Volume Title: The Economics of Artificial Intelligence: An Agenda

Volume Authors/Editors: Ajay Agrawal, Joshua Gans, and Avi Goldfarb, editors

Volume Publisher: University of Chicago Press

Volume ISBNs: 978-0-226-61333-8 (cloth); 978-0-226-61347-5 (electronic)

Volume URL: http://www.nber.org/books/agra-1

Conference Date: September 13–14, 2017

Publication Date: May 2019

Chapter Title: How Artificial Intelligence and Machine Learning Can Impact Market Design

Chapter Author(s): Paul R. Milgrom, Steven Tadelis

Chapter URL: http://www.nber.org/chapters/c14008

Chapter pages in book: (p. 567 – 585)

23

How Artificial Intelligence and Machine Learning Can Impact Market Design

Paul R. Milgrom and Steven Tadelis

23.1 Introduction

For millennia, markets have played a key role in providing individuals and businesses with the opportunity to gain from trade. More often than not, markets require structure and a variety of intuitional support to operate efficiently. For example, auctions have become a commonly used mechanism to generate gains from trade when price discovery is essential. Research in the area now commonly referred to as market design, going back to Vickrey (1961), demonstrated that it is critical to design auctions and market institutions more broadly in order to achieve efficient outcomes (see, e.g., Milgrom 2017; Roth 2015).

Any market designer needs to understand some fundamental details of the transactions that are expected to be consummated in order to design the most effective and efficient market structure to support these transactions. For example, the National Resident Matching Program, which matches doctors to hospital residencies, was originally designed in an era when nearly all doctors were men and wives followed them to their residencies. It needed to be redesigned in the 1990s to accommodate the needs of couples, when men and women doctors could no longer be assigned jobs in different cities. Even

Paul R. Milgrom is the Shirley and Leonard Ely Professor of Humanities and Sciences in the Department of Economics at Stanford University and professor, by courtesy, at both the Department of Management Science and Engineering and the Graduate School of Business. Steven Tadelis holds the James J. and Marianne B. Lowrey Chair in Business and is professor of economics, business, and public policy at the University of California, Berkeley, and a research associate of the National Bureau of Economic Research.

For acknowledgments, sources of research support, and disclosure of the authors' material financial relationships, if any, please see http://www.nber.org/chapters/c14008.ack.

something as mundane as the sale of a farm when a farmer dies requires knowledge of the structure and decisions about whether to sell the whole farm as a unit, or to separate the house for sale as a weekend retreat while selling the land to neighboring farmers, or selling the forest separately to a wildlife preservation fund.

In complex environments, it can be difficult to understand the underlying characteristics of transactions, and it is challenging to learn enough about them in order to design the best institutions to efficiently generate gains from trade. For example, consider the recent growth of online advertising exchanges that match advertisers with online ads. Many ads are allocated to advertisers using real-time auctions. But how should publishers design these auctions in order to make the best use of their advertising space, and how can they maximize the returns to their activities? Based on the early theoretical auction design work of Myerson (1981), Ostrovsky and Schwartz (2017) have shown that a little bit of market design in the form of setting better reserve prices can have a dramatic impact on the profits an online ad platform can earn.

But how can market designers learn the characteristics necessary to set optimal, or at least better, reserve prices? Or, more generally, how can market designers better learn the environment of their markets? In response to these challenges, artificial intelligence (AI) and machine learning are emerging as important tools for market design. Retailers and marketplaces such as eBay, TaoBao, Amazon, Uber, and many others are mining their vast amounts of data to identify patterns that help them create better experiences for their customers and increase the efficiency of their markets. By having better prediction tools, these and other companies can predict and better manage sophisticated and dynamic market environments. The improved forecasting that AI and machine-learning algorithms provide help marketplaces and retailers better anticipate consumer demand and producer supply as well as help target products and activities to finer segmented markets.

Turning back to markets for online advertising, two-sided markets such as Google, which match advertisers with consumers, are not only using AI to set reserve prices and segment consumers into finer categories for ad targeting, but they also develop AI-based tools to help advertisers bid on ads. In April 2017 Google introduced "Smart Bidding," a product based on AI and machine learning that helps advertisers bid automatically on ads based on ad conversions so they can better determine their optimal bids. Google explained that the algorithms use vast amounts of data and continually refine models of users' conversion to better spend an advertiser's dollars to where they bring in the highest conversion.

Another important application of AI's strength in improving forecasting to help markets operate more efficiently is in electricity markets. To operate efficiently, electricity market makers such as California's Independent System Operator must engage in demand and supply forecasting. An inaccurate forecast in the power grid can dramatically affect market outcomes causing high variance in prices, or worse, blackouts. By better predicting demand and supply, market makers can better allocate power generation to the most efficient power sources and maintain a more stable market.

As the examples above demonstrate, the applications of AI algorithms to market design are already widespread and diverse. Given the infancy of the technology, it is a safe bet that AI will play a growing role in the design and implementation of markets over a wide range of applications. In what follows, we describe several less obvious ways in which AI has played a key role in the operation of markets.

23.2 Machine Learning and the Incentive Auction

In the first part of the twentieth century, the most important infrastructure projects for the United States related to transportation and energy infrastructure. By the early twenty-first century, however, it was not just people and goods that needed to be transported in high volumes, but also information. The emergence of mobile devices, WiFi networks, video on demand, the Internet of Things, services supplied through the cloud, and much more has already created the need for major investments in the communication network, and with 5G technologies just around the corner, more is coming.

Wireless communications, however, depend on infrastructure and other resources. The wireless communication rate depends on the channel capacity, which in turn depends jointly on the communication technology used and the amount of radio spectrum bandwidth devoted to it. To encourage growth in bandwidth and the rapid develop of new uses, the Obama White House in 2010 issued its National Broadband Plan. That plan set a goal of freeing a huge amount of bandwidth from older, less productive uses to be used instead as part of the modern data highway system.

In 2016–2017, the US Federal Communications Commission (FCC) designed and ran an auction market to do part of that job. The radio spectrum licenses that it sold in that auction raised about $20 billion in gross revenue. As part of the process of making room for those new licenses, the FCC purchased TV broadcast rights for about $10 billion, and incurred nearly $3 billion in costs to move other broadcasters to new TV channels. Some 84MHz of spectrum was made available in total, including 70MHz for wireless broadband and 14MHz for unlicensed uses. This section describes the processes that were used, and the role of AI and machine learning to improve the underlying algorithms that supported this market.

Reallocating spectrum from one use to another is, in general, neither easy nor straightforward, in either the planning or the implementation (Leyton-Brown, Milgrom, and Segal 2017). Planning such a change can involve surprisingly hard computational challenges, and the implementation requires high levels of coordination. In particular, the reallocation of a portion of the spectrum band that had been used for UHF broadcast television required deciding how many channels to clear, which stations would cease broadcasting (to make room for the new uses), what TV channels would be assigned to the remaining stations that continued to broadcast, how to time the changes to avoid interference during the transition, and to assure that the TV tower teams, which would replace the old broadcast equipment, had sufficient capacity, and so on. Several of the computations involved are, in principle, nondeterministic polynomial time (NP)-hard, making this a particularly complex market-design problem. One of the most critical algorithms used for this process—the "feasibility checker"—was developed with the aid of machine-learning methods.

But why reallocate and reassign TV stations at all? Broadcast television changed enormously in the late twentieth century. In the early days of television, all viewing was of over-the-air broadcasts using an analog technology. Over the decades that followed, cable and satellite services expanded so much that, by 2010, more than 90 percent of the US population was reached by these alternative services. Standard definition TV signals were replaced by high definition and, eventually, 4K signals. Digital television and tuners reduced the importance of channel assignments, so that the channel used by consumers/viewers did not need to match the channel used by the broadcaster. Digital encoding made more efficient use of the band and it became possible to use multiplexing, so that what was once a single standard-definition broadcast channel could carry multiple high-definition broadcasts. Marginal spectrum had fallen in value compared to the alternative uses.

Still, the reallocation from television broadcasting would be daunting and beyond what an ordinary market mechanism could likely achieve. The signal from each of thousands of TV broadcast towers across the United States can interfere with potential uses for about 200 miles in every direction, so all of the broadcasts in any frequency needed to be cleared to make the frequencies available for new uses. Not only would it be necessary to coordinate among different areas of the United States, but coordination with Canada and Mexico would improve the allocation, too; most of the population of Canada lives, and most of its TV stations operate, within 200 miles of the US border. Because a frequency is not usable until virtually all of the relevant broadcasters have ceased operation, efficiency would demand that these changes would need to be coordinated in time, too; they should be roughly simultaneous. In addition, there needed to be coordination *across* frequencies. The reason is that we need to know in advance which channels will be cleared before the frequencies can be efficiently divided between uplink uses and downlink uses.

Among the many issues to be resolved, one would be how to determine which stations would continue to broadcast after the transition. If the goal were efficiency, then the problem can be formulated as maximizing the total value of the TV stations that continue to broadcast after the auction. Let N

be the set of all currently broadcasting TV stations and let $S \subseteq N$ be a subset of those TV stations. Let $\mathcal{C}$ be the set of available channels to which to assign stations after the auction, and let $\varnothing$ denote the null assignment for a station that does not continue to broadcast. A channel assignment is a mapping $A : N \rightarrow \mathcal{C} \cup \{\varnothing\}$. The constraints on the channels available for assignment are to ones that rule out interference between pairs of TV stations, taking the form: $A(n_1) = c_1 \Rightarrow A(n_2) \neq c_2$ for some $(c_1,c_2) \in \mathcal{C}^2$. Each such constraint is described by a fourtuple: (n_1,c_1n_2,c_2). There were more than a million such constraints in the FCC's problem. A channel assignment is feasible if it satisfies all the interference constraints; let $\mathcal{A}$ denote the feasible set of assignments. A set of stations S' can be feasibly assigned to continue broadcasting, which we denote by $S' \in \mathcal{F}(\mathcal{C})$, if there exists some feasible channel assignment $\mathrm{A} \in \mathcal{A}$ such that $\varnothing \notin A(S')$.

Most of the interference constraints took a special form. Those constraints assert that no two stations that are geographic neighbors can be assigned to the same channel. Let us call such stations "linked" and denote the relationship by $(n_1,n_2) \in L$. For such a pair of stations, the constraint can be written as: $A(n_1) = A(n_2) \Rightarrow A(n_1) = \varnothing$. These are the *cochannel interference constraints*. One can think of (N,L) as defining a graph with nodes N and arcs L. If the cochannel constraints were the only ones, then determining whether $S' \in \mathcal{F}$ would amount to deciding whether there exists a way to assign channels in $\mathcal{C}$ to the stations in N so that no two linked nodes are on the same channel.

Figure 23.1 shows the graph of the cochannel interference constraints for the United States and Canada. The constraint graph is most dense in the eastern half of the United States and along the Pacific Coast.

In the special case of cochannel constraints, the problem of checking the feasibility of a set of stations is a standard *graph-coloring* problem. The problem is to decide whether it is possible to assign a color (channel) to each node (station) in the graph so that no two linked nodes are given the same color. Graph-coloring is in the class of NP-complete problems, for which there is no known algorithm that is guaranteed to be fast, and for which it is commonly hypothesized[1] that worst-case solution time grows exponentially in the problem size. Since the general station assignment problem includes the graph coloring problem, it, too, is NP-complete, and can become intractable at scales such as that of the FCC's problem.

The problem that the FCC would ideally like to solve using an auction is to maximize the value of the stations that remain on-air to broadcast, given the reduced set of channels $\mathcal{C}$. If the value of station j is v_j, the problem can be formulated as follows:

$$\max_{S \in \mathcal{F}(\mathcal{C})} \sum_{j \in S} v_j.$$

1. The standard computer science hypothesis that $P \neq NP$ implies that no fast algorithm exists for NP-complete problems.

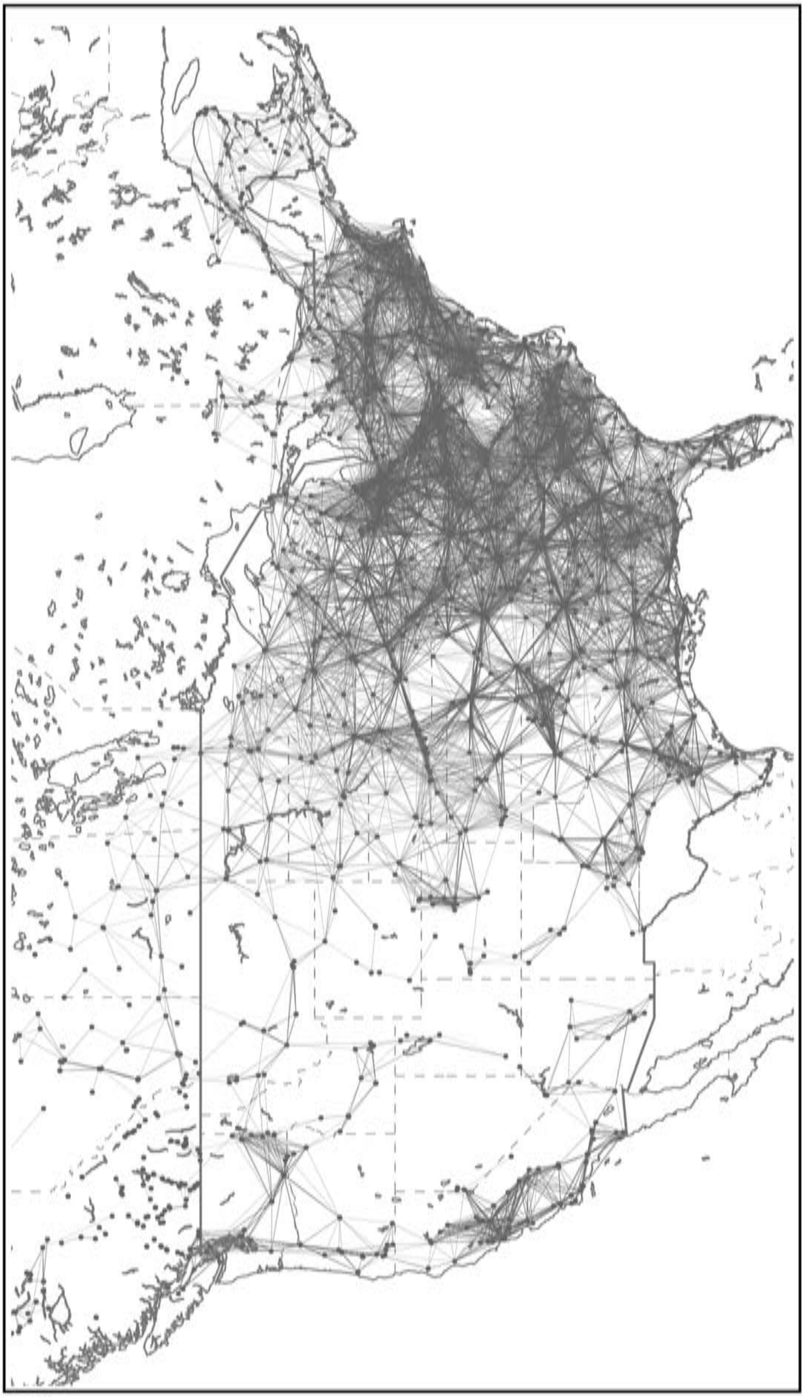

Fig. 23.1 Cochannel interference graph for spectrum reallocation

This problem is very hard. Indeed, as we have just argued, even checking the condition $S \in \mathcal{F}(\mathcal{C})$ is NP-complete, and solving exactly the related optimization is even harder in practice. Computational experiments suggest that with weeks of computation approximate optimization is possible, but with an optimization shortfall that can be a few percent.

For a TV station owner, it would be daunting to formulate a bid in an auction in which even the auctioneer, with all the bids in hand, would find it challenging to determine the winners. Faced with such a problem, some station owners might choose not to participate. That concern led the FCC staff to prefer a strategy-proof design, in which the optimal bid for the owner of a single station is relatively simple, at least in concept: compute your station's value and bid that amount. As is well known, there is a unique strategy-proof auction that optimizes the allocation and pays zero to the losers: the Vickrey auction. According to the Vickrey rules, if the auctioneer purchases the broadcast rights from station j, it must pay the owner this price:

$$p_i = \left(\max_{S \in \mathcal{F}(C)} \sum_{j \in S} v_j \right) - \left(\max_{\substack{S \in \mathcal{F}(C) \\ i \notin S}} \sum_{j \in S} v_j \right).$$

For a winning station i, the Vickrey price p_i will be larger than the station value. With roughly 2,000 stations to include in the optimization, a 1 percent error in either of the two maximizations would result in a pricing error for p_i equal to about 2,000 percent of the value of an average station. Such huge potential pricing errors would likely raise hackles among some of the potential bidders.

One way to put the problem of the Vickrey auction into sharp relief is to imagine the letter that the FCC might write to broadcasters to encourage their participation:

> Dear Mr. Broadcaster:
>
> We have heard your concerns about the complexity of the spectrum reallocation process. You may even be unsure about whether to participate or how much to bid. To make things as easy as possible for you, we have adopted a Nobel Prize–winning auction procedure called the "Vickrey auction." In this auction, all you need to do is to tell us what your broadcast rights are worth to you. We'll figure out whether you are a winner and, if so, how much to pay to buy your rights. The rules will ensure that it is in your interest to report truthfully. That is the magic of the Vickrey auction!
>
> The computations that we do will be very hard ones, and we cannot guarantee that they will be exactly correct.

Such a letter would leave many stations owners uncomfortable and unsure about whether to participate. The FCC decided to adopt a different design.

What we describe here is a simplified version of the design, in which the broadcasters' only choices are whether to sell their rights or to reject the

FCC's offer and continue to broadcast. Each individual broadcaster was comforted by the assurance that it could bid this way, even if it had additional options, too.[2]

In the simplified auction, each bidder i was quoted a price $p_i(t)$ at each round t of the auction that decreased from round-to-round. In each round, the bidder could "exit," rejecting the current price and keeping its broadcast rights, or it could accept the current price. After a round t of bidding, stations were processed one at a time. When station i was processed, the auction software would use its *feasibility checker* to attempt to determine whether it could feasibly assign station i to continue broadcasting, given the other stations that had already exited and to which a channel must be assigned. This is the generalized graph-coloring problem, mentioned earlier. If the software timed out, or if it determined that it is impossible to assign the station, then the station would become a winner and be paid $p_i(t-1)$. Otherwise, its price would be reduced to $p_i(t)$ and it would exit or continue, according to the bidder's instructions. It would be obvious to a station owner that, regardless of the pricing formula and of how the software performed, its optimal choice when its value is v_i is to exit if $p_i(t) < v_i$ and otherwise to continue.[3]

The theory of clock auctions of this sort for problems with hard computations has been developed by Milgrom and Segal (2017), who also report simulations showing high performance in terms of efficiency and remarkably low costs of procuring TV broadcast rights.

The performance of this descending auction design depends deeply on the quality of the feasibility checker. Based on early simulations, our rough estimate was the each 1 percent of failures in feasibility checking would add about 1.5 percent—or about \$150 million—to the cost of procuring the broadcast rights. So, solving most of the problems very fast became a high priority for the auction-design team.

As a theoretical proposition, any known algorithm for feasibility checking in the spectrum-packing problem has worst-case performance that grows exponentially in the size of the problem. Nevertheless, if we know the distribution of likely problems, there can still be algorithms that are fast with

2. In the actual auction, some broadcasters also had the option to switch from a UHF TV channel to a channel in the high VHF band, or one in the low VHF band (the so-called HVHF and LVHF options).

3. The pricing formula that the FCC used for each station was $p_i(t) = (\text{Pop}_i\text{Links}_i)^{0.5}q(t)$. In this formula, $q(t)$ is the "base clock price" that scaled the price offers to all the bidders. This price began at a high level $q(0)$ to encourage participation, and it declined round-by-round during the auction; Pop_i denotes the population of the area served by the station, which stands in for the value of the station. By linking prices to population served, the auctioneer is able to offer higher prices to valuable stations in high-population areas that it might need to acquire for a successful auction; Links_i measured the number of other stations to which station i was linked in the interference graph. It was hoped that, by including this term in the pricing formula, the auction would be able to offer higher prices to and buy the rights of stations that pose particularly difficult problems by interfering with many other stations.

high probability. But how can we know the distribution and how can such an algorithm be found?

The FCC auction used a feasibility checker developed by a team of Auctionomics researchers at the University of British Columbia, led by Professor Kevin Leyton-Brown. There were many steps in the development, as reported by Newman, Fréchette, and Leyton-Brown (forthcoming), but here we emphasize the role of machine learning. Auctionomics' goal was to be able to solve 99 percent of the problem instances in one minute or less.

The development effort began by simulating the planned auction to generate feasibility problems like those that might be encountered in a real auction. Running many simulations generated about 1.4 million problem instances that could be used for training and testing a feasibility-checking algorithm. The first step of the analysis was to formulate the problem as mixed integer programs and test standard commercial software—CPLEX and Gurobi—to see how close those could come to meeting the performance objectives. The answer was: not close. Using a 100-seconds cutoff, Gurobi could solve only about 10 percent of the problems and CPLEX only about 25 percent. These were not nearly good enough for decent performance in a real-time auction.

Next, the same problems were formulated as satisfiability problems and tested using seventeen research solvers that had participated in recent SAT-solving tournaments. These were better, but none could solve as many as two-thirds of the problems within the same 100-second cutoff. The goal remained 99 percent in sixty seconds.

The next step was to use automated algorithm configuration, a procedure developed by Hutter, Hoos, and Leyton-Brown (2011) and applied in this setting by Leyton-Brown and his students at the University of British Columbia. The idea is to start with a highly parameterized algorithm for solving satisfiability problems[4] and to train a random forest model of the algorithm performance, given the parameters. To do that, we first ran simulated auctions with what we regarded as plausible behavior by the bidders to generate a large data set of representative problems. Then, we solved those problems using a variety of different parameter settings to determine the distribution of solution times for each vector of parameters. This generated a data set with parameters and performance measures. Two of the most interesting performance characteristics were the median run time and

4. There are no known algorithms for NP-complete problems that are guaranteed to be fast, so the best existing algorithms are all heuristics. These algorithms weight various characteristics of the problem to decide about such things as the order in which to check different branches of a search tree. These weights are among the parameters that can be set and adapted to work well for a particular class of problems, such as those that arise in the incentive auction application. The particular software algorithm that we used was CLASP, which had more than 100 exposed parameters that could be modified.

the fraction of instances solved within one minute. Then, using a Bayesian model, we incorporated uncertainty in which the experimenter "believes" that the actual performance is normally distributed with a mean determined by the random forest and a variance that depends on the distance of the parameter vector from the nearest points in the data set. Next, the system identifies the parameter vector that maximizes the expected improvement in performance, given the mean and variance of the prior and the performance of the best-known parameter vector. Finally, the system tests the actual performance for the identified parameters and adds that as an observation to the data set. Proceeding iteratively, the system identifies more parameters to test, investigates them, and adds them to the data to improve the model accuracy until the time budget is exhausted.

Eventually, this machine-learning method leads to diminishing returns to time invested. One can then create a new data set from the instances on which the parameterized algorithm was "slow," for example, taking more than fifteen seconds to solve. By training a new algorithm on those instances, and running the two parameterized algorithms in parallel, the machine-learning techniques led to dramatic improvements in performance.

For the actual auction, several other problem-specific tricks were also applied to contribute to the speed-up. For example, to some extent it proved possible to decompose the full problem into smaller problems, to reuse old solutions as starting points for a search, to store partial solutions that might help guide solutions of further problems, and so on. In the end, the full set of techniques and tricks resulted in a very fast feasibility checker that solved all but a tiny fraction of the relevant problems within the allotted time.

23.3 Using AI to Promote Trust in Online Marketplaces

Online marketplaces such as eBay, Taobao, Airbnb, and many others have grown dramatically since their inception just over two decades ago, providing businesses and individuals with previously unavailable opportunities to purchase or profit from online trading. Wholesalers and retailers can market their goods or get rid of excess inventory; consumers can easily search marketplaces for whatever is on their mind, alleviating the need for businesses to invest in their own e-commerce website; individuals transform items they no longer use into cash; and more recently, the so called "gig economy" is comprised of marketplaces that allow individuals to share their time or assets across different productive activities and earn extra income.

The amazing success of online marketplaces was not fully anticipated, primarily because of the hazards of anonymous trade and asymmetric information. Namely, how can strangers who have never transacted with one another, and who may be thousands of miles apart, be willing to trust each other? Trust on both sides of the market is essential for parties to be willing to transact and for a marketplace to succeed. The early success of eBay is often attributed to the innovation of introducing its famous feedback and reputation mechanism, which was adopted in one form or another by practically every other marketplace that came after eBay. These online feedback and reputation mechanisms provide a modern-day version of more ancient reputation mechanisms used in the physical marketplaces that were the medieval trade fairs of Europe (see Milgrom, North, and Weingast 1990).

Still, recent studies have shown that online reputation measures of marketplace sellers, which are based on buyer-generated feedback, don't accurately reflect their actual performance. Indeed, a growing literature has shown that user-generated feedback mechanisms are often biased, suffer from "grade inflation," and can be prone to manipulation by sellers.[5] For example, the average percent positive for sellers on eBay is about 99.4 percent, with a median of 100 percent. This causes a challenge to interpret the true levels of satisfaction on online marketplaces.

A natural question emerges: Can online marketplaces use the treasure trove of data it collects to measure the quality of a transaction and predict which sellers will provide a better service to their buyers? It has become widely known that all online marketplaces, as well as other web-based services, collect vast amounts of data as part of the process of trade. Some refer to this as the "exhausts data" generated by the millions of transactions, searches, and browsing that occur on these marketplaces daily. By leveraging this data, marketplaces can create an environment that would promote trust, not unlike the ways in which institutions emerged in the medieval trade fairs of Europe that helped foster trust. The scope for market design goes far beyond the more mainstream application like setting rules of bidding and reserve prices for auctions or designing tiers of services, and in our view, includes the design of mechanisms that help foster trust in marketplaces. What follows are two examples from recent research that show some of the many ways that marketplaces can apply AI to the data they generate to help create more trust and better experiences for their customers.

23.3.1 Using AI to Assess the Quality of Sellers

One of the ways that online marketplaces help participants build trust is by letting them communicate through online messaging platforms. For example, on eBay buyers can contact sellers to ask them questions about their products, which may be particularly useful for used or unique products for which buyers may want to get more refined information than is listed. Similarly, Airbnb allows potential renters to send messages to hosts and ask questions about the property that may not be answered in the original listing.

Using Natural Language Processing (NLP), a mature area in AI, market-

5. On bias and grade inflation see, for example, Nosko and Tadelis (2015), Zervas, Proserpio, and Byers (2015), and Filippas, Horton, and Golden (2017). On seller manipulation of feedback scores see, for example, Mayzlin, Dover, and Chevalier (2014) and Xu et al. (2015).

places can mine the data generated by these messages in order to better predict the kind of features that customers value. However, there may also be subtler ways to apply AI to manage the quality of marketplaces. The messaging platforms are not restricted to pretransaction inquiries, but also offer the parties to send messages to each other *after* the transaction has been completed. An obvious question then emerges: How could a marketplace analyze the messages sent between buyers and sellers post the transaction to infer something about the quality of the transaction that feedback doesn't seem to capture?

This question was posed and answered in a recent paper by Masterov, Mayer, and Tadelis (2015) using internal data from eBay's marketplace. The analysis they performed was divided into two stages. In the first stage, the goal was to see if NLP can identify transactions that went bad when there was an independent indication that the buyer was unhappy. To do this, they collected internal data from transactions in which messages were sent from the buyer to the seller after the transaction was completed, and matched it with another internal data source that recorded actions by buyers indicating that the buyer had a poor experience with the transactions. Actions that indicate an unhappy buyer include a buyer claiming that the item was not received, or that the item was significantly not as described, or leaves negative or neutral feedback, to name a few.

The simple NLP approach they use creates a "poor-experience" indicator as the target (dependent variable) that the machine-learning model will try to predict, and uses the messages' content as the independent variables. In its simplest form and as a proof of concept, a regular expression search was used that included a standard list of negative words such as "annoyed," "dissatisfied," "damaged," or "negative feedback" to identify a message as negative. If none of the designated terms appeared, then the message was considered neutral. Using this classification, they grouped transactions into three distinct types: (a) no posttransaction messages from buyer to seller, (b) one or more negative messages, or (c) one or more neutral messages with no negative messages.

Figure 23.2, which appears in Masterov, Mayer, and Tadelis (2015), describes the distribution of transactions with the different message classifications together with their association with poor experiences. The x-axis of figure 23.1 shows that approximately 85 percent of transactions fall into the benign first category of no posttransaction messages. Buyers sent at least one message in the remaining 15 percent of all transactions, evenly split between negative and neutral messages. The top of the y-axis shows the poor experience rate for each message type. When no messages are exchanged, only 4 percent of buyers report a poor experience. Whenever a neutral message is sent, the rate of poor experiences jumps to 13 percent, and if the message's content was negative, over one-third of buyers express a poor experience.

In the second stage of the analysis, Masterov, Mayer, and Tadelis (2015)

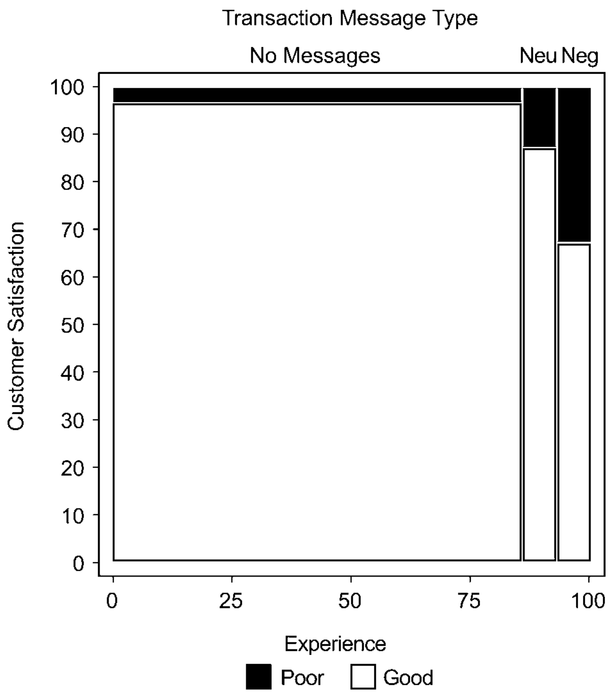

Fig. 23.2 Message content and poor experiences on eBay

Source: Masterov et al. 2015. ©2015 Association for Computing Machinery, Inc. Reprinted by permission. https://doi.org/10.1145/2764468.2764499.

used the fact that negative messages are associated with poor experiences to construct a novel measure of seller quality based on the idea that sellers who receive a higher frequency of negative messages are worse sellers. For example, imagine that seller A and seller B both sold 100 items and that seller A had five transactions with at least one negative message, while seller B had eight such transactions. The implied quality score of seller A is then 0.05 while that of seller B is 0.08, and the premise is that seller B is a worse seller than seller A. Masterov, Mayer, and Tadelis (2015) show that the relationship between this ratio, which is calculated for every seller at any point in time using aggregated negative messages from past sales, and the likelihood that a current transaction will result in a poor experience, is monotonically increasing.

This simple exercise is a proof of concept that shows that by using the message data and a simple natural language processing AI procedure, they were able to better predict which sellers will create poor experiences than one can infer from the very inflated feedback data. eBay is not unique in allowing the parties to exchange messages and the lessons from this research are easily generalizable to other marketplaces. The key is that there is information in communication between market participants, and past communication can help identify and predict the sellers or products that will cause buyers poor experiences and negatively impact the overall trust in the marketplace.

23.2.2 Using AI to Create a Market for Feedback

Aside from the fact that feedback is often inflated as described earlier, another problem with feedback is that not all buyers choose not to leave feedback at all. In fact, through the lens of mainstream economic theory, it is surprising that a significant fraction of online consumers leave feedback. After all, it is a selfless act that requires time, and it creates a classic free-rider problem. Furthermore, because potential buyers are attracted to buy from sellers or products that already have an established good track record, this creates a "cold-start" problem: new sellers (or products) with no feedback will face a barrier-to-entry in that buyers will be hesitant to give them a fair shot. How could we solve these free-rider and cold-start problems?

These questions were analyzed in a recent paper by Li, Tadelis, and Zhow (2016) using a unique and novel implementation of a market for feedback on the huge Chinese marketplace Taobao where they let sellers pay buyers to leave them feedback. Naturally, one may be concerned about allowing sellers to pay for feedback as it seems like a practice in which they will only pay for good feedback and suppress any bad feedback, which would not add any value in promoting trust. However, Taobao implemented a clever use of NLP to solve this problem: it is the platform, using an NLP AI model, that decides whether feedback is relevant and not the seller who pays for the feedback. Hence, the reward to the buyer for leaving feedback was actually managed by the marketplace, and was handed out for informative feedback rather than for positive feedback.

Specifically, in March 2012, Taobao launched a "Rebate-for-Feedback" (RFF) feature through which sellers can set a rebate value for any item they sell (cash back or store coupon) as a reward for a buyer's feedback. If a seller chooses this option, then Taobao guarantees that the rebate will be transferred from the seller's account to a buyer who leaves high-quality feedback. Importantly, feedback quality only depends on how informative it is, rather than whether the feedback is positive or negative. Taobao measures the quality of feedback with a NLP algorithm that examines the comment's content and length and finds out whether key features of the item are mentioned. Hence, the marketplace manages the market for feedback by forcing the seller to deposit at Taobao a certain amount for a chosen period, so that funds are guaranteed for buyers who meet the rebate criterion, which itself is determined by Taobao.[6]

6. According to a Taobao survey (published in March 2012), 64.8 percent of buyers believed that they will be more willing to buy items that have the RFF feature, and 84.2 percent of buyers believed that the RFF option will make them more likely to write detailed comments.

Taobao's motivation behind the RFF mechanism was to promote more informative feedback, but as Li, Tadelis, and Zhow (2016) noted, economic theory offers some insights into how the RFF feature can act as a potent signaling mechanism that will further separate higher- from lower-quality sellers and products. To see this, recall the literature launched by Nelson (1970) who suggested that advertising acts as a signal of quality. According to the theory, advertising—which is a form of burning money—acts as a signal that attracts buyers who correctly believe that only high-quality sellers will choose to advertise. Incentive compatibility is achieved through repeat purchases: buyers who purchase and experience the products of advertisers will return in the future only if the goods sold are of high enough quality. The cost of advertising can be high enough to deter low-quality sellers from being willing to spend the money and sell only once because those sellers will not attract repeat customers, and still low enough to leave profits for higher-quality sellers. Hence, ads act as signals that separate high-quality sellers, and in turn attract buyers to their products.

As Li, Tadelis, and Zhow (2016) argue, the RFF mechanism plays a similar signaling role as ads do. Assuming that consumers express their experiences truthfully in written feedback, any consumer who buys a product and is given incentives to leave feedback will leave positive feedback only if the buying experience was satisfactory. Hence, a seller will offer RFF incentives to buyers only if the seller expects to receive positive feedback, and this will happen only if the seller will provide high quality. If a seller knows that their goods and services are unsatisfactory, then paying for feedback will generate negative feedback that will harm the low-quality seller. Equilibrium behavior then implies that RFF, as a signal of high quality, will attract more buyers and result in more sales. The role of AI was precisely to reward buyers for information, not for positive feedback.

Li, Tadelis, and Zhou (2016) proceeded to analyze data from the period where the RFF mechanism was featured and confirmed that first, as expected, more feedback was left in response to the incentives provided by the RFF feature. More important, the additional feedback did not exhibit any biases, suggesting that the NLP algorithms used were able to create the kind of screening needed to select informative feedback. Also, the predictions of the simple signaling story were borne out in the data, suggesting that using NLP to support a novel market for feedback did indeed solve both the free-rider problem and the cold-start problem that can hamper the growth of online marketplaces.

23.4 Using AI to Reduce Search Frictions

An important application of AI and machine learning in online marketplaces is the way in which potential buyers engage with the site and proceed to search for products or services. Search engines that power the search of products online are based on a variety of AI algorithms that are trained to maximize what the provider believes to be the right objective. Often this boils down to conversion, under the belief that the sooner a consumer converts a search to a purchase, the happier the consumer is both in the short and the long run. The rationale is simply that search itself is a friction, and hence, maximizing the successful conversion of search activity to a purchase reduces this friction.

This is not inconsistent with economic theory that has modeled search as an inevitable costly process that separates consumers from the products they want. The canonical search models in economics either build on the seminal work of Stigler (1961), who assumes that consumers sample a fixed number of stores and choose to buy the lowest priced item, or more often, on the models of McCall (1970) and Mortensen (1970), who posit that a model of sequential search is a better description of consumer search behavior. In both modeling approaches consumers know exactly what they wish to buy.

However, it turns out that unlike the simplistic models of search employed in economic theory, where consumers know what they are looking for and the activity of search is just a costly friction, in reality, people's search behavior is rich and varied. A recent paper by Blake, Nosko, and Tadelis (2016) uses comprehensive data from eBay to shed light on the search process with minimal modeling assumptions. Their data show that consumers search significantly more than other studies—which had limited access to search behavior over time—have suggested.

Furthermore, search often proceeds from the vague to the specific. For example, early in a search a user may use the query "watch," then refine it to "men's watch," and later add further qualifying words such as color, shape, strap type, and more. This suggests that consumers often learn about their own tastes, and what product characteristics exist, as part of the search process. Indeed, Blake et al. (2016) show that the average number of terms in the query rises over time, and the propensity to use the default-ranking algorithm declines over time as users move to more focused searches like price sorting.

These observations suggest that marketplaces and retailers alike could design their online search algorithms to understand search intent so as to better serve their consumers. If a consumer is in the earlier, exploratory phases of the search process, then offering some breadth will help the consumer better learn their tastes and the options available in the market. But when the consumer is driven to purchase something particular, offering a narrower set of products that match the consumer's preferences would be better. Hence, machine learning and AI can play an instrumental role in recognizing customer intent.

Artificial intelligence and machine learning cannot only help predict a customer's intent, but given the large heterogeneity on consumer tastes, AI can help a marketplace or retailer better segment the many customers into groups that can be better served with tailored information. Of course, the idea of using AI for more refined customer segmentation, or even personalized experiences, also raises concerns about price discrimination. For example, in 2012 the *Wall Street Journal* reported that "Orbitz Worldwide Inc. has found that people who use . . . Mac computers spend as much as 30% more a night on hotels, so the online travel agency is starting to show them different, and sometimes costlier, travel options than Windows visitors see. The Orbitz effort, which is in its early stages, demonstrates how tracking people's online activities can use even seemingly innocuous information—in this case, the fact that customers are visiting Orbitz.com from a Mac—to start predicting their tastes and spending habits."[7]

Whether these practices of employing consumer data and AI will help or harm consumers is not obvious, as it is well known from economic theory that price discrimination can either increase or reduce consumer welfare. If, on average, Mac users prefer staying at fancier and more expensive hotels because owning a Mac is correlated with higher income and tastes for luxury, then the Orbitz practice is beneficial because it shows people what they want to see and reduces search frictions. However, if this is just a way to extract more surplus from consumers who are less price sensitive, but do not necessarily care for the snazzier hotel rooms, then it harms these consumers.

There is currently a lot of interest in policy circles regarding the potential harms to consumers from AI-based price discrimination and market segmentation. McSweeny and O'Dea (2017) suggest that once AI is used to create more targeted market segments, this may not only have implications only for consumer welfare, but for antitrust policy and market definitions for mergers. But, as Gal and Elkin-Koren (2017) suggest, the same AI-targeting tools used by retailers and marketplaces to better segment consumers may be developed into tools for consumers that will help them shop for better deals and limit the ways in which marketplaces and retailers can engage in price discrimination.

23.5 Concluding Remarks

In its early years, classical economic theory paid little attention to market frictions and treated information and computation as free. That theory led to conclusions about efficiency, competitive prices for most goods, and full employment of valuable resources. To address the failures of that theory, economists began to study models with search frictions, which predicted that price competition would be attenuated, that some workers and resources

7. See "On Orbitz, Mac Users Steered to Pricier Hotels," Dana Mattioli, *The Wall Street Journal*, Aug. 23, 2012. https://www.wsj.com/articles/SB10001424052702304458604577488822667325882.

could remain unemployed, and that it could be costly to distinguish reliable trading partners from others. They also built markets for complex resource-allocation problems in which computations and some communications were centralized, lifting the burden of coordination from individual market participants.

With these as the key frictions in the traditional economy, AI holds enormous potential to improve efficiency. In this chapter, we have described some of the ways that AI can overcome computational barriers, reduce search frictions, and distinguish reliable partners. These are among the most important causes of inefficiency in traditional economies, and there is no longer any question that AI is helping to overcome them, with the promise of widespread benefits for all of us. As Roth (2002) noted, market designers "cannot work only with the simple conceptual models used for theoretical insights into the general working of markets. Instead, market design calls for an engineering approach." Artificial intelligence has already proven to be a valuable tool in the economist-as-engineer tool box.

References

Blake, T., C. Nosko, and S. Tadelis. 2016. "Returns to Consumer Search: Evidence from eBay."*17th ACM Conference on Electronic Commerce* (EC 2016), 531–45.

Filippas, A., J. J. Horton, and J. M. Golden. 2017. "Reputation in the Long-Run." Working paper, Stern School of Business, New York University.

Gal, M. S., and N. Elkin-Koren. 2017. "Algorithmic Consumers." *Harvard Journal of Law & Technology* 30:1–45.

Hutter, F., H. Hoos, and K. Leyton-Brown. 2011. "Sequential Model-Based Optimization for General Algorithm Configuration." Conference paper, International Conference on Learning and Intelligent Optimization. https://link.springer.com/chapter/10.1007/978-3-642-25566-3_40.

Leyton-Brown, K., P. R. Milgrom, and I. Segal. 2017. "Economics and Computer Science of a Radio Spectrum Reallocation." *Proceedings of the National Academy of Sciences* 114 (28): 7202–09. www.pnas.org/cgi/doi/10.1073/pnas.1701997114.

Li, L. I., S. Tadelis, and X. Zhou. 2016. "Buying Reputation as a Signal of Quality: Evidence from an Online Marketplace." NBER Working Paper no. 22584, Cambridge, MA.

Masterov, D. V., U. F. Mayer, and S. Tadelis. 2015. "Canary in the E-commerce Coal Mine: Detecting and Predicting Poor Experiences Using Buyer-to-Seller Messages." In *Proceedings of the Sixteenth ACM Conference on Economics and Computation*, EC '15, 81–93.

Mayzlin, D., Y. Dover, and J. Chevalier. 2014. "Promotional Reviews: An Empirical Investigation of Online Review Manipulation." *American Economic Review* 104 (8): 2421–55.

McCall, J. J. 1970. "Economics of Information and Job Search." *Quarterly Journal of Economics* 84 (1): 113–26.

McSweeny, T., and B. O'Dea. 2017. "The Implications of Algorithmic Pricing for

Coordinated Effects, Analysis and Price Discrimination Markets in Antitrust Enforcement." *Antitrust* 32 (1): 75–81.
Milgrom, P. R. 2017. *Discovering Prices: Auction Design in Markets with Complex Constraints.* New York: Columbia University Press.
Milgrom, P. R., D. C. North, and B. R. Weingast. 1990. "The Role of Institutions in the Revival of Trade: The Law Merchant, Private Judges, and the Champagne Fairs." *Economics and Politics* 2 (1): 1–23.
Milgrom, P. R., and I. Segal. 2017. "Deferred Acceptance Auctions and Radio Spectrum Reallocation." Working paper.
Mortensen, D. T. 1970. "Job Search, the Duration of Unemployment and the Phillips Curve." *American Economic Review* 60 (5): 847–62.
Myerson, R. B. 1981. "Optimal Auction Design." *Mathematics of Operations Research* 6 (1): 58–73.
Nelson, P. 1970. "Information and Consumer Behavior." *Journal of Political Economy* 78 (2): 311–29.
Newman, N., A. Fréchette, and K. Leyton-Brown. Forthcoming. "Deep Optimization for Spectrum Repacking." *Communications of the ACM* (CACM).
Nosko, C., and S. Tadelis. 2015. "The Limits of Reputation in Platform Markets: An Empirical Analysis and Field Experiment." NBER Working Paper no. 20830, Cambridge, MA.
Ostrovsky, M., and M. Schwartz. 2017. "Reserve Prices in Internet Advertising: A Field Experiment." Working paper.
Stigler, G. J. 1961. "The Economics of Information." *Journal of Political Economy* 69 (3): 213–25.
Roth, A. E. 2002. "The Economist as Engineer: Game Theory, Experimentation, and Computation as Tools for Design Economics." *Econometrica* 70 (4): 1341–78.
Roth, A. E. 2015. *Who Gets What—and Why: The New Economics of Matchmaking and Market Design*. New York: Houghton Mifflin Harcourt.
Vickrey, W. 1961. "Counterspeculation, Auctions, and Competitive Sealed Tenders." *Journal of Finance, American Finance Association* 16 (1): 8–37.
Xu, H., D. Liu, H. Wang, and A. Stavrou. 2015. "E-commerce Reputation Manipulation: The Emergence of Reputation-Escalation-as-a-Service." *Proceedings of 24th World Wide Web Conference* (WWW 2015):1296–306.
Zervas, G., D. Proserpio, and J. W. Byers. 2015. "A First Look at Online Reputation on Airbnb, Where Every Stay Is Above Average." Working paper, Boston University.