Optimal Bargaining on eBay Using Reinforcement Learning*

Etan A. Green† E. Barry Plunkett

University of Pennsylvania

Abstract

Bargaining is ubiquitous. How can people bargain better? We train a reinforcement learning agent to bargain optimally in "Best Offer" listings on eBay, and we characterize its behavior in a manner that humans can use. As a buyer, the agent starts lower than human buyers and bargains longer. As the seller, the agent interprets offers as signals—of the buyer's willingness to pay and of the item's desirability—that human sellers ignore. Simple strategies derived from these agents purchase more items for lower prices than human buyers and sell more items for higher prices than human sellers.

*Replication materials: `https://github.com/bpiv400/eBay`. We thank Matthew Backus, Thomas Blake, Bradley Larsen, and Steven Tadelis for posting the dataset that underlies this paper; Ashton Anderson, Matthew Backus, Hamsa Bastani, Jake Hofman, Yuval Rottenstreich, Maurice Schweitzer, Joseph Simmons, and Lyle Ungar for comments on previous drafts; seminar participants at Microsoft Research, UCSD, and Data Colada; and the Dean's Fund and Analytics Initiative at Wharton for generous financial support.

†To whom correspondence should be addressed: `etangr@upenn.edu`

Introduction

This paper identifies optimal bargaining strategies in a marketplace with millions of items: "Best Offer" listings on eBay. In a Best Offer listing, the seller advertises a list price and then haggles with buyers—a format common to markets for real estate, automobiles, and business-to-business transactions, among others.

Standard approaches to studying bargaining are ill-equipped to prescribe an optimal strategy for bargaining against humans. The theoretical literature focuses on equilibrium predictions, not best responses to boundedly rational opponents [1]. Laboratory experiments are designed to evaluate a limited number of pre-determined strategies, not to find the best one among a broad set [e.g., 2, 3]. And while field studies describe how people bargain [4–9], none as yet prescribes a strategy that performs better.

We train a reinforcement learning (RL) agent to bargain optimally on eBay. Inspired by theories of animal behavior, RL agents learn through trial and error—by taking actions, observing outcomes, and reinforcing actions that lead to desirable outcomes. This training process culminates in a strategy (often called a policy) that maps the offer history for a given listing to an offer that maximizes the eventual payoff. We find optimal strategies for buyers and sellers, and we characterize these strategies in a manner that humans can use.

As a buyer, the agent bargains more aggressively than humans. The agent buyer begins with lower first offers, and it makes more offers—and it buys more items at lower prices as a result.

As the seller, the agent interprets signals that humans ignore—and sells more items for higher prices as a result. In particular, the agent rejects high first offers, as these signal the buyer's willingness to pay more, and it rejects early first offers, which signal the item's desirability. By contrast, human sellers accept high first offers irrespective of their timing.

Humans frequently play poor strategies—buyers by making their best offer first, rather than last, and sellers by countering with small concessions. Experience is a partial salve: more experienced human buyers (but not sellers) behave more like the agent.

A different approach to training game-playing RL agents

A striking feature of recent game-playing RL agents, such as those for chess and Go [10], is that they learn to beat the best humans without ever playing a human. Instead, they play against copies of themselves, learning best responses to best-responding opponents and converging to equilibrium strategies. This works because each of these games is two-player and zero-sum. For (symmetric) two-player zero-sum games, the minimax theorem guarantees that a Nash equilibrium strategy will not lose more often than it wins—regardless of how its opponent plays.

No such guarantee applies to games that are not two-player and zero-sum, such as bargaining. On eBay, the seller may bargain with multiple buyers, and a seller and buyer share a surplus only if they reach an agreement.

To train agents that bargain optimally against humans, we train them to best respond to humans, not to copies of themselves. The strategies that the agents learn, while beatable, are designed to exploit human weaknesses—and in so doing, suggest ways for humans to improve.

Ideally, the agents would train directly on eBay. The difficulty is that RL agents achieve superhuman performance only after a volume of interaction that is infeasible in the real world.[1] We circumvent this issue by training neural networks to mimic the behavior of human buyers and sellers as observed in a large dataset of Best Offer listings [5]. We then train the agents by pitting them against these simulated humans.

The value of this approach rests on two conditions. First, the agents must confine their search to strategies that humans sometimes play; otherwise, the response from the simulated humans will be unreliable. We ensure this by constraining the agents to offers that humans regularly make. Second, the simulated humans must credibly mimic humans. This is partly verifiable, by comparing the behavior of real and simulated humans, as we do. But it also rests on an assumption that the simulated humans are trained on the information that actual humans observe. Online contexts such as ours make this assumption credible. On eBay, the buyer and seller communicate through a minimal interface, and eBay records much of what they see.[2]

Simulating Best Offer listings on eBay

Best Offer listings on eBay follow a rigid structure.[3] The seller sets a list price, at which a buyer may purchase the item immediately. If a buyer offers less than the list price (in turn 1), the seller may accept, counter, or reject (in turn 2). After the seller rejects or counters, the buyer may accept, counter, or walk (i.e., end the negotiation). A *thread* between the seller and one buyer consists of at most three offers per side, in alternating fashion, followed by a take-it-or-leave-it decision for the buyer (in turn 7).[4]

We train neural networks to mimic buyer and seller behavior as observed in a dataset comprising the universe of Best Offer listings on eBay over a year-long period. Along with features that describe the listing (e.g., the list price), the data contain complete offer histories for all threads. Collectively, we train nearly two dozen neural networks (further described in the Appendix). One model predicts a distribution of arrival times. Another predicts a distribution of first offers. Separate models predict distributions over the response time and offer in subsequent turns. We use the trained models to simulate threads by sampling buyer arrivals, first offers, seller responses, and so on until the item sells or the listing expires. We do not simulate attributes of the listing; instead, we enforce the attributes chosen by the human seller, including the list price.

To validate the simulated buyers and sellers, we first simulate each listing once in a holdout

[1] For example, AlphaZero eclipsed the Stockfish engine in chess after 1.2 billion moves [10].

[2] This assumption does not hold perfectly. For instance, while we observe the number of photos of each item, we do not observe the photos themselves.

[3] This structure reflects the period when the data were generated, during 2012 and 2013.

[4] After a thread ends, the buyer and seller rarely interact again. In the data, 91% of buyer-seller interactions are unique to one thread.

sample. For a wide variety of outcomes, the distribution from the simulations matches the distribution from the data almost perfectly (Figures A7-A14 in the Appendix). Second, we train a neural network to discriminate between simulated threads and those observed in the data. For each thread, this discriminator observes features of the listing, which are identical for simulated and real threads, as well as the offer path, which may vary. The discriminator largely fails to distinguish between real and simulated threads. In a separate holdout sample, the discriminator correctly classifies 53% of threads, barely higher than the random-guessing benchmark of 50% (Figure A15).[5] Put differently, the simulated humans generate counterfactual offer paths that are nearly indistinguishable from real ones.

The agent buyer

The agent buyer is a neural network that maps features the buyer observes (of both the listing and the offer history so far) to an offer that maximizes the buyer's surplus. This surplus is 0 if the agent does not purchase the item; otherwise, it is the difference between the value of the item and the purchase price. We train many agents, each of which values items at a different multiple, λ, of the list price. The agent makes offers on behalf of one buyer selected at random in each listing. The analyses that follow compare agents in simulated listings to randomly selected threads from the data. (See the Appendix for more detail.)

Bargaining strategies that yield larger discounts tend to result in fewer purchases, as brinkmanship jeopardizes agreement. Hence, we evaluate buyers based on the discounts they achieve and the rate at which they purchase items (Figure 1a). Human buyers purchase 68% of the items for which they either pay full price upfront or make an offer, and they average a discount off the list price of 18%, or $16.74, on those purchases.

The agents buy more items for less. For instance, the $\lambda = 1 - \epsilon$ agent, who is willing to pay any amount below the list price, purchases 79% of items at an average discount of 34%, or $33.08, off the list price.[6] Simple strategies that condition the offer only on the turn (Figure 1b) approach the performance of the agents. Consider a strategy that offers two-thirds of the list price and accepts whatever the seller offers in response. This strategy purchases 95% of items at an average discount of 20%, comparable to the $\lambda = 3$ agent.

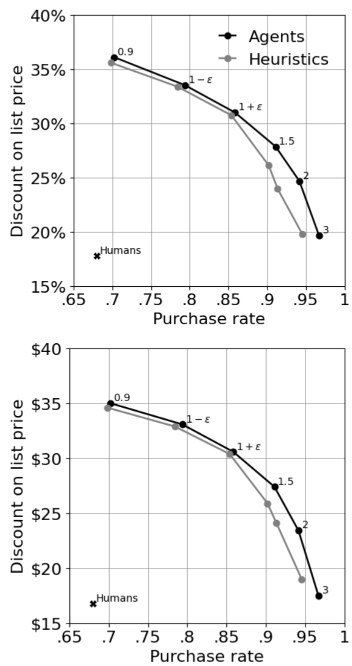
Humans realize paltry discounts in part because they pay full price upfront 23% of the time (Figure 2a). By contrast, $\lambda \leq 1$ agents never pay the list price (on any turn), and even the $\lambda = 3$ agent, who realizes a large surplus at any sale price, pays full price on just 17% of its first turns. Paying the list price upfront sacrifices a large discount to insure against a small risk of losing the item. In the data, sellers respond to 96% of first offers, with an average discount of 18% off the

[5] To test whether the discriminator can differentiate between threads generated by different processes, we train a separate neural network to discriminate between threads from the data and those from simulations in which one human buyer is replaced at random by an agent buyer (further described in the Appendix). This "placebo" discriminator achieves an accuracy of 99%.

[6] Using a more conservative baseline than the item's list price, this agent pays $5.89 less, on average, than the highest buyer offer observed in the data (Figure A16).

Figure 1: The agents and corresponding heuristics purchase more items than humans, and at lower prices.

(a) Average percentage (top) and dollar (bottom) discount off the list price on purchases.



Note: The agents are indexed by λ—i.e., their valuation as a multiple of the list price. A $\lambda = 1 - \epsilon$ agent is willing to pay up to, but not including, the list price, whereas a $\lambda = 1 + \epsilon$ agent is also willing to pay the list price.

(b) Heuristics that approximate the agents. The turn-1 offer is the fraction of the list price. In later turns, the offer is the fractional distance from the buyer's previous offer to the seller's most recent offer.

	Turn			
λ	1	3	5	7
0.9	1/2	1/5	1/4	acc
$1-\epsilon$	1/2	2/5	2/5	acc
$1+\epsilon$	1/2	2/5	1/2	acc
1.5	3/5	1/2	acc	
2	2/3	1/2	acc	
3	2/3	acc		

Note: For $\lambda < 1$, the heuristic accepts in turn 7 only if the seller's final offer is less than λ times the list price.

list price.

When human buyers bargain, they make high first offers, averaging 62% of the list price (Figure 2c). By contrast, agents willing to pay up to 150% of the list price always start at 50% of the list price, which is the lowest first offer we commonly observe in the data, and as a result, the lowest first offer we allow the agents to make. Even an agent that values items at twice the list price starts lower than humans, on average.

Despite making high first offers, human buyers purchase relatively few items. This is because humans often make their best offer first. When the seller does not accept the buyer's first offer, human buyers walk 44% of the time. By contrast, the agents never walk before the final turn. Of their 3 allotted offers, human buyers use 1.2 on average, compared to 1.3 for the $\lambda = 3$ agent and 2.3 for the bargain-hunting $\lambda = 0.9$ agent (Figure 2b).

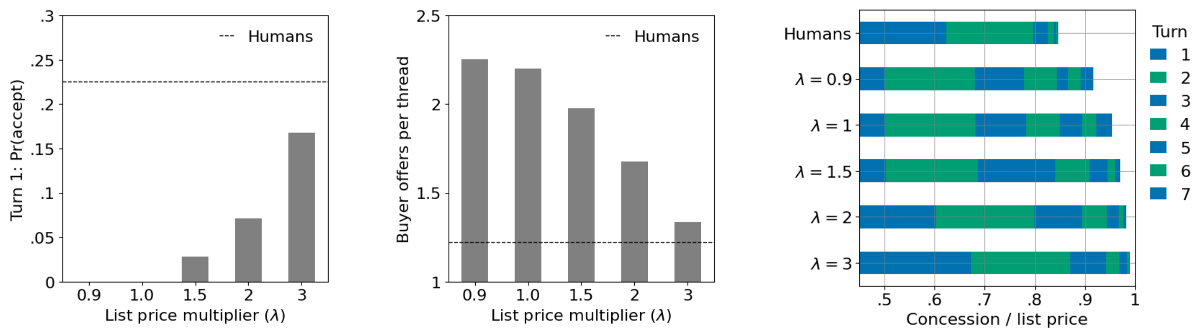
By starting lower and bargaining longer, the agents extract larger discounts (Figure 2c). For human buyers, the seller's concessions total 19% of the list price on average, of which just 1.5 percentage points comes from the seller's second and third offers combined. For the $\lambda = 1$ agent,

Figure 2: The agents bargain more often than human buyers (a). When they bargain, the agents make more offers (b), make lower first offers (c), and extract larger concessions from sellers in later turns (c). Humans' aversion to bargaining is less consistent with bargaining costs (d) than with inexperience (e).

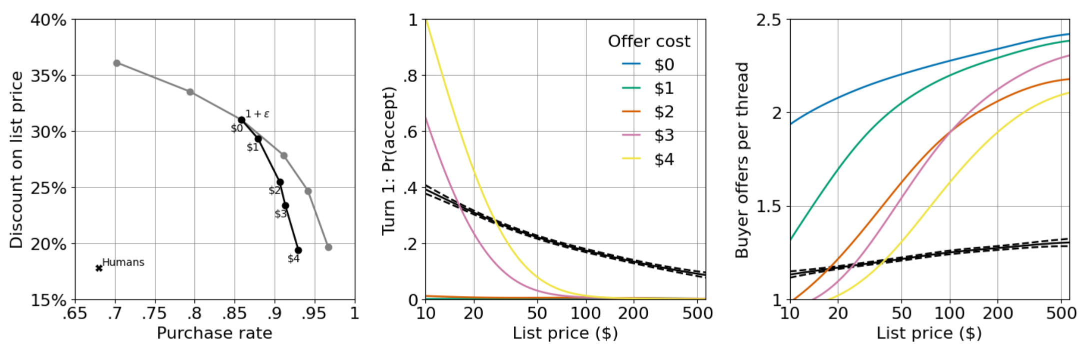
(a) Rate at which buyers pay full price upfront.

(b) Average number of offers made by the buyer.

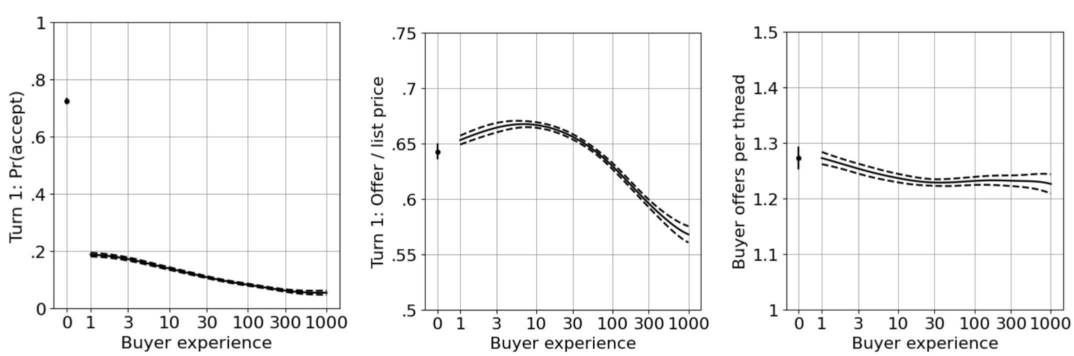
(c) Amount conceded by turn when the buyer bargains.



(d) We retrain the $\lambda = 1 + \epsilon$ agent with a fixed dollar cost for every offer it makes (left). Offer costs induce the agent to pay full price upfront (middle) and stop bargaining earlier (right)—but only for inexpensive listings. Human buyers are relatively insensitive to the list price.



(e) More experienced human buyers bargain more often (left) and make lower first offers when they bargain (middle), but they make the same number of offers on average as inexperienced buyers (right).



the seller's concessions total 28% of the list price on average, of which 9.6 percentage points comes from the seller's last two offers.

These late concessions close the gap between the parties and increase the probability of a sale. At the end of a thread, the average gap between the $\lambda = 1$ agent and the seller is just 4.6% of the list price (including purchases, which have a gap of 0). Threads with human buyers conclude with the parties separated by 15% of the list price, on average.

Humans' apparent aversion to bargaining could be explained by the time and effort costs of bargaining, which our agents do not bear. To explore this, we train agents that incur a fixed cost for every offer they make (Figure 2d). Moderate offer costs rationalize an aversion to bargaining—but only for inexpensive items. For expensive items, the potential discount exceeds these costs, and the penalized agents always bargain and frequently make multiple offers. Presumably, human buyers who find bargaining costly will only bargain for expensive items. However, human buyers are not particularly sensitive to the list price. For items listed at $10, the potential savings are slim—negotiated sale prices average $7—yet a majority of human buyers bargain over these items. For items listed at $500, the average savings on negotiated sales exceed $100, yet 10% of human buyers still pay full price upfront. Similarly, the number of offers made by human buyers is relatively unresponsive to the list price.

Alternatively, some human buyers may choose not to bargain because they underestimate the benefits. Consistent with this explanation, more experienced human buyers behave more like the agents (Figure 2e). Buyers who have not previously made an offer in a Best Offer listing do so only a quarter of the time, whereas those who have previously made an offer for just a single item bargain 80% of the time.[7] Presumably, uninitiated buyers do not realize that sellers offer large discounts to almost anyone who asks. When they bargain, more experienced buyers make lower first offers than less experienced buyers, but they do not bargain any longer.

The agent seller

The agent seller is a neural network that maps features the seller observes to an offer that maximizes the payoff when the listing ends. If the item sells, the payoff is the sale price. If the listing expires without a sale, the payoff is the residual value of the item. We train two agents: an impatient agent, who values every item at zero (and who is thus highly motivated to sell); and a more patient agent, who values each item at a continuation value that we estimate by simulating each listing many times. (See the Appendix for more detail.) The analyses that follow compare the agent seller in simulated listings to human sellers in the data—for listings in which the seller responds to a buyer offer. This excludes listings that expire without a buyer arrival or for which the first buyer pays the list price immediately.

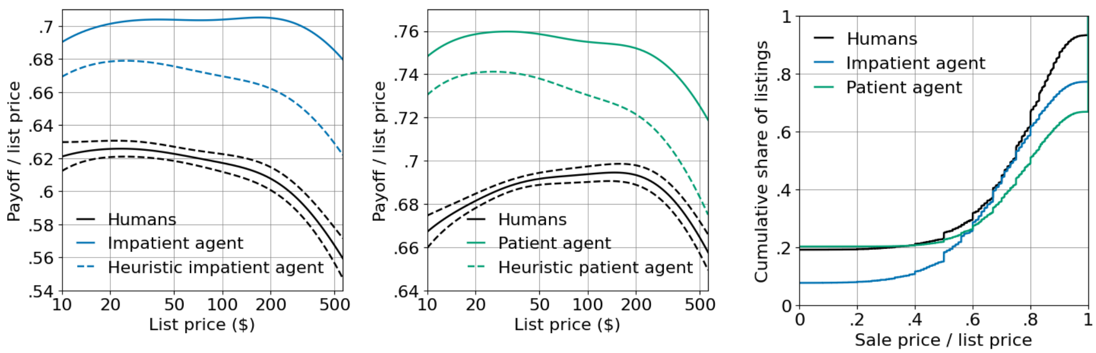
Both agents achieve higher payoffs than human sellers at every list price (Figure 3a), largely because the agents sell more items for full price (Figure 3b). The patient agent sells a similar

[7] Estimated with buyer fixed-effects, going from zero to one listing of experience decreases the probability that a buyer pays full price upfront by 8 percentage points ($|t| = 27.6$).

Figure 3: Both agents achieve higher payoffs than human sellers at all list prices, and they sell more items for full price.

(a) Average normalized payoffs by list price when unsold listings are valued at 0 (left) or at their estimated market values (right).

(b) Cumulative distribution of normalized sale prices.



Note: Figure 6 diagrams the heuristic agents.

Note: Unsold listings have a sale price of 0. Figure A19 shows corresponding plots for the heuristic agents.

share of items as human sellers, yet the patient agent receives the list price five times as often as human sellers. The impatient agent sells more items than human sellers—and still receives full price more than three times as often as human sellers.

The agents sell more items for the list price by rejecting generous first offers. Whereas human sellers typically accept first offers that are close to the list price, both agents frequently reject them—and they reject more generous first offers at *higher* rates (Figure 4a), a pattern that holds for all list prices (Figure A21). In essence, the agents wager that they can improve upon generous first offers. This bet is particularly risky for the impatient agent, who could end up with nothing. Even so, it rejects first offers for 90% of the list price six times as often as human sellers.

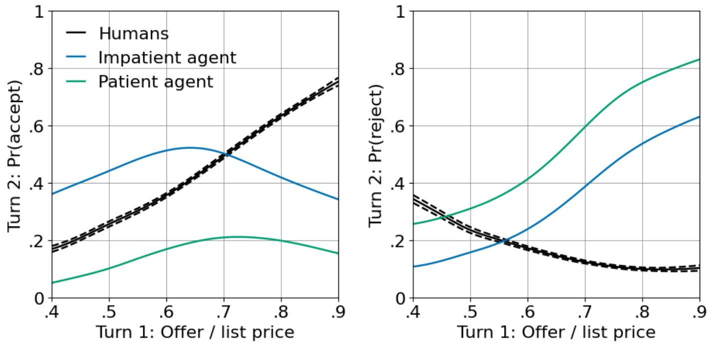
One reason that these bets pay off is because rejecting generous first offers induces buyers to pay full price (Figure 4b). The more generous the first offer, the more likely the buyer is to pay full price after a rejection, a pattern that holds for all list prices (Figure A22). A generous first offer signals the buyer's willingness to pay more, and rejecting that offer separates buyers who are willing to pay full price from those who are not.

Buyers appear to interpret a rejection as an indication that the list price is firm, and a concession as an indication that the price is negotiable. Buyers are sharply more likely to pay full price, and less likely to counter, after their first offers are rejected than after the seller responds with a small concession (Figure 4c).

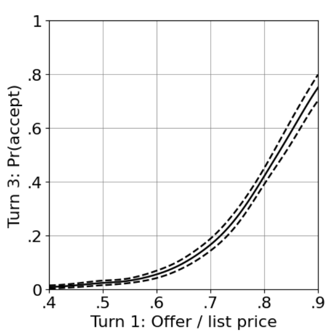
Small concessions appear to be a dominated strategy. Neither agent ever makes a counteroffer that is closer to its previous offer than to the buyer's current offer. By contrast, both amateur

Figure 4: The agents reject generous first offers, and buyers often respond by paying full price.

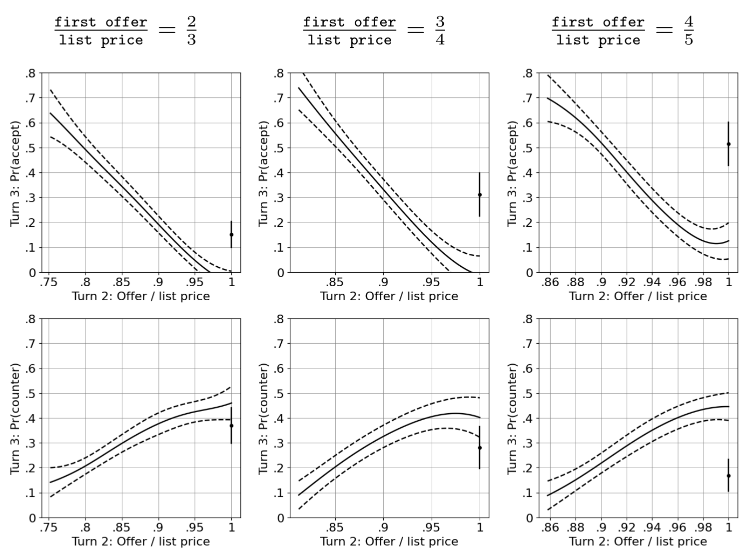
(a) Rates at which sellers accept (left) and reject (right) first offers. Human sellers accept generous first offers, whereas both agents reject them.



(b) The rate at which buyers accept (i.e., pay full price) after the seller rejects the first offer, in the data. Buyers who make more generous first offers are more willing to pay full price.



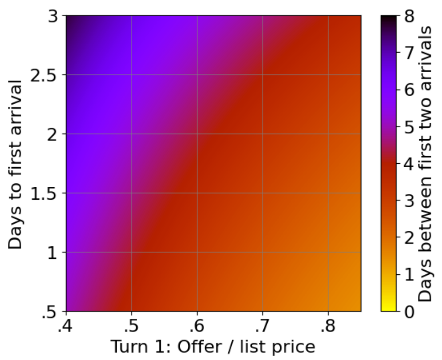
(c) Buyer accept (top) and counteroffer (bottom) rates after the seller responds to the first offer, for common first offers. Buyers are discontinuously more likely to accept, and less likely to counter, after the seller rejects the first offer than after the seller offers a small concession, particularly when the first offer is more generous.



and professional human sellers do so for a majority of counteroffers.[8] Rather than reject, humans concede a little. Rather than concede a little, the agents reject (or concede a lot).

Figure 5: A generous first offer by an early arriving first buyer indicates that the item is popular. Whereas human buyers are sensitive to these attributes, human sellers—both amateur and professional—are not.

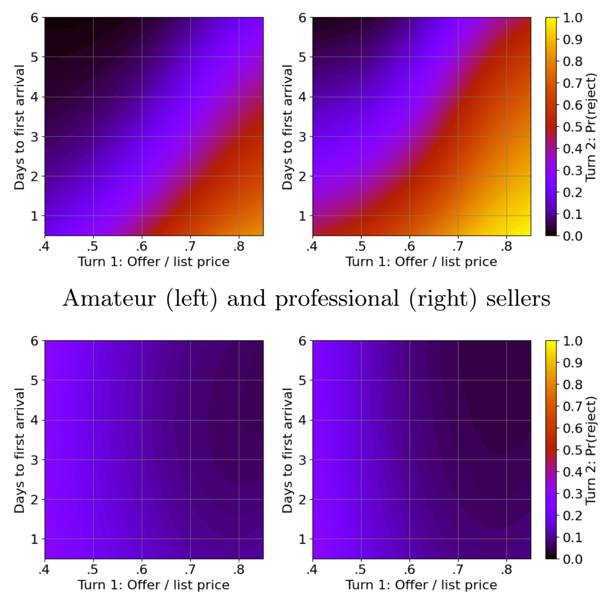
(a) Expected number of days between the first two buyer arrivals, in the data. Generous and early first offers predict that a second buyer will arrive soon.



Note: Assumes a Poisson arrival process in which λ is log-linear in the first offer, the time to the first arrival, and an interaction between them.

(b) Rates at which agents and humans reject first offers.

Impatient (left) and patient (right) agents



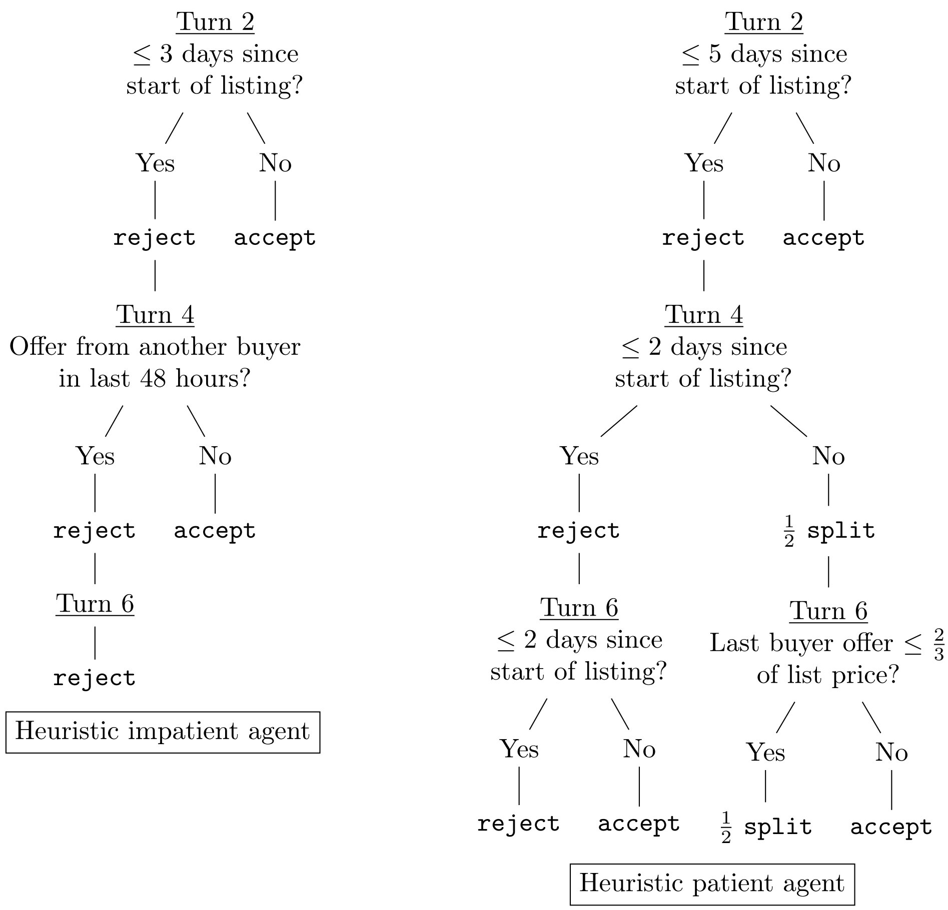
A second reason to reject generous first offers is that generous—and early—first offers predict that another buyer will arrive soon (Figure 5a). This relationship insures against the downside risk of rejecting a generous first offer—i.e., that the buyer walks.

The agents are sensitive to the timing of the first buyer's arrival in a way that human sellers are not (Figure 5b). Both agents frequently reject early and generous first offers, and rarely reject late and ungenerous ones. By contrast, human sellers are insensitive to the timing of the buyer's arrival (and reject more generous first offers at *lower* rates). Here, we split human sellers into amateurs and professionals, both of whom reject infrequently regardless of the timing of the offer. In other words, we find no evidence that professional sellers behave more similarly to our agents than amateurs.

It is difficult to overstate the importance of rejecting early first offers in each agent's strategy. We train a *heuristic* model for each agent (Figure 6): a one-cut decision tree that best predicts its offers in each turn, conditional on its offers in previous turns. Despite this simplicity, both heuristics achieve higher payoffs than human sellers at all list prices (Figure 3a). Both heuristics

[8] Professional sellers are those with an eBay store. The median number of prior Best Offer listings for sellers with and without a store is 39 and 4, respectively.

Figure 6: Heuristic versions of the impatient (left) and patient (right) agents. In turn 2, both heuristic agents condition their behavior only on the time since the item was listed, rejecting early and accepting late.



Note: $\frac{1}{2}$ `split` is an offer that splits the difference between the buyer and seller's most recent offers.

also treat first offers similarly, by rejecting those that arrive early and accepting those that arrive late. These heuristics bound a set of simple strategies that human sellers can use to better respond to buyers.

Discussion

A long-standing debate in artificial intelligence is whether algorithms should be designed to replace or augment humans [11, 12]. These goals need not be in opposition: an agent that outperforms humans can also help humans improve. We train an algorithm to beat humans at a real-world, economic game, and we characterize its behavior in a manner that humans can use.[9]

Our approach is subject to some limitations. First, we do not consider complex payoff functions (e.g., that incorporate the possibility of future interactions). Second, we do not explore all possible strategies (e.g., amending the list price), lest the agents explore beyond the confines of the data.

Strategies learned on eBay may not be optimal elsewhere. For instance, the agent seller infers the desirability of items from offers because demand for items on eBay (e.g., bespoke antiques) is otherwise hard to predict. In markets where items are more comparable (e.g., real estate), sellers may be able to predict demand from past sales.

Future work could evaluate the agents on eBay, either by substituting a human with an agent, or by describing the heuristic agents to humans and observing changes in their strategies and outcomes. Some caution is warranted, as these strategies only work on the margin. If buyers and sellers both behave like the agents, threads will last longer, sellers will make large concessions late, and buyers will capture a majority of the surplus.

Separately, future experimental work should break the convention of pairing a seller with one buyer. Our results speak to the importance of predicting future buyer arrivals—and hence, to the limitations of studies in which the number of buyers is arbitrarily restricted to one. For the same reason, future theoretical analysis of bargaining should relax the common assumption of fixed valuations. Buyer offers inform sellers about demand for their items, and sellers should learn from these signals.

Economists have begun using machine learning to solve "prediction problems" [14], such as whether to detain a defendant before trial [15]. We use reinforcement learning to solve a dynamic problem, requiring a sequence of decisions rather than a single choice. Many interesting problems in economics are dynamic in nature, such as how to allocate capital to investments or organs to patients. Reinforcement learning offers the prospect of solving these real-world games.

[9] Agents that mimic humans, rather than beat them, can also help humans improve. In a complex game like chess, recommendations from agents trained to mimic slightly better humans can be more digestible than those from agents trained to beat the best humans [13].

References

[1] C. F. Camerer, *Journal of Economic Perspectives* **11**, 167 (1997).

[2] C. F. Camerer, G. Nave, A. Smith, *Management Science* **65**, 1867 (2019).

[3] P.-H. Lin, *et al.*, *Nature Human Behaviour* **4**, 917 (2020).

[4] M. Backus, T. Blake, S. Tadelis, *Journal of Political Economy* **127**, 1599 (2019).

[5] M. Backus, T. Blake, B. Larsen, S. Tadelis, *The Quarterly Journal of Economics* **135**, 1319 (2020).

[6] M. Backus, T. Blake, J. Pettus, S. Tadelis, *NBER* (2020).

[7] B. J. Larsen, *The Review of Economic Studies* **88**, 851 (2021).

[8] D. Keniston, *et al.*, *NBER* (2021).

[9] B. Larsen, J. Freyberger, *NBER* (2021).

[10] D. Silver, *et al.*, *Science* **362**, 1140 (2018).

[11] H. A. Simon, A. Newell, *Operations research* **6**, 1 (1958).

[12] D. C. Engelbart, *Augmenting human intellect: A conceptual framework* (1962).

[13] R. McIlroy-Young, S. Sen, J. Kleinberg, A. Anderson, KDD (2020), p. 1677–1687.

[14] J. Kleinberg, J. Ludwig, S. Mullainathan, Z. Obermeyer, *American Economic Review* **105**, 491 (2015).

[15] J. Kleinberg, H. Lakkaraju, J. Leskovec, J. Ludwig, S. Mullainathan, *The Quarterly Journal of Economics* **133**, 237 (2018).