Deciding What's Fair: Challenges of Applying Reinforcement Learning in Online Marketplaces

Andrew Chong
UC Berkeley School of Information
qchong@berkeley.edu

Abstract

Reinforcement learning (RL) techniques offer a versatile and powerful extension to the toolkit for computer scientists and marketplace designers for their use in online marketplaces. As the use of RL techniques continues to expand, their application in online marketplaces raise questions of their appropriate use, particularly around issues of fairness and market transparency. I argue that the use of RL techniques, alongside similar calls in domains such as automated vehicle systems, is a problem of sociotechnical specification that faces a set of normative and regulatory challenges unique to marketplaces. I provide a selective overview of the RL literature as applied to markets to illustrate challenges associated with the use of RL techniques in online marketplaces. I conclude with a discussion of capacity-building in research and institutions that is required in order to maximize benefits from algorithmically managed marketplaces for stakeholders and broader society.

1 Introduction

Economic interaction increasingly occurs in marketplace environments managed by platform firms. These mediated market environments, constructed and managed with code, connect a wide range of market participants - from app developers and buyers to consumers and gig workers. Many of these managed marketplaces are integral to the business models of platform companies, and buoyed by network effects, frequently dominate their respective industries - for example, Google Ads in online advertising, Amazon Marketplace in e-commerce, and Uber Marketplace in ridesharing.

In 2002, Roth remarked in "The Economist as Engineer" that computational and experimental methods would become increasingly important for economists engaged in market design, particularly around online marketplaces. He argued that "as marketplaces proliferate on the web... a great deal of market design is going to be done by computer programmers" - with economists having "an opportunity to learn a lot from the markets that result." (p. 1343) He cautioned that "[d]esigners therefore cannot work only with the simple conceptual models used for theoretical insights into the general working of markets," arguing instead that market design in these environments would require "an engineering approach," relying heavily on computation and experimentation. (p. 1341)

Roth's call to economists in 2002 was prescient, but also relatively sanguine about the increasing role platform companies play in managing markets. Today, as online marketplaces have continued to expand, major platform companies make a myriad of decisions on behalf of market participants, actively shaping market outcomes and determining allocative outcomes between different groups of market participants. And while platform companies must strive to ensure marketplaces perform well for participants, they also face strong incentives to experiment with how markets are constructed and algorithmically managed to serve their own interests.

Such marketplace decisions made by platform companies encompass what prices are show to different users, such as in the use of in-app targeted promotions, what quality characteristics are offered to

35th Conference on Neural Information Processing Systems (NeurIPS 2021), Sydney, Australia.

different users, such as wait time in ridesharing, as well as the use of a range of other targeted behavioral nudges and interventions on market participants - a trend Ezrachi and Stucke (2016) describe as "behavioral discrimination." Other key market decisions include how search results are returned, such as returning listings in AirBnb, what affordances are available to individual users, such as the ability to ask for more details on tasks in freelance platform TaskRabbit, and how buyers and sellers are matched - such as between riders and drivers, as in the case of ridesharing.

Within marketplace environments, platform firms manage complex market processes that must continually take into account a range of factors and be dynamically updated. Platforms seek to optimize or maintain various outcomes, including increasing market share, maintaining quality consistency for users, keeping users on the platform, and of course, profit. Internal models take into account factors like reservation wages or prices of workers or sellers, willingness to pay for buyers, and a wide range of other data platforms collect, such as purchase history, result of past experiments, cross-site data and OS/browser characteristics. Particularly where platforms benefit from significant market power, partly due to cross-platform network effects intrinsic to marketplaces, a range of firm marketplace practices have drawn scrutiny from regulators, as well as users themselves.

Within this setting, agent-based modeling has many advantages, with growing literature examining applying RL approaches in marketplace environments. Most crucially, agent-based modeling provides tractability over analytical solutions for complex, dynamic environments, requiring only that the relevant problem space can be specified in term of states, actions and a reward function. Marks (2006) argues that applying agent-based modeling to marketplaces has many other advantages, including the ease of integrating changes to user behavior in learning models. He also argues that in "operating, real-time.. market[s]," "continual shocks might never allow the system to approach, let alone, reach the equilibrium" - advocating models of market engineering that reflect a significant departure from the theoretical literature prioritizing "equilibrium characterization." (p. 1354) Such an actively managed vision of a marketplace is at odds with the celebrated closed-form analytical results in the mainstream economics literature, and more akin, as Roth argues, to an actively managed engineering system. Roth contrasts the shift in approach from that of physics to applied engineering.

This paper provides a selective overview of applications of reinforcement learning in markets and discusses their implications for the increasing use of RL techniques in online marketplaces. I argue that the use of RL techniques, alongside similar calls in domains such as automated vehicle systems, is a problem of sociotechnical specification that faces a set of normative and regulatory challenges unique to marketplaces. I provide a selective overview of applications of RL in online marketplaces to illustrate challenges faced. I conclude with a discussion of capacity-building in research and institutions that is required in order to maximize benefits from algorithmically managed marketplaces for stakeholders and broader society.

2 Modeling the marketplace environment using reinforcement learning

Broadly, reinforcement learning can be described as an approach to modeling behavior where an agent may "learn an optimal action policy in a sequential decision process, through repeated experience." (Charpentier et al., 2020, p. 1) Reinforcement learning is attractive due to the relative simplicity of its set-up, which only requires that the problem space can be defined in terms of agent *states*, *actions* that enable traversal across states, and the definition of a *reward function* that enables the agent to determine which states are preferable. Such models are notably versatile, with the described agent depending on the model set-up.

First, recent applications of RL in online marketplaces model *AI agents* that choose what prices, results or promotions to supply to individual market users on digital platforms. Wang et al. (2015), for instance, utilize a deep reinforcement learning strategy for personalized targeting that varies both the amount and timing of personalized promotions to maximize revenue. Shi et al. (2018) present an alternate model for commodity search in Taobao, a Chinese online retail platform, where the agent is a search engine, with market users simulated as part of the environment to determine which sets of products to return to users.

Second, RL techniques have had a far longer history in applications to model *individual firms* as agents, examining firm pricing strategies that most often are not customized to individual users. Such work has been categorized under revenue or yield management in operations research, and best known in its application to the airline industry. Gosavi et al. (2002), for instance, adopt a RL approach to maximize airline revenue taking into account variables such as multiple fare classes and overbooking. More recently, Liu et al. used data from e-commerce platform Alibaba Inc. to develop a deep reinforcement learning model to provide markdown pricing and daily pricing across thousands of products. Such models may be used customize prices for different groups, for instance, based on destinations and fare class for airline tickets, but due to the choice of agent and model specification, have not been used widely to model personalized price targeting or discrimination.

Other work that model firms as agents have sought to use RL techniques to examine inter-firm dynamics, allowing simulation and tractability of more complex games that do not have analytical solutions, such as in dynamic oligopoly. (Charpentier et al., 2020) Kastius and Schlosser (2021), for instance, model the use of RL by individual firms in duopoly settings, contrasting the effectiveness of different algorithms (Deep Q-Networks and Soft Actor Critic). In their model, they find such techniques can find equilibrium solutions in oligopoly settings that are "usually intractable due to the curse of dimensionality." (p. 1) In their model, they also find that use of RL algorithms by firms can result in collusion even without "direct communication." (p. 1) Waltman and Kaymak (2008) examine the use of Q-learning to model the learning behavior of firms in Cournot oligopoly games, finding, similarly, that collusive behavior can emerge even where "there is no possibility of explicit communication between firms." (p. 3275) While such theoretical results around potential algorithmic collusion are of interest given the increase in RL techniques for revenue management by firms, empirical validation remains unclear in actively managed, complex marketplaces with multiple outcomes for optimization.

Third, RL techniques have modeled *individual users* themselves as agents, seeking to understand how consumers learn and make decisions. Hopkins (2007), for instance, examines whether consumer behavior can best be modeled as "more sophisticated belief-based models" or as "very simple reinforcement learning models," (p. 349) finding that the latter might better explain trends such as consumer lock-in towards purchasing inferior goods. Laibson et al. (2009) model investor saving rate decisions with a naive reinforcement learning heuristic, finding that the model accounts robustly for patterns of saving behavior, suggesting that "individual investors chase their own historical returns and shy away from their own historical return variance when making 401(k) savings rate decisions." (p. 2532)

These contrasting approaches illustrate the considerable flexibility RL techniques have in applications to markets. In the next section, I discuss some key issues that arise in applications of RL in online marketplaces, particularly where they are used to customize market choices and information to different groups or individuals.

3 Problem of sociotechnical specification: Norms and regulations around marketplaces

As marketplace environments are modeled as agent-based problems, they come up against what Gilbert (2021) has described as a problem of sociotechnical specification. Using the application of RL in autonomous vehicle systems as an example, Gilbert discusses the need for "deliberative" mechanisms to achieve consensus around normative definition in agent-based models, arguing for the need to develop "interfaces" with social institutions so that the "definitions of states, actions, and rewards are responsibly indexed to the concerns of stakeholders." Broadly, such a stance has also been discussed in the HCI literature in market design. Lampinen and Brown (2017) discusses the importance of accounting for stakeholder values in market design, where "markets are not [framed] as free-standing, naturally occurring systems, but rather as human artifacts which are actively designed and shaped." (p. 4332) Drawing from Roth's work in market design, they discuss market design concepts of thickness, congestion, safety, stability, but also of repugnance - where people may morally object to specific types of market transactions that they do not themselves engage in but may find morally repugnant, such as the market for kidney sales.

As platform companies seek to design policies based on agent-based models, they come up against two main areas of sociotechnical specification. First, marketplace policies may go against consumer norms and expectations about how marketplaces should function. For instance, in multiple studies, researchers have generally found that consumers disapprove of most forms of price discrimination (Maxwell & Garbarino, 2010; Poort & Borgesius, 2019), presenting a direct challenge for marketplace policies based on the use of AI agents for price targeting for different users. Second, marketplace policies must navigate existing laws and regulations around markets, notably in consumer protection, data privacy, antitrust, and anti-discrimination. For example, in the United States, a variety of credit, housing and employment laws disallow personalized pricing based on protected characteristics. (See Directorate, 2018) In the EU, in addition to existing laws around consumer protection and data privacy, regulators are considering rules that require companies to inform consumers of how prices are set. The Consumer Rights Directive, for instance, includes language on how online marketplaces must "inform consumers of the main parameters determining ranking of offers presented to them and the relative importance of these parameters as opposed to other parameters." (De Streel & Jacques, 2019) Srinivasan (2020) has additionally argued that the rules around financial marketplaces should also apply to ad sales markets operated by Google as custodians of market information. In the following, I discuss three key characteristics of reinforcement learning and their relevance to the problem of sociotechnical specification.

3.1 Dimensions and inputs for maximization

In defining potential *actions* in agent-based models, model developers define the scope by which platforms operating marketplaces may seek to maximize outcomes. For instance, Wang et al. (2015) develop a model for personalized targeting for mobile promotions where prices and timing are customized for different consumers to maximize revenue for the firm. Such forms of price targeting comes up against consumer norms and expectation for how markets should function, particularly around price discrimination. Both Maxwell & Gabarino (2010) and Poort & Borgesius (2019) find that consumers disapprove of most forms of price discrimination, finding that most consumers feel that a single firm should provide the same prices to all its customers.

Models may also be developed to engage in various forms of price, quality or behavioral discrimination, or use data - such as estimates of a user's willingness to pay or response to past promotions - that users may find unfair or deceptive, which would fall under the FTC's mandate to challenge "unfair or deceptive acts or practices" under Section 5b) of the FTC Act in the United States. For instance, AI agents in Wang et al. takes into account individual consumers' past response to promotions to determine which prices to provide to consumers.

3.2 Reward function

In agent-based models, reward functions are defined to quantify desirable states for agents, and policies are compared in terms of long-term outcomes such as revenue. For instance, an AI pricing agent may have as its reward function surplus extracted from consumers, which is then maximized by the agent. Wang et al. (2015), for instance, seeks to maximize overall long-term revenue for the firm by varying timing and prices in targeted promotions provided to users. In a marketplace environment, such simulations, esp. where operated by a for-profit platform company, are unlikely to result in policies that maximize total market surplus for buyers and sellers. Instead, marketplace companies may experiment with the use of pricing AI agents to maximally extract surplus from buyers and sellers.

In the U.S., antitrust seeks to protect the competitive process as a means of increasing output and lowering prices for consumers. Regulators are less concerned with the ability of platform firms to extract surplus as long as there is sufficient competition between platforms. To the extent that many platforms hold significant market power in their respective industries due to marketplaces benefiting from natural cross-platform network effects, the ability of platform firms to extract surplus through price, quality or behavioral discrimination using RL techniques becomes of greater concern. As regulation continues to evolve, policymakers and users may demand greater transparency over what outcomes RL models are optimizing for within online marketplaces.

3.3 Explainability and transparency

Agent-based models have been particularly attractive in their ability to provide tractable solutions to dynamic, complex environments through the use of simulations, as opposed to closed-form analytical solutions. Such models pose problems of explainability and transparency for both everyday users but also regulators and researchers, since outcomes arrive from simulations that often suffer from a variety of robustness issues. Marketplace firms that apply reinforcement learning models face the task of demonstrating that RL-derived policies benefit stakeholders, and that their derived policies do not contravene existing rules and regulations around marketplaces, particularly around antitrust.

As a useful comparison, while the use of game theory has been influential in providing theoretical foundations for when antitrust intervention may be necessary, empirical application of such models in the U.S. courts remains limited, with - as Hovenkamp discusses - various scholars remaining opposed to their use in policy. (p. 347) Similarly, should agent-based models become increasingly used by platform companies, it will be noteworthy to see the extent to which they will be invoked in antitrust proceedings.

Increased use of such models by platform companies will complicate the ability of jurors and judges to assess antitrust violations. Well-known jurist and law and economics scholar Richard Posner describes in 2001 that issues of technical complexity were increasingly entangled with issues of antitrust, and bemoaned the difficulty of finding "truly neutral competent experts" who were not themselves "employed by or have other financial ties to firms involved in or potentially affected by antitrust litigation in this sector." (p. 937) Should RL techniques become increasingly used to manage complex, dynamic market environments, such issues of inadequate technical expertise will recur as jurors and judges are asked to evaluate whether complex, simulated models for managing marketplaces meet various criteria. Among these key decisions include whether a firm's policies are sufficiently increasing output or maximizing welfare for consumers, or meets the rule of reason in the Sherman Act - an assessment of whether the pro-competitive effects of a firm practice outweigh its anti-competitive effects.

4 Conclusion

Reinforcement learning techniques are the latest extension to the toolkit computer scientists and marketplace designers possess for use in online marketplaces. Due to their flexibility and tractability, RL techniques offer advantages for modeling actively managed marketplace environments, but also, as discussed, introduce a range of concerns. In order to maximize benefits from their application in algorithmically managed marketplaces, more research is needed to examine how such agent-based models may be used to maximize social welfare outcomes in marketplace environments. Additionally, new deliberative mechanisms and institutions are required to determine normative constraints around the implementation of such models in marketplace environments, as well as work to improve explainability and transparency to stakeholders and regulators, who may otherwise resist the implementation of such models if they cannot be demonstrated to improve outcomes for stakeholders and broader society.

References

Afsar, M. M., Crump, T., & Far, B. (2021). Reinforcement learning based recommender systems: A survey. arXiv preprint arXiv:2101.06286.

Charpentier, A., Elie, R., & Remlinger, C. (2021). Reinforcement learning in economics and finance. Computational Economics, 1-38.

Choi, J. J., Laibson, D., Madrian, B. C., & Metrick, A. (2009). Reinforcement learning and savings behavior. The Journal of finance, 64(6), 2515-2534.

Directorate for Financial and Enterprise Affairs Competition Committee (2018, Nov). Personalised Pricing in the Digital Era-Note by the United States, ORGANISATION FOR ECON. COOP. DEV.

De Streel, A., Jacques, F. (2019). Personalised pricing and EU law.

Ezrachi, A., & Stucke, M. E. (2016). The rise of behavioural discrimination.

Gilbert, T. K. (2021, Jan). Mapping the Political Economy of Reinforcement Learning Systems: The Case of Autonomous Vehicles. Simons Institute Newsletter.

Gosavii, A., Bandla, N., & Das, T. K. (2002). A reinforcement learning approach to a single leg airline revenue management problem with multiple fare classes and overbooking. IIE transactions, 34(9), 729-742.

Hovenkamp, H. (1995). Law and Economics in the United States: a brief historical survey. Cambridge Journal of Economics, 19, 331-331.

Kastius, A., & Schlosser, R. (2021). Dynamic pricing under competition using reinforcement learning. Journal of Revenue and Pricing Management, 1-14.

Lampinen, A., & Brown, B. (2017, May). Market design for HCI: Successes and failures of peer-to-peer exchange platforms. In Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (pp. 4331-4343).

Liu, J., Zhang, Y., Wang, X., Deng, Y., & Wu, X. (2019). Dynamic Pricing on E-commerce Platform with Deep Reinforcement Learning. arXiv preprint arXiv:1912.02572.

Marks, R. (2006). Market design using agent-based models. Handbook of computational economics, 2, 1339-1380.

Maxwell, S., & Garbarino, E. (2010). The identification of social norms of price discrimination on the internet. Journal of Product Brand Management.

Poort, J., & Zuiderveen Borgesius, F. J. (2019). Does everyone have a price? Understanding people's attitude towards online and offline price discrimination. Internet Policy Review, 8(1), 1-20.

Posner, R. A. (2000). Antitrust in the new economy. Antitrust Lj, 68, 925.

Srinivasan, D. (2020). Why Google Dominates Advertising Markets. Stan. Tech. L. Rev., 24, 55.

Roth, A. E. (2002). The economist as engineer: Game theory, experimentation, and computation as tools for design economics. Econometrica, 70(4), 1341-1378.

Shi, J. C., Yu, Y., Da, Q., Chen, S. Y., & Zeng, A. X. (2019, July). Virtual-taobao: Virtualizing real-world online retail environment for reinforcement learning. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 33, No. 01, pp. 4902-4909).

Waltman, L., & Kaymak, U. (2008). Q-learning agents in a Cournot oligopoly model. Journal of Economic Dynamics and Control, 32(10), 3275-3293.

Wang, W., Li, B., & Luo, X. (2019). AI Agents for Sequential Promotions: Combining Deep Reinforcement Learning and Dynamic Field Experimentation. Fox School of Business Research Paper Forthcoming.

Zhang, K., Yang, Z., & Başar, T. (2021). Multi-agent reinforcement learning: A selective overview of theories and algorithms. Handbook of Reinforcement Learning and Control, 321-384.