Recency, Records, and Recaps: Learning and Nonequilibrium Behavior in a Simple Decision Problem

DREW FUDENBERG and ALEXANDER PEYSAKHOVICH, Harvard University

Nash equilibrium takes optimization as a primitive, but suboptimal behavior can persist in simple stochastic decision problems. This has motivated the development of other equilibrium concepts such as cursed equilibrium and behavioral equilibrium. We experimentally study a simple adverse selection (or "lemons") problem and find that learning models that heavily discount past information (i.e., display recency bias) explain patterns of behavior better than Nash, cursed, or behavioral equilibrium. Providing counterfactual information or a record of past outcomes does little to aid convergence to optimal strategies, but providing sample averages ("recaps") gets individuals most of the way to optimality. Thus, recency effects are not solely due to limited memory but stem from some other form of cognitive constraints. Our results show the importance of going beyond static optimization and incorporating features of human learning into economic models used in both understanding phenomena and designing market institutions.

Categories and Subject Descriptors: J.4 [**Social and Behavioral Sciences**]: Economics

General Terms: Economics, Theory, Human Factors

Additional Key Words and Phrases: Learning, behavioral economics, recency, equilibrium concepts

ACM Reference Format:
Drew Fudenberg and Alexander Peysakhovich. 2016. Recency, records, and recaps: Learning and nonequilibrium behavior in a simple decision problem. ACM Trans. Econom. Comput. 4, 4, Article 23 (August 2016), 18 pages.
DOI: http://dx.doi.org/10.1145/2956581

1. INTRODUCTION

Understanding when repeat experience can lead individuals to optimal behavior is crucial to the success of game theory and behavioral economics. Equilibrium analysis assumes that all individuals choose optimal strategies, whereas much research in behavioral economics shows that people often have predictable biases in single-shot decisions [Kahneman and Tversky 2000]. However, many important economic decisions involve repetition and the chance to learn from past mistakes, and this may often alleviate the effects of behavioral biases. Two classes of economic models allow for persistent mistakes: "self-confirming" models allow mistakes only about "off-path" events (e.g., an individual may believe that a particular restaurant serves terrible food, never patronize it, and thus never learn about her mistaken belief). More "behavioral"

This work was supported by the National Science Foundation under grants SES 0951462 and 1258665. The work was completed when A. Peysakhovich was a graduate student and research scientist at Harvard University.
Authors' addresses: D. Fudenberg, MIT, Department of Economics, 77 Massachusetts Ave, E52-418. Cambridge, MA 02139; email: drewf@mit.edu; A. Peysakhovich, Facebook Artificial Intelligence Research, 770 Broadway, New York, NY 10003; email: alex.peys@gmail.com.

DOI: http://dx.doi.org/10.1145/2956581

models suppose that even "on-path" mistakes can persist long enough to be treated as equilibrium phenomena.

We steer a middle course and argue that on-path mistakes do occur and can sometimes persist. Importantly, the well-documented tendency to discount older information (i.e., recency bias [Erev and Haruvy 2013]) plays a key role in determining whether mistakes are permanent or temporary. In particular, recency bias can imply behavior that looks very different from equilibrium predictions even in stationary stochastic environments. We demonstrate this phenomenon in a series of experiments in which participants face a simple stochastic decision problem. A simple learning model with high recency better organizes behavior across experiments than three well-known equilibrium concepts. Additionally, insights gained from the learning model suggest an intervention that markedly improves payoffs and is generalizable to other situations.

Our stochastic decision problem is the simplified version of the classic lemons model [Akerlof 1970], the additive lemons problem (ALP) [Esponda 2008]. In the ALP, a seller is endowed with an object with a value known to the seller and unknown to the buyer, and the object is worth a fixed amount more to the buyer than the seller. The buyer makes a single take-it-or-leave-it offer to the seller, who accepts or rejects it. Buyers receive full feedback (and the corresponding payoff) if their offer is accepted and receive a payoff of 0 and no feedback if their offer is rejected.

The ALP with subjects acting as buyers and computers playing sellers makes a useful laboratory model for studying the persistence of mistakes. The calculation of the buyer's optimal strategy requires individuals to use conditional expectations, something that many people find difficult and/or unintuitive. This often leads to large deviations from optimal behavior the first time individuals play the game. Allowing them to play the ALP repeatedly allows us to examine the persistence of these mistakes and the effects of manipulations on convergence to optimality.

Our baseline ALP lets us generate predictions using Nash equilibrium (NE); cursed equilibrium (CE); behavioral equilibrium (BE); and a simple learning model, temporal difference reinforcement learning (TDRL) [Sutton and Barto 1998].[1] In our first experiment, subjects were randomly assigned to one of four conditions: two payoff structures for the lemons problem (high or low "value added") were crossed with two information conditions. In the "informed" condition, participants were told the prior distribution of seller valuations; in the "uninformed" condition, they were not given any information about this distribution of values other than its support.[2]

These four conditions were chosen so that Nash, cursed, and behavioral equilibrium have clear and distinct predictions. In our first experiment, we find that none of these equilibrium concepts provide a good fit to data. In contrast, a very simple learning model with relatively high recency organizes the aggregate behavioral patterns relatively parsimoniously. In addition, we find direct evidence for recency: individuals react strongly to last period outcomes even after experience with the decision problem.

We then consider a succession of treatments with different feedback structures. In a second experiment, buyers are informed of the object's value regardless of whether their bid is accepted or rejected. This allows us to test whether behavior in the main treatment comes from incorrect expectations about the value of the rejected items. Providing this additional feedback has very little effect, and our qualitative findings are unchanged. Because this treatment makes the information subjects receive exogenous to their actions, it also permits a cleaner test of recency effects, which we again confirm.

[1]To use simulation methods, we need to fix a functional form for the learning model; we chose TDRL for its simplicity.

[2]This latter structure corresponds to the assumptions of Esponda [2008], who argues that it seems a better description of many field settings.

Recency effects are very powerful in the ALP because a single experience with a strategy contains very little information about whether that strategy is successful. Thus, when participants heavily discount the past's information, they are not able to learn the optimal behavior. In our final experiment, we ask whether this discounting of past information is a result of limited memory or of more complicated cognitive constraints.

To answer this, we consider two treatments. In the "more information" condition, participants play the ALP against 10 sellers simultaneously. Each round, buyers make a single-offer decision that applies to all 10 sellers. At the end of a round, participants receive feedback about each of the 10 transactions: what the seller's value was, whether the offer was accepted, and the buyer's profits on that transaction. The "simple information" condition has identical rules. However, instead of receiving fully detailed feedback on each transaction, participants are told their average profit out of the 10 transactions and average values of the objects that they actually purchased.

Providing more information has little effect, but providing the information in the pithy, more readily understood form of averages ("recaps") significantly improves the subjects' payoffs. This suggests that recency effects may not simply be an issue of "memory space" but also the (lack of) computational resources to construct useful summary statistics from multiple pieces of data. Exploring these computational constraints is an important avenue for future research.

2. THEORY

2.1. Nash Equilibrium in the Additive Lemons Problem

To investigate the effects of different information and feedback conditions on learning, payoffs, and convergence or nonconvergence of behavior to optimality, we focus on the ALP as introduced in Samuelson and Bazerman [1985] and further studied by Esponda [2008]. In this game, there are two players: a buyer and a seller. The seller begins with an object of value v drawn from a uniform distribution between 0 and 10; this value is known to the seller but is unknown to the buyer. The buyer makes a single take-it-or-leave-it offer b to the seller. If the seller accepts this offer, the buyer receives the object and pays b to the seller. The object is worth $v+k$ to the buyer, and thus there is a gain from the occurrence of trade.

This game has a unique NE in weakly undominated strategies: it is weakly dominant for the seller to accept all offers below v and reject all offers above v. Because the seller has a dominant strategy, we transform the ALP into a single-person decision for the rest of our study. The buyer's optimization problem is thus

$$\max_b \Pr(v \leq b)[E(v + k \mid v \leq b) - b] = b[k - b/2].$$

Solving the maximization shows that the optimal bid $b_{NE} = k$. Thus, in NE, buyers offer k every round and sellers accept when $v < k$ and reject if $v > k$.

We chose the ALP for several reasons. First, lemons problems are familiar to economists. Second, the ALP is easy to describe to subjects but also tends to elicit suboptimal first responses due to failures of probabilistic reasoning.[3] Additionally, the ALP can be played repeatedly in a short amount of time. We will focus on two payoff conditions: a "low added value" condition where $k = 3$ and a "high added value" condition where $k = 6$.

[3]The key to this failure is that the expectation in the buyer's maximization problem is a *conditional* expectation. To make an optimal decision, the buyer needs to take into account that if a bid of b is accepted, the item's value must lie below v. There is a large amount of experimental evidence that individuals frequently fail to make this correction in many decisions of interest, including common value auctions [Kagel and Levin 1986], the Monty Hall problem [Tor and Bazerman 2003], and strategic voting games [Guarnaschelli et al. 2000].

The ALP is very similar to the Acquire a Company (ACG) game [Samuelson and Bazerman 1985]. The ACG has the same extensive form, but the value to the buyer has the multiplicative form kv instead of the additive form $v+k$ that we consider here. In the ACG, for $k>2$ the optimal bid is 10 and for $k<2$ the optimal bid is 0. There has been a large amount of research on this game showing that when $k<2$, individuals fail to play the optimal strategy, even with learning opportunities [Ball et al. 1991]. However, the fact that the optimal bid is on the boundary is a significant confound here, given the aversion of individuals for corner solutions [Rubinstein et al. 1993]. Our specification of the ALP avoids this confound, as for any value of k the optimal solution is interior.[4]

2.2. Other Equilibrium Concepts

NE requires that each player's strategy is a best response to the true distribution of opponents' play and thus implies that the buyers in the ALP should make the optimal bid. Some alternative equilibrium concepts maintain the assumption that players correctly interpret and process the information that they receive and best respond to this information while allowing players to have incorrect beliefs, provided that those beliefs are consistent with their observations, so that players can only have wrong beliefs "off the equilibrium path" [Battigalli and Guatoli 1997; Dekel et al. 1999, 2004; Fudenberg and Levine 1993]. We focus here on a particular example of such a concept: BE [Esponda 2008].

A variety of behavioral experiments show that mistakes in probabilistic reasoning are fairly common [Samuelson and Bazerman 1985; Rubinstein et al. 1993; Tor and Bazerman 2003; Guarnaschelli et al. 2000; Charness and Levin 2009]. This motivates equilibrium concepts that allow or require individuals to make mistakes in updating beliefs about opponents' play and computing the associated best responses. In particular, CE allows for a specific type of mistake in computing conditional expectations without distinguishing between on-path and off-path errors [Eyster and Rabin 2005].

BE and CE make different predictions of behavior in the ALP. In addition, they suggest different causes for deviations from optimal play. We now discussthese predictions.

2.2.1. Behavioral Equilibrium. The solution concept of BE is developed specifically for the ALP in Esponda [2008]. This concept is meant to model settings where (1) individuals need to learn the distribution of Nature's moves (i.e., values) at the same time that they learn the distribution of opponent's play, and (2) buyers do not see the seller's value when the seller rejects the object.[5]

In our setting, BE can be expressed as a two-tuple (p^*, b_{BE}), where p^* is a probability distribution on the interval [0,10]. BE imposes two conditions on this tuple. First, b_{BE} must be optimal for the buyer given distribution p^* and the belief that sellers play optimal strategies. Second, p^* must be consistent with what buyers observe in equilibrium so that $p^*(A)$ for any subinterval A of the interval $[0, b_{BE}]$ must coincide with the true probability (in this case, uniform) of A.

However, no restrictions are placed on what probabilities p^* may place on the distribution of values that buyers never actually see. Given these two conditions, BE is a set valued solution concept with the property that $b_{BE} \leq b_{NE}$. Thus, BE predicts that

[4]In the online supplemental materials (see http:/alexpeys.github.io), we show the results of an experiment that confirm the presence of corner aversion in our experimental paradigm.

[5]BE allows for two types of agents: naive agents whose beliefs are only required to be self-confirming but do not know the distribution of Nature's move [Dekel et al. 2004] and sophisticated agents who know the payoff functions of the other players, as in rationalizable self-confirming equilibrium [Dekel et al. 1999]. In our formulation of the ALP, the buyers are told the seller's strategy, so the two types of agent are equivalent. When the ALP is formulated as a game, the sophisticated agents deduce that the seller will not accept a price below their value, but the naive agents need not do so.

buyers cannot persistently overbid in either of the payoff conditions that we examine in the ALP: if they were able to do so, they would learn that it would be better to make the NE bid instead. However, buyers can persistently underbid if they have overly pessimistic beliefs about the distribution of values above their bid.

2.2.2. Cursed Equilibrium. Individuals often fail to deal correctly with conditional probabilities. This assumption is built into the equilibrium concept of fully cursed equilibrium (CE) [Eyster and Rabin 2005]. Formally, CE assumes that when individuals in a Bayesian game optimize, they completely ignore the correlation between their other players' types and their strategies.[6]

As in the work of Charness and Levin [2009], we adapt CE by supposing that participants treat the computer as a "player" so that the buyer's maximization problem replaces the conditional expectation of the value v with its unconditional expectation so that the buyer solves

$$max_b \Pr(v \leq b)[E(v) + k - b]$$

and the CE bid is

$$b_{CE} = (5 + k)/2.$$

Note that this leads to overbidding (relative to the best response) if $k < 5$ and underbidding when $k > 5$. Thus, CE predicts overbidding in the low added value conditions ($k = 3$) and underbidding in high added value conditions ($k = 6$).

As noted earlier, the predictions of CE do not depend on whether or not players are told the distribution of Nature's moves or on the sort of feedback that they receive in the course of repeated trials. Another property of CE is that in many games, including the ALP, the payoff that players expect to receive in equilibrium does not match the actual payoffs that they will receive. Thus, to the extent that CE is meant to describe behavior that persists when subjects have experience (as the "equilibrium" part of its name suggests), it implies that individuals have permanently incorrect yet stable beliefs about their expected payoffs.

2.3. Learning Dynamics with Recency

A common argument given for the use of equilibrium analysis is that equilibrium arises as the long-run result of a nonequilibrium learning process [Fudenberg and Kreps 1995, 1998]. However, there is a substantial amount of evidence both from the lab (e.g., Camerer [2003]) and the field (e.g., Agarwal et al. [2008] and Malmandier and Nagel [2011]) that individuals react strongly to recently experienced outcomes and discount past information. Individuals who display such "recency effects" will not converge to using a single strategy in a stochastic environment and thus will be poorly described by an equilibrium model. Therefore, it is interesting to explore the use of learning dynamics to generate predictions in place of an equilibrium concept [Roth and Erev 1995].[7] In general, the details of those distributions depend on the specifics

[6]In applying CE to the lemons problem [Eyster and Rabin 2005], use a refinement to restrict off-path play that is analogous to our assumption that the sellers do not use weakly dominated strategies. The same authors also propose the notion of partially cursed equilibrium, in which beliefs are a convex combination of the fully cursed beliefs and those in the NE.

[7]Recency has been incorporated into both belief-based and reinforcement-based models of learning by adding a parameter that controls the speed of informational discounting (e.g., see Cheung and Friedman [1997], Fudenberg and Levine [1998], Sutton and Barto [1998], Camerer and Ho [1999], and Benaim et al. [2009]). Recency effects have also been modeled by supposing that individuals "sample" a set of experiences either with all experiences in the recent past weighted equally (Fudenberg and Levine [2014] relate this to informational discounting) or with more recent experiences being more likely to be sampled (see Nevo and Erev [2012]). Most of these learning models converge to an ergodic distribution.

of the model, but it is easier to characterize the effect of recency in some limit cases. At one extreme, with very little recency, each of a large number of past outcomes has approximately equal weight; thus, in a stationary decision environment, we expect each individual to obtain close to the optimal payoff.[8] On the other hand, the most extreme case of recency is to play a best response to last period's information. In the ALP, if the seller's value today is expected to be exactly the same as that of yesterday, then the optimal bid equals yesterday's value; this implies that for both the $k = 3$ and $k = 6$ versions of the ALP, the population average bid will be the unconditional expectation of the seller's value, which is 5.[9]

In practice, we do not expect observed behavior to correspond to either of these limits but instead to reflect an intermediate weight on recency, so we would like to know the aggregate implications of such intermediate weights in our two conditions. To get a sense of this, we now specialize to a specific model that is easy to work with—the TDRL model [Sutton and Barto 1998]. This model has a single parameter that controls the rate at which information from past observations is discounted. Although more complex learning models fit various data better, variations of TDRL have been shown to fit human and animal learning behavior reasonably well (Glimcher et al. [2008] provide a survey), and we believe that the qualitative effect of recency on the aggregate distribution of play will be roughly the same for many of the alternative models.

TDRL works as follows: for each action a, the agent begins at time 1 with a valuation $v_1(a)$, which we assume is chosen randomly.[10] In each period t, individuals use a logit choice function, so they choose action a with likelihood proportional to $\exp(Gv_t(a))$. Here, G represents the degree of maximization; note that as G goes to infinity, the probability of the action with the highest value goes to 1, so the choice function approximates maximization, whereas as G goes to 0, all actions are chosen with equal probability.

After each choice, individuals receive feedback and update their valuations. In the case of the ALP, individuals receive different feedback depending on whether their offer is accepted or not. We deal with these cases in turn.

First, suppose that the individual's offer is accepted. The individual then sees the seller's valuation for the object in that round. In our TDRL model, individuals update their valuations for action a according to

$$v_t(a) = v_{t-1}(a) + A(r_{t-1}(a) - v_{t-1}(a)),$$

where $r_t(a)$ is what the payoff would have been from choosing action a in that round. The basic idea is simple: the function $v(a)$ measures the value assigned to action a. The term in parentheses represents the *prediction error*. If it is positive, this means that a did better than expected, and conversely if it is negative, then a did worse than expected; the value $v(a)$ is then incremented upward or downward accordingly. Parameter A is the *learning rate*—the higher it is, the more responsive individuals are to recent rounds.

Note that in our model, individuals generate payoffs and update valuations even for bids that they did not choose. This requires individuals to be able to compute the counterfactual payoffs when information sets are censored. Here, we assume that if individuals bid a and are rejected, they correctly infer that the computer's value v was

[8]This assumes either that the agents play all actions with positive probability, as in smooth fictitious play, or that they receive counterfactual information about the payoffs of the actions that they did not use.

[9]Note that if bids are restricted to be integers, the optimal response is to bid the smallest integer larger than the realized computer value; this predicts an average bid of 5.5 in the ALP.

[10]With high recency, the impact of the initial values dissipates in a few rounds.

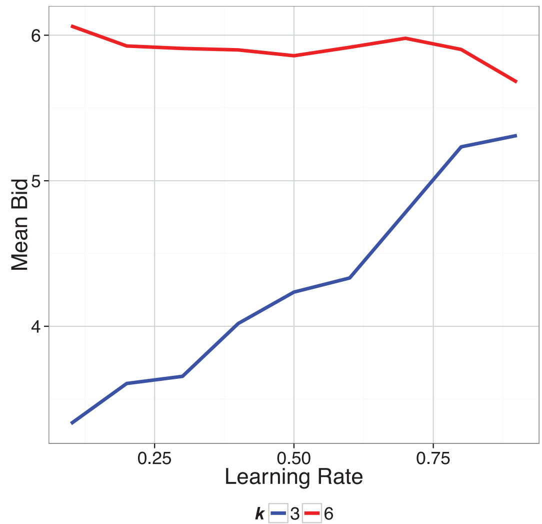


Fig. 1. TDRL mean predicted bid as a function of learning rate and condition.

above their bid a, draw a random value v from the interval $[a, 10]$, and update their valuations as if this hypothetical v were the true computer value.[11]

We simulate $N = 1{,}000$ agents playing 30 rounds of the ALP. Figure 1 shows the average simulated behavior in the final round for two different k values as well as different values of A and $G = 1$.[12] We note that we still see average offers very close to optimal for $k = 6$; however, in the $k = 3$ condition, low levels of A are required for approximately optimal play (see Figure 1).

There are many other learning rules considered in the literature. We believe that any "reasonable" rule with high recency will yield similar results, as in the limit of high recency, any such rule will just best respond to the last period.

3. GENERAL EXPERIMENTAL DESIGN

3.1. Subject Recruitment

All experimental participants were recruited online using the labor market of Amazon Mechanical Turk (MTurk).[13] In all of the following studies, individuals read the instructions for the games and answered a comprehension quiz. Individuals who failed the comprehension quiz were not allowed to participate in the study; reported participant numbers are for those who passed the quiz.[14]

[11]In the case where subjects do observe v even when their bid is rejected, we assume that they update with the observed value of v. Note that this generates exactly the same distributions of behavior. We will see in experiment 2 that providing the subjects with full feedback does not change the distribution of bids.

[12]The simulation results are robust to changes in the noise parameter. Because we are concerned with the first moment of the distribution of bids, changing noise mostly affects the dispersion of bids rather than the mean.

[13]Although one might worry about the lack of control in an online platform, several studies have demonstrated the validity of psychological and economics experiments conducted on the MTurk platform [Simons and Chabris 2012; Peysakhovich and Karmarkar 2015; Peysakhovich and Rand 2015; Rand et al. 2014a; Peysakhovich et al. 2014; Imas 2014]. All recruited subjects were US based.

[14]In our studies, the failure rate on the quizzes was approximately 25%, which is slightly higher than the rate in the preceding studies (typically 10% to 20%). However, our game is much more complicated than simple games such as the one-shot Public Goods game (see online materials for instructions and example comprehension quizzes).

All experiments were incentive compatible: participants earned points during the course of the experiment that were converted into USD. Participants earned a show-up fee of 50 cents and could earn up to $2 extra depending on their performance. All games were played for points, and participants were given an initial point balance to offset potential losses. Experiments took between 10 and 17 minutes, and each subject participated in only one experiment in this series.

4. EXPERIMENT 1

4.1. Design

We recruited $N = 190$ participants to play 30 rounds of the ALP. In each round, participants made a bid, restricted to the integers 0 through 10 to a computerized seller who played the dominant strategy. Participants were informed of the seller's strategy in the instructions. If a participant's bid was accepted, he or she received full feedback about the round including the value of v and their payoff. If a participant's bid was rejected, he or she was informed about this and received no additional information.

We varied two parameters to form four conditions. First, we varied the level of k, setting it equal either to 3 or 6. Second, we varied whether participants were informed of the distribution of seller values. In one case, they were told that v is distributed uniformly between 0 and 10. In the other, they were informed that there was a distribution, but not what it was. This gave us four conditions, which let us "score" the fit of each of the theories discussed earlier. Individuals were randomized into a single condition.

The theories discussed previously give clear hypotheses about what should happen in each of these treatments. CE predicts overbidding when $k = 3$ and underbidding when $k = 6$ (CE predicts bids of 4 and 5.5, respectively). BE is a set-valued solution concept; it rules out overbidding, allows underbidding in the uniformed condition, and predicts optimal bids in the informed treatments. Finally, simulations of the TDRL model predict that with high recency we should see higher than CE overbidding when $k = 3$ and almost NE behavior when $k = 6$. Table I summarizes these hypotheses.

4.2. Results

Figure 2 shows time courses of average offers binned by five-round blocks. Participants behave nearly optimally in $k = 6$ conditions, but even toward the end of the experiment, they behave quite suboptimally in the $k = 3$ conditions. To show that the misoptimization is economically significant, we look at the payoff consequences of these decisions. We define the efficiency of an individual decision as the expected payoff as a percentage of the expected payoff of the optimal strategy. Figure 2 also shows that this misoptimization does affect earned payoffs substantially: average efficiency in the last third of the game is only approximately 10% in the $k = 3$ conditions.[15]

There is no effect of being informed about the true distribution of values either in the averages (regressions show an insignificant effect of being informed; see the supplement) or in the full distribution of behavior in the last round (pooled across k conditions, Kolmogorov-Smirnov test $p = .68$).[16]

All reported statistics that do not include a specific test come from regression analysis with errors clustered at the participant level. We report full regression tables in the

[15]The bid distributions look centered (see online materials for full distributions and additional analyses), and the aggregate overbidding in $k = 3$ conditions reflects overbids by many subjects. In particular, the distribution of behavior does not correspond to a mixture of some subjects optimizing and others choosing at random due to inattention.

[16]One possible explanation of this is that subjects who are not told the true distribution expect it to be uniform distribution.

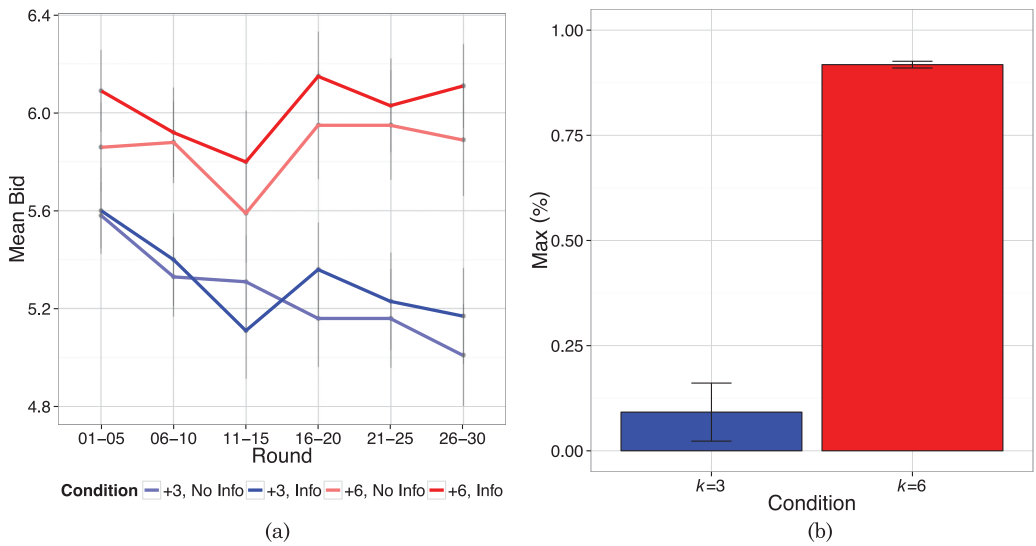


Fig. 2. Results of experiment 1. (a) Average bids by condition smoothed at 5-round blocks. (b) Mean expected payoff of offers in the last 10 rounds by condition.

supplemental materials available on our web sites. In addition, all error bars included in figures include the same cluster corrections.

Bids are higher in the $k = 6$ condition, as would be expected. This is true in even in the first round (t-test $p < .001$). Thus, participants do condition their initial play on this payoff-relevant parameter. To increase statistical power, we now pool across informed and uninformed conditions.

We first focus on the $k = 3$ condition. By the last 10 rounds, *aggregate* behavior appears to have converged: in regressions, the significance of a round number on a bid disappears when we restrict the sample to the last third of the game (mean first 5 rounds = 5.59; mean bid in last 10 rounds = 5.148). In addition, the distribution of bids in the first 10 rounds is significantly different from the distribution of bids in the last 10 rounds (Kolmogorov-Smirnov $p<.001$), but the distribution of behavior in rounds 11 through 20 is not significantly different from the distribution of behavior in rounds 21 through 30 (Kolmogorov-Smirnov $p = .43$). The average bid in the last 10 rounds is significantly above the optimal bid of 3 (clustered 95% confidence interval [4.88, 5.41]). There is a nonsignificant downward trend in bidding behavior in the $k = 3$ uninformed condition; however, this trend does not appear in any of our later experiments, so we attribute it to noise.

In the $k = 6$ condition, we see no experience effects. For symmetry with $k = 3$, we focus on the last 10 rounds. Here, bids are much closer to optimality and are not significantly different from the optimal bid of 6 (mean bid in last 10 rounds = 6.02 with a subject-level clustered 95% confidence interval given by [5.72, 6.27]).

We now turn to evaluating the performance of the theories, beginning with a comparison of the average bids with the predicted averages. There is substantial overbidding at $k = 3$,[17] no underbidding at $k = 6$, and very little difference between the informed and uninformed conditions. Thus, BE does not fit well with our data despite its substantial

[17]This underbidding cannot be explained simply by relaxing individual best response to incorporate random utility: the payoff functions of the ALP are relatively symmetric, so adding a small random utility term does not change the mean bid much. In particular, with logit best replies, the mean bid is below 3.5 for all but the most extreme values of the logit parameter.

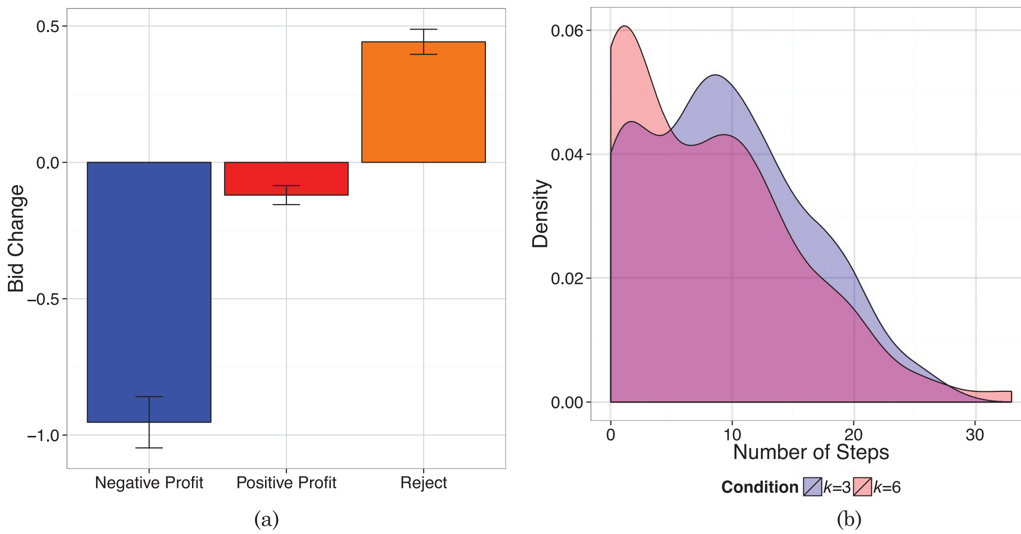


Fig. 3. Learning in experiment 1 persists even in the last 10 rounds. (a) Relationships between outcomes in a round and Δbid. (b) Distribution of steps taken by individuals (smoothed density).

intuitive appeal. The substantial overbidding in the $k = 3$ conditions is qualitatively consistent with CE, but the overbidding is even higher than CE predicts, and significantly so (mean bid in last 10 rounds = 5.148; 95% confidence interval clustered at participant level [4.88, 5.41]).[18] In addition, we do not see the underbidding in the $k = 6$ conditions that CE predicts (mean bid in last 10 rounds = 6.02 with subject clustered 95% confidence interval given by [5.72, 6.27]).

Finally, we turn to the TDRL model. We first discuss whether the model matches patterns in the aggregate data: as in TDRL simulations with high recency, we see that aggregate behavior exhibits extreme overbidding in the $k = 3$ conditions and optimal behavior in the $k = 6$ conditions.

Next, we look at the dynamics of behavior. Because both CE and BE are equilibrium concepts, they make predictions about aggregate behavior once subjects have enough experience/feedback for equilibrium to roughly approximate their behavior. However, these models do not make predictions about how behavior should change between rounds before the equilibrium is reached, and they predict little change in play once subjects have enough experience. In contrast, any learning model with a high weight on recent outcomes predicts there should be nonrandom changes in individual behavior between rounds and that this nonstationarity should continue even when individuals have received feedback on a substantial number of past plays.

To look for this individual-level effect, we define a variable called Δbid as the offer in round t minus the offer in round *t-1*. We then look at how Δbid is affected by what happens in round *t-1*, with the prediction that good outcomes of accepted bids should lead individuals to revise their bid upward, bad outcomes should lead individuals to revise their bids downward, and rejections (which indicate that the computer had a high value that round) should lead individuals to (on average) revise their bid upward. Again, we restrict this analysis to the last third of all rounds, where aggregate behavior has converged.

Figure 3 shows Δbid as a function of outcomes in a last round. We look at three bids: when an individual's bid was accepted and earned a positive profit, when bids were

[18] Eyster and Rabin [2005] find a similar effect when trying to fit CE models to some experimental data.

accepted and yielded a loss, and when bids were rejected. The figure shows that there is strong relationship between the previous period's outcome and Δbid. Regressions (see regression tables in the supplement) confirm the statistical significance of this effect. Additionally, we can look at what happens when an offer is rejected: individuals raise their offer by .405 points (95% confidence interval [.302, .508]) the next round.

Although the TDRL model with high recency describes first-order patterns in the data well, a high recency parameter implies a very strong behavioral response in the next round's offer (1 for 1 in the limit case of extreme recency), and we do not see such a strong response in the individual-level regressions. We could improve the fit of TDRL by adding additional parameters, but we are content to sacrifice in-sample fit for portability and simplicity. TDLR does better than either of the equilibrium concepts at organizing the general patterns in our experiments and can provide intuition about the effects of recency bias on the ALP and other learning scenarios.[19]

To test whether this pattern is driven by a small subset of individuals or is representative, we define a *step* as moving a bid up or down 1 point. We then look at the number of steps that individuals take in the last 10 rounds of the ALP (see Figure 3). If the recency results are driven by a small number of individuals, then we should expect to see a large mass of individuals at 0. If the results are representative, we should expect to see a smaller mass at 0 and most people taking multiple steps. Between 65% ($k = 6$) and 80% ($k = 3$) of participants' offer behavior exhibits persistent variance, even in the last third of experimental rounds. This finding is hard to reconcile with any sort of equilibrium analysis.[20]

In experiment 3, we show that our learning model is also useful in that it suggests particular interventions that can lead individuals closer to optimal behavior. Before turning to this, we present another experiment that tests the robustness of our results and further demonstrates the prevalence of recency-based learning.

5. EXPERIMENT 2: FULL FEEDBACK

The next experiment is designed to control for a potential confound in experiment 1: we saw that the average Δbid in a round in which an offer was accepted was $-.270$, so the observed overbidding primarily occurs due to individuals moving their bid upward after a rejected offer. One potential explanation for this is misperceptions about the value of v conditional on rejection. To check for this, as well as replicate our original results, we performed a second experiment.

5.1. Design

We recruited 75 new participants to play $k = 3$ and $k = 6$ conditions with one twist: whereas in experiment 1 participants simply received a rejected message if their offer was not accepted, participants now received full feedback about the seller's value v whether their offer was accepted or not. Participants were not ex-ante informed of the seller's distribution of values.

[19]It is true that CE/BE/NE are zero-parameter models, whereas TDRL has the learning rate as a free parameter. However, TDRL makes additional predictions about how play changes over time and so is "more falsifiable" than the equilibrium models.

[20]This variance does not appear to decrease during the course of the experiment. The average absolute value of Δbid is 1.02 in the first 10 rounds and .98 in the last 10 rounds; the difference is not significant (two-sided clustered t-test $p = .58$). We can also consider those individuals who show no variance in their behavior in the last 10 rounds. The fact that these individuals' behavior is constant is consistent with some equilibrium notion, but even among this group, there is substantial overbidding in the $k = 3$ condition (see supplemental materials for full histograms). Thus, the assumption of rational Bayesians is a poor fit among this group as well.

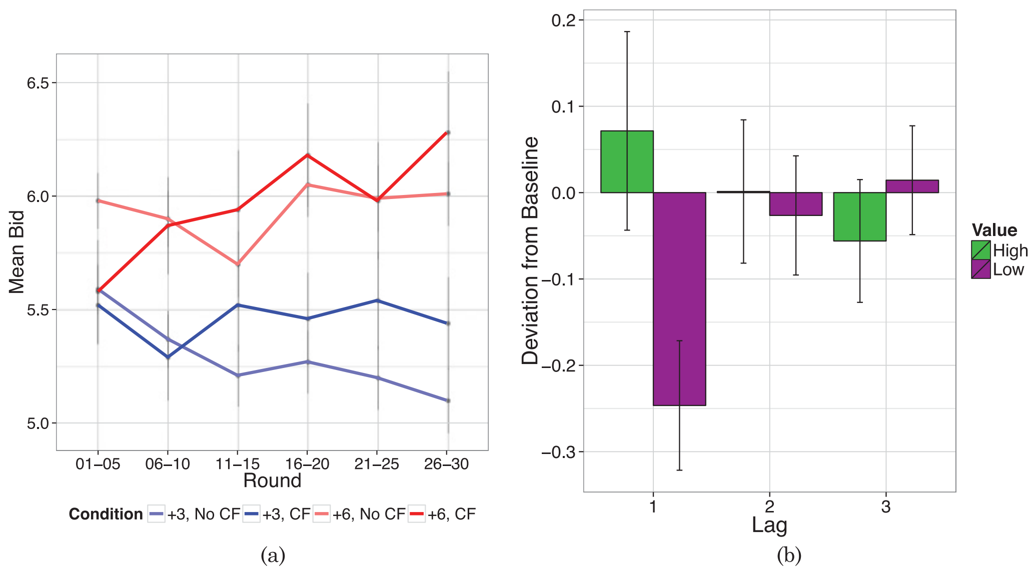


Fig. 4. Results of experiment 2. (a) Counterfactual information does not help individuals optimize. (b) Individuals respond to experiencing high/low outcomes in round *t-1* but much less so to experiencing high/low outcomes in rounds *t-2* or *t-3*.

5.2. Results

Comparing the data from the full feedback experiment to the behavior from experiment 1, we see little difference between behaviors of individuals who have counterfactual information versus those who do not (Figure 4). If anything, the individuals with counterfactual information do slightly worse (overbid more) in the $k = 3$ condition, but this difference is not significant in regressions (see supplemental material). As before, the aggregate outcomes are not driven by outliers, and there is large volatility in most individuals' behavior (see supplemental materials).

The full feedback experiment lets us perform a reduced form test of recency effects. In the baseline experiment, information that individuals received was partially endogenous (high bids were much more likely to get accepted). However, with full feedback, the computer's value v acts like an exogenous shock in round t. In a monotone learning model, higher values of v increase potential valuations of higher bids and low values of v decrease valuations of higher bids. Thus, we expect a monotone relationship between bids at time t and histories of observed computer values v.[21]

We can see a recency effect very starkly even in the last 10 rounds. We first split realized computer values into very low (values of 3 or below, bottom 30% of realizations) or very high (values of 7 or bigger, top 30% of realizations). We then take the average bid of each individual over the last 10 rounds and set that as the individual's "baseline." We then look for the effect of observing a very high or very low value in round t on

[21]To see the intuition behind this claim, suppose that an individual has observed a history of computer values $(c_1, c_2, \ldots, c_{t-1})$. The individual thus has valuations for each bid b and an associated expected bid that can be easily derived from the logit choice formulation. Now, consider marginally changing an observed valuation c_i upward. This will increase the valuation of each bid $b > c_i$ and not affect the valuation of each bid $b < c_i$ (because the value of that bid in that period was 0 anyway). Thus, this will increase the expected bid of the individual.

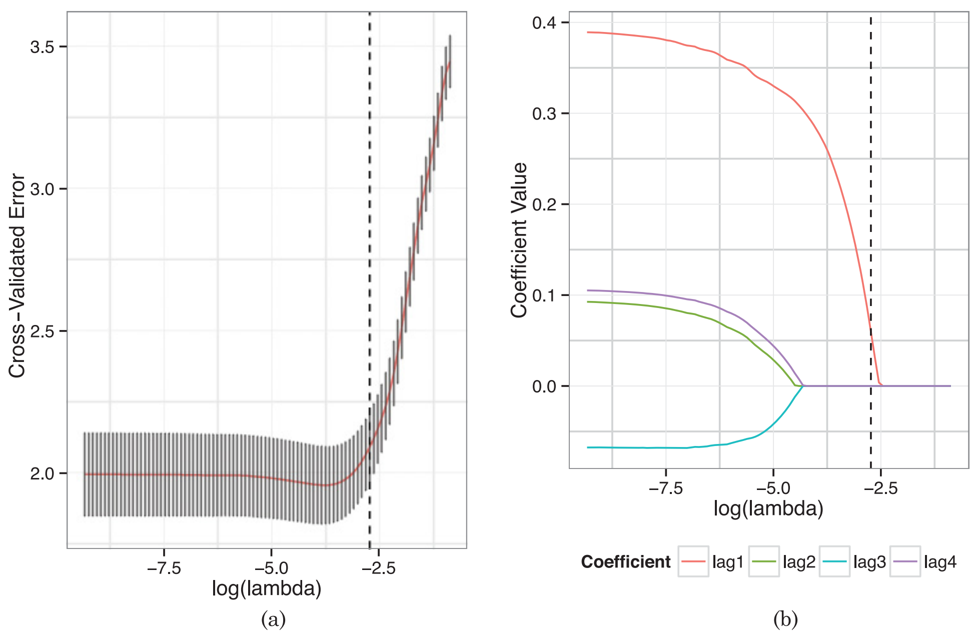


Fig. 5. (a) Cross validation selects a moderate penalty level for the lasso model. Dotted lines indicate the lambda chosen by the 1 standard deviation criterion. (b) In the model trained with this penalty level, only a single lag is selected to be included in the feature set. Even at low levels of regularization, lags beyond the first have very little influence on the model's predictions.

round *t+1, t+2* and *t+3* deviations from this average bid. Here, a positive deviation represents a higher than average (for that individual) bid, and a negative deviation represents a lower than average bid.

Figure 4 shows there is a large effect on behavior in the *t+1* round and no statistically appreciable effect on the *t+2* or *t+3* round.

An alternative way to quantify recency effects is to regress bids at time *t* on lagged experiences and then use some feature selection criterion to choose an "optimal" number of lags. If there are strong recency effects, the selected feature set should include a relatively small number of lags.

To perform feature selection, we use L_1 regularized linear regression or "lasso." In this procedure, the linear regression coefficients are chosen to minimize the sum of the squared residuals plus a coefficient times the sum of the absolute values of the coefficients (i.e., a penalty for model complexity). We choose the penalty level using cross validation and the "one standard deviation criterion" [Tibshirani 1996; Hastie et al. 2009].

Because we are interested in predicting individual changes in strategy in response to new information, we include individual-level dummies to capture an individual's average strategy. In addition, we do our cross validation at the round level (thus, each fold of the cross validation is one bid for each individual, and because we use the last 10 rounds of behavior, we are left with 10 folds).

Figure 5 shows the results of our analysis. Cross validation selects a model involving only a single lag of information as the best model for predicting individual bid behavior in the last 10 rounds. Even with a very small regularization penalty, information beyond the first round is not selected. Other forms of model selection (e.g., the Bayesian information criterion; see the online appendix for a fuller description) also implicate a single lag model. This provides further evidence for a strong recency bias.[22]

6. EXPERIMENT 3: RECAPS

So far, we have compared the predictions of various equilibrium concepts with that of a simple learning model and have shown that learning with recency better organizes existing data. In addition to organizing existing data, good models help us gain intuition about situations that we have not yet seen and give us the ability to design welfare-improving interventions. We now show that considering the implication of recency for learning delivers insights that equilibrium models do not.

Why does recency bias imply that suboptimal behavior should persist in our experiments? The ALP's feedback structure is such that a relatively small sample of outcomes typically does not reveal the optimal bid. Thus, high recency acts as a barrier toward learning optimal behavior in this setting. This suggests a prescription for intervention: increasing the number of outcomes that subjects observe simultaneously should help them make better decisions. To test this hypothesis, we performed another experiment.

6.1. Design

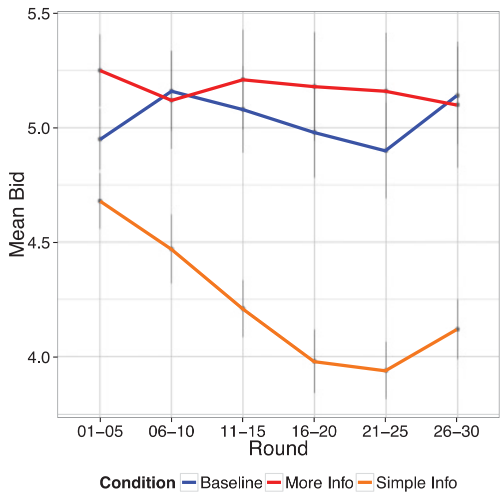
We recruited $N = 273$ more participants. In experiment 3, participants were assigned to one of three ALP conditions, all with $k = 3$. The control condition simply replicated the $k = 3$ condition from experiment 2. In the more information condition, participants played the ALP against 10 sellers simultaneously. Sellers' object values were determined independently. Each round (of 30) buyers made a single-offer decision that applied to all 10 sellers (who, as before, played the optimal strategy). Participants were informed of all of this. At the end of a round, participants received feedback about each of their transactions simultaneously: what the seller's value was, whether the offer was accepted, and the buyer's profits on that transaction.

There is much existing evidence that in addition to having limited memory, individuals also have limited computational resources [Miller 1956]. Thus, one may expect that more information is only useful if it is in easily "digestible" form. To look for evidence of computational constraints, we added a simple information condition. This condition was almost identical to the more information condition; individuals played 30 rounds with 10 sellers simultaneously and made a single offer that applied to each seller. However, instead of receiving fully detailed feedback on each transaction, participants received pithy recaps: they were told their average profit out of the 10 transactions and average values of the objects that they actually purchased (see the online appendix for examples of feedback screens).

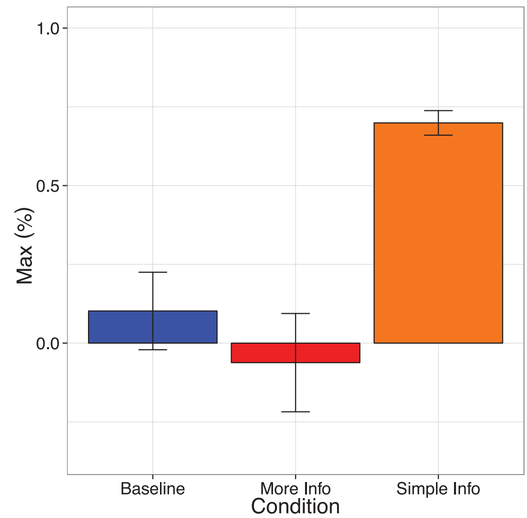
6.2. Results

Figure 6 shows the average offers in the experiment binned in five-round increments. We see that the addition of more information does not seem to help individuals converge to optimal behavior (round 21 through 30 mean offer in control = 5.02; mean offer in more information = 5.22). However, simple information in the form of pithy recaps does appear to be useful (round 21 through 30 mean offer = 4.02), which suggests that the learning problems caused by recency do not stem solely from limited memory.

[22]In addition to recency effects, research on learning and memory also identifies a primacy effect: first or initial experiences tend to be recalled more vividly and hence have a large impact on behavior later [Erev and Haruvy 2013]. To test for the effects of primacy on behavior. we remove the subject-level fixed effects from our regressions and add a term for "computer value observed in first round" to the regression while, as earlier, retaining one lagged computer value. Doing so, we find that the effect of the first-round experience is insignificant, so primacy effects do not seem to be important here.



(a)



(b)

Fig. 6. (a) Average bids by condition smoothed at 5-round blocks. (b) Mean expected payoff of offers in the last 10 rounds by condition.

This latter finding is consistent with that of Bereby-Meyer and Grosskopf [2008], who find that recaps are helpful in their version of the ACG, although they do not compare recaps with the more information condition.

Although individuals in the simple information condition still offer above the NE offer in the final round (one-sided t-test $p < .01$), they perform significantly better in economic terms (see Figure 5) than in the control and more information treatments. In addition to comparing sample averages, we can also see what effect the simple information condition has on the full distribution of behavior: we see that our treatment seems to drive the whole distribution of bids toward the optimum and decreases their variance (test for equal variance in the last 10 rounds of baseline versus the simple information condition rejects equal variance, $p < .001$; additional analyses are available on the authors' Web sites).

We also use the control condition to test whether many participants do not understand and/or are not paying much attention to the experiment. We see an effect of the payoff parameter k in each of our experiments, even in the first round, and these effects go in the expected direction (higher k leads to higher bids), so at least some of our respondents did pay attention. We can further investigate whether inattention is driving our results by using data on response times. (We did not include a response time measure for each individual decision in experiments 1 and 2, but we do include one here.) In the control condition, the average participant spends 4.21 seconds on each decision screen (bottom decile of participants spend on average 2.42; top decile of participants spend on average 5.8). We also see no correlation between average response time and bids, earnings, or bid variance across experiments (linear regression clustered on the participant gives $p = .7$, .55, and .8, respectively). Note that this also implies that response time does not predict the extent of recency bias, as higher recency bias mechanically correlates negatively with earned payoff.

7. CONCLUSION

Our results demonstrate that explicitly dynamic models of behavior can yield insights in ways that equilibrium models cannot. None of the equilibrium concepts (NE/CE/BE) that we consider are able to capture the full variation of behavior in the ALP. By contrast, a learning model with high recency fits aggregate behavior across treatments well. In addition, considering the dynamics of the learning process gives us intuition about interventions via feedback structure to help nudge individual behavior closer to optimum.

Our experiments show that computational, not just memory, constraints may contribute to the persistence of suboptimal behavior. This reinforces earlier arguments that recency effects are in part driven by computational constraints [Hertwig and Pleskac 2010] and suggests that here, at least memory load is not a primary driver of recency. Our results thus support incorporating more accurate representations of computational limits and other forms of bounded rationality into existing learning models.

There is a debate about whether findings from learning experiments such as ours can be applied to understand behavior in the field. Individuals may have computational constraints, but in the field they often have access to technological aids. This is argued strongly in Levine [2012]:

> Even before we all had personal computers, we had pieces of paper that could be used not only for keeping track of information—but for making calculations as well. For most decisions of interest to economists these external helpers play a critical role.

Such technologies can provide recaps and thus help to guide individuals toward optimal decisions. On the other hand, there is evidence of significant economic costs due to incomplete learning and recency bias in contexts such as credit card late fees [Agarwal et al. 2008], stock market participation [Malmendier and Nagel 2011], and IPO investment [Kaustia and Knüpfer 2008]. These findings suggest that even when recaps and record-keeping devices are available, they may not be utilized. At this point, however, the case seems far from settled. Further studies of the effectiveness of recaps in the lab and the field could have both scientific and social benefit.

There has been a recent explosion of interest in combining insights from economics with those from behavioral science to build better working institutions in the real world (e.g., Johnson and Goldstein [2003], Roth [2002], Kraft-Todd et al. [2015], Shafir [2013], Rand et al. [2014b], and Ouss and Peysakhovich [2015]). It is our hope that our results can contribute to this important conversation.

ACKNOWLEDGMENTS

We thank Douglas Bernheim, Gary Charness, Ignacio Esponda, Armin Falk, Simon Gaechter, Ed Glaeser, Moshe Hoffman, David Laibson, David K. Levine, Sendhil Mullainathan, Matthew Rabin, David G. Rand, Ran Spiegler, Andrei Shleifer, Josh Schwartzstein, and Shoshana Vasserman for helpful comments and the Harvard Program for Evolutionary Dynamics for its hospitality.

REFERENCES

S. Agarwal, J. C. Driscoll, X. Gabaix, and D. Laibson. 2008. *Learning in the Credit Card Market*. National Bureau of Economic Research, Cambridge, MA.

G. A. Akerlof. 1970. The market for “lemons”: Quality uncertainty and the market mechanism. *Quarterly Journal of Economics* 84, 488–500.

S. B. Ball, M. H. Bazerman, and J. S. Carroll. 1991. An evaluation of learning in the bilateral winner’s curse. *Organizational Behavior and Human Decision Processes* 48, 1–22.

P. Battigalli and D. Guatoli (Eds.). 1997. Conjectural equilibria and rationalizability in a game with incomplete information. In *Decisions, Games and Markets*. Kluwer, Norwell, MA.

M. Benaïm, J. Hofbauer, and E. Hopkins. 2009. Learning in games with unstable equilibria. *Journal of Economic Theory* 144, 1694–1709.

Y. Bereby-Meyer and B. Grosskopf. 2008. Overcoming the winner's curse: An adaptive learning perspective. *Journal of Behavioral Decision Making* 21, 15–27.

C. Camerer. 2003. *Behavioral Game Theory: Experiments in Strategic Interaction*. Princeton University Press, Princeton, NJ.

C. Camerer and T. Ho. 1999. Experience-weighted attraction learning in normal form games. *Econometrica* 67, 827–874.

G. Charness and D. Levin. 2009. The origin of the winner's curse: A laboratory study. *American Economic Journal: Microeconomics* 1, 207–236.

Y. Cheung and D. Friedman. 1997. Individual learning in normal form games: Some laboratory results. *Games and Economic Behavior* 19, 46–76

E. Dekel, D. Fudenberg, and D. K. Levine. 1999. Payoff information and self-confirming equilibrium. *Journal of Economic Theory* 89, 165–185.

E. Dekel, D. Fudenberg, and D. K. Levine. 2004. Learning to play Bayesian games. *Games and Economic Behavior* 46, 282–303.

I. Erev and E. Haruvy. 2013. Learning and the economics of small decisions. In *The Handbook of Experimental Economics* (2nd ed.), A. E. Roth and J. Kagel (Eds.). Forthcoming.

I. Esponda. 2008. Behavioral equilibrium in economies with adverse selection. *American Economic Review* 98, 1269–1291.

E. Eyster and M. Rabin. 2005. Cursed equilibrium. *Econometrica* 73, 1623–1672.

D. Fudenberg and D. Kreps. 1995. Learning in extensive games, I: Self-confirming equilibria. *Games and Economic Behavior* 8, 20–55.

D. Fudenberg and D. K. Levine. 1993. Self-confirming equilibrium. *Econometrica* 61, 523–545.

D. Fudenberg and D. K. Levine. 1998. *The Theory of Learning in Games*. MIT Press, Cambridge MA.

D. Fudenberg and D. K. Levine. 2014. Learning with recency bias. *Proceedings of the National Academy of Science* 111, 10826–10829.

P. W. Glimcher, E. Fehr, C. Camerer, and R. A. Poldrack. 2008. *Neuroeconomics: Decision Making and the Brain*. Academic Press, Cambridge, MA.

S. Guarnaschelli, R. D. Mckelvey, and T. R. Palfrey. 2000. An experimental study of jury decision rules. *American Political Science Review* 94, 407–423.

T. Hastie, R. Tibshirani, and J. Friedman. 2009. *The Elements of Statistical Learning*. Springer New York.

R. Hertwig and T. J. Pleskac. 2010. Decisions from experience: Why small samples? *Cognition* 115, 225–237.

A. Imas. 2014. The Realization Effect: Risk-Taking After Realized Versus Paper Losses. Available at http://papers.ssrn.com/sol3/papers.cmf?abstract_id=2403865.

E. J. Johnson and D. G. Goldstein. 2003. Do defaults save lives? *Science* 302, 1338–1339.

J. H. Kagel and D. Levin. 1986. The winner's curse and public information in common value auctions. *American Economic Review* 76, 894–920.

D. Kahneman and A. Tversky (Eds.). 2000. *Choices, Values, and Frames*. Cambridge University Press, Cambridge, MA.

M. Kaustia and S. Knüpfer. 2008. Do investors overweight personal experience? Evidence from IPO subscriptions. *Journal of Finance* 63, 2679–2702.

G. Kraft-Todd, S. Bhanot, E. Yoeli, and D. Rand. 2015. Promoting cooperation in the field. *Current Opinion in Behavioral Sciences* 3, 96–101.

D. K. Levine. 2012. *Is Behavioral Economics Doomed? The Ordinary Versus the Extraordinary*. Open Book Publishers, Cambridge, UK.

U. Malmendier and S. Nagel. 2011. Depression babies: Do macroeconomic experiences affect risk taking? *Quarterly Journal of Economics* 126, 373–416.

G. A. Miller. 1956. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review* 63, 81–97.

I. Nevo and I. Erev. 2012. On surprise, change, and the effect of recent outcomes. *Frontiers in Psychology* 3, 24.

A. Ouss and A. Peysakhovich. 2015. When Punishment Doesn't Pay: 'Cold Glow' and Decisions to Punish. Available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2247446

A. Peysakhovich and U. Karmarkar. 2015. Asymmetric Effects of Favorable and Unfavorable Information on Decision-Making Under Ambiguity. Available at http://pubsonline.informs.org/doi/abs/10.1287/mnsc.2015.2233

A. Peysakhovich, M. Nowak, and D. Rand. 2014. Humans display a "cooperative phenotype" that is domain general and temporally stable. *Nature Communications* 5, Article No. 4939.

A. Peysakhovich and D. Rand. 2015. Habits of Virtue: Creating Norms of Cooperation and Defection in the Laboratory. Available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2294242

D. G. Rand, A. Peysakhovich, G. Kraft-Todd, G. Newman, O. Wurzbacher, M. A. Nowak, and J. Greene. 2014a. Social heuristics shape intuitive cooperation. *Nature Communications* 5, Article No. 3677.

D. G. Rand, E. Yoeli, and M. Hoffman. 2014b. Harnessing reciprocity to promote cooperation and the provisioning of public goods. *Policy Insights from the Behavioral and Brain Sciences* 1, 1, 263–269.

A. E. Roth. 2002. The economist as engineer: Game theory, experimentation, and computation as tools for design economics. *Econometrica* 7, 4, 1341–1378.

A. E. Roth and I. Erev. 1995. Learning in extensive-form games: Experimental data and simple dynamic models in the intermediate term. *Games and Economic Behavior* 8, 164–212.

A. Rubinstein, A. Tversky, and D. Heller. 1993. *Naive Strategies in Zero-Sum Games*. Sackler Institute of Economic Studies, Ramat Aviv, Tel-Aviv.

W. F. Samuelson and M. H. Bazerman. 1985. Negotiation under the winner's curse. *Research in Experimental Economics* 3, 105–138.

D. J. Simons and C. F. Chabris. 2012. Common (mis)beliefs about memory: A replication and comparison of telephone and Mechanical Turk survey methods. *PLoS One* 7, 12, E51876.

E. Shafir (Ed.): 2013. *The Behavioral Foundations of Public Policy*. Princeton University Press, Princeton, NJ.

R. S. Sutton and A. G. Barto. 1998. *Reinforcement Learning: An Introduction*. Cambridge University Press, Cambridge, UK.

R. Tibshirani. 1996. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 267–288.

A. Tor and M. H. Bazerman. 2003. Focusing failures in competitive environments: Explaining decision errors in the Monty Hall game, the Acquiring a Company problem, and multiparty ultimatums. *Journal of Behavioral Decision Making* 16, 353–374.

Received May 2015; revised July 2015; accepted October 2015