true distribution of treatment effects, the optimal policy would approve all drugs. Meanwhile, robust policies approve no drugs at all.

Third, I consider a model of performance pay. An employer incentivizes an employee to exert costly effort by paying wages contingent on observed performance. If both participants know the distribution of performance conditional on effort, the optimal contract only compensates the employee at extreme levels of performance. The payment conditional on extreme performance may be extreme as well. In my model, this does not work well. Given limited data, it is not clear that it is worth investing effort for a small chance of receiving a bonus. To address this, the penalized contract caps and flattens the wage schedule. As before, wages are zero until performance reaches a threshold, at which point they jump and remain flat. But the threshold is lower and the payments smaller, allowing the agent to obtain moderate wages for moderate performance.

Finally, I consider a model of product bundling. A firm has several products for sale and wants to sell them in a way that maximizes expected profit. Here, the penalized policy favors selling large bundles of products, or even bundling all products together into a grand bundle. The reason for this is that consumers learn about their value for the product through reviews. If there are many products, but few reviews per product, consumers can be confident in the value of the grand bundle while being uncertain about the value of any given product. In that case, all else equal, it is easier to convince consumers to buy the bundle. I contrast this conclusion with prior work that recommends selling all items separately (Carroll 2017).

Related Literature This work contributes to three research efforts. For robust mechanism design, it is a principled way to interpolate between two extremes: the common prior and prior-freeness. In that respect, I share a goal with Artemov et al. (2013) and Ollár and Penta (2017), although my methods are quite different. For learning in games, I provide a convenient behavioral assumption that relies on tools from statistical learning theory. Liang (2020) studies a similar model but takes a more abstract approach to modeling agent behavior. For data-driven mechanism design, I study scenarios in which the agent is learning from data, not just the policymaker. Related work includes Camara et al. (2020), Cummings et al. (2020), and Immorlica, Mao, et al. (2020).

I leave a more detailed discussion of the prior literature to Section 3.7.

Organization. I introduce the model in Section 3.2. Section 3.3 formalizes the behavioral assumptions, with Rademacher complexity defined in Subsection 3.3.2 and sample privacy in Subsection 3.3.3. Section 3.4 defines the penalized policy. Section 3.5 proposes and evaluates the data-driven penalized policy. Section 3.6 presents four illustrative examples. Section 3.7 discusses related literature. Section 3.8 concludes. Omitted proofs can be found in Appendix **??**.

3.2 Model

This paper studies Stackelberg games of incomplete information. There are two players: a policymaker and an agent. The policymaker moves first and commits to a policy $p \in \mathcal{P}$. After observing the policy p, the agent chooses a response $r \in \mathcal{R}$, or possibly a mixed response $\pi^r \in \Delta(\mathcal{R})$. Finally, the state of nature $s \in \mathcal{S}$ is drawn from a distribution $\pi^s \in \Delta(S)$. The policymaker's utility is given by $u^P(p, r, s)$ and the agent's utility by $u^A(p, r, s)$.

So far, this setup is quite general. It can be used to model a variety of single-agent problems in mechanism design, contract design, information design, and other areas of interest to economists. Features like transfers, messages between the policymaker and agent, and asymmetric information about the state can be incorporated by defining the policy and response spaces appropriately.[2]

I maintain the following assumption throughout the paper.

Assumption 15. *The space $\mathcal{P}$ consists of $n_{\mathcal{P}} < \infty$ policies and utility functions u^A, u^P are bounded.*

The reader is welcome to think of $\mathcal{P}$ as a discrete approximation to an infinite policy space. Assumption 15 also ensures that, for any fixed policy p, the maximum variation in each participant's utility function is finite. Let $i \in \{A, P\}$ indicate the agent or policymaker,

[2] For example, if the policymaker has access to transfers, then the policy p must specify the transfers made (if any). Alternatively, if the agent has private information about some aspect of the state s, then her response r should map that information to some action. Finally, if the agent reports her private information to the policymaker, then p should map that report to some action.

and define

$$\Delta^i(p) = \sup_{r,s} u^i(p,r,s) - \inf_{r,s} u^i(p,r,s) < \infty \tag{3.1}$$

I will impose additional regularity assumptions, as needed, later on in the paper.

Common Knowledge. One way to solve this model is to assume that both the policymaker and the agent know the distribution π^s, or equivalently, that π^s represents their shared prior belief over the state s. Here, the agent is free to choose any response r that maximizes her expected utility, i.e.

$$r \in \arg\max_{r'} \mathrm{E}_{\pi^s}\left[u^A(p,r',s)\right] \tag{3.2}$$

In turn, the policymaker chooses a policy p that maximizes his expected utility after taking into account how the agent will respond.

If equation (C.14) has multiple solutions, then the policymaker may not know which response r the agent will choose. One way to deal with this is to specify a tie-breaking rule that determines which response r the agent chooses when she is indifferent between multiple responses.[3] Another approach is to remain agnostic to how the agent breaks ties and optimize against the worst-case best response. In that case, the policymaker can guarantee an expected utility of

$$\begin{aligned} \mathrm{CK}(\pi^s) = &\max_p \min_r \mathrm{E}_{\pi^s}\left[u^P(p,r,s)\right] \\ \text{s.t.} \quad & r \in \arg\max_{r'} \mathrm{E}_{\pi^s}\left[u^A(p,r',s)\right] \end{aligned} \tag{3.3}$$

[3] A typical tie-breaking rule assumes that the agent breaks ties in favor of the policymaker. This assumption is convenient in games with an infinite policy space because it ensures that an optimal policy exists. Furthermore, it is often innocuous insofar as small perturbations of the optimal policy can be used to break indifferences in favor of the policymaker with negligible loss of optimality. Here, the policy space is finite so existence of an optimal policy is not an issue. Furthermore, in games where the favorable tie-breaking assumption is innocuous, the common prior benchmark will be roughly the same regardless of whether I assume favorable tie-breaking or not.

I refer to $\mathrm{CK}(\pi^s)$ as the *common knowledge benchmark*.

The common knowledge assumption has two obvious drawbacks. First, the optimal policy will generally depend on the true distribution π^s and the policymaker may not know π^s. If our goal is to consult the policymaker, we cannot recommend a policy to him that relies on information that he does not have access to. Second, the agent may not know the true distribution π^s. In that case, she may choose responses r that are suboptimal in the sense that they do not solve equation (3.2), and the policy that solves (3.3) may no longer be optimal.

Similar issues arise when the distribution π^s is interpreted as a common prior belief. The policymaker may not be willing to specify a prior belief over the state s, or be concerned that his beliefs are too uninformed. The agent may not agree with the policymaker's prior belief, especially if that belief is not well-informed.

Prior-Free Solution Concepts. Another way to solve this model is to not make any assumptions about the agent's beliefs $\tilde{\pi}$. These approaches are called prior-free, belief-free, or belief-robust.

Suppose the policymaker knows the true distribution π^s, but does not know anything about the agent's beliefs. In that case, the *maximin policy* maximizes the policymaker's worst-case expected utility. That is,

$$\begin{aligned} \mathrm{MM}(\pi^s) &= \max_p \min_{\tilde{\pi}^s, r} \mathrm{E}_{\pi^s}\left[u^P(p, r, s)\right] \\ \text{s.t.} \quad & r \in \arg\max_{r'} \mathrm{E}_{\tilde{\pi}^s}\left[u^A(p, r', s)\right] \end{aligned} \tag{3.4}$$

The objective is evaluated with respect to the worst-case belief $\tilde{\pi}^s \in \Delta(\mathcal{S})$.

There are alternatives to the maximin policy. For example, the *minimax regret policy* minimizes the policymaker's worst-case regret from following policy p, rather than the policy

p' that would have been optimal given the agent's response r. That is,

$$\begin{aligned} \mathrm{MR}(\pi^s) = \min_p \max_{\tilde{\pi}^s, r} & \left(\max_{p', r'} \mathrm{E}_{\pi^s}\left[u^P(p', r', s)\right] - \mathrm{E}_{\pi^s}\left[u^P(p, r, s)\right] \right) \qquad (3.5) \\ \text{s.t.} \quad & r, r' \in \arg\max_{r'} \mathrm{E}_{\tilde{\pi}^s}\left[u^A(p, r', s)\right] \end{aligned}$$

Again, the objective is evaluated with respect to the worst-case belief $\tilde{\pi}^s \in \Delta(\mathcal{S})$.

The advantage of maximin and minimax regret policies are that they do not require the agent to know the true distribution π^s, or to agree with the policymaker's beliefs. The disadvantage is that they can be extremely conservative: $\mathrm{MM}(\pi^s)$ or $\mathrm{MR}(\pi^s)$ may only be a small fraction of $\mathrm{CK}(\pi^s)$. This is especially true in applications like contract design and Bayesian persuasion where the agent's optimal response varies considerably in her beliefs over the state.

Common Dataset. I propose a third way to solve this model: one that avoids key drawbacks of the two existing approaches. Rather than assume common knowledge of the distribution, I assume access to a common dataset. Formally, a dataset consists of n i.i.d. observations of the state, i.e.

$$S_1, \ldots, S_n \sim \pi^s$$

Each participant's strategy is now a statistical decision rule. The *policymaker's strategy* maps the dataset to a distribution over policies:

$$\sigma_n^P : \mathcal{S}^n \to \Delta(\mathcal{P})$$

The *realized policy* is

$$P_n \sim \sigma_n^P(S_1, \ldots, S_n)$$

The *agent's strategy* maps the dataset and the policymaker's policy to a distribution over responses:

$$\sigma_n^A : \mathcal{S}^n \times \mathcal{P} \to \Delta(\mathcal{R})$$

The *realized response* is

$$R_n \sim \sigma_n^A(S_1, \ldots, S_n, P_n)$$

Both P_n and R_n are random variables, because (i) they depend on the random sample and (ii) the policymaker and agent may use mixed strategies.

Remark 5. It is worth emphasizing that this dataset is ideal in several ways.

1. First, the states of nature S_i are directly observed. This is a good starting point because the data clearly identifies the distribution π^s. In practice, however, the state of nature may not be observed directly, and the distribution π^s may not be point identified by the available data. That raises the additional hurdle of partial identification.

2. Second, the observations S_i are drawn independently from the true distribution π^s. This is a standard assumption, but may not hold in dynamic environments where historical data does not fully reflect the present. It is possible to drop this assumption with an alternative approach developed in Camara et al. (2020). However, that approach requires the policymaker to interact repeatedly with the same agent. Here, I only require a single interaction.

3. Third, the dataset is available to both participants. More precisely, any data that the policymaker uses must also be available to the agent (it is not a problem if the agent has access to additional data). In some cases, the policymaker may be able to guarantee that this assumption holds by sharing his data with the agent. In other cases, the policymaker may prefer to keep the data confidential.

These are important caveats, but at least they are explicit. By being explicit about what data is available, it is possible to have a productive discussion over what the participants are likely to know and what they may not know. In contrast, common knowledge assumptions bypass any discussion of how the policymaker and agent actually arrived at said knowledge.

In the next section, I propose new rationality assumptions that restrict the agent's strategy σ_n^A.

3.3 Agent's Behavior

I propose new rationality assumptions that restrict the agent's behavior. The basic idea is that, since she has access to a dataset, the agent should not underperform standard heuristics for learning from data. I formalize this assumption by introducing *feasible regret bounds*. Then I provide sufficient conditions for a regret bound to be feasible.

The key object of interest is the agent's *regret*, which captures how suboptimal her strategy is.[4]

Definition 42. *The agent's* regret *is the difference between her optimal expected utility and the expected utility she achieves by following her strategy σ_n^A. Formally,*

$$\text{Regret}_n^A\big(\sigma_n^A, \sigma_n^P, \pi^s\big) = \max_r \mathrm{E}_{\pi^s}\big[u^A(P_n, r, s)\big] - \mathrm{E}_{\pi^s}\big[u^A(P_n, R_n, s)\big]$$

To be clear, expectations are taken with respect to the state s, the realized sample $S_1, \ldots, S_n$, and any internal randomization associated with mixed strategies. The sample $S_1, \ldots, S_n$ comes in through the realized response R_n and policy P_n.

It turns out that a refined notion of regret will be more useful. The policymaker not only cares about whether the agent makes mistakes, but also about how those mistakes are correlated with the policies that he chooses.

[4] This is not the same as the ex-post notion of regret used in the literature on learning in games.

Definition 43. *The agent's* conditional regret *is her regret conditional on the realized policy $P_n = p$.*

$$\text{Regret}_n^A\left(\sigma_n^A, \sigma_n^P, \pi^s \mid P_n = p\right) = \max_r \mathrm{E}_{\pi^s}[u(p, r, s)] - \mathrm{E}_{\pi^s}[u(p, R_n, s) \mid P_n = p]$$

To be clear, this is not related to contextual regret, or regret conditioned on private information (as in Camara et al. 2020). The realized policy does not convey any information about the distribution π^s that the agent does not already have access to. However, the realized response R_n and the policy P_n may be correlated because they depend on the same random sample. So the conditional expectation

$$\mathrm{E}_{\pi^s}[u(p, R_n, s) \mid P_n = p]$$

may be different from the unconditional expectation used to define regret.

3.3.1 Regret Bounds

A *regret bound* $B\left(\sigma_n^P, \pi^s, p\right)$ is a function of the policymaker's strategy σ_n^P, the realized policy $P_n = p$ and the distribution π^s that bounds the agent's conditional regret. The agent is allowed to use any strategy σ_n^A that satisfies

$$\text{Regret}_n^A\left(\sigma_n^A, \sigma_n^P, \pi^s \mid P_n = p\right) \leq B\left(\sigma_n^P, \pi^s, p\right)$$

This is an upper bound: it is always possible that the agent outperforms this bound. In particular, an agent that knows the true distribution π^s will satisfy this bound because her regret is zero.

Remark 6. Allowing the agent to obtain positive regret (or conditional regret) is one way to relax the prior knowledge assumption. I argue that this approach has several important advantages.

1. I do not require the researcher to specify a set of prior beliefs that the agent could reasonably possess (see e.g. Artemov et al. 2013, Ollár and Penta 2017).[5] This is difficult to do without guidance on which prior beliefs are reasonable. In contrast, I can provide guidance for how to choose regret bounds by considering measures of statistical complexity.

2. I do not rule out the possibility that the agent is Bayesian or that she has prior knowledge about the distribution π^s that goes beyond the common dataset. For example, the unbiased inference procedures in Salant and Cherry (2020) rule out this possibility. Regret bounds are intended to relax standard assumptions, not to contradict them.

3. I do not require the agent to be Bayesian: she is welcome to use strategies σ_n^A that are not consistent with Bayesian updating, as long as they satisfy the regret bound. Non-Bayesian (e.g. frequentist) methods for learning from data are commonly used in practice, including by empirical economists. It is reassuring that regret bounds allow for this possibility.

Not all regret bounds are compelling. I restrict attention to *feasible* regret bounds, which the agent can satisfy even if she has no prior knowledge of the distribution π^s.

Definition 44. *A regret bound B is* feasible *given policymaker's strategy σ_n^P if there exists an agent strategy $\tilde{\sigma}_n^A$ such that, for all policies p and distributions $\tilde{\pi}^s$,*

$$\mathrm{Regret}_n^A\left(\tilde{\sigma}_n^A, \sigma_n^P, \tilde{\pi}^s \mid P_n = p\right) \leq B\left(\sigma_n^P, \tilde{\pi}^s, p\right)$$

[5] This often involves committing to a metric ρ on the space of beliefs, and assuming that the agent's beliefs are within some distance ϵ of the true distribution. That approach is consistent with a regret bound for reasonable choices of ρ where optimizing against approximately correct beliefs leads to approximately optimal responses.

The strategy $\tilde{\sigma}_n^A$ may be distinct from the agent's actual strategy σ_n^A.

Feasible regret bounds make for reasonable rationality assumptions. After all, an agent whose strategy σ_n^A systematically underperforms a feasible regret bound would benefit from deviating to the strategy $\tilde{\sigma}_n^A$ that satisfies it. This deviation would reduce her regret, or equivalently, increase her expected utility according to the true distribution. It would require no prior knowledge of the distribution π^s, since $\tilde{\sigma}_n^A$ does not depend on π^s.

Infeasible regret bounds, however, make for dubious rationality assumptions. In that case, there does not exist a single strategy $\tilde{\sigma}_n^A$ that satisfies the regret bound across all distributions $\tilde{\pi}^s$. It is still possible for the agent to satisfy the regret bound under the true distribution π^s, but she would need prior knowledge about π^s that goes beyond the dataset. Although I do not want to *rule out* the possibility that the agent has prior knowledge, I also do not want to *require* prior knowledge.

3.3.2 Rademacher Complexity

In this subsection, I provide a sufficient condition for a regret bound to be feasible in the special case where the policymaker ignores the data. I begin by defining Rademacher complexity, a key concept from statistical learning theory, and stating the sufficient condition. Then I provide intuition.

A Rademacher random variable σ is uniformly distributed over $\{-1, 1\}$.

Definition 45 (Bartlett and Mendelson 2003)**.** *The* Rademacher complexity *induced by policy p is*

$$\mathcal{RC}_n^A(p, \pi^s) = \mathrm{E}_{\pi^s}\left[\max_r \frac{1}{n}\sum_{i=1}^n \sigma_i \cdot u^A(p, r, S_i)\right]$$

where $\sigma_1, \ldots, \sigma_n$ are i.i.d. Rademacher random variables.

The policymaker's strategy σ_n^P is *constant* if there exists a distribution $\pi^p \in \Delta(\mathcal{P})$ such that, for all sample realizations $S_1, \ldots, S_n$,

$$\sigma_n^P(S_1, \ldots, S_n) = \pi^p$$

Proposition 7. *Let the policymaker's strategy σ_n^P be constant. Then B is a feasible regret bound if*

$$\forall p, \tilde{\pi}^s, \quad B\left(\sigma_n^P, \tilde{\pi}^s, p\right) \geq 4\mathcal{RC}_n^A(p, \tilde{\pi}^s)$$

The Rademacher complexity bounds the agent's regret if she uses a particular strategy called *empirical utility maximization.* This strategy will play the role of $\tilde{\sigma}_n^A$ in the definition of feasibility.

Definition 46. Empirical utility maximization $\hat{\sigma}_n^A$ *is an agent strategy. It chooses the response r that maximizes the agent's expected utility with respect to the empirical distribution $\hat{\pi}^s$. Formally,*

$$\hat{\sigma}_n\left(S_1, \ldots, S_n\right) \in \arg\max_r \frac{1}{n} \sum_{i=1}^n u^A\left(P_n, r, S_i\right)$$

Let $\hat{R}_n = \hat{\sigma}_n\left(S_1, \ldots, S_n\right)$ denote the empirical utility maximizer.

If the policymaker's strategy σ_n^P is constant, then

$$\forall p, \tilde{\pi}^s, \quad \text{Regret}_n^A\left(\hat{\sigma}_n^A, p, \tilde{\pi}^s\right) \leq 4\mathcal{RC}_n^A(p, \tilde{\pi}^s) \tag{3.6}$$

This follows immediately from well-known results in Bartlett and Mendelson (2003).

Intuitively, Rademacher complexity measures the potential for empirical utility maximization to overfit to sampling noise. First, it trivializes the agent's learning problem by randomizing the sign of the agent's utility function. That is, it replaces the utility function $u^A(p, r, s)$ with $\sigma \cdot u^A(p, r, s)$, where σ is a Rademacher random variable. This modified learning problem is trivial since all responses are equally good: expected utility is zero for every response r, i.e.

$$\mathrm{E}_{\pi^s}\left[\sigma \cdot u^A(p, r, s)\right] = 0$$

Second, Rademacher complexity asks how much the empirical utility maximizer $\hat{R}_n$ will overfit to this modified problem. Formally, the empirical utility is

$$\frac{1}{n}\sum_{i=1}^{n} \sigma_i \cdot u^A\left(p, r, S_i\right)$$

where σ_i are i.i.d. Rademacher random variables. For any given response r, the empirical utility is zero in expectation. On the other hand, the empirical utility of $\hat{R}_n$ will generally have a positive expected value in finite samples. This expected value is the Rademacher complexity, and reflects how severely the agent can be misled by the sampling noise in $\sigma_1, \ldots, \sigma_n$.

Note that the Rademacher complexity will typically converge to zero as $n \to \infty$. In other words, the agent's learning problem typically becomes easier as she obtains more data. In addition, the Rademacher complexity generally depends on the policy p. The policymaker can increase or decrease the Rademacher complexity through his choice of policies.

3.3.3 Sample Privacy

In this subsection, I generalize the sufficient condition for a regret bound to be feasible. The condition in proposition 7 only applies when the policymaker ignores the data. I show that a similar condition applies as long as the policymaker does not use the data too aggressively. I formalize this requirement through a new concept called *sample privacy.*

Definition 47. *The policymaker's strategy σ_n^P satisfies* (ϵ, δ)-sample privacy *if there exists an event $E \subseteq \mathcal{S}^n$ where (i) E occurs high probability, i.e.*

$$\Pr_{\pi^s}[(S_1, \ldots, S_n) \in E] \geq 1 - \delta$$

and (ii) after conditioning on E, the sample is nearly independent of the realized policy, i.e.

$$(S_1, \ldots, S_n) \in E \implies \Pr_{\pi^s}[P_n = p \mid S_1, \ldots, S_n] \leq e^{\epsilon} \cdot \Pr_{\pi^s}[P_n = p \mid E]$$

Note that the realized policy P_n may be random even after conditioning on the sample $S_1, \ldots, S_n$, if the policymaker uses a mixed strategy.

The concept is closely related to *differential privacy* (Dwork, McSherry, et al. 2006). When applied to the policymaker's strategy σ_n^P, differential privacy ensures that the realized policy P_n does not change much when any one observation $S_i = s$ is replaced with another value $S_i = s'$. In contrast, sample privacy ensures that the policy P_n does not change much when the entire sample $S_1, \ldots, S_n$ is dropped and replaced with a new sample $S_1', \ldots, S_n'$. Critically, the new sample is drawn from the same distribution π^s as the original sample.

If the policymaker's strategy σ_n^P is constant then it satisfies $(0, 0)$-sample privacy. This is because the realized policy P_n is independent of the sample $S_1, \ldots, S_n$. Indeed, proposition 7 applies to any strategy σ_n^P that satisfies $(0, 0)$-sample privacy. From that perspective, proposition 8 shows that the lower bound in proposition 7 increases smoothly as the parameters (ϵ, δ) increase.

Proposition 8. *Let the policymaker's strategy σ_n^P satisfy (ϵ, δ)-sample privacy. Then B is a feasible regret bound if*

$$\forall p, \tilde{\pi}^s, \quad B\left(\sigma_n^P, \tilde{\pi}^s, p\right) \geq 4e^{\epsilon} \cdot \mathcal{RC}_n^A(p, \tilde{\pi}^s) + \delta \cdot \Delta^A(p)$$

The proof shows that if the policymaker's strategy σ_n^P satisfies sample privacy then conditioning on $P_n = p$ does not have a meaningful impact on any moment of the sample. In particular, the agent's conditional regret is a moment of the sample. There is a similar result in differential privacy (Dwork and A. Roth 2014, section 2.3.1).

The fact that sample privacy is needed at all suggests an important difference between a standard learning problem and the *concurrent learning* problem that I study. From the agent's perspective, a standard learning problem is one in which her utility function $u^A(p, r, s)$ does not depend on the realized sample $S_1, \ldots, S_n$. But when the policymaker is learning concurrently, the agent's utility function $u^A(P_n, r, s)$ depends on the sample through the realized policy P_n. Standard techniques for bounding the agent's regret may no longer apply. I give a concrete example of this in section 3.6.2.

3.4 Penalized Policy

This model formalizes a sense in which complex policies are undesirable. As policies become more complex, feasible regret bounds tend to loosen, and the agent's behavior becomes less predictable. I propose a *penalized policy* that balances the advantages of complexity with the disadvantages of unpredictability. I show that penalization can lead to new insights in four illustrative examples.

First, I restrict attention to regret bounds B that take on a particular form. I require a distribution-free upper bound on the Rademacher complexity, i.e.

$$\overline{\mathcal{RC}}_n^A(p) \geq \max_{\pi^s} \mathcal{RC}_n^A(p, \pi^s) \tag{3.7}$$

It follows immediately from proposition 8 that

$$B\left(\sigma_n^P, p, \pi^s\right) := e^{\epsilon} \cdot \overline{\mathcal{RC}}_n^A(p) + \delta \cdot \Delta^A(p)$$

is a feasible regret bound.

Remark 7. There are many well-known distribution-free upper bounds on the Rademacher complexity, based on measures like the VC dimension, Pollard's pseudo-dimension, and the covering number. For example, a useful bound due to Massart (2000) applies to finite response spaces with $n_{\mathcal{R}} < \infty$ elements. In that case,

$$\overline{\mathcal{RC}}_n^A(p) := \Delta^A(p) \cdot \sqrt{\frac{2 \ln n_{\mathcal{R}}}{n}}$$

is an upper bound on the Rademacher complexity.

These regret bounds are convenient because they only depend on the policy p and privacy parameters (ϵ, δ). Assuming I can guarantee sample privacy, this makes it possible to evaluate the policymaker's utility from a given policy p without having to consider the strategy σ_n^P that generated that policy. In particular, I can evaluate the policymaker's utility with respect to the worst-case mixed response π^r that satisfies the agent's regret bound:

$$\begin{aligned} \mathrm{WC}_n(p, \epsilon, \delta, \pi^s) = &\min_{\pi^r} \mathrm{E}_{\pi^s, \pi^r}\left[u^P\left(p, r, s\right)\right] \\ &\text{s.t.} \quad \max_{r'} \mathrm{E}_{\pi^s}\left[u^A\left(p, r', s\right)\right] - \mathrm{E}_{\pi^s, \pi^r}\left[u^A\left(p, r, s\right)\right] \leq 4e^{\epsilon} \cdot \overline{\mathcal{RC}}_n^A(p) + \delta \cdot \Delta^A(p) \end{aligned} \tag{3.8}$$

The mixed response π^r reflects the marginal distribution of the realized response R_n conditional on the realized policy $P_n = p$.[6] Intuitively, an agent may choose an optimal response with high probability, when the realized sample is representative, and a suboptimal response with low probability, when the realized sample is misleading, and still be nearly-optimal in expectation.

If the policymaker knew the true distribution π^s, he could solve for the *optimal penalized policy*. In this case, the agent is still learning from the dataset, but the policymaker can ignore the realized sample and guarantee $(0, 0)$-sample privacy.

[6] Only the marginal distribution is relevant because the sample does not directly enter into the policymaker's utility. The sample only affects the distribution of the realized policy P_n, which I have already conditioned on.

Definition 48. *An* optimal penalized policy *is any policy p that solves*

$$\mathrm{OP}_n(\pi^s) = \max_p \mathrm{WC}_n(p, 0, 0, \pi^s) \tag{3.9}$$

In section 3.5, I develop a strategy for the policymaker that approximates the optimal penalized policy by learning from the available data. For now, I focus on the optimal penalized policy.

I call these penalized policies because the worst-case objective implicitly penalizes policies that the agent perceives as more complex, as measured by the Rademacher complexity. As in penalized regression and other forms of regularization in statistics, penalization biases the policymaker towards policies that are less complex. However, there is a key difference. In statistics, policies are considered complex if they make the policymaker's learning problem hard. Here, policies are considered complex because they make the agent's learning problem hard.

From the perspective of microeconomic theory, penalization is interesting because it highlights a new tradeoff related to policy complexity. The more complex a policy, the looser the regret bound, and the less predictable the agent's behavior. If the policymaker is ambiguity-averse, he will tend to choose policies that are less complex than if we had assumed common knowledge. This is especially pronounced when the sample size n is small.

3.5 Data-Driven Penalized Policy

It is useful to study the optimal penalized policy to understand the qualitative effect of penalization on policymaking, but the optimal penalized policy is not feasible unless the policymaker knows the distribution π^s in advance. In this section, I show that it is possible to approximate the optimal penalized policy using the available data.

I construct a strategy $\hat{P}_n$ for the policymaker. This modifies the naive estimator that evaluates the policymaker's worst-case utility with respect to the empirical distribution $\hat{\pi}^s$, i.e.

$$\begin{aligned} \mathrm{WC}_n(p,\epsilon,\delta,\hat{\pi}^s) = & \min_{\pi^r} \mathrm{E}_{\hat{\pi}^s,\pi^r}\left[u^P\left(p,r,s\right)\right] \\ \text{s.t.} \quad & \max_{r'} \mathrm{E}_{\hat{\pi}^s}\left[u^A\left(p,r',s\right)\right] - \mathrm{E}_{\hat{\pi}^s,\pi^r}\left[u^A\left(p,r,s\right)\right] \le 4e^{\epsilon}\cdot\overline{\mathcal{RC}}_n^A(p)+\delta \end{aligned} \tag{3.10}$$

For many problems in statistics, econometrics, and machine learning, replacing the true distribution with the empirical distribution is enough to come up with a good estimator. In this model, where the agent is learning in addition to the policymaker, two changes are needed.

First, I conservatively estimate the set of mixed responses that meet the agent's regret bound. The naive estimator assumes that the agent's empirical regret, i.e.

$$\max_{r'} \mathrm{E}_{\hat{\pi}^s}\left[u^A\left(p,r',s\right)\right] - \mathrm{E}_{\hat{\pi}^s,\pi^r}\left[u^A\left(p,r,s\right)\right]$$

satisfies the regret bound B. However, B is a bound on the agent's regret evaluated with respect to the true distribution π^s. If the sample is unrepresentative, mixed responses π^r that satisfy the regret bound may violate the constraint in equation (3.10) because their empirical regret overestimates their true regret. In that case, the constraint rules out mixed strategies that the agent might use.

I can address this by adding a *buffer* to the regret bound, so that the empirical regret minus the buffer is unlikely to overestimate the true regret. Let $\alpha \in (0,1)$ be a tuning parameter. Define the buffer as follows:

$$\mathrm{BFR}_n = 8\sqrt{\frac{2\ln 4}{n}} + 8\sqrt{-\frac{2\ln\exp(-n^{\alpha})}{n}}$$

Lemma 17. *With probability exceeding $1 - n_{\mathcal{P}} \exp(-n^{\alpha})$, the buffer exceeds the difference between empirical and true regret. That is,*

$$\begin{aligned} 4\mathcal{RC}_n^A(p, \pi^s) + \mathrm{BFR}_n \geq \Big| &\left(\max_{r'} \mathrm{E}_{\hat{\pi}^s}\left[u^A\left(p, r', s\right)\right] - \mathrm{E}_{\hat{\pi}^s, \pi^r}\left[u^A\left(p, r, s\right)\right]\right) \\ &- \left(\max_{r'} \mathrm{E}_{\pi^s}\left[u^A\left(p, r', s\right)\right] - \mathrm{E}_{\pi^s, \pi^r}\left[u^A\left(p, r, s\right)\right]\right) \Big| \end{aligned}$$

for all mixed responses π^r and policies p,

Second, I introduce white noise into the objective $\mathrm{WC}_n(\cdot)$ order to control the sample privacy of $\hat{P}_n$. The challenge here is that there is an inherent circularity. Sample privacy is a property of the policymaker's strategy $\hat{\sigma}_n^P$, but in order to define $\hat{\sigma}_n^P$ I need to specify the privacy parameters (ϵ, δ) in the agent's regret bound. I address this challenge in theorem 8.

More concretely, I ensure sample privacy by adapting the exponential mechanism proposed by Manski and Tetenov (2007). Let $\beta \in (0, \alpha/2)$ be another tuning parameter. I add noise from the Gumbel distribution into the policymaker's objective function, i.e.

$$\nu_n(p) \sim \textsc{Gumbel}\left(0, n^{-\beta}\right)$$

For any given $\epsilon > 0$, this estimator will satisfy (ϵ, δ_n)-privacy, where δ_n depends on tuning parameters α, β, ϵ and is decreasing exponentially in n. More precisely,

$$\delta_n = n_{\mathcal{P}} \exp\left(-\frac{\epsilon^2}{2K^2} \cdot n^{\alpha - 2\beta}\right) \tag{3.11}$$

where the constant K is defined by

$$K := \max_p \max\left\{\frac{2\Delta^P(p)^2}{8\sqrt{2}}, \Delta^P(p)\right\} \tag{3.12}$$

Incorporating this into equation (**??**) gives a noisy, conservative estimate of the worst-case utility.

$$\begin{aligned} \widehat{\mathrm{WC}}_n(p) = & \min_{\pi^r} \mathrm{E}_{\hat{\pi}^s,\pi^r}\left[u^P\left(p,r,s\right)\right] + \nu_n(p) \\ \text{s.t.} \quad & \max_{r'} \mathrm{E}_{\hat{\pi}^s}\left[u^A\left(p,r',s\right)\right] - \mathrm{E}_{\hat{\pi}^s,\pi^r}\left[u^A\left(p,r,s\right)\right] \\ & \leq \left(4e^{\epsilon}+4\right) \cdot \overline{\mathcal{RC}}_n^A(p) + \delta_n \cdot \Delta^A(p) + \mathrm{BFR}_n \end{aligned} \tag{3.13}$$

Definition 49. *The data-driven penalized policy $\hat{P}_n$ solves*

$$\hat{P}_n \in \arg\max_p \widehat{\mathrm{WC}}_n(p) \tag{3.14}$$

To summarize, $\hat{P}_n$ depends on three tuning parameters. The parameter $\alpha \in (0,1)$ controls how quickly the buffer BFR_n vanishes as n grows. The parameter $\beta \in (0, \alpha/2)$ controls how quickly the privacy-preserving noise vanishes as n grows. The parameter $\epsilon > 0$ controls how the two dimensions of sample privacy are balanced; decreasing ϵ means increasing δ_n, and vice-versa.

The next theorem verifies that this estimator obtains the privacy guarantees that were assumed in its definition. It holds for any fixed $\epsilon > 0$, with δ_n defined according to equation (3.11).

Theorem 8. *The estimator $\hat{P}_n$ guarantees (ϵ, δ_n)-sample privacy.*

This result holds under very weak assumptions on the underlying game (assumption 15), which makes it challenging to prove. I outline the proof in subsection 3.5.3.

3.5.1 Convergence

I show that, in the limit as the sample size grows, the policymaker's payoff under $\hat{P}_n$ converges to his optimal payoff in a model where the distribution is common knowledge. That is, common knowledge is the limiting phenomenon under $\hat{P}_n$, despite the challenges that arise when the policymaker and agent are learning simultaneously.

This requires a regularity condition that ensures that the agent and policymaker can learn the optimal response and policy in the easy case where their opponent is not also learning. The condition ensures that the Rademacher complexity of both the agent and policymaker's learning problem vanishes at a reasonable rate as the sample size grows.

Definition 50. *The policymaker's Rademacher complexity is defined over all policy-response pairs, i.e.*

$$\mathcal{RC}_n^P(\pi^s) = \mathrm{E}_{\pi^s}\left[\sup_{p,r} \frac{1}{n}\sum_{i=1}^{n} \sigma_i \cdot u^P(p, r, S_i)\right]$$

where $\sigma_1, \ldots, \sigma_n$ are i.i.d. Rademacher random variables.[7]

Assumption 16. *The agent and policymaker's Rademacher complexity vanish at the typical $\tilde{O}(n^{-1/2})$ rate, where the tilde in $\tilde{O}$ means "up to log factors". More precisely:*

1. *There exists a constant K^A such that*

$$\overline{\mathcal{RC}}_n^A(p) \leq K^A n^{-1/2} \log n$$

for all policies p and sample sizes n, where K^A does not depend on p or n.

2. *There exists a constant K^P such that*

$$\mathcal{RC}_n^P(\pi^s) \leq K^P n^{-1/2} \log n$$

for all distributions π^s and sample sizes n, where K^P does not depend on π^s or n.

[7] I use this quantity to bound the generalization error of $\hat{P}_n$. More precisely, for any given policy p, the generalization error is the difference between the performance of policy p according to the empirical distribution $\hat{\pi}^s$ and its actual performance under the true distribution π^s. Note that this error will generally depend on the agent's response r. To account for this, the policymaker's Rademacher complexity takes a supremum over responses r in addition to policies p. This ensures that the bound on the generalization error holds regardless of what the agent's response is or how it is influenced by the data.

This assumption is both easy to satisfy and stronger than necessary. It is easy to satisfy because it holds whenever the agent's response space is finite. It also holds whenever the relevant optimization problems have a finite VC dimension, or a finite pseudo-dimension. And the assumption is stronger than necessary because the $\tilde{O}(n^{-1/2})$ rate is not needed for proposition 9. I do make use of this rate for theorem 8, in order to give concrete rates of convergence for my estimator.

Proposition 9 establishes convergence. It refers to three quantities of interest. First, the common knowledge benchmark $\mathrm{CK}(\pi^s)$ (equation 3.3) describes the policymaker's optimal payoff when the distribution π^s is common knowledge. Second, the optimal penalized benchmark $\mathrm{OP}_n(\pi^s)$ (equation 3.9) describes the policymaker's optimal payoff when he knows the distribution π^s but the agent is still learning. Third, the performance of the data-driven penalized policy $\hat{P}_n$ is given by

$$\mathrm{E}_{\pi^s}\left[\mathrm{WC}_n\left(\hat{P}_n, \epsilon, \delta_n, \pi^s\right)\right]$$

It is easy to see that

$$\mathrm{E}_{\pi^s}\left[\mathrm{WC}_n\left(\hat{P}_n, \epsilon_n, \delta_n, \pi^s\right)\right] \leq \mathrm{OP}_n(\pi^s) \leq \mathrm{CK}(\pi^s)$$

I show that all three quantities coincide in the limit as $n \to \infty$.

Proposition 9. *Both the performance of $\hat{P}_n$ and the optimal penalized benchmark converge to the common knowledge benchmark, as $n \to \infty$. That is,*

$$\lim_{n\to\infty} \mathrm{E}_{\pi^s}\left[\mathrm{WC}_n\left(\hat{P}_n, \epsilon, \delta_n, \pi^s\right)\right] = \lim_{n\to\infty} \mathrm{OP}_n(\pi^s) = \mathrm{CK}(\pi^s)$$

I outline the proof in subsection 3.5.4.

3.5.2 Rate of Convergence

Establishing convergence is not enough to argue that the data-driven penalized policy $\hat{P}_n$ is a practical solution to the policymaker's problem. At a minimum, $\hat{P}_n$ should have a reasonable rate of convergence. In this subsection, I show that the rate of convergence of $\hat{P}_n$ is approximately as good as the rate of convergence of the optimal penalized policy. To do this, I need a richness assumption that rules out cases where the agent is indifferent between all of her responses.

First, I observe that there are games in which $\hat{P}_n$ cannot possibly have a good rate of convergence. This is because even the optimal penalized policy has a poor rate of convergence.

Proposition 10. *For any $\gamma > 0$, there exists a game where the optimal penalized benchmark has at best an $n^{-\gamma}$ rate of convergence. That is,*

$$\mathrm{OP}_n(\pi^s) = \mathrm{CK}(\pi^s) - \Omega\left(n^{-\gamma}\right)$$

Furthermore, this game satisfies all of my regularity assumptions.

This result appears to be quite pessimistic, but it is not all that surprising. There happen to be games where the policymaker cannot guarantee the ideal outcome unless the agent has very precise distributional knowledge. In these cases, slow rates of convergence are inevitable. But that speaks to the fundamentals of the game itself, rather than to the quality of $\hat{P}_n$ as a strategy.

A more instructive question is to ask how $\hat{P}_n$'s rate of convergence compares to optimal penalized benchmark.

Definition 51. *Let $\gamma > 0$ be the largest real number such that the optimal penalized bench-*

mark converges to the common knowledge benchmark at the rate $n^{-\gamma}$. That is,

$$\mathrm{OP}_n(\pi^s) = \mathrm{CK}(\pi^s) - O(n^{-\gamma})$$

Loosely, the richness assumption requires that, for any distribution $\tilde{\pi}^s$ and policy p, the agent's best response r must be strictly better than her worst response r'.

Assumption 17. *For every distribution $\tilde{\pi}^s$ and policy p, there exist responses r, r' such that*

$$\mathrm{E}_{\tilde{\pi}^s}\left[\left(u^A(p,r,s) - u^A(p,r',s)\right)^2\right] \geq C^2$$

for some constant $C > 0$ that does not depend on $\tilde{\pi}^s, p, r, r'$.

Theorem 9 says that $\hat{P}_n$'s rate of convergence is approximately optimal. I characterize the rate of convergence in terms of its tuning parameters and then show how to optimize them.

Theorem 9. *The performance of $\hat{P}_n$ converges to the common knowledge benchmark at the rate:*

$$\mathrm{E}_{\pi^s}\left[\mathrm{WC}_n\left(\hat{P}_n, \epsilon, \delta_n, \pi^s\right)\right] = \mathrm{CK}(\pi^s) - O\left(n^{-\min(\gamma(1-\alpha),\beta)}\right)$$

I outline the proof in subsection 3.5.4.

The next corollary describes the rate of convergence when the tuning parameters are optimized. Generally, my estimator will only approximate the optimal rate of convergence $n^{-\gamma}$, with the difference reflecting the cost of sample privacy.

Corollary 4. *For any $\varepsilon > 0$, there exist parameter values α, β such that*

$$\mathrm{E}_{\pi^s}\left[\mathrm{WC}_n\left(\hat{P}_n, \epsilon, \delta_n, \pi^s\right)\right] = \mathrm{CK}(\pi^s) - O\left(n^{\frac{\gamma}{1+2\gamma}-\varepsilon}\right)$$

Proof. Set $\beta = \gamma/(1+2\gamma)$ and let α be slightly larger than 2β. □

For example, suppose that the optimal penalized benchmark converges at a typical $n^{-1/2}$ rate. Then $\hat{P}_n$'s rate of convergence can be set arbitrarily close to $n^{-1/4}$.

Remark 8. Theorem 9 can be understood as a possibility result. It says that $\hat{P}_n$ achieves a particular rate of convergence. But it does not claim that this rate of convergence is the best possible. It relies on explicit finite sample bounds, but does not evaluate whether they are tight enough to be useful in practice. It provides limited guidance on how to choose the tuning parameter α in finite samples, and no guidance on how to choose ϵ because it does not affect the rate of convergence.[8] Answering these questions effectively would likely require putting more structure on the underlying game.

3.5.3 Proof Outline of Theorem 8

The proof of theorem 8 relies on four lemmas, some of which are applications of known results. The key challenge is that $\widehat{\mathrm{WC}}_n(\cdot)$ is not a friendly object. It is a constrained minimization problem where the empirical distribution enters into both the objective and the constraint. And it has little structure, because I made few assumptions on the game between the policymaker and the agent.

The first step is to show that the objective $\widehat{\mathrm{WC}}_n(\cdot)$ falls within a distance t of its mean, with high probability. This will be used to establish that $\widehat{\mathrm{WC}}_n(\cdot)$ satisfies a sample privacy property, which immediately implies that $\hat{P}_n$ satisfies that property. Intuitively, if $\widehat{\mathrm{WC}}_n(\cdot)$ varies substantially with the sample, then a substantial amount of noise $\nu_n(\cdot)$ will be needed to ensure privacy. This first step will limit how much $\widehat{\mathrm{WC}}_n(\cdot)$ varies with the sample.

I rely on a concentration inequality due to McDiarmid (1989), applied to this setting. It

[8] All else equal, it is better if ϵ is small, but it does not need to be small in order for the results to hold. It only needs to be constant. The reason is that the empirical regret bound depends on ϵ through

$$e^{\epsilon} \cdot \overline{\mathcal{RC}}_n^A(p)$$

This term vanishes as n grows because the Rademacher complexity vanishes as n grows.

relies on a bounded differences property that I will substantiate later on in this proof.

Lemma 18 (McDiarmid 1989)**.** *Suppose that $\widehat{\mathrm{WC}}_n(p)$ has the bounded differences property, where changing the ith sample realization from s to s' will change its value by at most c. Formally,*

$$\widehat{\mathrm{WC}}_n(p \mid S_1, \ldots, S_{i-1}, s, S_{i+1}, S_n) - \widehat{\mathrm{WC}}_n(p \mid S_1, \ldots, S_{i-1}, s', S_{i+1}, S_n) \leq c \tag{3.15}$$

Then the following concentration inequality holds. For any $t > 0$,

$$\Pr\left[\widehat{\mathrm{WC}}_n(p) - \mathrm{E}\left[\widehat{\mathrm{WC}}_n(p)\right] \geq t\right] \leq \exp\left(-\frac{2t^2}{nc^2}\right)$$

where the probability and expectation are over the sampling process.

It follows from McDiarmid's inequality and the union bound that

$$\Pr\left[\exists p \in \mathcal{P}, \widehat{\mathrm{WC}}_n(p) - \mathrm{E}\left[\widehat{\mathrm{WC}}_n(p)\right] \geq t\right] \leq n_{\mathcal{P}} \exp\left(-\frac{2t^2}{nc^2}\right) \tag{3.16}$$

where $n_{\mathcal{P}}$ is the number of policies in $\mathcal{P}$.

Condition (3.16) is enough to guarantee sample privacy. This follows from the same reasoning that Manski and Tetenov (2007) use to establish differential privacy of the exponential mechanism.

Lemma 19. *Let c ensure the bounded differences property (3.15). For any $t > 0$, $\hat{P}_n$ satisfies (ϵ, δ)-sample privacy where*

$$\epsilon = 2tn^{\beta} \quad \text{and} \quad \delta = n_{\mathcal{P}} \exp\left(-\frac{2t^2}{nc^2}\right)$$

To establish the bounded differences property, I rely on the robustness lemma of Camara et al. (2020). Keep in mind that changing a sample realization S_i affects not only the objective in $\widehat{\mathrm{WC}}_n(p)$, but also the constraint. Since the game between the policymaker and agent is largely arbitrary, the impact of tightening or relaxing the constraint can be difficult to capture. However, it is possible to derive a loose bound that does not depend at all on the underlying structure of the game. It only relies on the fact that the worst case is being evaluated with respect to mixed strategies.

To state the robustness lemma, I need additional notation. Consider the policymaker's worst-case utility when the agent's regret is bounded by a constant $b \geq 0$, i.e,

$$\begin{aligned} \mathrm{WC}(p, b, \pi^s) = &\min_{\pi^r} \mathrm{E}_{\pi^s,\pi^r}\left[u^P\left(p, r, s\right)\right] \\ \text{s.t.} \quad &\max_{r'} \mathrm{E}_{\pi^s}\left[u^A\left(p, r', s\right)\right] - \mathrm{E}_{\pi^s,\pi^r}\left[u^A\left(p, r, s\right)\right] \leq b \end{aligned} \tag{3.17}$$

Lemma 20 (Camara et al. 2020)**.** *The worst-case utility* $\mathrm{WC}(p, b, \pi^s)$ *decreases smoothly in the bound* b*. That is, for any constants* $b' > b > 0$*,*

$$\mathrm{WC}(p, b', \pi^s) \geq \mathrm{WC}(p, b, \pi^s) - \Delta^A(p)\left(\frac{b'-b}{b}\right)$$

I use this result to establish and quantify the bounded differences property, as follows.

Lemma 21. *The random variable* $\widehat{\mathrm{WC}}_n(p)$ *satisfies the bounded differences property as long as*

$$c \geq \Delta^A(p)\left(\frac{2\Delta^P(p)\cdot n^{-1}}{(4e^{\epsilon}+4)\cdot \overline{\mathcal{RC}}_n^A(p) + \delta + \mathrm{BFR}_n}\right) + \Delta^P(p)\cdot n^{-1}$$

Now, I can define the parameter c. Recall from lemma 19 that δ depends on c. To avoid a circular definition, c should not depend on δ. Moreover, c should not depend on the particular policy p, which depends on δ indirectly through the strategy $\hat{\sigma}_n^P$. By lemma 21 and simple inequalities, it is sufficient to set

$$c \geq \max_p \left(\Delta^A(p) \left(\frac{2\Delta^P(p) \cdot n^{-1}}{(4e^\epsilon + 4) \cdot \overline{\mathcal{RC}}_n^A(p) + \mathrm{BFR}_n} \right) + \Delta^P(p) \cdot n^{-1} \right)$$

Technically, I could define c as the right-hand side on this expression and be done. But I will use a slightly looser bound for the sake of interpretability. Recall the constant K defined in equation (3.12). It is sufficient to set

$$c := Kn^{-\frac{1+\alpha}{2}}$$

The last step of the proof is to derive δ_n. Recall from lemma 19 that $\hat{P}_n$ satisfies (ϵ, δ)-sample privacy where, after plugging in the value of c,

$$\epsilon = 2tn^\beta \quad \text{and} \quad \delta = n_{\mathcal{P}} \exp\left(-\frac{2t^2}{K^2} \cdot n^\alpha \right)$$

Since this holds for any value $t > 0$, I can invert the equation $\epsilon = 2tn^\beta$ to find the value $t = \epsilon/(2n^\beta)$ that keeps the parameter ϵ constant. Plugging this value of t into δ gives me δ_n, as defined in equation (3.11). This completes the proof of theorem 8.

3.5.4 Proof Outline of Theorem 9

The proof of theorem 9 has to contend with the same high-level challenge as the proof of theorem 8: namely, $\widehat{\mathrm{WC}}_n(\cdot)$ is not a friendly object. First, I prove proposition 9, which establishes convergence. Then I build on this argument to prove theorem 9.

To prove proposition 9, I need some way to characterize $\hat{P}_n$'s performance. I do this by substituting its actual performance, which is hard to describe, with its estimated performance $\widehat{\mathrm{WC}}_n\left(\hat{P}_n\right)$. The estimated performance is easier to work with because that is what $\hat{P}_n$ is maximizing. The next lemma shows that this substitution is justified.

Lemma 22. *The estimated performance $\widehat{\mathrm{WC}}_n\left(\hat{P}_n\right)$ determines a lower bound on the actual performance of $\hat{P}_n$. More precisely,*

$$\mathrm{E}_{\pi^s}\left[\mathrm{WC}_n\left(\hat{P}_n, \epsilon, \delta_n, \pi^s\right)\right] \geq \widehat{\mathrm{WC}}_n\left(\hat{P}_n\right) - n_{\mathcal{P}}\sqrt{3}\cdot n^{-\beta} - 2\mathcal{RC}_n^P(\pi^s) - n_{\mathcal{P}}\exp(-n^{\alpha})\cdot \max_p \Delta^P(p)$$

This lower bound reflects three observations. First, by construction, $\widehat{\mathrm{WC}}_n(p)$ erred on the side of being too conservative with respect to the agent's empirical regret bound. All else equal, this would mean that $\widehat{\mathrm{WC}}_n(p)$ should lower bound $\mathrm{WC}_n(p)$ with high probability. Second, $\widehat{\mathrm{WC}}_n(p)$ involves sampling noise. This can lead to generalization error, where the policymaker expects a policy to perform better than it does. I can bound the generalization error using the policymaker's Rademacher complexity. Finally, $\widehat{\mathrm{WC}}_n(p)$ involves privacy-preserving noise. By construction, this is vanishing as the sample size grows.

In light of this lemma, in order to prove proposition 9 it is enough to show that

$$\lim_{n\to\infty} \mathrm{E}_{\pi^s}\left[\widehat{\mathrm{WC}}_n\left(\hat{P}_n\right)\right] = \mathrm{CK}(\pi^s)$$

This is true because the privacy-preserving noise is vanishing as $n \to \infty$, the policymaker's empirical utility is converging in probability to his expected utility according to the true distribution, and the regret bound is vanishing as $n \to \infty$. In particular, Berge's maximum theorem implies that the policymaker's worst-case utility is continuous with respect to the regret bound.

Next, I turn to theorem 9.

I just showed that the empirical regret bound in the definition of $\widehat{\mathrm{WC}}_n(p)$ (3.13) is vanishing as the sample size grows. Similarly, the regret bound in the definition of $\mathrm{WC}_n(p, 0, 0, \pi^s)$ is vanishing as the sample size grows. For a given sample size n, these bounds will take on different values. In general, they will shrink at different rates. And one bound involves empirical regret while the other involves regret with respect to the true distribution.

Despite these differences, we need to compare $\widehat{\mathrm{WC}}_n(p)$ with $\mathrm{WC}_n(p, 0, 0, \pi^s)$ to prove a result. As in theorem 8, these objects are too abstract to characterize directly. However, I can compare one abstract object with another: $\widehat{\mathrm{WC}}_n(p)$ for sample size n with $\mathrm{WC}_m(p, 0, 0, \pi^s)$ for a smaller sample size m. The idea is that if m is sufficiently small compared to n, then the regret bound for $\mathrm{WC}_m(p, 0, 0, \pi^s)$ is more conservative than the empirical regret bound for $\widehat{\mathrm{WC}}_n(p)$, even though the latter would be much more conservative if $m = n$. After accounting for some other differences, I can show that $\widehat{\mathrm{WC}}_n(p)$ is comparable to $\mathrm{WC}_m(p, 0, 0, \pi^s)$. Since I know the rate of convergence for the latter in m, I can determine the rate of convergence for the former in n.

The next lemma formalizes this argument.

Lemma 23. *The estimated performance $\widehat{\mathrm{WC}}_n\left(\hat{P}_n\right)$ with sample size n is comparable to the strategically-regularized benchmark with sample size*

$$m = \Theta(n^{1-\alpha})$$

evaluated with respect to the true distribution. More precisely,

$$\mathrm{E}_{\pi^s}\left[\widehat{\mathrm{WC}}_n\left(\hat{P}_n\right)\right] \geq \mathrm{SR}_m(\pi^s) - n_{\mathcal{P}}\sqrt{3} \cdot n^{-\beta} - 2\mathcal{RC}_n^P(\pi^s) - n_{\mathcal{P}} \exp(-n^\alpha) \cdot \max_p \Delta^P(p)$$

This result relies critically on another lemma that may be of independent interest. It is a lower bound on Rademacher complexity that makes use of Khintchine's inequality. This is the only part of the proof that relies on assumption 17.

Lemma 24. *Recall the constant C in assumption 17. For any policy p and distribution π^s,* *the Rademacher complexity is bounded below by*

$$\mathcal{RC}_n^A(p, \pi^s) \geq \frac{C}{2\sqrt{2n}} \tag{3.18}$$

Combining lemma 23 with lemma 22, we have

$$\begin{aligned}
\mathrm{E}_{\pi^s}\left[\mathrm{WC}_n\left(\hat{P}_n, \epsilon, \delta_n, \pi^s\right)\right] &\geq \mathrm{SR}_m(\pi^s) - O\left(n^{-\beta}\right) - O\left(n^{-1/2}\right) \\
&= \mathrm{CK}(\pi^s) - O\left(m^{-\gamma}\right) - O\left(n^{-\beta}\right) - O\left(n^{-1/2}\right) \\
&= \mathrm{CK}(\pi^s) - O\left(m^{-\gamma}\right) - O\left(n^{-\beta}\right) \\
&= \mathrm{CK}(\pi^s) - O\left(n^{-\gamma(1-\alpha)}\right) - O\left(n^{-\beta}\right) \\
&= \mathrm{CK}(\pi^s) - O\left(n^{-\min(\gamma(1-\alpha),\beta)}\right)
\end{aligned}$$

The second line follows from assumption 51. The third line removes lower-order terms. The fourth line plugs in the value $m = \Theta\left(n^{2(1-\alpha)}\right)$. The last line takes the maximum over the two rightmost terms, and completes the proof of theorem 9.

3.6 Illustrative Examples

I argue that penalization can lead to new insights in four illustrative examples: vaccine distribution, prescription drug approval, performance-based pay, and product bundling. These examples are not intended to be as general or realistic as possible. Instead, they are meant to convey a core insight that motivates the use of penalized policies in similar applications.

3.6.1 Vaccine Distribution

I present a model of vaccine distribution where penalization can motivate a common practice: insisting on statistically-significant clinical trial results before delivering medical treatments.

Decision-making based on statistical significance is hard to justify (e.g. Wasserstein and Lazar 2016) and conflicts with the recommendations of the treatment choice literature (e.g. Manski 2019). Likewise, existing solution concepts (e.g. common knowledge, maximin, and minimax regret) do not support this practice. But penalization does. The intuition is that, when vaccine quality is not common knowledge, skepticism among the population can undermine a vaccine rollout. Since vaccine distribution involves fixed costs, it may be better to wait until clinical trial results are sufficiently persuasive before attempting to vaccinate the population.

Model. Consider a town of m agents that is afflicted by a disease. A new vaccine is being developed to treat this disease. However, in order to distribute this vaccine, the policymaker must invest in a treatment center at a fixed cost c. Given the treatment center, the policymaker can treat each agent at zero marginal cost. Therefore, the policymaker must decide whether to provide treatments ($T = 1$) at cost c, or not to treat ($T = 0$).

An agent's outcome Y depends on both whether she is treated, and whether she complies with the treatment. Let $C = 1$ indicate compliance, and $C = 0$ indicate noncompliance. Let Y_1 denote her outcome conditional on being successfully treated and let Y_0 denote her outcome otherwise. For simplicity, I assume that the agent has no private information about her outcome, so that compliance C is independent of the outcomes Y_0 and Y_1.

It remains to specify payoffs and the dataset. The agent tries to maximize her expected outcome:

$$\mathrm{E}[Y \mid C, T] = \mathrm{E}[Y_0 + C \cdot T \cdot (Y_1 - Y_0) \mid C, T] = \omega_0 + \omega_1 \cdot C \cdot T$$

The parameter ω_1 is called the average treatment effect (ATE). The policymaker wants to maximize the expected welfare minus costs, i.e.

$$m\mathrm{E}[Y \mid C, T] - c$$

Both the policymaker and the agents have access to clinical trial data where compliance is guaranteed. This includes n treated outcomes Y_1^i and n untreated outcomes Y_0^i. The key summary statistic is the sample average treatment effect, i.e.

$$\hat{\omega}_1 = \frac{1}{n}\sum_{i=1}^{n} Y_1^i - \frac{1}{n}\sum_{i=1}^{n} Y_0^i$$

This is a sufficient statistic for both the policymaker and the agent to optimize.

Existing Solution Concepts. To establish that penalization leads to a new insight, I first need to evaluate what existing solution concepts recommend.

Claim 1. *The optimal common knowledge policy treats iff the ATE exceeds the per-capita cost, i.e.*

$$\omega_1 \geq \frac{c}{m}$$

Proof. The agent complies with the treatment iff $\omega_1 \geq 0$ and, given compliance, the policymaker prefers to treat whenever the ATE exceeds the per-capita costs. □

Claim 2. *The maximin policy never treats.*

Proof. If the policymaker treats, he incurs a cost $c > 0$ and the agents may not comply anyways, leading to a negative payoff in the worst case. By not treating, he guarantees a payoff of zero. □

Claim 3. *The minimax regret policy never treats iff the ATE exceeds twice the per-capita cost, i.e.*

$$\omega_1 \geq \frac{2c}{m}$$

Proof. If the policymaker does not treat, the maximum regret $m\omega_1 - c$ occurs when the agents would have complied with treatment. If he does treat the population, the maximum regret c occurs when the agents decide not to comply. The policymaker treats iff $m\omega_1 - c \leq c$, or $\omega_1 \leq 2c/m$. □

None of these solutions justify statistically-significant clinical trial results as a precondition for treatment. Furthermore, this conclusion does not depend on the assumption that the policymaker knows the true distribution. The literature on treatment choice has studied problems like these, under the assumption that compliance is independent of sample size. It recommends statistical decision rules like empirical welfare maximization (e.g. Manski 2004b, Stoye 2009, Kitagawa and Tetenov 2018, Mbakop and Tabord-Meehan 2021). In this case, the empirical welfare maximizer under common knowledge would treat iff the sample ATE exceeds cost per-capita, i.e.

$$\hat{\omega}_1 \geq \frac{c}{m} \tag{3.19}$$

This rule follows the preponderence of the evidence but does not insist on statistical significance. In particular, the threshold for treatment does not vary with the sample size n.

Optimal Penalized Policy. Penalization will motivate a form of statistical significance because, in this model, a treatment is only successful if agents agree to comply with said treatment. When clinical trial results are not statististically significant, agents may be sufficiently uncertain that they decide not to comply, even if compliance is optimal under the true distribution.

To substantiate this intuition, I solve for the optimal penalized policy. To define a regret bound, I assume that the ATE is bounded, i.e. $\omega_1 \in [\underline{\omega}_1, \bar{\omega}_1]$ where $\bar{\omega}_1 > c/m > \underline{\omega}_1$, and I define

$$\overline{\mathcal{RC}}_n^A(T=0) := 0 \quad \text{and} \quad \overline{\mathcal{RC}}_n^A(T=1) := \frac{(\bar{\omega}_1 - \underline{\omega}_1)\sqrt{2\ln 2}}{\sqrt{n}}$$

This is a valid upper bound on the Rademacher complexity (see Massart's lemma in remark

7).

Claim 4. *The optimal penalized policy treats iff*

$$\omega_1 \geq \frac{c}{m} + O\left(n^{-1/2}\right)$$

Proof. The policymaker never treats if $\omega_1 < c/m$, since his payoff would be negative regardless of whether the agents comply. Suppose $\omega \geq c/m$. Given treatment, the maximum probability q of non-compliance that satisfies the regret bound solves

$$q \cdot \omega_1 + (1-q) \cdot 0 = \frac{(\bar{\omega}_1 - \underline{\omega}_1)\sqrt{2\ln 2}}{\sqrt{n}}$$

where the left-hand side is the agent's regret and the right-hand side is the regret bound. This leaves

$$q = \frac{(\bar{\omega}_1 - \underline{\omega}_1)\sqrt{2\ln 2}}{\omega_1\sqrt{n}}$$

The policymaker's worst-case payoff from treatment is $(1-q)m \cdot \omega_1 - c$. Plugging in the value of q, he prefers to treat iff

$$m\omega_1 \left(1 - \frac{(\bar{\omega}_1 - \underline{\omega}_1)\sqrt{2\ln 2}}{\omega_1\sqrt{n}}\right) - c \geq 0$$

This simplifies to

$$\omega_1 \geq \frac{c}{m} + \frac{(\bar{\omega}_1 - \underline{\omega}_1)\sqrt{2\ln 2}}{\sqrt{n}}$$

□

The $O(n^{-1/2})$ term is analogous to a critical value. In that sense, the optimal penalized policy insists that the difference between the ATE and the per-capita cost exceeds that critical value. Of course, this requires the policymaker to know the true distribution. In general, he would have to evaluate the sample ATE rather than the true ATE, in line with the data-driven penalized policy developed in section 3.5. But the essential intuition survives.

3.6.2 Prescription Drug Approval

In a model of prescription drug approval by a regulator, I show that strategic regularization restricts doctors' ability to prescribe drugs that have not been proven effective in clinical trials. The threshold for approval increases as more drugs are approved. This is similar to stepwise methods for multiple hypothesis testing (e.g. Holm 1979, Romano and Wolf 2005).

In contrast, in models with a common prior or rational expectations, the optimal policy is to approve all drugs. Essentially, this delegates the decision to doctors, who are better informed than the regulator. In my model, however, doctors may prescribe ineffective drugs if the clinical trial returns a "false positive" where an ineffective drug appears to be effective by random chance. Limiting the number of drugs approved can reduce the risk of false positives, and provide better welfare guarantees.

Model. A population of patients is afflicted by a disease. There are m treatments available, as well as a placebo. As in the previous example (section 3.6.1), let $\omega_j \in [\underline{\omega}, \bar{\omega}]$ be the average treatment effect (ATE) of treatment j. The placebo's treatment effect is normalized to zero. In addition, patients incur a private cost $c_j \in [0, \bar{c}]$ from treatment j, where the maximum cost $\bar{c} > \bar{\omega}$ exceeds the maximum treatment effect. For example, this could represent the patient's copay for a prescription drug. The patients are nonstrategic and accept whatever treatment is offered.

There is a regulator (policymaker) who approves treatments and a doctor (agent) who prescribes them. Formally, the regulator specifies a set $\mathcal{A} \subseteq \{1, \dots, m\}$ of approved treatments. Then the doctor either prescribes a treatment $j \in \mathcal{A}$ to a given patient, or prescribes the placebo $j = \emptyset$. Formally, the doctors response r is a choice function, mapping sets $\mathcal{A}$ to treatments $j \in \mathcal{A} \cup \{\emptyset\}$.

Both participants have access to clinical trial data with sample size n, where $\hat{\omega}_j$ is the sample ATE of treatment j. Both participants want to maximize the patient's expected outcome minus costs, i.e. $\omega_j - c_j$ for the chosen treatment j. But the doctor has an informational advantage. She knows patient costs c_j at the time of treatment choice, but the regulator does not.

There are many reasonable ways to accommodate the uncertainty in the costs c_j. I assume that the policymaker evaluates his worst-case regret with respect to these costs (not necessarily with respect to the agent's beliefs). That is,

$$u^P(\mathcal{A}, r, \vec{\omega}) = \max_{\vec{c}} \left(\max_{\mathcal{A}'} \left(\omega_{r(\mathcal{A}')} - c_{r(\mathcal{A}')} \right) - \left(\omega_{r(\mathcal{A})} - c_{r(\mathcal{A})} \right) \right)$$

Existing Solution Concepts. I first establish what existing solution concepts recommend, before moving onto the optimal penalized policy.

Claim 5. *The optimal common knowledge policy approves all treatments.*

Proof. If the ATEs are common knowledge, the doctor will choose the treatment j that maximizes outcome minus costs, i.e. $\omega_j - c_j$. The regulator prefers this to any other policy. If he excludes a treatment j, it is always possible that treatment j was the only treatment with low costs, e.g. where $c_j = 0$ and $c_i = \bar{c}$ for $i \neq j$. In that case, the regulator may regret excluding treatment j. □

These same conclusion holds if the regulator does not know the true disribution, as long as the doctor does. In that case, the regulator has even more incentive to defer to the doctor.

Claim 6. *The maximin policy approves no treatments.*

Proof. Suppose the approved set $\mathcal{A}$ is nonempty. The worst case outcome occurs when all treatments $j \in \mathcal{A}$ share the maximum cost, $c_j = \bar{c}$, and the doctor chooses the worst treatment $j \in \mathcal{A}$ because she believes it to be highly effective. The regulator's utility is

$$\min_{j \in \mathcal{A}} \omega_j - \bar{c}$$

This is always negative, since $\omega_j < \bar{c}$. It is better to not approve any treatments, since this at least guarantees non-negative utility for the regulator. □

Claim 7. *The minimax regret policy approves no treatments.*

Proof. Suppose that all treatments have zero cost and the doctor believes they have zero effectiveness. Let j^* be the treatment with the highest ATE.

Suppose the regulator deviates from approvals $\mathcal{A}$ to approvals $\mathcal{A}'$ that includes j^*. The doctor breaks her indifference in favor of the placebo when presented with $\mathcal{A}$, but breaks her indifference in favor of treatment j^* when presented with $\mathcal{A}'$. The regulator's regret is given by $\omega_{j^*} = \max_j \omega_j$. Therefore, his worst-case regret is bounded below by ω_{j^*}. In contrast, approving no treatments guarantees that the regulator's regret is bounded above by ω_{j^*}. Therefore, approving no treatments minimizes worst-case regret. □

Optimal Penalized Policy. Penalization will motivate policies that are less extreme and more realistic, where the regulator approves some treatments but not all. The intuition is that the doctor faces a multiple testing problem. The more treatments are approved, the greater the chances of a false positive – a treatment that appears to be effective but is not. Approving too many treatments can cause doctors to prescribe false positives with high probability.

To substantiate this intuition, I solve for the optimal penalized policy, where

$$\overline{\mathcal{RC}}_n^A(\mathcal{A}) := \frac{(\bar{\omega} - \underline{\omega})\sqrt{2\ln\left(|\mathcal{A}| + 1\right)}}{\sqrt{n}}$$

As in the previous subsection, this bound follows from Massart's lemma (remark 7). Note that the regret bound is increasing in the $|\mathcal{A}|$ of approved treatments.[9]

The optimal penalized policy begins by ordering treatments according to their ATE. Let $\omega_{(k)}$ denote the k^{th} highest ATE.

Claim 8. *The optimal penalized policy approves the k^{th} best treatment iff*

$$\omega_{(k)} \geq \frac{(\bar{\omega} - \underline{\omega})\sqrt{2\ln(k+1)}}{\sqrt{n}}$$

Proof. To derive the optimal penalized policy, I refer to two terms that reflect the regulator's regret with respect from the unknown cost. First, the regulator's regret from not approving treatment j will be ω_j in the worst case. This occurs when treatment j has zero cost and all other treatments have maximum cost. It follows that worst-case regret is at least $\max_{j \notin \mathcal{A}} \omega_j$.

Second, the regulator's regret from approving a treatment j will be $\overline{\mathcal{RC}}_n^A(\mathcal{A})$ in the worst case. Suppose $\mathcal{A}$ is nonempty. This level of regret can be achieved by setting the cost of all treatments that the doctor does not choose to $\bar{c}$, and the cost of the treatment j that the doctor does choose to $c_j = \omega_j + \overline{\mathcal{RC}}_n^A(\mathcal{A})$. The doctor will be indifferent between treatment j and the placebo. At the same time, the $\overline{\mathcal{RC}}_n^A(\mathcal{A})$ is an upper bound on the regulator's regret, because that is an upper bound on the agent's regret and both participants care about welfare.

The optimal penalized policy minimizes the larger of these two regret terms. It begins

[9] The agent's response space is technically larger than $|\mathcal{A}| + 1$ since it consists of maps from $\mathcal{A}$ to either a treatment $j \in \mathcal{A}$ or the placebo. But for a fixed policy $\mathcal{A}$, it is without loss to restrict attention to $|\mathcal{A}| + 1$ unique actions.

by ordering treatments according to their ATE. The regulator approves the best treatment iff

$$\omega_{(1)} \geq \frac{(\bar{\omega} - \underline{\omega})\sqrt{2}}{\sqrt{n}}$$

Similarly, the regulator approves the kth best treatment iff

$$\omega_{(k)} \geq \frac{(\bar{\omega} - \underline{\omega})\sqrt{2\ln(k+1)}}{\sqrt{n}}$$

□

As in section 3.6.1, treatments are approved only if they reach a critical value. More treatments are approved as the sample size n grows, since the critical value decreases according to $O(n^{-1/2})$. Moreover, the critical value is increasing in the number of treatments already approved. As I mentioned earlier in this section, this is similar to stepwise methods for multiple hypothesis testing. That is not surprising, because the motivation for limiting the number of approved drugs is precisely to ensure that doctors are not misled by false positives.

Role of Sample Privacy. In Subsection 3.3.3, I claimed that bounds from statistical learning theory on the agent's regret may not be valid if the policymaker is also learning from data. This model of prescription drug approval makes that clear. I elaborate below.

The doctor's regret from expected utility maximization can vary significantly based on how the regulator uses the sample. Suppose the regulator approves a single treatment j independently of the sample. The doctor's regret will be on the order of $O\left(n^{-1/2}\right)$. That is the bound implied by Massart's lemma (remark 7). Alternatively, suppose that regulator only approves the empirical utility maximizer, i.e. $\arg\max_j \hat{\omega}_j$. In that case, the doctor's choice is the same as if all treatments had been approved, and her regret will be on the order of $O\left(n^{-1/2}\log m\right)$. Again, that is the bound implied by Massart's lemma (remark 7).

The dependence on m reflects a multiple testing problem that arises even though the agent only has two choices: the approved treatment and the placebo. It arises because the choice presented to her is correlated with the sample.

Sample privacy means that, with some probability, the regulator is not selecting the k treatments whose sample ATE is highest. More generally, it is well-known that using data to determine the hypotheses one wants to test will threaten the validity of those tests. There are several ways to get around this. For example, the regulator could use his prior beliefs to fix the k most promising treatments that he wants to evaluate and approve. In contrast, the approach that I develop in Section 3.5.1 is data-driven, and uses sample privacy to control how aggressively the data is used.

3.6.3 Performance Pay

Third, I consider a model of performance pay. An employer incentivizes an employee to exert costly effort by paying wages contingent on observed performance. Here, penalization caps and flattens the wage schedule. Wages are zero until performance reaches some threshold. At that point, they jump and remain flat. The employee receives moderate pay, but with high probability.

In contrast, the common knowledge solution recommends very high pay, but only if the employee attains the best possible performance is obtained. In the penalized model, this does not work well. If historical data is limited, it may not be obvious to the employee that it is worth investing effort for a small chance of receiving the bonus. Flatter contracts may be less potent, but they provide clearer incentives.

This insight contributes to the recent literature in robust contract design (e.g. Carroll 2015; Carroll and Meng 2016a,b; Dütting et al. 2019) that, in turn, builds on previous attempts to explain the ubiquity of simple contracts (e.g. Holmstrom and Milgrom 1987, 1991). In particular, Holmstrom and Milgrom (1991) motivate fixed wages when the principal is unable to measure some dimensions of the agent's performance. The optimal penalized contract developed here suggests that learnability may be another motivation.[10]

Model. I define a simple principal-agent problem, along the lines of Sappington (1983). A principal wants to incentivize the agent to take desirable actions. The timing of the game is as follows: first, the principal commits to a wage schedule or contract w; second, the agent takes a hidden action a; third, the principal observes an outcome x and pays the agent a wage $w(x)$.

Let $\mathcal{X} = \{x_1, \ldots, x_m\} \subseteq \mathbb{R}$ be a finite outcome space, in increasing order. Outcomes are determined, stochastically, by the agent's hidden action. Let $A = \{0, 1\}$ be a binary action space, where $a = 0$ indicates no effort and $a = 1$ indicates effort. The agent incurs a cost of effort $c > 0$. Let π_0^x be the outcome distribution conditional on no effort. Let π_1^x be the outcome distribution conditional on effort. I assume both distributions have full support. Let $X^a \sim \pi_a^x$ denote the realized outcome when the agent takes action a.

I restrict attention to conditional distributions that satisfy a monotone likelihood ratio property.

Assumption 18. *The likelihood ratio $\ell(x) = \pi_1^x(x)/\pi_0^x(x)$ is weakly increasing in x.*

Before the agent acts, the principal commits to a wage function $w : \mathcal{X} \to \mathbb{R}_+$.[11] By definition, the wage function satisfies *limited liability* (i.e. $w(x) \geq 0$ for all $x \in \mathcal{X}$). Both the agent and the principal are risk-neutral. Given action a and outcome x, the agent's utility is $w(x) - c \cdot a$ while the principal's utility is $x - w(x)$.

[10] Relatedly, Valenzuela-Stookey (2020) proposes an axiomatic model where an agent's evaluation of lotteries may be imprecise. He motivates his axioms in a frequentist model of learning. He applies his representation towards a general principal-agent problem where, if the agent is suitably cautious, optimal contracts correspond to step functions. In the principal-agent model that I consider, the optimal contract is a step function even in the standard case where the agent knows the true distribution.

[11] The reader is welcome to treat wages as dollar-valued in order to ensure that the policy space is finite, as specified in assumption 15. This does not meaningfully affect the analysis.

Let $X_1^a, \ldots, Y_n^a \sim \pi_a^x$ be i.i.d. samples of outcomes for each action $a \in \{0, 1\}$. For example, this could be data that a manager collects through personal experimentation, or through costly monitoring of past employees, and shares with his current employees.

Existing Solution Concepts. I first establish what existing solution concepts recommend, before moving onto the optimal penalized policy. These results will refer to a *zero contract* w, which sets, for all $x_i \in \mathcal{X}$, $w(x_i) = 0$.

The optimal common knowledge contract makes a recommendation that seems extreme: pay the agent if and only if the realized performance is the highest possible.

Claim 9. *The optimal common knowledge contract is either the zero contract, or sets*

$$\forall x_i \in \mathcal{X} \quad w(x_i) = \begin{cases} \frac{c}{\pi_1^x(x_i) - \pi_0^x(x_i)} & i = m \\ 0 & i < m \end{cases}$$

For the the prior-free solution concepts, on the other hand, there is often no contract that guarantees effort by the agent.

Claim 10. *The maximin contract is the zero contract.*

Proof. Suppose the agent beliefs that the outcome distribution does not depend on effort, i.e. $\pi_0^x = x_1^x$. Then no contract can incentivize her to put in effort. Since the outcome distribution has full support, any non-zero contract would make positive payments for no change in effort. □

Claim 11. *Let π_1^x be the uniform distribution. If the number of outcomes m is sufficiently large, then the minimax regret contract is the zero contract.*

Proof. Let w be the minimax regret contract. Suppose the agent believes that putting in effort guarantees outcome x_i and not putting in effort guarantees outcome $x_j < x_i$ (satisfying assumption 18). Let $q_i = \min_{i'} \Pr_{\pi_1^x}[x_{i'}]$ be the probability of outcome x_i The principal can incentivize the agent by paying her cost c conditional on outcome x_i. This costs him $q_i c$ in expectation, and he benefits from the surplus generated by the agent's effort.

If the contract w does not incentivize effort under these beliefs, then regret is the surplus $\mathrm{E}_{\pi^x}[X^1 - X^0]$ plus the expected wages $\mathrm{E}_{\pi^x}[w(X^1)]$ minus $q_i c$. I claim that the contract w cannot incentivize effort when m is large. In order to guarantee effort across all choices of x_i, the contract would need to ensure that $w(x_i) = w(x_j) + c$ for all $i, j < i$. In particular, set $j = i - 1$ and invoke limited liability to see that $w(x_i) = c \cdot i$. As $m \to \infty$, the wages paid at any $i > m/2$ grows to infinity as well. This contradicts the optimality of the contract w.

It follows that, for large m, regret is the surplus $\mathrm{E}_{\pi^x}[X^1 - X^0]$ plus the expected wages $\mathrm{E}_{\pi^x}[w(X^1)]$ minus $q_i c$. For the uniform distribution, $q_i = q$ does not depend on the contract, To minimize regret it suffices to minimize the expected wages, by letting w be the zero contract. □

Optimal Penalized Contract. The optimal penalized policy caps and flattens the optimal common knowledge contract. This makes the incentives for effort less potent, but it also makes them clearer. Given limited data, the agent is still able to determine that effort is optimal.

To substantiate this intuition, I solve for the optimal penalized policy, where

$$\overline{\mathcal{RC}}_n^A(w) := \left(\max_i w(x_i)\right) \cdot \sqrt{2\ln 2}\sqrt{n}$$

As in the previous subsections, this bound follows from Massart's lemma (remark 7). Note that the regret bound is increasing in the *maximum wage*, i.e. $\max_i w(x_i)$.

Claim 12. *The optimal penalized contract is a threshold contract. More precisely, it optimizes the following contract across all maximum wages $\bar{w}$ and probabilities q of effort.*

Define

$$\alpha_j = \frac{c + \bar{w}\left(\frac{4}{q} \cdot \sqrt{\frac{2\log 2}{n}} - \sum_{i=j+1}^{m} \left(\pi_1^x(x_i) - \pi_0^x(x_i)\right)\right)}{\pi_1^x(x_j) - \pi_0^x(x_j)}$$

Let k be the largest j such that $\ell(x_j) > 1$ and $\alpha_j \leq \bar{w}$. If no such integer exists, set $w(x) = 0$ for all outcomes $x \in \mathcal{X}$. Otherwise, set $w(x_i)$ to equal $\bar{w}$ if $i > k$, α_k if $i = k$, and 0 if $i < k$.

The optimal penalized contract identifies the maximum wage as an hindrance to learnability. When wages conditional on high performance are very large, small changes in the perceived probability of high performance conditional on effort can have a large impact on the perceived utility of effort. For that reason, the principal would have to add a premium to the agent's wages in order to ensure that the agent puts in effort. This makes the contract actuarially less appealing. On the other hand, if the maximum wage is capped, then the perceived utility of effort is less sensitive to beliefs. The premium needed to incentivize effort is smaller. The optimal penalized policy balances the benefits of higher maximum wages with the costs identified here.

3.6.4 Product Bundling

Finally, I consider a model of product bundling. A firm has several products for sale and wants to sell them in a way that maximizes expected profit. When the sample size is small, the penalized policy only offers large bundles of products, rather than selling them separately. This contrasts with prior work that recommends selling items separately (e.g. Carroll 2017).

In my model, the reason for bundling is that consumers learn about their value for the product through reviews. If there are many products, but few reviews per product, consumers can be confident in the value of a large bundle while being uncertain about the value of any given product. In that case, all else equal, it is easier to convince consumers to buy the bundle.

Model. There are n goods, one seller, and one buyer. The buyer has preferences over bundles $x \in \{0,1\}^n$ of goods. The seller can offer a menu M consisting of bundles x at prices p, i.e. $(x,p) \in M$. The seller cannot prevent the buyer from buying more than one good. If $(x,p),(x',p') \in M$ then for bundle

$$x'' = (\max\{x_1, x'_1\}, \ldots, \max\{x_m, x'_m\})$$

there exists a price $p'' \leq p + p'$ such that $(x'', p'') \in M$. The buyer's value for each of the goods is described by a vector $v \in [0, \bar{v}]^n$. Her utility from choosing a bundle $(x,p) \in M$ is

$$u^A(M,(x,p),v) = v \cdot x - p$$

The seller cares about profits and, for simplicity, has zero costs of production. His utility is p.

To relax the assumption that the buyer knows her value from product j, I assume that she knows her ranking relative to other agents. More precisely, she knows the quantile q_j of her value v_j in the marginal distribution π^v_j. Given menu M and a chosen bundle $(x,p) \in M$, the buyer at quantile $q = (q_1, \ldots, q_m)$ receives a payoff of

$$U^A\left(q, M, (x,p), \pi^v\right) := \sum_{j=1}^{m} x_j \cdot \inf\left\{u \in \mathbb{R} \mid q \leq \Pr_{\pi^v}[v_j \leq u]\right\} - p$$

The quantile is her private type. If she knows π^v_j then she knows the value she derives from product j. If she is uncertain about π^v_j, then she faces uncertainty about her value. For example, a Netflix user might understand that she is particularly predisposed towards science fiction, but not know whether the quality of a particular science fiction movie.

The following assumption will be convenient later on.

Assumption 19. *The marginal distributions π^v_j have well-defined density functions f_j.*

There exists a constant $K > 0$ where $f_i(v_j) \geq K$ is bounded below by that constant on its support $[0, \bar{v}]$.

I assume that there is a common dataset consisting of product reviews. Each review of product j is a single observation of the value v_j of the product to the user that reviewed it. Each product has n reviews. This data identifies the marginal distributions π^v_j, but not the joint distribution of the value profiles v. This is because I do not observe a n users observing m products, but rather nm users observing 1 product each. Arguably, this better reflects the kind of review data that would be available in practice.

Since only the marginal distributions are identified by the available data, I do not assume that the seller knows the joint distribution. Instead, I follow Carroll (2017) in assuming that the seller knows (at most) the marginal distributions. He evaluates his profits with respect to the worst-case joint distribution that is consistent with the true marginal distributions.

Existing Solution Concepts. I first establish what existing solution concepts recommend, before moving onto the optimal penalized policy. I rely on Carroll's (2017) characterization of the optimal menu when the marginal distributions are common knowledge.

Claim 13 (Carroll 2017)**.** *Selling each product separately is an optimal common knowledge menu. That is, each product j can be purchased individually at a price p_j. The price of a bundle x is the sum of prices $\sum_{j=1}^{m} x_j p_j$ of the constituent goods.*

The maxmin criterion is useless here because it fails to distinguish between any menus.

Claim 14. *Every menu is a maxmin menu.*

Proof. Suppose the buyer believes that the distribution π^{v^j} assigns probability one to $v^j = 0$. In that case, no menu can guarantee a positive payoff. □

The minimax regret criterion is only marginally more useful. It does not distinguish between menus that only sell the grand bundle and menus that also sell the goods separately.

It specifies a price for the grand bundle that depends on the maximum value $\bar{v}$. If we allow $\bar{v} \to \infty$, the price of any bundle x in the menu grows to infinity, regardless of whether consumers are likely to have these extreme values.

Claim 15. *A menu is a minimax regret menu iff the price of the grand bundle is $m\bar{v}/2$.*

Proof. Let M be a menu. Suppose the buyer believes that the distributions π^{v^j} assigns probability one to $v^j = \bar{v}$. The seller regrets not choosing a menu M' that offers the grand bundle at price $m\bar{v}$. Specifically, his regret from M is $m\bar{v}$ minus the price of the grand bundle in M.

Suppose the buyer believes that the distributions π^{v^j} assigns probability one to values that make the buyer exactly indifferent between buying and not buying the grand bundle in M. In particular, by the definition of menus, I can set values such that the buyer is exactly indifferent between buying and not buying any bundle $(x, p) \in M$. She breaks indifferences in favor of not buying. The seller regrets not choosing a menu M' that offers the grand bundle at an ϵ discount, for arbitrarily small $\epsilon > 0$. This would generate profits arbitrarily close to the price of the grand bundle in M, whereas M generates zero profits. Therefore, the seller's regret from M is price of the grand bundle in M.

Altogether, the maximum regret is minimized when the menu M splits the difference and offers the grand bundle at price $m\bar{v}/2$. □

Optimal Penalized Menu. In contrast to the previous results, the optimal penalized policy will recommend bundling when the sample size is small. To formalize this, I need to define the penalized policy in a setting where the agent cares about quantile, rather than expected utility.

First, I redefine regret. Let x_n denote the bundle that the buyer purchases, as a function of the realized sample. Given menu M, the buyer's *quantile regret* is

$$\text{Q-Regret}(M, q, \pi^v) = \max_{(x,p)\in M} U^A\left(q, M, (x,p), \pi^v\right) - U^A\left(q, M, x_n, \pi^v\right)$$

Next, I specify a bound B on quantile regret. Let $H > 0$ be a constant and let K be defined as in assumption 19. Then

$$B(M, q, \pi^v) := 4HK^{-1}\sqrt{\frac{\ln 2 + \ln|M| + \ln n}{n}}$$

The following proposition implies that this regret bound is feasible in the spirit of Definition 44.

Claim 16. *There exists a constant $H > 0$ and a buyer's strategy x_n such that*

$$\text{Q-Regret}(M, q, \pi^v) \leq 4HK^{-1}\sqrt{\frac{\ln 2 + \ln|M| + \ln n}{n}}$$

The constant H varies with model parameters ($\bar{v}$, m, and K) but does not depend on π^s or M.

The key feature of the regret bound B is its dependence on the menu size $|M|$. This has immediate implications. If the seller only offers the grand bundle, then $|M| = 1$ and the term $\ln|M| = 0$ in the bound disappears. On the other hand, if the seller sells every item separately, then $|M| = 2^m$ and the term $\ln|M| = m$ in the bound is equal to the number of products. As such, if there are many products, selling separately could mean much less predictable buyer behavior compared to only offering the grand bundle.

The next claim formalizes this intuition with a limiting argument. It refers to the menu M^S in which all goods are sold separately (so $|M^S| = 2^m$).

Claim 17. *Let M be an optimal penalized menu. For every sample size n and fraction*

$\alpha \in (0, 1)$, *there exists a sufficiently large number* m *of products such that the menu size* $|M|$ *is less than an* α*-fraction of the menu size from selling separately, i.e.*

$$\frac{|M|}{|M^S|} \leq \alpha$$

Proof. Suppose for contradiction that $|M| \geq \alpha|M^S|$, i.e. $|M| \geq \alpha 2^m$. As $m \to \infty$, the regret bound grows to infinity since $|M|$ grows to infinity. This causes the worst-case payoff to fall to zero. However, offering only the grand bundle at a suitable price sets $|M| = 1$ and ensures a positive profit (by assumption 19). This contradicts the optimality of M.

□

3.7 Related Literature

This work contributes to three research efforts. For robust mechanism design, it is a principled way to interpolate between two extremes: the common prior and prior-freeness. For learning in games, it provides a convenient behavioral assumption that does not rely on agents using a particular model or estimator. For data-driven mechanism design, it extends existing work to settings where the agent is learning, not just the policymaker. Below, I elaborate on each of these contributions.

Robust Mechanism Design. Robust mechanism design tries to relax the common prior assumption, as well as other knowledge assumptions used in mechanism design.

Initially, this literature focused on prior-free solution concepts that assumed no distributional knowledge whatsoever. Early on, Bergemann and Morris (2005) and Chung and Ely (2007) gave prior-free foundations for ex post incentive compatibility as a solution concept.[12] These papers worked with Harsanyi type spaces, where type profiles encode both

[12] Later contributions moved beyond ex post incentive compatibility (e.g. Börgers and Smith 2014, Börgers

the distribution of the state (or payoff type) as well as agents' higher-order beliefs. They sought to implement a social choice correspondence in any Bayes-Nash equilibrium of any type space.[13]

Prior-free solution concepts and the common prior assumption are two extreme cases, but there is a rich terrain that lies between them. For example, Oury and Tercieux (2012) propose *continuous implementation*. Given a type space that satisfies the common prior, the designer wants to implement a social choice correspondence in all type spaces that are arbitrarily close to the original one. Other researchers take a similar approach (e.g. Meyer-ter-Vehn and Morris 2011, Jehiel, Meyer-ter-Vehn, and Moldovanu 2012). Alternatively, Artemov et al. (2013) assume Δ-rationalizability (Battigalli and Siniscalchi 2003), where it is commonly known that the state distribution belongs to some pre-specified set Δ. In a similar spirit, Ollár and Penta (2017) assume that only pre-specified moments of the state distribution are common knowledge.

My work can also be seen as straddling the divide between prior-freeness on the one hand and the common prior on the other.[14] It is clearly inspired by the robustness literature, but its methods are different. Rather than specify a moment restriction or set Δ of plausible distributions, I assume that the agent has access to a dataset with sample size n. The parameter n controls how knowledgeable the policymaker and agent are supposed to be. It is a principled way to interpolate between prior-freeness ($n = 0$) and the common prior ($n = \infty$). This is true regardless of whether the dataset is interpreted literally (as in this paper), or as a stylized model of shared experience.

2017).

[13] Prior-free approaches also developed in algorithmic game theory (Goldberg et al. 2006). This work focused on prior-free approximations to Bayesian-optimal mechanisms, whereas the economics literature focused on worst-case optimal mechanisms and characterizing which social choice correspondences were implementable.

[14] Granted, I avoid some of the issues related to strategic uncertainty that the prior literature has to deal with, due to my focus on single-agent mechanism design problems. But my learning-theoretic approach to relaxing the common prior assumption also seems to be compatible with multi-agent settings (c.f. Liang 2020).

My model has three advantages relative to the nearest alternatives, namely Artemov et al. (2013) and Ollár and Penta (2017). First, it has few tuning parameters. The only parameter related to beliefs is the sample size n. Second, it makes it easier to decide "how much" robustness is required. I posit that researchers find it easier to gauge whether a sample size (or rate of convergence) is reasonable for their setting of interest, compared to an arbitrary set of beliefs or a set of moment restrictions. Third, it has a clear learning foundation. This is important because "robust" predictions can be quite sensitive to how one departs from the common prior assumption, so we need a good justification if we want to prioritize one over the other.[15]

Learning in Games. The literature on learning in games tries to replace prior knowledge or equilibrium assumptions with a more explicit process of learning from historical data. It is useful to divide this literature along two dimensions. First, whether data arises from repeated interaction (i.e. online learning) or random sampling (i.e. batch learning). Second, whether agents are learning about each other's strategies or about the state of nature. I am primarily concerned with models where agents learn about the state of nature through random sampling.

Liang (2020) is particularly relevant to my work. The author also studies incomplete information games where agents learn about the state through a finite dataset. In her model, agents adopt learning rules from a prespecified class of learning rules. If the learning rules are consistent, and converge uniformly, then predicted behavior is compatible with the common prior assumption in the limit as the sample size grows. In finite samples, predictions that hold under the common prior assumption (like the no-trade theorem) might not be necessarily true.

[15] For example, if departures are defined using the product topology on the universal type space, then even small departures from the common prior can lead to drastic changes in predicted behavior (Lipman 2003; Rubinstein 1989; Weinstein and Yildiz 2007). In contrast, under the strategic topology, small departures from the common prior lead to small changes in predicted behavior (Chen et al. 2010; E. Dekel et al. 2006).

However, this paper differs from Liang (2020) in two respects. First, I commit to a particular class of learning rules: those that satisfy my regret bound. This class contains all learning rules that perform at least as well as empirical utility maximization, given the true distribution. By identifying a natural class of learning rules, I reduce the burden on researchers who want to use Liang's method. Second, my goal is policy design rather than predicting behavior. The new insights from my model do not come from agents learning per se. Instead, they come from the fact that policy choices can impact how quickly agents learn.

Researchers have also looked at the implications of statistical complexity for economic behavior. Some of this work considers the trade-off, from the agent's perspective, of choosing more or less complex statistical models to estimate (e.g. Al-Najjar and Pai 2014, Olea et al. 2021). Other work studies models of bounded rationality that can be motivated as a response to statistical complexity (e.g. Valenzuela-Stookey 2020, Jehiel 2005). In contrast, my work looks at how policy choices can make the agent's learning problem more or less complex. Furthermore, I do not assume that the agent is frequentist or that she relies on a particular statistical model.

Finally, researchers have also studied environments where agents' beliefs may not converge. This could be due to bounded rationality (Aragones et al. 2005; Haghtalab et al. 2021) or the fact that the environment is hopelessly complicated (Mailath and L. Samuelson 2020; Al-Najjar 2009). In particular, Al-Najjar (2009) relies on the notion of VC dimension, which is closely linked to the notion of Rademacher complexity used in this paper. Aside from this, however, the focus of these papers is different from my own. My assumptions imply that the agent can learn her optimal response to any policy, given sufficient data.

Data-driven Mechanism Design. The literature on data-driven mechanism design fuses robust mechanism design with learning in games. The goal is to combine microeconomic theory with data to provide more concrete policy recommendations. There is not much interaction between this literature and prior work in structural econometrics, presumably because it is driven by a community of computer scientists and microeconomic theorists, rather than empirical economists. As a result, the methods are somewhat different. The focus tends to be decision-theoretic, in line with Manski (2019), and there is less emphasis on estimating model parameters per se.

One prominent line of work studies the sample complexity of auction design. Here, the auctioneer lacks prior knowledge of the distribution of bidder values. Instead, he has access to a dataset, usually consisting of i.i.d. draws from the value distribution. A typical question is how many draws are needed in order for the auctioneer to guarantee near-optimal revenue with high probability (e.g. Balcan, Blum, Hartline, et al. 2008, Cole and Roughgarden 2014). Many of these papers rely on measures of learning complexity, like covering numbers (Balcan, Blum, Hartline, et al. 2008), pseudo-dimension (J. Morgenstern and Roughgarden 2015), and Rademacher complexity (Syrgkanis 2017). Gonçalves and Furtado (2020) use more familiar econometric methods towards a similar end.

These papers are focused on the auctioneer's learning problem, but ignore the bidders' learning problems. This is possible because of the application that they emphasize: auctions with dominant strategies, where agents have independent private values. In that context, there is no need for agents to learn about the value distribution. But this is not true in general. For example, in models with interdependent values, implementating reasonable outcomes in dominant strategies may be impossible (Jehiel, Meyer-ter-Vehn, Moldovanu, and Zame 2006). Alternatively, consider problems like monopoly regulation, contract design, or Bayesian persuasion. In these problems, the agent's optimal action depends on her beliefs over a hidden state of nature.

There is some prior work where both the policymaker and the agents are learning from data. For example, Camara et al. (2020) also study a single-agent policy design problem.

Their model incorporates online learning, where data is generated over time through repeated interaction, and the data-generating process is arbitrary. In contrast, I consider batch learning, where data is generated from random sampling. Furthermore, Cummings et al. (2020) and Immorlica, Mao, et al. (2020) consider agents that are learning from i.i.d. samples, respectively, in models of price discrimination and social learning. Both papers assume that agents' beliefs converge to the true distribution at a reasonable rate. In contrast, I assume that the agent's regret converges to zero at a reasonable rate. Although these assumptions should be mutually compatible, the advantage of regret bounds is that they make explicit how policies affect the complexity of the agent's learning problem.

3.8 Conclusion

I proposed a modeling assumption that replaces common knowledge with a common dataset. I studied this in the context of incomplete-information games where a policymaker commits to a policy, an agent responds, and both are able to learn from the available data. I formalized the modeling assumption using concepts like regret, Rademacher complexity, and sample privacy. I showed that policies that are too complex in a precise sense can be suboptimal because they lead to unpredictable behavior. I proposed penalized policies and motivated them through theoretical guarantees and illustrative examples.

The most important – and challenging – direction for future work is to turn this methodology towards real applications. One approach is to find highly-structured, data-rich settings where the data-driven penalized policy developed in this paper can actually be used. Particularly promising areas may lie in education or healthcare, where rich value-added measures have been used for policies like teacher compensation. Another approach is to treat penalization as desirable even in the absence of an explicit dataset. Here, the dataset would be seen as a metaphor for experiences that shape the agent's beliefs. Lab experiments could be used to determine whether policies that are more complex (in the sense of this paper) actually lead to suboptimal responses.

Bringing this method to any serious application will likely also require theoretical extensions. For example, there may be settings where the distribution is only partially identified, where there are multiple agents interacting strategically, or where more efficient estimators can take advantage of particular problem structure. These are all worthwhile open questions.

Bibliography

Abeler, J. & Marklein, F. (2016, December). Fungibility, Labels, and Consumption. *Journal of the European Economic Association*, *15*(1), 99–127.

Adleman, L. (1978). Two theorems on random polynomial time. In *19th annual symposium on foundations of computer science* (pp. 75–83).

Akbarpour, M., Kominers, S. D., Li, S., & Milgrom, P. (2021, March). *Investment incentives in near-optimal mechanisms.*

Akbarpour, M. & Li, S. (2020). Credible auctions: a trilemma. *Econometrica*, *88*(2), 425–467.

Alon, N., Lingas, A., & Wahlén, M. (2007). Approximating the maximum clique minor and some subgraph homeomorphism problems. *Theoretical Computer Science*, *374*(1), 149–158.

Anderlini, L. & Sabourian, H. (1995). Cooperation and effective computability. *Econometrica*, *63*(6), 1337–1369.

Anunrojwong, J., Iyer, K., & Manshadi, V. (2020). Information design for congested social services: optimal need-based persuasion. In *Proceedings of the 21st acm conference on economics and computation* (pp. 349–350). EC '20. Virtual Event, Hungary: Association for Computing Machinery.

Apesteguia, J. & Ballester, M. A. (2010). The computational complexity of rationalizing behavior. *Journal of Mathematical Economics*, *46*(3), 356–363.

Aragones, E., Gilboa, I., Postlewaite, A., & Schmeidler, D. (2005, December). Fact-free learning. *American Economic Review*, *95*(5), 1355–1368.

Arora, R., Dekel, O., & Tewari, A. (2012). Online bandit learning against an adaptive adversary: from regret to policy regret. In *Proceedings of the 29th international coference on international conference on machine learning* (pp. 1747–1754). ICML'12. Edinburgh, Scotland: Omnipress.

Arora, R., Dinitz, M., Marinov, T. V., & Mohri, M. (2018). Policy regret in repeated games. In *Proceedings of the 32nd international conference on neural information processing systems* (pp. 6733–6742). NIPS'18. Montréal, Canada: Curran Associates Inc.

Arora, S. & Barak, B. (2009). *Computational complexity: a modern approach.* Cambridge University Press.

Artemov, G., Kunimoto, T., & Serrano, R. (2013). Robust virtual implementation: Toward a reinterpretation of the Wilson doctrine. *Journal of Economic Theory*, *148*(2), 424–447.

Balcan, M.-F., Blum, A., Haghtalab, N., & Procaccia, A. D. (2015). Commitment without regrets: online learning in stackelberg security games. In *Proceedings of the sixteenth acm conference on economics and computation* (pp. 61–78). EC '15. Portland, Oregon, USA: Association for Computing Machinery.

Balcan, M.-F., Blum, A., Hartline, J. D., & Mansour, Y. (2008). Reducing mechanism design to algorithm design via machine learning. *Journal of Computer and System Sciences*, *74*(8), 1245–1270.

Barberis, N., Huang, M., & Thaler, R. H. (2006, September). Individual preferences, monetary gambles, and stock market participation: a case for narrow framing. *American Economic Review*, *96*(4), 1069–1090.

Bartlett, P. L. & Mendelson, S. (2003, March). Rademacher and gaussian complexities: risk bounds and structural results. *J. Mach. Learn. Res. 3*, 463–482.

Battigalli, P. & Siniscalchi, M. (2003). Rationalization and Incomplete Information. *The B.E. Journal of Theoretical Economics*, *3*(1), 1–46.

Benartzi, S. & Thaler, R. H. (1995). Myopic loss aversion and the equity premium puzzle. *The Quarterly Journal of Economics*, *110*(1), 73–92.

Bennett, C. H. & Gill, J. (1981). Relative to a Random Oracle A, PA != NPA != co-NPA with Probability 1. *SIAM Journal on Computing*, *10*(1), 96–113.

Bergemann, D., Brooks, B., & Morris, S. (2017). First-price auctions with general information structures: implications for bidding and revenue. *Econometrica*, *85*(1), 107–143.

Bergemann, D. & Morris, S. (2005). Robust mechanism design. *Econometrica*, *73*(6), 1771–1813.

Bergemann, D. & Morris, S. (2013). Robust predictions in games with incomplete information. *Econometrica*, *81*(4), 1251–1308.

Blum, A., Gunasekar, S., Lykouris, T., & Srebro, N. (2018). On preserving non-discrimination when combining expert advice. In *Proceedings of the 32nd international conference on neural information processing systems* (pp. 8386–8397). NIPS'18. Montréal, Canada: Curran Associates Inc.

Blum, A., Hajiaghayi, M., Ligett, K., & Roth, A. (2008). Regret minimization and the price of total anarchy. In *Proceedings of the fortieth annual acm symposium on theory of computing* (pp. 373–382). STOC '08. Victoria, British Columbia, Canada: ACM.

Blum, A. & Mansour, Y. (2007, December). From external to internal regret. *J. Mach. Learn. Res. 8*, 1307–1324.

Blume, L., Brandenburger, A., & Dekel, E. (1989, May). An overview of lexicographic choice under uncertainty. *Ann. Oper. Res. 19*(1-4), 231–246.

Börgers, T. (2017, June). (no) foundations of dominant-strategy mechanisms: a comment on chung and ely (2007). *Review of Economic Design*, *21*(2), 73–82.

Börgers, T. & Smith, D. (2014, May). Robust mechanism design and dominant strategy voting rules. *Theoretical Economics*, *9*(2), 339–360.

Boutilier, C. (2012). Eliciting forecasts from self-interested experts: scoring rules for decision makers. In *Proceedings of the 11th international conference on autonomous agents and multiagent systems - volume 2* (pp. 737–744). AAMAS '12. Valencia, Spain: International Foundation for Autonomous Agents and Multiagent Systems.

Braverman, M., Mao, J., Schneider, J., & Weinberg, M. (2018). Selling to a no-regret buyer. In *Proceedings of the 2018 acm conference on economics and computation* (pp. 523–538). EC '18. Ithaca, NY, USA: ACM.

Brown, J. R., Kapteyn, A., Luttmer, E. F. P., Mitchell, O. S., & Samek, A. (2021, July). Behavioral Impediments to Valuing Annuities: Complexity and Choice Bracketing. *The Review of Economics and Statistics*, *103*(3), 533–546.

Buşoniu, L., Babuška, R., & De Schutter, B. (2010). Multi-agent reinforcement learning: an overview. In D. Srinivasan & L. C. Jain (Eds.), *Innovations in multi-agent systems and applications - 1* (pp. 183–221). Berlin, Heidelberg: Springer Berlin Heidelberg.

Cai, L. & Juedes, D. (2003). On the existence of subexponential parameterized algorithms. *Journal of Computer and System Sciences*, *67*(4), 789–807.

Camara, M. K., Hartline, J. D., & Johnsen, A. (2020). Mechanisms for a no-regret agent: beyond the common prior. In *2020 ieee 61st annual symposium on foundations of computer science (focs)* (pp. 259–270).

Camerer, C., Babcock, L., Loewenstein, G., & Thaler, R. H. (1997, May). Labor Supply of New York City Cabdrivers: One Day at a Time. *The Quarterly Journal of Economics*, *112*(2), 407–441.

Carroll, G. (2015, February). Robustness and linear contracts. *American Economic Review*, *105*(2), 536–63.

Carroll, G. (2017). Robustness and separation in multidimensional screening. *Econometrica*, *85*(2), 453–488.

Carroll, G. & Meng, D. (2016a). Locally robust contracts for moral hazard. *Journal of Mathematical Economics*, *62*, 36–51.

Carroll, G. & Meng, D. (2016b). Robust contracting with additive noise. *Journal of Economic Theory*, *166*, 586–604.

Cesa-Bianchi, N. & Lugosi, G. (2006). *Prediction, learning, and games*. New York, NY, USA: Cambridge University Press.

Chen, Y.-C., di Tillio, A., Faingold, E., & Xiong, S. (2010). Uniform topologies on types. *Theoretical Economics*, *5*(3), 445–478.

Choi, J. J., Laibson, D., & Madrian, B. C. (2009, December). Mental accounting in portfolio choice: evidence from a flypaper effect. *American Economic Review*, *99*(5), 2085–95.

Chung, K.-S. & Ely, J. C. (2007). Foundations of dominant-strategy mechanisms. *The Review of Economic Studies*, *74*(2), 447–476.

Cole, R. & Roughgarden, T. (2014). The sample complexity of revenue maximization. In *Proceedings of the forty-sixth annual acm symposium on theory of computing* (pp. 243–252). STOC '14. New York, New York: ACM.

Cook, S. A. (1971). The complexity of theorem-proving procedures. In *Proceedings of the third annual acm symposium on theory of computing* (pp. 151–158). STOC '71. Shaker Heights, Ohio, USA: ACM.

Cummings, R., Devanur, N. R., Huang, Z., & Wang, X. (2020). Algorithmic price discrimination. In *Proceedings of the Thirty-First Annual ACM-SIAM Symposium on Discrete Algorithms*. SODA '20. Salt Lake City, Utah, USA.

Das, S., Kamenica, E., & Mirka, R. (2017, October). Reducing congestion through information design. In *2017 55th annual allerton conference on communication, control, and computing (allerton)* (pp. 1279–1284).

Daskalakis, C. & Syrgkanis, V. (2016). Learning in auctions: regret is hard, envy is easy. In *2016 ieee 57th annual symposium on foundations of computer science (focs)* (pp. 219–228).

Daskalakis, C., Goldberg, P. W., & Papadimitriou, C. H. (2009). The complexity of computing a nash equilibrium. *SIAM Journal on Computing*, *39*(1), 195–259.

Dekel, E., Fudenberg, D., & Morris, S. (2006). Topologies on types. *Theoretical Economics*, *1*(3), 275–309.

Deng, Y., Schneider, J., & Sivan, B. (2019). Strategizing against no-regret learners. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, & R. Garnett (Eds.), *Advances in neural information processing systems 32* (pp. 1579–1587). Curran Associates, Inc.

Devanur, N. R., Peres, Y., & Sivan, B. (2019). Perfect bayesian equilibria in repeated sales. *Games and Economic Behavior*, *118*, 570–588.

Dudík, M., Haghtalab, N., Luo, H., Schapire, R. E., Syrgkanis, V., & Vaughan, J. W. (2017). Oracle-efficient online learning and auction design. In *2017 ieee 58th annual symposium on foundations of computer science (focs)* (pp. 528–539).

Dughmi, S. & Xu, H. (2016). Algorithmic Bayesian persuasion. In *Proceedings of the forty-eighth annual acm symposium on theory of computing* (pp. 412–425). STOC '16. Cambridge, MA, USA: ACM.

Dütting, P., Roughgarden, T., & Talgam-Cohen, I. (2019). Simple versus optimal contracts. In *Proceedings of the 2019 acm conference on economics and computation* (pp. 369–387). EC '19. Phoenix, AZ, USA: ACM.

Dwork, C., McSherry, F., Nissim, K., & Smith, A. (2006). Calibrating noise to sensitivity in private data analysis. In S. Halevi & T. Rabin (Eds.), *Theory of cryptography* (pp. 265–284). Berlin, Heidelberg: Springer Berlin Heidelberg.

Dwork, C. & Roth, A. (2014, August). The algorithmic foundations of differential privacy. *Found. Trends Theor. Comput. Sci. 9*(3-4), 211–407.

Echenique, F., Golovin, D., & Wierman, A. (2011). A revealed preference approach to computational complexity in economics. In *Proceedings of the 12th acm conference on electronic commerce* (pp. 101–110). EC '11. San Jose, California, USA: Association for Computing Machinery.

Ely, J. C. & Szydlowski, M. (2020). Moving the goalposts. *Journal of Political Economy*, *128*(2), 468–506.

Eppstein, D. (2009). Finding large clique minors is hard. *J. Graph Algorithms Appl. 13*(2), 197–204.

Ergin, H. & Sarver, T. (2010). A unique costly contemplation representation. *Econometrica*, *78*(4), 1285–1339.

Feng, Y. & Hartline, J. D. (2018, October). *An end-to-end argument in mechanism design (prior-independent auctions for budgeted agents)*. Los Alamitos, CA, USA: IEEE Computer Society.

Fortnow, L. & Vohra, R. V. (2009). The complexity of forecast testing. *Econometrica*, *77*(1), 93–105.

Foster, D. P. & Vohra, R. V. (1997). Calibrated learning and correlated equilibrium. *Games and Economic Behavior*, *21*(1), 40–55.

Gabaix, X. (2014, September). A Sparsity-Based Model of Bounded Rationality. *The Quarterly Journal of Economics*, *129*(4), 1661–1710.

Gabaix, X., Laibson, D., Moloche, G., & Weinberg, S. (2006, September). Costly information acquisition: experimental analysis of a boundedly rational model. *American Economic Review*, *96*(4), 1043–1068.

Gale, D. & Shapley, L. S. (1962). College admissions and the stability of marriage. *The American Mathematical Monthly*, *69*(1), 9–15.

Garey, M., Johnson, D., & Stockmeyer, L. (1976). Some simplified np-complete graph problems. *Theoretical Computer Science*, *1*(3), 237–267.

Gasarch, W. I. (2019, March). Guest column: the third p=?np poll. *SIGACT News*, *50*(1), 38–59.

Gilboa, I., Postlewaite, A., & Schmeidler, D. (2021). The complexity of the consumer problem. *Research in Economics*, *75*(1), 96–103.

Gilboa, I. & Zemel, E. (1989). Nash and correlated equilibria: some complexity considerations. *Games and Economic Behavior*, *1*(1), 80–93.

Gofer, E. & Mansour, Y. (2016, April). Lower bounds on individual sequence regret. *Machine Learning*, *103*(1), 1–26.

Goldberg, A. V., Hartline, J. D., Karlin, A. R., Saks, M., & Wright, A. (2006). Competitive auctions. *Games and Economic Behavior*, *55*(2), 242–269. Mini Special Issue: Electronic Market Design.

Goldstein, I. & Leitner, Y. (2018). Stress tests and information disclosure. *Journal of Economic Theory*, *177*, 34–69.

Gonçalves, D. & Furtado, B. (2020, August). *Statistical mechanism design: robust pricing and reliable projections.*

Haagerup, U. (1981). The best constants in the khintchine inequality. *Studia Mathematica*, *70*(3), 231–283.

Hadwiger, H. (1943). Über eine klassifikation der streckenkomplexe. *Vierteljahrsschr Naturforsch, Ges. Zurich*, *88*, 133–142.

Haghtalab, N., Jackson, M. O., & Procaccia, A. D. (2021). Belief polarization in a complex world: a learning theory perspective. *Proceedings of the National Academy of Sciences*, *118*(19).

Hart, S. & Mas-Colell, A. (2001). A general class of adaptive strategies. *Journal of Economic Theory*, *98*(1), 26–54.

Hartline, J. D., Johnsen, A., Nekipelov, D., & Zoeter, O. (2019). Dashboard mechanisms for online marketplaces. In *Proceedings of the 2019 acm conference on economics and computation* (pp. 591–592). EC '19. Phoenix, AZ, USA: ACM.

Hartline, J. D. & Lucier, B. (2015, October). Non-optimal mechanism design. *American Economic Review*, *105*(10), 3102–24.

Hartline, J. D., Syrgkanis, V., & Tardos, É. (2015). No-regret learning in bayesian games. In *Proceedings of the 28th international conference on neural information processing systems - volume 2* (pp. 3061–3069). NIPS'15. Montreal, Canada: MIT Press.

Hastings, J. & Shapiro, J. M. (2018, December). How are snap benefits spent? evidence from a retail panel. *American Economic Review*, *108*(12), 3493–3540.

Hązła, J., Jadbabaie, A., Mossel, E., & Rahimian, M. A. (2021). Bayesian decision making in groups is hard. *Operations Research*, *69*(2), 632–654.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, *6*(2), 65–70.

Holmstrom, B. & Milgrom, P. (1987). Aggregation and linearity in the provision of intertemporal incentives. *Econometrica*, *55*(2), 303–328.

Holmstrom, B. & Milgrom, P. (1991). Multitask principal-agent analyses: incentive contracts, asset ownership, and job design. *Journal of Law, Economics, & Organization*, *7*, 24–52.

Hossain, T. & Morgan, J. (2006). ...plus shipping and handling: revenue (non) equivalence in field experiments on ebay. *The B.E. Journal of Economic Analysis & Policy*, *5*(2), 1–30.

Hu, J. & Wellman, M. P. (1998). Multiagent reinforcement learning: theoretical framework and an algorithm. In *Proceedings of the fifteenth international conference on machine learning* (pp. 242–250). ICML '98. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.

Immorlica, N., Lucier, B., Pountourakis, E., & Taggart, S. (2017). Repeated sales with multiple strategic buyers. (pp. 167–168). EC '17. Cambridge, Massachusetts, USA: Association for Computing Machinery.

Immorlica, N., Mao, J., Slivkins, A., & Wu, Z. S. (2020). Incentivizing exploration with selective data disclosure. In *Proceedings of the 21st acm conference on economics and computation* (pp. 647–648). EC '20. Virtual Event, Hungary: Association for Computing Machinery.

Jakobsen, A. M. (2020, May). A model of complex contracts. *American Economic Review*, *110*(5), 1243–73.

Jehiel, P. (2005). Analogy-based expectation equilibrium. *Journal of Economic Theory*, *123*(2), 81–104.

Jehiel, P., Meyer-ter-Vehn, M., & Moldovanu, B. (2012). Locally robust implementation and its limits. *Journal of Economic Theory*, *147*(6), 2439–2452.

Jehiel, P., Meyer-ter-Vehn, M., Moldovanu, B., & Zame, W. R. (2006). The limits of ex post implementation. *Econometrica*, *74*(3), 585–610.

Johnson, D. S. (1974). Approximation algorithms for combinatorial problems. *Journal of Computer and System Sciences*, *9*(3), 256–278.

Jose, V. R. R., Nau, R. F., & Winkler, R. L. (2008). Scoring rules, generalized entropy, and utility maximization. *Operations Research*, *56*(5), 1146–1157.

Kamenica, E. & Gentzkow, M. (2011, October). Bayesian persuasion. *American Economic Review*, *101*(6), 2590–2615.

Karp, R. M. (1972). Reducibility among combinatorial problems. In R. E. Miller, J. W. Thatcher, & J. D. Bohlinger (Eds.), *Complexity of computer computations: proceedings of a symposium on the complexity of computer computations* (pp. 85–103). Boston, MA: Springer US.

Karp, R. M. & Lipton, R. J. (1980). Some connections between nonuniform and uniform complexity classes. In *Proceedings of the twelfth annual acm symposium on theory of computing* (pp. 302–309). STOC '80. Los Angeles, California, USA: Association for Computing Machinery.

Kearns, M., Mansour, Y., & Ng, A. Y. (1999). Approximate planning in large pomdps via reusable trajectories. In *Proceedings of the 12th international conference on neural information processing systems* (pp. 1001–1007). NIPS'99. Denver, CO: MIT Press.

Kitagawa, T. & Tetenov, A. (2018). Who should be treated? empirical welfare maximization methods for treatment choice. *Econometrica*, *86*(2), 591–616.

Koch, A. K. & Nafziger, J. (2016). Goals and bracketing under mental accounting. *Journal of Economic Theory*, *162*, 305–351.

Koch, A. K. & Nafziger, J. (2019). Correlates of narrow bracketing. *The Scandinavian Journal of Economics*, *121*(4), 1441–1472.

Kohli, R., Krishnamurti, R., & Mirchandani, P. (1994, May). The minimum satisfiability problem. *SIAM J. Discret. Math.* *7*(2), 275–283.

Kostochka, A. V. (1984). Lower bound of the hadwiger number of graphs by their average degree. *Combinatorica*, *4*(4), 307–316.

Köszegi, B. & Matějka, F. (2020, January). Choice Simplification: A Theory of Mental Budgeting and Naive Diversification. *The Quarterly Journal of Economics*, *135*(2), 1153–1207.

Lian, C. (2020, December). A Theory of Narrow Thinking. *The Review of Economic Studies*, *88*(5), 2344–2374.

Liang, A. (2020, July). *Games of incomplete information played by statisticians.*

Lipman, B. L. (1999). Decision theory without logical omniscience: toward an axiomatic framework for bounded rationality. *The Review of Economic Studies*, *66*(2), 339–361.

Lipman, B. L. (2003). Finite order implications of common priors. *Econometrica*, *71*(4), 1255–1267.

Littlestone, N. & Warmuth, M. (1994). The weighted majority algorithm. *Information and Computation*, *108*(2), 212–261.

Littman, M. L. (1994). Markov games as a framework for multi-agent reinforcement learning. In *Proceedings of the eleventh international conference on international conference on machine learning* (pp. 157–163). ICML'94. New Brunswick, NJ, USA: Morgan Kaufmann Publishers Inc.

Mahajan, M. & Raman, V. (1999). Parameterizing above guaranteed values: maxsat and maxcut. *Journal of Algorithms*, *31*(2), 335–354.

Mailath, G. J. & Samuelson, L. (2020, May). Learning under diverse world views: model-based inference. *American Economic Review*, *110*(5), 1464–1501.

Mandler, M. (2015). Rational agents are the quickest. *Journal of Economic Theory*, *155*, 206–233.

Mandler, M., Manzini, P., & Mariotti, M. (2012). A million answers to twenty questions: choosing by checklist. *Journal of Economic Theory*, *147*(1), 71–92.

Manski, C. F. (1993, January). Adolescent econometricians: how do youth infer the returns to schooling? In *Studies of supply and demand in higher education* (pp. 43–60). University of Chicago Press.

Manski, C. F. (2004a). Measuring expectations. *Econometrica*, *72*(5), 1329–1376.

Manski, C. F. (2004b). Statistical treatment rules for heterogeneous populations. *Econometrica*, *72*(4), 1221–1246.

Manski, C. F. (2011, August). Actualist rationality. *Theory and Decision*, *71*(2), 195–210.

Manski, C. F. (2019, December). *Econometrics for decision making: building foundations sketched by haavelmo and wald* (Working Paper No. 26596). National Bureau of Economic Research.

Manski, C. F. & Tetenov, A. (2007). Admissible treatment rules for a risk-averse planner with experimental data on an innovation. *Journal of Statistical Planning and Inference*, *137*(6), 1998–2010.

Mansour, Y., Slivkins, A., Syrgkanis, V., & Wu, Z. S. (2016). Bayesian exploration: incentivizing exploration in bayesian games. In *Proceedings of the 2016 acm conference on economics and computation* (p. 661). EC '16. Maastricht, The Netherlands.

Martin, V. (2017). When to quit: narrow bracketing and reference dependence in taxi drivers. *Journal of Economic Behavior & Organization*, *144*, 166–187.

Mas-Collell, A., Whinston, M., & Green, J. R. (1995). *Microeconomic theory*. Oxford University Press.

Massart, P. (2000). Some applications of concentration inequalities to statistics. *Annales de la Faculté des sciences de Toulouse: Mathématiques*, *9*(2), 245–303.

Mbakop, E. & Tabord-Meehan, M. (2021). Model selection for treatment choice: penalized welfare maximization. *Econometrica*, *89*(2), 825–848.

McCarthy, J. (1956). Measures of the value of information. *Proceedings of the National Academy of Sciences*, *42*(9), 654–655.

McDiarmid, C. (1989). On the method of bounded differences. In J. Siemons (Ed.), *Surveys in combinatorics, 1989: invited papers at the twelfth british combinatorial conference* (pp. 148–188). London Mathematical Society Lecture Note Series. Cambridge University Press.

Meyer-ter-Vehn, M. & Morris, S. (2011). The robustness of robust implementation. *Journal of Economic Theory*, *146*(5), 2093–2104.

Mirrlees, J. A. (1971). An exploration in the theory of optimum income taxation. *The Review of Economic Studies*, *38*(2), 175–208.

Morgenstern, J. & Roughgarden, T. (2015). The pseudo-dimension of near-optimal auctions. In *Proceedings of the 28th international conference on neural information processing systems - volume 1* (pp. 136–144). NIPS'15. Montreal, Canada: MIT Press.

Myerson, R. B. (1981). Optimal auction design. *Mathematics of Operations Research*, *6*(1), 58–73.

Al-Najjar, N. I. (2009). Decision makers as statisticians: diversity, ambiguity, and learning. *Econometrica*, *77*(5), 1371–1401.

Al-Najjar, N. I. & Pai, M. M. (2014). Coarse decision making and overfitting. *Journal of Economic Theory*, *150*, 467–486.

Nekipelov, D., Syrgkanis, V., & Tardos, É. (2015). Econometrics for learning agents. In *Proceedings of the sixteenth acm conference on economics and computation* (pp. 1–18). EC '15. Portland, Oregon, USA: ACM.

Nisan, N. & Ronen, A. (2001). Algorithmic mechanism design. *Games and Economic Behavior*, *35*(1), 166–196.

Olea, J. L. M., Ortoleva, P., Pai, M. M., & Prat, A. (2021, February). Competing models.

Ollár, M. & Penta, A. (2017, August). Full implementation and belief restrictions. *American Economic Review*, *107*(8), 2243–77.

Oury, M. & Tercieux, O. (2012). Continuous implementation. *Econometrica*, *80*(4), 1605–1637.

Papadimitriou, C. H., Vempala, S. S., Mitropolsky, D., Collins, M., & Maass, W. (2020). Brain computation by assemblies of neurons. *Proceedings of the National Academy of Sciences*, *117*(25), 14464–14472.

Rabin, M. & Weizsäcker, G. (2009, September). Narrow bracketing and dominated choices. *American Economic Review*, *99*(4), 1508–43.

Read, D., Loewenstein, G., & Rabin, M. (1999, December). Choice bracketing. *Journal of Risk and Uncertainty*, *19*(1), 171–197.

Richter, M. K. & Wong, K.-C. (1999a). Computable preference and utility. *Journal of Mathematical Economics*, *32*(3), 339–354.

Richter, M. K. & Wong, K.-C. (1999b). Non-computability of competitive equilibrium. *Economic Theory*, *14*(1), 1–27.

Romano, J. P. & Wolf, M. (2005). Stepwise multiple testing as formalized data snooping. *Econometrica*, *73*(4), 1237–1282.

Ross, S. A. (1973). The economic theory of agency: the principal's problem. *The American Economic Review*, *63*(2), 134–139.

Roth, A. E. (1982). The economics of matching: stability and incentives. *Mathematics of Operations Research*, *7*(4), 617–628.

Roughgarden, T. (2021). *Beyond the worst-case analysis of algorithms*. Cambridge University Press.

Rubinstein, A. (1986). Finite automata play the repeated prisoner's dilemma. *Journal of Economic Theory*, *39*(1), 83–96.

Rubinstein, A. (1989). The electronic mail game: strategic behavior under "almost common knowledge". *The American Economic Review*, *79*(3), 385–391.

Rubinstein, A. (1993). On price recognition and computational complexity in a monopolistic model. *Journal of Political Economy*, *101*(3), 473–484.

Ryabko, D. & Hutter, M. (2008). On the possibility of learning in reactive environments with arbitrary dependence. *Theoretical Computer Science*, *405*(3), 274–284.

Salant, Y. & Cherry, J. (2020). Statistical inference in games. *Econometrica*, *88*(4), 1725–1752.

Samuelson, P. A. (1938). A note on the pure theory of consumer's behaviour. *Economica*, *5*(17), 61–71.

Sappington, D. (1983). Limited liability contracts between principal and agent. *Journal of Economic Theory*, *29*(1), 1–21.

Schaefer, T. J. (1978). The complexity of satisfiability problems. In *Proceedings of the tenth annual acm symposium on theory of computing* (pp. 216–226). STOC '78. San Diego, California, USA: Association for Computing Machinery.

Simonsohn, U. & Gino, F. (2013). Daily horizons: evidence of narrow bracketing in judgment from 10 years of m.b.a. admissions interviews. *Psychological Science*, *24*(2), 219–224.

Spence, M. & Zeckhauser, R. (1971). Insurance, information, and individual action. *The American Economic Review*, *61*(2), 380–387.

Stoye, J. (2009). Minimax regret treatment choice with finite samples. *Journal of Econometrics*, *151*(1), 7081.

Stracke, R., Kerschbamer, R., & Sunde, U. (2017). Coping with complexity: experimental evidence for narrow bracketing in multi-stage contests. *European Economic Review*, *98*, 264–281.

Syrgkanis, V. (2017). A sample complexity measure with applications to learning optimal auctions. In *Proceedings of the 31st international conference on neural information processing systems* (pp. 5358–5365). NIPS'17. Long Beach, California, USA: Curran Associates Inc.

Szekeres, G. & Wilf, H. S. (1968). An inequality for the chromatic number of a graph. *Journal of Combinatorial Theory*, *4*(1), 1–3.

Thaler, R. H. (1985). Mental accounting and consumer choice. *Marketing Science*, *4*(3), 199–214.

Tversky, A. & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, *211*(4481), 453–458.

Uther, W. & Veloso, M. (2003, January). *Adversarial reinforcement learning.*

Valenzuela-Stookey, Q. (2020, September). *Subjective complexity under uncertainty.*

Vickrey, W. (1961). Counterspeculation, auctions, and competitive sealed tenders. *The Journal of Finance*, *16*(1), 8–37.

von Neumann, J. & Morgenstern, O. (1944). *Theory of games and economic behavior.* Princeton University Press.

Wald, A. (1950). *Statistical decision functions.* Wiley: New York.

Wasserstein, R. L. & Lazar, N. A. (2016). The asa statement on p-values: context, process, and purpose. *The American Statistician*, *70*(2), 129–133.

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

Wilson, A. (2014). Bounded memory and biases in information processing. *Econometrica*, *82*(6), 2257–2294.

Wilson, R. (1987). Game-theoretic analyses of trading processes. In T. F. Bewley (Ed.), *Advances in economic theory: fifth world congress* (pp. 33–70). Econometric Society Monographs. Cambridge University Press.

Zhang, M. (2021). A theory of choice bracketing under risk. In *Proceedings of the 22nd acm conference on economics and computation* (pp. 886–887). EC '21. Budapest, Hungary: Association for Computing Machinery.

Appendix A

Omissions from Chapter 1

A.1 Proof of Lemmas 1 and 6

I begin by proving lemma 6. Recall definition 25 and inequality (1.9). If u is (i, j, n)-separable then there cannot exist a violation of (i, j, n)-separability. Applying (i, j, n)-separability, the first line of inequality(1.9) becomes

$$\begin{aligned}
&u_i\left(\ldots, x_{i-1}, a_1, x_{i+1}, \ldots, x_{j-1}, x_{j+1}, \ldots\right) \\
+&u_j\left(\ldots, x_{i-1}, x_{i+1}, \ldots, x_{j-1}, a_2, x_{j+1}, \ldots\right) \\
+&u_i\left(\ldots, x_{i-1}, b_1, x_{i+1} \ldots, x_{j-1}, x_{j+1}, \ldots\right) \\
+&u_j\left(\ldots, x_{i-1}, x_{i+1} \ldots, x_{j-1}, b_2, x_{j+1}, \ldots\right)
\end{aligned}$$

while the second line becomes

$$\begin{aligned}
&u_i\left(\ldots, x_{i-1}, a_1, x_{i+1}, \ldots, x_{j-1}, x_{j+1}, \ldots\right) \\
+&u_j\left(\ldots, x_{i-1}, x_{i+1}, \ldots, x_{j-1}, b_2, x_{j+1}, \ldots\right) \\
+&u_i\left(\ldots, x_{i-1}, b_1, x_{i+1}, \ldots, x_{j-1}, x_{j+1}, \ldots\right) \\
+&u_j\left(\ldots, x_{i-1}, x_{i+1}, \ldots, x_{j-1}, a_2, x_{j+1}, \ldots\right)
\end{aligned}$$

These two expressions are the same, up to reordering of terms.

Next, consider the converse. Suppose that u does not have a violation of (i, j, n)-

separability. I claim that u is (i,j,n)-separable. Note that inequality (1.9) becomes an equality, where for all values (x,a_1,a_2,b_1,b_2),

$$\begin{aligned} &u\left(\ldots,x_{i-1},a_1,x_{i+1},\ldots,x_{j-1},a_2,x_{j+1},\ldots\right)+u\left(\ldots,x_{i-1},b_1,x_{i+1}\ldots,x_{j-1},b_2,x_{j+1},\ldots\right)\\ &=u\left(\ldots,x_{i-1},a_1,x_{i+1},\ldots,x_{j-1},b_2,x_{j+1},\ldots\right)+u\left(\ldots,x_{i-1},b_1,x_{i+1},\ldots,x_{j-1},a_2,x_{j+1},\ldots\right)\end{aligned}$$

If we set

$$a_1 := x_i \quad a_2 := x_j \quad b_1 := 0 \quad b_2 := 0$$

and rearrange terms, then

$$u(x) = u\left(\ldots,x_{j-1},0,x_{j+1},\ldots\right)+u\left(\ldots,x_{i-1},0,x_{i+1},\ldots\right)-u\left(\ldots,x_{i-1},0,x_{i+1}\ldots,x_{j-1},0,x_{j+1},\ldots\right)$$

This satisfies the definition of (i,j,n)-separability. This completes the proof of lemma 6.

Next, consider lemma 1. I want to show that if u is symmetric, it is additively separable iff there exists no violation (x,a_1,a_2,b_1,b_2) of (i,j,n)-separability where $a := a_1 = a_2$ and $b := b_1 = b_2$. One direction is immediate: if u is additively separable then it is (i,j,n)-separable, and therefore has no violation of (i,j,n)-separability. In the other direction, non-existence of violations implies

$$\begin{aligned} &u\left(\ldots,x_{i-1},a,x_{i+1},\ldots,x_{j-1},a,x_{j+1},\ldots\right)+u\left(\ldots,x_{i-1},b,x_{i+1}\ldots,x_{j-1},b,x_{j+1},\ldots\right)\\ &=u\left(\ldots,x_{i-1},a,x_{i+1},\ldots,x_{j-1},b,x_{j+1},\ldots\right)+u\left(\ldots,x_{i-1},b,x_{i+1},\ldots,x_{j-1},a,x_{j+1},\ldots\right)\end{aligned}$$

Note that

$$u\left(\ldots,x_{i-1},a,x_{i+1},\ldots,x_{j-1},b,x_{j+1},\ldots\right)=u\left(\ldots,x_{i-1},b,x_{i+1},\ldots,x_{j-1},a,x_{j+1},\ldots\right)$$

by symmetry. Applying this to the previous equation and rearranging yields

$$\begin{aligned} u\left(\ldots, x_{i-1}, a, x_{i+1}, \ldots, x_{j-1}, b, x_{j+1}, \ldots\right) = & \frac{1}{2} \cdot u\left(\ldots, x_{i-1}, a, x_{i+1} \ldots, x_{j-1}, a, x_{j+1}, \ldots\right) \\ & + \frac{1}{2} \cdot u\left(\ldots, x_{i-1}, b, x_{i+1} \ldots, x_{j-1}, b, x_{j+1}, \ldots\right) \end{aligned}$$

This implies (i, j, n)-separability. Since u is (i, j, n)-separable for all i, j, n, it is additively separable. This completes the proof of lemma 1.

A.2 Proof of Lemma 2

In section 1.3.3, I proved Lemma 2 for the maximum and quadratic utility functions. My goal here is to extend the construction for the quadratic utility function to any symmetric utility function u where there exist constants $a, b \in \mathbb{Q}$ and an outcome $x \in \mathcal{X}$ such that

$$u(a, a, x_3, x_4, \ldots) + u(b, b, x_3, x_4, \ldots) > u(a, b, x_3, x_4, \ldots) + u(b, a, x_3, x_4, \ldots) \tag{A.1}$$

Note that a, b, x can be described in $O(1)$ time, since x is N-dimensional where N does not change with the number of variables n. For convenience, let $n \geq N$. This is without loss since the asymptotic runtime is determined by $n \to \infty$.

I prove the result for two separate cases, which are collectively exhaustive.

Case 1. In the first case,

$$u(a, a, x_3, x_4, \ldots) \neq u(b, b, x_3, x_4, \ldots)$$

Without loss of generality, assume

$$u(a, a, x_3, x_4, \ldots) < u(b, b, x_3, x_4, \ldots) \tag{A.2}$$

It follows from this and condition (A.1) that

$$u(a, b, x_3, x_4, \ldots) < u(b, b, x_3, x_4, \ldots) \tag{A.3}$$

I modify the division of the sample space Ω depicted in figure 1.3.3. For each clause j, let the first subinterval have length α/m. Let the second and third subintervals have length $(1 - \alpha)/(2m)$.

For each clause j, let x^j be a permutation of $(a, b, x_3, x_4, \ldots)$. The values a, b will coincide with the two dimensions i where variable i is represented in clause j. Formally, if $v_{j_1} = v_i$ or $v_{j_1} = \neg v_i$ then $x^j_i = a$. If $v_{j_2} = v_i$ or $v_{j_2} = \neg v_i$ then $x^j_i = b$. Otherwise, the sequence x^j is in the same order as x, i.e. x_3 precedes x_4, x_4 precedes x_5, and so forth.

When ω falls into the first subinterval associated with clause j, set

$$X^T_i(\omega) = \begin{cases} b & v_{j_1} = v_i \\ a & v_{j_1} = \neg v_i \\ b & v_{j_2} = v_i \\ a & v_{j_2} = \neg v_i \\ x^j_i & \text{otherwise} \end{cases} \qquad X^F_i(\omega) = \begin{cases} b & v_{j_1} = \neg v_i \\ a & v_{j_1} = v_i \\ b & v_{j_2} = \neg v_i \\ a & v_{j_2} = v_i \\ x^j_i & \text{otherwise} \end{cases} \tag{A.4}$$

When ω falls into the second subinterval associated with clause j, set

$$X_i^T(\omega) = \begin{cases} b & \neg v_{j_1} = v_i \\ a & \neg v_{j_1} = \neg v_i \\ b & v_{j_2} = v_i \\ a & v_{j_2} = \neg v_i \\ x_i^j & \text{otherwise} \end{cases} \qquad X_i^F(\omega) = \begin{cases} b & \neg v_{j_1} = \neg v_i \\ a & \neg v_{j_1} = v_i \\ b & v_{j_2} = \neg v_i \\ a & v_{j_2} = v_i \\ x_i^j & \text{otherwise} \end{cases} \tag{A.5}$$

When ω falls into the third subinterval associated with clause j, set

$$X_i^T(\omega) = \begin{cases} b & v_{j_1} = v_i \\ a & v_{j_1} = \neg v_i \\ b & \neg v_{j_2} = v_i \\ a & \neg v_{j_2} = \neg v_i \\ x_i^j & \text{otherwise} \end{cases} \qquad X_i^F(\omega) = \begin{cases} b & v_{j_1} = \neg v_i \\ a & v_{j_1} = v_i \\ b & \neg v_{j_2} = \neg v_i \\ a & \neg v_{j_2} = v_i \\ x_i^j & \text{otherwise} \end{cases} \tag{A.6}$$

Now, consider the decisionmaker's expected utility from lottery X, conditioned on the interval associated with clause j. That is,

$$\mathrm{E}\left[u\left(X_1,\ldots,X_n\right) \mid \omega \in \left[\frac{j-1}{m}, \frac{j}{m}\right)\right] \tag{A.7}$$

I will use the fact that u is symmetric to reorder $X_1,\ldots,X_n$ as needed. When the assignment $g(X)$ makes both variables in clause j true, expected utility (A.7) becomes

$$A := \frac{1}{3}\left(\alpha u\left(b, b, x_3, x_4, \ldots\right) + (1-\alpha) u\left(a, b, x_3, x_4, \ldots\right)\right) \tag{A.8}$$

When the assignment $g(X)$ makes v_{j_1} true but v_{j_2} false, expected utility (A.7) becomes

$$B := \frac{1}{3}\left(\alpha u\left(b, a, x_3, x_4, \ldots\right) + \frac{1-\alpha}{2}\cdot u\left(b, b, x_3, x_4, \ldots\right) + \frac{1-\alpha}{2}\cdot u\left(a, a, x_3, x_4, \ldots\right)\right) \tag{A.9}$$

Since u is symmetric, this is also true if $g(X)$ makes v_{j_2} true but v_{j_1} false. Finally, when $g(X)$ makes neither entry in clause j true, expected utility (A.7) becomes

$$C := \frac{1}{3}\left(\alpha u\left(a, a, x_3, x_4, \ldots\right) + (1-\alpha) u\left(a, b, x_3, x_4, \ldots\right)\right) \tag{A.10}$$

When $\alpha = 1$, $A > B$. This follows from condition (A.3). When $\alpha = 0$, $B > A$. This follows from condition (A.1). For any $\alpha > 0$, $A > C$. This follows from condition (A.2).

The expressions A, B, C are continuous in α. It follows from this and the observations in the previous paragraph that there exists a value $\alpha \in (0, 1)$ such that $A = B > C$. The value of α depends on the choice correspondence c via the revealed utility function u, but it does not depend on n or the boolean formula BF. From here, as in the quadratic utility case of section 1.3.3, it follows that maximizing expected utility is equivalent to MAX 2-SAT.

Case 2. In the second case,

$$u(a, a, x_3, x_4, \ldots) = u(b, b, x_3, x_4, \ldots) \tag{A.11}$$

It follows from this and condition (A.1) that

$$u(a, b, x_3, x_4, \ldots) < u(b, b, x_3, x_4, \ldots) \tag{A.12}$$

My previous construction no longer works, since it would imply $A = C$. However, I can repair the argument with a similar construction.

Let each interval in the sample space Ω associated with clause j be divided into four subintervals, rather than three. The first two subintervals have length $\alpha/(2m)$. The last two subintervals have length $(1-\alpha)/(2m)$. When ω falls into the first subinterval associated with clause j, set

$$X_i^T(\omega) = \begin{cases} b & v_{j_1} = v_i \\ b & v_{j_1} = \neg v_i \\ b & v_{j_2} = v_i \\ a & v_{j_2} = \neg v_i \\ x_i^j & \text{otherwise} \end{cases} \qquad X_i^F(\omega) = \begin{cases} b & v_{j_1} = \neg v_i \\ b & v_{j_1} = v_i \\ b & v_{j_2} = \neg v_i \\ a & v_{j_2} = v_i \\ x_i^j & \text{otherwise} \end{cases}$$

Intuitively, this mirrors equation (A.4), except that the first assertion v_{j_1} in clause j is automatically true. When ω falls into the second subinterval associated with clause j, set

$$X_i^T(\omega) = \begin{cases} b & v_{j_1} = v_i \\ a & v_{j_1} = \neg v_i \\ b & v_{j_2} = v_i \\ b & v_{j_2} = \neg v_i \\ x_i^j & \text{otherwise} \end{cases} \qquad X_i^F(\omega) = \begin{cases} b & v_{j_1} = \neg v_i \\ a & v_{j_1} = v_i \\ b & v_{j_2} = \neg v_i \\ b & v_{j_2} = v_i \\ x_i^j & \text{otherwise} \end{cases}$$

Intuitively, this mirrors equation (A.4), except that the second assertion v_{j_2} in clause j is automatically true. When ω falls into the third subinterval associated with clause j, define $X_i^T(\omega), X_i^F(\omega)$ according to equation (A.5). When ω falls into the fourth subinterval associated with clause j, define $X_i^T(\omega), X_i^F(\omega)$ according to equation (A.6).

Now, consider the decisionmaker's expected utility from lottery X, conditioned on the interval associated with clause j. When the assignment $g(X)$ makes both variables in clause j true, expected utility (A.7) becomes

$$A := \frac{1}{3}\left(\alpha u\left(b, b, x_3, x_4, \ldots\right) + (1-\alpha) u\left(a, b, x_3, x_4, \ldots\right)\right) \tag{A.13}$$

When the assignment $g(X)$ makes v_{j_1} true but v_{j_2} false, expected utility (A.7) becomes

$$\begin{aligned} B := &\frac{1}{3}\Big(\frac{\alpha}{2} \cdot u\left(b, a, x_3, x_4, \ldots\right) + \frac{\alpha}{2} \cdot u\left(b, b, x_3, x_4, \ldots\right) \\ &+ \frac{1-\alpha}{2} \cdot u\left(b, b, x_3, x_4, \ldots\right) + \frac{1-\alpha}{2} \cdot u\left(a, a, x_3, x_4, \ldots\right)\Big) \end{aligned}$$

Since u is symmetric, this is also true if $g(X)$ makes v_{j_2} true but v_{j_1} false. Finally, when $g(X)$ makes neither entry in clause j true, expected utility (A.7) becomes

$$C := \frac{1}{3}\left(\alpha u\left(b, a, x_3, x_4, \ldots\right) + (1-\alpha) u\left(a, b, x_3, x_4, \ldots\right)\right) \tag{A.14}$$

When $\alpha = 1$, $A > B$. This follows from condition (A.12). When $\alpha = 0$, $B > A$. This follows from condition (A.1). For any $\alpha > 0$, $A > C$. This follows from condition (A.12).

The expressions A, B, C are continuous in α. It follows from this and the observations in the previous paragraph that there exists a value $\alpha \in (0,1)$ such that $A = B > C$. As in case 1, this implies that maximizing expected utility is equivalent to MAX 2-SAT.

A.3 Proof of Lemma 3

The argument is similar to the proof of Lemma 2. Let u be any symmetric utility function where there exist constants $a, b \in \mathbb{Q}$ and an N-dimensional outcome $x \in \mathcal{X}$ such that

$$u(a, a, x_3, x_4, \ldots) + u(b, b, x_3, x_4, \ldots) < u(a, b, x_3, x_4, \ldots) + u(b, a, x_3, x_4, \ldots) \tag{A.15}$$

As before, assume without loss that $n \geq N$, and consider the following two cases.

Case 1. In the first case,

$$u(a, a, x_3, x_4, \ldots) \neq u(b, b, x_3, x_4, \ldots)$$

Without loss of generality, assume

$$u(a, a, x_3, x_4, \ldots) < u(b, b, x_3, x_4, \ldots) \tag{A.16}$$

It follows from this and condition (A.15) that

$$u(a, b, x_3, x_4, \ldots) > u(a, a, x_3, x_4, \ldots) \tag{A.17}$$

The construction is almost identical to the construction in case 1 of the proof of Lemma 2. The only exception is that, in the definitions of X_i^T and X_i^F, I replace any value a with b, and any value b with a. As before, consider the decisionmaker's expected utility from lottery X, conditioned on the interval associated with clause j. When the assignment $g(X)$ makes both variables in clause j true, the conditional expected utility becomes

$$A := \frac{1}{3}\left(\alpha u\left(a, a, x_3, x_4, \ldots\right) + (1-\alpha) u\left(b, a, x_3, x_4, \ldots\right)\right) \tag{A.18}$$

When the assignment $g(X)$ makes v_{j_1} true but v_{j_2} false, the conditional expected utility becomes

$$B := \frac{1}{3}\left(\alpha u\left(a, b, x_3, x_4, \ldots\right) + \frac{1-\alpha}{2} \cdot u\left(a, a, x_3, x_4, \ldots\right) + \frac{1-\alpha}{2} \cdot u\left(b, b, x_3, x_4, \ldots\right)\right) \tag{A.19}$$

Since u is symmetric, this is also true if $g(X)$ makes v_{j_2} true but v_{j_1} false. Finally, when

$g(X)$ makes neither entry in clause j true, the conditional expected utility becomes

$$C := \frac{1}{3}\left(\alpha u\left(b, b, x_3, x_4, \ldots\right) + (1-\alpha) u\left(b, a, x_3, x_4, \ldots\right)\right) \tag{A.20}$$

When $\alpha = 1$, $B > A$. This follows from condition (A.17). When $\alpha = 0$, $A > B$. This follows from condition (A.15). For any $\alpha > 0$, $C > A$. This follows from condition (A.16).

The expressions A, B, C are continuous in α. It follows from this and the observations in the previous paragraph that there exists a value $\alpha \in (0, 1)$ such that $A = B < C$. For this value of α, the expected utility conditioned on the interval associated with clause j is equal to A iff the assignment $g(X)$ makes clause j true. Otherwise, it is equal to $C > A$.

If the assignment $g(X)$ makes k_T clauses true and k_F clauses false, then the unconditional expected utility is $Ak_T + Ck_F$. Since $C > A$, this is proportional to the number of clauses j that are false. Maximizing expected utility is equivalent to maximizing the number of clauses that are false. In turn, this is equivalent to minimizing the number of clauses that are true. Therefore, if $X \in c(M)$ then the assignment $g(X)$ solves MIN 2-SAT.

Case 2. In the second case,

$$u(a, a, x_3, x_4, \ldots) = u(b, b, x_3, x_4, \ldots) \tag{A.21}$$

It follows from this and condition (A.15) that

$$u(a, b, x_3, x_4, \ldots) > u(a, a, x_3, x_4, \ldots) \tag{A.22}$$

The construction is almost identical to the construction in case 2 of the proof of Lemma 2. The only exception is that, in the definitions of X_i^T and X_i^F, I replace any value a with b, and any value b with a. As before, consider the decisionmaker's expected utility from lottery

X, conditioned on the interval associated with clause j. When the assignment $g(X)$ makes both variables in clause j true, expected utility (A.7) becomes

$$A := \frac{1}{3}\left(\alpha u\left(a,a,x_3,x_4,\ldots\right) + (1-\alpha)u\left(b,a,x_3,x_4,\ldots\right)\right) \tag{A.23}$$

When the assignment $g(X)$ makes v_{j_1} true but v_{j_2} false, the conditional expected utility becomes

$$\begin{aligned}B :=& \frac{1}{3}\Big(\frac{\alpha}{2}\cdot u\left(a,b,x_3,x_4,\ldots\right) + \frac{\alpha}{2}\cdot u\left(a,a,x_3,x_4,\ldots\right)\\ &+ \frac{1-\alpha}{2}\cdot u\left(b,b,x_3,x_4,\ldots\right) + \frac{1-\alpha}{2}\cdot u\left(a,a,x_3,x_4,\ldots\right)\Big)\end{aligned}$$

Since u is symmetric, this is also true if $g(X)$ makes v_{j_2} true but v_{j_1} false. Finally, when $g(X)$ makes neither entry in clause j true, the conditional expected utility becomes

$$C := \frac{1}{3}\left(\alpha u\left(a,b,x_3,x_4,\ldots\right) + (1-\alpha)u\left(b,a,x_3,x_4,\ldots\right)\right) \tag{A.24}$$

When $\alpha = 1$, $B > A$. This follows from condition (A.22). When $\alpha = 0$, $A > B$. This follows from condition (A.15). For any $\alpha > 0$, $C > A$. This follows from condition (A.22).

The expressions A, B, C are continuous in α. It follows from this and the observations in the previous paragraph that there exists a value $\alpha \in (0,1)$ such that $A = B < C$. As in case 1, this implies that maximizing expected utility is equivalent to MIN 2-SAT.

A.4 Proof of Lemma 4

Consider the outcome $\bar{x}^n$ that maximizes utility $u(x)$ across all n-dimensional outcomes x. Similarly, consider the outcome $\underline{x}^n$ that minimizes utility $u(x)$ across all n-dimensional outcomes x.

Given an outcome x and parameter ϵ, the Turing machine performs the following computation. Let $k = \lfloor 1/\epsilon \rfloor$. Construct a grid

$$Y = \{\epsilon, 2\epsilon, \ldots, (k-1)\epsilon, k\epsilon\}$$

For every $y \in Y$, define a lottery X^y as follows. When $\omega \leq y$, $X^y(\omega) = \underline{x}^n$. Otherwise, $X^y(\omega) = \bar{x}^n$. Finally, output the largest value $y \in Y$ such that

$$x \in c\left(\{x, X^y\}\right)$$

This is well-defined by assumption 2, which ensures that binary menus are represented in the collection $\mathcal{M}$. Moreover, this can be done in polynomial time since c is strongly tractable.

A.5 Proof of Lemma 5

Let u be a continuous utility function where $d_n = \text{Had}(G_n(u))$. Let M denote an n-dimensional product menu. Let BF denote a boolean formula with d_n variables $v_1, \ldots, v_{d_n}$. Suppose there exists a $O(\text{poly}(n))$-time algorithm that maximizes expected utility in any menu M. I want to find a $O(\text{poly}(n))$-time algorithm that solves MAX 2-SAT for any boolean formula .

There are two main steps to this proof. In step 1, I construct an auxilliary formula BF' with n variables $v'_1, \ldots, v'_n$, using polynomial-size advice. This will be an instance of a weighted MAX 2-SAT problem, where weights are allowed to be negative. A solution to this auxilliary problem with correspond to a solution to the original problem. In step 2, I will reduce the weighted MAX 2-SAT problem to expected utility maximization, using polynomial-size advice. This will be similar to the proof of Lemmas 2 and 3. It follows that the solving MAX 2-SAT for the original formula is weakly tractable if expected utility maximization is weakly tractable.

Step 1. Let $\tilde{G}_n(u)$ be the largest complete minor of $G_n(u)$. By definition, this has d_n nodes. Let k be an arbitrary node in $\tilde{G}_n(u)$. By definition of the graph minor, there is a subset of nodes in $G_n(u)$ whose edges were contracted to form k. Let τ denote the size of this subset, and let $k_1, \ldots, k_\tau$ denote the nodes themselves.

First, I add clauses to the auxilliary formula BF' that represent clauses in the original formula BF. Consider a clause CL_j in the original formula BF. Let v_i be a variable represented in CL_j, which corresponds to node $k^{j,i}$ in $\tilde{G}_n(u)$. For each clause j and pair of variables (say, i and $-i$), choose nodes $h^{j,i} \in \{k_1^{j,i}, \ldots, k_\tau^{j,i}\}$ such that that $h^{j,i}$ and $h^{j,-i}$ share an edge in the inseparability graph $G_n(u)$. I claim that it is always possible to find such a pair. Since $\tilde{G}_n(u)$ is a complete graph, there is an edge between nodes $k^{j,i}$ and $k^{j,-i}$ in $\tilde{G}_n(u)$. Since $\tilde{G}_n(u)$ was produced by edge contractions, that edge $(k^{j,i}, k^{j,-i})$ can exist only if they represent nodes that share an edge in $G_n(u)$. This proves the claim.

I have identified the variables $v'_{h^{j,i}}$ and $v'_{h^{j,-i}}$, but not yet added a clause. Recall from lemma 6 that since $h^{j,i}$ and $h^{j,-i}$ share an edge in the inseparability graph $G_n(u)$, there is a violation of $((h^{j,i}, h^{j,-i}), n)$-separability. That violation consists of an n-dimensional outcome x^j and quadruple $a_1^j, a_2^j, b_1^j, b_2^j \in [0, 1]$. Which clauses I add depends on the direction of that violation. For convenience, for arbitrary $a, b \in [0, 1]$, let

$$\tilde{u}^j(a, b) := u\left(\ldots, x^j_{h^{j,i}-1}, a, x^j_{h^{j,i}+1}, \ldots, x^j_{h^{j,-i}-1}, b, x^j_{h^{j,-i}+1}, \ldots\right)$$

There are two cases to consider.

1. Suppose that

$$\tilde{u}^j(a_1^j, a_2^j) + \tilde{u}^j(b_1^j, b_2^j) > \tilde{u}^j(a_1^j, b_2^j) + \tilde{u}^j(b_1^j, a_2^j)$$

   Add the clause $v'_{h^{j,i}} \vee v'_{h^{j,-i}}$ to the auxilliary formula, with weight 1.

2. Suppose that

$$\tilde{u}^j(a_1^j, a_2^j) + \tilde{u}^j(b_1^j, b_2^j) < \tilde{u}^j(a_1^j, b_2^j) + \tilde{u}^j(b_1^j, a_2^j)$$

   Add three clauses to the auxilliary formula: $\neg v'_{h^{j,i}} \vee \neg v'_{h^{j,-i}}$, $\neg v'_{h^{j,i}} \vee v'_{h^{j,-i}}$, and $v'_{h^{j,i}} \vee \neg v'_{h^{j,-i}}$. Each of these clauses has weight -1.

For intuition, compare the three clauses in case 2 to the clause $v'_{h^{j,i}} \vee v'_{h^{j,-i}}$ in case 1. The case 1 clause is true if and only if exactly two of the three case 2 clauses are satisfied. The case 1 clause is false if and only if all three of the case 2 clauses are satisfied. Therefore, the unweighted case 2 clauses are a way to represent the assertion that the case 1 clause is false. By adding weight -1, this effectively becomes an assertion that the case 1 clause is true.

Next, I add clauses to the auxilliary formula BF' that capture the constraint that, for any node k in $G'_n(u)$, we have $v'_{k_i} = v'_{k_j}$ for all $i, j \leq \tau$. Without loss of generality, suppose that $k_1, \ldots, k_n$ are ordered in a way where k_i has an edge with k_{i+1} in the inseparability graph $G_n(u)$. This is always possible since node k was created by contracting a sequence of edges (k_i, k_{i+1}) in $G_n(u)$. Let $x^{k_i}, (a_1^{k_i}, a_2^{k_i}, b_1^{k_i}, b_2^{k_i})$ be a violation of (k_i, k_{i+1}, n)-separability, and let

$$\tilde{u}^{k_i}(a, b) := u\left(\ldots, x^{k_i}_{k_i - 1}, a, x^{k_i}_{k_i + 1}, \ldots, x^{k_i}_{k_{i+1} - 1}, b, x^{k_i}_{k_{i+1} + 1}, \ldots\right)$$

Let $\gamma > 0$ be a constant that I define later. As before, there are two cases.

1. Suppose that

$$\tilde{u}^{k_i}(a_1^{k_i}, a_2^{k_i}) + \tilde{u}^j(b_1^{k_i}, b_2^{k_i}) > \tilde{u}^j(a_1^{k_i}, b_2^{k_i}) + \tilde{u}^j(b_1^{k_i}, a_2^{k_i})$$

   Add the clauses $v'_{k_i} \vee \neg v'_{k_j}$ and $\neg v'_{k_i} \vee v'_{k_j}$ to the auxilliary formula BF'. Each has weight γ.

   Note that an assignment where $v'_{k_i} \neq v'_{k_{i+1}}$ will make one of the two clauses false, whereas an assignment where $v'_{k_i} = v'_{k_{i+1}}$ will make both clauses true. All else equal, since clauses have positive weight $\gamma > 0$, weighted MAX 2-SAT prefers to set $v'_{k_i} = v'_{k_{i+1}}$.

2. Suppose that

$$\tilde{u}^{k_i}(a_1^{k_i}, a_2^{k_i}) + \tilde{u}^j(b_1^{k_i}, b_2^{k_i}) < \tilde{u}^j(a_1^{k_i}, b_2^{k_i}) + \tilde{u}^j(b_1^{k_i}, a_2^{k_i})$$

Add the clauses $v'_{k_i} \vee v'_{k_j}$ and $\neg v'_{k_i} \vee \neg v'_{k_j}$ to the auxilliary formula BF'. Each has weight $-\gamma$.

Note that an assignment where $v'_{k_i} \neq v'_{k_{i+1}}$ will make both of the two clauses true, whereas an assignment where $v'_{k_i} = v'_{k_{i+1}}$ will make only one clause true. All else equal, since clauses have negative weight $-\gamma$, weighted MAX 2-SAT *still* prefers to set $v'_{k_i} = v'_{k_{i+1}}$.

Intuitively, if the weight γ is large enough, then weighted MAX 2-SAT will prioritize $v'_{k_i} = v'_{k_{i+1}}$ over satisfying any of the other clauses in BF'. Since this applies for all $i = 1, \ldots, \tau$, this will ensure that $v'_{k_i} = v'_{k_j}$ for all $i, j \leq \tau$.

I have added all the clauses and only need to specify the weight parameter γ of the clauses that represent constraints. Let there be m clauses in the original formula BF. Let $\gamma := m + 1$. Let m_1 be the number of clauses j in BF that fall into case 1 above, and let m_2 be the number that fall into case 2. Observe that $m_1 + m_2 = m$. Let n_0 be the number of nodes in $G_n(u)$ that were deleted to form the minor $\tilde{G}_n(u)$. Let $n_1 := n - n_0$. In that case, any assignment that satisfies $v'_{k_i} = v'_{k_j}$ for all i, j, k has a weighted value of at least

$$2(n_1 - 1)(m + 1) - 2m_2$$

even if no other clauses are satisfied. Here, $2(n_1 - 1)$ is the number of clauses that represent constraints, multiplied by their weight $m + 1$. Among the clauses in BF' that represent clauses in BF, at least two of the three case 2 clauses are always satisfied; this adds weight $-2m_2$.

In contrast, any assignment where $v'_{k_i} \neq v'_{k_j}$ for some i, j, k has a weighted value of at most

$$2(n_1 - 1)(m + 1) - (m + 1) - 2m_2 + m$$

Here, either $\neg v'_{k_i} \vee v'_{k_j}$ or $v'_{k_i} \vee \neg v'_{k_j}$. The fact that one of these clauses is false implies a weighted loss of $m + 1$. In the ideal case where all case 1 clauses are true and case 2 clauses are false adds a weight of $-2m_2 + m$. This is not enough to compensate for the violation of the constraint.

It follows that the constraint $v'_{k_i} = v'_{k_j}$ is satisfied in any assignment that solves weighted MAX 2-SAT. Given this constraint, any assignment in BF has a corresponding assignment in BF' where setting $v_k = \text{true}$ is equivalent to setting $v'_{k_i} = \text{true}$ for all $i = 1, \ldots, \tau$. If the assignment in BF satisfies some number m_0 of clauses, then the assignment in BF' as a weighted value of

$$2(n_1 - 1)(m + 1) - 2m_2 + m_0$$

by construction. Holding the formula BF fixed, this is proportional to m_0. That is, the number of clauses satisfied in BF is proportional to weighted value in BF'.

It follows that a solution to weighted MAX 2-SAT for the auxiliary formula BF' can be efficiently transformed into a solution to MAX 2-SAT for the original formula BF. Furthermore, the auxilliary formula BF' can be constructed in $O(\text{poly}(n))$ time, given advice that describes the inseparability graph $G_n(u)$, the composition of its largest complete minor $\tilde{G}_n(u)$, and all the violations of (i, j, n)-separability.

Step 2. Having described the auxilliary problem, it remains to construct a menu such that expected utility maximization corresponds to solving weighted MAX 2-SAT. Essentially, I want to recreate the argument that I used in Lemmas 2 and 3.

I begin by splitting the sample space into intervals that represent clauses CL'_j in BF'. Let m' be the number of clauses in BF'. By construction, each clause CL'_j has some weight w_j. Let $\beta_j \geq 0$ be a constant that will be defined later. Associate each clause CL'_j with an interval of length

$$l_j = \frac{\beta_j |w_j|}{\sum_{l=1}^{m'} \beta_l |w_l|}$$

Split each of these intervals into four subintervals. I will specify their widths later.

As in Lemmas 2 and 3, I define partial menus $M_i = \{X_i^T, X_i^F\}$. Suppose that $\omega \in \Omega$ falls into the interval associated with clause j of BF'. By construction of BF', there exists a violation

$$\tilde{u}^j(a_1^j, a_2^j) + \tilde{u}^j(b_1^j, b_2^j) \neq \tilde{u}^j(a_1^j, b_2^j) + \tilde{u}^j(b_1^j, a_2^j)$$

There are four cases to consider, corresponding to cases in Lemmas 2 and 3.

1. Suppose that

$$\tilde{u}^j(a_1^j, a_2^j) + \tilde{u}^j(b_1^j, b_2^j) > \tilde{u}^j(a_1^j, b_2^j) + \tilde{u}^j(b_1^j, a_2^j)$$

where $\tilde{u}^j(a_1^j, a_2^j) \neq \tilde{u}^j(b_1^j, b_2^j)$. Assume without loss of generality that $\tilde{u}^j(b_1^j, b_2^j) > \tilde{u}^j(a_1^j, a_2^j)$.

The construction is analogous to that in case 1 of Lemma 2. When ω falls into the first two subintervals associated with clause j, set

$$X_i^T(\omega) = \begin{cases} b_1^j & v_{j_1} = v_i \\ a_1^j & v_{j_1} = \neg v_i \\ b_2^j & v_{j_2} = v_i \\ a_2^j & v_{j_2} = \neg v_i \\ x_i^j & \text{otherwise} \end{cases} \qquad X_i^F(\omega) = \begin{cases} b_1^j & v_{j_1} = \neg v_i \\ a_1^j & v_{j_1} = v_i \\ b_2^j & v_{j_2} = \neg v_i \\ a_2^j & v_{j_2} = v_i \\ x_i^j & \text{otherwise} \end{cases} \tag{A.25}$$

When ω falls into the third subinterval associated with clause j, set

$$X_i^T(\omega) = \begin{cases} b_1^j & \neg v_{j_1} = v_i \\ a_1^j & \neg v_{j_1} = \neg v_i \\ b_2^j & v_{j_2} = v_i \\ a_2^j & v_{j_2} = \neg v_i \\ x_i^j & \text{otherwise} \end{cases} \qquad X_i^F(\omega) = \begin{cases} b_1^j & \neg v_{j_1} = \neg v_i \\ a_1^j & \neg v_{j_1} = v_i \\ b_2^j & v_{j_2} = \neg v_i \\ a_2^j & v_{j_2} = v_i \\ x_i^j & \text{otherwise} \end{cases} \tag{A.26}$$

When ω falls into the fourth subinterval associated with clause j, set

$$X_i^T(\omega) = \begin{cases} b_1^j & v_{j_1} = v_i \\ a_1^j & v_{j_1} = \neg v_i \\ b_2^j & \neg v_{j_2} = v_i \\ a_2^j & \neg v_{j_2} = \neg v_i \\ x_i^j & \text{otherwise} \end{cases} \qquad X_i^F(\omega) = \begin{cases} b_1^j & v_{j_1} = \neg v_i \\ a_1^j & v_{j_1} = v_i \\ b_2^j & \neg v_{j_2} = \neg v_i \\ a_2^j & \neg v_{j_2} = v_i \\ x_i^j & \text{otherwise} \end{cases} \tag{A.27}$$

Next, I specify the lengths of the subintervals. Recall that the length of the interval associated with clause j is l_j. Let $\alpha \in [0, l_j]$ be a constant. Let the first two subintervals each have width α/l_j. Let the last two subintervals each have width $(1-\alpha)/l_j$. As in case 1 of Lemma 2, there exist constants α and $A > C$ so that expected utility from lottery $X \in M$ conditional on the interval associated with clause j is some constant A when the assignment $g(X)$ makes clause j true, and C otherwise. There are at most n^2 unique such constants, one for every pair of nodes in $G_n(u)$, and I take this as advice.

Finally, I specify the length of the interval associated with clause j by letting $\beta_j = 1/(A-C)$. This ensures that the probability of this is proportional to $|w_j|/(A-C)$. All else equal, the effect of choosing an assignment X that makes clause j true is to increase the unconditional expected utility from $C|w_j|/(A-C)$ to $A|w_j|(A-C)$. The difference is $|w_j|$. This is precisely the weight that the weighted MAX 2-SAT problem assigned to clause j.

2. Suppose that

$$\tilde{u}^j(a_1^j, a_2^j) + \tilde{u}^j(b_1^j, b_2^j) > \tilde{u}^j(a_1^j, b_2^j) + \tilde{u}^j(b_1^j, a_2^j)$$

where $\tilde{u}^j(a_1^j, a_2^j) = \tilde{u}^j(b_1^j, b_2^j)$.

The construction is analogous to that in case 2 of Lemma 2. When ω falls into the first subinterval associated with clause j, set

$$X_i^T(\omega) = \begin{cases} b_1^j & v_{j_1} = v_i \\ b_1^j & v_{j_1} = \neg v_i \\ b_2^j & v_{j_2} = v_i \\ a_2^j & v_{j_2} = \neg v_i \\ x_i^j & \text{otherwise} \end{cases} \qquad X_i^F(\omega) = \begin{cases} b_1^j & v_{j_1} = \neg v_i \\ b_1^j & v_{j_1} = v_i \\ b_2^j & v_{j_2} = \neg v_i \\ a_2^j & v_{j_2} = v_i \\ x_i^j & \text{otherwise} \end{cases}$$

When ω falls into the second subinterval associated with clause j, set

$$X_i^T(\omega) = \begin{cases} b_1^j & v_{j_1} = v_i \\ a_1^j & v_{j_1} = \neg v_i \\ b_2^j & v_{j_2} = v_i \\ b_2^j & v_{j_2} = \neg v_i \\ x_i^j & \text{otherwise} \end{cases} \qquad X_i^F(\omega) = \begin{cases} b_1^j & v_{j_1} = \neg v_i \\ a_1^j & v_{j_1} = v_i \\ b_2^j & v_{j_2} = \neg v_i \\ b_2^j & v_{j_2} = v_i \\ x_i^j & \text{otherwise} \end{cases}$$

When ω falls into the third or fourth subintervals associated with clause j, define $X_i^T(\omega)$ in the same way as in the previous case.

As before, let $\alpha \in [0, l_j]$ be a constant. Let the first two subintervals each have width α/l_j. Let the last two subintervals each have width $(1-\alpha)/l_j$. Let α, A, C be the constants from case 2 of Lemma 2, which I take as advice. Let $\beta_j = 1/(A-C)$.

3. Suppose that

$$\tilde{u}^j(a_1^j, a_2^j) + \tilde{u}^j(b_1^j, b_2^j) < \tilde{u}^j(a_1^j, b_2^j) + \tilde{u}^j(b_1^j, a_2^j) \tag{A.28}$$

where $\tilde{u}^j(a_1^j, a_2^j) \neq \tilde{u}^j(b_1^j, b_2^j)$. This case follows from the construction in case 1 of Lemma 3 in the same way that case 1 follows from the construction in case 1 of Lemma 2. The only difference is that, since $C > A$, I define $\beta_j = 1/(C-A)$.

All else equal, the effect of choosing an assignment X that makes clause j *false* is to increase the unconditional expected utility from by $|w_j|$. By construction, $w_j = -|w_j|$ whenever (A.28) holds. Therefore, the effect of choosing an assignment X that makes clause j *true* is to increase the unconditional expected utility from by w_j. This is precisely the weight that the weighted MAX 2-SAT problem assigned to clause j.

4. Suppose that

$$\tilde{u}^j(a_1^j, a_2^j) + \tilde{u}^j(b_1^j, b_2^j) < \tilde{u}^j(a_1^j, b_2^j) + \tilde{u}^j(b_1^j, a_2^j)$$

   where $\tilde{u}^j(a_1^j, a_2^j) = \tilde{u}^j(b_1^j, b_2^j)$. This case follows from the construction in case 2 of Lemma 3 in the same way that case 2 follows from the construction in case 2 of Lemma 2. The only difference is that, since $C > A$, I define $\beta_j = 1/(C - A)$.

By construction of the menu M, the expected utility of lottery X is proportional to the weighted value of the assignment $g(X)$. Therefore, expected utility maximization solves the weighted MAX 2-SAT problem for the auxilliary formula BF'. Since the menu M and the assignment $g(X)$ can be computed in polynomial-time with polynomial-size advice, this completes step 2. In turn, step 2 completes the proof of Lemma 5.

A.6 Proof of Corollaries 1 and 2

Suppose a weakly tractable choice correspondence maximizes expected utility, where

$$d_n := \text{Had}(G_n(u))$$

Lemma 5 provides a $O(n^k \cdot \text{poly}(n))$-time algorithm to solve MAX k-SAT for any boolean formula with at most d_n variables.

Corollary 1 follows almost immediately. Fix an integer n' such that $d_{n'} = n$. Lemma 5 provides a $O(\text{poly}(n'))$-time algorithm to solve MAX 2-SAT for any boolean formula with at most n variables. I claim that this runtime is also polynomial in n, which proves the corollary. Since $d_n = \Omega(\text{poly}(n))$, $d_n \geq Cn^\alpha$ for some constants C, α. It follows that $n' \leq C^{-1}n^{1/\alpha}$, which implies $n' = O(\text{poly}(n))$. The composition of polynomials is polynomial. This proves the claim.

Corollary 2 is a bit more involved. Fix an integer n' such that $d_{n'} = n$. Lemma 5 provides a $O(\text{poly}(n'))$-time algorithm to solve MAX 2-SAT for any boolean formula with at most n variables. First, I claim that this runtime is subexponential in n. Since $d_n = \omega(\log n)$, $n' = o(2^n)$. Therefore, the runtime is $o(\text{poly}(2^n))$, or $o(2^n)$. This proves the claim.

Second, I show that there exists a subexponential-time algorithm for 3-SAT. Unfortunately, applying the standard reduction from 3-SAT to MAX 2-SAT only yields a $o(\text{poly}(2^{n^2}))$ algorithm for 3-SAT. For that reason, I take an alternate approach. Let $k \geq 0$ be an integer and consider a decision variant of the MAX 2-SAT that asks whether there exists an assignment that satisfies at least k clauses. I claim there exists an algorithm that solves this problem with runtime

$$O\left(2^{o(k)} \cdot \text{poly}(n)\right) \tag{A.29}$$

This can be used to construct an $O(2^{o(n)})$-time algorithm for 3-SAT (Cai and Juedes 2003, Corollary 4.2). The algorithm for the decision variant of MAX 2-SAT for has two cases, which depend on the number m of clauses in the boolean formula.

1. Suppose $k \leq m/2$. Then there always exists an assignment that satisfies at least k clauses (Mahajan and Raman 1999, Proposition 5). The algorithm should always output "true". The runtime is $O(\text{poly}(n))$, which is the amount of time it takes to verify that $k \leq m/2$. This is consistent with equation (A.29).

2. Suppose $k > m/2$. Since there are m clauses and each clause has at most two literals, there can be at most $2m$ unique variables represented in the boolean formula. Run the subexponential-time algorithm for MAX 2-SAT for these $2m$ variables and evaluate whether at least k clauses are satisfied. The runtime is

   $$O\left(2^{o(m)}\right)$$

   which is consistent with equation (A.29) since $k = \Omega(m)$.

This proves the claim, and the corollary.

A.7 Proof of Lemma 7

Let $G := G_n(u)$ be an undirected graph. I claim that

$$\mathrm{Had}(G) = O(\log n) \iff \mathrm{cdgn}(G) = O(\log n)$$

There are two directions to prove.

1. First, let $d = \mathrm{cdgn}(G)$. By definition, there exists a minor G' where every node in G' has degree that is equal to or greater than d. Let $\mathrm{avg}(G)$ be the average degree across all nodes in G'. Clearly, $\mathrm{avg}(G) \geq d$. It follows from Kostochka (1984) that G' has a complete minor G'' containing

$$\Omega\left(\mathrm{avg}(G')/\sqrt{\log \mathrm{avg}(G')}\right)$$

nodes. Since G' is a minor of G, G'' is also a (complete) minor of G. Altogether, this implies

$$\mathrm{Had}(G) = \Omega\left(d/\sqrt{\log d}\right)$$

In particular, if $\mathrm{Had}(G) = O(\log n)$ then $d = O(\log n)$.

2. Second, let $d = \mathrm{Had}(G)$. By definition, there is a complete minor G' with d nodes. Since G' is complete, every node in G' has degree d. Therefore, $\mathrm{cdgn}(G) \geq d$. In particular, if $\mathrm{cdgn}(G) = O(\log n)$ then $d = O(\log n)$.

A.8 Proof of Lemma 8

Let G be an undirected graph where $\text{cdgn}(G) = d$. Let the directed graph $\vec{G}$ have n nodes. I construct it as follows.

1. Find a node i in the original graph G that has degree less than or equal to d. This is always possible since G is a minor of itself, and the contraction degeneracy requires all minors of G to have a node with degree less than or equal to d. Searching over nodes i and evaluating their degree takes $O(n^2)$ time.

2. If node j shares an edge with node i in the undirected graph G, let $\vec{G}$ have a directed edge from node i to node j. By definition, this leaves node i with no more than d outgoing edges. Searching over nodes j takes $O(n)$ time.

3. Delete node i from G. Return to step 1 if G is not empty. This occurs at most $O(n)$ times.

This algorithm has runtime $O(n^3)$.

The only remaining property to verify is that $\vec{G}$ is acyclic. This holds because step 1 visits each node i exactly once, and step 2 only creates an edge from node i to a node j that has not yet been visited. Any path in $\vec{G}$ must be strictly increasing in the order in which step 1 visits nodes. This rules out cycles, which must begin and end with the same node.

A.9 Proof of Lemma 9

I begin by identifying a particular node $k \in F$. Construct a minor G' of the graph G as follows.

1. Starting with the graph G, find an edge (i, j) where $i \in F$ is in the frontier and $j \notin F$ is not.

2. Modify G by contracting the edge (i, j) into a new node i'.

3. Modify the frontier F by removing i and replacing it with i'.

4. Repeat step 1 with the modified graph G and frontier F, until no suitable edges remain.

5. Delete all remaining nodes $i \notin F$. None of these nodes are connected with the frontier F, or there would have been another edge to conrtact in step 4.

Henceforth, let G be the original graph and F the original frontier. For every node $i \in F$ there exists a contracted node i' in the minor G'. By construction, the minor G' has an edge between nodes i' and nodes j' iff one of the following is true.

1. G has an edge between nodes i and j.

2. There is a path in G from i to j that does not go through the frontier F.

Let $d = \text{cdgn}(G)$. By the definition of contraction degeneracy, there exists a node k' in the minor G' with at most d edges. Let $k \in F$ be the node in G that k' represents.

Suppose the algorithm were to visit node k in step 5. First, consider the indirect influencers $i \in I_k$. By definition of I_k, there exists a path from i to k that does not pass through F. Next, consider the nodes k' and i' in G', representing nodes k and $i \in I_k$ in G. There is an edge between k' and i' in G' since, as I just showed, there is a path in G from i to j that does not go through the frontier F. This path will be contracted in the procedure used to define G', until only an edge between k' and i' remains. However, I defined k' as a node that has at most d edges in G'. Since there are at most d nodes i' in G' that share an edge with k', there can be at most d nodes $i \in I_k$. This completes the proof, since I have identified a node k in G where $|I_k| \leq d$.

A.10 Proof of Lemma 10

I only need to verify that the definition of indirect influencers in algorithm 2 is consistent with the definition in 1. This is because the definition of the successors is left arbitrary in algorithm 1, and the definition of the predecessors is identical in both algorithms 1 and 2.

Let I_i be the indirect influencers of algorithm 1. Formally, I_i is the subset of unvisited coordinates j where there is a predecessor $k \in P_i$ whose choice $X_k^*(\cdot)$ depends on X_j. Let I_i' be the indirect influencers of algorithm 2. Formally, I_i' consists of the frontier nodes $j \in F$ where where G contains a path between i and j that does not pass through F.

It is sufficient to show that $I_i \subseteq I_i'$. The fact that I_i' may contain nodes that are not in I_i is immaterial. Algorithm 2 only uses I_i' as an argument for choices $X_{P_i}^*(\cdot)$ of i's predecessors. Any node in $I_i' \setminus I_i$ is superfluous insofar as it does not actually affect the function $X_{P_i}^*(\cdot)$.

To show that $I_i \subseteq I_i'$, consider any node $j \in I_i$. By definition, there is a predecessor $k \in P_i$ whose choice $X_k^*(\cdot)$ depends on X_j.

First, I claim that $j \in F$. This follows from the fact that the choice $X_k^*(\cdot)$ can only depend on partial lotteries associated with frontier nodes. Recall that step 6 of algorithm 2 calls step 6 of algorithm 1. This ensures, at each iteration, that choice $X_k^*(\cdot)$ remains a function of partial lotteries associated unvisited nodes that are either (i) successors or (ii) indirect influencers of some visited node. Successors of visited nodes are added to the frontier F in step 7, and only removed after they are visited. Indirect influencers are in the frontier by definition, and only removed after they are visited. Therefore, at each iteration, $X_k^*(\cdot)$ remains a function of partial lotteries associated with frontier nodes.

Second, I claim that there exists a path in G between i and j that does not pass through the frontier F. By definition, k is a visited node where $X_k^*(\cdot)$ depends on both X_i and X_j. Since k is visited, $k \notin F$. Therefore, it suffices to show that there are paths in G between i and k, as well as k and j. The argument is the same in each case.

I claim there exists a path between i and k that does not pass through F. Obsevre that, in order for $X_k^*(\cdot)$ to depend on X_i, it must have been defined (or redefined) in step 6 of a previous iteration of algorithm 2. Let h be the node visited during that previous iteration. There are two cases.

1. If $h = k$, then X_k^* is being defined for the first time. In order for $X_k^*(\cdot)$ to depend on X_i, it must be the case that $i \in S_k \cup I'_k$. By definition of S_k, if $i \in S_k$ then i and j share an edge in G. This means there is a path in G between i and k that does not pass through F, vacuously, because it does not have any interior nodes. Alternatively, suppose $i \in I'_k$. By definition of I'_k, there is a path from i to k that does not pass through F.

2. Suppose $h \neq k$. Before $X_k^*(\cdot)$ depended on X_i, it depended on X_h. Then X_h was replaced with X_h^*, which depended on X_i. Since the algorithm visits h in an earlier iteration, it is no longer in the frontier F by the time the algorithm visits i. Therefore, there is a path from $k \notin F$ to $h \notin F$, and it suffices to find a path from h to i that does not pass through F.

In the second case, I can repeat this argument with node h taking the role of node k. There are only $n < \infty$ nodes, so eventually it will be the case that $h = k$.

This completes the argument. I have shown that $j \in F$ and there exists a path in G between i and j that does not pass through the frontier F. Therefore, $j \in I'_i$. This is what I sought to show.

A.11 Proof of Lemma 11

Consider step 5 of algorithm 2. In the iteration where node i is visited, the algorithm defines

$$X_i^*\left(X_{S_i}, X_{I_i}\right) \in \arg\max_{X_i \in M_i} \mathrm{E}\big[u\left(X_i, X_{S_i}, X_{P_i}^*\left(X_i, X_{I_i}\right), 0, 0, \ldots\right)\big]$$

To prove this result, it is enough to show that $X_i^*(\cdot)$ is consistent with expected utility maximization in the following sense. If the decisionmaker is constrained to lotteries $X' \in M$ where $X'_{S_i} = X_{S_i}$ and $X'_{I_i} = X_{I_i}$, her optimal choice X' should satisfy $X'_i = X_i^*(X_{S_i}, X_{I_i})$. If that holds, the optimality of algorithm 1 follows from the optimality of dynamic programming.

I begin by establishing a useful property. Suppose that the (undirected) inseparability graph $G := G_n(u)$ has an edge between nodes i and j. I claim that $j \in S_i \cup P_i$. To prove this claim, there are three cases to consider.

1. The algorithm has not yet visited node j. Then $j \in S_i$ is a successor of i, by definition.

2. The algorithm has already visited node j and there is an edge in $\vec{G}$ from j to i. Then $i \in S_j$ is a successor of j, by definition. Since $X_j^*(\cdot)$ depends on X_{S_j}, it depends on X_i. Therefore, $j \in P_i$ is a predecessor of i.

3. The algorithm has already visited node j and there is an edge in $\vec{G}$ from i to j. This case is somewhat more involved. First, note that $i \leq j$. This follows from the fact that there is an edge from i to j implies that i precedes j in the topological order of step 2.

   Next, consider the iteration of step 5 that visits node j. Step 4 visits i before j, but i has not been visited yet, so it must be the case that step 5d skipped over i. This only occurs when there are too many indirect influencers of i, i.e. $|I_i| > d$.

   I claim that $i \in F$. In that case, $i \in I_j$ because there is a path in G between i and j that does not pass through F. This path simply consists of the edge (i, j). Since $X_j^*(\cdot)$

depends on X_{I_j}, it depends on X_i. Therefore, $j \in P_i$ is a predecessor of i.

Suppose for contradiction that $i \notin F$. Let node $k \in I_i$ be an indirect influencer of i. By definition, there is a path in G from k to i that does not pass through F. This path can be extended to j by passing through i. Since $i \notin F$, the extended path does not pass through F. Therefore, $k \in I_j$ is an indirect influencer of j. This implies $|I_j| \geq |I_j| > d$. That contradicts the fact that node j is being visited, since step 5d would skip over it.

It follows from these three cases that $j \in S_i \cup P_i$.

By the preceding argument, any node $j \notin S_i \cup P_i$ must not share an edge with i in the inseparability graph G. By definition of the inseparability graph, this means that u is (i, j, n)-separable. Applying the definition of separability for each $j \notin S_i \cup P_i$, I can represent the utility function as

$$u(x) = u_i\left(x_i, x_{P_i}, x_{S_i}\right) + u_{-i}\left(x_{-i}\right)$$

For the purpose of maximizing expected utility, the function u_{-i} is irrelevant. Therefore, setting $X_j = 0$ for $j \notin S_i \cup P_i$ is without loss of optimality.

A.12 Proof of Lemma 12

To establish the runtime, I analyze each step of the algorithm.

1. Step 1 can be done in $O(\text{poly}(n))$ time, by lemma 8.

2. Step 2 can be done in $O(n^2)$ time, using standard algorithms for topological sorting.

3. Step 3 can be done in $O(1)$ time.

4. Step 4 can be done in $O(n)$ time.

5. Step 5 can be done in $O(n^4)$ time.

   (a) Step 5a can be done in $O(n^2)$ time by searching through all edges.

   (b) Step 5b can be done in $O(n^2)$ time by searching through all edges.

   (c) Step 5c can be done in $O(n^3)$ time. This involves checking up to n nodes $j \in F$. For each j, I need to evaluate whether there exists a path in G between i and j. This can be done in $O(n^2)$ time by breadth-first search.

   (d) Step 5d can be done in $O(n)$ time by searching through the set I_i of indirect influencers. This repeats step 5 at most n times before either (i) moving on to step 6 or (ii) reaching an error. Lemma 9 guarantees that it will not reach an error.

6. Step 6 has two parts.

   (a) First, it runs step 5 of algorithm 1. This step involves an optimization problem. Since u is efficiently computable and the sample space is split into m intervals, evaluating expected utility for a given lottery takes $O(m \cdot \text{poly}(n))$ time. For each $X_{S_i \cup I_i} \in M_{S_i \cup I_i}$, I consider up to k alternative partial lotteries $X_i \in M_i$. I claim that the set $M_{S_i \cup I_i}$ has up to k^{2d} elements, which implies that step 5 takes $O(k^{2d+1} m \cdot \text{poly}(n))$ time.

   To show that $M_{S_i \cup I_i}$ has no more than k^{2d} elements, it suffices to show that $|S_i \cup I_i| \leq 2d$. Step 5d ensures that $I_i \leq d$. The successors S_i can be split into two parts. The first part consists of unvisited nodes j where $\vec{G}$ contains an edge from i to j. There are at most d nodes of this kind, by step 1. The second part of S_i consists of unvisited nodes j where $\vec{G}$ contains an edge from j to i. Let $i' := j$ and $j' := i$. Restated, node j' is visited before node i' and $\vec{G}$ contains an edge from i' to j'. In bullet 3 of the proof of lemma 11, I showed that this implies $i' \in I_{j'}$.

Restated in my original notation, $j \in I_i$. Therefore, these nodes j were already counted among the d nodes in I_i. To summarize, there are at most $2d$ nodes in $S_i \cup I_i$.

(b) Second, it runs step 6 of algorithm 1. This iterates over $O(n)$ predecessors $j \in P_i$. For each j, it needs to redefine X_j^* for up to k^{2d} elements of $M_{S_i \cup I_i}$. Each redefinition can be done in $O(k^{d+1})$ time by looking up the values of X_j^* for different arguments X_i, X_{I_i}. Overall, this takes $O(nk^{3d+1})$ time.

7. Step 7 can be done in $O(n)$ time. It returns to step 4 at most n times.

8. The output can be described in $O(nm)$ space.

Combining all these steps yields a runtime that satisfies the bound (1.12).

A.13 Proof of Proposition 4

First, I show that when the utility function u is Hadwiger separable, maximizing expected utility is consistent with relatively narrow dynamic choice bracketing. Lemma 10 implies that algorithm 2 is a special case of algorithm 1. Lemma 11 implies that algorithm 2 maximizes expected utility. Let $d_n = \text{cdgn}(G_n(u))$. I showed in the proof of Lemma 12 (bullet 6a) that, for each node i in $G_n(u)$, $S_i \cup I_i$ has no more than $2d_n$ elements. I showed in lemma 7 that $d_n = O(\log n)$ if u is Hadwiger separable. Therefore, algorithm 2 is dynamic choice bracketing with bracket size $2d_n = O(\log n)$. By definition, this is relatively narrow.

Next, I show that if relatively narrow dynamic choice bracketing is rational, then it maximizes expected utility with respect to a Hadwiger separable utility function. I assumed the NU-ETH for this result in order to use Theorem 2. Note that algorithm 1 can be solved in polynomial time with polynomial-size advice, where the advice includes the order in which nodes are visited and the set of successors. This follows from an argument similar to Lemma

12. Therefore, Theorem 2 implies that the revealed utility function u must be Hadwiger separable. I emphasize that this argument is not circular because the proof of Theorem 2 did not make use of Proposition 4. Finally, note that the NU-ETH is sufficient but may not be necessary. It may be possible to prove this result directly without invoking Theorem 2.

A.14 Proof of Lemma 14

Let X^* be the output of the greedy algorithm on product menu M. Consider a decisionmaker who runs the greedy algorithm for $i-1$ iteratons, choosing $X_1^*, \ldots, X_{i-1}^*$, but chooses the remaining lotteries $X_i, \ldots, X_n$ optimally. Formally, define

$$\mathrm{OPT}_i := \max_{X_{>i} \in M_{>i}} \mathrm{E}[\bar{u}\,(X_1^*, \ldots, X_i^*, X_{i+1}, \ldots, X_n, 0, 0, \ldots)]$$

Observe that

$$\mathrm{OPT}_0 = \max_{X \in M} \mathrm{E}[\bar{u}\,(X_1, \ldots, X_n, 0, 0, \ldots)]$$

is simply expected utility maximization, whereas

$$\mathrm{OPT}_n = \mathrm{E}[\bar{u}\,(X_1^*, \ldots, X_n^*, 0, 0, \ldots)]$$

is the expected utility obtained by the greedy algorithm. The goal is to show that

$$2 \cdot \mathrm{OPT}_n \geq \mathrm{OPT}_0 \tag{A.30}$$

Next, consider the added value from choosing X_i^* in iteration i of the greedy algorithm. This corresponds to the expected value of a random variable $\Delta_i : \Omega \to \mathbb{R}$, where

$$\Delta_i := \bar{u}\left(X_1^*, \ldots, X_{i-1}^*, X_i^*, 0, 0, \ldots\right)$$

$$- \bar{u}\left(X_1^*, \ldots, X_{i-1}^*, 0, 0, 0, \ldots\right)$$

Since this is the added value of the greedy algorithm in each iteration, we have

$$\mathrm{OPT}_n = \sum_{i=1}^{n} \mathrm{E}[\Delta_i] \tag{A.31}$$

I claim that the added value $\mathrm{E}[\Delta_i]$ exceeds the lost value from a simple deviation from the optimal solution to OPT_{i-1}, where one chooses X_i^* instead of the optimal X_i. Formally,

$$\begin{aligned}
\Delta_i &\geq \bar{u}\left(X_1^*, \ldots, X_{i-1}^*, X_i, 0, 0, \ldots\right) \\
&\quad - \bar{u}\left(X_1^*, \ldots, X_{i-1}^*, 0, 0, 0, \ldots\right) \\
&\geq \bar{u}\left(X_1^*, \ldots, X_{i-1}^*, X_i, X_{i+1} \ldots, X_n, 0, 0, \ldots\right) \\
&\quad - \bar{u}\left(X_1^*, \ldots, X_{i-1}^*, 0, X_{i+1}, \ldots, X_n, 0, 0, \ldots\right) \\
&\geq \bar{u}\left(X_1^*, \ldots, X_{i-1}^*, X_i, X_{i+1} \ldots, X_n, 0, 0, \ldots\right) \\
&\quad - \bar{u}\left(X_1^*, \ldots, X_{i-1}^*, X_i^*, X_{i+1}, \ldots, X_n, 0, 0, \ldots\right)
\end{aligned} \tag{A.32}$$

This holds for any partial lotteries $X_{i+1}, \ldots, X_n$. The first inequality follows from construction of the greedy algorithm. The second inequality follows from the diminishing returns, where the analog to x'' in definition 25 is

$$x'' = \left(0, \ldots, 0, 0, X_{i+1}^*(x), \ldots, X_n^*(x), 0, 0 \ldots\right)$$

The third inequality follows from the fact that $\bar{u}$ is non-decreasing, since $X_i^* \geq 0$.

It follows from inequality (A.32) that

$$\mathrm{OPT}_i \leq \mathrm{E}[\Delta_{i+1}] + \mathrm{E}\left[\bar{u}\left(X_1^*, \ldots, X_{i-1}^*, X_i^*, X_{i+1}, \ldots, X_n, 0, 0, \ldots\right)\right]$$

when $X_{i+1}, \ldots, X_n$ are defined as the arguments that obtain OPT_i. By definition, OPT_{i+1} is an upper bound for the second term of the right-hand side. Therefore,

$$\mathrm{OPT}_i \leq \mathrm{E}[\Delta_{i+1}] + \mathrm{OPT}_{i+1}$$

Apply this inequality recursively to show that

$$\begin{aligned} \mathrm{OPT}_0 &\leq \mathrm{OPT}_n + \sum_{i=1}^{n} \mathrm{E}[\Delta_i] \\ &= 2 \cdot \mathrm{OPT}_n \end{aligned}$$

where the second line follows from equation (A.31). This establishes inequality (A.30).

A.15 Randomized Approximation Algorithm

This approximation algorithm follows a heuristic called randomized rounding. It begins by setting up a mixed integer programming formulation of expected utility maximization for the maximum utility function $u(x) = \max_i x_i$. It solves a linear programming relaxation in polynomial time. Then, as needed, it randomly rounds the real-valued solution to the linear programming relaxation to a nearby integer-valued solution to the original mixed integer programming problem. A similar algorithm has been used to obtain a $(1 - 1/e)$-approximation to MAX 2-SAT, and I show that it can obtain the same approximation for this particular expected utility maximization problem.

I begin by addressing the use of randomization. I have modeled the decisionmaker as a Turing machine that makes deterministic choices. Alternatively, I could have modeled her as a probabilistic Turing machine that makes stochastic choices. Of course, stochastic choice would violate the assumption that the agent always maximizes expected utility for some utility function u, except in the trivial case where she randomizes over choices that she is indifferent to. However, it is natural to ask whether the decisionmaker can efficiently generate stochastic choices that match a deterministic choice correspondence c with high probability. This question can be addressed by modeling the decisionmaker as a probabilistic Turing machine.

I can formulate a probalistic relaxation of tractability as follows. The complexity class BPP refers to decision problems that can be solved with bounded error in polynomial time. It consists of problems in which there exists a polynomial-time Turing machine that gives the correct answer with probability greater than or equal to 2/3. Given a choice correspondence c, consider a decision problem D_c that asks whether a given lottery $X \in M$ is chosen, i.e. $X \in c(M)$. A choice correspondence c is tractable in a probabilistic sense if $D_c \in \text{BPP}$.

I claim that my results do not change if I relax tractability in this way, as long as $\text{NP} \not\subset \text{P/poly}$. Keep in mind that I already assume this to prove 4, and assume a stronger conjecture to prove Theorem 2. To prove the claim, note that Adleman's theorem implies $\text{BPP} \subseteq \text{P/poly}$ (Adleman 1978, Bennett and Gill 1981). However, I have already argued in Corollary 1 of Theorem 2 that for utility functions u where

$$\text{Had}(G_n(u)) = \Omega(\text{poly}(n))$$

expected utility maximization does not belong to P/poly as long as $\text{NP} \not\subset \text{P/poly}$. Therefore, it cannot belong to BPP. If I assume the NU-ETH hypothesis, Theorem 2 implies the somewhat stronger result that expected utility maximization belongs to P/poly as long as u is Hadwiger separable. Again, this means that it cannot belong to BPP.

Having shown that my results are robust to the use of probabilistic Turing machines, I can now turn to approximation algorithms that use randomization. First, I want to construct a mixed integer programming formulation of

$$\max_{X \in M} \mathrm{E}_\omega[\max\{X_1(\omega), \ldots, X_n(\omega)\}]$$

By assumption 1, I can restrict attention to a finite number of representative points $\omega \in \Omega$ in the sample space. Without loss of generality, assume they occur with equal probability. The number of these points is polynomial in the description of the product menu M. For convenience, whenever I refer to ω in this proof, let ω denote a point in this finite set.

Let X_i^j denote the jth element of M_i. Formally, define the mixed integer programming formulation as

$$\max \sum_\omega u_\omega \quad \text{subject to}$$

$$d_{i,\omega} \in \{0,1\}, \quad \forall i, \omega$$

$$\sum_{i=1}^{n} d_{i,\omega} = 1, \quad \forall \omega$$

$$p_{i,j} \in \{0,1\}, \quad \forall i, j$$

$$\sum_{j=1}^{|M_i|} p_{i,j} = 1, \quad \forall i$$

$$u_\omega \leq \sum_{i=1}^{n} d_{i,\omega} \sum_{j=1}^{|M_i|} p_{i,j} \cdot X_i^j(\omega)$$

Consider the following linear programming relaxation:

$$\max \sum_\omega u_\omega \quad \text{subject to}$$

$$d_{i,\omega} \in [0,1], \quad \forall i, \omega$$

$$\sum_{i=1}^{n} d_{i,\omega} = 1, \quad \forall \omega$$

$$p_{i,j} \in [0,1], \quad \forall i, j$$

$$\sum_{j=1}^{|M_i|} p_{i,j} = 1, \quad \forall i$$

$$u_\omega \leq \sum_{i=1}^{n} d_{i,\omega} \sum_{j=1}^{|M_i|} p_{i,j} \cdot X_i^j(\omega)$$

This can be solved in polynomial-time because the number of variables and constraints is polynomial in the description length of the menu M.

Let $u^*_\omega, d^*_{i,\omega}, p^*_{i,j}$ be the solution to the linear programming problem. The randomized rounding algorithm chooses X_i^j with probability $p_{i,j}$.

Next, I show that the randomized rounding algorithm obtains a constant approximation to the mixed integer programming problem. First, observe that

$$\begin{aligned}
\Pr_{p^*}\left[\max_i X_i(\omega) \leq t\right] &= \prod_{i=1}^{n} \Pr_{p^*}[X_i(\omega) \leq t] \\
&\leq \left[\frac{1}{n}\sum_{i=1}^{n} \Pr_{p^*}[X_i(\omega) \leq t]\right]^n \\
&= \left[1 - \frac{1}{n}\sum_{i=1}^{n} \Pr_{p^*}[X_i(\omega) > t]\right]^n \\
&\leq \left[1 - \frac{1}{n}\max_i \Pr_{p^*}[X_i(\omega) > t]\right]^n
\end{aligned}$$

where the last inequality follows from the fact that

$$\sum_{i=1}^{n} \Pr_{p^*}[X_i(\omega) > t] \geq \max_i \Pr_{p^*}[X_i(\omega) > t]$$

Next, observe that

$$\Pr_{p^*}\left[\max_i X_i(\omega) > t\right] \geq 1 - \left[1 - \frac{1}{n}\max_i \Pr_{p^*}[X_i(\omega) > t]\right]^n$$

$$
\begin{aligned}
&\geq 1 - \left[1 - \frac{1}{n}\right]^n \max_i \mathrm{Pr}_{p^*}[X_i(\omega) > t] \\
&\geq \left(1 - \frac{1}{e}\right) \max_i \mathrm{Pr}_{p^*}[X_i(\omega) > t] \\
&\geq \left(1 - \frac{1}{e}\right) \sum_{i=1}^{n} d^*_{i,\omega} \mathrm{Pr}_{p^*}[X_i(\omega) > t]
\end{aligned}
$$

Finally, note that

$$\mathrm{E}_{p^*}\left[\max_i X_i(\omega)\right] = \int_0^1 \mathrm{Pr}_{p^*}\left[\max_i X_i(\omega) > t\right] dt$$

This is a well-known property of expectations. By linearity of integration and the inequality above, we have

$$
\begin{aligned}
\int_0^1 \mathrm{Pr}_{p^*}\left[\max_i X_i(\omega) > t\right] dt &\geq \left(1 - \frac{1}{e}\right) \sum_{i=1}^{n} d^*_{i,\omega} \int_0^1 \mathrm{Pr}_{p^*}[X_i(\omega) > t] dt \\
\mathrm{E}_{p^*}\left[\max_i X_i(\omega)\right] &\geq \left(1 - \frac{1}{e}\right) \sum_{i=1}^{n} d^*_{i,\omega} \mathrm{E}_{p^*}[X_i(\omega)] \\
&= \left(1 - \frac{1}{e}\right) \sum_{i=1}^{n} d^*_{i,\omega} \sum_{j=1}^{|M_i|} p^*_{i,j} X_i^j(\omega) \\
&\geq \left(1 - \frac{1}{e}\right) u^*_\omega
\end{aligned}
$$

Finally, sum over the states ω to obtain

$$\mathrm{E}_{\omega,p^*}\left[\max_i X_i(\omega)\right] \geq \left(1 - \frac{1}{e}\right) \sum_\omega u^*_\omega$$

The right-hand side is the optimum of the linear programming relaxation. Since it is a relaxation, this implies it is an upper bound for the optimum of the mixed integer programming problem, which is equal to the maximum expected utility. Therefore, this inequality implies that randomized rounding gives a $(1 - {}^1\!/\!{}_e)$-approximation.

Appendix B

Omissions from Chapter 2

B.1 Special Cases

B.1.1 Bayesian Persuasion

There is an informed sender (i.e. principal) and an uninformed receiver (i.e. agent). The principal designs the process by which information is revealed to the agent. Let $\mathcal{M}$ be a finite set of messages that he can send. Knowing that the agent will react to an informative message, the principal attempts to persuade the agent towards actions that he prefers. Let $\mathcal{A}$ be a finite set of actions that the agent can take. The agent chooses a response $r : \mathcal{M} \to \mathcal{A}$ that maps messages to actions.

A policy is an information structure $p : \mathcal{Y} \to \Delta(\mathcal{M})$. That is, an information structure $p_t(y_t)$ describes the probability of a message m_t being sent, conditional on the state being y_t. The agent receives the message m_t and takes action $a_t = r_t(m_t)$. While the agent may not know the process that generated the state y_t, she understands the process p_t that generates the message m_t conditional on the state. Armed with this understanding, she can infer something about the state y_t based on the message m_t.

All that remains is to specify payoffs. Let $u : \mathcal{A} \times \mathcal{Y} \to \mathbb{R}$ be the agent's utility function from a given action in a given state. Similarly, let $v : \mathcal{A} \times \mathcal{Y} \to \mathbb{R}$ be the principal's utility. In the previous subsection, the utility functions U, V depended on the triple (r, p, y) rather than the pair (a, y). To reconcile the two models, we let participants evaluate (r, p, y) by their expected utility conditional on the state. Formally,

$$U(r, p, y) = \sum_{m \in \mathcal{M}} p(m, y) \cdot u(r(m), y) \quad \text{and} \quad V(r, p, y) = \sum_{m \in \mathcal{M}} p(m, y) \cdot v(r(m), y)$$

When the state is fixed, the residual variation in utility is due to the fact that messages are drawn randomly from the distribution $p(y)$. These distributions are common knowledge because the agent observes the principal's policy p before taking an action. Indeed, the fact that the principal commits to an information structure is the defining feature of Bayesian persuasion.

Example 6. [Judge-Prosecutor Game] The state space is $\mathcal{Y} = \{\text{Innocent}, \text{Guilty}\}$ and the action space is $\mathcal{A} = \{\text{Convict}, \text{Acquit}\}$. The judge has 0-1 utility u and prefers to convict if the defendant is guilty and acquit if the defendant is innocent. Regardless of the state, the prosecutor's utility v is 1 following a conviction and 0 following an acquittal.

This example satisfies regularity (7) with the discrete metric on $\mathcal{R}$, the l_1-metric on $\mathcal{P}$, and $K^U_{\mathcal{R}} = K^V_{\mathcal{R}} = K^U_{\mathcal{P}} = K^V_{\mathcal{P}} = 1$.

The worst-case policy $p^*(\pi, \epsilon)$ sends the message "convict" whenever the defendant is guilty. If the defendant is innocent, it sends the message "convict" with probability

$$q = \max\left\{1, \min\left\{0, \frac{p - \epsilon}{1 - p}\right\}\right\}$$

The cost of ϵ-robustness $\Delta(\pi, \epsilon) = O(\epsilon)$ decreases smoothly with ϵ. This game satisfies assumption **??** since q is increasing in p (and hence convex combinations of distributions p will yield q that is bounded between the ϵ-robust policies for the extremal distributions, which are close by assumption).

The worst-case policy $p^\dagger(\pi, \epsilon)$ for an unknown private signal is full transparency. The cost of informational robustness, i.e. $\nabla(\pi, 0)$, is the difference between the principal's value under the common prior π and his payoff under full transparency. This game satisfies assumption **??** with $M_1 = 1$ and $M_2 = O(\epsilon)$. It trivially satisfies assumptions **??** and **??** since $p^\dagger$ is constant.

B.1.2 Contract Design

In classic models of moral hazard, the principal incentivizes an agent to put effort into a task the principal cares about. The timing of the game is as follows: (1) the principal commits to a contract, (2) the agent takes a hidden action, (3) nature randomly chooses an outcome, (4) the agent is paid based on the outcome, (5) the game concludes. For concreteness, we consider the limited liability model due to Sappington (1983) where both participants are risk-neutral but the principal is not allowed to charge the agent. This model has been popularized by recent work in robust contract design (see e.g. Carroll 2015, Dütting et al. 2019).

Formally, let $\mathcal{R}$ be a finite set of actions that the agent can take. Let $\mathcal{O}$ be a finite set of outcomes o. The principal observes the outcome but not the action. The state $y : \mathcal{R} \to \mathcal{O}$ describes how actions map to outcomes. The employer commits to a contract $p : \mathcal{O} \to [0, \bar{p}]$ that specifies a non-negative payment for each outcome. The cost function $c : \mathcal{R} \to \mathbb{R}$ describes how costly it is for the agent to take a particular action. The agent's utility function is

$$U(r, p, y) = p(y(r)) - c(r)$$

The benefit function $b : \mathcal{O} \to \mathbb{R}$ describes how beneficial a given outcome is to the principal. The principal's utility function is

$$V(r, p, y) = b(y(r)) - p(y(r))$$

Through the contract p, the principal can incentivize the agent to take actions that, depending on the state, will lead to a more beneficial outcome.

Example 7. The agent is given a task of unknown difficulty. There are two actions $\mathcal{A} = \{\text{work}, \text{shirk}\}$, two outcomes $\mathcal{O} = \{\text{success}, \text{failure}\}$, and three states $\mathcal{Y} = \{\text{trivial}, \text{moderate}, \text{impossible}\}$. In the trivial state, both actions lead to success. In the impossible state, both actions lead to failure. In the moderate state, work leads to success and shirk leads to failure.

The principal's benefits are $b(\text{success}) = 2$ and $b(\text{failure}) = 0$. The agent's costs are $c(\text{work}) = 1$ and $c(\text{shirk}) = 0$. In the impossible and trivial states, the optimal contract pays nothing after both outcpmes and the agent will shirk. In the moderate state, the optimal contract pays $p(\text{success}) = 5$ to cover the agent's costs if she works, otherwise $p(\text{failure}) = 0$. Generally, if the principal pays the agent after success, the agent will have to take into account the risk that the task turns out to be impossible (where work induces costs without any payment) or trivial (where work is not required for payment). To incentivize work, the contract must compensate the agent accordingly.

This example satisfies regularity (7) with U, V normalized, the discrete metric on $\mathcal{R}$, the sup-norm-metric on $\mathcal{P}$, and $K^U_{\mathcal{R}} = K^V_{\mathcal{R}} = K^U_{\mathcal{P}} = K^V_{\mathcal{P}} = 1$.

The worst-case policy $p^*(\pi, \epsilon)$ sets $p(\text{failure}) = 0$ and

$$p(\text{success}) = \frac{c(\text{work}) - c(\text{shirk}) + \epsilon}{\pi(\text{moderate})}$$

so long as $p(\text{success}) \leq \bar{p}$ and the principal's π-expected payoff is greater than zero when the agent works. Otherwise, the worst-case policy sets all transfers to zero. The cost of ϵ-robustness $\Delta(\pi, \epsilon) = O(\epsilon)$ decreases smoothly with ϵ.

This game satisfies assumption **??**. To see this, note that as long as working is strictly more costly than shirking, the optimal policies that induce effort are bounded away from the optimal policies that do not. Among the policies that do not induce effort, convex combinations of the distribution will not make inducing effort desirable. Among policies that do induce effort, the fact that the payments following success are decreasing in π(moderate) means (as in the last section) that convex combinations of distributions lead to optimal policies that are between the extremal policies.

The worst-case policy $p^{\dagger}(\pi, \epsilon)$ for an unknown private signal is the same as the optimal policy under a common prior without a private signal. The cost of informational robustness, i.e. $\nabla(\pi, 0)$, is the difference between the principal's value when the agent only works in the "moderate" state and the principal pays her cost of effort conditional on success (assuming the principal prefers this to shirking with zero transfers) and his value in the common prior game without a private signal. This game satisfies assumption **??** with $M_1 = O(\epsilon)$ and $M_2 = 0$.

B.2 Agent's Learning Problem

Upper bounds on external regret are often viewed as compelling assumptions (e.g. Nekipelov et al. 2015, Braverman et al. 2018) because there exist relatively simple algorithms that guarantee vanishing ER as $T \to \infty$. For example, the exponential weights algorithm (a.k.a. hedge algorithm, exponentiated gradient algorithm) satisfies no-ER. In contrast, our behavioral assumptions – e.g. no-FCIR – may appear daunting, insofar as the agent must solve a learning problem with a context space that is exponential in the number of alternative mechanisms, $|\Sigma_0|$. When both the sequence of states $y_{1:T}$ and the learner L are particularly pathological, no-FCIR may indeed be too strong an assumption. When the the learner satisfies additional properties, or sequence of states has some stochastic structure (e.g. is i.i.d. or Markov), no-FCIR may be more reasonable.

In this section, we make one simple observation. There exists a learner that guarantees no-FCIR (and hence no-CIR) for the agent under our mechanism from theorem 6, with the best rate of convergence we can hope for.

Definition 52 (CFL)**.** *Suppose the principal publicizes the forecast π_t in every period t.*[1] *The* common forecast learner *(CFL) sets*

$$r_t \in \arg\max_{r \in \mathcal{R}} \mathrm{E}_{y \sim \pi_t}[U(r_t, p_t, y_t)]$$

Proposition 11 verifies that the CFL satisfies the behavioral assumptions of theorem 6.

Proposition 11. *Let σ^* be the mechanism from theorem 6. Then the CFL satisfies ϵ-bounded FCIR* (10) *in expectation, i.e.*

$$\mathrm{E}_{L,\sigma^*}[\mathrm{FCIR}] \leq \epsilon = \tilde{O}\left(\frac{1}{T^{1/4}}\sqrt{|\mathcal{Y}||\mathcal{F}|}\right) + O\left(\sqrt{|\mathcal{Y}|\delta_{\mathcal{F}}}\right)$$

and $\tilde{\epsilon}$-lower-bounded FER (11)*, where $\tilde{\epsilon} = 0$. Moreover, if the agent uses CFL, the principal's regret bound in theorem 6 applies regardless of whether alignment* (20) *holds.*

Note that these rates preserve the $T^{1/4}$ convergence rate (up to $\delta_{\mathcal{F}}$ error) that is present in all of our mechanisms and reflects miscalibration of the principal. In that sense, the fact that the agent is also learning does not deteriorate the principal's performance at all. Although this has not been our emphasis so far, it would be interesting to see whether (or identify conditions under which) other simple learning algorithms satisfy our behavioral assumptions with decent rates of convergence.

B.3 Calibrated Forecasting

In this appendix, we describe our forecasting algorithm and bound its miscalibration.

[1] In our view, part of the principal's objective is to make the agent's problem as simple as possible. From a worst-case perspective, there is no benefit to hiding this information. With that said, we see no reason why this result should not apply under the weaker assumption that the principal's forecasting algorithm is public knowledge.

A linearly homogeneous, differentiable function H is *strongly convex* with parameter ξ if

$$H(\pi) \geq H(\tilde{\pi}) + \nabla H(\tilde{\pi}) \cdot (\pi - \tilde{\pi}) + \frac{\xi}{2}\|\pi - \tilde{\pi}\|_2^2$$

The gradient of H describes a *proper scoring rule* $S(\pi) = \nabla H(\pi)$ where $H(\pi) = \pi \cdot S(\pi)$ (McCarthy 1956). A scoring rule $S : \Delta(\mathcal{Y}) \to \mathbb{R}^{\mathcal{Y}}$ is proper if the report $\tilde{\pi}$ that maximizes the π-expected score is the distribution π. Strong convexity of H can be thought of as sharpening the incentives for truth-telling (Boutilier 2012).

Specifically, consider the quadratic scoring rule (see e.g. Jose et al. 2008)

$$S_y(\pi) = 2\pi(y) - \sum_{\tilde{y} \in \mathcal{Y}} \pi(\tilde{y})^2$$

where $H(\pi) = \|\pi\|_2^2$ is strongly convex with $\xi = 2$.

Recall that the mechanism σ^* is supposed to be nonresponsive. As a consequence, we cannot determine the principal's beliefs π_t in a given period based on his historical payoffs. To ensure that the beliefs π_t are well-calibrated, we consider an auxilliary online learning problem based on a scoring rule S. In period t, the principal makes a prediction π_t with loss function $S_{y_t}(\pi_t)$. Specifically, the predictions come from the discretized set of priors $\mathcal{F}_1$, formed by choosing a representative element π from each set in the partition $\mathcal{S}_{\mathcal{F}}$. In terms of the score, this approximation has limited cost. Let $\pi = [\hat{\pi}_F]_{\mathcal{F}_1}$ be the belief $\pi \in \mathcal{F}_1$ that is closest to the empirical distribution $\hat{\pi}_F$. Then

$$\begin{aligned} S_y(\hat{\pi}_F) - S_y(\pi) &\leq S_y(\hat{\pi}_F) - S_y(\hat{\pi}_F - \delta_{\mathcal{F}}) \\ &= 2\hat{\pi}_F(y) - \sum_{\tilde{y} \in \mathcal{Y}} \hat{\pi}_F(\tilde{y})^2 - 2(\hat{\pi}_F(y) - \delta_{\mathcal{F}}) + \sum_{\tilde{y} \in \mathcal{Y}} (\hat{\pi}_F(\tilde{y}) - \delta_{\mathcal{F}})^2 \\ &= 2\delta_{\mathcal{F}} - \sum_{\tilde{y} \in \mathcal{Y}} \hat{\pi}_F(\tilde{y})^2 + \sum_{\tilde{y} \in \mathcal{Y}} (\hat{\pi}_F(\tilde{y}) - \delta_{\mathcal{F}})^2 \end{aligned}$$

$$\leq 2\delta_{\mathcal{F}} \tag{B.1}$$

where $\hat{\pi}_F - \delta_{\mathcal{F}}$ is shorthand notation for the vector $(\hat{\pi}_F(y) - \delta_{\mathcal{F}})_{y \in \mathcal{Y}}$.

In this auxilliary problem, the exponential weights algorithm (see e.g. Cesa-Bianchi and Lugosi 2006) obtains expected external regret at most

$$\sqrt{2T \log |S_{\mathcal{F}}|}$$

relative to the best-in-hindsight $\pi_F^* \in \mathcal{F}_1$. A reduction due to Blum and Mansour (2007) (theorem 5) translates this into a bound on expected internal regret of

$$|S_{\mathcal{F}}|\sqrt{2T \log |S_{\mathcal{F}}|}$$

relative to the best-in-hindsight $\pi_F^* \in \mathcal{F}_1$. Combine this with the maximum approximation error (B.1) to bound the expected internal regret relative to the best-in-hindsight contextual belief $\pi \in \Delta(\mathcal{Y})$, which must be the empirical distribution $\hat{\pi}_F$ since S is proper. Specifically,

$$|S_{\mathcal{F}}|\sqrt{2T \log |S_{\mathcal{F}}|} + 2T\delta_{\mathcal{F}} \geq \mathrm{E}\left[\sum_{\pi \in \mathcal{F}_1} n_F \hat{\pi}_F \cdot (S(\hat{\pi}_F) - S(\pi))\right] \tag{B.2}$$

where n_F is the number of periods t where $[\pi_t]_{\mathcal{F}_1} = F_t$. This is a statement about the expected scoring loss, where the expectation reflects randomization in the algorithm. Our next result, lemma 25, translates this into a statement about the l_1 distance between the principal's belief π and the empirical distribution $\hat{\pi}_F$.

Lemma 25. *Let S be a proper scoring rule where the optimal expected score H is ξ-strongly convex. Then*

$$\sqrt{\frac{2|\mathcal{Y}|\kappa}{\xi}} \geq \frac{1}{T} \sum_{\pi \in \mathcal{F}_1} n_F d_1(\pi, \hat{\pi}_F)$$

where

$$\kappa = \frac{1}{T} \sum_{\pi \in \mathcal{F}_1} n_F \hat{\pi}_F \cdot (S(\hat{\pi}_F) - S(\pi))$$

Proof. Consider the principal's π-expected regret from predicting $\tilde{\pi}$:

$$\begin{aligned}
\pi \cdot (S(\pi) - S(\tilde{\pi})) &= H(\pi) - \pi \cdot \nabla H(\tilde{\pi}) \\
&\geq H(\tilde{\pi}) - \nabla H(\tilde{\pi}) \cdot \tilde{\pi} + \frac{\xi}{2} \|\pi - \tilde{\pi}\|_2^2 \\
&= \frac{\xi}{2} \|\pi - \tilde{\pi}\|_2^2 \\
&\geq \frac{\xi}{2} \left(\frac{1}{\sqrt{|\mathcal{Y}|}} \|\pi - \tilde{\pi}\|_1 \right)^2 \\
&= \frac{\xi}{2|\mathcal{Y}|} \|\pi - \tilde{\pi}\|_1^2
\end{aligned}$$

where the second-to-last line follows from $\| \cdot \|_1 \leq |\mathcal{Y}|^{1/2} \| \cdot \|_2$. It follows that his regret in the auxilliary problem satisfies

$$\kappa \geq \frac{1}{T} \sum_{\pi \in \mathcal{F}_1} n_F \frac{\xi}{2|\mathcal{Y}|} d_1 \left(\pi, \hat{\pi}_F\right)^2$$

where, implicitly, F is the forecast context such that $\pi \in F$. Take the square root of both sides of this inequality:

$$\begin{aligned}
\sqrt{\kappa} &\geq \sqrt{\frac{1}{T} \sum_{\pi \in \mathcal{F}_1} n_F \frac{\xi}{2|\mathcal{Y}|} d_1 \left(\pi, \hat{\pi}_F\right)^2} \\
&\geq \frac{1}{\sqrt{T}} \cdot \frac{1}{\sqrt{T}} \sum_{\pi \in \mathcal{F}_1} n_F \sqrt{\frac{\xi}{2|\mathcal{Y}|}} d_1 \left(\pi, \hat{\pi}_F\right) \\
&\geq \sqrt{\frac{\xi}{2|\mathcal{Y}|}} \cdot \frac{1}{T} \sum_{\pi \in \mathcal{F}_1} n_F d_1 \left(\pi, \hat{\pi}_F\right)
\end{aligned}$$

where the first line is the l^2 norm of a vector with T entries, the second line is the l^1 norm of that same vector, and the inequality follows from $\|\cdot\|_1 \leq T^{1/2}\|\cdot\|_2$. Collapse these inequalities and rearrange terms to obtain the desired result. □

If we use the quadratic scoring rule, this lemma implies

$$\mathrm{E}\left[\frac{1}{T}\sum_{\pi\in\mathcal{F}_1} n_F d_1(\pi,\hat{\pi}_F)\right] \leq \sqrt{|\mathcal{Y}||S_{\mathcal{F}}|\sqrt{\frac{2\log|S_{\mathcal{F}}|}{T}} + 2|\mathcal{Y}|\delta_{\mathcal{F}}} \tag{B.3}$$

To optimize this bound, up to log factors, set $\delta_{\mathcal{F}} = \left(\frac{1}{T}\right)^{\frac{1}{2|\mathcal{Y}|+2}}$, assuming $|S_{\mathcal{F}}| = \left(\left(\frac{1}{\delta_{\mathcal{F}}}\right)^{|\mathcal{Y}|}\right)$.

B.4 Generalized Results

Upper bounds on CIR constitute our rationality assumptions for the agent. However, our results also rely on informational assumptions. Sections 2.4, 2.5, and 2.6 consider environments that differ primarily by how "informed" the agent appears, relative to the principal. In all three cases, however, we require the agent to be at least as informed as the principal. What is the principal's information? Recall that our mechanisms σ^* will be forecast mechanisms. A calibrated learning algorithm – which we specify later on – will produce a sequence of forecasts $\pi_1, \ldots, \pi_T$. It is possible that these forecasts will become correlated with the state, e.g. if there is a trend in the data. We do not rule this out; however, if our forecasts inadvertently pick up useful information, this information should be available to the agent as well (either implicitly or because we publish π_t along with p_t).

The notion of forecastwise regret (and forecastwise CIR) formalizes what we mean by the principal's "information" being available to the agent. The agent's benchmark includes the principal's forecast as additional context. Formally, define the forecast space $\mathcal{F} = \Delta(\mathcal{Y})$. Fix a small constant $\delta_{\mathcal{F}} > 0$ and consider a finite partition $S_{\mathcal{F}}$ of $\mathcal{F}$ where $\pi, \tilde{\pi} \in F \in S_{\mathcal{F}}$ implies $d_\infty(\pi, \tilde{\pi}) \leq \delta_{\mathcal{F}}$. Let $\mathcal{F}_1 \subseteq \mathcal{F}$ contain a single distribution $\pi \in F$ for every $F \in S_{\mathcal{F}}$.

Definition 53. *Let the information partition combine the forecast and CIR context, i.e.*

$$\mathcal{I} = S_{\mathcal{F}} \times (S_{\mathcal{R}})^{\Sigma}$$

and let the information $I_t \in \mathcal{I}$ in period t be the unique set that satisfies

$$(\pi_t, r_t^*, (r_t^p)_{p \in \mathcal{P}_0}) \in I_t$$

Definition 54 (FCIR)**.** *The agent's* forecastwise CIR *relative to a modification rule $h : \mathcal{I} \to \mathcal{R}$ is*

$$\text{FCIR}(h) = \frac{1}{T} \sum_{t=1}^{T} (U(h(I_t), p_t, y_t) - U(r_t, p_t, y_t))$$

The FCIR relative to the best-in-hindsight modification rule is $\text{FCIR} = \max_{h:\mathcal{I}\to\mathcal{R}} \text{FCIR}(h)$.

To state our assumption, we need to define a forecastwise version of ER, just as we defined a forecastwise version of CIR at the end of section 2.3. Let the forecast context $F_t \in S_{\mathcal{F}}$ in period t be the unique set that satisfies $\pi_t \in F_t$.

Definition 55 (FER)**.** *The agent's* forecastwise external regret *relative to a strategy $h : S_{\mathcal{F}} \to \mathcal{R}$ is*

$$\text{FER}(h) = \frac{1}{T} \sum_{t=1}^{T} (U(h(F_t), p_t, y_t) - U(r_t, p_t, y_t))$$

The FER relative to the best-in-hindsight strategy is $\text{FER} = \max_{h:\mathcal{F}\to\mathcal{R}} \text{FER}(h)$.

Theorem 10. *Assume regularity (7) and ϵ-bounded FCIR (10). There exists a nonresponsive mechanism σ^* parameterized by the agent's learner L and a constant $\bar{\epsilon} > 0$ such that*

1. *The principal's regret is bounded, i.e.*

$$\mathrm{E}_{\sigma^*}[\text{PR}] \leq \frac{1}{T} \sum_{I \in \mathcal{I}} n_I \Delta(\hat{\pi}_I, \bar{\epsilon})$$

$$+\frac{1}{\bar{\epsilon}}\left(O(\epsilon)+\tilde{O}\left(T^{-1/4}\sqrt{|\mathcal{Y}||S_{\mathcal{F}}||\mathcal{S}_R|^{(|\Sigma_0|+|S_{\mathcal{P}}|)/2}}\right)+O\left(\delta_{\mathcal{F}}^{1/2}\right)+O(\delta_{\mathcal{P}})\right)$$

Assumption 20 (Alignment)**.** *The stage game is* (ϵ, M_1, M_2)*-aligned if, for all signals* γ*,*

$$\underbrace{\left(\phi_{p^*(\pi,\epsilon)}(\pi,\epsilon)-\alpha_{p^*(\pi,\epsilon)}(\pi,\gamma,\epsilon)\right)}_{\textit{maximum downside of } \gamma \textit{ for the principal}}$$
$$\leq M_1 \underbrace{\max_{r,r_J\in\mathcal{R}} \mathrm{E}_{y\sim\pi}\left[\mathrm{E}_{J\sim\tilde{\gamma}(\cdot,y)}[U(r_J,p^*(\pi,\epsilon),y)]-U(r,p^*(\pi,\epsilon),y)\right]}_{\textit{usefulness of } \gamma \textit{ to the agent}}+M_2$$

and, for all policies $p\in\mathcal{P}_0$*,*

$$\underbrace{\left(\beta_p(\pi,\gamma,\epsilon)-\phi_{p^*(\pi,\epsilon)}(\pi,\epsilon)\right)}_{\textit{maximum upside of } \gamma \textit{ for the principal}} \leq M_1 \underbrace{\max_{r,r_J\in\mathcal{R}} \mathrm{E}_{y\sim\pi}\left[\mathrm{E}_{J\sim\tilde{\gamma}(\cdot,y)}[U(r_J,p,y)]-U(r,p,y)\right]}_{\textit{usefulness of } \gamma \textit{ to the agent}}+M_2$$

Theorem 11. *Assume regularity* (7)*,* ϵ*-bounded FCIR* (10)*,* $\tilde{\epsilon}$*-lower-bounded FER* (11)*, and* $(\bar{\epsilon}, M_1, M_2)$*-alignment* (20)*. There exists a nonresponsive mechanism* σ^* *parameterized by* $\bar{\epsilon}$ *such that*

1. *The principal's regret is bounded, i.e.*

$$\begin{aligned}\mathrm{E}_{\sigma^*}[\mathrm{PR}] \leq & \frac{1}{T}\sum_{F\in S_{\mathcal{F}}} n_F\Delta(\hat{\pi}_F,\bar{\epsilon})+\frac{1}{\bar{\epsilon}}\left(O(\epsilon)+\tilde{O}\left(T^{-1/4}\sqrt{|\mathcal{Y}||S_{\mathcal{F}}|}\right)+O\left(\sqrt{\delta_{\mathcal{F}}}\right)+O(\delta_{\mathcal{P}})\right)\\ & +O(\tilde{\epsilon})+M_1\left(O(\tilde{\epsilon})+O(\epsilon)+\tilde{O}\left(T^{-1/4}\sqrt{|\mathcal{Y}||S_{\mathcal{F}}|}\right)+O\left(\sqrt{\delta_{\mathcal{F}}}\right)+O(\delta_{\mathcal{P}})\right)\\ & +O(M_2)\end{aligned}$$

Theorem 12. *Assume regularity* (7) *and* ϵ*-bounded FCIR* (10)*. There exists a nonresponsive mechanism* σ^* *parameterized by a constant* $\bar{\epsilon}>0$ *such that*

1. *The principal's regret is bounded, i.e.*

$$\mathrm{E}_{\sigma^*}[\mathrm{PR}] \leq \frac{1}{T} \sum_{F \in S_{\mathcal{F}}} n_F \Delta(\hat{\pi}_F, \bar{\epsilon}) + \frac{1}{\bar{\epsilon}} \left(O(\epsilon) + \tilde{O}\left(T^{-1/4} \sqrt{|\mathcal{Y}||S_{\mathcal{F}}|} \right) + O\left(\delta_{\mathcal{F}}^{1/2} \right) + O(\delta_{\mathcal{P}}) \right)$$

B.5 Omitted Proofs

B.5.1 Proof of Propositions 5 and 6

Recall that a policy p_t in period t can affect the agent's behavior μ_τ in period $\tau > t$. This raises the prospect that a mistake today can cause irreversible damage to the principal's average utility. By definition, the principal will regret that mistake. This would make the principal's problem infeasible, in that he cannot guarantee low regret for himself.

Generally-speaking, regret bounds can bypass this problem if they restrict how much the agent's response r_t depends on the policy history $p_{1:t-1}$. This is reasonable a priori, since the policy history $p_{1:t-1}$ appears irrelevant to the agent's problem. It neither affects nor predicts the state y_t, except through its dependence on $y_{1:t-1}$. It is not needed as a predictor of the policy p_t because the agent observes p_t before choosing a response.[2] For these reasons, it seems that the agent can only make herself worse off by allowing irrelevant variables like $p_{1:t-1}$ to affect her response r_t. This would be true if our notion of good performance overall had clear implications for behavior in each period, so that unecessary variation in behavior implies a departure from optimality. Unfortunately, there are various kinds of behavior that obtain low ER. The agent can easily switch between these behaviors while still satisfying no-ER. In the process, she can cause substantial benefit or harm to the principal.

To clarify the problem, we present several examples of learners that we regard as patho-

[2] For example, if the agent were Bayesian then there would be no dependence on $p_{1:t}$ at all. If the agent used the exponential weights algorithm then there would only be an indirect dependence, since that algorithm depends on the agent's historical payoffs and these, in turn, depend on the policy history.

logical. These are implicit counterexamples to the proposition that no-ER constraints are sufficient for no-regret mechanism design.

Our counterexamples are closely related to the pathological phenomenon of "superefficiency" in statistics. Suppose we are trying to estimate the mean θ of the normal random variable $X \sim N(\theta, 1)$, given an i.i.d. random sample $X_1, \ldots, X_n$. Our objective is to minimize the mean square error, but this depends on the parameter θ. A typical solution is the maximum likelihood estimator (MLE), which in this case outputs the sample mean $n^{-1} \sum_{i=1}^{n} X_i$. For reasons that are unimportant to our discussion, MLE is considered "efficient". However, it is easy to find an estimator that outperforms MLE. For example, a wild-ass guess (WAG) ignores the data and outputs θ^*. If it happens to be the case that $\theta = \theta^*$ then this estimator is optimal.

In the following example, we construct a learner that alternates between a WAG-like predictor and a MLE-like predictor depending on a seemingly irrelevant choice by the principal.

Example 8. [Selective Superefficiency] Consider a learner L that is capable – either by ingenuity or dumb luck – of predicting the state sequence $y_{1:T}$ perfectly. However, the learner uses this ability only selectively, depending on the state y_1 and policy p_1 in the first period. Despite this seemingly irrational behavior, the learner satisfies vanishing external regret.

Let $P \subsetneq \mathcal{P}$ be a nonempty subset of policies. Let $Y \subsetneq \mathcal{Y}$ be a nonempty subset of states. Let $r*$ be the best-in-hindsight response by time T. That is, consider some $r^* \in \mathcal{F}$ that happens to be best-in-hindsight given the realized state sequence $y_{1:T}$ but will not be best-in-hindsight uniformly over all state sequences. Given $y_{1:T}$, define the learner L as follows:

1. If $y_1 \in Y$ and $p_1 \in P$ then use response r^*
2. If $y_1 \in Y$ and $p_1 \notin P$ then use the response that happens to be optimal given y_t.

3. If $y_1 \notin Y$ and $p_1 \in P$ then use the response that happens to be optimal given y_t.

4. If $y_1 \notin Y$ and $p_1 \notin P$ then use response r^*

In cases 1 and 4, the learner follows the best-in-hindsight response and therefore achieves zero regret. In cases 2 and 3, the learner acts optimally ex post and therefore achieves non-positive regret.

Nonetheless, no mechanism can guarantee no-regret for the principal. Suppose the mechanism chooses $p_1 \in P$. If it turns out that $y_1 \in Y$ then the agent will follow π^*. Otherwise, the agent will be superefficient. Were the mechanism to deviate to $p_1 \notin P$, the situation would be reversed. These constitute permanent changes in the agent's behavior. Suppose one type of behavior is "better" for the principal than another. It is always possible in hindsight that the mechanism's first-period policy was the one that led to the "worse" type of behavior.

Can further assumptions rule out this kind of behavior? Again, consider the analogy with statistics. The WAG estimator – always predict θ^* – will perform very poorly in the counterfactual world where $\theta^* \neq \theta$. Formally, this estimator is not "consistent". Similarly, the learner from example 8 does not guarantee vanishing average external regret under counterfactual state sequences. This reflects a peculiar unresponsiveness to the data.

Unfortunately, imposing consistency or no-regret on all sequences does not rule out these kinds of pathologies. Consider Hodge's (superefficient) estimator, which outputs θ^* unless there is sufficient evidence that $\theta \neq \theta^*$. In that case, it outputs the sample mean. If "sufficient evidence" is defined carefully, this estimator will outperform MLE when $\theta = \theta^*$ and will asympotically match MLE otherwise. We can patch up example 8 in a similar way.

Example 9. [Selective Superefficiency, revised] We want to modify the learner in example 8 to ensure no-regret on all counterfactual state sequences $\tilde{y}_{1:T}$. This is straightforward. In period $t+1$, if the history $\tilde{y}_{1:t}$ matches the presumed sequence $y_{1:t}$ exactly, then proceed as before. Otherwise, follow any no-ER algorithm. This guarantees vanishing average regret as long as the regret from the first period t where $\tilde{y}_{1:t} \neq y_{1:t}$ is bounded – which it is, since our utility functions are bounded.

Therefore, for any sequence $y_{1:t}$ there is a learner that satisfies no-regret on all sequences, but exhibits the pathological behavior from example 8 on the realized sequence.

Statisticians deal with superefficiency by arguing that it generically fails to occur. That is, any alternative estimator will weakly underperform MLE on Lebesgue-almost all values θ. For example, we can view Hodge's estimator as asymptotically equivalent to MLE whenever $\theta \neq \theta^*$. In our setting, attempting this argument would necessitate a definition of genericity for sequences of states. While we can provide various definitions, none seem especially compelling.[3]

Another natural restriction to impose is that the learner should not outperform the experts. That is, given the sequence $y_{1:T}$, we only consider learners whose regret at period T is non-negative. Clearly, this rules out the learners in examples 8 and 9, which may obtain negative regret in the sequences where it predicts the state perfectly and acts on that information. Unfortunately, this does not rule out the broader phenomenon, as the following example illustrates.

Example 10. [Selective Superinefficiency]

As in example 8, define a learner L that appears capable of predicting the state sequence $y_{1:T}$ perfectly. This learner will continue to use this ability selectively. Moreover, when the learner uses this ability, she does not always use it to her advantage. Instead, with probability $1-q$ she uses it to her own disadvantage. When q is chosen correctly, the learner satisfies zero regret for all mechanisms.

[3] For example, one could assign equal measure to each sequence $y_{1:T}$ in the set $\mathcal{Y}^T$ of all possible sequences. By this measure, the measure of any constant sequence $(y, \ldots, y)$ would converge to zero as $T \to \infty$. Yet it does not seem unreasonable a priori that the world should persist in a fixed state. Alternatively, one could assign equal measure to all permutations of a given sequence $y_{1:T}$. This would effectively return us to an i.i.d. setting.

Let $P \subsetneq \mathcal{P}$ be a nonempty subset of policies. Let $Y \subsetneq \mathcal{Y}$ be a nonempty subset of states. Let r^* be the best-in-hindsight response by time T. Let $r_t^\dagger$ be the response that happens to be optimal for y_t. Let $\tilde{r}_t$ be the response that minimizes the agent's utility when the state is y_t. Given $y_{1:T}$, define the learner L as follows:

1. If $y_1 \in Y$ and $p_1 \in P$ then use response r^*
2. If $y_1 \in Y$ and $p_1 \notin P$ then use $r_t^\dagger$ with probability q and $\tilde{r}_t$ with probability $1 - q$
3. If $y_1 \notin Y$ and $p_1 \in P$ then use $r_t^\dagger$ with probability q and $\tilde{r}_t$ with probability $1 - q$
4. If $y_1 \notin Y$ and $p_1 \notin P$ then use prior π^*

In cases 1 and 4, the learner follows the best-in-hindsight prior and therefore achieves zero regret. In cases 2 and 3, as long as $\tilde{r}_t$ underperforms r^* in every period t, by continuity there exists a probability q such that the agent achieves zero regret.

This learner now satisfies both an upper bound and a lower bound on regret. Nonetheless, our difficulties remain. Just as in example 8, the first-period policy can cause permanent changes in the agent's behavior. It is always possible in hindsight that the mechanism's first-period policy was the one that led to the "worse" type of behavior.

We can use this example to prove proposition **??** (proposition **??** is a corollary). In the Bayesian persuasion example, let $y_{2:T}$ be drawn i.i.d. where the defendant is guilty with probability $q = 0.5 - \epsilon$ for a very small $\epsilon > 0$. If the principal chooses p_1 correctly, he can persuade the agent to convict with probability near one. Otherwise, the agent convicts with probability near 0.5.

In the contract theory example, let $y_{2:T}$ be be drawn i.i.d. from some distribution where the principal would find it optimal to pay the agent in the stage game, but both states occur with positive probability. If the principal chooses p_1 correctly, he can pay the agent her cost of effort and achieve the first-best outcome (agent works iff working is effective). Otherwise, the principal has to compensate the agent for her cost of effort in states where working is ineffective.

The fundamental problem with the learners in examples 8, 9, 10 is not that they are well-informed. After all, in some settings we might reasonably expect the agent to be better informed than the analyst. The problem is that they fail to consistently and fully exploit the private information that they clearly possess. Bounds on counterfactual internal regret capture this failure to exploit information and rule out these kinds of pathological behaviors.

Example 11. Returning to example 8, consider two constant mechanisms σ^p and $\sigma^{\tilde{p}}$ where $p \in P$ and $\tilde{p} \notin P$. Regardless of the state sequence $y_{1:T}$, exactly one of these mechanisms (say σ^p) will cause the agent to predict the state perfectly while the other will cause the agent to follow the best-in-hindsight prior. The agent's behavior r^p following σ^p will differ across periods y_t, y_τ if and only if $y_t \neq y_\tau$, while the behavior $r^{\tilde{p}}$ following $\sigma^{\tilde{p}}$ remains constant throughout. The vector $(r^p, r^{\tilde{p}})$ will therefore differ across periods y_t, y_τ if and only if $y_t \neq y_\tau$.

If we require the agent to have no-contextual regret where the context is $(r^p, r^{\tilde{p}})$, it is equivalent to requiring her to predict the state perfectly even if the principal uses $\sigma^{\tilde{p}}$. This is essentially the context used to define CIR. The learner guarantees no-CIR under mechanism σ^p, because it predicts the state perfectly. However, it does not predict the state perfectly under mechanism $\sigma^{\tilde{p}}$, so in this case the agent accumulates CIR.

B.5.2 Proof of Lemmas 26 and 27 and Additional Results

The lemmas in this section will be used repeatedly in the proofs of theorems 5, 6, and 7.

Proof of Lemmas 26 and 27

Lemma 26 states that for any policy p, information structure γ, constants $\epsilon, \tilde{\epsilon} > 0$, and distribution π, we have

$$\alpha_p(\pi, \gamma, \epsilon + \tilde{\epsilon}) \geq \alpha_p(\pi, \gamma, \epsilon) - \frac{\tilde{\epsilon}}{\epsilon} \quad \text{and} \quad \beta_p(\pi, \gamma, \epsilon + \tilde{\epsilon}) \leq \beta_p(\pi, \gamma, \epsilon) + \frac{\tilde{\epsilon}}{\epsilon}$$

Note that lemma 27 is just a special case where the information structure γ is uninformative. To prove this, define

$$B(\pi, \gamma, \epsilon) = \left\{ \mu_J \in \Delta(\mathcal{R}) \mid \epsilon \geq \max_{\tilde{r}_J \in \mathcal{R}} \mathrm{E}_{y \sim \pi} \left[\mathrm{E}_{J \sim \gamma(\cdot, y)}[U(\tilde{r}_J, p, y) - \mathrm{E}_{r \sim \mu_J}[U(r, p, y)]] \right] \right\}$$

and recall that

$$\alpha_p(\pi, \gamma, \epsilon) = \min_{\mu_J \in B(\pi, \gamma, \epsilon)} \mathrm{E}_{y \sim \pi} \left[\mathrm{E}_{J \sim \gamma(\cdot, y)}[\mathrm{E}_{r \sim \mu_J}[V(r, p, y)]] \right]$$

Note that $\alpha_p(\pi, \gamma, \epsilon)$ is decreasing and convex in ϵ. Convexity follows from the fact that $\mu_J \in B(\pi, \gamma, \epsilon)$ and $\tilde{\mu}_J \in B(\pi, \gamma, \tilde{\epsilon})$ implies $\lambda \mu_J + (1 - \lambda)\tilde{\mu}_J \in B(\pi, \gamma, \lambda\epsilon + (1 - \lambda)\tilde{\epsilon})$. Therefore,

$$\alpha_p(\pi, \gamma, \lambda\epsilon + (1 - \lambda)\tilde{\epsilon}) \leq \lambda\alpha_p(\pi, \gamma, \epsilon) + (1 - \lambda)\alpha_p(\pi, \gamma, \tilde{\epsilon})$$

Consider any supporting line of α_p at ϵ. It is bounded above by α_p, by definition. Therefore, its slope is at most

$$\frac{\alpha_p(\pi, \gamma, 0) - \alpha_p(\pi, \gamma, \epsilon)}{\epsilon} \leq \frac{1}{\epsilon}$$

since α_p is bounded in the unit interval by our regularity assumption. Therefore, the supporting line will underestimate $\alpha_p(\pi, \gamma, \epsilon + \tilde{\epsilon})$ by at most $\tilde{\epsilon}/\epsilon$ and at least zero. This implies our bound. The argument for β_p is analogous after we observe that it is increasing and concave in ϵ.

Bounds for Misspecified Distributions

The following lemma states that the principal's worst-case utility α_p is not too sensitive to changes in the distribution, for any fixed policy p.

Lemma 26. *For any policy p, information structure γ, constant $\epsilon > 0$, and distributions $\pi, \tilde{\pi}$, we have*

$$\alpha_p(\pi, \gamma, \epsilon) \geq \alpha_p(\tilde{\pi}, \gamma, \epsilon) - \frac{2d_1(\pi, \tilde{\pi})}{\epsilon} - d_1(\pi, \tilde{\pi})$$

Proof. Note that $B(\pi, \gamma, \epsilon) \subseteq B(\tilde{\pi}, \gamma, \epsilon + 2d_1(\pi, \tilde{\pi}))$ since (1) for any $\tilde{r}_J$,

$$\mathrm{E}_{y\sim\pi}\big[\mathrm{E}_{J\sim\gamma(\cdot,y)}[U(\tilde{r}_J, p, y)]\big] \geq \mathrm{E}_{y\sim\tilde{\pi}}\big[\mathrm{E}_{J\sim\gamma(\cdot,y)}[U(\tilde{r}_J, p, y)]\big] - d_1(\pi, \tilde{\pi})$$

and (2), for any μ_J,

$$\mathrm{E}_{y\sim\pi}\big[\mathrm{E}_{J\sim\gamma(\cdot,y)}\big[\mathrm{E}_{r\sim\mu_J^*}[U(r, p, y)]\big]\big] \leq \mathrm{E}_{y\sim\tilde{\pi}}\big[\mathrm{E}_{J\sim\gamma(\cdot,y)}\big[\mathrm{E}_{r\sim\mu_J^*}[U(r, p, y)]\big]\big] + d_1(\pi, \tilde{\pi})$$

Therefore,

$$\begin{aligned}
\alpha_p(\pi, \gamma, \epsilon) &\geq \min_{\mu_J\in B(\tilde{\pi},\gamma,\epsilon+2d_1(\pi,\tilde{\pi}))} \mathrm{E}_{y\sim\pi}\big[\mathrm{E}_{J\sim\gamma(\cdot,y)}[\mathrm{E}_{r\sim\mu_J}[V(r, p, y)]]\big] \\
&\geq \min_{\mu_J\in B(\tilde{\pi},\gamma,\epsilon+2d_1(\pi,\tilde{\pi}))} \mathrm{E}_{y\sim\tilde{\pi}}\big[\mathrm{E}_{J\sim\gamma(\cdot,y)}[\mathrm{E}_{r\sim\mu_J}[V(r, p, y)]]\big] - d_1(\pi, \tilde{\pi}) \\
&= \alpha_p(\tilde{\pi}, \gamma, \epsilon + 2d_1(\pi, \tilde{\pi})) - d_1(\pi, \tilde{\pi}) \\
&\geq \alpha_p(\tilde{\pi}, \gamma, \epsilon) - \frac{2d_1(\pi, \tilde{\pi})}{\epsilon} - d_1(\pi, \tilde{\pi})
\end{aligned}$$

□

The following lemma states that the ϵ-robust policy for a distribution $\tilde{\pi}$ that is near the true distribution π will perform almost as well as the ϵ-robust policy for the true distribution π.

Lemma 27. *For any $\epsilon > 0$ and distributions $\pi, \tilde{\pi}$, we have*

$$\alpha_{p^*(\tilde{\pi},\epsilon)}(\pi, \epsilon) \geq \alpha_{p^*(\pi,\epsilon)}(\pi, \epsilon) - \frac{4d_1(\pi, \tilde{\pi})}{\epsilon} - 2d_1(\pi, \tilde{\pi})$$

Proof. First, observe that

$$\begin{aligned} \alpha_{p^*(\pi,\epsilon)}(\pi, \epsilon) &\leq \alpha_{p^*(\pi,\epsilon)}(\tilde{\pi}, \epsilon) + \frac{2d_1(\pi, \tilde{\pi})}{\epsilon} + d_1(\pi, \tilde{\pi}) \\ &\leq \alpha_{p^*(\tilde{\pi},\epsilon)}(\tilde{\pi}, \epsilon) + \frac{2d_1(\pi, \tilde{\pi})}{\epsilon} + d_1(\pi, \tilde{\pi}) \end{aligned}$$

Next, observe that

$$\alpha_{p^*(\tilde{\pi},\epsilon)}(\tilde{\pi}, \epsilon) \leq \alpha_{p^*(\tilde{\pi},\epsilon)}(\pi, \epsilon) + \frac{2d_1(\pi, \tilde{\pi})}{\epsilon} + d_1(\pi, \tilde{\pi})$$

Collapse these inequalities to obtain the desired result. □

The following lemma states that the ϵ-informationally-robust policy for a distribution $\tilde{\pi}$ that is near the true distribution π will provide a similar guarantee against the worst-case information structure γ as the ϵ-informationally-robust policy for the true distribution π.

Lemma 28. *For any $\epsilon > 0$ and distributions $\pi, \tilde{\pi}$, we have*

$$\inf_{\gamma} \alpha_{p^\dagger(\pi,\epsilon)}(\pi, \gamma, \epsilon) \leq \inf_{\gamma} \alpha_{p^\dagger(\tilde{\pi},\epsilon)}(\pi, \gamma, \epsilon) + \frac{4d_1(\pi, \tilde{\pi})}{\epsilon} + 2d_1(\pi, \tilde{\pi})$$

Proof. First, observe that

$$\alpha_{p^\dagger(\pi,\epsilon)}(\pi, \gamma, \epsilon) \leq \alpha_{p^\dagger(\pi,\epsilon)}(\tilde{\pi}, \gamma, \epsilon) + \frac{2d_1(\pi, \tilde{\pi})}{\epsilon} + d_1(\pi, \tilde{\pi})$$

which implies

$$\begin{aligned}\inf_{\gamma} \alpha_{p^{\dagger}(\pi,\epsilon)}(\pi,\gamma,\epsilon) &\leq \inf_{\gamma} \alpha_{p^{\dagger}(\pi,\epsilon)}(\tilde{\pi},\gamma,\epsilon) + \frac{2d_1(\pi,\tilde{\pi})}{\epsilon} + d_1(\pi,\tilde{\pi}) \\ &\leq \inf_{\gamma} \alpha_{p^{\dagger}(\tilde{\pi},\epsilon)}(\tilde{\pi},\gamma,\epsilon) + \frac{2d_1(\pi,\tilde{\pi})}{\epsilon} + d_1(\pi,\tilde{\pi})\end{aligned}$$

Next, observe that

$$\alpha_{p^{\dagger}(\tilde{\pi},\epsilon)}(\tilde{\pi},\gamma,\epsilon) \leq \alpha_{p^{\dagger}(\tilde{\pi},\epsilon)}(\pi,\gamma,\epsilon) + \frac{2d_1(\pi,\tilde{\pi})}{\epsilon} + d_1(\pi,\tilde{\pi})$$

which implies

$$\inf_{\gamma} \alpha_{p^{\dagger}(\tilde{\pi},\epsilon)}(\tilde{\pi},\gamma,\epsilon) \leq \inf_{\gamma} \alpha_{p^{\dagger}(\tilde{\pi},\epsilon)}(\pi,\gamma,\epsilon) + \frac{2d_1(\pi,\tilde{\pi})}{\epsilon} + d_1(\pi,\tilde{\pi})$$

Collapse these inequalities to obtain the desired result. □

B.5.3 Proof of Theorem 10

Assume access to a forecast $\pi_t \in \Delta(\mathcal{Y})$ for every period t. We will define this later. In period t, the mechanism computes the policy $p^*(\pi_t,\bar{\epsilon})$ that maximizes the worst-case payoff in the $\bar{\epsilon}$-robust stage game, treating the forecast π_t as the common prior. That is,

$$p^*(\pi_t,\bar{\epsilon}) \in \arg\max_{p\in\mathcal{P}} \alpha_p(\pi_t,\bar{\epsilon})$$

The mechanism chooses p_t as follows. Let P be the unique policy context that includes the policy $p^*(\pi_t,\bar{\epsilon}) \in P$. Let $p_t = p_P$, where $p_P \in \mathcal{P}_1$ is the representative element of P.

We will refer to the average regret accumulated in each forecast context F, i.e.

$$\epsilon_F = \max_{r \in \mathcal{R}} \frac{1}{n_F} \sum_{t \in F} (U(r, p_F, y_t) - U(r_t, p_F, y_t))$$

where $p_F = p_P$ for the unique policy context P associated with forecast context F. We will also refer to the average regret accumulated in each information context $I \in \mathcal{I}$, i.e.

$$\epsilon_I = \max_{r \in \mathcal{R}} \frac{1}{n_I} \sum_{t \in I} (U(r, p_I, y_t) - U(r_t, p_I, y_t))$$

where $p_I = p_F$ for the unique forecast context F associated with information I. Note that

$$\epsilon_F = \frac{1}{n_F} \sum_{I \in \mathcal{I}_F} n_I \epsilon_I = \frac{1}{n_F} \sum_{I \in \mathcal{I}_F} \max_{r \in \mathcal{R}} \frac{1}{n_I} \sum_{t \in I} (U(r, p_F, y_t) - U(r_t, p_F, y_t))$$

The next two lemmas imply an upper bound on the principal's regret in terms of the quantity

$$\iota \geq \frac{1}{T} \sum_{t=1}^{T} d_1(\pi_t, \hat{\pi}_I)$$

that measures the discrepancy between the forecast π_t and the empirical distribution $\hat{\pi}_I$ conditioned on the agent's information I. Lemma 29 is a lower bound on the principal's payoff under σ^*. Lemma 30 is an upper bound on the his payoff under any constant $\sigma^p \in \Sigma_0$.

Lemma 29. *Suppose the principal runs σ^*. Then*

$$\frac{1}{T} \sum_{I \in \mathcal{I}} \sum_{t \in I} V(r_t, p_t, y_t) \geq \max_{p \in \mathcal{P}} \alpha_p(\hat{\pi}_I, \bar{\epsilon}) - \left(\frac{\epsilon + 4\iota + K^U_{\mathcal{R}} \delta_{\mathcal{R}} + 2K^U_{\mathcal{P}} \delta_{\mathcal{P}}}{\bar{\epsilon}} \right) - \left(2\iota + K^V_{\mathcal{R}} \delta_{\mathcal{R}} + K^V_{\mathcal{P}} \delta_{\mathcal{P}}\right)$$

Proof. Let r_I be a representative element in the response context R associated with information I under mechanism σ^*. By regularity,

$$\begin{aligned}\epsilon_I &\geq \max_{r\in\mathcal{R}} \frac{1}{n_I}\sum_{t\in I}\left(U(r,p_I,y_t) - U(r_I,p_I,y_t)\right) - K^U_{\mathcal{R}}\delta_{\mathcal{R}} \\ &= \max_{r\in\mathcal{R}} \mathrm{E}_{y\sim\hat{\pi}_I}[U(r,p_I,y) - U(r_I,p_I,y)] - K^U_{\mathcal{R}}\delta_{\mathcal{R}}\end{aligned}$$

It follows, by regularity and definition of α, that

$$\begin{aligned}\frac{1}{n_I}\sum_{t\in I} V(r_t,p_I,y_t) &\geq \mathrm{E}_{y\sim\hat{\pi}_I}[V(r_I,p_I,y)] - K^V_{\mathcal{R}}\delta_{\mathcal{R}} \\ &\geq \alpha_{p_I}(\hat{\pi}_I,\epsilon_I + K^U_{\mathcal{R}}\delta_{\mathcal{R}}) - K^V_{\mathcal{R}}\delta_{\mathcal{R}}\end{aligned}$$

Summing over information $I \in \mathcal{I}$ and using lemma 15, we obtain

$$\begin{aligned}\frac{1}{T}\sum_{I\in\mathcal{I}}\sum_{t\in I} V(r_t,p_I,y_t) &\geq \frac{1}{T}\sum_{I\in\mathcal{I}}\sum_{t\in I}\alpha_{p_I}(\hat{\pi}_I,\epsilon_I + K^U_{\mathcal{R}}\delta_{\mathcal{R}}) - K^V_{\mathcal{R}}\delta_{\mathcal{R}} \\ &\geq \frac{1}{T}\sum_{I\in\mathcal{I}}\sum_{t\in I}\alpha_{p^*(\pi_t,\bar{\epsilon})}(\hat{\pi}_I,\epsilon_I + K^U_{\mathcal{R}}\delta_{\mathcal{R}} + 2K^U_{\mathcal{P}}\delta_{\mathcal{P}}) - K^V_{\mathcal{R}}\delta_{\mathcal{R}} - K^V_{\mathcal{P}}\delta_{\mathcal{P}} \\ &\geq \frac{1}{T}\sum_{I\in\mathcal{I}}\sum_{t\in I}\left(\alpha_{p^*(\pi_t,\bar{\epsilon})}(\hat{\pi}_I,\bar{\epsilon}) - \frac{\epsilon_I + K^U_{\mathcal{R}}\delta_{\mathcal{R}} + 2K^U_{\mathcal{P}}\delta_{\mathcal{P}}}{\bar{\epsilon}}\right) - K^V_{\mathcal{R}}\delta_{\mathcal{R}} - K^V_{\mathcal{P}}\delta_{\mathcal{P}} \\ &\geq \frac{1}{T}\sum_{I\in\mathcal{I}}\sum_{t\in I}\alpha_{p^*(\pi_t,\bar{\epsilon})}(\hat{\pi}_I,\bar{\epsilon}) - \left(\frac{\epsilon + K^U_{\mathcal{R}}\delta_{\mathcal{R}} + 2K^U_{\mathcal{P}}\delta_{\mathcal{P}}}{\bar{\epsilon}}\right) - K^V_{\mathcal{R}}\delta_{\mathcal{R}} - K^V_{\mathcal{P}}\delta_{\mathcal{P}}\end{aligned}$$

Focus on the first term, i.e.

$$\begin{aligned}\frac{1}{T}\sum_{I\in\mathcal{I}}\sum_{t\in I}\alpha_{p^*(\pi_t,\bar{\epsilon})}(\hat{\pi}_I,\bar{\epsilon}) &\geq \frac{1}{T}\sum_{I\in\mathcal{I}}\sum_{t\in I}\left(\alpha_{p^*(\hat{\pi}_I,\bar{\epsilon})}(\hat{\pi}_I,\bar{\epsilon}) - \frac{4d_1(\pi_t,\hat{\pi}_I)}{\bar{\epsilon}} - 2d_1(\pi_t,\hat{\pi}_I)\right) \\ &\geq \frac{1}{T}\sum_{I\in\mathcal{I}} n_I\alpha_{p^*(\hat{\pi}_I,\bar{\epsilon})}(\hat{\pi}_I,\bar{\epsilon}) - \frac{4\iota}{\bar{\epsilon}} - 2\iota\end{aligned}$$

Collapsing these inequalities gives us the desired bound. □

Lemma 30. *Suppose the principal runs some constant mechanism $\sigma^p \in \Sigma_0$. Then*

$$\frac{1}{T}\sum_{t=1}^{T} V(r_t, p, y_t) \leq \frac{1}{T}\sum_{I\in\mathcal{I}} n_I \max_{\tilde{p}\in\mathcal{P}} \beta_{\tilde{p}}(\hat{\pi}_I, \bar{\epsilon}) + \left(\frac{\epsilon + K^U_{\mathcal{R}}\delta_{\mathcal{R}}}{\bar{\epsilon}}\right) + K^V_{\mathcal{R}}\delta_{\mathcal{R}}$$

Proof. Let r_I be a representative element in the response context R associated with information I under σ^p. By regularity,

$$\begin{aligned}\epsilon_I &\geq \max_{r\in\mathcal{R}} \frac{1}{n_I}\sum_{t\in I} (U(r, p_I, y_t) - U(r_I, p_I, y_t)) - K^U_{\mathcal{R}}\delta_{\mathcal{R}} \\ &= \max_{r\in\mathcal{R}} \mathrm{E}_{y\sim\hat{\pi}_I}[U(r, p_I, y) - U(r_I, p_I, y)] - K^U_{\mathcal{R}}\delta_{\mathcal{R}}\end{aligned}$$

It follows, by regularity and definition of β, that

$$\begin{aligned}\frac{1}{n_I}\sum_{t\in I} V(r_t, p, y_t) &\leq \mathrm{E}_{y\sim\hat{\pi}_I}[V(r_I, p, y)] + K^V_{\mathcal{R}}\delta_{\mathcal{R}} \\ &\leq \alpha_p(\hat{\pi}_I, \epsilon_I + K^U_{\mathcal{R}}\delta_{\mathcal{R}}) + K^V_{\mathcal{R}}\delta_{\mathcal{R}}\end{aligned}$$

Summing over information $I \in \mathcal{I}$ and using lemma 15, we obtain

$$\begin{aligned}\frac{1}{T}\sum_{I\in\mathcal{I}}\sum_{t\in I} V(r_t, p, y_t) &\leq \frac{1}{T}\sum_{I\in\mathcal{I}} n_I \beta_p(\hat{\pi}_I, \epsilon_I + K^U_{\mathcal{R}}\delta_{\mathcal{R}}) + K^V_{\mathcal{R}}\delta_{\mathcal{R}} \\ &\leq \frac{1}{T}\sum_{I\in\mathcal{I}} n_I \left(\beta_p(\hat{\pi}_I, \bar{\epsilon}) + \frac{\epsilon_I + K^U_{\mathcal{R}}\delta_{\mathcal{R}}}{\bar{\epsilon}}\right) + K^V_{\mathcal{R}}\delta_{\mathcal{R}} \\ &\leq \frac{1}{T}\sum_{I\in\mathcal{I}} n_I \beta_p(\hat{\pi}_I, \bar{\epsilon}) + \left(\frac{\epsilon + K^U_{\mathcal{R}}\delta_{\mathcal{R}}}{\bar{\epsilon}}\right) + K^V_{\mathcal{R}}\delta_{\mathcal{R}} \\ &\leq \frac{1}{T}\sum_{I\in\mathcal{I}} n_I \max_{\tilde{p}\in\mathcal{P}} \beta_{\tilde{p}}(\hat{\pi}_I, \bar{\epsilon}) + \left(\frac{\epsilon + K^U_{\mathcal{R}}\delta_{\mathcal{R}}}{\bar{\epsilon}}\right) + K^V_{\mathcal{R}}\delta_{\mathcal{R}}\end{aligned}$$

This is the desired bound. □

From these two lemmas, it immediately follows that

$$\mathrm{PR} \leq \frac{1}{T}\sum_{I\in\mathcal{I}} n_I \Delta(\hat{\pi}_I, \bar{\epsilon}) + 2\left(\frac{\epsilon + 2\iota + K^U_{\mathcal{R}}\delta_{\mathcal{R}} + K^U_{\mathcal{P}}\delta_{\mathcal{P}}}{\bar{\epsilon}}\right) + \left(2\iota + 2K^V_{\mathcal{R}}\delta_{\mathcal{R}} + K^V_{\mathcal{P}}\delta_{\mathcal{P}}\right)$$

Therefore, to bound the principal's regret, all that remains is to bound ι.

Set $\mathcal{C} = S_{\mathcal{R}}^{|\mathcal{P}_0|+|\mathcal{P}_1|}$ where $C_t(\mathcal{P}_1)$ is the vector describing the agent's response contexts R_t under policy history $p^*_{1:t-1}$ and policy choices $p_t \in \mathcal{P}_1$. Note that this is different from the behavior context that we used to define the agent's information, which refers to the response context R_t under policy history $p^*_{1:t}$. Because we are currently designing the mechanism, we cannot refer to p^*_t without attempting to solve a fixed point problem that may not have a solution.

We use the algorithm from appendix **??** to generate π_t, with a modification: run it separately for each context C_t. Adapting equation (B.3), we obtain

$$\begin{aligned}
\mathrm{E}_{\sigma^*}\left[\frac{1}{T}\sum_{C\in\mathcal{C}}\sum_{F\in\mathcal{F}} n_{F,C} d_1(\pi_t, \hat{\pi}_C)\right] &\leq \frac{1}{T}\sum_{C\in\mathcal{C}} n_C \sqrt{|\mathcal{Y}||\mathcal{F}|\sqrt{\frac{2\log|\mathcal{F}|}{n_C}} + 2|\mathcal{Y}|\delta_{\mathcal{F}}} \\
&\leq \frac{1}{T}\sum_{C\in\mathcal{C}} n_C^{3/4}\sqrt{|\mathcal{Y}||\mathcal{F}|\sqrt{2\log|\mathcal{F}|}} + \sqrt{2|\mathcal{Y}|\delta_{\mathcal{F}}} \\
&\leq \frac{1}{T}\sum_{C\in\mathcal{C}} \left(\frac{T}{|\mathcal{C}|}\right)^{3/4} \sqrt{|\mathcal{Y}||\mathcal{F}|\sqrt{2\log|\mathcal{F}|}} + \sqrt{2|\mathcal{Y}|\delta_{\mathcal{F}}} \\
&= \left(\frac{|\mathcal{C}|}{T}\right)^{1/4} \sqrt{|\mathcal{Y}||\mathcal{F}|\sqrt{2\log|\mathcal{F}|}} + \sqrt{2|\mathcal{Y}|\delta_{\mathcal{F}}}
\end{aligned}$$

where $n_{F,C}$ is the number of periods t where $C_t = C$ and $\pi_t \in F$.

Consider any two periods t, τ where $I_t = I_\tau$ but $C_t \neq C_\tau$. Since $I_t = I_\tau$ and information includes the forecast as context, we know that $\pi_t = \pi_\tau$. Now, consider

$$n_{F_t,C_t} d_1(\pi_t, \hat{\pi}_{C_t}) + n_{F_\tau,C_\tau} d_1(\pi_\tau, \hat{\pi}_{C_\tau})$$

$$= n_{F_t,C_t} d_1(\pi_t, \hat{\pi}_{C_t}) + n_{F_t,C_\tau} d_1(\pi_t, \hat{\pi}_{C_\tau})$$

$$\geq \left(n_{F_t,C_t} + n_{F_t,C_\tau}\right) d_1 \left(\pi_t, \frac{1}{n_{F_t,C_t} + n_{F_t,C_\tau}} \left(n_{F_t,C_t}\hat{\pi}_{C_t} + n_{F_t,C_\tau} d_1(\pi_t, \hat{\pi}_{C_\tau})\right)\right)$$

by subadditivity and homogeneity of norms. By continuing this process of combining contexts, we find

$$\frac{1}{T}\sum_{C\in\mathcal{C}}\sum_{F\in\mathcal{F}} n_{F,C} d_1(\pi_t, \hat{\pi}_C) \geq \frac{1}{T}\sum_{I\in\mathcal{I}} n_I d_1(\pi_t, \hat{\pi}_I) = \iota$$

Therefore, the earlier miscalibration bound applies to $\mathrm{E}_{\sigma^*}[\iota]$ as well. Finally, we obtain our bound on the expected principal's regret.

$$\begin{aligned}
\mathrm{E}_{\sigma^*}[\mathrm{PR}] \leq & \frac{1}{T}\sum_{I\in\mathcal{I}} n_I \Delta(\hat{\pi}_I, \bar{\epsilon}) \\
& + 2\left(\frac{\epsilon + 2\left(\frac{|\mathcal{C}|}{T}\right)^{1/4}\sqrt{|\mathcal{Y}||\mathcal{F}|\sqrt{2\log|\mathcal{F}|}} + 2\sqrt{2|\mathcal{Y}|\delta_{\mathcal{F}}} + K^U_{\mathcal{R}}\delta_{\mathcal{R}} + K^U_{\mathcal{P}}\delta_{\mathcal{P}}}{\bar{\epsilon}}\right) \\
& + \left(2\left(\frac{|\mathcal{C}|}{T}\right)^{1/4}\sqrt{|\mathcal{Y}||\mathcal{F}|\sqrt{2\log|\mathcal{F}|}} + 2\sqrt{2|\mathcal{Y}|\delta_{\mathcal{F}}} + 2K^V_{\mathcal{R}}\delta_{\mathcal{R}} + K^V_{\mathcal{P}}\delta_{\mathcal{P}}\right)
\end{aligned}$$

B.5.4 Proof of Theorem 11

Define

$$\hat{\gamma}_P(I, y) = \frac{n_I \hat{\pi}_I(y)}{n_F \hat{\pi}_F(y)} \cdot \mathbf{1}(I \in \mathcal{I}_F)$$

as the empirical information structure conditional on forecast context F. This definition follows from Bayes' rule.

Assume access to a forecast $\pi_t \in \Delta(\mathcal{Y})$ for every period t. We will define this later. In period t, the mechanism computes the policy $p^*(\pi_t, \bar{\epsilon})$ that maximizes the worst-case payoff in the $\bar{\epsilon}$-robust stage game, treating the forecast π_t as the common prior. That is,

$$p^*(\pi_t, \bar{\epsilon}) \in \arg\max_{p \in \mathcal{P}} \alpha_p(\pi_t, \bar{\epsilon})$$

The mechanism chooses p_t as follows. Let P be the unique policy context that includes the policy $p^*(\pi_t, \bar{\epsilon}) \in P$. Let $p_t = p_P$, where $p_P \in \mathcal{P}_1$ is the representative element of P.

The next two lemmas imply an upper bound on the principal's regret in terms of the quantity

$$\iota \geq \frac{1}{T} \sum_{t=1}^{T} d_1(\pi_t, \hat{\pi}_F)$$

that measures the discrepancy between the forecast π_t and the empirical distribution $\hat{\pi}_F$ conditioned on the forecast context F. Lemma 31 is a lower bound on the principal's payoff under σ^*. Lemma 32 is an upper bound on the his payoff under any constant $\sigma^p \in \Sigma_0$.

Lemma 31. *Suppose the principal runs the mechanism σ^*. Then*

$$\begin{aligned}\frac{1}{T}\sum_{t=1}^{T} V(r_t, p_t, y_t) \geq & \frac{1}{T}\sum_{F \in \mathcal{F}} n_F \max_{p \in \mathcal{P}} \alpha_p(\hat{\pi}_F, \bar{\epsilon}) - \left(\frac{\epsilon + 6\iota + K_{\mathcal{R}}^U \delta_{\mathcal{R}} + 2K_{\mathcal{P}}^U \delta_{\mathcal{P}}}{\bar{\epsilon}}\right) \\ & - \left(M_1(\epsilon + \tilde{\epsilon} + 2\iota + 2K_{\mathcal{P}}^U \delta_{\mathcal{P}}) + M_2 + 3\iota + K_{\mathcal{R}}^V \delta_{\mathcal{R}} + K_{\mathcal{P}}^V \delta_{\mathcal{P}}\right)\end{aligned}$$

Proof. Let r_I be a representative element in the response context R associated with information I under mechanism σ^*. By regularity,

$$\begin{aligned}\epsilon_I &\geq \max_{r \in \mathcal{R}} \sum_{t \in I} (U(r, p_I, y_t) - U(r_I, p_I, y_t)) - K_{\mathcal{R}}^U \delta_{\mathcal{R}} \\ &= \max_{r \in \mathcal{R}} \mathrm{E}_{y \sim \hat{\pi}_I}[U(r, p_I, y) - U(r_I, p_I, y)] - K_{\mathcal{R}}^U \delta_{\mathcal{R}}\end{aligned}$$

It follows, by regularity and definition of α, that

$$\begin{aligned}\frac{1}{n_I}\sum_{t\in I} V(r_t, p_I, y_t) &\geq \mathrm{E}_{y\sim\hat{\pi}_I}[V(r_I, p_I, y)] - K^V_{\mathcal{R}}\delta_{\mathcal{R}} \\ &\geq \alpha_{p_I}(\hat{\pi}_I, \epsilon_I + K^U_{\mathcal{R}}\delta_{\mathcal{R}}) - K^V_{\mathcal{R}}\delta_{\mathcal{R}}\end{aligned}$$

Summing over information $I \in \mathcal{I}_F$ and using lemma 15, we obtain

$$\begin{aligned}\frac{1}{n_F}\sum_{I\in\mathcal{I}_F}\sum_{t\in I} V(r_t, p_F, y_t) &\geq \frac{1}{n_F}\sum_{I\in\mathcal{I}_F}\sum_{t\in I}\alpha_{p_F}(\hat{\pi}_I, \epsilon_I + K^U_{\mathcal{R}}\delta_{\mathcal{R}}) - K^V_{\mathcal{R}}\delta_{\mathcal{R}} \\ &\geq \frac{1}{n_F}\sum_{I\in\mathcal{I}_F}\sum_{t\in I}\alpha_{p^*(\pi_t,\bar{\epsilon})}(\hat{\pi}_I, \epsilon_I + K^U_{\mathcal{R}}\delta_{\mathcal{R}} + 2K^U_{\mathcal{P}}\delta_{\mathcal{P}}) \\ &- K^V_{\mathcal{R}}\delta_{\mathcal{R}} - K^V_{\mathcal{P}}\delta_{\mathcal{P}} \\ &\geq \frac{1}{n_F}\sum_{I\in\mathcal{I}_F}\sum_{t\in I}\left(\alpha_{p^*(\pi_t,\bar{\epsilon})}(\hat{\pi}_I, \bar{\epsilon}) - \frac{\epsilon_I + K^U_{\mathcal{R}}\delta_{\mathcal{R}} + 2K^U_{\mathcal{P}}\delta_{\mathcal{P}}}{\bar{\epsilon}}\right) \\ &- K^V_{\mathcal{R}}\delta_{\mathcal{R}} - K^V_{\mathcal{P}}\delta_{\mathcal{P}} \\ &= \frac{1}{n_F}\sum_{I\in\mathcal{I}_F}\sum_{t\in I}\alpha_{p^*(\pi_t,\bar{\epsilon})}(\hat{\pi}_I, \bar{\epsilon}) - \left(\frac{\epsilon_F + K^U_{\mathcal{R}}\delta_{\mathcal{R}} + 2K^U_{\mathcal{P}}\delta_{\mathcal{P}}}{\bar{\epsilon}}\right) \\ &- K^V_{\mathcal{R}}\delta_{\mathcal{R}} - K^V_{\mathcal{P}}\delta_{\mathcal{P}}\end{aligned}$$

So far, we have a lower bound for the principal's payoff that nearly matches the principal's worst-case payoff in the stage game if the agent had information structure $\hat{\gamma}_F$ in each forecast context. Furthermore, we know that this information structure cannot be particularly useful to the agent. Define

$$-\tilde{\epsilon}_F = \max_{r\in\mathcal{R}}\frac{1}{n_F}\sum_{I\in\mathcal{I}_F}\sum_{t\in I}(U(r, p_F, y_t) - U(r_t, p_F, y_t))$$

as the (possibly negative) regret accumulated in forecast context F relative to the best-in-hindsight response, rather than the best-in-hindsight function from information to responses.

Let π_F be the forecast associated with forecast context F. Note that

$$\begin{aligned}
\epsilon_F + \tilde{\epsilon}_F &= \min_{\tilde{r}\in\mathcal{R}} \frac{1}{n_F} \sum_{I\in\mathcal{I}_F} \max_{r\in\mathcal{R}} \sum_{t\in I} \left(U(r, p_F, y_t) - U(\tilde{r}, p_F, y_t)\right) \\
&\geq \min_{\tilde{r}\in\mathcal{R}} \frac{1}{n_F} \sum_{I\in\mathcal{I}_F} \max_{r\in\mathcal{R}} \sum_{t\in I} \left(U(r, p^*(\pi_F,\bar{\epsilon}), y_t) - U(\tilde{r}, p^*(\pi_F,\bar{\epsilon}), y_t)\right) - 2K^U_{\mathcal{P}}\delta_{\mathcal{P}} \\
&= \min_{r}\max_{r_J} \mathrm{E}_{y\sim\hat{\pi}_F}\big[\mathrm{E}_{J\sim\hat{\gamma}_F(\cdot,y)}[U(r_J, p^*(\pi_F,\bar{\epsilon}), y) - U(r, p^*(\pi_F,\bar{\epsilon}), y)]\big] - 2K^U_{\mathcal{P}}\delta_{\mathcal{P}} \\
&\geq \min_{r}\max_{r_J} \mathrm{E}_{y\sim\pi_F}\big[\mathrm{E}_{J\sim\hat{\gamma}_F(\cdot,y)}[U(r_J, p^*(\pi_F,\bar{\epsilon}), y) - U(r, p^*(\pi_F,\bar{\epsilon}), y)]\big] \\
&- 2d_1(\pi_F, \hat{\pi}_F) - 2K^U_{\mathcal{P}}\delta_{\mathcal{P}}
\end{aligned}$$

It follows from assumption 20 that

$$M_1(\epsilon_F + \tilde{\epsilon}_F + 2d_1(\pi_F,\hat{\pi}_F) + 2K^U_{\mathcal{P}}\delta_{\mathcal{P}}) + M_2 \geq \alpha_{p^*(\pi_F,\bar{\epsilon})}(\pi_F,\bar{\epsilon}) - \alpha_{p^*(\pi_F,\bar{\epsilon})}(\pi_F,\hat{\gamma}_F,\bar{\epsilon})$$

which can be rewritten as

$$\alpha_{p^*(\pi_F,\bar{\epsilon})}(\pi_F,\hat{\gamma}_F,\bar{\epsilon}) \geq \alpha_{p^*(\pi_F,\bar{\epsilon})}(\pi_F,\bar{\epsilon}) - \big(M_1(\epsilon_F + \tilde{\epsilon}_F + 2d_1(\pi_F,\hat{\pi}_F) + 2K^U_{\mathcal{P}}\delta_{\mathcal{P}}) + M_2\big) \tag{B.4}$$

Next, we relate our lower bound on the principal's payoff to the term $\alpha_{p^*(\pi_F,\bar{\epsilon})}(\pi_F,\hat{\gamma}_F,\bar{\epsilon})$. Note that

$$\begin{aligned}
&\frac{1}{n_F}\sum_{I\in\mathcal{I}_F} n_I \alpha_{p^*(\pi_F,\bar{\epsilon})}(\hat{\pi}_I,\bar{\epsilon}) \\
&= \mathrm{E}_{y\sim\hat{\pi}_F}\big[\mathrm{E}_{J\sim\hat{\gamma}(\cdot,y)}\big[\alpha_{p^*(\pi_F,\bar{\epsilon})}(\hat{\pi}_{F|J},\bar{\epsilon})\big]\big] \\
&\geq \min_{e_J} \mathrm{E}_{y\sim\hat{\pi}_F}\big[\mathrm{E}_{J\sim\hat{\gamma}(\cdot,y)}\big[\alpha_{p^*(\pi_F,\bar{\epsilon})}(\pi_{F|J}, e_J)\big]\big] \quad \text{s.t.} \quad \bar{\epsilon} = \mathrm{E}_{y\sim\pi_F}[\mathrm{E}_{J\sim\hat{\gamma}\cdot,y}[e_J]] \\
&= \alpha_{p^*(\pi_F,\bar{\epsilon})}(\hat{\pi}_F,\hat{\gamma}_F,\bar{\epsilon}) \\
&\geq \alpha_{p^*(\pi_F,\bar{\epsilon})}(\pi_F,\hat{\gamma}_F,\bar{\epsilon}) - \frac{2d_1(\pi_F,\hat{\pi}_F)}{\bar{\epsilon}} - d_1(\pi_F,\hat{\pi}_F)
\end{aligned}$$

So combining this with inequality (B.4) gives

$$
\begin{aligned}
&\frac{1}{n_F}\sum_{I\in\mathcal{I}_F} n_I \alpha_{p^*(\pi_F,\bar{\epsilon})}(\hat{\pi}_I,\bar{\epsilon})\\
&\geq \alpha_{p^*(\pi_F,\bar{\epsilon})}(\pi_F,\bar{\epsilon}) - \big(M_1(\epsilon_F + \tilde{\epsilon}_F + 2d_1(\pi_F,\hat{\pi}_F) + 2K^U_{\mathcal{P}}\delta_{\mathcal{P}}) + M_2\big)\\
&-\left(\frac{2d_1(\pi_F,\hat{\pi}_F)}{\bar{\epsilon}}\right) - d_1(\pi_F,\hat{\pi}_F)\\
&\geq \alpha_{p^*(\hat{\pi}_F,\bar{\epsilon})}(\hat{\pi}_F,\bar{\epsilon}) - \big(M_1(\epsilon_F + \tilde{\epsilon}_F + 2d_1(\pi_F,\hat{\pi}_F) + 2K^U_{\mathcal{P}}\delta_{\mathcal{P}}) + M_2\big)\\
&-\left(\frac{6d_1(\pi_F,\hat{\pi}_F)}{\bar{\epsilon}}\right) - 3d_1(\pi_F,\hat{\pi}_F)
\end{aligned}
$$

Collapsing these inequalities gives us

$$
\begin{aligned}
&\frac{1}{n_F}\sum_{t\in F} V(r_t,p_F,y_t)\\
&\geq \alpha_{p^*(\hat{\pi}_F,\bar{\epsilon})}(\hat{\pi}_F,\bar{\epsilon}) - \left(\frac{\epsilon_F + 6d_1(\pi_F,\hat{\pi}_F) + K^U_{\mathcal{R}}\delta_{\mathcal{R}} + 2K^U_{\mathcal{P}}\delta_{\mathcal{P}}}{\bar{\epsilon}}\right)\\
&-\big(M_1(\epsilon_F + \tilde{\epsilon}_F + 2d_1(\pi_F,\hat{\pi}_F) + 2K^U_{\mathcal{P}}\delta_{\mathcal{P}}) + M_2 + 3d_1(\pi_F,\hat{\pi}_F) + K^V_{\mathcal{R}}\delta_{\mathcal{R}} + K^V_{\mathcal{P}}\delta_{\mathcal{P}}\big)
\end{aligned}
$$

Summing over forecast contexts $F \in \mathcal{F}$ gives us the desired result. □

Lemma 32. *Suppose the principal runs some constant mechanism $\sigma^p \in \Sigma_0$. Then*

$$
\frac{1}{T}\sum_{t=1}^{T} V(r_t,p,y_t) \leq \frac{1}{T}\sum_{F\in\mathcal{F}} n_F \max_{\tilde{p}\in\mathcal{P}} \beta_{\tilde{p}}(\hat{\pi}_F,\bar{\epsilon}) + \left(\frac{\epsilon + K^U_{\mathcal{R}}\delta_{\mathcal{R}}}{\bar{\epsilon}}\right) + \big(M_1(\epsilon+\tilde{\epsilon}) + M_2 + K^V_{\mathcal{R}}\delta_{\mathcal{R}}\big)
$$

Proof. Let r_I be a representative element in the response context R associated with information I under mechanism σ^p. By regularity,

$$
\begin{aligned}
\epsilon_I &\geq \max_{r\in\mathcal{R}} \sum_{t\in I} \left(U(r,p,y_t) - U(r_I,p,y_t)\right) - K^U_{\mathcal{R}}\delta_{\mathcal{R}}\\
&= \max_{r\in\mathcal{R}} \mathrm{E}_{y\sim\hat{\pi}_I}[U(r,p,y) - U(r_I,p,y)] - K^U_{\mathcal{R}}\delta_{\mathcal{R}}
\end{aligned}
$$

It follows, by regularity and definition of β, that

$$\begin{aligned}\frac{1}{n_I}\sum_{t\in I} V(r_t,p,y_t) &\le \mathrm{E}_{y\sim\hat{\pi}_I}[V(r_I,p,y)] + K^V_{\mathcal{R}}\delta_{\mathcal{R}}\\ &\le \beta_p(\hat{\pi}_I,\epsilon_I + K^U_{\mathcal{R}}\delta_{\mathcal{R}}) + K^V_{\mathcal{R}}\delta_{\mathcal{R}}\end{aligned}$$

Summing over information $I \in \mathcal{I}_F$ and using lemma 15, we obtain

$$\begin{aligned}\frac{1}{n_F}\sum_{I\in\mathcal{I}_F}\sum_{t\in I} V(r_t,p,y_t) &\le \frac{1}{n_F}\sum_{I\in\mathcal{I}_F} n_I\beta_p(\hat{\pi}_I,\epsilon_I + K^U_{\mathcal{R}}\delta_{\mathcal{R}}) + K^V_{\mathcal{R}}\delta_{\mathcal{R}}\\ &\le \frac{1}{n_F}\sum_{I\in\mathcal{I}_F} n_I\left(\beta_p(\hat{\pi}_I,\bar{\epsilon}) + \frac{\epsilon_I + K^U_{\mathcal{R}}\delta_{\mathcal{R}}}{\bar{\epsilon}}\right) + K^V_{\mathcal{R}}\delta_{\mathcal{R}}\\ &= \frac{1}{n_F}\sum_{I\in\mathcal{I}_F} n_I\beta_p(\hat{\pi}_I,\bar{\epsilon}) + \left(\frac{\epsilon_F + K^U_{\mathcal{R}}\delta_{\mathcal{R}}}{\bar{\epsilon}}\right) + K^V_{\mathcal{R}}\delta_{\mathcal{R}}\end{aligned}$$

So far, we have an upper bound for the principal's payoff that nearly matches the principal's worst-case payoff in the stage game if the agent had information structure $\hat{\gamma}_F$ in each forecast context. Furthermore, we know that this information structure cannot be particularly useful to the agent. Define

$$-\tilde{\epsilon}_F = \max_{r\in\mathcal{R}}\frac{1}{n_F}\sum_{I\in\mathcal{I}_F}\sum_{t\in I}\left(U(r,p,y_t) - U(r_t,p,y_t)\right)$$

as the (possibly negative) regret accumulated in forecast context F relative to the best-in-hindsight response, rather than the best-in-hindsight function from information to responses. Let π_F be the forecast associated with forecast context F. Note that

$$\epsilon_F + \tilde{\epsilon}_F = \min_{\tilde{r}\in\mathcal{R}}\frac{1}{n_F}\sum_{I\in\mathcal{I}_F}\max_{r\in\mathcal{R}}\sum_{t\in I}\left(U(r,p,y_t) - U(\tilde{r},p,y_t)\right)$$

It follows from assumption 20 that

$$M_1(\epsilon_F + \tilde{\epsilon}_F) + M_2 \geq \beta_p(\hat{\pi}_F, \hat{\gamma}_F, \bar{\epsilon}) - \max_{\tilde{p}\in\mathcal{P}} \beta_{\tilde{p}}(\hat{\pi}_F, \bar{\epsilon})$$

which can be rewritten as

$$\beta_p(\hat{\pi}_F, \hat{\gamma}_F, \bar{\epsilon}) \leq \max_{\tilde{p}\in\mathcal{P}} \beta_{\tilde{p}} + M_1(\epsilon_F + \tilde{\epsilon}_F) + M_2$$

Next, we relate our upper bound on the principal's payoff to the term $\beta_p(\hat{\pi}_F, \hat{\gamma}_F, \bar{\epsilon})$. Note that

$$\begin{aligned}\frac{1}{n_F}\sum_{I\in\mathcal{I}_F} n_I \beta_p(\hat{\pi}_I, \bar{\epsilon}) &= \mathrm{E}_{y\sim\hat{\pi}_F}\big[\mathrm{E}_{J\sim\hat{\gamma}(\cdot,y)}\big[\beta_p(\hat{\pi}_{F|J}, \bar{\epsilon})\big]\big] \\ &\leq \max_{e_J} \mathrm{E}_{y\sim\hat{\pi}_F}\big[\mathrm{E}_{J\sim\hat{\gamma}(\cdot,y)}\big[\beta_p(\pi_{F|J}, e_J)\big]\big] \quad \text{s.t.} \quad \bar{\epsilon} = \mathrm{E}_{y\sim\pi_F}[\mathrm{E}_{J\sim\cdot,\hat{y}}[e_J]] \\ &= \beta_p(\hat{\pi}_F, \hat{\gamma}_F, \bar{\epsilon})\end{aligned}$$

Collapsing these inequalities gives us

$$\frac{1}{n_F}\sum_{t\in F} V(r_t, p, y_t) \leq \max_{\tilde{p}\in\mathcal{P}} \beta_{\tilde{p}}(\hat{\pi}_F, \bar{\epsilon}) + \left(\frac{\epsilon_F + K^U_{\mathcal{R}}\delta_{\mathcal{R}}}{\bar{\epsilon}}\right) + \big(M_1(\epsilon_F + \tilde{\epsilon}_F) + M_2 + K^V_{\mathcal{R}}\delta_{\mathcal{R}}\big)$$

Summing over forecast contexts $F \in \mathcal{F}$ gives us the desired result. □

From these two lemmas, it immediately follows that

$$\begin{aligned}\text{PR} \leq &\frac{1}{T}\sum_{F\in\mathcal{F}} n_F \Delta(\hat{\pi}_F, \bar{\epsilon}) + \left(\frac{2\epsilon + 6\iota + 2K^U_{\mathcal{R}}\delta_{\mathcal{R}} + 2K^U_{\mathcal{P}}\delta_{\mathcal{P}}}{\bar{\epsilon}}\right) \\ &+ \big(M_1(2\epsilon + 2\tilde{\epsilon} + 2\iota + 2K^U_{\mathcal{P}}\delta_{\mathcal{P}}) + 2M_2 + 3\iota + 2K^V_{\mathcal{R}}\delta_{\mathcal{R}} + K^V_{\mathcal{P}}\delta_{\mathcal{P}}\big)\end{aligned}$$

Therefore, to bound the principal's regret, all that remains is to bound ι. If we use the algorithm from appendix **??** to generate π_t, this follows directly from equation (B.3), which states

$$\mathrm{E}_{\sigma^*}[\iota] \le \sqrt{|\mathcal{Y}||\mathcal{F}|\sqrt{\frac{2\log|\mathcal{F}|}{T}} + 2|\mathcal{Y}|\delta_{\mathcal{F}}}$$

Finally, we obtain our bound on the expected principal's regret.

$$\begin{aligned}\mathrm{E}_{\sigma^*}[\mathrm{PR}] \le & \frac{1}{T}\sum_{F\in\mathcal{F}} n_F \Delta(\hat{\pi}_F, \bar{\epsilon}) + \left(\frac{2\epsilon + 6\sqrt{|\mathcal{Y}||\mathcal{F}|\sqrt{\frac{2\log|\mathcal{F}|}{T}} + 2|\mathcal{Y}|\delta_{\mathcal{F}}} + 2K^U_{\mathcal{R}}\delta_{\mathcal{R}} + 2K^U_{\mathcal{P}}\delta_{\mathcal{P}}}{\bar{\epsilon}}\right) \\ & + M_1\left(2\epsilon + 2\tilde{\epsilon} + 2\sqrt{|\mathcal{Y}||\mathcal{F}|\sqrt{\frac{2\log|\mathcal{F}|}{T}} + 2|\mathcal{Y}|\delta_{\mathcal{F}}} + 2K^U_{\mathcal{P}}\delta_{\mathcal{P}}\right) \\ & + 2M_2 + 3\sqrt{|\mathcal{Y}||\mathcal{F}|\sqrt{\frac{2\log|\mathcal{F}|}{T}} + 2|\mathcal{Y}|\delta_{\mathcal{F}}} + 2K^V_{\mathcal{R}}\delta_{\mathcal{R}} + K^V_{\mathcal{P}}\delta_{\mathcal{P}}\end{aligned}$$

B.5.5 Proof of Theorem 12

Assume access to a forecast π_t for every period t. We will define this later. The mechanism chooses p_t as follows. Let P be the unique policy context that includes the policy $p^{\dagger}(\pi_t, \bar{\epsilon}) \in P$. Let $p_t = p_P$, where $p_P \in \mathcal{P}_1$ is the representative element of P.

The next two lemmas imply an upper bound on the principal's regret in terms of the quantity

$$\iota \ge \frac{1}{T}\sum_{t=1}^{T} d_1(\pi_t, \hat{\pi}_F)$$

that measures the discrepancy between the forecast π_t and the empirical distribution $\hat{\pi}_F$ conditioned on the forecast context F. Lemma 33 is a lower bound on the principal's payoff under σ^*. Lemma 34 is an upper bound on the his payoff under any constant $\sigma^p \in \Sigma_0$.

Lemma 33. *Suppose the principal runs the mechanism σ^*. Then*

$$\begin{aligned}\frac{1}{T}\sum_{t=1}^{T} V(r_t, p_t, y_t) \geq & \frac{1}{T}\sum_{F\in\mathcal{F}} n_F \inf_{\gamma} \max_{p\in\mathcal{P}} \alpha_p(\hat{\pi}_F, \gamma, \bar{\epsilon}) - \left(\frac{\epsilon + 4\iota + K^U_{\mathcal{R}}\delta_{\mathcal{R}} + 2K^U_{\mathcal{P}}\delta_{\mathcal{P}}}{\bar{\epsilon}}\right) \\ & - \left(2\iota + K^V_{\mathcal{R}}\delta_{\mathcal{R}} + K^V_{\mathcal{P}}\delta_{\mathcal{P}}\right)\end{aligned}$$

Proof. Let r_I be a representative element in the response context R associated with information I under mechanism σ^*. By regularity,

$$\begin{aligned}\epsilon_I &\geq \max_{r\in\mathcal{R}} \sum_{t\in I} \left(U(r, p_I, y_t) - U(r_I, p_I, y_t)\right) - K^U_{\mathcal{R}}\delta_{\mathcal{R}} \\ &= \max_{r\in\mathcal{R}} \mathrm{E}_{y\sim\hat{\pi}_I}[U(r, p_I, y) - U(r_I, p_I, y)] - K^U_{\mathcal{R}}\delta_{\mathcal{R}}\end{aligned}$$

It follows, by regularity and definition of α, that

$$\begin{aligned}\frac{1}{n_I}\sum_{t\in I} V(r_t, p_I, y_t) &\geq \mathrm{E}_{y\sim\hat{\pi}_I}[V(r_I, p_I, y)] - K^V_{\mathcal{R}}\delta_{\mathcal{R}} \\ &\geq \alpha_{p_I}(\hat{\pi}_I, \epsilon_I + K^U_{\mathcal{R}}\delta_{\mathcal{R}}) - K^V_{\mathcal{R}}\delta_{\mathcal{R}}\end{aligned}$$

Summing over information $I \in \mathcal{I}_F$, we obtain

$$\begin{aligned}\frac{1}{n_F}\sum_{I\in\mathcal{I}_F}\sum_{t\in I} V(r_t, p_F, y_t) &\geq \frac{1}{n_F}\sum_{I\in\mathcal{I}_F}\sum_{t\in I} \alpha_{p_F}(\hat{\pi}_I, \epsilon_I + K^U_{\mathcal{R}}\delta_{\mathcal{R}}) - K^V_{\mathcal{R}}\delta_{\mathcal{R}} \\ &\geq \frac{1}{n_F}\sum_{I\in\mathcal{I}_F}\sum_{t\in I} \alpha_{p^\dagger(\pi_t,\bar{\epsilon})}(\hat{\pi}_I, \epsilon_I + K^U_{\mathcal{R}}\delta_{\mathcal{R}} + 2K^U_{\mathcal{P}}\delta_{\mathcal{P}}) - K^V_{\mathcal{R}}\delta_{\mathcal{R}} - K^V_{\mathcal{P}}\delta_{\mathcal{P}} \\ &\geq \frac{1}{n_F}\sum_{I\in\mathcal{I}_F}\sum_{t\in I} \left(\alpha_{p^\dagger(\pi_t,\bar{\epsilon})}(\hat{\pi}_I, \bar{\epsilon}) - \frac{\epsilon_I + K^U_{\mathcal{R}}\delta_{\mathcal{R}} + 2K^U_{\mathcal{P}}\delta_{\mathcal{P}}}{\bar{\epsilon}}\right) \\ &- K^V_{\mathcal{R}}\delta_{\mathcal{R}} - K^V_{\mathcal{P}}\delta_{\mathcal{P}} \\ &= \frac{1}{n_F}\sum_{I\in\mathcal{I}_F}\sum_{t\in I} \alpha_{p^\dagger(\pi_t,\bar{\epsilon})}(\hat{\pi}_I, \bar{\epsilon}) - \left(\frac{\epsilon_F + K^U_{\mathcal{R}}\delta_{\mathcal{R}} + 2K^U_{\mathcal{P}}\delta_{\mathcal{P}}}{\bar{\epsilon}}\right)\end{aligned}$$

$$- K^V_{\mathcal{R}}\delta_{\mathcal{R}} - K^V_{\mathcal{P}}\delta_{\mathcal{P}}$$

Focus on the first term, i.e.

$$\begin{aligned}
\frac{1}{n_F}\sum_{I\in\mathcal{I}_F} n_I \alpha_{p^\dagger(\pi_F,\bar\epsilon)}(\hat\pi_I,\bar\epsilon) &= \mathrm{E}_{y\sim\hat\pi_F}\left[\mathrm{E}_{J\sim\hat\gamma(\cdot,y)}\left[\alpha_{p^\dagger(\pi_F,\bar\epsilon)}(\hat\pi_{F|J},\bar\epsilon)\right]\right] \\
&\geq \min_{e_J} \mathrm{E}_{y\sim\hat\pi_F}\left[\mathrm{E}_{J\sim\hat\gamma(\cdot,y)}\left[\alpha_{p^\dagger(\pi_F,\bar\epsilon)}(\hat\pi_{F|J},e_J)\right]\right] \quad \text{s.t.} \quad \bar\epsilon = \mathrm{E}_{y\sim\hat\pi_F}[\mathrm{E}_{J\sim\hat\gamma\cdot,y}[e_J]] \\
&= \alpha_{p^\dagger(\pi_F,\bar\epsilon)}(\hat\pi_F,\hat\gamma_F,\bar\epsilon) \\
&\geq \inf_\gamma \alpha_{p^\dagger(\pi_F,\bar\epsilon)}(\hat\pi_F,\gamma,\bar\epsilon) \\
&\geq \inf_\gamma \alpha_{p^\dagger(\hat\pi_F,\bar\epsilon)}(\hat\pi_F,\gamma,\bar\epsilon) - \frac{4d_1(\pi_F,\hat\pi_F)}{\bar\epsilon} - 2d_1(\pi_F,\hat\pi_F)
\end{aligned}$$

Collapsing these inequalities gives us

$$\begin{aligned}
\frac{1}{n_F}\sum_{I\in\mathcal{I}_F}\sum_{t\in I} V(r_t,p_F,y_t) \geq& \inf_\gamma \alpha_{p^\dagger(\hat\pi_F,\bar\epsilon)}(\hat\pi_F,\gamma,\bar\epsilon) - \left(\frac{\epsilon_F + 4d_1(\pi_F,\hat\pi_F) + K^U_{\mathcal{R}}\delta_{\mathcal{R}} + 2K^U_{\mathcal{P}}\delta_{\mathcal{P}}}{\bar\epsilon}\right) \\
&- \left(2d_1(\pi_F,\hat\pi_F) + K^V_{\mathcal{R}}\delta_{\mathcal{R}} + K^V_{\mathcal{P}}\delta_{\mathcal{P}}\right)
\end{aligned}$$

Summing over forecast contexts $F\in\mathcal{F}$ gives us the desired result. □

Lemma 34. *Suppose the principal runs some constant mechanism $\sigma^p \in \Sigma_0$. Then*

$$\frac{1}{T}\sum_{t=1}^{T} V(r_t,p,y_t) \leq \sum_{F\in\mathcal{F}} n_F \max_{\tilde p\in\mathcal{P}} \beta_{\tilde p}(\hat\pi_F,\hat\gamma_F,\bar\epsilon) + \left(\frac{\epsilon + K^U_{\mathcal{R}}\delta_{\mathcal{R}}}{\bar\epsilon}\right) + K^V_{\mathcal{R}}\delta_{\mathcal{R}}$$

Proof. Let r_I be a representative element in the response context R associated with information I under mechanism σ^p. By regularity,

$$\epsilon_I \geq \max_{r\in\mathcal{R}} \sum_{t\in I} \left(U(r,p,y_t) - U(r_I,p,y_t)\right) - K^U_{\mathcal{R}}\delta_{\mathcal{R}}$$

$$= \max_{r\in\mathcal{R}} \mathrm{E}_{y\sim\hat{\pi}_I}[U(r,p,y) - U(r_I,p,y)] - K^U_{\mathcal{R}}\delta_{\mathcal{R}}$$

It follows, by regularity and definition of β, that

$$\begin{aligned}\frac{1}{n_I}\sum_{t\in I} V(r_t,p,y_t) &\le \mathrm{E}_{y\sim\hat{\pi}_I}[V(r_I,p,y)] + K^V_{\mathcal{R}}\delta_{\mathcal{R}} \\ &\le \beta_p(\hat{\pi}_I, \epsilon_I + K^U_{\mathcal{R}}\delta_{\mathcal{R}}) + K^V_{\mathcal{R}}\delta_{\mathcal{R}}\end{aligned}$$

Summing over information $I \in \mathcal{I}_F$, we obtain

$$\begin{aligned}\frac{1}{n_F}\sum_{I\in\mathcal{I}_F}\sum_{t\in I} V(r_t,p,y_t) &\le \frac{1}{n_F}\sum_{I\in\mathcal{I}_F} n_I\beta_p(\hat{\pi}_I, \epsilon_I + K^U_{\mathcal{R}}\delta_{\mathcal{R}}) + K^V_{\mathcal{R}}\delta_{\mathcal{R}} \\ &\le \frac{1}{n_F}\sum_{I\in\mathcal{I}_F} n_I\left(\beta_p(\hat{\pi}_I,\bar{\epsilon}) + \frac{\epsilon_I + K^U_{\mathcal{R}}\delta_{\mathcal{R}}}{\bar{\epsilon}}\right) + K^V_{\mathcal{R}}\delta_{\mathcal{R}} \\ &= \frac{1}{n_F}\sum_{I\in\mathcal{I}_F} n_I\beta_p(\hat{\pi}_I,\bar{\epsilon}) + \left(\frac{\epsilon_F + K^U_{\mathcal{R}}\delta_{\mathcal{R}}}{\bar{\epsilon}}\right) + K^V_{\mathcal{R}}\delta_{\mathcal{R}}\end{aligned}$$

Focus on the first term, i.e.

$$\begin{aligned}\frac{1}{n_F}\sum_{I\in\mathcal{I}_F} n_I\beta_p(\hat{\pi}_I,\bar{\epsilon}) &= \mathrm{E}_{y\sim\hat{\pi}_F}\left[\mathrm{E}_{J\sim\hat{\gamma}(\cdot,y)}\left[\beta_p(\hat{\pi}_{F|J},\bar{\epsilon})\right]\right] \\ &\le \max_{e_J} \mathrm{E}_{y\sim\hat{\pi}_F}\left[\mathrm{E}_{J\sim\hat{\gamma}(\cdot,y)}\left[\beta_p(\hat{\pi}_{F|J},e_J)\right]\right] \quad \text{s.t.} \quad \bar{\epsilon} = \mathrm{E}_{y\sim\hat{\pi}_F}[\mathrm{E}_{J\sim\hat{\gamma}\cdot,y}[e_J]] \\ &= \beta_p(\hat{\pi}_F,\hat{\gamma}_F,\bar{\epsilon}) \\ &\le \max_{\tilde{p}\in\mathcal{P}} \beta_{\tilde{p}}(\hat{\pi}_F,\hat{\gamma}_F,\bar{\epsilon})\end{aligned}$$

Collapsing these inequalities gives us

$$\frac{1}{n_F}\sum_{I\in\mathcal{I}_F}\sum_{t\in I} V(r_t,p,y_t) \le \max_{\tilde{p}\in\mathcal{P}} \beta_{\tilde{p}}(\hat{\pi}_F,\hat{\gamma}_F,\bar{\epsilon}) + \left(\frac{\epsilon_F + K^U_{\mathcal{R}}\delta_{\mathcal{R}}}{\bar{\epsilon}}\right) + K^V_{\mathcal{R}}\delta_{\mathcal{R}}$$

Summing over forecast contexts $F \in \mathcal{F}$ gives us the desired result. □

From these two lemmas, it immediately follows that

$$\mathrm{PR} \leq \sum_{F \in \mathcal{F}} n_F \nabla(\hat{\pi}_F, \bar{\epsilon}) + 2\left(\frac{\epsilon + 2\iota + K^U_{\mathcal{R}}\delta_{\mathcal{R}} + K^U_{\mathcal{P}}\delta_{\mathcal{P}}}{\bar{\epsilon}}\right) + \left(2\iota + 2K^V_{\mathcal{R}}\delta_{\mathcal{R}} + K^V_{\mathcal{P}}\delta_{\mathcal{P}}\right)$$

Therefore, to bound the principal's regret, all that remains is to bound ι. If we use the algorithm from appendix **??** to generate π_t, this follows directly from equation (B.3), which states

$$\mathrm{E}_{\sigma^*}[\iota] \leq \sqrt{|\mathcal{Y}||\mathcal{F}|\sqrt{\frac{2\log|\mathcal{F}|}{T}}} + 2|\mathcal{Y}|\delta_{\mathcal{F}}$$

Finally, we obtain our bound on the expected principal's regret.

$$\begin{aligned}\mathrm{E}_{\sigma^*}[\mathrm{PR}] \leq & \frac{1}{T}\sum_{F \in \mathcal{F}} n_F \nabla(\hat{\pi}_F, \bar{\epsilon}) + 2\left(\frac{\epsilon + 2\sqrt{|\mathcal{Y}||\mathcal{F}|\sqrt{\frac{2\log|\mathcal{F}|}{T}}} + 2|\mathcal{Y}|\delta_{\mathcal{F}} + K^U_{\mathcal{R}}\delta_{\mathcal{R}} + K^U_{\mathcal{P}}\delta_{\mathcal{P}}}{\bar{\epsilon}}\right) \\ & + \left(2\sqrt{|\mathcal{Y}||\mathcal{F}|\sqrt{\frac{2\log|\mathcal{F}|}{T}}} + 2|\mathcal{Y}|\delta_{\mathcal{F}} + 2K^V_{\mathcal{R}}\delta_{\mathcal{R}} + K^V_{\mathcal{P}}\delta_{\mathcal{P}}\right)\end{aligned}$$

B.5.6 Proof of Theorem 5

We adapt the proof of theorem 10 to prove theorem 5. This will require only relatively minor changes. Let $\hat{\pi}_{I,F}$ denote the empirical distribution among periods $t \in I \cap F$. Let $n_{I,F}$ indicate the number of such periods. Let π_F denote the (unique) forecast associated with forecast context F. Previously, we defined

$$\iota \geq \frac{1}{T}\sum_{t=1}^{T} d_1(\pi_t, \hat{\pi}_I)$$

Now, we define

$$\iota \geq \frac{1}{T} \sum_{t=1}^{T} d_1(\pi_t, \hat{\pi}_{I,F})$$

Begin at the last line of lemma 29, where it says "focus on the first term". Rewrite that first term as

$$\frac{1}{T} \sum_{I \in \mathcal{I}} \sum_{F \in \mathcal{F}} \sum_{t \in I \cap F} \alpha_{p^*(\pi_F, \bar{\epsilon})}(\hat{\pi}_I, \bar{\epsilon})$$

Now we switch $\hat{\pi}_I$ with π_I, i.e. the convex combination of forecasts,

$$\pi_I = \frac{1}{n_I} \sum_{F \in \mathcal{F}} n_{I,F} \pi_F$$

By lemma 26, this gives us

$$\frac{1}{T} \sum_{I \in \mathcal{I}} \sum_{F \in \mathcal{F}} \sum_{t \in I \cap F} \alpha_{p^*(\pi_F, \bar{\epsilon})}(\hat{\pi}_I, \bar{\epsilon})$$

$$\geq \frac{1}{T} \sum_{I \in \mathcal{I}} \sum_{F \in \mathcal{F}} \sum_{t \in I \cap F} \left(\alpha_{p^*(\pi_F, \bar{\epsilon})}(\pi_I, \bar{\epsilon}) - \frac{2 d_1(\pi_I, \hat{\pi}_I)}{\bar{\epsilon}} - d_1(\pi_I, \hat{\pi}_I) \right)$$

Note that every forecast π_F leads to a policy $p^*(\pi_F, \bar{\epsilon})$ that is in the policy context P associated with information I. By assumption **??**,

$$\geq \frac{1}{T} \sum_{I \in \mathcal{I}} \sum_{F \in \mathcal{F}} \sum_{t \in I \cap F} \left(\alpha_{p^*(\pi_I, \bar{\epsilon})}(\pi_I, \bar{\epsilon}) - \frac{2 d_1(\pi_I, \hat{\pi}_I)}{\bar{\epsilon}} - d_1(\pi_I, \hat{\pi}_I) - O(\delta_{\mathcal{P}}) \right)$$

Now we apply lemma 26 again,

$$\geq \frac{1}{T} \sum_{I \in \mathcal{I}} \sum_{F \in \mathcal{F}} \sum_{t \in I \cap F} \left(\alpha_{p^*(\pi_I, \bar{\epsilon})}(\hat{\pi}_I, \bar{\epsilon}) - \frac{4 d_1(\pi_I, \hat{\pi}_I)}{\bar{\epsilon}} - 2 d_1(\pi_I, \hat{\pi}_I) - O(\delta_{\mathcal{P}}) \right)$$

and then lemma 27

$$\geq \frac{1}{T}\sum_{I\in\mathcal{I}}\sum_{F\in\mathcal{F}}\sum_{t\in I\cap F}\left(\alpha_{p^*(\hat{\pi}_I,\bar{\epsilon})}(\hat{\pi}_I,\bar{\epsilon}) - \frac{8d_1(\pi_I,\hat{\pi}_I)}{\bar{\epsilon}} - 4d_1(\pi_I,\hat{\pi}_I) - O(\delta_{\mathcal{P}})\right)$$

By the homogeneity and subadditivity of the l_1 norm,

$$d_1(\pi_I,\hat{\pi}_I) \leq \frac{1}{n_I}\sum_{i=1}^{n} n_{I,F} d_1(\pi_{I,F},\hat{\pi}_{I,F})$$

which gives us

$$\geq \frac{1}{T}\sum_{I\in\mathcal{I}}\sum_{F\in\mathcal{F}}\sum_{t\in I\cap F}\left(\alpha_{p^*(\hat{\pi}_I,\bar{\epsilon})}(\hat{\pi}_I,\bar{\epsilon}) - \frac{8d_1(\pi_{I,F},\hat{\pi}_{I,F})}{\bar{\epsilon}} - 4d_1(\pi_{I,F},\hat{\pi}_{I,F}) - O(\delta_{\mathcal{P}})\right)$$

$$\geq \frac{1}{T}\sum_{I\in\mathcal{I}}\sum_{F\in\mathcal{F}}\sum_{t\in I\cap F}\left(\alpha_{p^*(\hat{\pi}_I,\bar{\epsilon})}(\hat{\pi}_I,\bar{\epsilon})\right) - \frac{8\iota}{\bar{\epsilon}} - 4\iota - O(\delta_{\mathcal{P}})$$

This is essentially where we were by the end of lemma 29, with the addition of an $O(\delta_{\mathcal{P}})$ term and slightly different constants.

Lemma 30 requires no change. The discussion following lemma 30 requires very little change. Find the line that begins with “Consider any two periods”. We rewrite as follows. Consider any two periods t,τ where $I_t = I_\tau$ and $F_t = F_\tau$ but $C_t \neq C_\tau$. Since $F_t = F_\tau$ we know that $\pi_t = \pi_\tau$. Now, consider

$$n_{F_t,C_t} d_1(\pi_t,\hat{\pi}_{C_t}) + n_{F_\tau,C_\tau} d_1(\pi_\tau,\hat{\pi}_{C_\tau})$$

$$= n_{F_t,C_t} d_1(\pi_t,\hat{\pi}_{C_t}) + n_{F_t,C_\tau} d_1(\pi_t,\hat{\pi}_{C_\tau})$$

$$\geq \left(n_{F_t,C_t} + n_{F_t,C_\tau}\right) d_1\left(\pi_t, \frac{1}{n_{F_t,C_t} + n_{F_t,C_\tau}}\left(n_{F_t,C_t}\hat{\pi}_{C_t} + n_{F_t,C_\tau} d_1(\pi_t,\hat{\pi}_{C_\tau})\right)\right)$$

by subadditivity and homogeneity of norms. By continuing this process of combining contexts, we find

$$\frac{1}{T}\sum_{C\in\mathcal{C}}\sum_{F\in\mathcal{F}} n_{F,C} d_1(\pi_t, \hat{\pi}_C) \geq \frac{1}{T}\sum_{I\in\mathcal{I}}\sum_{F\in\mathcal{F}} n_{I,F} d_1(\pi_t, \hat{\pi}_{I,F}) = \iota$$

Therefore, our bound holds except with the addition of an $O(\delta_{\mathcal{P}})$ term and slightly different constants.

B.5.7 Proof of Theorem 6

Define

$$\pi_P = \frac{1}{n_P}\sum_{F\in\mathcal{F}} n_{P,F}\pi_F$$

It is straightforward to adapt the proof of theorem 11. Replace all reference to $p^*(\pi_F, \bar{\epsilon})$ with $p^*(\pi_P, \bar{\epsilon})$. This changes U and V (and all derived terms, like α) by at most $O(\delta_{\mathcal{P}})$, by assumption **??**. Replace all remaining references of forecast contexts F to policy contexts P. It remains to verify that

$$\iota \geq \frac{1}{T}\sum_{F\in\mathcal{F}}\sum_{t\in F} d_1(\pi_F, \hat{\pi}_F) \geq \frac{1}{T}\sum_{P\in\mathcal{P}}\sum_{t\in F} d_1(\pi_P, \hat{\pi}_P)$$

which follows from the homogeneity and subadditivity of the l_1 norm, and the fact that $\pi_P, \hat{\pi}_P$ are convex combinations of $\pi_F, \hat{\pi}_F$ for $F \subseteq P$.

B.5.8 Proof of Theorem 7

Define

$$\pi_P = \frac{1}{n_P}\sum_{F\in\mathcal{F}} n_{P,F}\pi_F$$

It is straightforward to adapt the proof of theorem 12. Replace all reference to $p^\dagger(\pi_F, \bar{\epsilon})$ with $p^\dagger(\pi_P, \bar{\epsilon})$. This changes U and V (and all derived terms, like α) by at most $O(\delta_{\mathcal{P}})$, by assumption **??**. Replace all remaining references of forecast contexts F to policy contexts P. It remains to verify that

$$\iota \geq \frac{1}{T} \sum_{F \in \mathcal{F}} \sum_{t \in F} d_1(\pi_F, \hat{\pi}_F) \geq \frac{1}{T} \sum_{P \in \mathcal{P}} \sum_{t \in F} d_1(\pi_P, \hat{\pi}_P)$$

which follows from the homogeneity and subadditivity of the l_1 norm, and the fact that $\pi_P, \hat{\pi}_P$ are convex combinations of $\pi_F, \hat{\pi}_F$ for $F \subseteq P$.

Appendix C

Omissions from Chapter 3

C.1 Proof of Proposition 8

This proof will be slightly more general than the proposition statement. Let $f : \mathcal{S}^n \to \mathbb{R}_+$ be an arbitrary function with upper bound $\bar{f}$. Suppose that there is an upper bound

$$\mathrm{E}_{\pi^s}[f(S_1, \ldots, S_n)] \leq B$$

The goal is to find a similar upper bound on

$$\mathrm{E}_{\pi^s}[f(S_1, \ldots, S_n) \mid P_n = p]$$

assuming that P_n is (ϵ, δ)-private.

First, I use the privacy property to show that bound the expected value of f conditional on the realized policy P_n and the event E in the definition of sample privacy. I show that this is not very different from the expected value of f conditioned on just the event E.

$$\begin{aligned}
&\mathrm{E}_{\pi^s}[f(S_1, \ldots, S_n) \mid P_n = p, E] \\
&= \sum_{(S_1, \ldots, S_n) \in E} \mathrm{Pr}_{\pi^s}[S_1, \ldots, S_n \mid P_n = p, E] \cdot f(S_1, \ldots, S_n) \\
&= \sum_{(S_1, \ldots, S_n) \in E} \frac{\mathrm{Pr}_{\pi^s}[S_1, \ldots, S_n \mid E] \cdot \mathrm{Pr}_{\pi^s}[P_n = p \mid S_1, \ldots, S_n]}{\mathrm{Pr}_{\pi^s}[P_n = p \mid E]} \cdot f(S_1, \ldots, S_n)
\end{aligned}$$

$$
\begin{aligned}
&\leq \sum_{(S_1,\ldots,S_n)\in E} \frac{\Pr_{\pi^s}[S_1,\ldots,S_n \mid E]\cdot e^{\epsilon}\cdot \Pr_{\pi^s}[P_n = p \mid E]}{\Pr_{\pi^s}[P_n = p \mid E]} \cdot f\left(S_1,\ldots,S_n\right) \\
&= \sum_{(S_1,\ldots,S_n)\in E} e^{\epsilon}\cdot \Pr_{\pi^s}[S_1,\ldots,S_n \mid E]\cdot f\left(S_1,\ldots,S_n\right) \\
&= e^{\epsilon}\cdot \mathrm{E}_{\pi^s}[f\left(S_1,\ldots,S_n\right) \mid E]
\end{aligned}
$$

Next, I use the upper bound to show that

$$
\begin{aligned}
B &\geq \mathrm{E}_{\pi^s}[f\left(S_1,\ldots,S_n\right)] \\
&= (1-\delta)\cdot \mathrm{E}_{\pi^s}[f\left(S_1,\ldots,S_n\right) \mid E] + \delta\cdot \mathrm{E}_{\pi^s}[f\left(S_1,\ldots,S_n\right) \mid \neg E] \\
&\geq (1-\delta)\cdot \mathrm{E}_{\pi^s}[f\left(S_1,\ldots,S_n\right) \mid E] \\
&\geq (1-\delta)\cdot e^{-\epsilon}\cdot \mathrm{E}_{\pi^s}[f\left(S_1,\ldots,S_n\right) \mid E, P_n = p] \\
&\geq (1-\delta)\cdot e^{-\epsilon}\cdot \mathrm{E}_{\pi^s}[f\left(S_1,\ldots,S_n\right) \mid E, P_n = p] \\
&\quad + \delta\cdot e^{-\epsilon}\cdot \mathrm{E}_{\pi^s}\left[f\left(S_1,\ldots,S_n\right) - \bar{f} \mid \neg E, P_n = p\right] \\
&= e^{-\epsilon}\left(\mathrm{E}_{\pi^s}[f\left(S_1,\ldots,S_n\right) \mid P_n = p] - \delta\cdot \bar{f}\right)
\end{aligned}
$$

Finally, I rearrange the lower bound on B to obtain the desired result, i.e.

$$
\mathrm{E}_{\pi^s}[f\left(S_1,\ldots,S_n\right) \mid P_n = p] \leq e^{\epsilon}\cdot B + \delta\cdot \bar{f}
$$

C.2 Proof of Lemma 17

I want to show that

$$
\begin{aligned}
4\mathcal{RC}_n^A(p,\pi^s) + \mathrm{BFR}_n \geq \Big| &\left(\max_{r'} \mathrm{E}_{\hat{\pi}^s}\left[u^A\left(p,r',s\right)\right] - \mathrm{E}_{\hat{\pi}^s,\pi^r}\left[u^A\left(p,r,s\right)\right]\right) \\
&- \left(\max_{r'} \mathrm{E}_{\pi^s}\left[u^A\left(p,r',s\right)\right] - \mathrm{E}_{\pi^s,\pi^r}\left[u^A\left(p,r,s\right)\right]\right) \Big|
\end{aligned}
$$

for all mixed responses π^r and policies p, with probability no less than $1 - n_{\mathcal{P}} \exp(-n^\alpha)$. For this purpose, it suffices to bound two quantities. First, observe that

$$\mathrm{E}_{\pi^s,\pi^r}\left[u^A\left(p,r,s\right)\right] - \mathrm{E}_{\hat{\pi}^s_n,\pi^r}\left[u^A\left(p,r,s\right)\right] \leq 2\overline{\mathcal{RC}}^A_n(p) + 4\sqrt{\frac{2\ln(4/\kappa)}{n}} \tag{C.1}$$

with probability $1 - \kappa$. This is the typical way of expressing regret bounds based on the Rademacher complexity (see e.g. Bartlett and Mendelson 2003). Second, observe that

$$\max_{r'} \mathrm{E}_{\hat{\pi}^s_n}\left[u^A\left(p,r',s\right)\right] - \max_{r'} \mathrm{E}_{\pi^s}\left[u^A\left(p,r',s\right)\right] \leq 2\overline{\mathcal{RC}}^A_n(p) + 4\sqrt{\frac{2\ln(4/\kappa)}{n}} \tag{C.2}$$

This follows from the fact that

$$\mathrm{E}_{\hat{\pi}^s_n}\left[u^A\left(p,r^*,s\right)\right] - \mathrm{E}_{\pi^s}\left[u^A\left(p,r^*,s\right)\right] \leq 2\overline{\mathcal{RC}}^A_n(p) + 4\sqrt{\frac{2\ln(4/\kappa)}{n}}$$

where $r^* \in \arg\max_r \mathrm{E}_{\hat{\pi}^s_n}\left[u^A\left(p,r,s\right)\right]$, and

$$\mathrm{E}_{\pi^s}\left[u^A\left(p,r^*,s\right)\right] - \max_{r'} \mathrm{E}_{\pi^s}\left[u^A\left(p,r',s\right)\right] \leq 0$$

so

$$\begin{aligned}&\mathrm{E}_{\hat{\pi}^s_n}\left[u^A\left(p,r^*,s\right)\right] - \mathrm{E}_{\pi^s}\left[u^A\left(p,r^*,s\right)\right] \\ &+ \mathrm{E}_{\pi^s}\left[u^A\left(p,r^*,s\right)\right] - \max_{r'} \mathrm{E}_{\pi^s}\left[u^A\left(p,r',s\right)\right] \leq 2\overline{\mathcal{RC}}^A_n(p) + 4\sqrt{\frac{2\ln(4/\kappa)}{n}}\end{aligned}$$

Adding together the last two inequalities gives inequality (C.2). Adding together inequalities (C.1) and (C.2) gives the desired result. Applying the union bound across all policies p, the result holds with probability $1 - n_{\mathcal{P}}\kappa$. Furthermore, the probability is uniform over all responses r, and therefore uniform across all mixed responses π^r. All that remains is to derive the probability κ and buffer BFR_n. Set $\kappa = \exp(-n^\alpha)$. Note that

$$\begin{aligned}
4\sqrt{\frac{2\ln(4/\kappa)}{n}} &= 4\sqrt{\frac{2\ln(4\exp(-n^\alpha))}{n}} \\
&= 4\sqrt{\frac{2\ln 4 - 2\ln(\exp(-n^\alpha))}{n}} \\
&\leq 4\sqrt{\frac{2\ln 4}{n}} + 4\sqrt{-\frac{2\ln(\exp(-n^\alpha))}{n}} \\
&\leq \mathrm{BFR}_n
\end{aligned}$$

C.3 Proof of Lemma 19

Let $E \subseteq \mathcal{S}^n$ be the set of all sample realizations $S_1, \ldots, S_n$ where

$$\widehat{\mathrm{WC}}_n(p) - \mathrm{E}\left[\widehat{\mathrm{WC}}_n(p)\right] \leq t$$

By lemma 18, $\Pr_{\pi^s}[E] \geq 1 - \delta$ where

$$\delta = \exp\left(-\frac{2t^2}{nc^2}\right)$$

To establish sample privacy, I need to show that

$$\Pr_{\pi^s}\left[\hat{P}_n = p \mid S_1, \ldots, S_n\right] \leq e^\epsilon \cdot \Pr_{\pi^s}\left[\hat{P}_n = p, E\right]$$

for any sample realizations $(S_1, \ldots, S_n) \in E$. Let the sample $(S'_1, \ldots, S'_n) \in E$ minimize

$$\Pr_{\pi^s}\left[\hat{P}_n = p \mid S'_1, \ldots, S'_n\right]$$

so that it suffices to show

$$\Pr_{\pi^s}\left[\hat{P}_n = p \mid S_1, \ldots, S_n\right] \leq e^{\epsilon} \cdot \Pr_{\pi^s}\left[\hat{P}_n = p \mid S_1', \ldots, S_n'\right] \tag{C.3}$$

I can characterize the distribution of $\hat{P}_n$ using standard results that link Gumbel error terms with exponential weights.

$$\Pr_{\pi^s}\left[\hat{P}_n = p \mid S_1, \ldots, S_n\right] = \frac{\exp\left(n^{\beta} \cdot \widehat{\mathrm{WC}}_n(p \mid S_1 \ldots, S_n)\right)}{\sum_{p'} \exp\left(n^{\beta} \cdot \widehat{\mathrm{WC}}_n(p' \mid S_1 \ldots, S_n)\right)}$$

Using this, I can rewrite equation (C.3) and manipulate it as follows.

$$\begin{aligned}
&\frac{\Pr_{\pi^s}\left[\hat{P}_n = p \mid S_1, \ldots, S_n\right]}{\Pr_{\pi^s}\left[\hat{P}_n = p \mid S_1', \ldots, S_n'\right]} \\
&= \frac{\exp\left(n^{\beta} \cdot \widehat{\mathrm{WC}}_n(p \mid S_1 \ldots, S_n)\right)}{\exp\left(n^{\beta} \cdot \widehat{\mathrm{WC}}_n(p \mid S_1' \ldots, S_n')\right)} \cdot \frac{\sum_{p'} \exp\left(n^{\beta} \cdot \widehat{\mathrm{WC}}_n(p' \mid S_1' \ldots, S_n')\right)}{\sum_{p'} \exp\left(n^{\beta} \cdot \widehat{\mathrm{WC}}_n(p' \mid S_1 \ldots, S_n)\right)} \\
&\leq \exp\left(tn^{\beta}\right) \cdot \frac{\sum_{p'} \exp\left(n^{\beta} \cdot \widehat{\mathrm{WC}}_n(p' \mid S_1' \ldots, S_n')\right)}{\sum_{p'} \exp\left(n^{\beta} \cdot \widehat{\mathrm{WC}}_n(p' \mid S_1 \ldots, S_n)\right)} \\
&\leq \exp\left(tn^{\beta}\right) \cdot \exp\left(tn^{\beta}\right) \cdot \frac{\sum_{p'} \exp\left(n^{\beta} \cdot \widehat{\mathrm{WC}}_n(p' \mid S_1 \ldots, S_n)\right)}{\sum_{p'} \exp\left(n^{\beta} \cdot \widehat{\mathrm{WC}}_n(p' \mid S_1 \ldots, S_n)\right)} \\
&\leq \exp\left(2tn^{\beta}\right)
\end{aligned}$$

Therefore, $\hat{P}_n$ is (ϵ, δ)-private when $\epsilon = 2tn^{\beta}$.

C.4 Proof of Lemma 21

Recall the definition of $\mathrm{WC}(p, b, \pi^s)$ (3.17). This represents the policymaker's worst-case utility when the agent's regret is bounded by a constant $b \geq 0$. Note that

$$\widehat{\mathrm{WC}}_n(p) = \mathrm{WC}(p, b, \pi^s) \quad \text{where} \quad b = (4e^\epsilon + 4) \cdot \overline{\mathcal{RC}}_n^A(p) + \delta + \mathrm{BFR}_n$$

Let $\hat{\pi}^s$ be the empirical distribution. Let $\tilde{\pi}^s$ be a modified empirical distribution where $S_i = s'$ instead of $S_i = s$. As I shift from $\hat{\pi}^s$ to $\tilde{\pi}^s$, the agent's empirical regret changes by at most $2\Delta^P(p) \cdot n^{-1}$. In particular, for any mixed response π^r,

$$\max_{r'} \mathrm{E}_{\hat{\pi}^s}\left[u^A\left(p, r', s\right)\right] - \mathrm{E}_{\hat{\pi}^s, \pi^r}\left[u^A\left(p, r, s\right)\right] \leq b$$

implies

$$\max_{r'} \mathrm{E}_{\tilde{\pi}^s}\left[u^A\left(p, r', s\right)\right] - \mathrm{E}_{\tilde{\pi}^s, \pi^r}\left[u^A\left(p, r, s\right)\right] \leq b + 2\Delta^P(p) \cdot n^{-1}$$

Likewise, the policymaker's empirical utility changes by at most $\Delta^P(p) \cdot n^{-1}$. It follows from these two observations that

$$\widehat{\mathrm{WC}}_n(p \mid \hat{\pi}^s) \geq \mathrm{WC}\big(p, b + 2\Delta^P(p) \cdot n^{-1}, \tilde{\pi}^s\big) - \Delta^P(p) \cdot n^{-1}$$

where the notation $\widehat{\mathrm{WC}}_n(p \mid \hat{\pi}^s)$ is used to emphasize that $\widehat{\mathrm{WC}}_n(p)$ is being evaluated with respect to the empirical distribution $\hat{\pi}^s$. By the robustness lemma (20),

$$\mathrm{WC}\big(p, b + 2\Delta^P(p) \cdot n^{-1}, \tilde{\pi}^s\big) \geq \mathrm{WC}(p, b, \tilde{\pi}^s) - \Delta^A(p)\left(\frac{2\Delta^P(p) \cdot n^{-1}}{b}\right)$$

By definition, $\widehat{\mathrm{WC}}_n(p \mid \tilde{\pi}^s) = \mathrm{WC}(p, b, \tilde{\pi}^s)$. It follows that

$$\widehat{\mathrm{WC}}_n(p \mid \tilde{\pi}^s) - \widehat{\mathrm{WC}}_n(p \mid \hat{\pi}^s) \leq \Delta^A(p) \left(\frac{2\Delta^P(p) \cdot n^{-1}}{b} \right) + \Delta^P(p) \cdot n^{-1}$$

Therefore, $\widehat{\mathrm{WC}}_n(p)$ satisfies the bounded differences property as long as

$$c \geq \Delta^A(p) \left(\frac{2\Delta^P(p) \cdot n^{-1}}{b} \right) + \Delta^P(p) \cdot n^{-1}$$

C.5 Proof of Lemma 22

I begin by making some observations and introducing some notation. Recall the empirical regret bound in the definition of $\widehat{\mathrm{WC}}_n(p)$ (3.13).

$$\max_{r'} \mathrm{E}_{\hat{\pi}^s} \left[u^A\left(p, r', s\right) \right] - \mathrm{E}_{\hat{\pi}^s, \pi^r} \left[u^A\left(p, r, s\right) \right] \leq (4e^{\epsilon} + 4) \cdot \overline{\mathcal{RC}}_n^A(p) + \delta_n \cdot \Delta^A(p) + \mathrm{BFR}_n \quad \text{(C.4)}$$

By lemma 17, any mixed response that satisfies this bound also satisfies

$$\max_{r'} \mathrm{E}_{\pi^s} \left[u^A\left(p, r', s\right) \right] - \mathrm{E}_{\pi^s, \pi^r} \left[u^A\left(p, r, s\right) \right] \leq 4e^{\epsilon} \cdot \overline{\mathcal{RC}}_n^A(p) + \delta_n \cdot \Delta^A(p)$$

with probability $1 - n_{\mathcal{P}} \exp(-n^{\alpha})$, where the expectations are evaluated with respect to the true distribution π^s. This gives an upper bound for $\widehat{\mathrm{WC}}_n(p)$ with high probability, i.e.

$$\begin{aligned} f(p, \hat{\pi}^s) = &\min_{\pi^r} \mathrm{E}_{\hat{\pi}^s, \pi^r} \left[u^P\left(p, r, s\right) \right] + \nu_n(p) \\ &\text{s.t.} \quad \max_{r'} \mathrm{E}_{\pi^s} \left[u^A\left(p, r', s\right) \right] - \mathrm{E}_{\pi^s, \pi^r} \left[u^A\left(p, r, s\right) \right] \leq 4e^{\epsilon} \cdot \overline{\mathcal{RC}}_n^A(p) + \delta_n \cdot \Delta^A(p) \end{aligned} \quad \text{(C.5)}$$

Let $\tilde{\pi}^r(p, \hat{\pi}^s)$ be the solution to the minimization problem (C.5). It is important to note that the set of feasible mixed responses no longer depends on the sample.

This proof consists of three parts. First, I want to bound the expected gap between $f(\hat{P}_n, \hat{\pi}^s)$ and $\widehat{\mathrm{WC}}_n\left(\hat{P}_n\right)$. It follows from the preceding discussion that

$$\mathrm{E}_{\pi^s}\left[f(\hat{P}_n, \hat{\pi}^s) - \widehat{\mathrm{WC}}_n\left(\hat{P}_n\right)\right] \geq -n_{\mathcal{P}} \exp(-n^{\alpha}) \cdot \max_p \Delta^P(p)$$

Next, I want to bound the expected gap between $f(p, \hat{\pi}^s)$ and $f(p, \pi^s)$, i.e.

$$\begin{aligned}
&\mathrm{E}_{\pi^s}\left[\max_p \left(f(p, \hat{\pi}^s) - f(p, \pi^s)\right)\right] \\
&= \ \mathrm{E}_{\pi^s}\left[\max_p \left(\mathrm{E}_{\hat{\pi}^s, \tilde{\pi}^r(p,\hat{\pi}^s)}\left[u^P\left(p, r, s\right)\right] - \mathrm{E}_{\pi^s, \tilde{\pi}^r(p,\pi^s)}\left[u^P\left(p, r, s\right)\right]\right)\right] \\
&\leq \ \mathrm{E}_{\pi^s}\left[\max_p \left(\mathrm{E}_{\hat{\pi}^s, \tilde{\pi}^r(p,\hat{\pi}^s)}\left[u^P\left(p, r, s\right)\right] - \mathrm{E}_{\hat{\pi}^s, \tilde{\pi}^r(p,\pi^s)}\left[u^P\left(p, r, s\right)\right]\right)\right] \\
&+ \mathrm{E}_{\pi^s}\left[\max_p \left(\mathrm{E}_{\hat{\pi}^s, \tilde{\pi}^r(p,\pi^s)}\left[u^P\left(p, r, s\right)\right] - \mathrm{E}_{\pi^s, \tilde{\pi}^r(p,\pi^s)}\left[u^P\left(p, r, s\right)\right]\right)\right] \\
&\leq \ \mathrm{E}_{\pi^s}\left[\max_p \left(\mathrm{E}_{\hat{\pi}^s, \tilde{\pi}^r(p,\pi^s)}\left[u^P\left(p, r, s\right)\right] - \mathrm{E}_{\pi^s, \tilde{\pi}^r(p,\pi^s)}\left[u^P\left(p, r, s\right)\right]\right)\right] \\
&\leq \ \mathrm{E}_{\pi^s}\left[\max_{p,\pi^r} \left(\mathrm{E}_{\hat{\pi}^s, \pi^r}\left[u^P\left(p, r, s\right)\right] - \mathrm{E}_{\pi^s, \pi^r}\left[u^P\left(p, r, s\right)\right]\right)\right] \\
&= \ \mathrm{E}_{\pi^s}\left[\max_{p,r} \left(\mathrm{E}_{\hat{\pi}^s}\left[u^P\left(p, r, s\right)\right] - \mathrm{E}_{\pi^s}\left[u^P\left(p, r, s\right)\right]\right)\right]
\end{aligned}$$

At this point, it follows from the standard symmetrization argument that

$$\mathrm{E}_{\pi^s}\left[\max_p \left(f(p, \hat{\pi}^s) - f(p, \pi^s)\right)\right] \leq 2\mathcal{RC}_n^P(\pi^s)$$

Finally, I want to bound the expected gap between $\mathrm{WC}_n(p, \epsilon, \delta_n, \pi^s)$ and $f(\hat{P}_n, \pi^s)$. Note that $\mathrm{WC}_n(p, \epsilon, \delta_n, \pi^s) = f(p, \pi^s) - \nu_n(p)$. Furthermore, note that

$$\begin{aligned}
\mathrm{E}_{\pi^s}\left[f(\hat{P}_n, \pi^s) - \mathrm{WC}_n\left(\hat{P}_n, \epsilon, \delta_n, \pi^s\right)\right] &= \mathrm{E}_{\pi^s}\left[\nu_n(\hat{P}_n)\right] \\
&\leq \mathrm{E}\left[\max_p \nu_n(p)\right]
\end{aligned}$$

$$
\begin{aligned}
&\le \mathrm{E}\left[\sum_p |\nu_n(p)|\right] \\
&= n_{\mathcal{P}} \mathrm{E}[|\nu_n(p)|] \\
&\le n_{\mathcal{P}} \sqrt{\mathrm{E}[|\nu_n(p)|^2]} \\
&\le n_{\mathcal{P}} \sqrt{\mathrm{E}[\nu_n(p)]^2 + \mathrm{Var}[\nu_n(p)]} \\
&\le n_{\mathcal{P}} \sqrt{n^{-2\beta} + 2n^{-2\beta}} \\
&\le n_{\mathcal{P}} \sqrt{3} \cdot n^{-\beta}
\end{aligned}
$$

Combining these three steps gives us the desired result.

$$
\begin{aligned}
&\mathrm{E}_{\pi^s}\left[\mathrm{WC}_n\left(\hat{P}_n, \epsilon, \delta_n, \pi^s\right) - \widehat{\mathrm{WC}}_n\left(\hat{P}_n\right)\right] \\
&= \mathrm{E}_{\pi^s}\left[\mathrm{WC}_n\left(\hat{P}_n, \epsilon, \delta_n, \pi^s\right) - f(\hat{P}_n, \pi^s)\right] - \mathrm{E}_{\pi^s}\left[f(\hat{P}_n, \hat{\pi}^s) - f(\hat{P}_n, \pi^s)\right] \\
&+ \mathrm{E}_{\pi^s}\left[f(\hat{P}_n, \hat{\pi}^s) - \widehat{\mathrm{WC}}_n\left(\hat{P}_n\right)\right] \\
&\ge -n_{\mathcal{P}}\sqrt{3} \cdot n^{-\beta} - \mathrm{E}_{\pi^s}\left[\max_p \left(f(p, \hat{\pi}^s) - f(p, \pi^s)\right)\right] - n_{\mathcal{P}} \exp(-n^{\alpha}) \cdot \max_p \Delta^P(p) \\
&\ge -n_{\mathcal{P}}\sqrt{3} \cdot n^{-\beta} - 2\mathcal{RC}_n^P(\pi^s) - n_{\mathcal{P}} \exp(-n^{\alpha}) \cdot \max_p \Delta^P(p)
\end{aligned}
$$

C.6 Proof of Proposition 10

This example will exhibit a simple game where

$$
\mathrm{OP}_n(\pi^s) = \mathrm{CK}(\pi^s) - \Omega\left(n^{-\gamma}\right)
$$

The agent faces an estimation problem and cares about her accuracy. The policy space is a singleton; it is irrelevant. The state space $\mathcal{S} = [0, 1]$ is the unit interval. The response space $\mathcal{R} = [0, 1]$ is also the unit interval. The agent's response $r \in [0, 1]$ is a prediction, subject to square loss, i.e.

$$u^A(p, r, s) = -(r - s)^2$$

The policymaker cares about the agent's accuracy with respect to a bliss point $s_0 \in [0, 1]$ that I will specify later. However, his sensitivity to inaccuracy is different from the agent, i.e.

$$u^P(p, r, s) = -|r - s_0|^{2\gamma}$$

I claim that there exists a distribution $\tilde{\pi}^s$ where the agent's regret bound is $\Omega(n^{-1})$. Let the bliss point $s_0 := \mathrm{E}_{\tilde{\pi}^s}[s]$ be the mean of s according to $\tilde{\pi}^s$. Let the distribution $\pi^s := \tilde{\pi}^s$. Existence follows from two observations. First, the mean square error of the maximum likelihood estimator is $O(n^{-1})$. Second, the maximum likelihood estimator is known be efficient.

To characterize the optimal penalized benchmark, I need to consider responses that satisfy the agent's regret bound. One such response is is $r_n = \mathrm{E}_{\tilde{\pi}^s}[s] + \Omega(n^{-1/2})$. The policymaker's expected utility under r_n must be at least as large as the optimal penalized benchmark, which is the worst case expected utility. That is,

$$\begin{aligned}
\mathrm{OP}_n(\pi^s) &\leq -|r_n - s_0|^{2\gamma} \\
&= -\left|\mathrm{E}_{\pi^s}[s] + \Omega(n^{-1/2}) - \mathrm{E}_{\tilde{\pi}^s}[s]\right|^{2\gamma} \\
&= -\left(\Omega(n^{-1/2})\right)^{2\gamma} \\
&= -\Omega(n^{-\gamma}) \qquad \text{(C.6)}
\end{aligned}$$

Next, consider the common knowledge benchmark. The agent will predict the mean, $r = \mathrm{E}_{\pi^s}[s]$, and the policymaker's expected utility will be

$$\mathrm{CK}(\pi^s) = |\mathrm{E}_{\pi^s}[s] - s_0|^{2\gamma} = |\mathrm{E}_{\tilde{\pi}^s}[s] - \mathrm{E}_{\tilde{\pi}^s}[s]|^{2\gamma} = 0 \qquad \text{(C.7)}$$

I can combine equations (C.6) and (C.7) to show

$$\mathrm{OP}_n(\pi^s) \leq \mathrm{CK}(\pi^s) - \Omega(n^{-\gamma})$$

This completes the first part of the proof.

Next, I need to verify that this game satisfies (in particular) assumption 16. To do this, I need to introduce the pseudodimension: a method for bounding the Rademacher complexity. The following definition is specialized to the agent's Rademacher complexity.

Definition 56. *A vector $(w_1, \ldots, w_n) \in \mathbb{R}^n$ is a* witness *for a vector $(S_1, \ldots, S_n)$ if, for any realizations $(\sigma_1, \ldots \sigma_n) \in \{-1, 1\}^n$, there exists a response r such that*

$$\operatorname{sign}\left(-\left(r - S_i\right)^2 - w_i\right) = \sigma_i \tag{C.8}$$

A vector $(S_1, \ldots, S_n)$ is shattered *if it has a witness $(w_1, \ldots, w_n)$. The* pseudo-dimension *is the largest integer m such that some vector $(S_1, \ldots, S_m)$ is shattered.*

Claim 18. *The pseudo-dimension is at most 2.*

Since the pseudo-dimension is bounded, the agent's Rademacher complexity is $\tilde{O}(n^{-1/2})$.

Proof. For the sake of contradiction, suppose that the vector $S_1, \ldots, S_n$ is shattered for $n > 2$. By condition (C.8), $\sigma_i = 1$ means that S_i is within some distance d_i of r, where d_i depends on w_i and γ. Define n intervals $I_1, \ldots, I_n$ where $I_i = [S_i - d_i, S_i + d_i]$. Then $\sigma_i = 1$ means $r \in I_i$, and $\sigma_i = 0$ means $r \notin I_i$. Let $f(r)$ be the set of intervals I_i such that $r \in I_i$. Each vector σ corresponds to a unique element in the range of $f(r)$.

I claim that the range of $f(r)$ has at most $2n+1$ elements. If we list the n left endpoints and the n right endpoints of intervals, in order, these define a different set of $2n+1$ intervals. Within each interval J, we can move r from the left to the right, without entering or exiting any interval I_i. Therefore, f is invariant over each interval J. Since there are at most $2n+1$ intervals J, the range of $f(r)$ must have at most $2n+1$ elements.

However, this leads to a contradiction. There are 2^n distinct values of the vector σ. But each vector σ must correspond to a unique element in the range of r, and there are only $2n+1$ such elements. When $n=3$, $2^n=8$ but $2n+1=7$. When $n>3$, the discrepancy is even larger. Therefore, the vector $S_1,\ldots,S_n$ does not have a witness when $n>3$. It follows from the definition that the pseudo-dimension is at most 2. □

Next, consider the policymaker's Rademacher complexity. Note that

$$\max_r \sum_{i=1}^{n} \sigma_i |r-s_0|^{2\gamma}$$

has only three possible solutions: $r=s_0$, $r=0$, or $r=1$. Without loss of generality, I can restrict the response space to $\{0,s_0,1\}$. It follows from Massart's finite lemma that the policymaker's Rademacher complexity is $O(n^{-1/2})$.

C.7 Proof of Lemma 23

Recall the empirical regret bound in the definition of $\widehat{\mathrm{WC}}_n(p)$ (3.13).

$$\max_{r'} \mathrm{E}_{\hat{\pi}^s}\left[u^A\left(p,r',s\right)\right] - \mathrm{E}_{\hat{\pi}^s,\pi^r}\left[u^A\left(p,r,s\right)\right] \leq (4e^{\epsilon}+4)\cdot\overline{\mathcal{RC}}_n^A(p)+\delta_n\cdot\Delta^A(p)+\mathrm{BFR}_n \quad (\mathrm{C.9})$$

I want to argue that every mixed response π^r that satisfies this empirical regret bound also satisfies the regret bound in the definition of $\mathrm{WC}_m(p,0,0,\pi^s)$, i.e.

$$\max_{r'} \mathrm{E}_{\pi^s}\left[u^A\left(p,r',s\right)\right] - \mathrm{E}_{\pi^s,\pi^r}\left[u^A\left(p,r,s\right)\right] \leq 4\cdot\mathcal{RC}_m^A(p,\pi^s) \quad (\mathrm{C.10})$$

At least, this should hold with high probability. Let π^r be a mixed response satisfying the empirical regret bound (C.9). By lemma 17, with probability $1 - n_{\mathcal{P}} \exp(-n^{\alpha})$, we have

$$\max_{r'} \mathrm{E}_{\pi^s}\left[u^A\left(p, r', s\right)\right] - \mathrm{E}_{\pi^s, \pi^r}\left[u^A\left(p, r, s\right)\right] \leq (4e^{\epsilon} + 8) \cdot \overline{\mathcal{RC}}_n^A(p) + \delta_n \cdot \Delta^A(p) + 2\mathrm{BFR}_n \quad \text{(C.11)}$$

This puts the agent's regret in terms of the true distribution. Next, I claim that the right-hand side of inequality (C.11) is $\Theta(m^{-1/2})$. There are three terms to consider. The first term is $\tilde{O}(n^{-1/2})$ by assumption 16. The second term is decreasing exponentially in n, since $\alpha < 2\beta$. The third term is $\Theta(n^{(\alpha-1)/2})$, and it is leading since $\alpha > 0$. Plugging in the value of m gives us $\Theta(m^{-1/2})$. Finally, note that as long as there is a sufficiently large constant in front of m, we have

$$\begin{aligned}(4e^{\epsilon} + 8) \cdot \overline{\mathcal{RC}}_n^A(p) + \delta_n \cdot \Delta^A(p) + 2\mathrm{BFR}_n &\leq 4 \cdot \frac{C}{2\sqrt{2m}} \\ &\leq 4 \cdot \mathcal{RC}_m^A(p, \pi^s) \qquad \text{(C.12)}\end{aligned}$$

where the last line follows from lemma 24. Combining inequalities (C.11) and (C.12) gives us the desired inequality (C.10), with probability $1 - n_{\mathcal{P}} \exp(-n^{\alpha})$.

I have established that the set of mixed responses that $\widehat{\mathrm{WC}}_n(p)$ minimizes over is, with high probability, a subset of the set of mixed responses that $\mathrm{WC}_m(p, 0, 0, \pi^s)$ minimizes over. All that remains is to compare the policymaker's objective under $\widehat{\mathrm{WC}}_n(p)$ with his objective under $\mathrm{WC}_m(p, 0, 0, \pi^s)$. This compares expected utility under the empirical distribution, plus privacy-preserving noise, to expected utility under the true distribution. But this is precisely the situation we found ourselves in during the proof of lemma 22. I can apply the same bounds here to complete the proof.

C.8 Proof of Lemma 24

Recall the definition of Rademacher complexity:

$$\begin{aligned}\mathcal{RC}_n^A(p,\pi^s) &= \frac{1}{n}\mathrm{E}_{\pi^s}\left[\max_r \sum_{i=1}^n \sigma_i \cdot u^A\left(p,r,S_i\right)\right]\\ &= \frac{1}{n}\mathrm{E}_{\pi^s}\left[\mathrm{E}_{\pi^s}\left[\max_r \sum_{i=1}^n \sigma_i \cdot u^A\left(p,r,S_i\right) \mid S_1,\ldots,S_n\right]\right]\end{aligned}$$

where the second equality follows from the law of iterated expectations. To bound the Rademacher complexity, it suffices to bound the interior expectation. Observe that

$$\begin{aligned}&\mathrm{E}_{\pi^s}\left[\max_r \sum_{i=1}^n \sigma_i \cdot u^A\left(p,r,S_i\right) \mid S_1,\ldots,S_n\right]\\ &= \max_{r'} \mathrm{E}_{\pi^s}\left[\max_r\left(\sum_{i=1}^n \sigma_i \cdot \left(u^A\left(p,r,S_i\right) - u^A\left(p,r',S_i\right)\right)\right) + \sum_{i=1}^n \sigma_i \cdot u^A\left(p,r',S_i\right) \mid S_1,\ldots,S_n\right]\\ &= \max_{r'} \mathrm{E}_{\pi^s}\left[\max_r\left(\sum_{i=1}^n \sigma_i \cdot \left(u^A\left(p,r,S_i\right) - u^A\left(p,r',S_i\right)\right)\right) \mid S_1,\ldots,S_n\right] \qquad \text{(C.13)}\\ &= \max_{r'} \mathrm{E}_{\pi^s}\left[\max_r\left(\sum_{i=1}^n \sigma_i \cdot \left(u^A\left(p,r,S_i\right) - u^A\left(p,r',S_i\right)\right)\right)^+ \mid S_1,\ldots,S_n\right] \qquad \text{(C.14)}\\ &\geq \max_{r,r'} \mathrm{E}_{\pi^s}\left[\left(\sum_{i=1}^n \sigma_i \cdot \left(u^A\left(p,r,S_i\right) - u^A\left(p,r',S_i\right)\right)\right)^+ \mid S_1,\ldots,S_n\right] \qquad \text{(C.15)}\\ &= \frac{1}{2}\max_{r,r'} \mathrm{E}_{\pi^s}\left[\left|\sum_{i=1}^n \sigma_i \cdot \left(u^A\left(p,r,S_i\right) - u^A\left(p,r',S_i\right)\right)\right| \mid S_1,\ldots,S_n\right] \qquad \text{(C.16)}\\ &\geq \frac{1}{2\sqrt{2}}\max_{r,r'}\sqrt{\sum_{i=1}^n \left(u^A\left(p,r,S_i\right) - u^A\left(p,r',S_i\right)\right)^2} \qquad \text{(C.17)}\\ &\geq \frac{C\sqrt{n}}{2\sqrt{2}}\end{aligned}$$

The first two equalities follow from algebraic manipulations. Line (C.13) follows from the fact that

$$\mathrm{E}_{\pi^s}\left[\sum_{i=1}^{n} \sigma_i \cdot u^A\left(p, r', S_i\right) \mid S_1, \ldots, S_n\right] = 0$$

Line (C.14) follows from the fact that setting $r = r'$ ensures that the interior sum is zero, so that the maximum over all r is non-negative. Line (C.15) follows from Jensen's inequality. Line (C.16) follows from the fact that the sum inside the expectation is symmetrically distributed around zero. To see this, let X be a symmetric random variable with mean zero. Then

$$\begin{aligned}
\mathrm{E}[|X|] &= \Pr[X = 0] \cdot 0 + \Pr[X > 0] \cdot \mathrm{E}[X \mid X \geq 0] + \Pr[X < 0] \cdot \mathrm{E}[-X \mid X < 0] \\
&= \Pr[X > 0] \cdot \mathrm{E}[X \mid X > 0] + \Pr[X > 0] \cdot \mathrm{E}[X \mid X > 0] \\
&= 2 \cdot \Pr[X > 0] \cdot \mathrm{E}[X \mid X > 0] \\
&= 2 \cdot \mathrm{E}\left[X^+\right]
\end{aligned}$$

Line (C.17) follows from Khintchine's inequality, with constants derived by Haagerup (1981). Finally, the last inequality follows from assumption 17.

C.9 Proof of Claim 9

This is a proof by contradiction. Let w be an effort-inducing contract that pays a positive wage $w(x_i) > 0$ for some outcome $i < m$. Consider a modified contract w' that pays $w'(x_i) = 0$ and

$$w'(x_m) = w(x_m) + \frac{\pi_1^x(x_i)}{\pi_1^x(x_m)} w(x_i)$$

Under contract w', the expected wages conditional on effort are

$$\begin{aligned}\sum_{j=1}^{m} \pi_1^x(x_j)w'(x_j) &= \pi_1^x(x_j)w'(x_j) + \pi_m^x w'(x_m) + \sum_{j\neq i,m} \pi_1^x(x_j)w'(x_j) \\ &= \pi_1^x(x_m)w(x_m) + \pi_1^x(x_m)\frac{\pi_1^x(x_i)}{\pi_1^x(x_m)}w(x_i) + \sum_{j\neq i,m} \pi_1^x(x_j)w'(x_j) \\ &= \sum_{j=1}^{m} \pi_1^x(x_j)w(x_j)\end{aligned}$$

That is, expected wages conditional on effort are the same for w and w'. However, expected wages conditional on effort are smaller for w' than for w. This follows from assumption 18, which implies that

$$\pi_0^x(x_m)\frac{\pi_1^x(x_i)}{\pi_1^x(x_m)} \leq \pi_1^x(x_i)$$

This inequality is strict unless $\pi_0^x = \pi_1^x$, in which case there is no effort-inducing contract anyways. Therefore, w' creates slack in the agent's incentive constraint without affecting the principal's utility. This allows the principal to slightly reduce wages, and be better off than under w.

C.10 Proof of Claim 12

Let w be an optimal contract with a maximum wage of $\bar{w}$. In the worst case, the agent will not put in effort if

$$\mathrm{E}_{\pi^x}\left[w(X^1) - w(X^0)\right] - c \leq 4\overline{\mathcal{RC}}_n^A(w)$$

Otherwise, the agent will put in effort with probability $1-q$ where

$$q = \frac{4\overline{\mathcal{RC}}_n^A(w)}{\mathrm{E}_{\pi^x}[w(X^1) - w(X^0)] - c}$$

This only depends on the contract through two quantities: $\bar{w}$ and $\mathrm{E}_{\pi^x}[w(X^1) - w(X^0)]$. Holding those quantities fixed, the agent's probability of effort is the same.

The rest of the proof essentially follows from arguments in the proof of Claim 9. There, I turned w into another contract w' that preserved the expected wages but increased

$$\mathrm{E}_{\pi^x}\left[w(X^1) - w(X^0)\right]$$

That was the common knowledge case. Here, because increasing $\mathrm{E}_{\pi^x}[w(X^1) - w(X^0)]$ increases q, the principal's payoff under contract w' is actually better than under w. That does not adversely affect the argument.

For any given maximum wage $\bar{w}$, the previous argument shows that $w(x_m) < \bar{w}$ implies $w(x_i) = 0$ for all outcomes $i < m$. More generally, as long as $\ell(x_j) > 1$, this argument can be used to show that $w(x_j) < \bar{w}$ implies $w(x_i) = 0$ for all outcomes $i < j$. It follows from inspection that ceases to be true when $\ell(x_j) < 1$. In fact, a similar argument shows that $w(x_j) = 0$ for any outcome x_j where $\ell(x_j) < 1$.

This establishes the fact that w is a threshold function. Next, suppose I want to ensure that the agent puts in effort with probability q. I need

$$\mathrm{E}_{\pi^x}\left[w(X^1) - w(X^0)\right] - c = \frac{4\overline{\mathcal{RC}}_n^A(w)}{q} \tag{C.18}$$

Suppose $w(x_j) = \bar{w}$ for all $i > j$ and $w(x_j) = 0$ for all $j < i$. The expression in the statement of the claim ensures that I increase $w(x_i)$ as much as necessary to guarantee equation (C.18).

C.11 Proof of Claim 16

The buyer's strategy is the *empirical quantile maximizer*, i.e.

$$x_n \in \max_{x \in M} U^A\left(q, M, x, \hat{\pi}^s\right)$$

I will bound her quantile regret under this strategy. By the Dvoretzky-Kiefer-Wolfowitz inequality,

$$\Pr_{\pi^v}\left[\sup_u \left|\Pr_{\hat{\pi}^v}\left[u^A\left(M, x, v\right) \leq u\right] - \Pr_{\pi^v}\left[u^A\left(M, x, v\right) \leq u\right]\right| \geq t\right] \leq 2 \exp\left(-2nt^2\right)$$

By assumption 19, this implies

$$\Pr_{\pi^v}\left[\left|U^A\left(q, M, x, \pi^v\right) - U^A\left(q, M, x, \pi^v\right)\right| \geq \frac{t}{K}\right] \leq 2 \exp\left(-2nt^2\right)$$

By the union bound,

$$\Pr_{\pi^v}\left[\max_{x \in M}\left|U^A\left(q, M, x, \pi^v\right) - U^A\left(q, M, x, \pi^v\right)\right| \geq \frac{t}{K}\right] \leq 2|M| \exp\left(-2nt^2\right)$$

Set the right-hand side equal to a constant γ and solve for t. This yields

$$t = \sqrt{\frac{\ln\left(\frac{2|M|}{\gamma}\right)}{2n}}$$

It follows that

$$\text{Q-Regret}(M, q, \pi^v) \leq \sqrt{\frac{2 \ln\left(\frac{2|M|}{\gamma}\right)}{K^2 n}} + \bar{v} m \cdot \gamma$$

Setting $\delta = 1/n$ to complete the proof.