Recency, Records and Recaps: Learning and Non-equilibrium Behavior in a Simple Decision Problem*

DREW FUDENBERG, Harvard University
ALEXANDER PEYSAKHOVICH, Facebook

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.

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

Author Keywords & Phrases: Learning, behavioral economics, recency, equilibrium concepts

1. INTRODUCTION

Understanding when repeat experience can lead individuals to optimal behavior is crucial for the success of game theory and behavioral economics. Equilibrium analysis assumes all individuals choose optimal strategies while much research in behavioral economics shows that people often have predictable biases in single-shot decisions [Kahneman & 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 (for example: an individual may believe a particular restaurant serves terrible food, never patronize it and thus never learn about her mistaken belief). More "behavioral" 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 & 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 experimentally.

This work is supported by the National Science Foundation under grants SES 0951462 and 1258665.
Author's addresses: D. Fudenberg, Economics Department, Harvard University; A. Peysakhovich, Data Science, Facebook.

EC '14, June 08 - 12 2014, Stanford , CA, USA
http://dx.doi.org/10.1145/2600057.2602872

*Both authors contributed equally to this work.

We conduct 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 this model suggest an intervention that markedly improves payoffs and is generalizable to other situations.

Our stochastic decision problem is the simplified version of 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 if their offer is accepted but receive 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 is generally unintuitive for many individuals, and leads to large deviations from optimal behavior in the first round. Allowing individuals 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 & 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 neither Nash equilibrium, behavioral equilibrium, nor cursed equilibrium fit well with the 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 in [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 ten 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 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 non-convergence of behavior to optimality, we focus on the *additive lemons problem* (ALP) as introduced in [Samuelson & 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, thus there is a gain from the occurrence of trade.

This game has a unique Nash equilibrium 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

$$\max_b P(b \geq v)\ (E(v \mid b \geq v) + k - b)$$

Solving the maximization shows that the optimal bid $b_{NE} = k$. Thus, in Nash equilibrium 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 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 $k=6$.

The ALP is very similar to the Acquire a Company Game (ACG; [Samuelson & 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 which shows 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

Nash equilibrium requires that each player's strategy is a best response to the true distribution of opponents' play, and so 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 they receive and best respond to this information, while allowing players to have incorrect beliefs provided those beliefs are consistent with their observations, so that players can only have wrong beliefs "off the equilibrium path" [Battigalli & Guatoli 1997, Dekel et al. 1999, 2004, Fudenberg & Levine 1993]. We focus here on a particular example of such a concept: behavioral equilibrium [Esponda 2008].

A variety of behavioral experiments show that mistakes in probabilistic reasoning are fairly common [Samuelson & Bazerman 1985, Rubinstein et al. 1993, Tor & Bazerman 2003, Guarnaschelli et al. 2000, Charness & 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, cursed equilibrium allows for a specific type of mistake in computing conditional expectations, without distinguishing between on-path and off-path errors [Eyster & Rabin 2000].

Behavioral equilibrium and cursed equilibrium make different predictions of behavior in the ALP. In addition, they suggest different causes for deviations from optimal play. We now discuss these predictions.

[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 & Levin 1986], the Monty Hall problem [Tor & Bazerman 2003] and strategic voting games [Guarnaschelli et al. 2000].

[4] In the appendix we show the results of an experiment that confirm the presence of corner aversion in our experimental paradigm.

Behavioral Equilibrium

The solution concept of behavioral equilibrium (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 don't 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 of 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 buyers cannot persistently overbid in either of the payoff conditions we examine in the ALP: Were they were 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.

Cursed Equilibrium

Individuals often fail to deal correctly with conditional probabilities. This assumption is built into the equilibrium concept of fully cursed equilibrium (CE) [20]. 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 [Charness & Levin 2009] we adapt CE by supposing that participants treat the computer as a "player," so that buyer's maximization problem replaces the conditional expectation of the value v with its *unconditional expectation*

$$\max_b P(b \geq v)\,(E(v) + k - b)$$

and the cursed equilibrium bid is

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

[5] BE allows for two types of agents: naïve 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 type 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 naïve agents need not do so.

[6] In applying cursed equilibrium to the lemons problem [Eyster & 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 Nash equilibrium.

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 above, 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 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 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 non-equilibrium learning process [Fudenberg & Kreps 1995, 1998]. However, there is a substantial amount of evidence both from the lab [Camerer 2003] and the field [Agarwal et al. 2008, Malmandier & 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 so will be poorly described by an equilibrium model. Thus it is interesting to explore the use of learning dynamics to generate predictions in place of an equilibrium concept [Roth & Erev 1995].[7] In general the details of those distributions depend on the specifics 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, so 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 yesterday's, 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

[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 (see e.g. [Cheung & Friedman 1997], [Fudenberg & Levine 1998], [Sutton & Barto 1998], [Camerer & Ho 1999], [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 & Levine 2013] relate this to informational discounting) or with more recent experiences being more likely to be sampled ([Nevo & Erev 2012]). Most of these learning models converge to an ergodic distribution.

[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 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.

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 temporal difference reinforcement learning model (TDRL; [Sutton & 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] provides a survey) 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, while 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 *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 above their bid *a*, draw a random value *v* from the interval [*a*, 10] and update their valuations as if this hypothetical *v* was the true computer value. [11]

We simulate N=1000 agents playing 30 rounds of the ALP. Figure 1 shows the average simulated behavior in the final round for two different *k* values

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

[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.

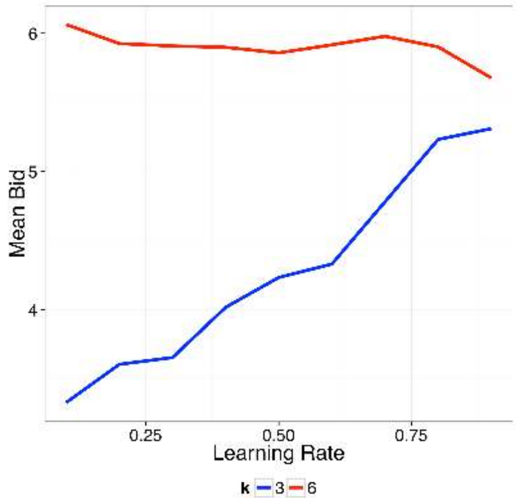

Figure 1: TDRL mean predicted bid as a function of learning rate and condition.

as well as different values of A and G=1.[12] We note that even at we still see average offers very close to optimal for *k=6* but in the *k=3* condition low levels of *A* are required for approximately optimal play.

3. GENERAL EXPERIMENTAL DESIGN

Subject Recruitment

All of the experimental participants were recruited online using the labor market Amazon's 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]

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.

[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 and 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, a number of studies have demonstrated the validity of psychological and economics experiments conducted on the MTurk platform (see [Simons & Chabris 2012, Peysakhovich & Karmarkar 2013, Peysakhovich & Rand 2013, Rand et al. *forthcoming*]). 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 studies above (typically 10-20%). 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).

4. EXPERIMENT 1

Design

We recruited N=190 participants to play 30 rounds of the additive lemons problem. 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, they received full feedback about the round including the value of *v* and their payoff. If a participant's bid was rejected, they were informed about this and received no additional information.

We varied two parameters to form 4 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 4 conditions, which let us "score" the fit of each of the theories discussed above. Individuals were randomized into a single condition.

The theories discussed above 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 with high recency higher than CE-overbidding in aggregate when *k=3* and almost NE behavior when *k=6* treatments. Table 1 summarizes these hypotheses.

Table 1: Hypotheses by condition in the Baseline Experiment

K=3, Informed			*K=3, Uninformed*		
CE	BE	TDRL	CE	BE	TDRL
> NE	NE	> NE	> NE	≤ NE	> NE
K=6, Informed			*K=6, Uninformed*		
CE	BE	TDRL	CE	BE	TDRL
< NE	NE	Close to NE	< NE	NE	Close to NE

Results

Figure 2 shows time courses of average offers binned by 5 round blocks. Participants behave nearly optimally in *k=6* conditions but even towards 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 1/3rd of the game is only approximately 10% in the *k=3* conditions.[15]

[15] The overall bid distributions look are centered (see supplemental materials), and the aggregate overbidding in *k=3* conditions reflects overbids by many subjects. In particular, the distribution of

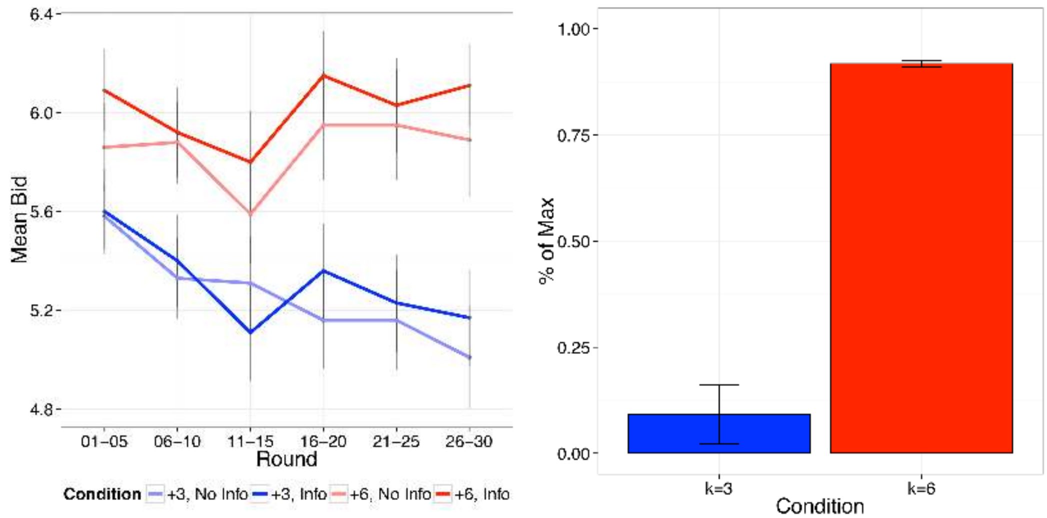

Figure 2: Results of experiment 1. A) Average bids by condition smoothed at 5 round blocks. B) Mean expected payoff of offers in last 10 rounds by condition.

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 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 supplemental materials available on our websites. 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 round number on bid disappears when we restrict the sample to the last 1/3 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 11-20 is not significantly different from the distribution of behavior in rounds 21-30 (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]).

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 subject-level clustered 95% confidence interval given by [5.72, 6.27]).

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.

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 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 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 *non-random* changes in individual behavior between rounds and that this non-stationarity 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 1/3 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 bins: when an individual's bid was accepted and earned a positive profit, when bids were 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 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]) next round.

[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.

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

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]

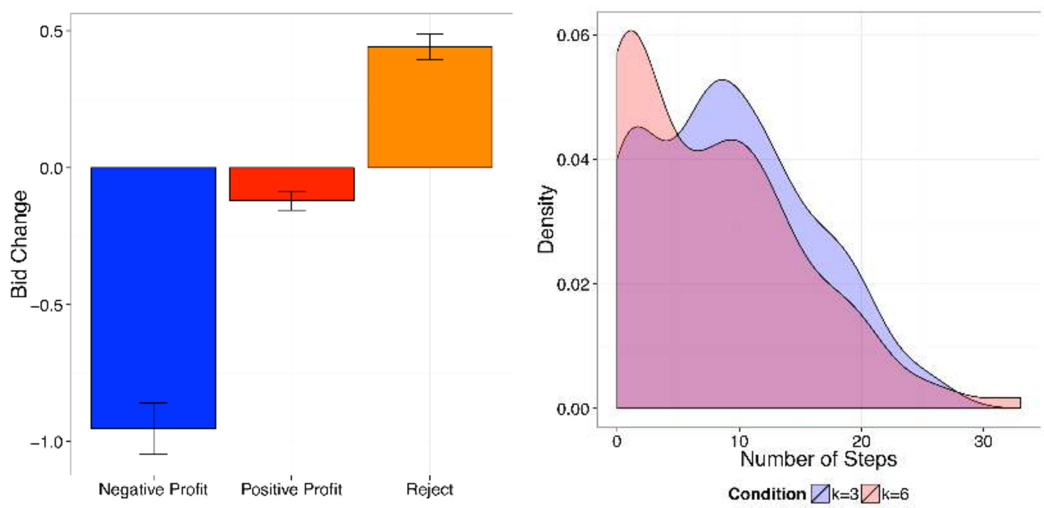

Figure 3: Learning in experiment 1 persists even in the last 10 rounds. A) Relationships between outcomes in a round and Δbid. B) Distribution of steps taken by individuals in each experimental condition (smoothed density).

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 (Figure 3). If the recency results were 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 1/3 of experimental rounds. This finding is hard to reconcile with any sort of equilibrium analysis.[20]

[19] It is true that CE/BE/NE are zero-parameter models while TDRL has the learning rate as a free parameter. However, TDRL makes additional predictions about how play changes over time 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). Thus the assumption of rational Bayesians is a poor fit among this group as well.

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.

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.

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

Results

Comparing the data from experiment FF (Full Feedback) to the behavior from experiment 1, we see little difference between behaviors of individuals who have counterfactual information vs. 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 on authors' websites).

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

[21] To see the intuition behind this claim, suppose that an individual has observed a history of computer values (c_1, c_2, ..., 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.

very high or very low value in round *t* on 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*st round, and no statistically appreciable effect on the *t+2nd* or *t+3*rd round.

An alternative way to quantify recency effects is to regress bids at time *t* on lagged experiences then use a model selection criterion to choose an "optimal" number of lags. If there are strong recency effects, the selected model should use a relatively small number of lags. Using either the Bayesian Information Criterion or lasso (see authors' websites) selects a regression model with a single lag. This provides further evidence for strong recency-biased learning in our participants.

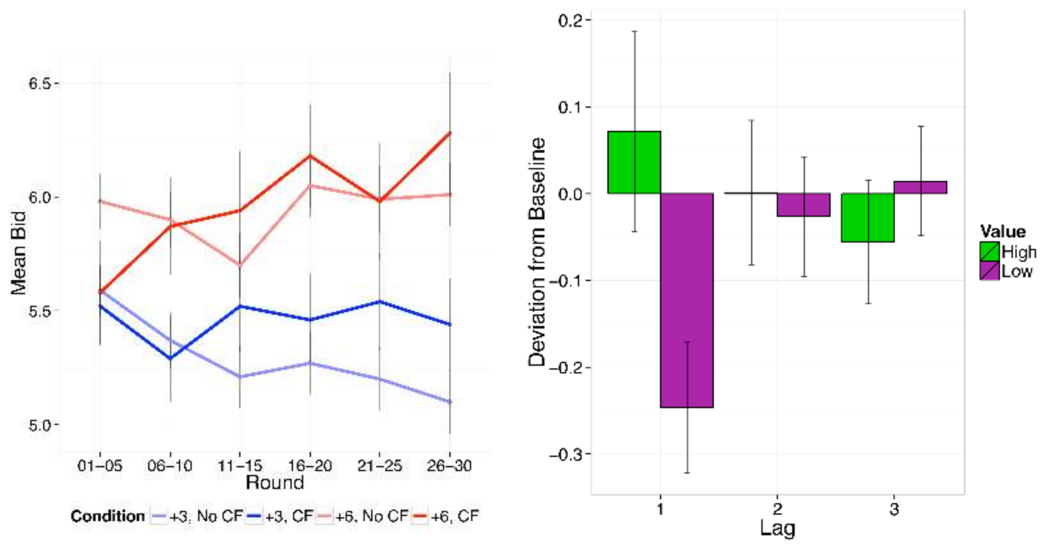

Figure 4: Results of experiment 2. A) Counterfactual information does not help individuals optimize. B) Individuals respond to experiencing high/low outcomes in round *t-1* but much less so to experiencing high/low outcomes in rounds *t-2* or *t-3*.

5. EXPERIMENT 3 : RECAPS

So far we have compared the predictions of various equilibrium concepts with that of a simple learning model and shown that learning with recency better organizes existing data. However, we do not build models just to organize existing data, rather good models help us gain intuition about situations we have not yet seen and give us the ability design welfare improving interventions. We now show that considering the dynamics of 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 doesn't reveal the optimal bid. Thus high recency acts as a barrier towards learning optimal behavior in this setting. This suggests a prescription for intervention: increasing the number of outcomes subjects observe simultaneously should help them make better decisions. To test this hypothesis, we performed another experiment.

Design

We recruited N=273 more participants. In experiment 3 participants were assigned to one of 3 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 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 they actually purchased (see online appendix for examples of feedback screens).

Results

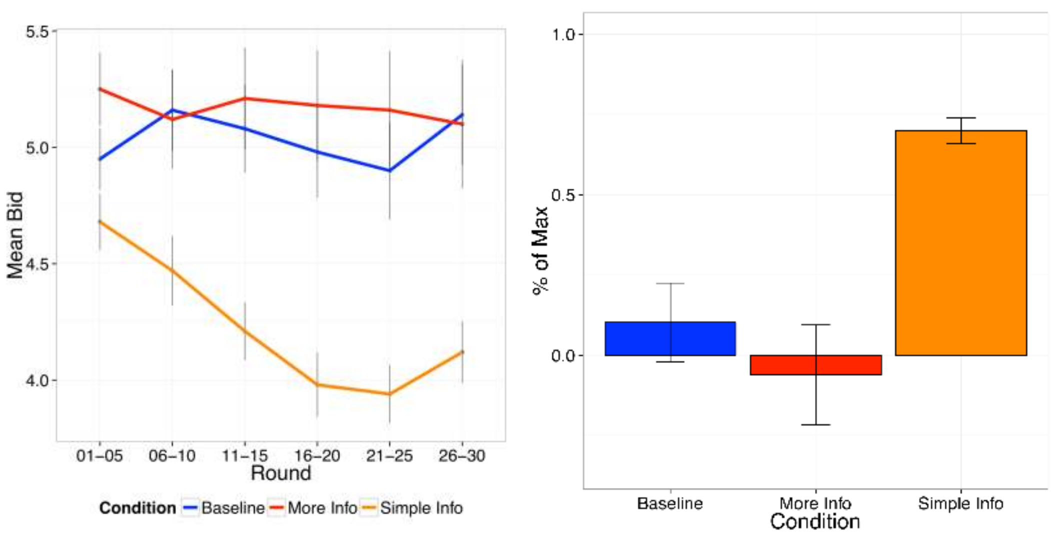

Figure 5: A) Average bids by condition smoothed at 5 round blocks. B) Mean expected payoff of offers in last 10 rounds by condition.

Figure 5 shows the average offers in the experiment binned in 5 round increments. We see that the addition of more information doesn't seem to help individuals converge to optimal behavior (round 21-30 mean offer in control =

5.02, mean offer in *more info* = 5.22). However, simple information in the form of pithy recaps does appear to be useful (rounds 21-30 mean offers = 4.02), which suggests that the learning problems caused by recency do not stem solely from limited memory.

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

Though 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 (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 towards the optimum as well as decreasing their variance (test for equal variance in last 10 rounds of baseline vs. simple info. condition rejects equal variance, $p<.001$, additional analyses available from authors' websites).

6. 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) 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, thinking dynamically 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 & Pleskae 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 guide individuals towards 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 & Nagel 2011] and IPO investment [Kaustia & Knupfer 2008]. These findings suggest that even when recaps and record-keeping devices are available they may not be utilized. At this point, though, 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.

7. ACKNOWLEDGEMENTS

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, the Harvard Program for Evolutionary Dynamics for its hospitality.

8. REFERENCES

AGARWAL, S., DRISCOLL, J. C., GABAIX, X., & LAIBSON, D. 2008. Learning in the credit card market: National Bureau of Economic Research.

AKERLOF, G. A. 1970. The market for" lemons": Quality uncertainty and the market mechanism. *The Quarterly Journal of Economics*, 84: 488-500.

BALL, S. B., BAZERMAN, M. H., & CARROLL, J. S. 1991. An evaluation of learning in the bilateral winner's curse. *Organizational Behavior and Human Decision Processes,* 48:1-22.

BATTIGALLI, P. & D. GUATOLI 1997. Conjectural Equilibria and Rationalizability in a Game with Incomplete Information in *Decisions, Games and Markets*, Kluwer Academic Publishers, Norwell, MA.

BENAÏM, M., J. HOFBAUER, & E. HOPKINS 2009. Learning in Games with Unstable Equilibria, *Journal of Economic Theory* 144: 1694-1709.

BEREBY-MEYER, Y., & GROSSKOPF, B. 2008. Overcoming the winner's curse: an adaptive learning perspective. *Journal of Behavioral Decision Making,* 21: 15-27.

CAMERER, C. 2003. *Behavioral game theory: Experiments in strategic interaction*. Princeton: Princeton University Press.

CAMERER, C., & HO, T. 1999. Experience-weighted Attraction Learning in Normal Form Games. *Econometrica,* 67*:* 827-874.

CHARNESS, G., & LEVIN, D. 2009. The origin of the winner's curse: a laboratory study. *American Economic Journal: Microeconomics,* 1: 207-236.

CHEUNG, Y. & D. FRIEDMAN 1997. Individual Learning in Normal Form Games: Some Laboratory Results, *Games and Economic Behavior* 19: 46-76

DEKEL, E., FUDENBERG, D., & LEVINE, D. K. 1999. Payoff information and self-confirming equilibrium. *Journal of Economic Theory,* 89: 165-185.

DEKEL, E., FUDENBERG, D., & LEVINE, D. K. 2004. Learning to play Bayesian games. *Games and Economic Behavior,* 46: 282-303.

EREV, I. & HARUVY, E. 2013. Learning and the economics of small decisions. *Forthcoming in The Handbook of Experimental Economics, Vol. 2* Roth, A.E. & Kagel, J. Eds.

ESPONDA, I. 2008. Behavioral equilibrium in economies with adverse selection. *American Economic Review*, 98: 1269-1291.

EYSTER, E., & RABIN, M. 2005. Cursed equilibrium. *Econometrica,* 73: 1623-1672.

FUDENBERG, D. & D. KREPS 1995. Learning in Extensive Games, I: Self-Confirming Equilibria, *Games and Economic Behavior,* 8: 20-55.

FUDENBERG, D. & LEVINE, D.K. 1998. *The theory of learning in games* MIT press, Cambridge MA

FUDENBERG, D., & LEVINE, D. K. 1993. Self-confirming equilibrium. *Econometrica* 61: 523-545.

FUDENBERG. D. & LEVINE, D.K. 2013. Learning with Recency Bias, mimeo,

GLIMCHER, P. W., FEHR, E., CAMERER, C., & POLDRACK, R. A. 2008. *Neuroeconomics: Decision making and the brain*:

GUARNASCHELLI, S., MCKELVEY, R. D., & PALFREY, T. R. 2000. An experimental study of jury decision rules. *American Political Science Review*, 94:407-423.

HERTWIG, R., & PLESKAC, T. J. 2010. Decisions from experience: Why small samples? *Cognition, 115*: 225-237

HORTON, J., RAND, DG & ZECKHAUSER, RJ. 2011. The online laboratory: Conducting experiments in a real labor market. *Experimental Economics,* 14: 399-425.

KAGEL, J. H., & LEVIN, D. 1986. The winner's curse and public information in common value auctions. *American Economic Review*, 76: 894-920.

KAHNEMAN, D. & TVERSKY, A. eds. 2000. *Choices, values, and frames*. Cambridge University Press, 2000.

KAUSTIA, M., & KNÜPFER, S. 2008. Do investors overweight personal experience? Evidence from IPO subscriptions. *The Journal of Finance,* 63: 2679-2702.

LEVINE, D. K. 2012. *Is Behavioral Economics Doomed?: The Ordinary Versus the Extraordinary*: Open

Book Publishers.
MALMENDIER, U., & NAGEL, S. 2011. Depression Babies: Do Macroeconomic Experiences Affect Risk Taking? *Quarterly Journal of Economics, 126*: 373-416.
MILLER, G. A. 1956. The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological Review,* 63: 81-97.
NEVO, I., & EREV, I. 2012. On surprise, change, and the effect of recent outcomes. Frontiers in Cognitive Science, 3.
OUSS, A., & PEYSAKHOVICH. A. 2013. When Punishment Doesn't Pay: 'Cold Glow' and Decisions to Punish. *SSRN Working Paper Series*.
PEYSAKHOVICH, A. & KARMARKAR, U. 2013. Asymmetric effects of favorable and unfavorable information on decision-making under ambiguity. *Mimeo*.
PEYSAKHOVICH, A. & RAND, D. 2013. Habits of virtue: creating norms of cooperation and defection in the laboratory. *SSRN Working Paper Series*.
RAND, D. G., PEYSAKHOVICH, A., KRAFT-TODD, G., NEWMAN, G., WURZBACHER, O., NOWAK, M.A. & GREENE, J. forthcoming. Social Heuristics Shape Intuitive Cooperation. *Nature Communications.*
ROTH, A. E., & EREV, I. 1995. Learning in extensive-form games: Experimental data and simple dynamic models in the intermediate term. *Games and economic behavior*, 8:164-212.
RUBINSTEIN, A., TVERSKY, A., & HELLER, D. 1993. Naive strategies in zero-sum games: Sackler Institute of Economic Studies.
SAMUELSON, W. F., & BAZERMAN, M. H. 1985. Negotiation under the winner's curse. *Research in experimental economics,* 3: 105-138.
SIMONS, D. J., & CHABRIS, C. F. 2012. Common Mis Beliefs about Memory: A Replication and Comparison of Telephone and Mechanical Turk Survey Methods. *PLoS One*, E51876
SUTTON, R. S., & BARTO, A. G. 1998. *Reinforcement learning: An introduction*. Cambridge Univ Press, Cambridge UK.
TOR, A., & BAZERMAN, M. H. 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.