Essays in Design Economics

James Michelson

July 2024

Department of Philosophy
Carnegie Mellon University
Pittsburgh, PA 15213

Thesis Committee:

Kevin Zollman, Chair	*Carnegie Mellon University*
Alexey Kushnir, Chair	*Carnegie Mellon University*
Vincent Conitzer	*Carnegie Mellon University*

Submitted in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in
Logic, Computation, and Methodology.

"Philosophers have hitherto only interpreted the world in various ways; the point, however, is to change it."
—Karl Marx, *Theses on Feuerbach*

"Yes, I do think we are simply the tellers of fables, but is that not wonderful?"
—Ariel Rubinstein, 'Dilemmas of an Economic Theorist'

Abstract

This thesis contains two essays in philosophy of science on the value and application of economic theory and one essay in microeconomics on optimal multidimensional auction design.

On the topic of the value of economic theory in design economics, I propose a philosophical reformulation of scientific realism that is appropriate to the scientific domain of economics (chapter 2). In virtue of this formulation, I argue that economic theory explains the successes of design economics.

In chapter 4, I advance a unified theory of observer effects across the social sciences, a problem I call *reflexive measurement*. This theoretical understanding of the role of reflexivity extends more generally to all aspects of science including prediction and theorizing. I also examine the role of economic theory in addressing this problem.

In addition to these two philosophical contributions, chapter 3 is an extended exploration of optimal auction design in the multidimensional setting of a single good with multiple quality levels using simulations. This work examines whether the *exclusive-buyer mechanism* is optimal and offers several conjectures concerning qualitative properties of optimal mechanisms in this analytically intractable setting.

Acknowledgments

I want to begin by thanking my friends and colleagues in Pittsburgh: Dejan Makovec, Brendan Fleig-Goldstein, Bele Wollesen, Mason Broxham, Andrew Warren, Gal Ben-Porath, Nil-Jana Akpinar, Fernando Larrain, Zeyu Tang, Sam Mark, Carly Medina, Max Siemers. Without you this journey would have been immeasurably worse.

My debt of gratitude also extends those to who were a little further away: Marina Dubova, Tim Cejka, Adrianna Kayla Lee, Agne Sabaliauskaite, Giovanna Vitelli, Hannah Schigutt, John Davis, Ola Topczewska, Rawle Michelson. Your support was invaluable.

I have also benefitted enormously from faculty who've advised and mentored me throughout my time in Pittsbrugh. David Danks was a fantanstic supervisor for my Masters degree and continued to offer sage advice throughout my time at CMU. David Childers has been source of invaluable help and patience as I've explored philosophical problems in economics well beyond the scope of this thesis. Vincent Conitzer was able to offer incredible advice and feedback in *both* auction theory and philosophy. Kevin Zollman has been my supervisor throughout my time at CMU and has been as unfailing in his generosity as he has been insightful in his feedback. Finally, Alexey Kushnir took a gamble on a philosopher with a programming habit early in my PhD. Without that decision, this thesis would not exist.

Pittsburgh has been a wonderful place to complete a PhD. I am grateful for the flora and fauna responsible for keeping me healthy and happy throughout my time here: from Greta, the groundhog living in my backyard, to peaceful summer evenings spent Schenley Park. Additionally, special thanks to Akshay Venkatesh, Lindsey Graff, Vyas Sekar, Christopher Philips, and Frank Pfenning. I pray your squash game lives to regret my departure.

Contents

Chapter 1

Introduction

1.1 Motivation

Suppose a famous auction house like Sotheby's[1] wishes to maximize its profits selling a work of art. Which auction format should it use? Auctions can be *open-bid* or *sealed-bid*, depending on whether bidders must reveal their bids publicly. An auction can also be *single-round* or *multi-round*, where bidders are allowed to bid multiple times, as in the case of English[2] or Dutch[3] auctions. Multiple goods may be sold, sometimes packaged or *bundled* together for a discount. *Reserve prices* for these goods can be posted so that a seller will only sell if bids are above them. In *all-pay* auctions, all bidders are required to pay irrespective of whether they submit the winning bid. Additionally, auctions can even require bidders pay an amount different than their bid, as in the case of a *second-price* auction[4], where the bidder with the highest bid pays wins the good(s) but only pays the value of the second highest bid.

Perhaps surprisingly, it turns out this question has an associated deep body of economic theory, stretching back to Nobel Laureate William Vickrey's (1961) seminal research demonstrating the revenue equivalence of first and second-price auctions. Auction theory—more generally, the theory of *mechanism design*—has illuminated many facets of the problem of designing auctions. Theoretical work has extended Vickrey's famous result, characterizing the entire class of profit-maximizing auctions for selling a single good (Myerson 1981). We know that if multiple goods are sold offering selling them together (Adams and Yellen 1976) or offering a lottery, where the bidder is unsure which good they end up with in the event they win (Thanassoulis 2004), can increase profits. Theoretical work has even extended our understanding of 'rational' behavior in auctions (Li 2017), adding to our scientific arsenal of concepts that we can use to characterize how people will respond to different institutional designs. These are only a few examples from the decades of research in theoretical economics that applies to designing auctions.

This theory also applies much more widely than one might initially guess. An auction can be represented mathematically by a pair of functions[5]. An *allocation function* maps bids to outcomes. A *transfer function* maps bids to payments. With this pair of functions it is possible to characterize the wide range of variations in auction format introduced above. Moreover, this mathematical representation covers a surprisingly wide range of design problems. Parallels can immediately be drawn to the problem of *price screening* faced by a monopolist[6]. Any bilateral trading platform (e.g., a stock market or exchange) can be characterized as a *double auction* with multiple buyers and sellers, offering an alternative perspective on the design of markets. The theory of auction design has even been applied to the problem of data acquisition and survey sampling (Roth and Schoenebeck 2012), an example covered in more detail in chapter 4.

[1] `https://www.sothebys.com/en`

[2] `https://en.wikipedia.org/wiki/English_auction`

[3] `https://en.wikipedia.org/wiki/Dutch_auction`

[4] Most commonly contrasted with a *first-price* auction, where the bidder with the highest bid wins the good(s) and pays the value of their bid.

[5] A comprehensive model of multidimensional auction design relevant to the microeconomic portion of this thesis is introduced in section 3.3.

[6] The problem of price screening is commonly known as *second-degree price discrimination*, where a seller offers a range of options to buyers in order to reveal buyers' private information concerning the value of the seller's goods.

What is the scientific value of this body of economic theory? Is it actually useful for designing institutions and policymaking? Economists and philosophers of science are divided on this question. Well-known, award-winning microeconomic theorists like Ariel Rubinstein believe economic theory merely amounts to a collection of fables (Rubinstein 2012) whereas other equally well-known, award-winning microeconomic theorists assert that this kind of scientific theory has predictive value and sharpens the intuitions of economists working on these problems (Roth and Wilson 2019). Philosophers of science are no less split, with some arguing that economic theory played an incredibly limited role in some of the major successes of auction design in the 1990s (Nik-Khah 2008) and others asserting this kind of theory is true in a strong, ontological sense (Ross 2008).

This thesis represents a philosopher of science's attempt to better understand a body of scientific theory combined with an attempt to contribute to it. The overarching goal is to illuminate the value and uses of economic theory for questions of design. The scientific portion of this thesis contributes to our theoretical understanding of auction design. The philosophical orientation of this thesis is prescriptive. The philosophical conclusions of this thesis directly entail recommendations regarding how to improve science. Thus, the novel philosophical contributions of this thesis are intended to be relevant for economists and social scientists interested in the foundational problems that concern the study of human behavior.

1.2 (A Very Brief) Introduction to Design Economics

Design economics is both a collection of institutional design problems as well as an approach to doing economics[7]. Economist Al Roth has called design economics "the part of economics intended to further the design and maintenance of markets and other economic institutions" (Roth 2002, p. 1341). It is often used synonymously with 'market design'. At its core, the theoretical content of design economics models the choice a principal (designer) makers concerning how to design an institution to achieve some goal. Common goals include profit maximization (optimality), welfare maximization, efficiency, stability, and equity. Oftentimes, this work is applied in orientation. Economists increasingly consult on questions of policy and empirical and computational work play a prominent role alongside economic theory (see, for first-hand accounts of this phenomenon, Binmore and Klemperer 2002; Roth and Wilson 2019; Sönmez 2023).

What are some examples of institutional design problems tackled by economists working in design economics? Auction design, introduced above, is a core topic in design economics. Both empirical and theoretical work in economics tries to understand the trade-offs that come with choosing one type of auction format over another. Matching problems are another major topic. These concern problems like matching students to schools, new doctors to hospitals, and organs to transplant recipients. The flexibility of economic theory in this domain means that it can be applied well beyond the setting that motivated its development. As mentioned above, auction theory also applies to problems of price discrimination and data acquisition, a topic that was pioneered by computer scientists (see, for example, Roth and Schoenebeck 2012; Cai, Daskalakis, and Papadimitriou 2015). Although economists aided policymakers on questions of institutional design in the 1990s, since then computer scientists have been increasingly involved in applying economic theory to problems in and adjacent to computer science[8]. Although the canonical problems of auction design and matching markets were responsible for establishing the domain of design economics, its successes have broadened its scope of application and attracted scientists from neighboring disciplines.

The intellectual origins of design economics can be traced to the 1960s and 70s. Early work on auctions (Vickrey 1961) and matching mechanisms (Gale and Shapley 1962) blossomed into substantial theoretical literature in mechanism design and matching theory. Mechanism design, in the spirit of Hurwicz (1972), became a discipline in its own right, which now boasts its own specialized journal, the *Review of Economic Design* (created in 1994). A similar story exists for the development of matching theory. Ultimately, both of these disciplines blossomed alongside related empirically oriented literature on the outcomes of particular forms of institutional design[9]. To understand the theoretical orientation of this type of scientific modeling,

[7]See section 2.3 for an extended introduction to this topic, along with consideration of two major empirical successes from the 1990s: the Federal Communication Commision's (FCC) 1994 radio spectrum auction and the National Resident Matching Program's (NRMP) redesign of the matching algorithm in 1996.

[8]Computer scientists have also been at the forefront of studying computational aspects of economic theory (see, for example, Nisan et al. 2007).

[9]See, for example, discussions by Roth (2002) and Myerson (2008).

one of the founders of mechanism design, Leonid Hurwicz, has described it as follows:

> in a design problem, the goal function is the main given, while the mechanism is the unknown. Therefore, the design problem is the inverse of traditional economic theory, which is typically devoted to the analysis of the performance of a given mechanism. (Hurwicz and Reiter 2006, p. 30)

Though economic theories of design economics do examine the performance of particular institutions, much theoretical work takes place in the same vein that Hurwicz outlined. As I will show in chapter 2, these approaches are complementary.

As noted above, in design economics empirical and computational work are seen as "natural complements" (Roth 2002, p. 1363) to theory. This has led to what some economists have called the 'engineering approach' to questions of institutional design. Al Roth has elaborated on this approach by way of analogy with the construction of suspension bridges:

> The simple theoretical model in which the only force is gravity, and beams are perfectly rigid, is elegant and general. But bridge design also concerns metallurgy and soil mechanics, and the sideways forces of water and wind. Many questions concerning these complications can't be answered analytically but must be explored using physical or computational models. These complications, and how they interact with the parts of the physics captured by the simple model, are the domain of the engineering literature. (Roth 2002, p. 1342)

The key idea behind the engineering approach is that computation and experimentation "fill[] the gaps between theory and design" (Roth 2002, p1374). Computational methods analyze settings that are too complex to solve analytically and laboratory experiments offer predictions about how people will behave in these environments. This kind of approach motivates the use of simulations in chapter 3 of this thesis.

The rise of design economics has been tied to a change in perspective in how economists view their subject matter. This change has even recently been lamented by economic theorists:

> If I had to name one major shift in the sensibilities of economic theorists in the past half century, a prime candidate would be the way we conceptualize markets—from quasi-natural phenomena admired from afar to manmade institutions whose design can be tweaked by economist-engineers. (Spiegler 2024, p137)

I believe this change reflects progress in the science of economics. The turn away from contemplating man-made institutions as "quasi-natural phenomena" characterizes a step in the evolution of the 'dismal science' of economics towards a better understanding of our (socially constructed) world. I hope to convince the reader that not only is the theoretical contents of this emerging scientific domain philosophically interesting in its own right but also that it can be fruitfully applied to address foundational problems in other areas of social science.

1.3 Structure of the Thesis

This thesis consists of three chapters: two in the philosophy of science, one focusing on economic theory in design economics (chapter 2) and the other on statistics in social science (chapter 4), as well as an additional chapter in auction theory (chapter 3). The chapters in this thesis can be seen as answering the following related questions:

- *'What is the scientific value of economic theory in design economics?'* (chapter 2)
- *'What is an example of a theoretical contribution in design economics?'* (chapter 3)
- *'What other problems might this body of theory help with?'* (chapter 4)

Although each chapter is self-contained, the two chapters in philosophy of science are ordered in the following way: the conclusion of chapter 2 on the value of economic theory in design economics—that economic theory explains the successes of design economics—strengthens the case for its application to problems of reflexive measurement in chapter 4. The recommendation to apply the theory of mechanism design to the problem of designing incentive-compatible measurements (4.5) follows from the conclusion concerning the value of this theory.

Chapter 2 offers a *minimal non-fable* account of economic theory in design economics. The goal of this chapter is to offer a philosophical refutation of the position that economic theory does not "produce conclu-

sions of real value" (Rubinstein 2012, p. 37). My argument draws inspiration from a classic philosophical argument that argues for the truth of scientific theories in virtue of their success—the *no-miracles* argument for scientific realism (Putnam 1975). However, this chapter adapts arguments for realism to scientific domains where notions of truth are a "non-starter" (Alexandrova and Northcott 2009, p. 328). Instead of arguing for the strong conclusion concerning the truth of economic theory in design economics, I argue for a much weaker thesis: economic theory explains the successes of design economics. I develop an account of the economic theory of design which I believe best accounts for the recent empirical successes of economics[10]. Unlike existing philosophical accounts of the role of economic theory in design economics (see, for example, Alexandrova and Northcott 2009; Ross 2008) the argument developed here is sensitive to contemporary contributions by economists and computer scientists as well as captures the subtle differences in the kinds of theory used to inform questions of institutional design.

The scientific contributions of this thesis can be found in chapter 3. These concern developing an improved understanding of optimal auction design in the multidimensional setting of a single good with multiple quality levels. Optimal multidimensional auction design problems are notoriously difficult to solve analytically and for this reason, the setting is explored using simulations. These contributions fall squarely in the tradition of design economics, which views computation and experiment as "natural complements" (Roth 2002, p. 1342) to theory. Using simulations, the optimality of the *exclusive-buyer mechanism*[11] is explored alongside other conjectures concerning qualitative characteristics of profit-maximizing mechanisms in this multidimensional setting. Surprisingly, I find the exclusive-buyer mechanism is optimal in settings where randomization is not required for profit maximization. Additionally, I find that the *exclusion region*—the measure of the type space which does not receive the good in equilibrium—does not change with the number of bidders and is sometimes measure zero. The contributions are generative in that they "suggest[] an agenda for theoretical work" (Roth 2002, p. 1363) and can be used to guide future theoretical research on optimal multidimensional auction design.

The final substantive chapter of the thesis advances a general philosophical account of observer effects in the social sciences (chapter 4). My account unifies almost a century of sustained research on this topic by both scientists and philosophers. In my view, observer effects can be understood as problems of *reflexive measurement*, which occur when people are aware of their status as objects of scientific investigation. Viewed at a sufficient level of philosophical abstraction, this characterization not only recovers well-known problems in the social sciences like 'Goodhart's Law' and the *Hawthorne effect* but also sheds new light on how to overcome them. A typical understanding of measurement error will fail to account for the distribution shift caused by a reflexive measurement. The conclusions of chapter 2 facilitate a connection to recent developments in theoretical computer science in the field of *incentive compatible learning*. I argue that the economic theory of design economics can address problems of reflexive measurement where other ascientific, "purely statistical" (Marget 1929, p. 319) approaches fail to do so.

[10]These successes are outlined in more detail in section 2.3 and a brief overview of design economics is provided below (section 1.2).

[11]The exclusive-buyer mechanism is formally defined in section 3.3.1. For a review of related work, see the literature review in section 3.2.

Chapter 2

A Realist Argument for Design Economics

2.1 Introduction

This chapter advances a *minimal non-fable* account of economic theory and its role in the emerging domain of 'design economics' (Roth 2018). Though closely related to arguments concerning philosophical realism and its relationship to scientific theory, this chapter eschews broader claims about the 'truth' of science in favor of more narrowly arguing for the rejection of a troubling, albeit common position, namely, that economic theory and modeling should not and can not "aspire toward purposefulness and... practical use" (Rubinstein 2012, p. 35). This *theory-as-fable* view is deeply worrying. If correct, the standard by which we—scientists, philosophers, the general public, etc—are to evaluate economic models is altogether divorced from their ability to materially improve our lives. Thankfully, I believe this position is not supported by the available evidence. To borrow jargon from a familiar philosophical argument for realism: it would indeed be miraculous if the successes of design economics rested on a body of theory that amounted to a mere collection of fables. This is the central thesis of present chapter.

This argument specifically applies to the subdomain of design economics within the broader discipline of economics. In the past few decades, the profession of economics has become increasingly involved in the design of institutions and markets. The ascendancy of design economics is well-documented across the economics profession, as well as outside of it[1]. Part of this rise has undoubtedly been due to some very high-profile success stories; examples that seem to indicate the application of microeconomic theory leads to some kind of empirical success. Awards like the Nobel Prize and the John Bates Clark medal have been given to economic theorists working in associated theoretical disciplines of mechanism design and matching theory on the basis of their technical contributions and "how basic research can subsequently generate inventions that benefit society" (The Royal Swedish Academy of Sciences 2020).

Despite the widely heralded success stories of design economics, some economists and philosophers of science are skeptical of the role of economic theory in contributing to these successes. Most notably, renowned microeconomic theorist Ariel Rubinstein has advanced a view of economic theory wherein theory ought not "engage in predictions or recommendations" (Rubinstein 2012, p. 36). On this view, economic theory is merely a collection of fables. This position is radical: fables do not yield predictions nor offer scientific insight or intuition. Even more forcefully, Rubinstein (2012, p. 37) asserts he is "obsessively occupied with denying any interpretation contending that economic models produce conclusions of real value." It is my belief this view is entirely wrongheaded.

My argument adapts a familiar philosophical argument for scientific realism to the domain of economics. The *no-miracles* argument for scientific realism (Putnam 1975) probabilistically infers the truth of a scientific theory from its success. This argument is particularly well-suited to explain the success of a natural science like physics; in the context of economics, some work is required to adapt realist concerns to the local domain

[1]See (Roth and Wilson 2019; Spiegler 2024) for a contemporary view of this rise by economists. From a complementary perspective in philosophy of science, see (Alexandrova and Northcott 2009; Guala 2001).

of economics. To advance a minimal non-fable account of economic theory in the domain of design economics I propose a modified (weaker) *explanatory challenge* version of the no-miracles argument with a similar non-miraculous inference. Insofar as successes accrue to a scientific domain it becomes increasingly unlikely that their success cannot be explained by a common feature that is not reducible to random chance. This version of the no-miracles argument is designed for scientific domains like economics where the truth of a scientific theory is difficult to establish.

Armed with the explanatory challenge version of the no-miracles argument, it is then a matter of determining the kinds of features that explain the success of design economics. Some obvious candidates emerge. The role of experiment and computation have been heralded as "natural complements" (Roth 2002, p. 1342) to theory and there is broad consensus between economists and philosophers of science concerning their importance in bringing about success. Additionally, personal and professional relationships between advising economists and policymakers are also considered. These are also widely considered important. What of economic theory and its role in explaining the success of design economics? Here, views diverge drastically. Some assert that the case for theory is wildly overblown (Nik-Khah 2008; Rubinstein 2012). Others take a more moderate view, believing it is useful in conjunction with other features like experiment and computation, as well as being circumscribed by point of application (Alexandrova and Northcott 2009). Some even assert theory has positive predictive value or sharpens the intuitions of economists entirely on its own (Roth and Wilson 2019).

This final story concerning the role of theory in honing economists' intuitions is by far the most common. It is even called the "mainstream paradigm for market design" (Sönmez 2023, p. 10) by those adjacent to it. It supports the view that the role of theory in explaining the successes of design economics is not reducible to random chance. Furthermore, a particular quality of theories of design economics supports this position. The *projective* quality of these theories is such that the world can be made to reflect or "mirror" (Guala 2001) these theories. Theories in design economics such as, say, auction design, model choices designers subsequently take. The world can then be made to look like the model. Taken together, these claims support the probabilistic inference that theory explains the successes of design economics. The alternative theory-as-fable view makes a miracle of this success.

This chapter is structured as follows. In section 2.2 I give a brief overview of philosophical debates in scientific realism, paying particular attention to the no-miracles argument for scientific realism. A modified, explanatory challenge version of the no-miracles argument is introduced in the specific context of the scientific domain of economics. Since this argument begins from the observations of a science's successes, section 2.3 introduces the science of design economics and highlights its successes, showcasing two notable success stories: the residency matching algorithm developed for new doctors in the United States and the auction design implemented by the FCC for their radio spectrum auctions. A consideration of features that might account for the successes of design economics then takes place in section 2.4. Here, experiment, computation, personal and professional relationships, as well as theory are all considered, and arguments for and against their role in explaining success is the explored. There is broad consensus on the role played by all features excluding theory. Theory is considered in detail in section 2.5. I find the theory-as-fable view untenable in light of both the testimonies of those who served as economic advisors on the successful instances of design economics and the projective quality of the theory. Finally, I conclude in section 2.6.

2.2 Explaining a Successful Science

What should we make of a successful scientific practice? Are we to believe its theoretical premises? Its ability to be successful again? These questions have been at the heart of philosophy of science's attempt to grapple with the philosophical doctrine of *scientific realism*—the idea that the theoretical content of a successful science is, in some sense, "true"—since the dawn of the Twentieth Century. Much of this preoccupation has centered on the scientific discipline of physics, which boasts a host of impressive scientific achievements. On the other hand, economics, the "dismal science", hasn't yet split the atom or found a way to reliably mitigate the woes of inflation or unemployment, let alone forecast them. However, design economics has slowly accrued a number of success stories that it can rightfully be proud of[2]. But how can we adapt

[2]These are covered in detail in Section 2.3 below.

considerations of realism to social scientific domains like economics? Answering this question is the focus of the present section.

It is helpful to sketch the philosophical doctrine of scientific realism in further detail since the position is quickly reached from observing a science's successes. This story is most straightforwardly understood in the context of physics; I will subsequently address realism in the context of economics. Scientific realism is the philosophical position that something about a science—its theories, structures, entities, experiments, etc—is true. The notion of 'truth' alluded to is up for debate. Sometimes a literal, semantic notion of truth is invoked whereas other times it suffices that a scientific theory is only "approximately true" (see Chakravartty 2017 for discussion). Furthermore, "[e]ven when our sciences have not yet got things right, the realist holds that we often get close to the truth. We aim at discovering the inner constitution of things and at knowing what inhabits the most distant reaches of the universe" (Hacking 1983, p. 21). Realists maintain that (1) there is a mind-indepedent reality and (2) our science has access to it. Taken together, a generic realist philosophical position towards science amounts to: "our best scientific theories give true or approximately true descriptions of observable and unobservable aspects of a mind-independent world" (Chakravartty 2017, §1.2).

In contrast to scientific realists, *scientific anti-realists* believe that "[scientific theories] are tools for thinking. Theories are adequate or useful or warranted or applicable, but no matter how much we admire the speculative and technological triumphs of natural science, we should not regard even its most telling theories as true" (Hacking 1983, p. 21). On this view, theoretical entities like electrons or quarks are convenient fictions that allow scientists to better understand the world. The same goes for theories like the theory of general relativity. It is not to be interpreted as a literally true description of the external world but as a tool for thinking. There are simply better and worse tools; none of them are true in any sense of the word. This division between realism and anti-realism is not a disagreement about whether something (a practice, theory, etc) is or isn't scientific, or whether some science actually works, but instead about whether the epistemological apparatus of science actually corresponds to something in the external world (i.e., it is "true" or "approximately true").

There are as many arguments for scientific realism as there are against it (for anti-realism). One of the most prominent defenses of scientific realism follows from the intuition that it would be unlikely if a successful science wasn't, in some sense, true. The more a science is successful, the more unlikely it would be if its underlying theory didn't latch onto something in the external world. This is known as the *no-miracles argument* for scientific realism. On this view, realism "is the only philosophy that doesn't make the success of science a miracle" (Putnam 1975, p. 73). A rough schematization of the argument is as follows:

$P1$ Science X is *successful*

$P2$ Science X has theoretical content Y

$P3$ If Y weren't *true*, the successes of X would be *miraculous*

$C1$ Y is true

Where the relevant notions of *success*, *truth*, and *miracle* require further explication. Different realist positions can be articulated on the basis of different understandings of these notions, as well as the links between them. Of all the arguments for scientific realism; however, the no-miracles argument puts the successes of a given science front and center: in the absence of such successes, no account of the "truth" of the scientific theory can be given.

By way of an example, we can consider a no-miracles argument for the truth of the theory of gravity in virtue of its correct prediction of the orbit of Halley's comet. Halley's comet is a short-period comet that is visible to the naked eye from Earth on average once every 76 years (NASA 2024). The theory of gravity developed by Isaac Newton was used by his contemporary Edmond Halley to correctly predict the next time the comet would be visible in his *Synopsis of the Astronomy of Comets* (1705). This correct prediction (success) is used to argue that the theory of gravity is true. A single correct prediction can be augmented by other successful predictions, each serving to render the "miraculous" inference increasingly unlikely. Of course, many philosophers of science committed to antirealist philosophical positions have argued against inferring the truth of a scientific theory from its success (see, for example, Stanford 2000); this example merely illustrates how a no-miracles argument proceeds. First, the theoretical component of a science is identified and its successes are established. Then, the truth of the theory is established in virtue of the

likelihood that the successes weren't due to random chance.

The no-miracles argument is designed to argue for the truth of a scientific theory in virtue of that theory's purported ability to generate scientific successes. In the context of economics, however, it makes little sense to speak of the truth of an economic theory in such frankly metaphysical terms. For starters, the idea that a fully rational, self-interested agent is a literally true description of anyone is a "non-starter" (Alexandrova and Northcott 2009, p. 328). Noted economists like Milton Friedman (1953) have famously made a virtue of unrealistic assumptions and their role in economic theory. Truth—in any form—is clearly not something can be inferred from any success in this scientific domain. How then should we interpret successes in the field of economics? It is helpful to reconsider the no-miracles argument. This argument can also be viewed as creating an *explanatory challenge* where the successes of a given science merits an explanation. What accounts for the successes? In the context of a science like economics, the goal is to yield a suitable explanation which does not make successful scientific practices unlikely.

The explanatory challenge argument begins, like the no-miracles argument above, from the observation that a science is successful. Instead of narrowly considering the role of theory in achieving that success (and its subsequent truthfulness) this argument allows for arbitrary features of the science to take the place of theory:

$P1$ Science X is *successful*

$P4$ Science X has common feature F

$P5$ It would be *miraculous* if F didn't explain the success of X

$C2$ F explains the success of X

Here, it is still necessary to both explicate a version of *success* and expound a probabilistic case for what constitutes a *miracle*. Notice, however, there is no notion of truth lurking in the conclusion. This argument is much weaker than a no-miracles argument for scientific realism. The challenge here is, unlike in the no-miracles argument for philosophical realism, merely to argue that some F contributes to the success of science X. There is no abductive-type argument for why future F is more important than any other feature F' or that F cannot contribute absent this other feature. Instead, the goal is simply to argue that it is unlikely feature F does not play a role in explaining[3] the success of science X.

It is helpful to unpack this argument in more detail. What constitutes an adequately circumscribed "science"? This question is not intended to raise the specter of the problem of scientific demarcation. Instead, it can be read as asking what merits the consideration of design economics independently from the rest of economics. Note, I do not consider design economics a separate science. As I understand it here, design economics is a collection of institutional design problems (concerning auctions, matching markets, etc). It is as much a sociological phenomenon as it is anything else: it is the collection of institutional design problems that have attracted members of the academic economics profession[4]. I am not concerned with whether this collection of problems is a distinct science. My goal is to explain why this collection of institutional design problems that have attracted academic economists has exhibited success. The use of "science" in the premises above is merely intended to collect the relevant success stories of design economics: these are the explananda that merit an explanans.

What of "features"? This framing is deliberately constructed to avoid the narrow focus on theory alone. The animating idea is simple: aspects of a science like its use and reliance on experiment and computation may also explain its success. In the case of design economics, for example, experimental work was conducted to assess bidders' behavior in controlled environments prior to the final version of the 1994 FCC auctions (Plott 1997) and experiments were used to confirm field observations concerning the stability of different matching algorithms (Kagel and Roth 2000). I will return to this point in more detail below. Crucially, however, I take the relevant features that explain a science's success to be those that are not reducible to random chance. If the collection of successful cases that are to be explained all took place on a Tuesday, this would not make the feature 'took place on a Tuesday' the relevant kind of explanation. This is because

[3]The argument is robust to all philosophical variants of "explanation" (see, for discussion, Woodward and Ross 2021). Informally, "explanation" is used here to mean "contributes to the success of science".

[4]Note, insofar as there are other design problems which have not attracted academic economists (say, for example, in the design of public transport networks, a problem that usually attracts civil engineers) they are not considered problems of design economics as investigated here.

the explanation is not necessary to explain a science's success. Whereas the original no-miracles argument establishes that the truth of a scientific theory is sufficient to explain its success (how else could we infer a miracle if the theory wasn't true?) the weaker explanatory challenge version merely asserts that a feature is necessary for explaining success.

Finally, how should we reason about the probabilistically about the nature of a "miracle"? The original formulation of the no-miracles argument has been criticized by philosophers of science as an instance of fallacious reasoning known as the 'base rate fallacy' (Magnus and Callender 2004). If the set of candidate theories from which a given theory is drawn overwhelmingly contains true theories, then the conditional probability the theory is true given it is successful is obviously very high. Similarly, if the likelihood of a given theory being true is low, then the resulting conditional probability is also low. Thus, "the no-miracles argument turns on neglecting this base rate" (Magnus and Callender 2004, p. 326). This contention challenges the nature of what constitutes a miracle with respect to the success of a science.

In the present context, there are two issues the base rate objection. First, it hinges on a restricted understanding of scientific truth: a scientific theory is either true or false. Insofar as a scientific theory cannot be characterized so starkly, this contention loses its force. Second, the notion of "miracle" required to establish that a common feature explains the successes of science can be significantly weakened without undermining the conclusion. As the number of successful instances of a scientific endeavor grows, insofar as they share common features, these features are increasingly *unlikely* to fail to explain the success. The colloquial use of the word "miracle" merely stands in for a probabilistic argument concerning whether a common feature of the science explains its successes. The weakened no-miracles argument requires identifying features that are not reducible to random chance; the issue at hand is whether a plausible story can be given for how these common features explain a science's successes. The objective of this argument is to show that, as the number of successful cases grows, common features they share explain the successful cases.

The explanatory challenge version of the no-miracles argument adapts the philosophical debate over scientific realism to scientific domains where a notion of truth is harder to establish. This project is connected to other attempts by philosophers of science to establish *local realisms* specific to scientific domains like economics (see Mäki 2009). The key conviction that animates this section is that there should exist some basis to ascertain whether a science that is not in the business of truth-telling is more or less 'real', in the sense of latching onto a mind-independent reality. If a scientific domain like economics can muster success stories that, say, reliably improve welfare or maximize revenue or result in efficient outcomes, it ought to be possible to dignify the common features of these successes with the recognition they deserve.

2.3 What is 'Design Economics'?

'Design economics' is typically understood as "the part of economics intended to further the design and maintenance of markets and other economic institutions" (Roth 2002, p1341). It is often used synonymously with the designation 'market design' (as in Roth 2018) although it is not limited in applicability to markets and marketplaces; the label also covers auction design, hiring practices, organ exchanges, and matching students to schools. At its core, design economics models the choice a principal (designer) makes concerning how to optimally, efficiently, or equitably design an institution. It is not limited to a single theoretical approach instead drawing (not exclusively) from social choice theory, mechanism design, and matching. Design economics aims to inform policy and other concrete decisions concerning questions of institutional design; it has a practical focus that distinguishes it from other branches of economic theory.

This practical focus entails that experimental and computational economics are "natural complements" (Roth 2002, p1342) to theory and thus, market design calls for an 'engineering approach', which noted economist Alvin Roth (2002, p1342) has elaborated by way of an analogy with the construction of suspension bridges:

> The simple theoretical model in which the only force is gravity, and beams are perfectly rigid, is elegant and general. But bridge design also concerns metallurgy and soil mechanics, and the sideways forces of water and wind. Many questions concerning these complications can't be answered analytically but must be explored using physical or computational models. These complications, and how they interact with the parts of the physics captured by the simple model, are the domain of the engineering literature.

The key idea behind the engineering approach is that computation and experimentation "fill[] the gaps between theory and design" (Roth 2002, p1374). Computational methods analyze settings that are too complex to solve analytically and laboratory experiments offer predictions about how people will behave in these environments.

In this chapter, I will cover two important case studies that are commonly considered *the* success stories of design economics: (1) the design of a labor clearinghouse for American doctors and (2) the design of the US Federal Communication Commission's (FCC) auction for different parts of the radio spectrum. Although both of these cases occurred in the 1990s, the formative decade where economists were presented with opportunities to help design the rules for complex markets for the first time, the intellectual origins of design economics can be traced to the 1960s and 70s. Early work on auctions (Vickrey 1961) and matching mechanisms (Gale and Shapley 1962) blossomed into substantial theoretical literature in mechanism design and matching theory. Mechanism design, in the spirit of Hurwicz (1972), became a discipline in its own right, which now boasts its own specialized journal, the *Review of Economic Design* (created in 1994). A similar story exists for the development of matching theory. Ultimately, both of these disciplines blossomed alongside related empirically oriented literature on the outcomes of particular forms of institutional design[5]

It is worth briefly highlighting the achievements of the 1990s since they represent the first significant achievements of design economics. Interested readers can consult (Roth 2002) for a longer discussion. First, we begin with the example of the market for new American doctors.

Example 1 (Entry-Level Labor Market for American Doctors)**.** *The entry-level position for an American doctor is called a residency. A good residency substantially influences the career paths of doctors and residents provide much of the labor force of hospitals. In the 1940s, competition for people and positions was so fierce it led to market failure, where the market 'unraveled' and students were being appointed to jobs a full two years before graduating from medical school (Roth 2002, p1346). Thus, students were hired before much indication of their medical school performance was known to their employer, a major source of inefficiency. This market failure was resolved in the 1950s by the introduction of a centralized clearinghouse called the National Resident Matching Program (NRMP) where students would submit rank-ordered lists of jobs and an algorithm was devised to produce a matching of students with hospitals. By 1951 (and lasting into the 1970s) over 95% of positions were filled through this match.*

However, by the 1990s, there was a crisis of confidence in the labor market. The issue was that the original algorithm devised by the NRMP was designed to match individuals and not married couples, who were increasingly graduating together from medical school. The key problem for the NRMP was that the matches generated by the original algorithm were not 'stable': there were pairs of individuals and hospitals who were not matched to each other that would prefer to be matched instead of their proposed matching. Empirical evidence suggests that stability is an important criterion for a successful clearinghouse. Thus, two economists (Roth and Peranson 1999) proposed an alternate design of the clearinghouses' algorithm which would generate a stable match for both individuals and couples. The design was completed in 1996 and implemented in 1997. It is used in a number of medical labor market clearinghouses across the world to this day.

The other major success story of the 1990s was the design of the FCC's spectrum auctions.

Example 2 (FCC Spectrum Auctions)**.** *In 1993 the United States (US) Congress directed the Federal Communications Commission (FCC) to design auctions to efficiently allocate radio spectrum licenses and raise money for the government. Until 1981, spectrum licenses had been historically allocated through a political process called "comparative hearings", whereas after 1981 they were allocated by lottery. Both procedures were characterized by lots of rent-seeking behavior and bureaucratic complications, leading to substantial delays, all of which Congress wanted to avoid. A further issue was that auctions for spectrum licenses had already been tried in other countries and, in the notable case of Austalia, evidenced 'gaming' where participants submitted multiple highest bids and withdrew their winning bids to acquire the license at a lower cost after the auction concluded. Since the FCC was concerned with mitigating this kind of behavior, they hired academic economist John McMillan (then at UCSD) to advise their staff. Additionally, academic economists were hired by communications companies and put forward proposals on behalf of their clients, many of which were adopted by the FCC.*

The resulting auction was carefully designed to avoid a number of pitfalls that concerned the FCC. The

[5]See, for example, discussions by Roth (2002) and Myerson (2008).

auction was open-bid and multi-round, allowing bidders to get a sense of what other bidders thought the licenses were worth ("price discovery"). This was designed to avoid the problem of the "winner's curse" where the bidder who overestimates the value of a license is more likely to submit the winning bid. To overcome the problem of complementarities—where acquiring licenses in combination can change their overall value—the idea of auctioning licenses one at a time was rejected. Additionally, to prevent bidders from concealing their intentions by delaying their bids, the FCC imposed an 'activity rule' that required bidders to continually bid on licenses they were interested in or lose their ability to do so in the future. These auctions raised an estimated $230 billion dollars in revenue for the US government by 2023 (Lee and Malamud 2023) and have been adopted all over the world.

These examples highlight the kinds of success that design economics can lay claim to. In the case of residency matching, matches that are stable for a large number of market participants is an example of what constitutes success. In the case of the spectrum license auctions, success concerned efficient auctions that result in substantial government revenue. These are not the only goals that can be designed for. Equity or welfare maximization can also be the targets of design. Moreover, the adoption of the approach of design economics has led to other successes across the world. It is important to sketch the extent of these successes since they provide a basis for arguing about which features contributed to the success.

As a point of departure, it is worth noting the NRMP and FCC continue to, respectively, use matching algorithms and auctions to this day. The original work of Roth and Peranson (1999) that developed a matching algorithm for matching new doctors has placed over 20,000 doctors per year and has been extended to many other medical fields as well as the market new lawyers (see Roth 2002, p. 1346). The FCC continues to use auctions to allocate its radio spectrum, most notably again in 2016-7 where the FCC ran an "incentive auction", designed to repurpose the spectrum for new uses, raising $19.8 billion in revenue for the US government (Federal Communications Commision 2017). Both matching algorithms and auctions have been developed and used across the world. The sale of the British 3G telecom licenses in 2000 was hailed as the "biggest ever" auction held on earth, raising £22.5 billion (approximately 2.5% of GNP) (Binmore and Klemperer 2002). Dozens of countries across the world have used auctions in increasingly more complicated settings[6]. Matching theory has also been used to rectify the lack of 'thickness'—enough buyers and sellers to produce satisfactory outcomes—in kidney exchanges in the United States (Roth 2007). There is too much variation across the institutions that have been designed (auctions, exchanges, matchings, etc), geographic settings, and types of success to fully cover here. What is clear is this: successful practical applications of design economics recur all over the world, where success is broadly understood to cover everything from revenue maximization to stability.

Before turning to how economics and philosophers of science might address the explanatory challenge posed by the successes of design economics, it is helpful to mention the role of experiment and computation in design economics. These features are "natural complements" (Roth 2002, p. 1342) to theory and are widely held to account for much of the success of design economics[7]. Experimental evidence on the stability of competing market mechanisms (drawn from the experiences of labor markets for doctors in the United Kingdom) was used to argue in favor of the particular matching algorithm adopted by the NRMP (Kagel and Roth 2000). In the case of the 1994 FCC auctions, experiments were conducted to investigate how bidder's behavior changes in light of different auction configurations (Plott 1997). In this case, an important aspect of the empirical evidence derived from experiments was to "teach researchers some of the ways in which it would be safe to perturb the final auction design" (Alexandrova and Northcott 2009, p. 320). Experiments are invaluable tools that economists use to narrow down the range of theoretical concerns that are relevant to a given problem and isolate causal effects.

Computational aspects of design economics also contribute to its successes in a number of ways. Al Roth (2002) identifies a number of ways computational techniques contributed to the design of the matching algorithm used by the NRMP (above). Computation was used to explore alternative algorithm designs and their performance on past data. Computation was also used to further generate and support conjectures about the stability of matchings under different configurations of inputs. Thus, computation can be used to "suggest[] an agenda for theoretical work" (Roth 2002, p. 1363) just as it can be used to give evidence for or

[6]See, for example, a recent overview of combinatorial auction designs used in practice (Palacios-Huerta, Parkes, and Steinberg 2022).

[7]For an economist's view of this, see Roth (2002). From the perspective of philosophy of science, see Alexandrova and Northcott (2009).

against theories[8]. Furthermore, computational methods help us "analyze games that may be too complex to solve analytically" (Roth 2002, p. 1374), offering insights into settings which lack formulation as analytically tractable economic models. Finally, computational methods have also been used to analyze the significance of particular aspects of the FCC's 2016-7 incentive auction *post hoc* (Newman et al. 2020), exploring the performance of alternative auction configurations on realistic models of bidder behavior.

In conclusion, design economics is a part of economics that concerns questions of institutional design, and tackles these questions in a fashion that makes use of experiment and computation to a large degree. It is been likened to a type of engineering, where a goal or objective is the target that an economist (engineer) tries to 'hit'. This section has outlined two prominent case studies that constitute the exemplary success stories of design economics, as well as sketch the extent to which they have been successfully replicated and extended around the world. There are large differences in opinion concerning how economists and philosophers interpret these successes. The next section covers how commentators have chosen to make sense of these successes.

2.4 Explaining the Success of Design Economics

What explains the successes of design economics? The preceding section was designed not only to showcase two important examples which are the touchstones for discussions concerning design economics' success but also to highlight how these successes have been replicated and extended across the world. While individual instances of success have been contested[9] the presentation above paints a more general picture of the successes of design economics: the varied goals that have been achieved, the extent of the successes, and the geographic scope of their implementation. It is this higher-level view of the successes of design economics that merits an explanation. This section covers how philosophers of science and economists would react to the explanatory challenge version of the no-miracles argument sketched above.

Before turning to possible explanations for the successes of design economics a negative conclusion must first be dismissed, namely, that the examples solicited in the preceding section are not examples of success. The animating idea of this objection is something along the lines of 'things could have gone better': auctions could have been more efficient, matches more stable, etc (for an example of this kind of criticism, see Ledyard, Porter, and Rangel 1997). There are two rejoinders to this objection. First, examples of abject failures are well-documented. Examples like the 2000 Swiss UMTS auction (Wolfstetter 2001) or the 1990 New Zealand spectrum auction (Milgrom 2004, §1.2.2) are clear examples of failure. In the former case, the outcome was inefficient, raised far less revenue than projected, and the final allocation outcome was deemed "unnecessarily random" (Milgrom 2004, p. 12). In the latter case, widely described as "flop", far fewer bidders participated than was expected, resulting in a much lower revenue for the Swiss government. The examples presented in the previous section are very clearly not instances of this kind of failure. To what extent they are successes can be explored after the fact with simulations (see Newman et al. 2020). The key insight is that, in the case of failures, alternative auction designs could have performed much better. Moreover, this fact is widely acknowledged. Insofar as it is unclear which designs could have performed better and we avoid obviously negative outcomes like few bidders, unstable matches, and low revenue (etc) these cases should be considered successes.

Additionally, it is possible to contend that although the previously canvassed examples are indeed successful, no explanation for their success is warranted. This is akin to asserting that their success is, effectively, random. A direct implication of this 'no explanation needed' view is that there is no feature (not reducible to random chance[10]) that explains why these cases were successful. The aforegoing presentation of the explanatory challenge version of the no-miracles argument is deliberately constructed to avoid this conclusion. As the set of successful instances of a science grows, the probability that some common, non-arbitrary feature does not explain their success shrinks. In contrast to the problem of base rates (Magnus and Callender 2004), this formulation of an argument for realism is not narrowly focused on the role of theory and its relation to truth in explaining success. Thus, as the number of successful cases grows, any common, non-arbitrary feature present in these cases is an increasingly promising candidate for explaining their success.

[8]This is what motivates chapter 2 of this PhD.

[9]I will discuss below the work of Edward Nik-Shah (2008) in uncovering the suppressed role of commercial interests in the FCC's 1994 spectrum auction from archival records.

[10]I will refer to features which are not reducible to random chance as *non-arbitrary* features.

What features might then explain the successes of design economics? Though it is possible to garner a wide variety of candidates, I will focus on three[11] sets of features: (1) personal and professional connections; (2) experiment and computation; and, (3) theory. The goal is to collect the ideas and opinions of philosophers of science and economists to better understand how these features might explain the success stories of design economics. Again, it bears emphasizing that establishing an explanatory connection between a feature and successful science is not *abductive*: there is no claim that a given feature is *the best possible explanation* for the success. Instead, the goal is merely to make the case that common, non-arbitrary features mattered in bringing about success. Furthermore, there is nothing that precludes the possibility these features only matter in conjunction. That theory might also require experiment, for example, is well within the conclusion of the argument established here.

The idea that personal connections between economists, industry stakeholders, and policymakers matter in explaining the success of design economics is widely echoed by economists who've served in advisory capacities. Notably, senior economic advisor to the FCC Evan Krewel extensively documents the key actors who were decisive in bringing economists into the FCC's auction design procedure (and their relationships) in the forward to Paul Milgrom's (2004) *Putting Auction Theory to Work*. Even Milgrom's character—his "integrity" (Milgrom 2004, p. xix)—is cited as important in his efforts to persuade the FCC to adopt his design. These narratives concerning the personal relations between academic economists and policymakers are also echoed by other economists who've had similar roles as economic advisors (see, for example, Roth 2002; Sönmez 2023). There is no straightforward means of ascertaining the extent of these relationships across all the successful instances of design economics; however, if it turns out that these cases have this feature of personal connections in common, then this should constitute a possible explanation for their success.

That good relations between consulting economists and policymakers is a feature that explains the successes of design economics will surprise no one who has worked in a government or administrative setting. However, the role of corporate interests in shaping the Overton Window of policy options has not gone unnoticed by commentators. Philosopher of science Edward Nik-Shah's unexpurgated account (2008) of the 1994 FFC spectrum auction from archival records shows the degree to which economists were used to legitimize outcomes that aligned with the incentives of those corporations that hired them. His conclusion could be no less ambiguous: "*corporate imperatives demonstratively played the decisive role in determining the auction*" (Nik-Khah 2008, 89, emphasis original). Ultimately, the extent to which this constitutes corruption is not relevant to my argument. However, the extent to which this occurred in other instances of success renders it a feature worthy of consideration for explaining that success.

The role of experiment and computation in design economics was already documented above. Economist Al Roth (2002; 2018) has made the case that experiment and computation are complements to theory and drawn comparisons between economics and the discipline of engineering in the natural sciences. Philosophers of science have echoed this perspective, most notably Alexandrova and Northcott (2009, p. 320), who persuasively argue one of the roles of experiments was to "teach researchers some of the ways in which it would safe to perturb the final auction design." Like Al Roth, they echo the idea of economists as 'engineers', this time drawing analogies between auction design and the "development of *racing cars*" (Alexandrova and Northcott 2009, 331, emphasis original). "Theoertical knowledge alone is not enough" they continue, "teams also have a huge testing programs, analogous to the experimental test beds of the [FCC's 1994] spectrum acution" (Alexandrova and Northcott 2009, p. 331).

It is hard to assess the ubiquity of computation and experiment across the successful cases garnered above. Clearly, leading economists working in design economics march to a similar tune: experiment and computation are everywhere heralded as important for the design of institutions (Roth 2002; Binmore and Klemperer 2002; Sönmez 2023; Milgrom 2004). Philosophers of science critical of the prominence of economic theory in attributions of the success of design economics are unified in their endorsement of the importance of experiment in explaining the success of the FCC's 1994 spectrum auction (Alexandrova and Northcott 2009; Nik-Khah 2008). Again, insofar as experiment and computation are common across the successful cases of design economics, the broad consensus concerning their importance supports the claim they explain that success. And given the claims by economists above, it seems likely this is in fact common to the success stories of design economics.

[11] There are undoubtedly more, many of paramount importance. However; these are commonly considered in the literature and (1) and (2) serve as preludes to the central claims of this chapter concerning (3) theory.

The role occupied by the features covered so far in explaining the successful instances of design economics would likely be granted by those critical of the narrative concerning the importance of theory in explaining those very same successes. The evidence mustered above establishes that the role of (1) personal and professional relationships and (2) computation and experiment is not *arbitrary*; it is not reducible to random chance. These features are, in some minimal capacity, necessary for success. The picture for (3) theory is much murkier. In the wake of the FCC's 1994 spectrum auction[12] much fanfare was made by academic economists concerning the crucial role played by theory in determining the success of the auction (see, for example, McAfee and McMillan 1996; McMillan 1994). This picture has subsequently been contested by a number of philosophers of science (Nik-Khah 2008; Alexandrova and Northcott 2009). It is worth outlining the relationship between scientific realism and economic theory—in the vein of a 'local realism' for economics (Mäki 2009) advocated above—to better understand how economists and philosophers of science understand the role of theory in the science of economics.

It is helpful to begin with Friedman's (1953) seminal essay 'The Methodology of Positive Economics' because the conclusion drawn here is sympathetic to his motivations. Furthermore, this represents a canonical understanding of the role of theory in contemporary economics. For Friedman, "[t]he ultimate goal of a positive science is the development of a "theory" or, "hypothesis" that yields valid and meaningful (i.e., not truistic) predictions about phenomena not yet observed." (Friedman 1953, p. 7). The notion of novel prediction alluded to here captures the animating idea of realism in the example of Halley's comet used in the preceding section[13]. For Friedman, the operative notion of success that characterizes the science of economics is *predictive success*. Notably, the success of a theory can be evaluated independently of how unrealistic its assumptions are:

> a theory cannot be tested by comparing its "assumptions" directly with "reality". Indeed, there is no meaningful way this can be done. Complete "realism" is clearly unattainable, and the question whether a theory is realistic "enough" can be settled only by seeing whether it yields predictions that are good enough for the purpose at hand or that are better than predictions from alternative theories." (Friedman 1953, p. 41)

Though there are many ambiguities throughout Friedman's writing regarding his use of 'realism' and 'realisticness' his essay "is transparently the manifesto of an engineer rather than a scientist" (Ross 2008, p. 740).

This starting point for evaluating the role of theory in explaining the successes of design economics is helpful in the context of design economics. Though Friedman's primary concern is predictive success (what is often called successful *novel prediction*), the outlook is manifestly that of a scientist concerned with using theory to change (improve) the world. His endorsement of economic theory despite its unrealistic assumptions is a clear indication that his understanding of scientific realism is one which is not centered on a notion of truth[14]. Friedman goes further, stating that attempts to make theoretical assumptions more realistic "is certain to render a theory useless" (Friedman 1953, p. 30). Commentators have noted that "[w]hile "realism" is a name for members in a set of philosophical doctrines, "*realisticness*" characterizes features of representations". Thus, Friedman's essay constitutes a plea for using false theory for instrumental purposes. This means that

> the locus of appropriate criticism of any chunk of economics does not mostly lie at the level of general philosophical description of method, but rather at the level of how the method is used and how its use is constrained and what results it produces. (Mäki 2009, p. 33)

It is then not hard to recover a Friedman-like endorsement of the role of theory in design economics. Though the assumptions of the theory are unrealistic—as noted above, the ascription of literal truth to assumptions of rationality and self-interest is a "non-starter" (Alexandrova and Northcott 2009, p. 328)—what matters about economic theory from the perspective of a doctrine of philosophical realism is what "results it produces" (above).

[12]This sentiment was also echoed by Binmore and Klemperer (2002) after the UK's 2000 3G spectrum auction.

[13]Friedman (1953, p. 9) even notes the role that retrodiction plays in establishing a successful economic theory

[14]More specifically, Maki notes that Friedman can be read as "thinking that the assumptions of his theory have a definite truth value—namely, that of false—and that being unrealistic in this sense is a good thing" (1992, p. 179). This point is incidental to the central claim above: the philosophical realism Friedman can be read as endorsing is one similar to that of the 'economist as engineer', except that his focus is on novel prediction and not outcomes like matching stability or revenue maximization.

Milton Friedman's views of the role of theory in economics are diametrically opposed to the influential contemporary microeconomist Ariel Rubinstein, who provocatively argues that economic theory is a merely 'collection of fables' (Rubinstein 2006; Rubinstein 2012). This economic *theory-as-fables* view is the foil for the present chapter. On this view, economic theory ought "not aspire toward purposefulness... [and] does not engage in predictions and recommendations" (Rubinstein 2012, pp. 35–6). He explicitly contrasts his approach to economic theory[15] with (1) a view of economic theory which aims to make predictions about the real world and (2) a view of economic theory which aims to sharpen an economist's perception or intuition. By his own admission, Rubinstein is "obsessively occupied with denying any interpretation contending that economic models produce conclusions of real value" (Rubinstein 2012, p. 37). For Rubinstein, the modest goal of a "teller[] of fables" (Rubinstein 2006, p. 882) is all an economic theorist should hope for.

Rubinstein's theory-as-fables approach is complemented by economists who maintain "a heightened awareness of the gap between reality and its representation, coupled with a detached, bemused attitude to this gap" (Spiegler 2024, p. 175). This *ironic* reading of economic theory is a product of the distance between the representation and reality: it shies away from "taking the model seriously" (Spiegler 2024, p. 176) in the sense of offering policy prescriptions or scientific predictions. Although the economics profession "has [historically] been willing to sustain the irony-suffused culture of economic theory" (Spiegler 2024, p. 176), a more recent anti-irony turn is connected to the rise of design economics:

> The increasing appeal of the "market design" field lies in its practitioners' ability to go through the regular motions of an economic-theory exercise while insisting on a straightforward, nonironic connection to an economic reality. The "economist as engineer," as Al Roth (2002) called it; irony is not meant to be an engineer's thing. Market design methodology focuses on tightly regulated economic environments whose actors are expected to follow rigid rules. As a result, the gap between model and reality appears small enough to curb the irony impulse. (Spiegler 2024, pp. 177–8)

In documenting this turn, Spiegler (2024) is correct to point out the aspects of economic behavior these models fail to capture; however, his own account of the "curb[ing]" of the irony impulse is, I believe, exactly in line with where I take the successes of design economics to lie, as I will return to below.

How would someone persuaded by the theory-as-fable view respond to the explanatory challenge version of the no-miracles argument? First, one could object to the probabilistic conclusion. There is simply insufficient evidence to establish, in any reasonable form, the "miraculous" inference. Unlikely, perhaps; miraculous, no. As the evidence grows (i.e., the number of successful instances of design economics increases) this view would be subject to change. (After all, at some point every unlikely occurrence yields a miraculous interpretation). Note; however, this path is not open to someone who is "obsessively occupied with denying... that economic models produce conclusions of real value" (Rubinstein, above) or shies away from "taking the model seriously" (Spiegler, above). Alternatively, on the theory-as-fables view, other features—experiment, computation, personal connections, etc—might explain the successes of design economics but theory falls short of the mark[16]. This is equivalent to claiming that the effect of theory is reducible to random chance: it has the same status as the feature corresponding to 'it took place on a Tuesday'.

Here is what Rubinstein has to say about the 1994 FCC spectrum auctions:

> I personally know some of the people who planned this tender and similar tenders. They are undoubtedly bright and intelligent. They are also people with two feet firmly on the ground. However, to the best of my understanding, they based their recommendations on basic intuitions and human simulations, and not on sophisticated models of game theory. I do not find any basis for claiming that it was game theory that helped them in planning the tender. At most, these advisors were intimately familiar with a specific type of strategic considerations that we often study in game theory. (Rubinstein 2012, p. 125)

Here, we find evidence for the view that "basic intuitions and human simulations" (above) formed the basis for recommendations given by economic theorists to the FCC. Here, not only is Rubinstein rejecting the view that (1) economic theory can serve "as a basis for making predictions about the real world" (Rubinstein 2012, p. 34) but also the weaker view (2) that the objective of economic theory is to "sharpen perception"

[15]Rubinstein is principally concerned with economic *models*; however, we can extend his view to economic theory, viewed simply as a collection of models.

[16]Note, an endorser of the theory-as-fable view would even have a hard time admitting that theory was useful only when present in conjunction with other features.

(Rubinstein 2012, p. 34).

Clearly, on the subject of the role of theory in explaining the successes of design economics, opinions differ widely. Supposing that experiment and computation are common features across the range of successes there seems to be broad consensus that they are non-arbitrary: they genuinely contribute to explaining the success stories of design economics in a manner that is not reducible to random chance. Without denying the importance of these other features in explaining the successes of design economics a case needs to be made for the non-arbitrariness of theory. This is the role of the proceeding section.

2.5 The Role of Theory

The theory-as-fable view of the role of economic theory in explaining the successes of design economics entails a radical conclusion: theory is, effectively, useless. But sustaining this conclusion in the face of the explanatory challenge version of the no-miracles argument is no simple feat. A *fable-ist*[17] cannot contend that theory explains the successful cases only conjunction with other features (for then it would be practically useful and lose its status as a mere fable). Furthermore, the fable-ist cannot even dispute the probabilistic inference concerning a theory's role in explaining success for otherwise they would allow for the possibility that the fable might one day become practical advice (given sufficient evidence). Rubinstein's (2012) canonical formulation of the theory-as-fable view subsumes the contributions of theory under those of "basic intuitions and simulations" (above). The goal of this section is to show why this move is wrongheaded.

The argument first proceeds by establishing the premises of the explanatory challenge version of the no-miracles argument. The successes have already been established ($P1$) so I first begin by isolating the common features of theory across these cases ($P4$). Second, multiple arguments are given for why the role of theory is not arbitrary with respect to explaining the successes of design economics ($P5$). In particular, the argument is constructed with reference to classic formulations of scientific realism in the context of the natural sciences as well as Rubinstein's own earlier work on interpretations of game theory. There turns out to be a surprising connection between the two, which, in my view, helps account for the role of theory in the successful instances of design economics.

Though game theory is common to all instances of design economics, the extent to which it is used varies considerably. Notably, economist Tayfun Sönmez (2023), advocates for the use of an *axiomatic methodology* which is derived from the work of Hervé Moulin (1988). On his account, "game theory, mechanism design, experimental, computational, and empirical techniques assum[e] supporting roles." (Sönmez 2023, p. 13). Though it is fair to say that game theory is common to the successful instances of design economics (supporting premise $P4$), it is clear the degree to which game theory is used (let alone the particular theoretical model) varies considerably. Certainly, some theoretical primitives (e.g., random variables, utility functions, solution concepts, etc) are common throughout the range of successful cases; specific economic models and theories such as particular auction designs (English, Dutch, etc) are not. At a minimum, there is enough theoretical overlap between diverse fields within the broader approach of design economics to conclude that, for example, game theory is common auction design and the problem of residency matching.

Another feature common to the multitude of theories behind the successful instances of design economics is that they model the choices the economist (designer) subsequently gets to take. As noted by philosopher of science Francesco Guala (2001, 456, emphasis original):

> [t]heory can be used to produce new technology, by shaping the social world so as to mirror a model in all its essential aspects. The 'idealised' character of a theory may thus be a virtue rather than a defect, as the explicit role of theory is to point to a possibility. Theory *projects*, rather than describing what is already there.

The key idea here is that the world can be shaped to "mirror a model" (above). Guala calls this aspect of theory *projective*. Theories of economic design can be formulated to reflect decisions economists get to take. Classic models like Myerson's (**myerson1981**) optimal auction design model yield an auction format (an allocation and payment function) which can be subsequently implemented by the auction designer. This insight extends to counterexamples in theory that point to impossibilities in practice (for example, Myerson and Satterthwaite 1983). Theoretical results reflect circumstances which might actually come to pass. And in coming to pass, these circumstances can be made to reflect the theory that brought them about. This

[17] Equivalently, an adherent of the theory-as-fable view.

fact is true in virtue of economists' ability to shape policy. This projective quality is common to the diverse economic models used across the range of successful cases that canvassed above. This is another common theoretical feature that these cases all share.

It is this "mirroring" that ultimately reduces the gap between reality and representation. This point was noted by Speigler (2024) above, who remarked that it resulted in curbing that "irony impulse". Yet this mirroring can yield an incredibly tight correspondence between model and reality such as the social sciences have never before witnessed. For example, Google uses auction theory to determine which advertisements its users see. Google's chief economist Hal Varian (2009) has even sketched the mathematical details of the problem that confronts Google. Notice; however, that any implementation will be entirely computational: the functional form of the auction used by Google approximates the real-valued functions used in the mathematical model arbitrarily well up to floating point error. This renders many of the models of design economics *isomorphic* to their implementations in reality. Rubinstein (1991) explicitly denies this possibility[18]; however, I believe this rejection of my thesis is partly a function of it predating the success stories of the 1990s and the growth in electronic marketplaces and auctions, from Google and eBay to AirBnb and Amazon.

Could a fable-ist still press the issue? Could they grant that theory can be made to mirror the world—even approximating the real numbers up to floating point precision—but nonetheless fails to explain the successes of design economics? To do so, in light of the arguments presented above, they would need to maintain that the common theoretical elements of the successful cases are (1) reducible to "basic intuitions" (Rubinstein 2012, above)[19] and (2) that insofar as they are isomorphic, they are isomorphic to the part of reality that does not matter.

With regards to (2) above, philosophers of science working in economics have previously noted that "even if we grant that an auction model is partially isomorphic with reality, that fact is not very helpful in itself" because we cannot know if the model "is isomorphic to the part of reality that actually matters" (Alexandrova and Northcott 2009, p. 309). Though this is undeniably true, the fact that the locus of the decision that a designer subsequently makes (the auction format, the matching algorithm, etc) can be modeled with the fidelity of the natural sciences—recall those earlier notions of literal, semantic or approximate truth as used in physics—represents a significant departure from a version of 'local realism' for economics which rejects truth as a "non-starter" (Alexandrova and Northcott 2009, above). I believe the reduction in the gap between reality and representation at the point of decision matters for bringing about the successes of design economics. This is certainly *a* part that "actually matters" (Alexandrova and Northcott 2009, above). It might not be the most significant feature in explaining success; it might be required in conjunction with many other features. However, the fact that (1) our theoretical models can be made to mirror the world and (2) that this mirroring is tight (up to isomorphism, in the case of electronic auctions) suggest that there are theoretical features of design economics that help explain its successes in ways that are non-artbirary.

With regards to (1), the idea that the role of theory in explaining the successes of design economics is reducible to "basic intuitions", a number of economists working on auction design and matching problems strongly disagree. In the context of designing a matching algorithm for the NRMP (covered above), Al Roth goes as far as stating (Roth 2002, p. 1372), "[i]t turned out that the simple theory [of market design] offered a surprisingly good guide to the design, and approximated the properties of the large, complex markets fairly well". On this account, theory is useful in the first sense dismissed by Rubinstein: it has genuine *predictive value*. This strong view of the role of theory echoes the role of theory articulated by Milton Friedman (1953).

This view is not shared by all economists; however. In fact, the dominant view seems to be that "the real value of the theory is in developing intuition" (McAfee and McMillan 1996, p. 172)—the second view that Rubinstein (2012) rejects. Though individual contexts matter in determining which aspects of theory are relevant (i.e., there is no "one size fits all" approach (Binmore and Klemperer 2002, p. C94)), the role of theory is to sharpen and guide the intuitions of the economist. This view is also advocated by Al Roth in conjunction with noted game theorist Robert Wilson (2019). It has even been dubbed the "mainstream paradigm for market design" by Tayfun Sönmez (2023, p. 10). Ultimately, this argument devolves into an

[18]More specifically, in his opinion, this feature is undesirable: "models are not supposed to be to isomorphic to reality" (Rubinstein 1991, p. 918). Here, he is clearly following Friedman's (1953) dictates on the role of unrealistic assumptions.

[19]As noted above, accepting that theory explains the successes of design economics in conjunction with another factor (e.g., "human simulations" (Rubinstein 2012, above)) still cedes too much ground: the resulting conclusion cannot be that economic theory is merely a collection of fables.

attempt to reconcile mutually contradictory assertions between those who "do not find any basis for claiming it was game theory that helped" (Rubinstein 2012, above) plan the successes of design economics and those who maintain the predictive value of theory or its role in shaping intuition (Roth and Wilson 2019).

In my view, fable-ist attempts to cling to the position that theory does not explain the successes of design economics are increasingly hard to maintain in the face of the growing number and extent of the successes. I find the probabilistic nature of the no-miracles argument compelling. The alternative—that economic theory is a mere collection of fables and is not useful—not only runs headlong into the contravening claims of those very economists who brought about the successes of design economics but fails to account for the projective quality of these theories. To maintain theory is *that* useless involves, effectively, invoking a miracle. And for this reason, I agree with the conclusion ($C2$) of the explanatory challenge version of the no-miracles argument: theory explains the success of design economics.

2.6 Conclusion

The successes of design economics are hard to deny: from auction design to matching new entrants in labor markets, the corner of economics concerned with institutional design has demonstrated a remarkable ability to achieve a diverse range of objectives. Though the track record is far from spotless, successful auctions and matching markets have been replicated the world over, growing in complexity and scale. I have argued that these successes warrant an explanation. What features of the science of design economics can account for its successes? In doing so, I have argued for a 'local realism' wherein we use the success of economics as a basis for grounding claims about economics' ability to interact with an external world.

Putting success at the center of a philosophical argument for (some form of) scientific realism straightforwardly recovers a no-miracles type argument for realism. However, instead of an argument for the truth of a scientific theory, here I focus on features of the science that contribute to explaining its success. And in the case of design economics, there are many such features. Those covered here include experiment and computation, as well as more "political" (Roth 2002, p. 1345) features such as personal and professional relationships. In my view, theory is also an important feature that explains the successes of design economics. This position has its critics. Notably, economic theorists like Ariel Rubinstein (2006; 2012) articulate something I call the *theory-as-fable* view. On this view, theory is not to be taken seriously. It is not intended for prediction. It cannot offer policy recommendations. It does not even sharpen the intuitions of economists. Instead, economists should be "satisfied even if the economic model is merely interesting" (Rubinstein 1982, p. 36). This deflationary view of economic theory removes it from contention as a potential feature that can explain the successes of designed economics.

Yet the fable-ist's position is not easy to maintain. First, there are clear theoretical commonalities in the range of successful instances of design economics. Despite differences in degrees of application, game theory is common to all of the successes of design economics. Furthermore, in my view, the *projective* quality of these theories and the tightness between their representations and reality renders the explanation non-arbitrary (not reducible to random chance). The probabilistic inference of the explanatory challenge version of the no-miracles argument is then not so easily dismissed by a fable-ist. They cannot claim the conclusion 'theory explains the successes of design economics' is merely unlikely (not miraculous) for then the fable might one day become practical advice. They cannot claim the conclusion is true in virtue of the fact that theory explains success only in conjunction with another feature. This would then make theory scientifically useful—a property fables do not possess! Thus, I do not think the falbe-ist's position is tenable. It cannot fit the facts: design economics has been successful and there exist theoretical commonalities which are not reducible to random chance. These commonalities help explain its success. To assert otherwise is to make a miracle of the contemporary science of economics.

I am far from alone in articulating this argument. Economists who have worked on the success stories of design economics espouse a range of alternative positions, from the view that "the real value of the theory is in developing intuition" (McAfee and McMillan 1996, p. 172)—again, this is the "mainstream paradigm for market design" (Sönmez 2023, p. 10)—to the stronger claim that that the theory of market design has predictive value (Roth 2002). Rubinstein is nonetheless insistent: what matters, if theory matters at all, is "basic intuitions" (Rubinstein 2012, above). Perhaps there is a way to reconcile Rubinstein's views with those who claim the role of theory is to sharpen intuition. Rubinstein might simply underestimate what

constitutes a "basic" intuition. In the words of economists Ken Binmore and Paul Klemperer (2002, p. C95), who consulted on the UK's 2000 3G spectrum auction,

> [b]ut perhaps the most important lesson of all is not to sell ourselves too cheap. Ideas that seem obvious to a trained economist are often quite new to layfolk.

Future research could show a reconciliation is possible. What might be "basic" for Rubinstein might nonetheless represent a genuine scientific insight for those economists consult with. Ultimately, addressing the extent of this possibility is beyond the scope of this chapter.

In my view, economist Ran Spiegler (2024) is correct to point out the subtle change in design economists' attitudes towards their work. He incisively writes,

> [i]f I had to name one major shift in the sensibilities of economic theorists in the past half century, a prime candidate would be the way we conceptualize markets—from quasi-natural phenomena admired from afar to manmade institutions whose design can be tweaked by economist-engineers. (Spiegler 2024, p137)

Unlike Spiegler, I believe this change in perspective is unequivocally a good thing for *science*. Markets, in a distinctly concrete sense, are socially constructed: we build them, regulate them, correct them, police them, etc. The change in sensibility brought about by design economics is mirrored in the projective quality of its theory. Economists model the control they exert over an environment or institution. This "tweaking" reflects a genuine rise in the power yielded by economists in the 21st Century. It remains to be seen whether this a good thing for *society* (Hitzig 2024).

Chapter 3

Simulations on Optimal Auctions in Multidimensional Settings

I explore the properties of optimal multi-dimensional auctions in a setting where a single object of multiple qualities is sold to several buyers. Using simulations, I test the hypothesis that the optimal mechanism is an *exclusive buyer mechanism*, where buyers compete to be the right to be the only buyer to choose between quality levels of a good. I find compelling evidence of the optimality of the exclusive buyer mechanism in multi-dimensional settings and explore other conjectures concerning the set of types excluded from the mechanism in equilibrium in optimal multidimensional auctions.

3.1 Introduction

Since the seminal work of Roger Myerson (1981), revenue-maximizing auctions in the case of a single dimension (e.g., good) are well known. However, little is known about optimal auctions in settings with multiple dimensions of value. The problem is incredibly complex: despite sustained research efforts for decades, only in the past few years have we discovered how to optimally sell two goods to a single buyer (Daskalakis, Deckelbaum, and Tzamos 2017). In this chapter, I focus on the multidimensional setting of a single good with multiple quality levels and explore whether a specific mechanism—the *exclusive buyer mechanism*—is optimal. Building off existing work at the intersection of economics and computer science, I adopt a novel approach involving extensive simulations to explore the optimality of this particular mechanism. The approach adopted here makes use of a new approximation algorithm to uncover the qualitative features of optimal mechanisms in the particular multidimensional setting of a single good with multiple quality levels[1].

The setting of a single good with multiple quality levels can be understood as corresponding to the case of purchasing a single airplane ticket where the 'quality level' might be: economy class, business class, or first class. A buyer might prefer, say, business over economy class. Each quality level can be (and very often is) given a different price and might have different value to the buyer. Crucially, unlike the multidimensional setting where multiple goods are sold by a seller to an arbitrary number of buyers, here since only one good is sold so 'bundling'—selling subsets of all offered goods for a discount—cannot occur. The absence of bundling renders this a much simpler setting and an ideal point of departure for investigating the qualitative features of optimal multidimensional auctions.

In this thesis chapter, I am particularly concerned with investigating the optimality of the exclusive buyer mechanism in the setting of selling a single good with multiple quality levels to multiple bidders. This mechanism can be understood as an (either a second price or ascending bid) auction with a reserve price for the exclusive right to be the buyer of one of the quality levels of the good. The intuition for why this mechanism might be optimal can be reduced to: when there is a single good for sale with multiple dimensions of value, it is enough to consider the bidder who values a single quality level more than any other bidder values any other quality level. The exclusive buyer mechanism has previously been explored in multidimensional settings (e.g., Brusco, Lopomo, and Marx 2011; Belloni, Lopomo, and Wang 2010) but

[1]The codebase for this project is available at `https://github.com/jmemich/optimal-auction-multidim`.

this thesis chapter represents the first sustained exploration of its optimality in a wide range of settings. This chapter develops the exclusive buyer mechanism in much broader generality, facilitating an extended investigation into its optimality in a broad range of multidimensional settings.

The approach adopted in this chapter is to eschew analytic results in settings where few have been forthcoming in favor of exploring these complex, analytically intractable settings using simulations to approximate optimal mechanisms. The key idea is simple. Computer scientists and economists have made significant strides developing approximation algorithms that can arbitrarily well approximate optimal revenue using linear programming techniques (e.g., Cai, Daskalakis, and Weinberg 2012; Belloni, Lopomo, and Wang 2010). Taking these algorithmic developments as a point of departure, it is possible to explore a wide class of settings where a single good with multiple quality levels is sold to better understand important qualitative features of the optimal mechanism. Once an approximation algorithm is run to uncover the optimal mechanism, a qualitative analysis of the key features of the optimal mechanism then proceeds. Important questions explored in this thesis concern: Do these approximation algorithms yield deterministic optimal mechanisms or is randomization always required for revenue maximization? Is a positive measure of buyers always excluded from the allocation in equilibrium? Are there simple, intuitive mechanisms that might characterize the results from the approximation algorithms? These and related questions are explored in this chapter. Ultimately, the goal is to investigate the optimality of the exclusive buyer mechanism using simulations.

Across a broad range of settings, I find evidence for the optimality of the exclusive buyer mechanism and explore several related conjectures about the qualitative characteristics of optimal mechanisms using simulations. This mechanism effectively recovers the revenue achieved by the optimal mechanism in all settings considered here. Additionally, the exclusion region—the set of types excluded by the mechanism in equilibrium—is identical for all N considered in each setting. Finally, although there is considerable evidence that the interim allocations of the optimal mechanism yielded by the approximation algorithm are qualitatively similar to those of the exclusive buyer mechanism, there are settings where the optimal mechanism involves randomization and the allocations of the exclusive buyer mechanism fail to capture the behavior of the optimal mechanism. These results are designed to guide future theoretical research by way of conjecture: support for the conjectures presented in this chapter should help future research undertake more promising avenues of research in a complex and challenging field.

The structure of this chapter proceeds as follows. In section (3.2), we review the existing literature on multidimensional mechanism design, approximation algorithms, and specific works in the setting of a single good with multiple quality levels. The problem is formally introduced in section (3.3), and a description of the exclusive buyer mechanism can be found in section (3.3.1). Section (3.4) describes the approach of the chapter and explores several hypotheses concerning optimal multidimensional mechanisms. Finally, I conclude and discuss these results in section (3.5), highlighting the implications of these results for theoretical microeconomics. An Appendix with a complete description of the approximation algorithm developed for this thesis can be found in section (7.1).

3.2 Literature Review

The specific case considered here of selling a single good with multiple quality levels to an arbitrary number of buyers is a special case of the more general multidimensional mechanism design problem of selling an arbitrary number of goods to an arbitrary number of buyers. Since the groundbreaking work of Roger Myerson (1981) who solved the optimal auction design problem in the case of single-dimensional types, economists have sought to characterize optimal auctions in the more general multidimensional setting, with limited success. At this juncture, it is widely accepted that "[e]ssentially, nothing has been known about optimal auctions in this [multidimensional] setting" (Kolesnikov et al. 2022, p1). This literature review covers historical and recent developments by economists and computer scientists who have sought to uncover characteristics of optimal mechanisms in multidimensional settings, with a particular focus on the case of a single good with multiple quality levels.

In the case of single-dimensional types, early work on optimal mechanism design demonstrated the optimality of deterministic mechanisms (i.e., reserve or 'take-it-or-leave-it' prices) (Myerson 1981; Riley and Zeckhauser 1983). These approaches leveraged the approach of integration-by-parts (as used in Mussa and

Rosen 1978) to solve the relaxed optimal mechanism design problem without directly considering incentive-compatibility constraints. These early results cannot be generalized to multi-dimensional settings because the integral solution to the optimization problem is path-dependent; any two points in the multi-dimensional typespace can be connected by a continuum of paths. Thus, a major breakthrough in multidimensional auction design came with the use of duality-based approaches to multidimensional screening developed by (Rochet and Choné 1998) which circumvents this problem.

The duality-based approach (Rochet and Choné 1998) builds on the single-dimensional nonlinear pricing framework of (Mussa and Rosen 1978), which was given its canonical formulation in multidimensional settings in (Wilson 1993) and (Armstrong 1996). This work takes as its point of departure the approach of (Mirrlees 1971) on optimal taxation and relies on results that establish the implementability of a decision rule in multidimensional settings (Rochet 1987). In this multidimensional screening problem, a few key findings emerge. The first is that 'bunching'—a situation where multiple types are treated identically in the optimal solution—is a "robust" feature of multidimensional screening (Rochet and Stole 2003; Rochet and Choné 1998). There are two types of bunching: in the first case, a set of types of positive measure are excluded from purchasing the goods in the optimal solution (this is commonly known as the 'exclusion region' of the type space); in the second case, a non-negligible set of types outside the exclusion region receive the same product although they have different tastes. In addition, the work of (Rochet and Choné 1998) illustrates that the optimal solution to multidimensional screening problems may involve "bundling" the goods, which involves selling multiple goods together.

In multi-item settings, authors have long sought to characterize when bundling multiple goods in a single contract is optimal for the seller. Bundling strategies available to a seller include 'pure' bundling, where only the bundle of all goods is offered to sellers, and 'mixed' bundling, where each different bundle of items is priced separately. Early results showed that offering mixed bundles strictly dominates offering pure bundles to the buyers (Adams and Yellen 1976; McAfee, McMillan, and Whinston 1989) and more recent results have demonstrated that randomized bundles may dominate mixed bundles (Thanassoulis 2004; Daskalakis, Deckelbaum, and Tzamos 2017). In these settings, the optimal menu of contracts may include infinitely many randomized bundles (Manelli and Vincent 2007; Hart and Nisan 2019). Additionally, recent work has demonstrated settings where simply offering only the grand bundle of all goods is optimal (Haghpanah and Hartline 2021).

In the past few years, major breakthroughs in optimal multidimensional mechanism design have come from the use of the methods of optimal transport applied to the optimization problems of microeconomic theory (see Ekeland 2010). These results (Daskalakis, Deckelbaum, and Tzamos 2017; Kolesnikov et al. 2022) greatly aid the *certification* of optimality: the techniques of optimal transport facilitate the identification of the dual of the seller's optimization problem from which a given mechanism's optimality can be verified. Thus, previously existing results that characterize optimal mechanisms in specific settings (for example, where valuations for two goods are i.i.d on $U[0,1]^2$ (Pavlov 2011; Manelli and Vincent 2006)) can be shown to be optimal using a novel, more general approach. The success of the tools of optimal transport in mechanism design is due to the success of a 'guess-and-verify' approach where one guesses a solution to the primal problem and then the dual solution plays the role of a certificate of optimality for the initial guess.

These breakthroughs which facilitate the certification of optimality are particularly helpful when viewed in light of the growth in work at the intersection of economics and computer science. One line of work (Chawla, Hartline, and Kleinberg 2007; Cai, Daskalakis, and Weinberg 2012; Cai, Devanur, and Weinberg 2016; Belloni, Lopomo, and Wang 2010; Alaei et al. 2019) provides an algorithmic approximation of optimal mechanisms in multidimensional settings. Here, the buyer's typespace is discretized, and linear programming techniques are used to approximate the optimal solution, often using simple mechanisms like posted prices. Work in this area aims to achieve a constant factor of the optimal revenue achievable by a Bayesian incentive-compatible mechanism through an approximation. Other work at the intersection of computer science and economics offers insights into the nature of the optimal mechanisms in multidimensional settings. These works show that in specific settings, optimal mechanisms contain only a few contract points (Wang and Tang 2014) or that menus with only a finite number of items cannot ensure any positive fraction of optimal revenue (Hart and Nisan 2019). Further work in *automated mechanism design* circumvents the need to determine the optimal mechanism manually in favor of creating the mechanism from the setting and objectives (Conitzer and Sandholm 2002; Conitzer and Sandholm 2004). This line of research is ongoing (see, for example, Conitzer et al. 2021) and has even prompted new avenues of research using neural networks to discover

optimal mechanisms (Dütting et al. 2024).

Returning to the specific context of multidimensional mechanism design in the case of a single good with multiple quality levels, the work of Belloni, Lopomo, and Wang (2010) provides insight into the character of the optimal mechanism in this particular setting. Applying their algorithm to concrete cases, they find a number of surprising results from their simulations. First, there is clear evidence that in the optimal solution, a measure-zero set of buyers is excluded from the allocation in equilibrium. This stands in marked contrast to results in the multi-item case which show that the optimal solution requires exclusion (Rochet and Choné 1998; Armstrong 1996). Second, their results indicate the optimality an *exclusive-buyer mechanism*: it performs "quite well" relative to the numerical optimal solutions and that it "shares many of its defining features with its one-dimensional counterparts" (Belloni, Lopomo, and Wang 2010, p1085-6), including being implementable in dominant strategies. This mechanism involves an auction (with a reserve price) among buyers for who gets to be the sole recipient of the good. A premium can then be paid for whichever quality grade the winning buyer desires. Interestingly, this mechanism is entirely deterministic in the single-bidder case. This is particularly surprising because, in the neighboring multi-item case, randomized allocations are widely considered necessary for revenue maximization (Daskalakis 2015).

The theoretical study of exclusive-buyer mechanisms originates from the phenomenon of 'contingent re-auctions' where sellers will modify objects sold to benefit themselves or the general public (Brusco, Lopomo, and Marx 2011). For example, in the context of the US Spectrum License Auction 73 held in 2008[2], the US government adopted a contingent re-auction format where it offered restricted spectrum licenses first, and committed to re-auction the licenses without many of the restrictions in the case the reserve prices were not met. Brusco, Lopomo, and Marx (2011) show that an exclusive-buyer mechanism can always be parameterized such that the mechanism induces the efficient outcome in dominant strategies. However, outside of a restrictive context where all bidders' valuations for the restricted object are a fixed percentage of the unrestricted object, no general results concerning the optimality of the mechanism are provided.

Analytic results concerning optimal multidimensional auctions for a single good with multiple quality levels and a single bidder are scarce. Notably, (Pavlov 2011) investigates the case where the bidder's valuations for the object are uniformly distributed on the unit square $[c, c+1]^2$. Pavlov finds that the optimal mechanism varies considerably with c and sometimes requires randomization for revenue maximization. This approach was further generalized in the work of (Thirumulanathan, Sundaresan, and Narahari 2019a) who study an almost identical case where a bidder's valuations are distributed uniformly on the rectangle $[c_1, c_1 + b_1] \times [c_2, c_2 + b_2]$. Similarly to (Pavlov 2011), the solution to the optimal mechanism design problem entails both deterministic and stochastic contracts. Surprisingly, however, (Thirumulanathan, Sundaresan, and Narahari 2019a) find evidence of settings where optimal mechanisms do not exclude a position measure of buyers. Additionally, a working paper by (Haghpanah and Hartline 2014) gives sufficient conditions for the optimality of posting a single, uniform price for all quality levels of a good, albeit in a restricted class of settings.

Analytic results for optimally selling substitute goods have also been given in the Hotelling model (Hotelling 1929) where two horizontally differentiated goods are located at the endpoints of a segment. In this setting, (Balestrieri, Izmalkov, and Leao 2020) find that stochastic contracts are part of the optimal mechanism. The economic intuition that arises from this body of research is clear: by offering a lottery over which good the bidder receives, a seller can offer a discount to entice marginal buyers who would otherwise choose the outside option. Similarly, (Loertscher and Muir 2023), find that in this setting randomization is required by the seller to maximize revenue. These results support earlier work (Thanassoulis 2004) which shows that in the standard auction design problem for two substitute goods, the seller can always increase revenue by including stochastic contracts alongside take-it-or-leave-it prices in the optimal mechanism. As noted above, these results support the view that in multidimensional settings randomization is required to maximize revenue (see Daskalakis 2015).

The approaches of (Pavlov 2011; Thirumulanathan, Sundaresan, and Narahari 2019a) to solving the mechanism design problem for the case of a single good with multiple quality levels follows the work of (Guesnerie and Laffont 1984), where optimal control theory is used to address the fact the measure of participating types endogenously depends on the mechanism. The optimal control theory approach[3] has also

[2]For more details see (Brusco, Lopomo, and Marx 2009).

[3]See (Basov 2005, §7) for an extended discussion of the different approaches to multidimensional mechanism design and their respective strengths and weaknesses.

been successfully applied to single-dimensional settings when the participation constraints are endogenously determined by the mechanism (Jullien 2000). Here, the bidder's reservation utility depends on their type. This approach generalizes to accommodate the fact that the measure of participating types in a given mechanism is endogenously determined (for example, in the multidimensional case of a single good with two quality levels and a single buyer considered by Pavlov 2011; Thirumulanathan, Sundaresan, and Narahari 2019a).

Finally, although it has long been believed that it is always profitable for the seller to exclude some measure of bidders (Rochet and Choné 1998; Armstrong 1996) in multidimensional settings, recent theoretical and computational work in auction design in these settings has challenged these conclusions. The original result of (Armstrong 1996) demonstrated that in multi-product settings with a single bidder, the seller benefits from always excluding a positive measure of bidder types. By relaxing Armstrong's strong assumptions about the bidder's utility function and the convexity of the type space this result has been extended and it has been shown that "exclusion is generically optimal in a large class of models" (Barelli et al. 2014, p. 75). The intuition is as follows: in a multidimensional screening problem of dimension m, when the seller raises the price by $\epsilon > 0$ then they earn extra profits of order $O(\epsilon)$ from the remaining bidder types but the measure types excluded from the mechanism is of order $O(\epsilon^m)$. However, simulation results from (Belloni, Lopomo, and Wang 2010) suggest that in certain asymmetric settings, this intuition fails and it is optimal for the seller not to exclude any bidder types. This finding is corroborated by the theoretical work of (Thirumulanathan, Sundaresan, and Narahari 2019a) where the optimal mechanism for a single bidder with valuations distributed uniformly on a rectangle will include settings without exclusion.

3.3 Model

In this section, I introduce the optimal multidimensional auction design problem for a single with multiple quality levels in addition to the specific *exclusive buyer mechanism*. The formulation of the problem adopted here is drawn from Belloni, Lopomo and Wang (2010) and is akin to other, canonical formulations of auction design in the multi-bidder, multi-item case (e.g., Cai, Devanur, and Weinberg 2016). There is one seller wishing to sell one item with $j = 1, \ldots, K$ quality levels to $i = 1, \ldots, N$ bidders. Bidder i's valuation (their type) of quality level j is denoted $X_j^i = [\underline{x}_j^i, \overline{x}_j^i] \subset \mathbb{R}_+$. Each bidder's vector of valuations is given by $X^i = \prod_j X_j^i$ and I will denote by $X = \prod_i X^i$. I will denote by X^{-i} the types of all bidders except for i. Bidder i's type is private information and is known only to themselves.

Bidder i's valuation for quality level j is distributed according to the cumulative density function F_j^i. The joint density of all bidders' valuations of all quality grades is denoted F, and, again, denote by F^{-i} the distribution of types of all bidders except bidder i. The joint density is known to the seller. It is assumed that F is continuously differentiable. Furthermore, as is common in the setting of Myerson (1981), it is assumed that the distributions of bidders' valuations are independent and identical. However, a bidder's valuations across quality grades may be correlated.

A crucial step to solving the optimal auction design problem was the use of the *revelation principle* which simplifies the search space for optimal mechanisms (see Myerson 1981, Lemma 1). The revelation principle allows the auction designer to restrict their attention to a class of mechanisms called *direct mechanisms*. Direct mechanisms are those where the bidders simultaneously and confidentially reveal their types to the seller and the seller decides who gets the object and how much each bidder must pay, as a function of their types.

Thus, a direct mechanism is described by a pair of functions (q, p). The *allocation function* $q : X \to [0,1]^{KN}$ specifies the probability $q_j^i(x)$ for some $x \in X$ that bidder i receives the good with quality level j. Note that in deterministic mechanisms $q_j^i(x) \in \{0, 1\}$ for all $x \in X$. The *price function* $p : X \to \mathbb{R}^N$ specifies the amount each bidder pays (bidders might be required to pay even if they do not receive the good, as occurs in an 'all-pay' auction).

The utility functions of the seller and bidders are risk-neutral and additively separable. The bidders' utilities are given by

$$u^i(x) = \sum_j x_j^i q_j^i(x) - p^i(x) \tag{3.1}$$

for all $x \in X$. Denote bidder i's expected utility as

$$U^i(x^i) = \int_{X^{-i}} u^i(x^i, x^{-i}) dF^{-i}(x^{-i}) \tag{3.2}$$

for all $x^i \in X^i$. I assume for simplicity of presentation that costs are zero. The seller's utility function is given by

$$u^0(x) = \sum_i p^i(x) - \sum_i \sum_j r_j q_j^i(x) \tag{3.3}$$

where r is the seller's value estimate for the object, which is most commonly interpreted as the reserve price. Thus, the seller's expected utility is given by

$$\int_X u^0(x) dF(x) \tag{3.4}$$

However, not every pair of functions (q, p) represents a *feasible* auction mechanism. There are three types of constraints[4] that must be imposed on (q, p).

First, since there is only one object to be allocated, the allocation function must satisfy the following feasibility conditions (F):

$$\sum_i \sum_j q_j^i(x) \leq 1 \text{ and } q_j^i(x) \geq 0 \tag{F}$$

for all $i = 1, \ldots, N$, $j = 1, \ldots, K$ and $x \in X$. Note that, in contrast to the multidimensional setting of a single good with multiple quality levels, in the multi-item case where the seller has K goods to sell the probability conditions are given by $\sum_i q_j^i(x) \leq 1$ and $q_j^i(x) \geq 0$ for all $i = 1, \ldots, N$, $j = 1, \ldots, K$ and $x \in X$.

Second, the mechanism (p, q) must be *individually rational* (IR) in the sense that every bidder has non-negative expected utility from participating in the mechanism. More formally,

$$U^i(x^i) \geq 0 \tag{IR}$$

for all bidders $i = 1, \ldots, N$ and all $x^i \in X^i$.

Third, the revelation mechanism can only be implemented if no bidder can expect to gain from lying about their type. If bidder i misrepresents their true type x^i with the lie $\widehat{x}^i$ their expected utility would be

$$\int_{X^{-i}} \sum_j x_j^i q_j^i(\widehat{x}^i, x^{-i}) - p^i(\widehat{x}^i, x^{-i}) dF^{-i}(x^{-i}) \tag{3.5}$$

Thus, in a direct mechanism it is necessary to ensure

$$U^i(x^i) \geq \int_{X^{-i}} \sum_j x_j^i q_j^i(\widehat{x}^i, x^{-i}) - p^i(\widehat{x}^i, x^{-i}) dF^{-i}(x^{-i}) \tag{IC}$$

for all $i = 1, \ldots, N$ and $x^i, \widehat{x}^i \in X^i$. This final condition is known as *Bayesian incentive compatibility.*

The revenue maximization problem faced by the seller is therefore

$$\begin{gathered} \max_{p,q} \int_X \left(\sum_i p^i(x) - \sum_i \sum_j r_j q_j^i(x) \right) dF(x) \\ \text{subject to } (F), (IR), (IC) \end{gathered} \tag{OPT}$$

which I shall refer to as OPT throughout.

It is possible to reformulate the objective function to remove the dependence of the dimensionality N. This is particularly helpful for designing approximation algorithms (see, for example, Belloni, Lopomo, and Wang 2010). This is done using *interim* variables (Q, U), obtained by integrating out all but one type of buyer. The interim probability that buyer i is awarded the object quality grade j is

$$Q_j^i(x^i) = \int_{X^{-i}} q_j^i(x^i, x^{-i}) dF^{-i}(x^{-i}) \tag{3.6}$$

[4]Here, we outline (IR) and (IC) constraints when the solution concept is a Bayesian Nash equilibrium.

and the interim expected utility of each buyer can therefore be defined as

$$U^i(x^i) = \sum_j x^i_j Q^i_j(x^i) - \int_{X^{-i}} p^i(x^i, x^{-i}) dF^{-i}(x^{-i}) \tag{3.7}$$

It is then possible to rewrite the above constraints that used the *ex-post* allocations and transfer functions (q, p) in terms of interim variables (Q, U). For each bidder i, the incentive-compatibility constraint IC becomes:

$$U^i(x^i) - U^i(\hat{x}^i) \geq \sum_j Q^i_j(x^i)(x^i_j - \hat{x}^i_j) \quad \forall x^i, \hat{x}^i \in X^i \times X^i \tag{IIC}$$

Additionally, the interim individual rationality constraint becomes:

$$U^i(x^i) \geq 0 \quad \forall x \in X \tag{IIR}$$

Thus, since we only consider the case where all bidder's valuations are identically distributed (and, therefore, we can omit the superscript i) the objective function OPT can be rewritten in terms of interim variables (Q, U):

$$\begin{gathered} \max_{Q,U} \int_X \left(\sum_j (x_j - r_j) Q_j(x) - U(x) \right) dF(x) \\ \text{subject to } (F), (IIR), (IIC) \end{gathered} \tag{OPT$_*$}$$

Furthermore, the feasibility constraints (F) can be rewritten using (Border 1991)[5].

3.3.1 Exclusive Buyer Mechanism

We can now develop a formal description of the exclusive buyer mechanism by considering an analog of the single dimensional case. In the single dimensional case, a good is allocated according to a bidder's *virtual value*, which is a function of the bidder's type representing the surplus that can be extracted from the bidder (see Myerson 1981). In the multidimensional case of a single good with multiple quality levels we can define a multidimensional analog of a single dimensional virtual value. For each bidder i and each quality grade j of the good we define a virtual value $\beta^i_j(x)$, which is also a function of each other bidder's types. For each bidder, we define $\beta^i = \max_j \beta^i_j$ as the maximum of the quality grade-specific virtual values. (Note, although the virtual values may depend on the reserve price, we omit the notational dependence on r for clarity.) The key idea behind an exclusive buyer mechanism is that the good is allocated to the bidder i with the largest β^i.

This formulation of the exclusive buyer mechanism has its origins in the work of Brusco, Lopomo, and Marx (2011), who first explored a similar mechanism in the specific case of two quality levels. Their mechanism can be understood as an auction where the buyers compete in a second price or ascending-bid auction (with reserve prices) for the right to be the only buyer and choose which quality grade to purchase. If a bidder wins the auction they can select between the lower quality grade of the item or the higher quality grade (and pay an additional price). Their mechanism was further elaborated in a follow up work (Belloni, Lopomo, and Wang 2010) from which a number of a number of conjectures in this chapter are drawn.

Formally, in our more general context, we can define the set of bidders who are allocated the good as follows. Allowing for ties, let M denote the set of bidders with the largest β^i:

$$M(x) = \{i \mid \beta^i > \beta^{i'} \ \forall i' \neq i \text{ and } \beta^i \geq 0\} \tag{3.8}$$

Then the allocation q is defined as:

$$q^i_j(x) = \begin{cases} \frac{1}{|M(x)|} & i \in M(x) \text{ and } \beta^i_j = \max_{j'} \beta^i_{j'} \\ 0 & \text{otherwise} \end{cases} \tag{3.9}$$

Notice the similarity with the canonical single dimensional formulation of Myerson (1981). Again, it is worth emphasizing that at this juncture we have neither constrained the shape of function β^i_j nor have we restricted

[5]See Appendix 7.1 for details about how this affects computational performance.

its domain (it may also depend on other bidders' types). However, in this chapter, we consider the specific case of *linear* virtual values defined by

$$\beta_j^i = x_j^i - r_j \tag{3.10}$$

where r_j is the reserve price associated with quality level j. Surprisingly, as will become clear in Section 3.4, this simple functional form is a promising point of departure to explore the optimality of the exclusive buyer mechanism.

We can now express the revenue as a function of the allocation. Supposing there are no ties, note that the *interim* or *expected* allocation Q_j^i can be written:

$$Q_j^i(x^i) = \int_{X^{-i}} q_j^i(x^i, x^{-i}) dF^{-i}(x^{-i}) \tag{3.11}$$

$$= \underbrace{\mathbb{1}\{\beta_j^i \geq \beta_{j'}^i,\ \forall j' \neq j \text{ and } \beta_j^i \geq 0\}}_{j \text{ is } i\text{'s preferred quality grade}} \cdot \underbrace{\int_{X^{-i}} \mathbb{1}\{\beta^i > \beta^{i'}\ \forall i \neq i'\} dF^{-i}(x^{-i})}_{\text{probability } i \text{ wins}} \tag{3.12}$$

$$= \mathbb{1}\{\beta_j^i \geq \beta_{j'}^i,\ \forall j' \neq j \text{ and } \beta_j^i \geq 0\} \cdot F(\max\{\overline{x}_1, r_1 + \beta_j^i\}, \ldots, x_j^i, \ldots, \max\{\overline{x}_J, r_J + \beta_j^i\})^{N-1} \tag{3.13}$$

Where we make use of the fact that, when bidders' valuations are independent and identically distributed, $F^{N-1}(x) = F(x)^{N-1}$.

The price p can be found by integrating along any path ℓ from $\mathbf{0} \to x$. Here, we consider rectangular paths formed by the standard basis of the vector space X. For example, in the case of a good with two quality levels, the price for bidder i is given by the path $\ell(x^i) : (0,0) \to (x_1^i, 0) \to (x_1^i, x_2^i)$. More generally,

$$p^i(x) = \max_j \int_{\ell(x^i)} q_j^i(\ell(t^i), x^{-i}) \cdot \ell'(t^i) dt^i \tag{3.14}$$

Thus, the interim or expected price P^i is given by

$$P^i(x^i) = P^i(0) + \max_j \int_{\ell(x^i)} Q_j^i(\ell(t^i)) \cdot \ell'(t^i) dt^i \tag{3.15}$$

Note that $P(0) = 0$ (i.e., the agent with the lowest type pays zero).

Finally, we can now rewrite the objective OPT as follows:

$$OPT = \max_{p,q} \int_X \left(\sum_i p^i(x) - \sum_i \sum_j r_j q_j^i(x) \right) dF(x) \tag{3.16}$$

$$= \max_{p,q} \int_{X^i} \int_{X^{-i}} \sum_i \big(p^i(x^i, x^{-i}) - r \cdot q^i(x^i, x^{-i})\big) dF^{-i}(x^{-i}) dF^i(x^i) \tag{3.17}$$

$$= \max_{p,q} \int_{X^i} \sum_i \big(\max_j \int_{\ell(x^i)} Q_j^i(\ell(t^i)) \cdot \ell'(t^i) dt^i - r \cdot Q^i(x^i) \big) dF^i(x^i) \tag{3.18}$$

Ultimately, the expression above captures the seller's optimization problem when bidder valuations are independent and identically distributed and their virtual values β_j^i are linear.

3.4 Conjectures and Simulations

In this section, I explore whether the exclusive buyer mechanism (exclusive buyer mechanism) approximates the optimal revenue in settings where analytic results are known or those where they can be approximated arbitrarily well algorithmically. The goal is to uncover the qualitative features of the optimal mechanism in the multidimensional setting of a single good with multiple quality levels. The structure of this section is as follows. First, we introduce the strategy adopted to investigate the qualitative features of optimal mechanisms in section (3.4.1). Next, in section (3.4.2), we outline the specific conjectures investigated in this chapter. Finally, we investigate the optimal mechanisms that result from our approximation algorithms in a wide range of settings in section (3.4.3).

3.4.1 Methodology

The strategy we adopt to investigate the qualitative properties of optimal mechanisms in multidimensional settings is:

1. For a given setting, determine the optimal mechanism from the results of running the approximation algorithm (outlined in Appendix 7.1).
2. Once the optimal mechanism is found, examine the discretized interim allocations yielded by the approximation algorithm. The idea is to visually represent the discretized allocation for a quality level of the good (e.g., Figure 1) and find a mathematical expression that corresponds to the approximately optimal mechanism.

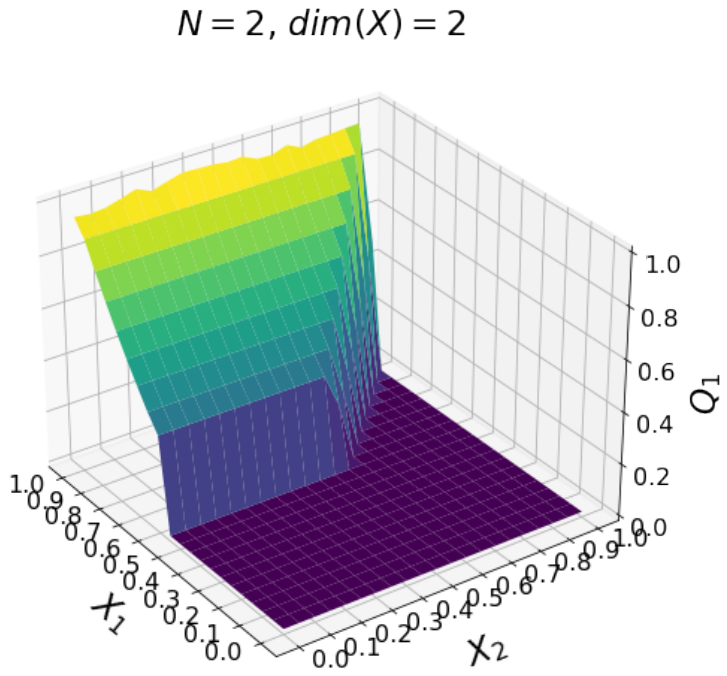

FIGURE 1: The interim allocation Q_1 for the first quality grade in the setting of symmetric, independent uniform setting of (Pavlov 2011) when $X \sim U[0,1]^2$. Notice, for example, the allocation is monotonic in values of X_1 and that the reserve price $p \in [.5, .6)$.

3. Based on the description of the exclusive buyer mechanism in Section 3.3.1, check if it matches the interim allocations yielded by the approximation algorithm. In Figure (1), the interim allocation can approximated by the interim allocation for the exclusive buyer mechanism (above):

$$Q_1(x;p) = \mathbb{1}\{\beta_1 > \beta_2 \text{ and } \beta_1 \geq 0\} F^{N-1}(x_1, \min\{\overline{x}_2, r_2 + \beta_1\}) \tag{3.19}$$

   Examination of the interim allocations of mechanisms yielded by the approximation algorithms is the basis for hypotheses concerning the optimal mechanisms.

By considering a wide enough range of cases and leveraging results from existing research in multidimensional mechanism design, we can develop some economic intuitions about the qualitative features of optimal mechanisms in a comparatively simpler multidimensional setting without worrying about the problem of bundling in the setting of multiple bidders *and* multiple goods. Since this approach is broadly susceptible to problems of discretization when approximating the optimal mechanism as well as precision issues with numerical computing, conclusions are, at best, a promising guide to developing further theoretical results. Thus, results are best interpreted as conjectures concerning the character of optimal mechanisms in multidimensional settings.

3.4.2 Conjectures

The principal conjecture investigated in this thesis chapter concerns the optimality of the exclusive buyer mechanism in the multidimensional setting of a single good with multiple quality levels:

Conjecture 1 (Revenue)**.** *The revenue of the exclusive buyer mechanism well-approximates the revenue of the optimal mechanism.*

By measuring the discrepancy between revenue from the exclusive buyer mechanism and that returned by the approximation algorithm it is possible to confirm or reject this conjecture.

It is also important to explore the interim allocations yielded by the approximation algorithm and compare them to those of the exclusive buyer mechanism. There should be visual confirmation that the optimal mechanism yielded by the algorithm is qualitatively similar to the exclusive buyer mechanism:

Conjecture 2 (Allocations)**.** *The allocation of the exclusive buyer mechanism well-approximates the allocation of the optimal mechanism yielded by the approximation algorithm.*

Note, if conjecture 2 holds, it automatically implies 1, since the allocations include the reserve price from which the revenue is calculated. However, since it is possible that the exclusive buyer mechanism approximates the revenue of the optimal mechanism but fails to share qualitative features of its allocation, it is helpful to distinguish between both conjectures.

Additionally, a surprising feature of some optimal mechanisms in the setting of a single good with multiple quality levels noted by several economists is that the set of types excluded by the allocation—the *exclusion region*—in equilibrium sometimes has measure zero (e.g., Thirumulanathan, Sundaresan, and Narahari 2019b; Belloni, Lopomo, and Wang 2010). This surprising finding stands in opposition to the result of (Armstrong 1996), where it was shown that in the case of a multiproduct monopolist that it is always optimal to exclude a positive measure of buyers. Thus, we explore under what circumstances the exclusion region is measure zero. Specifically, we conjecture:

Conjecture 3 (Measure Zero Exclusion Region)**.** *There exist multidimensional settings where a single good with multiple quality levels is sold to multiple bidders with a measure zero exclusion region.*

Finally, we conjecture that the exclusion region does not change with the number of bidders.

Conjecture 4 (Same Exclusion Region for all N)**.** *The exclusion region of the optimal mechanism in the multidimensional setting of a single good with multiple quality levels remains the same for $N = 1, 2, 3, \ldots$ bidders.*

3.4.3 Simulations

In what follows we explore conjectures 1, 2, 4 in the following contexts:

1. The **symmetric, independent, and uniform setting**, where buyers' valuations are independent across quality grades, uniformly and symmetrically distributed. I analyze the case where two bidders have identical valuations for a good with two quality grades, where each valuation is assumed to be distributed $X_1, X_2 \sim U[0, 1]$. Additionally, I analyze the case where $X_1, X_2 \sim U[2, 3]$ since, in contrast to the previous setting, it is known that in these settings the optimal mechanism involves randomization when $N = 1$. Analytic results in the single buyer case when $X \sim U[c, c + 1]^2$ are known (Pavlov 2011) and serve as a benchmark.
2. The **symmetric, independent, and non-uniform setting**. Here, it is desirable to see if the conclusions reached in the first setting extend to non-uniform distributions. In particular, we consider the case of the $Beta(\alpha, \beta)$ distribution (where $\alpha = 1, \beta = 2$), which was explored in (Daskalakis, Deckelbaum, and Tzamos 2017) in the context of multiple-goods.
3. The **symmetric, correlated setting**. It is unknown how arbitrary correlations between a buyer's dimensions of value affect the revenue-maximization problem faced by the auction designer in multidimensional settings. This setting aims to shed light on this problem by considering buyers with valuations drawn from $X_1 = X_2 = [0, 1]$ where the distribution of valuations is $f(x_1, x_2) = x_1 + x_2$.
4. The **asymmetric, independent, and uniform setting**. No analytic results are known in this similar setting; however, a very provisional analysis of the optimality of the exclusive buyer mechanism in this setting can be found in (Belloni, Lopomo, and Wang 2010), where $X_1 \sim U[6, 8], X_2 \sim U[9, 11]$ with

costs $c_1 = .9, c_2 = 5$. I replicate their analyses and extend their results in the context of the three conjectures proposed above.

5. The **asymmetric, independent, and non-uniform setting**, a direct extension of the above setting where the valuations for each quality level of the good are drawn from two different truncated normal distributions. In particular, I consider the case where $X_1 \sim truncnorm(\mu = 2.3, \sigma = 1, \underline{x}_1 = 2, \overline{x}_1 = 3)$ and $X_2 \sim truncnorm(\mu = 2.8, \sigma = .2, \underline{x}_2 = 2, \overline{x}_2 = 3)$.

Symmetric, independent, and uniform

Analytic results in the case of a single buyer exist in the symmetric, independent, and uniform setting considered here. (Pavlov 2011) studied the case of two substitute goods[6] independently and uniformly distributed on $U[c, c+1]^2$. In the specific case of a single buyer with valuations distributed according to $X \sim U[0,1]^2$ with zero costs, it is known that the optimal mechanism is deterministic and involves setting reserve price $p^* = \frac{1}{\sqrt{3}}$ for both goods (since the valuations are symmetric). Thus, the optimal allocation is given in Figure 2 and the auctioneer's revenue is simply $p^*(1 - p^{*2}) = 0.3849....$

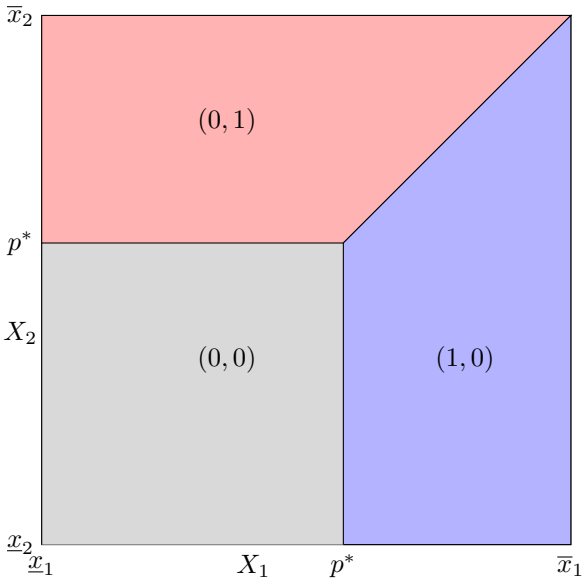

FIGURE 2: The optimal allocation of a single good with two quality symmetric levels to a single buyer with valuations $X \sim U[0,1]^2$ (Pavlov 2011). The area denoted $(0,0)$ is the 'exclusion region', where the good is not allocated. Note, in the setting of (Pavlov 2011), $p^* = \sqrt{1/3}$.

In the case where there is more than one bidder we need to rely on the approximation algorithm to study the qualitative features of the optimal mechanism. First, we can confirm the approximation algorithm yields similar revenue to that calculated by the appropriate exclusive buyer mechanism in this setting. These results are presented in Table 3.

Result Type	T	Revenue
approximation	5	0.66094...
approximation	10	0.625929...
approximation	15	0.612877...
approximation	20	0.606033...
exclusive buyer mechanism	50	0.589052...

TABLE 3: comparison of revenue generated by approximation algorithm with that of the exclusive buyer mechanism. T represents the number of intervals used to discretize each dimension of the buyers' valuations.

[6]Note that when there is a single buyer an equivalent interpretation of the setting with one good and multiple quality levels is that there are multiple goods but the buyer has unit demand.

Note that the revenue generated by the exclusive buyer mechanism was computed using the *ex-post* description of the auction. Additionally, the exclusive buyer mechanism's revenue was computed by numerical integration on a finder discretization grid as that used by the approximation algorithm for increased precision. The trend for different T implies that the approximation algorithm is converging to the result provided by the exclusive buyer mechanism. This supports Conjecture 1.

Using the *interim* allocation of the exclusive buyer mechanism described in section (3.3.1), we can plot the allocations against those returned by the approximation algorithm. These are presented side-by-side in Figure 4.

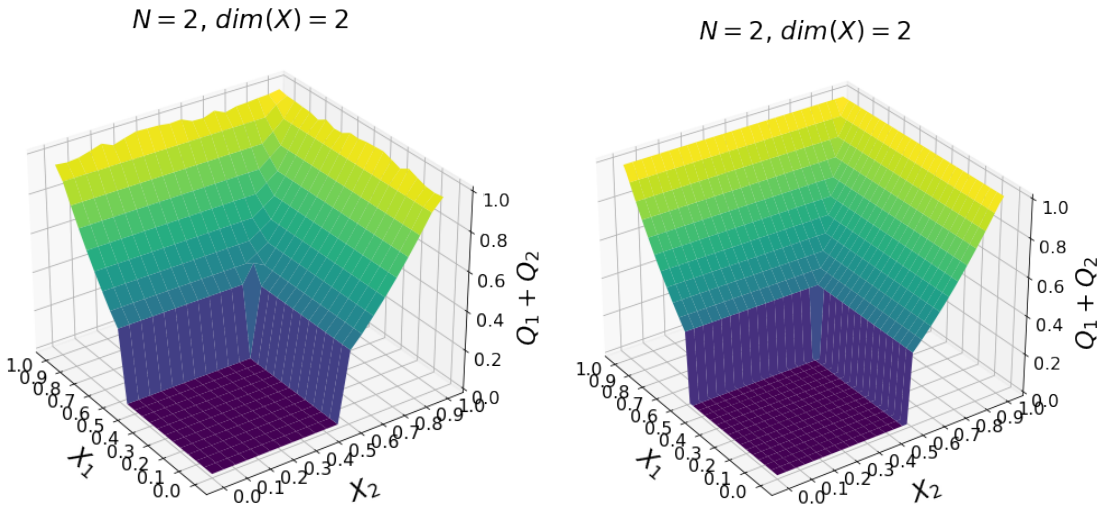

FIGURE 4: The allocations produced by the approximation algorithm (left) and exclusive buyer mechanism (right).

Conjecture 2 requires qualitatively evaluating the allocations returned from the approximation algorithm and those calculated from the exclusive buyer mechanism, without an obvious means to conclusively determine whether the conjecture is supported. However, the results in Figure 4 imply that exclusive buyer mechanism captures the behavior of the (approximately) optimal mechanism. Thus, these results offer support for Conjecture 2.

Notice that the exclusion region in Figure 4 is similar to that discovered in (Pavlov 2011). This is evidence against Conjecture 3 concerning the existence of measure zero exclusion regions. The price $p^* = \frac{1}{\sqrt{3}} = 0.577...$ that maximizes revenue in the case when $N = 1$ is consistent with the exclusion region defined by $p^* = 0.6$ returned by the approximation algorithm for the case of $N = 2$. (Note that when $[0, 1]$ is discretized into 20 intervals, the price is $p \in \{\dots, .5, .55, .6, \dots\}$ so the choice by the algorithm reflects its approximate optimality). Indeed, when $T = 20$, the approximation algorithm yields the same exclusion region for all of $N = 1, 2, 3$. This supports Conjecture 4.

Furthermore, it is helpful to investigate the behavior of the approximation algorithm when $X \sim U[2, 3]^2$, since we know from Pavlov (2011) that the optimal mechanism is stochastic. This is particularly important in the context of Conjecture 4 since it implies that the exclusive buyer mechanism will not be optimal in this case.

Again, first we confirm the approximation algorithm yields similar revenue to that calculated by the appropriate exclusive buyer mechanism in this setting. These results are presented in Table 5.

Result Type	T	Revenue
approximation	5	2.622409...
approximation	10	2.58287...
approximation	15	2.58287...
approximation	20	2.562314...
exclusive buyer mechanism	50	2.534499...

TABLE 5: comparison of revenue generated by approximation algorithm with that of the exclusive buyer

mechanism. Since the mechanism yielded by the approximation algorithm included randomization and, therefore, the menu of contract points included more than a single reserve price, the choice of reserve price for exclusive buyer mechanism was $r = 2.15$.

Here, the exclusive buyer mechanism yields a revenue similar to that of the optimal mechanism. The difference between the revenues is approximately 1%, despite the stochastic optimal mechanism. Furthermore, there is a clear indication that as T increases, the approximation algorithm converges to the revenue yielded by the exclusive buyer mechanism. This supports Conjecture 1.

We can plot the allocations of the optimal mechanism generated by the approximation algorithm with those of the exclusive buyer mechanism to assess our next conjecture. These are presented side-by-side in Figure 6.

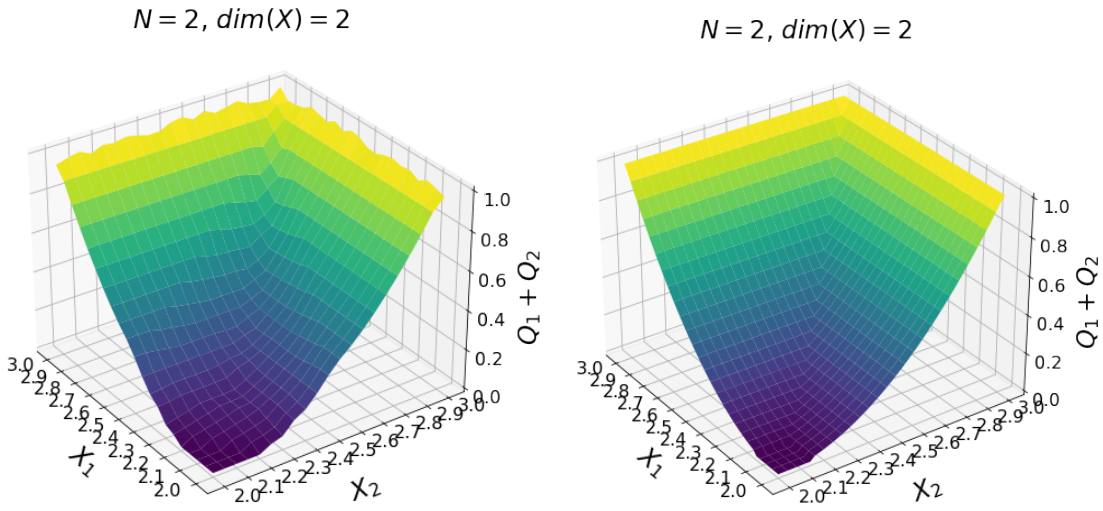

FIGURE 6: The allocations produced by the approximation algorithm (left) and exclusive buyer mechanism (right).

Although the shape of the allocations is similar in the region where $Q_1(x) + Q_2(x) > 0$, the exclusion region yielded by the optimal mechanism is triangular, indicating that the optimal mechanism includes randomization. In contrast, the exclusive buyer mechanism is a deterministic mechanism and has a rectangular exclusion region. Thus, this setting does not support Conjecture 2.

Again, the exclusion region in Figure 6 is similar to that discovered in (Pavlov 2011). This is notable because when $X \sim [c, c+1]^2$ there was randomization in the single bidder case. This is both evidence against Conjecture 3 concerning the existence of measure zero exclusion regions and evidence for Conjecture 4. Indeed, when $T = 20$, the approximation algorithm yields the same exclusion region for all of $N = 1, 2, 3$.

Symmetric, independent, and non-uniform setting

I examine the four conjectures in the context of a symmetric, independent, and non-uniform setting. Following the multi-unit example in (Daskalakis, Deckelbaum, and Tzamos 2017), I consider the case where $X_1, X_2 \sim Beta(\alpha, \beta)$ where $\alpha = 1, \beta = 2$. To the best of my knowledge, no prior work on analytic solutions to the optimal auction design problem exists in this setting. Therefore, we proceed by running the optimization algorithm and comparing the output of the algorithm to that provided by the exclusive buyer mechanism described above.

First, we compare the revenue generated by the approximation algorithm and the exclusive buyer mechanism. The results are presented in Table 7. Again, note the revenue from the exclusive buyer mechanism was calculated using *ex-post* allocations from a second-price auction. Although the revenue generated from the optimal mechanism is larger than the exclusive buyer mechanism (∼4%), the trend as T increases provides support for Conjecture 1.

Result Type	T	Revenue
approximation	5	0.448709...
approximation	10	0.418815...
approximation	15	0.406948...
approximation	20	0.400615...
exclusive buyer mechanism	50	0.385045...

TABLE 7: comparison of revenue generated by the approximation algorithm with that of the exclusive buyer mechanism when $X_1, X_2 \sim Beta(1, 2)$.

Next, we can compare the interim allocations from the approximation algorithm with those from the exclusive buyer mechanism. These are displayed graphically in Figure 8. There is a clear similarity between both allocations, supporting Conjecture 2.

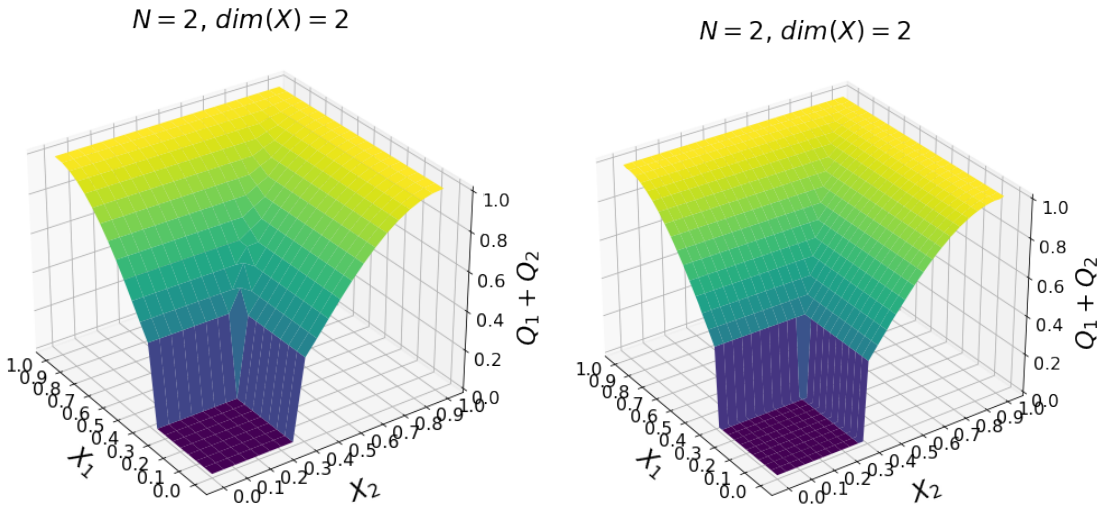

FIGURE 8: The allocations produced by the approximation algorithm (left) and exclusive buyer mechanism (right).

Note, these allocations offer a lack of support of Conjecture 3 concerning the existence of measure zero exclusion regions.

Finally, running the approximation algorithm for $N = 1, 2, 3$ confirms Conjecture 4. The exclusion region is the same for all N tested: the value of $p^* = 0.4$. (Recall, discretization of the grid into $T = 20$ intervals per quality level requires that $p^* \in \{\dots, 0.35, 0.4, 0.45, \dots\}$).

Thus, in conclusion, all conjectures are supported in the setting of symmetric, independent, and non-uniform distributions except for Conjecture 3 concerning the existence of a measure zero exclusion region.

Symmetric, correlated setting

In this setting, I allow for the valuations X_1 and X_2 to be correlated. Here, $X_1 = X_2 = [0, 1]$, where $X \sim F$ and $f(x_1, x_2) = x_1 + x_2$. This extension of the previous settings on the unit square facilitates a deeper understanding of correlated valuations in a familiar setting. As above, I proceed by running the approximation algorithm and comparing the results with those from the exclusive buyer mechanism.

With regard to revenue, we can see in Table 9 that the algorithm's revenue is, again, well-approximated by the exclusive buyer mechanism (∼3%). The trend as T increases is similar to the other settings: the revenue generated by the approximation algorithm converges to that yielded by the exclusive buyer mechanism. This supports Conjecture 1.

Result Type	T	Revenue
approximation	5	0.75159...
approximation	10	0.718683...
approximation	15	0.705905...
approximation	20	0.698962...
exclusive buyer mechanism	50	0.676192...

TABLE 9: comparison of revenue generated by the approximation algorithm with that of the exclusive buyer mechanism when $f(x_1, x_2) = x_1 + x_2$.

The allocations are also similar. The interim allocations from the approximation algorithm and the exclusive buyer mechanism are presented in Figure 10. This supports Conjecture 2 concerning the qualitative similarity of the optimal mechanism yielded by the approximation algorithm and that of the exclusive buyer mechanism.

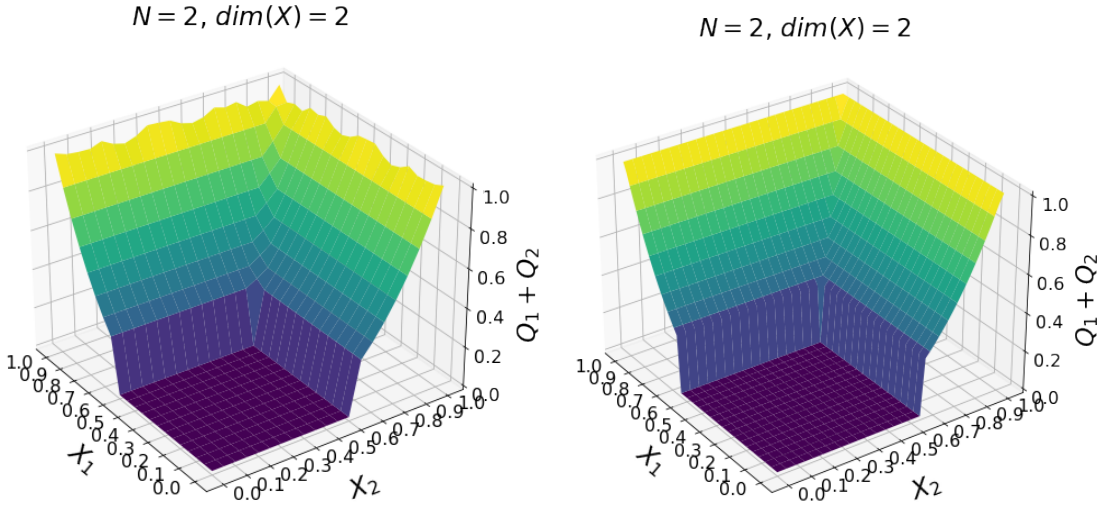

FIGURE 10: The allocations produced by the approximation algorithm (left) and the exclusive buyer mechanism (right).

Again, note the presence of an exclusion region undermines support for Conjecture 3 concerning the existence of settings without an exclusion region.

Finally, the exclusion regions in the symmetric, correlated, uniform setting where $X_1 = X_2 = [0, 1]$ and $f(x_1, x_2) = x_1 + x_2$ are the same for all $N = 1, 2, 3$. Note, the reserve price is $p^* = .65$ in this setting. This supports Conjecture 4.

As in the previous setting, all conjectures except for Conjecture 3 concerning the existence of a measure zero exclusion region are supported.

Asymmetric, independent, and uniform setting

In this setting, we consider the case where $X_1 \sim U[6, 8]$ and $X_2 \sim U[9, 11]$ as considered in (Belloni, Lopomo, and Wang 2010), who first studied the optimality of the exclusive buyer mechanism in multidimensional settings. Additionally, the costs associated with selling the first quality grade of the good are $c_1 = .9$ and the second quality grade are $c_2 = 5$. Since prior computational work exists assessing the optimality of the exclusive buyer mechanism, where possible, I can compare my findings here with those in (Belloni, Lopomo, and Wang 2010).

The revenue yielded by the approximation algorithm is similar to that of the exclusive buyer mechanism. The data are displayed in Table 11. These revenue numbers are consistent with those in (Belloni, Lopomo, and Wang 2010, Table 3), supporting Conjecture 1, namely, that the revenue of the optimal mechanism is well-approximated by the exclusive buyer mechanism.

Result Type	T	Revenue
approximation	5	6.02496...
approximation	10	5.941549...
approximation	15	5.91222...
approximation	20	5.893113...
exclusive buyer mechanism	50	5.805032...

TABLE 11: comparison of revenue generated by the approximation algorithm with that of the exclusive buyer mechanism in the setting of (Belloni, Lopomo, and Wang 2010).

With regard to the allocations, we can see a clear disparity between the interim allocations produced by the approximation algorithm and those of the exclusive buyer mechanism. The exclusive buyer mechanism's interim allocation for $Q_1(x) + Q_2(x)$ is much lower for large values of X_2 than that of the approximation algorithm. The allocations are presented in Figure 12.

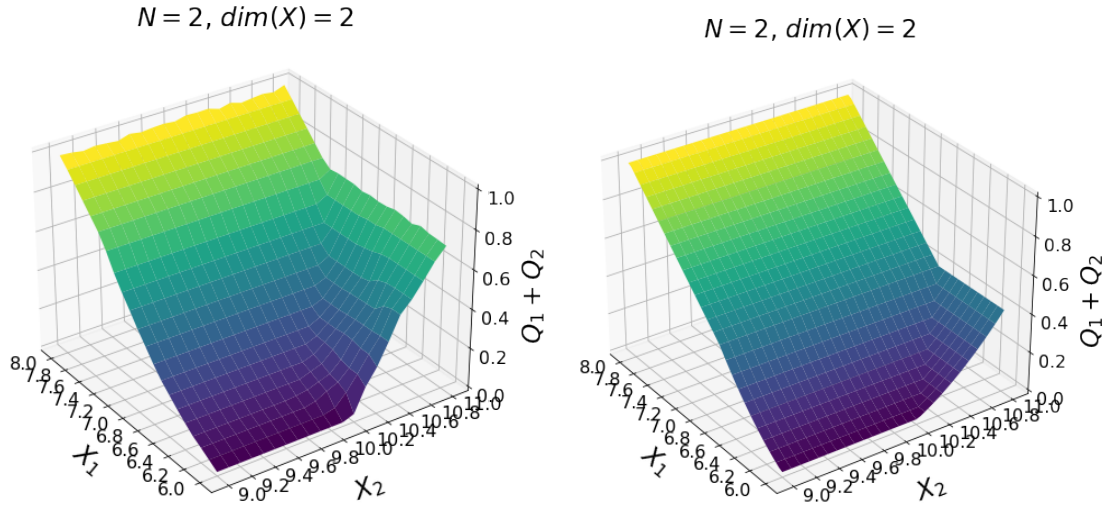

FIGURE 12: The allocations produced by the approximation algorithm (left) and the exclusive buyer mechanism (right).

This surprising feature of the exclusive buyer mechanism's interim allocation suggests that Conjecture 2 is *not* supported in the asymmetric, independent, and uniform setting. Here, although the optimal revenue is well approximated by the exclusive buyer mechanism, the qualitative features of the optimal mechanism are sufficiently different from those conjectured by the exclusive buyer mechanism. We will explore this result in more detail in the discussion section (3.5) below.

Notice, however, in this setting Conjecture 3 is supported. This is visible in Figure 13 below. This was also noted in the simulations (Belloni, Lopomo, and Wang 2010), suggesting it is robust to computational or approximation error. Additionally, Conjecture 4 is supported for all $N = 1, 2, 3$; however, it is important to note a number of surprising features of the optimal mechanism yielded by the approximation algorithm in this setting. Firstly, although the exclusion region is the same for all N, the *allocation* itself varies with the number of buyers. Secondly, there is evidence of randomization in the optimal mechanism yielded by the approximation algorithm.

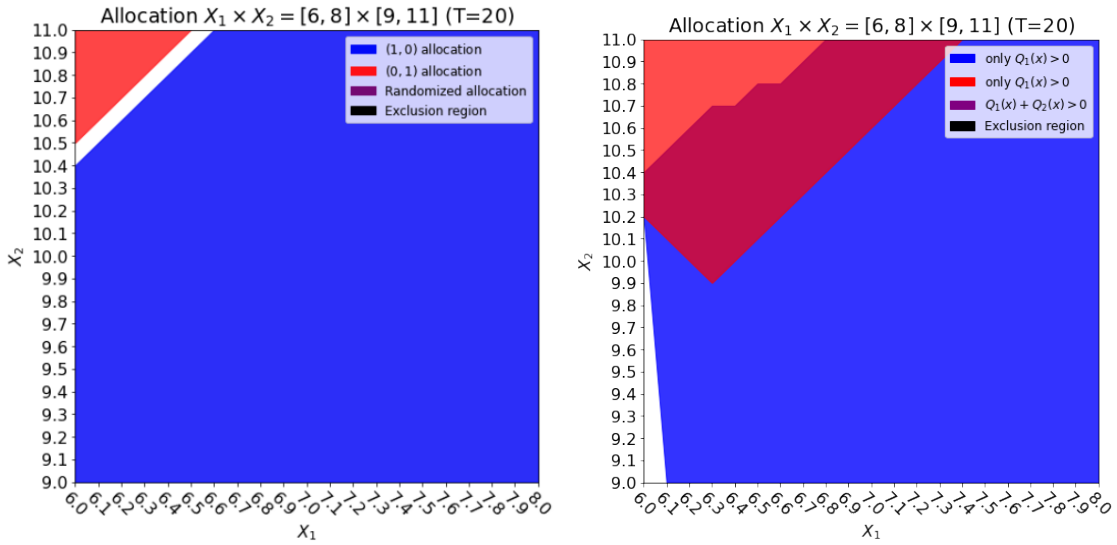

FIGURE 13: Graphs showing where the allocations for the first (Q_1) and second (Q_2) quality level of the good are non-zero for the optimal mechanism yielded by the approximation algorithm for $N = 1$ (left) and $N = 2$ (right).

Thus, in the symmetric correlated setting, although Conjectures 1, 3, 4 are supported, Conjecture 2 is not. The complexities of this result will be further explored below.

Asymmetric, independent, and non-uniform setting

I extend the investigation of asymmetric settings by considering the cases where buyers' valuations are distributed according to two different distributions. In particular, I consider the case where the valuations are drawn from $X_1 = X_2 = [2, 3]$ and the distribution of valuations is asymmetric, where $X_1 \sim truncnorm(\mu = 2.3, \sigma = 1)$ and $X_2 \sim truncnorm(\mu = 2.8, \sigma = .2)$. Note, in contrast to the previous asymmetric setting considered above, in this setting the sets from which the valuations are drawn are equal (i.e., $X_1 = X_2$) but the distributions are not (i.e., $f_1(x) \neq f_2(x)$). Again, no prior analytic results exist in this setting and therefore I proceed by running the approximation algorithm for the optimal auction and comparing the result to that of the exclusive buyer mechanism.

First, I compare the revenue generated by the approximation algorithm and the exclusive buyer mechanism. This is presented in Table 14. The similarity of the revenues generated by the approximation algorithm and the exclusive buyer mechanism lends support to Conjecture 1.

Result Type	T	Revenue
approximation	5	2.779996...
approximation	10	2.749451...
approximation	15	2.736171...
approximation	20	2.729342...
exclusive buyer mechanism	50	2.57921...

TABLE 14: comparison of revenue generated by the approximation algorithm with that of the exclusive buyer mechanism when $X_1 \sim truncnorm(\mu = 2.3, \sigma = 1, \underline{x}_1 = 2, \overline{x}_1 = 3)$ and $X_2 \sim truncnorm(\mu = 2.8, \sigma = .2, \underline{x}_2 = 2, \overline{x}_2 = 3)$.

When we compare the interim allocations generated by the approximation algorithm to those of the exclusive buyer mechanism in Figure 15, we can qualitatively see that the allocations are similar. This supports Conjecture 2.

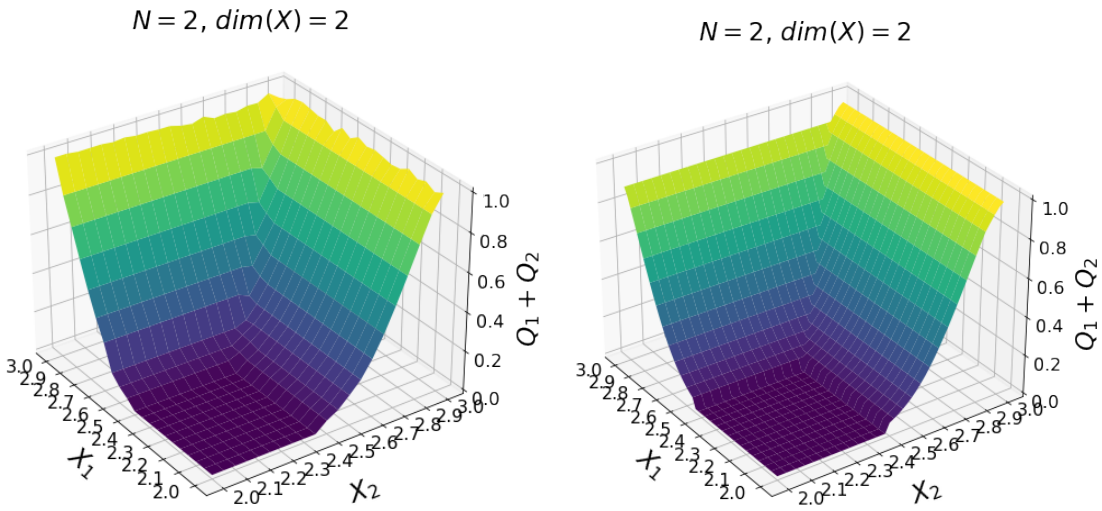

FIGURE 15: The allocations produced by the approximation algorithm (left) and exclusive buyer mechanism (right).

Additionally, note that the existence of an exclusion region in Figure 15 above undermines Conjecture 3.

Finally, running the approximation algorithm for all $N = 1, 2, 3$ confirms Conjecture 4. In this asymmetric setting $p_1^* = 2.55$ and $p_2^* = 2.5$ and the exclusion region remains the same for all N.

Thus, in conclusion, all conjectures except for Conjecture 3 concerning the existence of a measure zero exclusion region are supported in the setting of asymmetric, independent, and non-uniform distributions.

3.5 Discussion & Conclusion

Analytic results concerning the qualitative characteristics of optimal auctions in multidimensional settings are scarce. In this chapter, I explored the particular multidimensional setting of a single good with multiple quality levels. In particular, I investigated whether the exclusive buyer mechanism is optimal in this setting. To do this, I developed an approximation algorithm that facilitates the investigation of the optimal mechanism in multidimensional settings and compared the performance of the exclusive buyer mechanism to that of the optimal mechanism yielded by the approximation algorithm across a number of questions.

I explored four conjectures concerning the optimality of the exclusive buyer mechanism in the multidimensional setting of a single good with multiple quality levels. The rest of this section contains a discussion of how the results presented in Section 3.4.3 support or undermine these conjectures.

3.5.1 Conjecture 1 (Revenue)

This conjecture asserted that "the revenue of the exclusive buyer mechanism well-approximates the revenue of the optimal mechanism". Across all settings considered above, this conjecture is supported. However, support for this conjecture alone is far from sufficient to demonstrate the optimality of the exclusive buyer mechanism. It is well known that simple (deterministic) mechanisms can approximate optimal (stochastic) mechanisms up to some constant fraction of their revenue. This approximation can often be very close to the revenue yielded by the optimal mechanism. In the case where a single bidder's valuations are uniformly distributed on the unit square $X \sim [c, c+1]^2$, the gain from using a fully optimal mechanism over the best deterministic mechanism is at most 1.2% (Pavlov 2011, p11). Though in some settings considered here with $N = 2$ bidders, the gain is closer to 3-4%, the trend as T increases suggests that the final gain from using optimal mechanism over the exclusive buyer mechanism is closer to ∼1-2% in instances where randomization is required for optimality. Furthermore, the results are also consistent with previous simulation studies on the optimality of the exclusive buyer mechanism indicating it well-approximates the revenue generated by the optimal mechanism (Belloni, Lopomo, and Wang 2010). Thus, the results support Conjecture 1.

3.5.2 Conjecture 2 (Allocations)

Across all settings considered here, the interim allocations of the optimal mechanism yielded by the approximation algorithm share qualitative features with the exclusive buyer mechanism's allocations. However, there are multiple cases where the allocations are noticeably different. Thus, there is mixed support for Conjecture 2, which asserts that "the allocation of the exclusive buyer mechanism well-approximates the allocation of the optimal mechanism yielded by the approximation algorithm." I consider each of these cases in turn.

There is evidence of support for Conjecture 2 in the following settings:

- Symmetric, independent, and uniform setting ($X \sim U[0,1]^2$)
- Symmetric, independent, and non-uniform setting ($X \sim Beta(1,2)^2$)
- Symmetric, correlated, and uniform setting ($X \sim F, f(x_1, x_2) = x_1 + x_2$)
- Asymmetric, independent, and non-uniform setting ($X_1 \sim truncnorm(\mu = 2.3, \sigma = 1, \underline{x}_1 = 2, \overline{x}_1 = 3)$, $X_2 \sim truncnorm(\mu = 2.8, \sigma = .2, \underline{x}_2 = 2, \overline{x}_2 = 3)$)

In each of these settings, the reserve prices were consistent and the allocations were qualitatively similar in the region where $Q_1(x) + Q_2(x) > 0$. It is noteworthy that in each of these settings, the optimal mechanism is deterministic.

There is evidence against Conjecture 2 in the symmetric, independent, and uniform setting ($X \sim U[2,3]^2$) and the asymmetric, independent, and uniform setting first investigated by Belloni, Lopomo, and Wang (2010). In both these settings there is evidence of randomization in the optimal mechanism, which suggests the deterministic exclusive buyer mechanism is unable to approximate the optimal allocations. In the former case when $X \sim U[2,3]^2$ the exclusion region is not rectangular. In the latter case, the exclusion region has measure zero and, on closer inspection, evidence of randomization can be seen in Figure 16 where the allocations for each quality grade are shown separately. As can be seen in the graph of Q_1 on the left, a "fold" in the allocation occurs around the line $x_1 - c_1 = x_2 - c_2$ (recall, $c_1 = .9, c_2 = 5$). In this region, both $Q_1(x) > 0$ and $Q_2(x) > 0$.

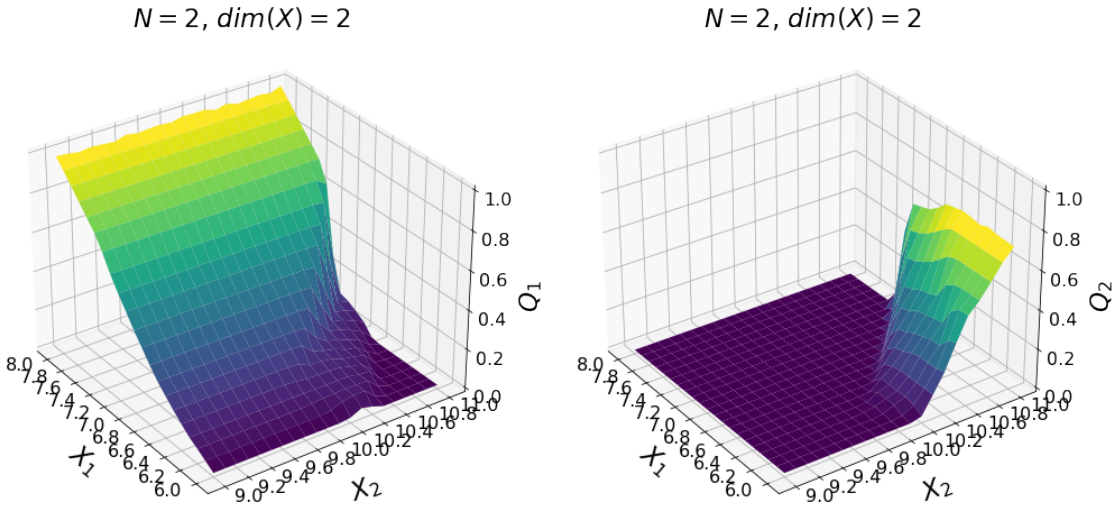

FIGURE 16: The allocation Q_1 for the first quality level (left) and Q_2 for the second quality level (right) of the good produced by the approximation algorithm when $N = 2$.

In conclusion, there is mixed support for Conjecture 2. When the optimal mechanism exhibits randomization, the allocations of the deterministic exclusive buyer mechanism do not well-approximate optimal allocations. However, it is possible to construct an *stochastic exclusive buyer mechanism* which includes lotteries. The idea is simple: again, each bidder bids to be the only bidder to choose between quality grades of the object. They can then purchase one of any number of lotteries determined by the seller. This mechanism was not considered in this chapter but future work should address whether it might be optimal in the

multidimensional setting of a single good with multiple quality grades.

3.5.3 Conjecture 3 (Measure Zero Exclusion Region)

Conjecture 3 concerns the existence of "multidimensional settings where a single good with multiple quality levels is sold to multiple bidders with a measure zero exclusion region". Only the asymmetric, independent, and uniform setting initially investigated by Belloni, Lopomo, and Wang (2010) provides evidence for this conjecture. This finding was also discussed in the work of Thirumulanathan, Sundaresan, and Narahar (2019a), who studied the multidimensional setting of a single good with multiple quality levels when bidder valuations are uniformly distributed on an arbitrary rectangle $[a, b] \times [c, d]$.

This finding stands in contrast to theoretical work in multidimensional mechanism design where it has been shown in the multi-unit case that it is always optimal to exclude a positive measure of bidder types (Armstrong 1996; Rochet and Choné 1998). Notice that in the example considered above the typespace $[6, 8] \times [9, 11]$ violates the assumption of strict convexity. This assumption is essential to establish Armstrong's (1996) proof. In the case of Rochet and Choné (1998), they require the cost function to be smooth, which is clearly violated in the setting above.

Figure 12 suggests that for the mechanism yielded by the approximation algorithm, there is an interval $6 \times [9, p^*]$ for some $p^* \in [9, 11]$ that receives zero utility in equilibrium. The measure of this interval is zero. However, contrary to (Armstrong 1996, Proposition 1), if one increases the price by ϵ, one would exclude a mass of types proportional to ϵ and not ϵ^2. This intuition suggests why optimal mechanisms might be characterized by measure zero exclusion regions in asymmetric settings.

3.5.4 Conjecture 4 (Same Exclusion Region for all N)

Conjecture 4 is supported in all settings considered here for $N = 1, 2, 3$. Computational intractability precludes the study of settings with more bidders but these initial results suggest a promising line of research where, for any given multidimensional setting of a single good with multiple quality levels and bidders with identical distributions of valuations, finding the set of values excluded from the mechanism in equilibrium when $N = 1$ can be used as a 'stepping-stone' to begin research for $N > 1$[7].

3.5.5 Conclusion

There is strong support for Conjectures 1 and 4 in all settings explored here. Thus, the exclusive buyer mechanism well-approximates the revenue of the optimal mechanism in a broad range of multidimensional settings where a single good with multiple quality levels is sold to multiple buyers. Additionally, the results indicate that, in this multidimensional context, the measure of types excluded from the mechanism in equilibrium is the same for any number of buyers. While there is strong evidence for Conjecture 2 concerning the similarity of the allocations in some settings, there is also evidence that the exclusive buyer mechanism fails to capture the qualitative behavior of the optimal mechanism when the optimal mechanism involves randomization, as in the asymmetric, independent, and uniform case explore in Belloni, Lopomo, and Wang (2010). Finally, only in the prior asymmetric, independent, and uniform settings is there support for Conjecture 3 concerning the absence of an exclusion region.

Further research into the possibility of a stochastic exclusive buyer mechanism should be investigated. Very few mechanisms have been shown to be generally optimal in the multidimensional context where a single good with multiple quality levels is sold to multiple buyers. Notably, research by Haghpanah and Hartline (2014, Theorem 9) indicates that a *favorite-outcome projection mechanism* is optimal, albeit in a restricted class of settings circumscribed by several strong assumptions. This chapter indicates that exclusive buyer mechanism might not only be optimal when randomization is not required for revenue-maximization but a further extension that includes stochastic contracts might be optimal more generally when randomization is required.

[7]Indeed, this approach motivated the decision to study how the exclusive buyer mechanism performs in the symmetric, independent, and uniform when $N = 2$ and $X \sim U[2, 3]^2$. Previous results (Pavlov 2011) show that in the single-bidder setting, randomization is required for optimality.

Finally, Conjecture 4 is especially important for guiding future research multidimensional auction design. It is a well-known problem that, despite significant advances in the mathematics of mechanism design which make use of optimal transport (see, for example, Ekeland 2010), proposing candidates for optimal auctions remains a major open problem. This is especially salient given the proliferation of duality results which suggest a 'guess-and-verify' approach (Daskalakis, Deckelbaum, and Tzamos 2017; Kolesnikov et al. 2022). The conjecture that optimal auctions in the single-bidder setting might share characteristics with those in multi-bidder settings can help guide future research in multidimensional auction design in the case of a single good with multiple quality levels.

Chapter 4

Reflexive Measurement

4.1 Introduction

Observer effects in the social sciences go by many names. It is not uncommon to speak of the 'Hawthorne effect' (Landsberger 1958) or an 'experimenter effect' (Rosenthal 1966) when the scientist or their science has a causal effect on its target of study. The idea that a measurement can causally affect the phenomenon it investigates is so widespread it is even enshrined in the common adage known as 'Goodhart's Law'[1]. Observer-type effects turn out to be ubiquitous across the social sciences, recognized and explored by economists and psychologists alike (Friedman 1953; Gergen 1973). Yet unified philosophical accounts of these diverse effects are rare despite their similar structure and widespread occurrence. For all that social scientists have labored long and hard to mitigate these problems in the course of their research, philosophers of science have neglected to put observer effects into broader philosophical perspective in a manner that can aid the practice of science.

In philosophy of science, the related idea of a 'self-fulfilling prophecy' is an ancient one, going back as far as the story of Oedipus, a mythical ancient Greek king whose best efforts to thwart an oracle's prophecy led to its tragic fulfillment. In its more modern guise philosophers of science have called the idea *reflexive prediction* and its counterpart notion in the social sciences is commonly traced back to sociologist Robert K. Merton's (1948) discussion of "self-fulfilling science". Like observer effects, this concept also captures the idea of the causal impact of science on what it studies. A canonical example of this phenomenon is a bank run: announcing an impending bank run may subsequently incite one. Contemporary philosophers of science have stressed the challenge reflexive prediction poses for theory development and testing in the social sciences (see Kopec 2011; Lowe 2018). Despite a number of highly distinguished accounts of the idea of reflexive prediction over the past century, the animating idea of the causal effect of science on what it studies and the challenge that it poses is rarely married to a discussion of measurement.

This essay offers a novel account of the concept of *reflexive measurement*. This account captures the salient features of diverse observer-type effects and recovers the intuition that a measurement is reflexive when agents are aware they are subjected to it. This is similar to the well-known idea of 'measurement as intervention' in the philosophy of economics (e.g., Morgan 2001). However, unlike existing versions of this account of measurement, a sensitivity to the different ways measurements causally affect their target of study necessitates revising the naive version of the measurement-as-intervention story. Measurements can sometimes fundamentally alter the phenomenon they investigate and other times only affect the data collected, leaving the underlying phenomenon unchanged. For example, in the context of survey research, it is common for respondents to lie about their preferences and opinions but the survey—the measurement instrument—does not causally affect the underlying phenomenon. This appreciation for the different ways measurements can causally affect the phenomenon they investigate sheds light on the deficiencies of the commonly used framework of de-biasing systematic measurement error for addressing the more fundamental problem of reflexive measurement.

Behind the account of reflexive measurement presented here is a reconceptualization of reflexivity in science writ large. Existing philosophical accounts of reflexive prediction would be of considerable help in

[1] 'When a measure becomes a target, it ceases to be a good measure' (Goodhart 1984).

better understanding the causal effects of measurements on their targets of study were it not for the fact that these insights are specifically tied to an understanding of prediction and rarely given in any form of generality. Thus, it is necessary to step back from specific scientific practices (e.g., prediction, measurement, theorizing) and ascertain how science—broadly understood as a sociological phenomenon—interacts with what it studies. A clear pattern then emerges. The causal effects of science occur only when agents are aware of the science that investigates their behavior. This, in a nutshell, is reflexivity.

The essay is structured as follows. In section 4.2, I review the major conceptual innovations on the topic of reflexive prediction by philosophers of science over the past century. Additionally, this review is supplemented by considerations of observer effects by experimental psychologists which are particularly germane to the topic of reflexive measurement. Section 4.3 collects these accounts and provides a general characterization of reflexivity in science. The central insight of this section is that agents' awareness of their position in a scientific study is the key causal pathway for reflexive effects. With a more general account of reflexivity in hand, the topic of measurement is then explored in section 4.4. Reflexive measurement is best captured by the measurement-as-intervention view, however, the naive understanding of this position requires modification in light of the distinction between data and phenomena. Section 4.6 concludes and offers directions for future work.

4.2 Literature Review

The idea that science causally affects its target of investigation has been long discussed in both science and philosophy. In philosophy of science alone, it is known as 'reflexivity', 'performativity', and sometimes 'reactivity'. However, this notion is mostly commonly discussed in the context of scientific theories and predictions. In this section, I provide a stylized overview of the conceptual development of this idea over the past century in order to subsequently develop an account of how measurements can casually affect what they measure. I draw from the well-developed concept of *reflexive prediction* in philosophy of science and supplement this understanding with developments in contemporary science which explicitly concern tackling the problem of the causal effect of science on what it studies. For a more even treatment of the development of reflexivity, albeit with less focus on psychology and contemporary economics, see the historical overview of (Mackinnon 2006).

Mid-twentieth century philosophical accounts of reflexivity following the sociologist Robert K. Merton's (1948) seminal account of 'self-fulfilling science' are isolated. Karl Popper (1953) briefly discussed a general formulation of the idea in the context of historicism in philosophy of social science and Ernest Nagel (1961) also noted the challenge this posed for theory construction in the social sciences. Subsequent accounts in the 1960s and 1970s focused more narrowly on the causal role predictions in the social sciences have on shaping their own truth-conditions (e.g., Buck 1963; Romanos 1973). These accounts focus heavily on the formulation and dissemination style (*FD-style*) of the prediction: whether the prediction was published in a newspaper or discussed on cable news, whether the prediction was public or private, etc. To different but ultimately similar degrees, these authors acknowledge that a single prediction will not, by itself, have a reflexive effect independent of how it comes to be known by those it makes predictions about.

These accounts from the 1960s and 1970s understood predictions in science as having a definite true/false truth value. A significant recent contribution by Kopec (2011) challenged this view, articulating a conception of probabilistic reflexive predictions. Here, a prediction is reflexive if it changes the probability of the event it predicts. The common use of applied statistics in developing predictions in the social sciences is better accommodated by this account. For example, social scientists have investigated whether public opinion polls indicating a favorite candidate in an upcoming election can increase that candidate's probability of winning (Rothschild and Malhotra 2014) and also whether the effects of election forecasts may depress voter turnout (Westwood, Messing, and Lelkes 2020). A narrow focus on the ultimate truth condition of the prediction misses the ways in which a prediction can nonetheless change individual behavior while leaving the result aggregate phenomena unchanged. Thus, even if a number of voters vote differently in light of public predictions, the result of an election may nonetheless remain unchanged.

A further criticism of existing philosophical work on reflexive prediction is that it fails to account of the idea that certain predictions may be more or less reflexive (Lowe 2018; Cejka 2022). The "Mertonian-derived, truth-centric notion of reflexive prediction" (Lowe 2018, p10) fails to capture the idea that, in some cases,

the effect of a reflexive prediction on (even the probability of) an event is "marginal at best" (Lowe 2018, p8). If an impending bank run is announced on the front page of the newspaper of record it has a vastly different effect than on the front page of a local student newspaper. The shift in focus to degrees of reflexivity represents a welcome change in our philosophical understanding of the many ways scientific predictions can interact with the social world.

Scientists have also developed their own accounts of reflexivity which are notably different from the philosophical views considered above. Contemporary work by economists on reflexivity has "situate[d] the concept in recent thinking on complex adaptive systems" (Beinhocker 2013, p331). This turn towards characterizing reflexivity in terms of systems-type thinking is associated with financier and investor George Soros, who has claimed that understanding the concept of reflexivity has enabled him to profit from his investments in financial markets (see, for example, Soros 2013). On this account, reflexivity is a property of systems. The systems-account emphasizes the interactions between agents and their environment, as well as explicitly conceptualizing agents' goals and cognitive abilities. An upshot of this account is that it arranges different systems along a 'spectrum of complexity' (Beinhocker 2013, p337) and enables the comparison of physical, human, and artificial systems in terms of reflexivity and complexity—an unusual feature of accounts of reflexivity.

Despite the common focus on reflexivity in economics by philosophers of science issues of reflexivity are found across many other sciences. It is helpful to also draw from the discipline of experimental psychology that has confronted reflexivity as a practical challenge. In doing so it becomes possible to develop a clearer overall picture of the causal effect of science on what it investigates. Phenomena like the 'experimenter effect' (Rosenthal 1966) and 'demand characteristics' (Orne 1962) have been well-known for decades and demonstrate the problems that come with either revealing information (e.g., a theoretical premise or expected result) to study participants during the course of research or participants 'guessing' the aims and objectives of the study and then adjusting their behavior accordingly. The implications of these findings constitute "a fundamental difference" (Gergen 1973, p313) between natural and social science.

Writing about the dangers of theories of psychology that are falsifiable "at will" by knowing study participants, mathematical psychologist R. Duncan Luce proposed the 'non-oxymoron criterion' (Luce 1995, p3) for theory-testing: scientists should be confident that their experimental design allows the theory to be tested despite the subject's knowledge of the theory. In other words, psychological hypotheses should not be able to be confirmed (disconfirmed) by the study participant at will. Psychologists have differed in their recommendations for how to avoid this. On the one hand, considering only "naive subjects" ensures that study participants are uninformed and therefore theories can be tested in "an uncontaminated way" (Gergen 1973, p313). On the other hand, we can directly address the self-interest of the subjects such that it "behooves the subjects to reveal their true preferences" (Luce 1995, p9). These views complement existing philosophical accounts by highlighting characteristics of study participants (e.g. how informed they are, their goals and desires) which contribute to the reflexivity of science.

In conclusion, it is important to bear in mind that in addition to philosophy of science and economics, disciplines as diverse as psychology (Luce 1995), political science (Rothschild and Malhotra 2014; Westwood, Messing, and Lelkes 2020), complex systems (Beinhocker 2013), and even theoretical computer science (Hardt et al. 2016; Perdomo et al. 2020) have all grappled with the issue that science causally effects its target of investigation in different forms. As I will detail below, observer effects are ubiquitous across scientific domains and only a broader understanding of these effects can do justice to the complexity of reflexivity in science. This literature review aims to surface some of the common concerns across these disciplines alongside the development of reflexivity as a key idea in contemporary philosophy of science. The next section will tie together these concerns into a general characterization of the concept of reflexivity.

4.3 Characterizing Reflexivity

Since almost all explicit definitions of reflexivity are inextricably tied to prediction in this section I propose a characterization of reflexivity which applies to all scientific practices. As such, it will necessarily be broader in scope and include more of scientific practice than is common on other accounts. The proposed account is closer in spirit to earlier attempts to understand reflexivity which attempted to grapple with the "complicated interaction between observer and observed" (Popper 1953, p14) at a high level of generality. Drawing from

the approach of Grunberg (1986), the account proposed here sheds light on the causal pathway by which reflexive effects manifest themselves. This serves to fix ideas for the discussion of measurement in subsequent sections. Additionally, this account extends to non-social scientific domains; a feature of reflexivity that has been widely under-appreciated by those who insist the concept uniquely applies to the study of humans and human behavior.

The motivating question for this more general account is: which scientific practices might be reflexive? All conceptions of reflexivity considered in the preceding section implicitly rely on a view of science that encompasses the social interaction between 'observer and observed' (Popper 1953) or 'scientist and study participant'(Gergen 1973; Luce 1995). A broad view requires us to consider science as a sociological phenomenon and allow our characterization of reflexivity to include facts concerning how the science in question interacts with its target of study. Who the scientist is or what institution they work at can have an outsized impact on the results of a scientific study. This is, effectively, a more general formulation of the 'formulation/dissemination-style' (*FD-style*) of reflexive predictions (Romanos 1973) which naturally extends to other scientific practices like measurement.

However, adopting broader sociological considerations is no small requirement. This means that irrespective of whether a specific scientific practice is known, the mere knowledge of the institution that carries it out can be sufficient to elicit a reflexive effect[2]. Consider that when Google dropped its "don't be evil" motto (Basu 2015) users may have felt the need to change their behavior when interacting with Google's products. Even without knowing the specific scientific practices Google was carrying out to investigate their users' behaviors, this might—in the broad sense of being a social interaction between observer and observed—constitute an instance of reflexivity for Google's study of its own users.

This further entails that the private/public distinction that animates so much of the reflexive prediction literature is no longer helpful (e.g., Buck 1963; Romanos 1973; Grunberg 1986). To see this consider the following example (adapted from Grunberg 1986):

Example 3 (Sumerian Economic Forecasts)**.** *The current Chairman of US Federal Reserve Jerome Powell delivers the Federal Reserve's annual economic forecasts on national television in ancient Sumerian with a presentation in cuneiform characters.*

Since, effectively, no one understands ancient Sumerian the forecast would be considered private. Setting aside the issue of market overreactions (e.g., Bondt and Thaler 1985), this would be an unprecedented action for a Chairman of the US Federal Reserve and may undermine investors' faith in the competence of major US financial institutions. This should constitute an instance of reflexivity in the same way that Google changing its motto should: the broader sociological context of a scientific practice can have enormous causal impacts on what it investigates.

The causal effect that science has on its target of study is clearly at the heart of all conceptions of reflexivity. On an overly simplistic view, reflexivity can even be understood as: the explicans causally affects the explicandum. Some kind of causal effect is clearly a necessary condition for the occurrence of reflexivity—on this, all philosophical accounts agree. But accounts of reflexivity differ in how they approach this. On one view, reflexivity is understood as a causal effect with a counterfactual component (Romanos 1973; Buck 1963). Another view emphasizes the causal effect of reflexivity as a stochastic phenomenon. A reflexive prediction, for example, changes the probability of an event occurring (Kopec 2011) and can even be ascertained by a test of statistical significance (Cejka 2022). A different kind of account gives definitions of reflexivity which omit causal language altogether in favor of clearly specifying the pathways along which the causal effects of reflexive science play out. Thus, for example, a reflexive prediction is "an utterance... made public in a language in terms that can be understood by the agents to whose behavior it refers and who therefore can by their actions either falsify or fulfill it" (Grunberg 1986, p476)[3]. A distinct advantage of this final approach is that it subsumes the causal effects of science on what it studies by specifying the mechanism by which agents might come to frustrate or fulfill a scientific prediction.

Before offering my own version of this type of account of reflexivity, it is important to be clear about the nature of "agents" that constitute part of the phenomenon investigated by scientists. In my view, the systems account of reflexivity (e.g., Beinhocker 2013; Soros 2013) correctly captures the important features of agency,

[2]There are even collateral effects from neighboring institutions or scientific practices. These are explored in the case of measurement in example 7.

[3]Alternatively, "in order to be reflexive it is sufficient for a public prediction to be partially believed" (Grunberg 1986, p484).

including agents' goals, cognitive capacities, and actions within the scope of a definition of reflexive system[4]. It is important to consider why this is particularly helpful. Firstly, note that reflexivity may characterize sciences that investigate collections of humans: organizations, governments, firms, etc., which act with a singular purpose. These can be modeled as agents. Secondly, it would be philosophically underwhelming to propose an account of reflexivity which rules out interesting cases like missiles (Grünbaum 1963) or thermostats (Beinhocker 2013) simply because the only agents to which the concept of reflexivity applies are human beings. It is desirable to simultaneously capture the intuition that there is something particularly philosophically interesting about the problems faced by social science but also that we should be open to discovering these problems in other scientific domains. Although there will certainly be disagreement over what constitutes agency, this ambiguity is a deliberate feature of the account presented here.

The account of reflexivity proposed here requires that the causal effect of observation or measurement or prediction—any form of scientific practice—on the target of inquiry be mediated through the awareness of the agents that constitute part of the phenomenon under investigation. Here, I am trying to generalize to all scientific practices the idea that the mechanism for a prediction to be reflexive is for it to be "partially believed" (Grunberg 1986, above). It is meaningless to speak of "belief" in the context of measurement. Some minimal degree of awareness of being observed is the relevant necessary condition for reflexivity. This requirement widens the scope of what is to be considered reflexive, as did the move to include the broad sociological context of scientific practice beyond, for example, individual public predictions. What ultimately matters for reflexivity is not how a given prediction, theory, or measurement was published or disseminated (i.e., its *FD-style*) but instead that the agents came to learn it.

The implications of this novel understanding can be fully seen in the following example.

Example 4 (Stoplight Example)**.** *A researcher aims to measure traffic patterns at an intersection. They stand on the side of the road noting the presence and absence of cars waiting at a stoplight. Inadvertently, however, they keep stepping on the cable that powers the stoplight, affecting the frequency with which the stoplight changes color.*

In this example, although the scientist causally intervenes in the target system they are seeking to study, the effect of this causal intervention is not a function of the agents' awareness of the science or scientist that studies their behavior. Thus, if the drivers of the cars (i.e., the agents) are unaware of the science that investigates their behavior, then the scientific study is not reflexive. This extends to the institution that the scientist is working for: if the agents are unaware not only of the scientist at the stoplight but of the broader sociological context in which they conduct their science, only then is science truly non-reflexive. As noted above, large corporations like Google and intelligence agencies like the CIA, which are often jokingly considered omniscient, will elicit reflexive effects even if the specific scientific investigations they carry out are unknown to those they study.

It is tempting to argue that awareness is also sufficient for reflexivity, however, I believe this kind of precise definition leads to the consideration of unhelpful counterexamples.

Example 5 (Alien Social Science)**.** *Assume only one species of aliens exists and consider that their alien social science which investigates human behavior on earth is entirely undetectable by us (i.e., has no causal impact we can discern). Despite this, some members of the public believe that aliens are real. Perhaps they have they have filled their imaginations with stories of Area 51 or watched too many X-Files episodes. Thus, they adapt and change their behavior in ways they think might frustrate alien social science.*

Since the account here argues in favor of considering collateral effects of institutions on earth (e.g., the FBI, Hollywood, etc.) a key part of reflexivity, then the alien social science is reflexive despite the lack of *any* causal effect on the phenomena it investigates.

This surprising conclusion is a feature of the inclusion of collateral reflexive effects (from, say, neighboring scientific institutions with bad reputations as covered in example 7 below) in the proposed characterization of reflexivity. When coupled with awareness as the appropriate causal pathway for mediating reflexivity renders the range of cases to which the designation of reflexivity applies very broad. This is partly by design:

[4]However, I do not believe the most promising path towards characterizing reflexivity is to "situate the concept in recent thinking on complex adaptive systems" (Beinhocker 2013, p331). Although there are undoubtedly good reasons to think about reflexivity in this manner for large-scale, complex phenomena like financial markets, the 'systems' approach is ill-suited to capture small-scale scientific investigations like laboratory studies (Luce 1995) and individual medical diagnoses (Hacking 1995). Especially since some systems accounts of reflexivity (e.g., Beinhocker 2013, p332) require that all reflexive systems be complex systems.

the goal of this account is to extend existing ideas and intuitions about reflexivity to (potentially) cover all scientific practices. Even Ian Hacking's (1995) seminal account of the causal effects of scientific theories themselves cannot be included in a discussion of reflexive predictions without significantly changing the scope of the argument (and all the relevant definitions of reflexive prediction). Treating awareness as a sufficient condition for reflexive entails that a science can be reflexive without causally affecting its target of study, a conclusion completely at odds with the animating idea of the characterization given here.

The characterization of reflexivity proposed here entails that almost all social scientific practice is reflexive. This feature of my characterization of reflexivity might strike the reader as unwelcome. Yet narrowly defining reflexivity in terms of "causal factors" (Romanos 1973; Buck 1963) or "changes in probability" (Kopec 2011) to pick out particular instances of reflexivity lands philosophers of science in the awkward position of assuming the role of scientists: determining what is and isn't reflexive in virtue of measurable effects. Moreover, recent developments in how to think about reflexivity emphasizing that reflexivity is a matter of degree (Lowe 2018) lend support to the idea that even minimal reflexive effects are still worthy of inclusion in the definition of reflexivity.

Ultimately, if reflexivity is defined by its effects it lands us with an arbitrary delineation of the term. Consider that a definition using the language of "causal factors" and "changes in probability" entails that the same prediction uttered in two almost identical circumstances might nonetheless result in one being reflexive while the other is not. These differing circumstances could be different days of the week, neighboring geographic regions, or even just differ in as much as a single study participant. Even more problematic is the fact that the absence of a discernible reflexive effect does not indicate the absence of reflexivity. A public prediction might result in exactly the same pattern of behavior (or probability of its occurrence) but the motivations for carrying out the behavior may have completely changed as a result of the prediction. A focus on the causal pathway by which reflexivity manifests allows philosophers of science to sidestep issues with ascertaining whether there is an appropriately reflexive causal effect for a given scientific practice.

Thankfully, social scientists are increasingly aware of the reflexive effects of their science. Contrary to earlier philosophical accounts of reflexivity which could only find a "a great deal of anecdotal evidence" (Grunberg 1986, p487) for the existence of reflexive predictions, a serious effort has been made to investigate the effects of public predictions in areas like election forecasting. Well-known election forecasts in the United States like Nate Silver's 538 website[5] which get national press coverage are now being investigated for their effects in depressing voter turnout (Westwood, Messing, and Lelkes 2020). Additionally, opinion polls indicating a favorite candidate in an upcoming election can increase the probability of that candidate winning (Rothschild and Malhotra 2014). The advantage of taking seriously the recommendation to treat reflexive effects as varying by degree (Lowe 2018) is that it leaves open the possibility of acknowledging that almost all social science is reflexive, though much of it might have barely any effect at all. Philosophers can offer a clear account of reflexivity and let scientists determine where it is appropriate to worry about it.

In summary: I provided a sociological picture of scientific practice, whereby a reflexive scientific practice can causally affect its target of inquiry when the cause is mediated through the agents' awareness. Although this condition is necessary for reflexivity, and, indeed, it is often sufficient, a definition should be avoided: it adds little to our scientific and philosophical understanding of a wide-ranging, complex phenomenon and only serves to distract us with far-fetched counterexamples. Furthermore, it is important not to define reflexivity by its effects. Today's election forecasts are reflexive, as are tomorrow's—irrespective of whether one elicits a causal effect and the other does not. What matters is whether the agents that comprise the phenomenon under investigation are aware of the prediction (and so it goes for measurement, etc.). Philosophers of science should let scientists determine the effects of reflexivity; our role is to clarify the phenomenon of reflexivity as one that does or does not apply to various scientific practices and domains. However, before instantiating this account in the novel context of measurement, it is important to consider the implications of this view for what kinds of science are reflexive.

4.4 Reflexive Measurement

Despite the enormous amounts of ink spilled by scientists lamenting the challenges of collecting accurate data from study participants in laboratories and surveys, this aspect of scientific practice has been mostly over-

[5] `https://fivethirtyeight.com`

looked by philosophers of science working on reflexivity. In this section, I instantiate the concept of reflexivity in the context of scientific measurement. Note, however, no new philosophical account of measurement is given in this section. Instead, the discussion of measurement sits closer to how a scientist encounters it. The focus of this section is on data collection since no measurement is possible without it. The heuristics to which scientists avail themselves to understand observer effects are exactly the level at which this account is pitched: it is an attempt to unify these solutions under a single philosophical perspective.

It is worth beginning with what is known about observer effects by scientists working across different fields. In its most general formulation across the social sciences, this is known as the 'Hawthorne effect' (see Landsberger 1958) or 'experimenter effect'[6] (Rosenthal 1966), where humans react to being observed and change their behavior in light of this observation. This general effect has been given a myriad of more specific formulations in different circumstances. To name a few of the most common found in scientific experiments: 'demand characteristics' are a phenomenon where study participants in an experiment act in ways they think the scientist desires (Orne 1962); the 'Pygmalion effect' is a psychological phenomenon whereby high expectations lead to improved performance (Rosenthal and Jacobson 1968); the 'John Henry effect' concerns the actions that study participants take on learning they are placed in a control group (as opposed to a treatment group) to overcome the disadvantage of being an experimental control (Colman 2008, p399). Outside of experiments, observer effects are commonly found in applied survey research: 'priming' occurs when a survey asks leading questions which can skew survey responses (Stantcheva 2022, §6.2); additionally, 'social desirability bias' is the phenomenon of surveys respondent lying or not sharing sensitive opinions (Krumpal 2013). Even 'Goodhart's Law' has its origins in the challenges faced by economists measuring economic indicators to set monetary policy (Goodhart 1984).

These disparate concerns all have the same root: the causal effects of scientists and their science on the target of study. Crucially, the kind of effects cataloged above are all mediated through the awareness of the agents studied. Study participants in laboratory experiments and respondents taking surveys are all fully aware of their role in scientific studies. Indeed, this is required for ethics approval. Thus, reflexive measurement can best be understood as the idea of 'measurement as intervention', which has long been known in philosophy of economics. Writing about the role of measurement instruments in economics, Mary Morgan shrewdly writes:

> "The ways in which the economic body is investigated and data are collected, categorized, analyzed, reduced, and reassembled amount to a set of experimental interventions—not in the economic process itself, but rather in the information collected from that process." (Morgan 2001, p237)

However, Morgan contends that the interventions do not causally affect the "economic process itself" and instead affect the "information collected from that process". In the context of much of contemporary economics, this seems apt, however, more broadly this account fails in other settings[7]. This insightful intuition about measurement instruments can be better appreciated by further considering the difference between 'data' and 'phenomena'.

A helpful philosophical account of the difference between data and phenomenon was developed by Jim Woodward (1989). Phenomena are "relatively stable and general features of the world which are potential objects of explanation and prediction by general theory", whereas data "by contrast, play the role of evidence for claims about phenomena" (Woodward 1989, p393-4). What matters in any scientific description or analysis of a phenomenon is that "the data should be *reliable evidence* for the phenomena in question" (Woodward 1989, p398, emphasis original). Furthermore,

> "Scientific investigation is typically carried on in a noisy environment; an environment in which the data we confront reflect the operation of many different causal factors, a number of which are due to the local, idiosyncratic features of the instruments we employ (including our senses) or the particular background situation in which we find ourselves." (Woodward 1989, p398)

In the context of the account of reflexive measurement introduced above (i.e., 'measurement as intervention'), data reflect the operation of measurement instruments and the broader sociological context in which scientists administer their measurement. Crucially, however, the causal effect of the measurement on what is being

[6]Though experimenter effects occur when study participants are not explicitly aware of them (e.g., through implicit cues that are registered subconsciously) the focus of this section is the reflexive effects of measurement. As per the characterization in the preceding section, these are only the effects that participants are aware of.

[7]See Example 6. I explore this in more detail below.

measured might affect the underlying phenomena itself and/or the data collected about it. To see this more clearly, I now consider two concrete examples.

To make vivid how an act of measurement may cause the phenomena under investigation to change, consider the following well-known experiment by psychologist Philip Zimbardo:

Example 6 (Stanford Prison Experiment)**.** *In 1971 a psychologist recruited participants for a "psychological study of prison life", which was a planned one-to-two week experiment that simulated prison life (see* Stanford Prison Experiment*, §2 for details). The goal was to assess the psychological effects of becoming a prisoner or prison guard.*

A full account of the Stanford Prison Experiment can be found in (Zimbardo 2008). It was prematurely ended after only 5 days since "prisoners were withdrawing and behaving in pathological ways, and... some of the guards were behaving sadistically" (*Stanford Prison Experiment*, §8). The ethics of this kind of experiment have been questioned: participants who simulated prisoners were deliberately made to feel humiliated (*Stanford Prison Experiment*, §3). Zimbardo's method of investigating the effects of simulated prisoner-guard has also been criticized as poor scientific practice (Texier 2019). Ultimately, it is abundantly clear that if the effects of the experiments are so pronounced as to induce a study participant to "[break] down and began to cry hysterically" (*Stanford Prison Experiment*, §8) then the measurement is reflexive in the sense of causally affecting the underlying phenomenon.

In contrast, a reflexive measurement can causally affect the data collected by a scientist without altering the underlying phenomenon under investigation. This is common in almost all forms of survey research where participants have the opportunity to lie or misrepresent their opinions. Here, a sensitive issue like the approval of a controversial political figure may be unaffected by a reflexive measurement but the data collected may be influenced by the study participants' reluctance to truthfully report their views. I believe this is the most promising way to realize Mary Morgan's insight that measurements of the economy are interventions "in the information collected from that process".

The next example provides a concrete case where the social context of the scientific practice can create exactly this kind of effect.

Example 7 (2020 US Census)**.** *In the run-up to the 2020 US Census, then-President of the United States Donald Trump made repeated remarks about the possibility of adding a citizenship question to the census (see Blake 2022). Subsequently, the Hispanic response rate was more than three times lower on the 2020 census than on the 2010 census*[8]*.*

This example highlights an important and often overlooked facet of scientific practice: *who* the scientist is or *what* institution they represent can have direct consequences on their ability to investigate phenomena in the face of reflexivity. Here, explicit condemnation of undocumented immigrants by former president Donald Trump likely had a causal role in more than tripling the number of those of Hispanic origin in the US who didn't complete the 2020 census compared to 2010. The underlying phenomena of interest investigated by the census (e.g., respondent's age, gender, income, etc.) don't change but the data collected are influenced by a powerful leader with the ability to use census data to create policies that leave undocumented immigrants worse off by deporting them.

We can be clearer about the particular phenomenon of reflexive measurement where a measurement casually affects either the data collected or the underlying phenomenon. The change in the distribution of the sample resulting from a reflexive measurement can be thought of as a kind of *distribution shift*[9]. The distribution shift of the sample represents its departure from the population-level data model. The measurement itself is an intervention that, for example, affects study participants' willingness to lie or conceal information in the face of a prying scientist. Notice this intervention only affects the sample since it is only the sample that is subjected to the measurement. Thus, this kind of reflexive measurement can be thought of as a kind of distribution shift where the sample distribution no longer represents the population distribution.

By way of contrast to this understanding of reflexive measurement, it is helpful to consider the commonly used approach of 'de-biasing' systematic measurement error. The systematic component of measurement error always occurs, with the same value, when the instrument is used in the same way in the same case (see Tal 2019). Thus, for example, we might say the systematic component of measurement error for a

[8]See Appendix 1 for calculation of this figure.

[9]This language is commonly used in computer science in the related context of *performative prediction* (Perdomo et al. 2020).

poorly worded survey question on political attitudes occurs when the responses are, for example, 'X% more left/right-leaning'.

Example 8 (De-biasing measurement error)**.** *Consider the case of a survey question asking about presidential approval in the US, which was answered by N respondents. The data $X_1, \ldots, X_N$ are assumed to come from a Gaussian distribution with unknown mean μ and variance σ. A scientist might then want to learn the value of μ. The statistical error associated with each random variable X_i is decomposed into a random and systematic component:*

$$\epsilon_i = \epsilon_i^{random} + \epsilon^{systematic} = \mu - X_i$$

Given further knowledge of the particulars of this domain of social inquiry, the scientist might impose additional assumptions about, for example, the shape of the distribution of the errors, or their covariance structure. These assumptions capture some of the flaws associated with a particular measurement instrument or measurement process. Ultimately, a scientist could make a post hoc correction for measurement error by subtracting off (often called 'de-biasing') the systematic component of the error:

$$X_i^{new} = X_i + \epsilon^{systematic}$$

Which will provide a more accurate (i.e., less biased) estimate of μ.

This example is paradigmatic of how measurement error is handled in the social sciences. The measurement instrument is biased and assumed to interact with those it's intended to measure in a uniform manner. The post hoc measurement error correction[10] can be read as, effectively, claiming the underlying data generating process measured by the instrument is actually a Gaussian distribution with mean $\mu - \epsilon^{\text{systematic}}$ and variance σ once the causal effect of the instrument on the data generating process is accounted for.

However, this kind of correction presumes the underlying sample distribution remains unchanged in the face of reflexivity except for a difference in means. In many settings, this may be a reasonable and accurate assumption. However, this framework of de-biasing error cannot account for changes in the higher moments (e.g., variance, skew, etc.) of the underlying distribution. Moreover, the type of statistical distribution itself might change, often considerably, as a result of the measurement. In example 7 above, some Hispanic subgroups who felt threatened by the increased condemnation of undocumented immigrants might be entirely missing from the resulting data. The problem of reflexive measurement is a deeper one than the framework of measurement error allows. This understanding of reflexivity and measurement error also applies more generally to de-biasing corrections in reflexive prediction (Cejka 2022). Sometimes these are appropriate responses to the problem of reflexivity in measurement, however, they are not a substitute for a general understanding of the problem of distribution shift induced by a measurement.

4.5 Distribution Shift and Mechanism Design

How best then to correct the sample distribution shift that results from a measurement? I think it is instructive to return to the observation by psychologist R. Duncan Luce who, when he advocated the 'non-oxymoron criterion' for theory-testing discussed above, noted that studies should be designed such that it "behooves the subjects to reveal their true preferences" (Luce 1995, p9). The account of reflexivity in the previous section focused on the causal pathway of agents' awareness, coupled with their goals, cognitive capabilities, and the actions available to them. Explicit concern with what agents want facilitates the possibility of designing measurements that induce a distribution shift such that accurate data are collected because it is in the study respondent's best interests.

Thus, we can think about whether measurements are, in a loose sense, incentive-compatible[11]. Incentive-compatible measurements induce minimal distribution shift such that the data collected are reliable evidence for the underlying phenomenon. This idea is related to that of *performative optimality* developed in (Perdomo

[10] Since $\epsilon^{\text{systematic}}$ is unknown a correction is only possible with an estimate of this quantity. Ascertaining whether or not this estimate is unbiased with respect to reflexive measurement effects is non-trivial. For a related discussion of this problem in the context of reflexive prediction see (Cejka 2022).

[11] This has a specific, technical meaning in the context of mechanism design. Here, I use it in the informal sense.

et al. 2020) to capture the distribution shift caused by performative predictions[12]. Two differences being: the account here presupposes neither a model nor some form of model retraining. A single measurement (a survey, a laboratory experiment, etc) should be designed in such a way as to be *reflexively optimal*. It should induce a distribution that is reliable evidence for the phenomenon under investigation (i.e., the sample distribution should be an accurate representation of the population distribution). Thus, it should incentivize truth-telling, discourage withholding relevant information, etc.

Further departing from (Perdomo et al. 2020), it is helpful to consider reflexive optimality an equilibrium notion[13] in the game-theoretic sense (e.g. Nash 1950). This is helpful for two reasons. First, it allows scientists to give an explicit model of agents' motivations and reasoning and how they interact with a measurement. As argued above, this is a key facet of understanding how reflexive science causally affects its target of study. Secondly, it facilitates the application of the techniques of mechanism design[14] to the problem of designing reflexively optimal measurement instruments. This turns out to be closely related to an active area of research in theoretical computer science called incentive-compatible learning. Here, the choice of statistical estimator or algorithm itself can induce people to adapt their behavior. Thus, it is possible to re-frame the choice of estimator or algorithm as one that induces truth-telling on behalf of those data are collected from. To better understand this approach, consider the following example from (Caragiannis, Procaccia, and Shah 2016):

Example 9 (Incentive-Compatible Mean Estimation)**.** *A statistician is trying to estimate the mean preferred temperature of occupants of a building. A sample of occupants are randomly selected and asked their preferred temperature. Consider the following scenarios.*

In one case, each person sampled is told that the estimator the statistician will use for their estimate of the population mean is the sample mean. Notice that if you have a preference for, say, warmer temperatures, you are best off lying about your preferred temperature to raise the sample average. This is because more extreme values will raise the sample average.

Suppose the statistician instead uses the sample median as his estimate of the population mean and this is communicated to each person in the sample. Even if you have a preference for much warmer temperatures, you no longer gain by lying since the median is robust to large outlier values (see Caragiannis, Procaccia, and Shah 2016 for extended discussion of this result).

In example 9 above, there is an explicit model of the relationship between the statistical estimator and the data collected in terms of benefits to people (i.e., their utility). The choice of statistical estimator (i.e., measurement instrument) is recast as a game theoretic problem whereby the statistician and the people in the sample play a game. The statistician wants to estimate the population mean. People in the sample will report their preferred temperature truthfully if they stand to benefit from it or can't benefit from lying. Thus, the statistician can use the *sample median* to estimate the *population mean* to achieve reflexive optimality. Note that despite the game-theoretic formulation of the problem the goal of statistical inference remains the same.

In this example, the sample distribution changes as a function of the estimator yet the underlying phenomenon of interest (people's preferred temperature) remains unchanged throughout. The measurement instrument (i.e., estimator) is chosen so as to induce a distribution shift which is more reliable evidence for people's preferred temperature. This framework of incentive-compatible estimators and algorithms has been extended to explicitly causal settings (Toulis et al. 2015), forecasting problems (Roughgarden and Schrijvers 2017), and even bandit-type exploration algorithms (Mansour, Slivkins, and Syrgkanis 2019). It makes the strong assumption that study participants know the functional form of the estimator and possess an ability to reason about how the actions they can take ultimately affect their welfare. However, it explicitly models the incentive structures faced by agents whose behavior is measured. This captures the key idea of the proposal that opened the section: scientists need to understand how their measurements affect the incentives of agents they collect data about.

[12]Their proposed definition is one of iterative convergence from model retraining (Perdomo et al. 2020, Definition 2.3). The use of the word 'prediction' should not confuse philosophers: the problem they consider is simultaneously a problem of measurement. Data are collected for retraining after each iteration of the model is deployed.

[13]The equilibrium notion of (Perdomo et al. 2020, p1) "coincide[s] with the stable points of [model] retraining" and does not reflect an understanding of why agents act they way they do. In contrast, recent work (Oesterheld et al. 2023) conceives of a truth-telling equilibrium in a performative prediction game as one induced by the self-interest of the participants. In the author's view, this latter contribution is the more promising approach to tackling the problem of reflexive measurement.

[14]See (Börgers 2015) for an introduction to the topic of mechanism design.

4.6 Conclusion

I have argued for a novel conception of reflexivity that puts the sociological practice of science at the center of our understanding of reflexivity. This move facilitates the consideration of the multitude of ways science and scientists causally affect their target of study. Reflexivity concerns a kind of causal effect that science has on its target of study where agents that comprise the phenomenon of interest are aware of the scrutiny they are subjected to. In the case of measurement, we can distinguish this measurement-as-intervention view by virtue of whether the measurement causally affects the underlying phenomenon or the data collected about it. The correct understanding of reflexive measurement is given by the concept of distribution shift, where the sample distribution is no longer an accurate representation of the population model due to the causal effect of the measurement.

A few points are worth emphasizing. Firstly, scientists who warn of the consequences of observer effects or self-fulfilling science often do not narrowly focus on either measurement or prediction but instead explore and investigate cases where they co-occur. In psychology, researchers (Gergen 1973; Luce 1995) consider how revealing a theoretical finding during a laboratory experiment can result in the study participants falsifying (or confirming) the experiment at will. In computer science, researchers have considered the effects of predictions on subsequent data collection (Perdomo et al. 2020). The account presented here offers a more general characterization of reflexivity which is more faithful to its varied manifestations. In my view, large swathes of science are entirely reflexive and yet, perhaps surprisingly, reflexive effects are often quite minor. Thankfully, scientists are increasingly aware of this phenomenon and have begun investigating specific occurrences of reflexivity in a far more thorough capacity than philosophers (e.g., Rothschild and Malhotra 2014; Westwood, Messing, and Lelkes 2020).

Additionally, an upshot of the characterization given in section 4.3 is that the terrain of the discussion concerning the presence of reflexivity, reactivity, and performativity in a scientific domain should shift from an antiquated "social" versus "natural" science framing to one instead marked by a deeper appreciation for the nature of agency. The question 'is a science reflexive?' is transformed into 'what is the nature of agency?' in virtue of the concern with awareness as the causal pathway along which reflexive effects materialize. Where opinions differ on the nature of agency, so too will they differ on the designations of reflexivity. Counterexamples to purely social scientific definitions of reflexivity like missiles (Grünbaum 1956) and thermostats (Beinhocker 2013) make this point abundantly clear. In my view, this is a welcome change. It moves from treating an existing academic division of labor as a primordial categorization of scientific practice to one instead informed by careful study of differing targets of inquiry.

Two extensions to the line of research initiated here are clear. Firstly, it is worth noting a significant omission from the present account is that of qualitative social science. Qualitative research techniques across the social sciences have become increasingly sophisticated (see, for example, King, Keohane, and Verba 2021). Further research outlining how the account presented here interacts with scientific practices like structured interviews and ethnographic research would be welcome. Secondly, it is an interesting and challenging proposition to extend the understanding of reflexive measurement as distribution shift to scientific practices like prediction and theory development. As introduced here, the concept is bound up with data collection, and extending the account here for other scientific practices may yield insights that aid scientists in overcoming the reflexive effects of science.

Almost half a century ago, political scientist Christopher Achen lamented the lack of understanding social scientists possess concerning how their measurement instruments investigate the world. He wrote:

> "[m]ajor improvements in our understanding of political thinking may therefore come to depend upon a considerably more advanced theoretical knowledge of our measuring instruments than we have yet mustered." (Achen 1975, p. 1231)

I have argued that part of the toolkit of modern science fails in the presence of a particular, pervasive type of measurement concern. It is my hope that this essay constitutes an accurate philosophical diagnosis of the problem of reflexivity and lays a conceptual foundation for future solutions to this problem in the context of measurement.

Chapter 5

Conclusion

5.1 Summary

In chapter 2 I propose a *minimal non-fable* account of economic theory in the domain of design economics. This account of economic theory does not make the success of design economics a miracle. To argue for this account, I adapt a no-miracles argument for scientific realism to the scientific domain of economics. With this *explanatory challenge* version of the no-miracles argument it becomes possible to philosophically interrogate the features of a successful science: are these features necessary to explain this science's successes? In the case of design economics, game theory is common to all its successes. Additionally, these theories are *projective* (Guala 2001) in the sense that the world can be made to mirror the models developed by economists. Moreover, as in the case of electronic auctions, the gap between representation and reality can be made arbitrarily tight. This accounts for my rejection of Ariel Rubinstein's conviction that economic theorists are "simply the tellers of fables" (Rubinstein 2006, p. 882).

Chapter 3 falls squarely in the tradition of design economics evaluated in the preceding chapter: the results are intended to "suggest[] an agenda for future theoretical work" (Roth 2002, p1363). I explore the properties of optimal multi-dimensional auctions in a setting where a single object of multiple qualities is sold to several buyers. Using simulations, I test the hypothesis that the optimal mechanism is an *exclusive-buyer mechanism*, where buyers compete for the right to be the only buyer to choose between quality levels of a good. I find that in most multidimensional settings considered in this thesis, the exclusive-buyer mechanism well-approximates the revenue generated by the optimal mechanism and qualitatively matches the optimal (interim) allocations. However, the exclusive-buyer mechanism is clearly not optimal when the optimal mechanism yielded by simulations is not deterministic. Additionally, I provide evidence for the optimality of mechanisms with measure zero exclusion regions and find consistent evidence that the exclusion region remains constant with the number of buyers.

Finally, in chapter 4, I offer a unified account of observer effects in the social sciences. I extend existing philosophical accounts of reflexivity to measurement, what I call the problem of *reflexive measurement*. Additionally, my account draws from the extensive writings of social scientists who confront this measurement challenge as a practical matter. I argue that the problem of reflexive measurement is akin to the well-known idea of 'measurement-as-intervention' in philosophy of economics (Morgan 2001). I contend that what matters for reflexivity in science is whether agents are aware they are the subject of scientific investigation. In light of this understanding of reflexive measurement, I explore the extent to which mechanism design and, in particular, *incentive compatible learning* can be used to mitigate the causal effects of social science on what it studies. I argue that the application of mechanism design to problems of measurement can be used to explicitly model the problem of misaligned incentives created when scientists investigate human behavior.

5.2 Future Work

The conclusions concerning optimal multidimensional auctions in chapter 3 invite further study of the exclusive-buyer mechanism. This mechanism is clearly optimal in cases where randomization is not re-

quired (conjecture 1) and captures the qualitative behavior of the allocations (conjecture 2). Furthermore, it is possible to extend the exclusive-buyer mechanism to include stochastic contracts, which entails that a more general formulation of this kind of mechanism may be optimal for the general multidimensional setting of a single good with multiple quality levels. The assumptions of identical bidder valuations and linear virtual values may drive the simulation results and alternative specifications should be explored. The exclusive-buyer mechanism is also compatible with measure zero exclusion regions in multidimensional settings, an unexpected phenomenon that merits further exploration.

The combined conclusions of chapters 2 and 4 point to the concrete application of economic theory to problems of measurement. Recent theoretical research on incentive-compatible estimators (Caragiannis, Procaccia, and Shah 2016) and data acquisition procedures (Bates et al. 2022; Roth and Schoenebeck 2012) has, to the best of my knowledge, yet to be applied in practice. Insofar as one can argue for the value of economic theory in bringing about the successes of design economics, this theory should be put to use in designing better measurements. To what extent 'designing' a measurement procedure is similar to the problems of design economics is an open question. In my view, answering this question is of paramount importance for social scientists seeking a "more advanced theoretical knowledge of our measuring instruments" (Achen 1975, p. 1231).

Chapter 6

Bibliography

Achen, Christopher H. (1975). "Mass Political Attitudes and the Survey Response". In: *American Political Science Review* 69.4, 1218–1231. DOI: 10.2307/1955282.

Adams, William James and Janet L. Yellen (1976). "Commodity Bundling and the Burden of Monopoly". In: *The Quarterly Journal of Economics* 90.3, pp. 475–498. ISSN: 00335533, 15314650. URL: http://www.jstor.org/stable/1886045 (visited on 11/06/2023).

Alaei, Saeed et al. (2019). "Efficient computation of optimal auctions via reduced forms". In: *Mathematics of Operations Research* 44.3, pp. 1058–1086.

Alexandrova, Anna and Robert Northcott (Mar. 2009). "306 Progress in Economics: Lessons from the Spectrum Auctions". In: *The Oxford Handbook of Philosophy of Economics*. Oxford University Press. ISBN: 9780195189254. DOI: 10.1093/oxfordhb/9780195189254.003.0011. URL: https://doi.org/10.1093/oxfordhb/9780195189254.003.0011.

Armstrong, Mark (1996). "Multiproduct nonlinear pricing". In: *Econometrica* 64.1, pp. 51–75.

Balestrieri, Filippo, Sergei Izmalkov, and Joao Leao (June 2020). "The Market for Surprises: Selling Substitute Goods Through Lotteries". In: *Journal of the European Economic Association* 19.1, pp. 509–535. ISSN: 1542-4766. DOI: 10.1093/jeea/jvaa021. URL: https://doi.org/10.1093/jeea/jvaa021.

Barelli, Paulo et al. (2014). "On the optimality of exclusion in multi-dimensional screening". In: *Journal of Mathematical Economics* 54, pp. 74–83. ISSN: 0304-4068. DOI: https://doi.org/10.1016/j.jmateco.2014.09.002. URL: https://www.sciencedirect.com/science/article/pii/S0304406814001128.

Basov, S. (2005). *Multidimensional Screening*. Studies in Economic Theory. Springer Berlin Heidelberg. ISBN: 9783540273134.

Basu, Tanya (Oct. 4, 2015). "New Google Parent Company Drops 'Don't Be Evil' Motto". In: *Time*. URL: https://time.com/4060575/alphabet-google-dont-be-evil/ (visited on 06/10/2023).

Bates, Stephen et al. (2022). *Principal-Agent Hypothesis Testing*. arXiv: 2205.06812 [cs.GT].

Beinhocker, Eric D. (2013). "Reflexivity, complexity, and the nature of social science". In: *Journal of Economic Methodology* 20.4, pp. 330–342. DOI: 10.1080/1350178X.2013.859403.

Belloni, Alexandre, Giuseppe Lopomo, and Shouqiang Wang (2010). "Multidimensional mechanism design: Finite-dimensional approximations and efficient computation". In: *Operations Research* 58.4-part-2, pp. 1079–1089.

Binmore, Ken and Paul Klemperer (2002). "The Biggest Auction Ever: The Sale of The British 3G Telecom Licenses". In: *The Economic Journal* 112.478, pp. C74–C96. DOI: https://doi.org/10.1111/1468-0297.00020.

Blake, Aaron (Jan. 18, 2022). "The audacious timeline of Trump's failed plot on the census and citizenship". In: *The Washington Post*. URL: https://www.washingtonpost.com/politics/2022/01/18/audacious-timeline-trumps-failed-plot-census-citizenship (visited on 06/05/2023).

Bondt, Werner F. M. De and Richard Thaler (1985). "Does the Stock Market Overreact?" In: *The Journal of Finance* 40.3, pp. 793–805. URL: http://www.jstor.org/stable/2327804.

Border, Kim C (1991). "Implementation of reduced form auctions: A geometric approach". In: *Econometrica: Journal of the Econometric Society* 59.4, pp. 1175–1187.

Brusco, Sandro, Giuseppe Lopomo, and Leslie M. Marx (2009). "The 'Google effect' in the FCC's 700MHz auction". In: *Information Economics and Policy* 21.2. Special Section on Auctions, pp. 101–114. ISSN: 0167-6245. DOI: https://doi.org/10.1016/j.infoecopol.2009.03.001.

— (2011). "The Economics of Contingent Re-auctions". In: *American Economic Journal: Microeconomics* 3.2, pp. 165–93. DOI: 10.1257/mic.3.2.165. URL: https://www.aeaweb.org/articles?id=10.1257/mic.3.2.165.

Buck, Roger C. (1963). "Reflexive Predictions". In: *Philosophy of Science* 30.4, pp. 359–369. DOI: 10.1086/287955.

Börgers, Tilman (2015). *An Introduction to the Theory of Mechanism Design*. Oxford University Press.

Cai, Yang, Constantinos Daskalakis, and Christos Papadimitriou (2015). "Optimum Statistical Estimation with Strategic Data Sources". en. In: *Conference on Learning Theory*. ISSN: 1938-7228. PMLR, pp. 280–296. URL: http://proceedings.mlr.press/v40/Cai15.html (visited on 07/29/2021).

Cai, Yang, Constantinos Daskalakis, and S. Matthew Weinberg (2012). "Optimal Multi-dimensional Mechanism Design: Reducing Revenue to Welfare Maximization". In: *2012 IEEE 53rd Annual Symposium on Foundations of Computer Science*, pp. 130–139. DOI: 10.1109/FOCS.2012.88.

Cai, Yang, Nikhil R. Devanur, and S. Matthew Weinberg (2016). "A Duality Based Unified Approach to Bayesian Mechanism Design". In: *Proceedings of the Forty-Eighth Annual ACM Symposium on Theory of Computing*. STOC '16. Cambridge, MA, USA: Association for Computing Machinery, 926–939. ISBN: 9781450341325. DOI: 10.1145/2897518.2897645. URL: https://doi.org/10.1145/2897518.2897645.

Caragiannis, Ioannis, Ariel Procaccia, and Nisarg Shah (2016). "Truthful Univariate Estimators". In: *Proceedings of The 33rd International Conference on Machine Learning*. Ed. by Maria Florina Balcan and Kilian Q. Weinberger. Vol. 48. Proceedings of Machine Learning Research. New York, New York, USA: PMLR, pp. 127–135.

Cejka, Timotej (2022). "Reflexivity of Predictions as a Statistical Bias". In: URL: http://philsci-archive.pitt.edu/21326/.

Chakravartty, Anjan (2017). "Scientific Realism". In: *The Stanford Encyclopedia of Philosophy*. Ed. by Edward N. Zalta. Summer 2017. Metaphysics Research Lab, Stanford University.

Chawla, Shuchi, Jason D. Hartline, and Robert Kleinberg (2007). "Algorithmic Pricing via Virtual Valuations". In: *Proceedings of the 8th ACM Conference on Electronic Commerce*. EC '07. San Diego, California, USA: Association for Computing Machinery, 243–251. ISBN: 9781595936530. DOI: 10.1145/1250910.1250946. URL: https://doi.org/10.1145/1250910.1250946.

Colman, Andrew M. (2008). *A Dictionary of Psychology*. Ed. by Andrew M. Colman. 3rd. Oxford University Press. ISBN: 9780199534067.

Conitzer, Vincent and Tuomas Sandholm (2002). "Complexity of mechanism design". In: *Proceedings of the Eighteenth Conference on Uncertainty in Artificial Intelligence*. UAI'02. Alberta, Canada: Morgan Kaufmann Publishers Inc., 103–110. ISBN: 1558608974.

— (2004). "Self-interested automated mechanism design and implications for optimal combinatorial auctions". In: *Proceedings of the 5th ACM Conference on Electronic Commerce*. EC '04. New York, NY, USA: Association for Computing Machinery, 132–141. ISBN: 1581137710. DOI: 10.1145/988772.988793. URL: https://doi.org/10.1145/988772.988793.

Conitzer, Vincent et al. (2021). "Automated Mechanism Design for Strategic Classification". In: *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*. KDD '21. Virtual Event, Singapore: Association for Computing Machinery, p. 1. ISBN: 9781450383325. DOI: 10.1145/3447548.3469650. URL: https://doi.org/10.1145/3447548.3469650.

Daskalakis, Constantinos (2015). "Multi-item auctions defying intuition?" In: *ACM SIGecom Exchanges* 14.1, pp. 41–75.

Daskalakis, Constantinos, Alan Deckelbaum, and Christos Tzamos (2017). "Strong duality for a multiple-good monopolist". In: *Econometrica* 85.3, pp. 735–767.

Dütting, Paul et al. (2024). "Optimal Auctions through Deep Learning: Advances in Differentiable Economics". In: *J. ACM* 71.1. ISSN: 0004-5411. DOI: 10.1145/3630749. URL: https://doi.org/10.1145/3630749.

Ekeland, Ivar (2010). "Notes on optimal transportation". In: *Economic Theory* 42.2, pp. 437–459. DOI: 10.1007/s00199-008-0426-9. URL: https://ideas.repec.org/a/spr/joecth/v42y2010i2p437-459.html.

Federal Communications Commision (May 2017). *Broadcast Incentive Auction and Post-Auction Transition*. URL: https://www.fcc.gov/about-fcc/fcc-initiatives/incentive-auctions.

Friedman, M. (1953). *Essays in Positive Economics*. A Phoenix book. Business economics. University of Chicago Press. ISBN: 9780226264035.

Gale, D. and L. S. Shapley (1962). "College Admissions and the Stability of Marriage". In: *The American Mathematical Monthly* 69.1, pp. 9–15. ISSN: 00029890, 19300972. URL: http://www.jstor.org/stable/2312726 (visited on 06/06/2024).

Gergen, Kenneth J. (1973). "Social Psychology as History". In: *Journal of Personality and Social Psychology* 26.2, pp. 309–320.

Goodhart, C. A. E. (1984). "Problems of Monetary Management: The UK Experience". In: *Monetary Theory and Practice: The UK Experience*. London: Macmillan Education UK. DOI: 10.1007/978-1-349-17295-5_4. URL: https://doi.org/10.1007/978-1-349-17295-5_4.

Grunberg, Emile (1986). "Predictability and Reflexivity". In: *The American Journal of Economics and Sociology* 45.4, pp. 475–488. DOI: https://doi.org/10.1111/j.1536-7150.1986.tb01946.x.

Grünbaum, Adolf (1956). "Historical Determinism, Social Activism, and Predictions in the Social Sciences". In: *The British Journal for the Philosophy of Science* 7.27, pp. 236–240. URL: http://www.jstor.org/stable/685878.

— (1963). "Comments on Professor Roger Buck's Paper "Reflexive Predictions"". In: *Philosophy of Science* 30.4, 370–372. DOI: 10.1086/287956.

Guala, Francesco (2001). "Building economic machines: The FCC auctions". In: *Studies in History and Philosophy of Science Part A* 32.3, pp. 453–477. ISSN: 0039-3681. DOI: https://doi.org/10.1016/S0039-3681(01)00008-5. URL: https://www.sciencedirect.com/science/article/pii/S0039368101000085.

Guesnerie, Roger and Jean-Jacques Laffont (1984). "A complete solution to a class of principal-agent problems with an application to the control of a self-managed firm". In: *Journal of Public Economics* 25.3, pp. 329–369. ISSN: 0047-2727. DOI: https://doi.org/10.1016/0047-2727(84)90060-4. URL: https://www.sciencedirect.com/science/article/pii/0047272784900604.

Hacking, Ian (1983). *Representing and Intervening: Introductory Topics in the Philosophy of Natural Science*. Cambridge University Press. DOI: 10.1017/CBO9780511814563.

— (1995). "The Looping Effects of Human Kinds". In: *Causal cognition: a multi-disciplinary debate*. Ed. by David Premack Dan Sperber and Ann James Premack. Chap. 12, pp. 351–383.

Haghpanah, Nima and Jason Hartline (2014). *Multi-dimensional Virtual Values and Second-degree Price Discrimination*. DOI: 10.48550/ARXIV.1404.1341. URL: https://arxiv.org/abs/1404.1341.

— (2021). "When Is Pure Bundling Optimal?" In: *Review of Economic Studies* 88.3, pp. 1127–1156.

Hardt, Moritz et al. (2016). "Strategic Classification". In: *Proceedings of the 2016 ACM Conference on Innovations in Theoretical Computer Science*. ITCS '16. Cambridge, Massachusetts, USA: Association for Computing Machinery, 111–122. ISBN: 9781450340571. DOI: 10.1145/2840728.2840730.

Hart, Sergiu and Noam Nisan (2019). "Selling multiple correlated goods: Revenue maximization and menu-size complexity". In: *Journal of Economic Theory* 183, pp. 991–1029. ISSN: 0022-0531. DOI: https://doi.org/10.1016/j.jet.2019.07.006. URL: https://www.sciencedirect.com/science/article/pii/S0022053119300717.

Hitzig, Zoë (2024). *The Normative Gap: Mechanism Design and Ideal Theories of Justice*. URL: https://scholar.harvard.edu/files/hitzig/files/finalrevision.pdf.

Hotelling, Harold (1929). "Stability in Competition". In: *The Economic Journal* 39.153, pp. 41–57. ISSN: 00130133, 14680297. URL: http://www.jstor.org/stable/2224214 (visited on 11/06/2023).

Hurwicz, L. and S. Reiter (2006). *Designing Economic Mechanisms*. Cambridge University Press. ISBN: 9781139454346.

Hurwicz, Leonid (1972). *On informationally decentralized systems*. Amsterdam [usw.]

Jullien, Bruno (2000). "Participation Constraints in Adverse Selection Models". In: *Journal of Economic Theory* 93.1, pp. 1–47. ISSN: 0022-0531. DOI: https://doi.org/10.1006/jeth.1999.2641. URL: https://www.sciencedirect.com/science/article/pii/S0022053199926418.

Kagel, John H. and Alvin E. Roth (Feb. 2000). "The Dynamics of Reorganization in Matching Markets: A Laboratory Experiment Motivated by a Natural Experiment*". In: *The Quarterly Journal of Economics* 115.1, pp. 201–235. ISSN: 0033-5533. DOI: 10.1162/003355300554719. URL: https://doi.org/10.1162/003355300554719.

King, G., R.O. Keohane, and S. Verba (2021). *Designing Social Inquiry: Scientific Inference in Qualitative Research, New Edition*. Princeton University Press. ISBN: 9780691224640.

Kolesnikov, Alexander et al. (2022). *Beckmann's approach to multi-item multi-bidder auctions*. DOI: 10.48550/ARXIV.2203.06837. URL: https://arxiv.org/abs/2203.06837.

Kopec, Matthew (2011). "A More Fulfilling (and Frustrating) Take on Reflexive Predictions". In: *Philosophy of Science* 78.5, pp. 1249–1259.

Krumpal, Ivar (2013). "Determinants of social desirability bias in sensitive surveys: a literature review". In: *Quality & Quantity* 47, 2025–2047. DOI: https://doi.org/10.1007/s11135-011-9640-9.

Landsberger, H. A. (1958). "Hawthorne Revisited. Management and the Worker, its Critics and Developments in Human Relations in Industry." In: *Cornell Studies in Industrial and Labor Relations* IX.

Ledyard, John O., David Porter, and Antonio Rangel (1997). "Experiments Testing Multiobject Allocation Mechanisms". In: *Journal of Economics & Management Strategy* 6.3, pp. 639–675. DOI: https://doi.org/10.1111/j.1430-9134.1997.00639.x. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1430-9134.1997.00639.x.

Lee, Nicol Turner and Jack Malamud (2023). "Reinstating the FCC's auction authority could save the Affordable Connectivity Program". In: *TechTank: Improving Technology Policy*. URL: https://www.brookings.edu/articles/reinstating-the-fccs-auction-authority-could-save-the-affordable-connectivity-program/#:~:text=With%20$230%20billion%20dollars%20of,to%20lapse%20earlier%20this%20year..

Li, Shengwu (2017). "Obviously Strategy-Proof Mechanisms". In: *American Economic Review* 107.11, 3257–87. DOI: 10.1257/aer.20160425. URL: https://www.aeaweb.org/articles?id=10.1257/aer.20160425.

Loertscher, Simon and Ellen Muir (2023). *Optimal Hotelling Auctions*.

Lowe, Charles (2018). "The Significance of Self-Fulfilling Science". In: *Philosophy of the Social Sciences* 48.4, pp. 343–363. DOI: 10.1177/0048393118767087.

Luce, R. Duncan (1995). "Four Tensions Concerning Mathematical Modeling in Psychology". In: *Annual Review of Psychology* 46.1, pp. 1–27. DOI: 10.1146/annurev.ps.46.020195.000245. URL: https://doi.org/10.1146/annurev.ps.46.020195.000245.

Mackinnon, Lauchlan (2006). "Appendix B: Reflexive Prediction". Unpublished PhD thesis. URL: https://www.researchgate.net/publication/37618514_Reflexive_Prediction_A_Literature_Review. PhD thesis. University of Queensland.

Magnus, P. D. and Craig Callender (2004). "Realist Ennui and the Base Rate Fallacy". In: *Philosophy of Science* 71.3, 320–338. DOI: 10.1086/421536.

Mäki, Ismo Uskali (1992). "Friedman and realism". In: *Research in the History of Economic Thought and Methodology* 10, pp. 171–195. ISSN: 0743-4154.

— (Mar. 2009). "Realistic Realism about Unrealistic Models". In: *The Oxford Handbook of Philosophy of Economics*. Oxford University Press. ISBN: 9780195189254. DOI: 10.1093/oxfordhb/9780195189254.003.0004.

Manelli, Alejandro M. and Daniel R. Vincent (2006). "Bundling as an optimal selling mechanism for a multiple-good monopolist". In: *Journal of Economic Theory* 127.1, pp. 1–35. ISSN: 0022-0531. DOI: https://doi.org/10.1016/j.jet.2005.08.007.

Manelli, Alejandro M and Daniel R Vincent (2007). "Multidimensional mechanism design: Revenue maximization and the multiple-good monopoly". In: *Journal of Economic theory* 137.1, pp. 153–185.

Mansour, Yishay, Aleksandrs Slivkins, and Vasilis Syrgkanis (2019). *Bayesian Incentive-Compatible Bandit Exploration*. arXiv: 1502.04147 [cs.GT].

Marget, Arthur W. (1929). "Morgenstern on the Methodology of Economic Forecasting". In: *Journal of Political Economy* 37.3, pp. 312–339. ISSN: 00223808, 1537534X. URL: http://www.jstor.org/stable/1824409 (visited on 02/16/2024).

McAfee, R. Preston and John McMillan (1996). "Analyzing the Airwaves Auction". In: *Journal of Economic Perspectives* 10.1, pp. 159–175. DOI: 10.1257/jep.10.1.159. URL: https://www.aeaweb.org/articles?id=10.1257/jep.10.1.159.

McAfee, R. Preston, John McMillan, and Michael D. Whinston (1989). "Multiproduct Monopoly, Commodity Bundling, and Correlation of Values". In: *The Quarterly Journal of Economics* 104.2, pp. 371–383. ISSN: 00335533, 15314650. URL: http://www.jstor.org/stable/2937852 (visited on 12/24/2022).

McMillan, John (1994). "Selling Spectrum Rights". In: *Journal of Economic Perspectives* 8.3, pp. 145–162. DOI: 10.1257/jep.8.3.145. URL: https://www.aeaweb.org/articles?id=10.1257/jep.8.3.145.

Merton, Robert K. (1948). "The Self-Fulfilling Prophecy". In: *Antioch Review* 8, pp. 193–210.

Milgrom, Paul (2004). *Putting Auction Theory to Work*. Churchill Lectures in Economics. Cambridge University Press. DOI: 10.1017/CBO9780511813825.

Mirrlees, J. A. (1971). "An Exploration in the Theory of Optimum Income Taxation". In: *The Review of Economic Studies* 38.2, pp. 175–208. URL: http://www.jstor.org/stable/2296779 (visited on 11/06/2023).

Morgan, Mary S. (Dec. 2001). "Making Measuring Instruments". In: *History of Political Economy* 33.1, pp. 235–251. DOI: 10.1215/00182702-33-Suppl_1-235.

Moulin, Hervé (1988). *Axioms of Cooperative Decision Making*. Econometric Society Monographs. Cambridge University Press.

Mussa, Michael and Sherwin Rosen (1978). "Monopoly and product quality". In: *Journal of Economic Theory* 18.2, pp. 301–317. ISSN: 0022-0531. DOI: https://doi.org/10.1016/0022-0531(78)90085-6. URL: https://www.sciencedirect.com/science/article/pii/0022053178900856.

Myerson, Roger B. (1981). "Optimal auction design". In: *Mathematics of Operations Research* 6.1, pp. 58–73.

— (2008). "Perspectives on Mechanism Design in Economic Theory". In: *American Economic Review* 98.3, pp. 586–603. DOI: 10.1257/aer.98.3.586. URL: https://www.aeaweb.org/articles?id=10.1257/aer.98.3.586.

Myerson, Roger B and Mark A Satterthwaite (1983). "Efficient mechanisms for bilateral trading". In: *Journal of Economic Theory* 29.2, pp. 265–281. ISSN: 0022-0531. DOI: https://doi.org/10.1016/0022-0531(83)90048-0. URL: https://www.sciencedirect.com/science/article/pii/0022053183900480.

Nagel, E. (1961). *The Structure of Science: Problems in the Logic of Scientific Explanation*. Donald F. Koch American Philosophy Collection. Harcourt, Brace & World. ISBN: 9780710018823.

NASA (2024). *1P/Halley*. URL: https://science.nasa.gov/solar-system/comets/1p-halley/.

Nash, John F. (1950). "Equilibrium points in ¡i¿n¡/i¿-person games". In: *Proceedings of the National Academy of Sciences* 36.1, pp. 48–49. DOI: 10.1073/pnas.36.1.48.

Newman, Neil et al. (2020). "Incentive Auction Design Alternatives: A Simulation Study". In: *Proceedings of the 21st ACM Conference on Economics and Computation*. EC '20. Virtual Event, Hungary: Association for Computing Machinery, 603–604. ISBN: 9781450379755. DOI: 10.1145/3391403.3399499. URL: https://doi.org/10.1145/3391403.3399499.

Nik-Khah, Edward (2008). "A tale of two auctions". In: *Journal of Institutional Economics* 4.1, 73–97. DOI: 10.1017/S1744137407000859.

Nisan, N. et al. (2007). *Algorithmic Game Theory*. Cambridge University Press. ISBN: 9781139466547.

Oesterheld, Caspar et al. (2023). "Incentivizing honest performative predictions with proper scoring rules". In: *Proceedings of the Thirty-Ninth Conference on Uncertainty in Artificial Intelligence*. Ed. by Robin J. Evans and Ilya Shpitser. Vol. 216. Proceedings of Machine Learning Research. PMLR, pp. 1564–1574. URL: https://proceedings.mlr.press/v216/oesterheld23a.html.

Orne, M. T. (1962). "On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications". In: *American Psychologist* 17.11, 776–783. DOI: 10.1037/h0043424.

Palacios-Huerta, Ignacio, David C Parkes, and Richard Steinberg (2022). "Combinatorial auctions in practice". In: *Available at SSRN 3844338*.

Patrick Cantwell Peter Davis, James Mulligan (2012). *DSSD 2010 CENSUS COVERAGE MEASUREMENT MEMORANDUM SERIES #2010-G-03*. US Census Bureau. URL: https://www2.census.gov/programs-surveys/decennial/2010/technical-documentation/methodology/g-series/g03.pdf.

Pavlov, Gregory (2011). "Optimal mechanism for selling two goods". In: *The BE Journal of Theoretical Economics* 11.1.

Perdomo, Juan et al. (2020). "Performative Prediction". In: *Proceedings of the 37th International Conference on Machine Learning*. Ed. by Hal Daumé III and Aarti Singh. Vol. 119. Proceedings of Machine Learning Research. PMLR, pp. 7599–7609. URL: https://proceedings.mlr.press/v119/perdomo20a.html.

Perron, Laurent and Vincent Furnon (Aug. 8, 2023). *OR-Tools*. Version v9.7. Google. URL: https://developers.google.com/optimization/.

Plott, Charles R. (1997). "Laboratory Experimental Testbeds: Application to the PCS Auction". In: *Journal of Economics & Management Strategy* 6.3, pp. 605–638. DOI: https://doi.org/10.1111/j.1430-9134.1997.00605.x. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1430-9134.1997.00605.x.

Popper, Karl (1953). *The Poverty of Historicism*. Harper Torchbooks.

Putnam, H. (1975). *Mathematics, Matter and Method: Volume 1, Philosophical Papers*. Cambridge Paperback Library. Cambridge University Press. ISBN: 9780521206655.

Riley, John and Richard Zeckhauser (1983). "Optimal Selling Strategies: When to Haggle, When to Hold Firm". In: *The Quarterly Journal of Economics* 98.2, pp. 267–289. URL: https://EconPapers.repec.org/RePEc:oup:qjecon:v:98:y:1983:i:2:p:267-289..

Rochet, Jean-Charles (1987). "A necessary and sufficient condition for rationalizability in a quasi-linear context". In: *Journal of Mathematical Economics* 16.2, pp. 191–200. ISSN: 0304-4068. DOI: https://doi.org/10.1016/0304-4068(87)90007-3. URL: https://www.sciencedirect.com/science/article/pii/0304406887900073.

Rochet, Jean-Charles and Philippe Choné (1998). "Ironing, sweeping, and multidimensional screening". In: *Econometrica* 66.4, pp. 783–826.

Rochet, Jean-Charles and Lars A. Stole (2003). "The Economics of Multidimensional Screening". In: *Advances in Economics and Econometrics: Theory and Applications, Eighth World Congress*. Ed. by Mathias Dewatripont, Lars Peter Hansen, and Stephen J.Editors Turnovsky. Vol. 1. Econometric Society Monographs. Cambridge University Press, 150–197. DOI: 10.1017/CBO9780511610240.006.

Romanos, George D. (1973). "Reflexive Predictions". In: *Philosophy of Science* 40.1, 97–109. DOI: 10.1086/288499.

Rosenthal, Robert (1966). *Experimenter effects in behavioral research*. Appleton-Century-Crofs.

Rosenthal, Robert and Lenore Jacobson (1968). "Pygmalion in the classroom". In: *The Urban Review* 3, pp. 16–20. DOI: https://doi.org/10.1007/BF02322211.

Ross, Don (2008). "Ontic Structural Realism and Economics". In: *Philosophy of Science* 75.5, 732–743. DOI: 10.1086/594518.

Roth, Aaron and Grant Schoenebeck (2012). *Conducting Truthful Surveys, Cheaply*. arXiv: 1203.0353 [cs.GT]. URL: https://arxiv.org/abs/1203.0353.

Roth, Alvin E. (2002). "The Economist as Engineer: Game Theory, Experimentation, and Computation as Tools for Design Economics". In: *Econometrica* 70.4, pp. 1341–1378. DOI: https://doi.org/10.1111/1468-0262.00335. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/1468-0262.00335. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/1468-0262.00335.

— (Oct. 2007). "The Art of Designing Markets". In: *Harvard Business Review*. URL: https://hbr.org/2007/10/the-art-of-designing-markets.

— (2018). "Marketplaces, Markets, and Market Design". In: *American Economic Review* 108.7, pp. 1609–58. DOI: 10.1257/aer.108.7.1609. URL: https://www.aeaweb.org/articles?id=10.1257/aer.108.7.1609.

Roth, Alvin E. and Elliott Peranson (1999). "The Redesign of the Matching Market for American Physicians: Some Engineering Aspects of Economic Design". In: *American Economic Review* 89.4, pp. 748–780. DOI: 10.1257/aer.89.4.748. URL: https://www.aeaweb.org/articles?id=10.1257/aer.89.4.748.

Roth, Alvin E. and Robert B. Wilson (2019). "How Market Design Emerged from Game Theory: A Mutual Interview". In: *Journal of Economic Perspectives* 33.3, pp. 118–43. DOI: 10.1257/jep.33.3.118. URL: https://www.aeaweb.org/articles?id=10.1257/jep.33.3.118.

Rothschild, David and Neil Malhotra (2014). "Are public opinion polls self-fulfilling prophecies?" In: *Research & Politics* 1.2, p. 2053168014547667. DOI: 10.1177/2053168014547667.

Roughgarden, Tim and Okke Schrijvers (2017). "Online Prediction with Selfish Experts". In: *Advances in Neural Information Processing Systems*. Ed. by I. Guyon et al. Vol. 30. Curran Associates, Inc.

Rubinstein, Ariel (1982). "Perfect Equilibrium in a Bargaining Model". In: *Econometrica* 50.1, pp. 97–109. ISSN: 00129682, 14680262. URL: http://www.jstor.org/stable/1912531 (visited on 04/24/2024).

— (1991). "Comments on the Interpretation of Game Theory". In: *Econometrica* 59.4, pp. 909–924. ISSN: 00129682, 14680262. URL: http://www.jstor.org/stable/2938166 (visited on 04/22/2024).

— (2006). "Dilemmas of an Economic Theorist". In: *Econometrica* 74.4, pp. 865–883. ISSN: 00129682, 14680262. URL: http://www.jstor.org/stable/3805911 (visited on 04/22/2024).

— (2012). *Economic Fables*. DOAB Directory of Open Access Books. Open Book. ISBN: 9781906924775.
Soros, George (2013). "Fallibility, reflexivity, and the human uncertainty principle". In: *Journal of Economic Methodology* 20.4, pp. 309–329. DOI: 10.1080/1350178X.2013.859415.
Spiegler, Ran (Apr. 2024). *The Curious Culture of Economic Theory*. The MIT Press. ISBN: 9780262379038. DOI: 10.7551/mitpress/14884.001.0001. eprint: https://direct.mit.edu/book-pdf/2362281/book_9780262379038.pdf. URL: https://doi.org/10.7551/mitpress/14884.001.0001.
Stanford, P. Kyle (2000). "An Antirealist Explanation of the Success of Science". In: *Philosophy of Science* 67.2, pp. 266–284. ISSN: 00318248, 1539767X. URL: http://www.jstor.org/stable/188724 (visited on 06/06/2024).
Stanford Prison Experiment. URL: https://www.prisonexp.org (visited on 07/11/2023).
Stantcheva, Stefanie (Oct. 2022). *How to Run Surveys: A guide to creating your own identifying variation and revealing the invisible*. URL: https://scholar.harvard.edu/files/stantcheva/files/How_to_run_surveys_Stantcheva.pdf (visited on 07/11/2023).
Sönmez, Tayfun (2023). *Minimalist Market Design: A Framework for Economists with Policy Aspirations*. arXiv: 2401.00307 [econ.GN].
Tal, Eran (2019). "Individuating quantities". In: *Philosophical Studies* 176.4, pp. 853–878. DOI: 10.1007/s11098-018-1216-2.
Texier, Thibault Le (2019). "Debunking the Stanford Prison Experiment". In: *American Psychologist* 74.7, pp. 823–839. DOI: https://doi.org/10.1037/amp0000401.
Thanassoulis, John (2004). "Haggling over substitutes". In: *Journal of Economic Theory* 117.2, pp. 217–245. ISSN: 0022-0531. DOI: https://doi.org/10.1016/j.jet.2003.09.002. URL: https://www.sciencedirect.com/science/article/pii/S0022053103003351.
The Royal Swedish Academy of Sciences (2020). *The Prize in Economic Sciences 2020: Popular Science Background*.
Thirumulanathan, D., Rajesh Sundaresan, and Y. Narahari (2019a). "On optimal mechanisms in the two-item single-buyer unit-demand setting". In: *Journal of Mathematical Economics* 82, pp. 31–60. ISSN: 0304-4068. DOI: https://doi.org/10.1016/j.jmateco.2019.01.005. URL: https://www.sciencedirect.com/science/article/pii/S030440681930014X.
— (2019b). "Optimal mechanisms for selling two items to a single buyer having uniformly distributed valuations". In: *Journal of Mathematical Economics* 82, pp. 1–30. ISSN: 0304-4068. DOI: https://doi.org/10.1016/j.jmateco.2019.01.004. URL: https://www.sciencedirect.com/science/article/pii/S0304406819300138.
Toulis, Panos et al. (2015). "Incentive-Compatible Experimental Design". In: *Proceedings of the Sixteenth ACM Conference on Economics and Computation*. DOI: 10.1145/2764468.2764525.
Varian, Hal R. (2009). "Online Ad Auctions". In: *American Economic Review* 99.2, pp. 430–34. DOI: 10.1257/aer.99.2.430. URL: https://www.aeaweb.org/articles?id=10.1257/aer.99.2.430.
Vickrey, William (1961). "COUNTERSPECULATION, AUCTIONS, AND COMPETITIVE SEALED TENDERS". In: *The Journal of Finance* 16.1, pp. 8–37. DOI: https://doi.org/10.1111/j.1540-6261.1961.tb02789.x.
Wang, Zihe and Pingzhong Tang (2014). "Optimal Mechanisms with Simple Menus". In: *Proceedings of the Fifteenth ACM Conference on Economics and Computation*. EC '14. Palo Alto, California, USA: Association for Computing Machinery, 227–240. ISBN: 9781450325653. DOI: 10.1145/2600057.2602863. URL: https://doi.org/10.1145/2600057.2602863.
Westwood, Sean Jeremy, Solomon Messing, and Yphtach Lelkes (2020). "Projecting Confidence: How the Probabilistic Horse Race Confuses and Demobilizes the Public". In: *The Journal of Politics* 82.4, pp. 1530–1544. DOI: 10.1086/708682.
Wilson, R.B. (1993). *Nonlinear Pricing*. Oxford University Press. ISBN: 9780195115826.
Wolfstetter, Elmar G. (2001). *The Swiss UMTS Spectrum Auction Flop: Bad Luck or Bad Design*. CESifo Working Paper Series 534. CESifo. URL: https://ideas.repec.org/p/ces/ceswps/_534.html.
Woodward, James and Lauren Ross (2021). "Scientific Explanation". In: *The Stanford Encyclopedia of Philosophy*. Ed. by Edward N. Zalta. Summer 2021. Metaphysics Research Lab, Stanford University.
Woodward, Jim (1989). "Data and Phenomena". In: *Synthese* 79.3, pp. 393–472.
Zimbardo, P.G. (2008). *The Lucifer Effect: How Good People Turn Evil*. Rider. ISBN: 9781846041037.

Chapter 7

Appendix

7.1 Approximation Algorithm for Optimal Multidimensional Auctions

I adopt and improve the original finite-dimensional approximation algorithm of (Belloni, Lopomo, and Wang 2010) by focusing on local and downward-sloping incentive-compatibility constraint (ICC) violations. Although these local constraints are often violated in this approximate setting, I drastically reduce the number of times *all* incentive-compatible constraints need to be checked.

In order to approximate an optimal solution to OPT_*, I discretize the type space X. Let T denote a positive integer that controls the granularity of the discretization. For each $j \in J$, let $X_T(j)$ denote the discretization of the interval $[\underline{x}_j, \overline{x}_j]$ given by $X_T(j) = \{\underline{x}_j, \underline{x}_j + \epsilon, \underline{x}_j + 2\epsilon, \ldots, \overline{x}_j\}$ where $\epsilon = \min_{j \in J}\{(\overline{x}_j - \underline{x}_j)/T\}$. Our discretized version of the type space X is given by $X_T := \prod_{j \in J} X_T(j)$. Furthermore, I define a probability density function on X_T by setting $\hat{f}(v) = f(v)/(\sum_{t \in X_T} f(t))$. I thus obtain a linear program which is a finite-dimensional approximation of OPT_* for each $T > 0$ by replacing X with X_T.

Belloni et al. (2010) use a plane-putting algorithm which works with a randomly chosen subset of incentive-compatibility (ICC) and Border (B) constraints at each iteration. They provide an efficient reduction in the growth in T of the Border constraints (B) from $O(2^{T^J})$ to $O(T^J \log(T^J))$ (Belloni, Lopomo, and Wang 2010, Lemma 10). We adopt their solution to checking (B) constraints; however, our approach to checking (ICC) constraints involves iteratively growing the 'local' region of the type space around each point v in the discretized set of types V_T. We do two things. First, all the immediately adjacent points in the discretized type space are always checked for incentive compatibility. Secondly, downwards-sloping points in the discretized type space are also checked. Furthermore, the downwards-sloping region of the type space grows until all (ICC) constraints are ultimately satisfied. This procedure is illustrated visually in Figure 17. Thus, for a fixed-size local region around each point in the discretized type space, we first satisfy *local* (ICC) and (B) constraints as in the iterative plane-cutting algorithm of (Belloni, Lopomo, and Wang 2010). Then we run the separation oracle with *all* (ICC) and (B) constraints. We then restart the solver with any previously violated constraints, this time increasing the size of the local region around each point in the discretized type space. This procedure iterates until no constraints are violated. This modified version of (Belloni, Lopomo, and Wang 2010)'s algorithm is described in Algorithm 1.

Our algorithm[1] is written in Python 3.10 and uses Google's open source linear programming solver 'GLOP' available in their `or-tools` package (Perron and Furnon 2023).

7.2 Census Non-Response Calculations

The 2020 census post-enumeration survey data can be found in the US Census data tables[2], where the 'Net Coverage Error for the Household Population in the United States by Race and Hispanic Origin' is given

[1] For more details see: `https://github.com/jmemich/optimal-auction-multidim`
[2] `https://data.census.gov/table`

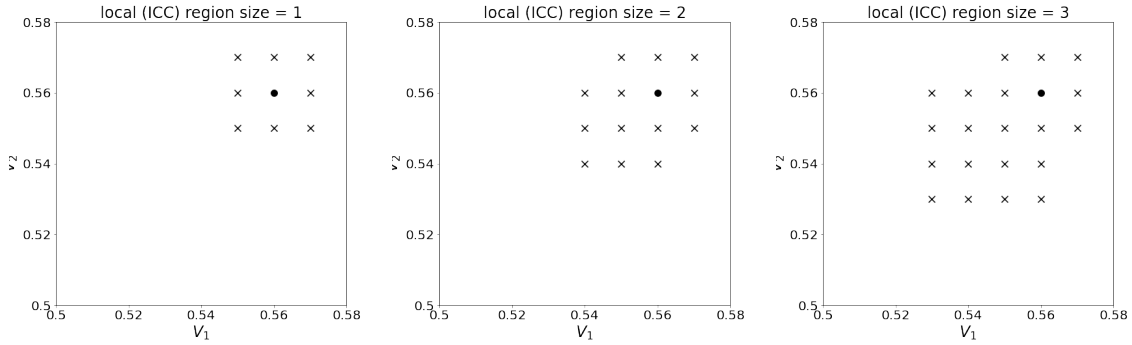

FIGURE 17: We iteratively grow the local region of the discretized type space checked for downwards-sloping constraint violations. Notice that immediately adjacent (ICC) constraints are always checked (×) when the local region increases in size around a point (•).

by the variable **C_RACEHISUS** and the net coverage error is estimated at -4.99%. The data for the 2010 US Census are not available on the census data tables, however, the official estimated net undercount of Hispanics was -1.54% (Patrick Cantwell 2012, p1).

$L = 1, S = \emptyset, A = \emptyset, \overline{OPT} = \infty$;
`violated_any_icc` $\leftarrow$ TRUE;
while *violated_any_icc* **do**
  `violated_local_icc` $\leftarrow$ TRUE;
  while *violated_local_icc* **do**
    $k = 1, A^k = A, S^k = S$;
    Solve the linear program associated with S^k. Let OPT^k denote the optimal value.;
    Solve the separation oracle using only local (ICC) constraints in region L. Let A^k denote all violated local (ICC) and (B) constraints.;
    if $A^k = \emptyset$ **then**
      `violated_local_icc` $\leftarrow$ FALSE;
      Break;
    end
    Select a subset $I^k \subset S^k$ of inactive (ICC) and (B) constraints;
    if $OPT^k < \overline{OPT}$ **then**
      $S^{k+1} \leftarrow (S^k \setminus I^k) \cup A^k \cup A$;
      $\overline{OPT} \leftarrow OPT^k$;
    else
      $S^{k+1} \leftarrow S^k \cup A^k \cup A$;
    end
    $k \leftarrow k + 1$;
  end
  Solve the separation oracle using all (ICC) constraints. Let A^* denote all violated (ICC) constraints.;
  if $A^* = \emptyset$ **then**
    `violated_any_icc` $\leftarrow$ FALSE;
    Break;
  end
  $A \leftarrow A \cup A^*$;
  $S \leftarrow S^k$;
  $L \leftarrow L + 1$;
end

Algorithm 1: Iterative plane-cutting algorithm with local and downwards-sloping (ICC) constraints