Recency, Records and Recaps: The effect of feedback on behavior in a simple decision problem*

Drew Fudenberg¹, Alexander Peysakhovich^{2,3}

Abstract

Suboptimal behavior can persist in simple stochastic decision problems. This has motivated the development of solution concepts such as cursed equilibrium (Eyster & Rabin 2005) and behavioral equilibrium (Esponda 2008). We experimentally study a simple adverse selection (or “lemons”) problem and find that learning models that heavily discount past information (ie. 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.

* Much of this work was done while Peysakhovich was a graduate student at the Harvard University Department of Economics. We thank the Harvard Program for Evolutionary Dynamics for its hospitality and the NSF for financial support.

¹Department of Economics, ²Program for Evolutionary Dynamics, Harvard University Cambridge, MA 02139, ³Department of Psychology, Yale University, New Haven, CT 06511

Introduction

People make mistakes. When does repeat experience lead individuals to optimal behavior and when can mistakes be persistent? This question lies at the heart of interpreting the potential market impacts of many findings from behavioral economics. Two classes of models allow the persistence of mistakes. In general, Bayesian-based models predict that only mistakes about “off-path” or otherwise unobserved events can persist in the long run. In contrast, many behavioral models assume that even “on-path” mistakes can be equilibrium phenomena (i.e. permanent or at least reasonably persistent even in the presence of feedback).

We steer a middle course and argue that “on path” mistakes do occur and can *sometimes* persist. Importantly, the well-documented phenomenon of *recency bias* (Erev & Haruvy 2013) plays a key role in determining whether mistakes are permanent or temporary.¹ In particular, recency bias can imply persistent suboptimal behavior in stochastic environments. To show this 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 across several decision problems.

Our stochastic decision problem is the simplified version of Akerlof’s (1970) classic lemons model introduced in Esponda (2008), the *additive lemons problem* (ALP). 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 seller² who accepts or rejects it. Buyers receive full feedback if their offer is accepted but receive no feedback if their offer is rejected.

¹ Recency bias refers to the tendency of individuals to discount past information. This is a useful heuristic in a non-stationary world where older observations will not necessarily be helpful (Sutton & Barto 1998). Recency bias has been incorporated into learning models explicitly by adding a term which controls the rate at which old information decays (Fudenberg & Levine 1998, Camerer 2003, Camerer & Ho 1999) or by allowing individuals to only condition their actions on a subset of the full history of outcomes (Erev & Haruvy 2013).

² Played in our experiments by a computer who always uses the dominant strategy.

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 (Tor & Bazerman 2003). 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).³ 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.⁴

These four conditions were chosen so that Nash, cursed and behavioral equilibrium have clear and distinct predictions: in the “low value added” payoff conditions CE predicts offers above NE offers while in the “high value added” conditions it predicts offers below NE. In contrast, BE predicts possible underbidding in conditions where participants are uninformed of the seller's valuation but coincides with NE predictions in conditions where the subjects were told the distribution of seller values. We use simulation methods to generate predictions for the TDRL model, and find that with relatively high recency the TDRL model predicts offers above NE in “low-added value” conditions but close-to optimal aggregate behavior in “high added value” conditions.

We then ask which model best organizes behavior across the four conditions. There is no underbidding in any of these conditions, more overbidding than (fully) cursed equilibrium predicts when it predicts overbidding, and almost optimal behavior where

³ To use simulation methods, we need to fix a functional form for the learning model. We choose the TDRL model for its simplicity, but the qualitative predictions come from the assumption of recency, not the chosen functional form.

⁴ This latter structure corresponds to the assumptions in Esponda (2008), who argues that it seems a better description of many field settings.

cursed equilibrium predicts underbidding, so neither Nash equilibrium, behavioral equilibrium, or cursed equilibrium fit well with the data. However, a 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, we add explicit counterfactual information: 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 information 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.

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 next 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 applied 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. In the *simple information* condition had identical rules. However, instead of receiving fully detailed feedback on each transaction, participants were 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.

Finally, to test the robustness of our findings, we study a slightly modified version of the Monty Hall problem. This problem also involves non-trivial probabilistic reasoning and, again, participants do not solve it correctly on the first try. Allowing for feedback we see that the effectiveness of recaps remains as before. Repeated play of the Monty Hall problem does not lead to optimal behavior but simple recaps lead to better decisions and higher payoffs.

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 formulated in 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 . Once this strategy is specified for the seller, the buyer's optimization problem is

$$\max_b \Pr(s \leq b)[E(v \mid s \leq b) - b] = \max_b (b/10)[k - b/2].$$

Solving the maximization problem shows that the optimal bid is $b^* = k$, so buyers offer k every round and sellers accept when $v < k$ and reject if $v > k$. Because the seller has a dominant strategy, we can transform the ALP into a single-person decision for the rest of our study.

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. 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 this correction in many decisions of interest, including common value auctions (Kagel & Levin 1986), the Monty Hall problem (Krauss & Wang 2003) and strategic voting games (Guarnashelli et al. 2000). Additionally, the ALP can be played repeatedly in a short amount of time.

The ALP is very similar to the Acquire a Company Game (ACG) introduced by (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 (Samuelson & Bazerman 1985) even with learning opportunities (Ball et al. 1991). Indeed, existing work is pessimistic (Bereby-Myer & Grosskopf 2008, Tor & Bazerman 2003) on whether any manipulations can help individuals converge to optimal behavior in this game. 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). The ALP, on the other hand, avoids this confound as for any value of k the optimal solution is interior. We examine the issue of corner-aversion empirically in experiment 3.

2.2 Predictions of Alternative Theories

Nash equilibrium makes the assumption of correct optimization on the part of the buyers. However, mistakes in the probabilistic reasoning required to arrive at the optimal solution are fairly common (Charness & Levin 2009, Kagel & Levin 1986, Krauss & Wang 2003, Samuelson & Bazerman 1985, Ball et al. 1991). Several classes of theories have been used to incorporate mistakes into the ALP and other economic problems. Some (eg. cursed equilibrium) are defined independently of the feedback subjects receive; they

are standard equilibrium concepts with non-standard utility functions or beliefs. Others theories, such as self-confirming equilibrium, maintain the assumption of Bayesian learning but use learning-theoretic foundations to explicitly take into account the feedback subjects receive (Fudenberg & Levine 1993a,b) and Fudenberg & Kreps (1995). Finally, a descriptive but less concise way to capture long run behavior is to use a fully specified dynamic learning model (for textbook treatments of this approach see Fudenberg & Levine 1998, Camerer 2003). Each theory makes different predictions behavior in the ALP and also suggests different causes for deviations from optimal play. We now discuss these theories in more detail. We then turn to experiments to see what each theory gets right and what it gets wrong.

Cursed Equilibrium

To incorporate subjects' failure to deal correctly with conditional probabilities, Eyster & Rabin (2005) introduce the cursed equilibrium (CE). CE assumes that when individuals in a Bayesian game optimize they completely ignore the correlation between their other players' types and their strategies.⁵

As a simple example, consider a game with two players, 1 and 2 and two states of Nature, A and B, with probabilities p and $1-p$ respectively. The states are observed by player 2 but not by player 1. Suppose player 2 plays action s when the state is A and plays r when the state is B. Then when player 1 sees action s , they should infer that the state is A. However, in CE player 1 ignores this information, and so optimizes against beliefs of the form "player 2 plays s with probability p , and r with probability $1-p$ regardless of the state of the world"

Following Esponda (2008), we adapt CE to the additive lemons problem by supposing that the buyer's maximization problem replaces the conditional expectation of the value v with its *unconditional expectation*.

$$\max_b \Pr(s \leq b)E(v - b) = \max_b (b / 10)[5 + k - b] .$$

⁵ In applying cursed equilibrium to the lemons problem Eyster & Rabin use a refinement to restrict off path play that is analogous to our assumption that the sellers do not use weakly dominated strategies. Eyster & Rabin 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.

so the cursed equilibrium bid here is

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

Note that this leads to overbidding (relative to the best response) if $k < 5$ and underbidding when $k > 5$.

As noted above, the predictions of CE do not depend on whether or not players were told the distribution of Nature's moves or on the sort of feedback they receive in the course of play. 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.

Behavioral Equilibrium

Esponda (2008) develops the solution concept of behavioral equilibrium (BE) specifically for the ALP. This concept is meant to model settings where (1) subjects 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.⁶

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 .

⁶ Thus BE combines aspects of rationalizable self confirming equilibrium (Dekel et al. 1999) with those of self-confirming equilibrium in Bayesian games (Dekel et al. 2004).

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. 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.

Temporal Difference Reinforcement Learning With Recency

We now turn to an explicitly dynamic point of view. To explore the effects of recency bias on individual decision-making it is helpful to have a concrete model that we can use for simulations. To do this, we choose a simple and extensively studied model: the temporal difference reinforcement learning model (TDRL). This model has a single parameter that controls the rate at which information from past observations is discounted. Variations of TDRL have been shown to fit human and animal learning behavior well (Glimcher et al. 2008) and neural evidence (Schultz 1998) suggest that the dopamine reward system may be performing some of the computations hypothesized by the model. For these reasons, we use the TDRL as our workhorse model; many of the conclusions reached here would be qualitatively unchanged under a richer and better fitting learning model which continues to incorporate recency (eg. Camerer & Ho 1999, Erev & Roth 1995, Roth & Erev 1998).

The TDRL model 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.⁷ In each period, individuals use a logit choice function, so they choose action a with probability

$$P(s_t = a) = \frac{\exp(\gamma v_t(a))}{\sum_{a' \in A} \exp(\gamma v_t(a'))}$$

⁷ Of course, a more realistic model would include a generating process for the initial valuations – that is, the source of the original mistake. However because the initial values are swamped by data fairly quickly we leave this out.

Here γ represents the degree of maximization; note that as gamma goes to infinity, choice approaches maximization whereas as gamma goes to 0 choice is completely random.

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 the TDRL model individuals update their valuations for action a according to

$$v_t(a) = v_{t-1}(a) + \alpha (PI_t - v_{t-1}(a))$$

where PI_t is the l *payoff* of the agent would have received from playing action a , in this 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. Thus the value of a is incremented upward or downward. Parameter α is the *learning rate* – the higher it is, the more responsive individuals are to recent rounds.

Note that this formula requires individuals to be able to compute the counterfactual payoffs to actions they did not play. How will updates occur when offers are rejected and individuals are not informed of the seller's valuation? For this model, we assume that individuals correctly infer that the computer's value v was above their bid b , draw a random number from the interval $[b, 10]$ and update their valuations as if this hypothetical v was the true value.⁸

This model predicts that aggregate behavior has converges to a stationary distribution when t is large. What does this distribution look like? To get some intuition, assume a continuous action space (so bids can be any real number between 0 and 10), a

⁸ This assumption does not drive our theoretical results. We also note that the data in experiment 1 reject the hypothesis that individuals do not update at all conditional on rejection. Finally, this asymmetry does not drive our experimental results (see experiment 2).

large population and gamma close to infinity (e.g. choice is essentially deterministic maximization). We now consider two limit cases on the learning rate.

First, suppose that α is close to 0. This means that a long history of outcomes carries approximately equal weights in determining the valuation function v_t . Thus, individuals will have, on average, correct valuations of each bid b and so aggregate behavior will be on average correct (and tightly concentrated around the optimal bid). On the other hand, consider a learning rate of 1, where behavior is just a best response to last period's information. Then, each individual simply bids the computer's value from last period and so the average population bid is equal to 5 ($E(v)$).

It is difficult to solve analytically for aggregate behavior for intermediate values of alpha, so we use simulations to derive predictions. We simulate $N=500$ agents playing 100 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 α and $\gamma=2$. The simulation results are qualitatively robust to variations in γ . We note that even at $\alpha = .9$ we still see average offers very close to optimal for $k=6$ but in the $k=3$ condition low levels of α are required for approximately optimal play.

Figure 1: Simulation results for the TDRL model. High recency (alpha close to 1) drives offer behavior to be very far from optimal in the $k=3$ condition but not in $k=6$ condition.

3. Experimental Design and Results

Subject Recruitment

All of the experimental participants were recruited online using the labor market Amazon’s Mechanical Turk (MTurk). 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 for example (Ouss & Peysakhovich 2013, Peysakhovich & Karmarkar 2013, Rand et al. 2012, Horton et al. 2011, Simons & Chabris 2012, Amir & Rand 2012, Peysakhovich & Rand 2013, Rand et al. 2013). All recruited subjects were US based (had US bank accounts).

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. Thus, all reported experiments use subjects who understood the rules of the experiment. Across all of our studies, approximately 75% of all participants passed the comprehension quizzes. This failure rate is slightly higher than

the typical rate in the studies above. However, our game is much more complicated than simple games such as the one shot Public Goods Game (see Appendix for instructions and 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. Experiments took between 10 and 17 minutes thus this amounted to an hourly wage of, on average, approximately \$9 per hour. No participant participated in any experiment in this series more than once.

Experiment 1: Baseline

We recruited $N=190$ participants to play 30 rounds of the additive lemons problem. In each round participants made a bid 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. Participants' payoffs were given in points which were converted into dollars and they began with a reserve of 50 points;—thus even if participants offered 9 points per round in the $k=3$ condition (where subjects could overpay the most) they would still, on average, not go broke over the course of the experiment.

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.

Our theories give clear hypotheses about what should happen in each of these treatments. CE predicts that we should see overbidding in $k=3$ and underbidding in $k=6$ treatments (CE predicts bids of 4 and 5.5 respectively). BE is a set-valued solution concept; it rules out overbidding, allows underbidding in the “uninformed” condition, and

and optimal bids in the informed treatments (as BE does not involve miscalculations of conditional probabilities or other sorts of misoptimization. Finally, simulations of the TDRL model with high recency predict overbidding (higher than CE-overbidding) in aggregate in $k=3$ treatments and almost NE behavior in $k=6$ treatments. Table 1 summarizes these hypotheses.

Table 1: Hypotheses by condition in the Baseline Experiment

<i>K=3, Informed</i>			<i>K=3, Uninformed</i>		
<u>CE</u>	<u>BE</u>	<u>TDRL</u>	<u>CE</u>	<u>BE</u>	<u>TDRL</u>
Above NE	Nash	Overbidding	Above NE	Below NE	Above NE
<i>K=6, Informed</i>			<i>K=6, Uninformed</i>		
<u>CE</u>	<u>BE</u>	<u>TDRL</u>	<u>CE</u>	<u>BE</u>	<u>TDRL</u>
Below NE	Nash	Close to NE	Below NE	Below NE	Close to NE

Results

Figure 2 shows time courses of average offers binned by 5 round blocks. There is a significant main effect of k across all rounds of the experiment, but there is no effect of being informed about the true distribution of values. This is confirmed in regressions (Table 2) which also show that there is a significant initial downward trend in the $v+3$ condition that is no longer present after 20 rounds. In the first 5 rounds of the game, our participants exhibit substantial and significant overbidding in the $k=3$ conditions (mean offer = 5.59; one-sided t-test $p < .001$ for offer > 3). Bids are much closer to optimality in the $k=6$ conditions (mean offer = 5.98) and are not significantly different from the optimal bid of 6 (two-sided t-test $p = .833$).

We also see a slight effect of experience, as bids in the last 5 rounds of the $k=3$ conditions are slightly lower than in the first 5 (mean offer = 5.096). This effect of experience is significant but disappears in the last third of the game, as shown in Table 3, where the significance of round number effect on bid disappears when we restrict the sample to the last 1/3 of the game. We can motivate this restriction via non-parametric tests: in the $k=3$ conditions the distribution of bids in the first 10 rounds is significantly

different from the distribution of bids in the last 10 rounds⁹ There are no experience effects in $k=6$ conditions.

One may ask whether this misoptimization seen in Figure 2 reflects mistakes by most subjects or is driven by some outliers. Figure 3 shows a histogram of bids in the last 10 rounds by condition. The distribution of behavior looks well centered and the aggregate overbidding in $k=3$ conditions is not driven by a few outliers.

Figure 2: Aggregate bid behavior by condition smoothed at 5 round blocks. Error bars represent standard errors clustered at the participant level.

⁹ Kolmogorov-Smirnov $p < .001$. The distribution of behavior in rounds 10-20 is not significantly different from the distribution of behavior in the last 10 rounds ($p = .43$).

Table 2: Effects of treatments on behavior.

	bid	bid
K=6	0.652 (0.164)***	0.594 (0.228)*
informed	0.103 (0.163)	0.052 (0.223)
round	-0.007 (0.003)*	-0.007 (0.003)*
(k=6)Xinformed		0.109 (0.327)
constant	5.341 (0.129)***	5.368 (0.146)***
R^2	0.04	0.04
N	5,700	5,700

Coefficients from OLS regressions.

Standard errors in parentheses clustered at participant level.

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

Figure 3: Bids in the last 10 rounds of experiment 1 by condition

Table 3: Experience effects in Experiment 1.

	All rounds	round > 20
K=6 dummy	0.337 (0.167)*	0.476 (0.599)
round	-0.016 (0.005)***	-0.011 (0.016)
(k=6)Xround	0.020 (0.006)**	0.015 (0.022)
constant	5.545 (0.115)***	5.423 (0.441)***
R^2	0.04	0.06
N	5,700	1,900

Coefficients from OLS regressions.

Standard errors in parentheses clustered at participant level.

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

Participants behave nearly optimally in $k=6$ conditions but continue to 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 4 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.

Figure 4: Mean efficiency of offers in last 10 rounds in experiment 1 by condition. Error bars represent standard errors clustered at participant level.

We now turn to the performance of the theories. First, we see no underbidding in any of our data, nor do we see any differences between informed and uninformed conditions. Thus, BE does not fit well with our data, despite its substantial intuitive appeal.

Second, we see substantial overbidding in the $k=3$ conditions, which is consistent with the key prediction of CE here, even though the overbidding is even higher than that CE predicts (Eyster & Rabin 2005 find a similar effect when trying to fit CE models to some instances of experimental data). On the other hand, we do not see the underbidding in the $k=6$ conditions that CE predicts.

Finally, we turn to the TDRL model. Rather than explicitly fitting parameters to the data, we explore qualitative features of the model's fit. First, as in TDRL simulations with high recency our high recency ($\alpha > .8$) 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 within a round once subjects have enough experience/feedback for equilibrium to roughly approximate their behavior, but 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 *predictable, non-random* changes in individual behavior between rounds and that these changes should continue even when individuals have played and received feedback on a substantial number of trials.

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. We restrict this analysis to the last 1/3 of all rounds, where aggregate behavior has converged.

Figure 5 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 .

We also look for these changes by regressing Δbid on outcomes in accepted rounds. We see a positive relationship between profit and changes Δbid . This relationship is both statistically and economically significant: – across conditions a 1 point increase in a round's profit increases bids in the next round by .19 points (Table 4). 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.

Figure 5: Relationships between outcomes in a round and changes in bid behavior in the last 10 rounds of experiment 1. Error bars represent standard errors clustered at participant level.

Table 4: Effects of information on Δbid

	Δbid	Δbid
roundprofit	0.191	0.340
	(0.026)***	(0.057)***
$K=6$		0.028
		(0.143)
(k=6)Xprofit		-0.207
		(0.068)**
constant	-0.576	-0.547
	(0.063)***	(0.076)***
R^2	0.08	0.11
N	1,014	1,014

Coefficients from OLS regressions.

Standard errors in parentheses clustered at participant level.

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

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 6). 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.

Figure 6: Number of steps taken by individuals by condition. Between 70 and 80% of individuals change their offer decisions between rounds.

Between 70 and 80% of participants' offer behavior exhibits (? Substantial? Just >0 isn't saying anything) variance, even in the last 1/3 of experimental rounds, which seems hard to reconcile with any sort of equilibrium analysis. What about the 20-30% of individuals who take make exactly the same bid in each of the last 10 rounds and so make 0 steps? . Figure 7 shows the bids that these individuals have converged to. The fact that these individuals' behavior is constant is consistent with equilibrium, but the substantial overbidding in the k=3 condition argues against the idea that these agents are rational Bayesians.

Figure 7: Bid behavior of individuals restricting to individuals whose strategies are constant over the last 10 rounds.

Thus we see that most subjects have a high responsiveness to recent outcome and that a simple model incorporating this recency effect does better at predicting aggregate behavior than the solution concepts of BE and CE. It is important to note that we do not claim to disprove the usefulness of these concepts. In particular, we believe that CE captures a real and important psychological phenomenon. However, we suggest that in repeated contexts CE is more useful as a guide to initial play before subjects get much feedback, while learning models are better for making medium and long-run predictions of most subjects' behavior. Even in this longer run, a concept like CE may act as a useful prediction for some fraction of the population, as shown by the 20-30% of our subjects who do not appear to adjust their behavior in response to feedback. Similarly, it is plausible that BE (and other forms of self-confirming equilibrium) do occur in the field, even though there is no evidence for them in the data we consider here.

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. However,

before turning to this, we present another experiment looking at the robustness of our results and further showing that extreme recency appears to exist in individuals' learning.

Experiment 2: Counter Factual Information

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 non-rational expectations about the value of v conditional on rejection. To check for this, as well as replicate our original results, 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 .

Results

Comparing the data from experiment CFI ("counter-factual information") to the behavior from experiment 1, we see little difference between behavior of individuals who have counterfactual information vs. those who do not (Figure 8). If anything, the individuals with counterfactual information do slightly worse (overbid more) in the $k=3$ condition, but this difference is not significant (Table 5).

As before, the aggregate outcomes are not driven by outliers. Figure 9 shows a histogram of bids in the last 10 rounds of the CFI experiment by condition – the data appear to be well centered. Thus, our results replicate and are the overbidding in $k=3$ conditions is not driven by the lack of information in rejection rounds.

We can also investigate whether most individuals appear to change their behavior from round to round in the CFI experiment. We define a step as in our baseline experiment and plot the number of steps individual behavior takes in the last 10 rounds. As in experiment 1 we find that approximately 80% of individuals indeed change their behavior from round to round (Figure 10).

Figure 8: The addition of counterfactual information does not help individuals converge to optimal behavior.

Figure 9: Bids in the last 10 rounds in CFI experiment.

Figure 10: Number of steps taken by individuals in the last 10 rounds of the CFI experiment.

Table 5: OLS Regressions pooling data from experiment 1 and experiment 2. Standard errors in parentheses clustered at participant level.

	bid	bid	bid
1=CFI experiment	0.108 (0.148)	0.108 (0.148)	0.174 (0.182)
round	-0.002 (0.003)	-0.012 (0.004)**	-0.012 (0.004)**
1=+6	0.614 (0.137)***	0.290 (0.145)*	0.329 (0.172)
+6Xround		0.021 (0.006)***	0.021 (0.006)***
+6XCFI			-0.138 (0.299)
constant	5.335 (0.108)***	5.489 (0.110)***	5.470 (0.118)***
R^2	0.03	0.04	0.04
N	7,950	7,950	7,950

Coefficients from OLS regressions.

Standard errors in parentheses clustered at participant level.

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

The CFI 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 counterfactual information, 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 . One way to quantify recency effects is to regress bid behavior at time t on lagged experienced

histories then use a model selection criterion to choose an “optimal” (from a model fit point of view) number of lags. If the recency hypothesis is correct, then the selected model should use a relatively small number of lags.

We use the Bayesian Information Criterion (BIC) to select the number of lags to include. The BIC of a fixed model is given by

$$-2\ln(L) + k*\ln(n)$$

where L is the likelihood of the model given the data, n is the number of observations (which we treat as number of subjects) and k is the number of model parameters. Thus, the BIC measure decreases in variance explained (model fit) and increases in number of free parameters ($k*\ln(n)$), so models with lower BIC values are “better” in terms of tradeoff between fit and number of free parameters (we refer the interested reader to (Posada & Buckley 2004) for more a more in-depth description).

Table 6 shows the results of this analysis. As predicted by the recency hypothesis, the linear regression model selected by the BIC involves only a single lag.¹⁰ This provides further evidence that a learning-based model with high recency (eg. TDRL) is an accurate descriptor of individual behavior.

¹⁰ Note that this does not mean that individuals literally discount all information beyond the last outcome. Rather this result means that the additional predictive power from looking at longer individual-level histories is small.

Table 6: BIC-based model selection.

	bid	bid	bid	bid
L1.v	0.078 (0.042)	0.075 (0.041)	0.078 (0.040)	0.079 (0.038)*
L1.(k=6)Xv	-0.092 (0.046)*	-0.090 (0.045)*	-0.093 (0.044)*	-0.096 (0.043)*
L2.v		0.017 (0.022)	0.012 (0.022)	0.012 (0.022)
L2.(k=6)Xv		-0.013 (0.029)	-0.008 (0.030)	-0.009 (0.030)
L3.v			0.030 (0.019)	0.029 (0.018)
L3.(k=6)Xv			-0.030 (0.028)	-0.027 (0.027)
L4.v				0.004 (0.028)
L4.(k=6)Xv				-0.017 (0.037)
constant	5.615 (0.123)***	5.568 (0.161)***	5.495 (0.159)***	5.519 (0.174)***
R^2	0.49	0.49	0.49	0.49
N	1,200	1,200	1,200	1,200
BIC	4063.22	4071.17	4077.78	4086.01

Coefficients from OLS regressions.

Standard errors in parentheses clustered at participant level.

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

Experiment 3: Recaps

We now show that an explicitly dynamic viewpoint is useful in not just describing the data but also in designing interventions to help individuals make better decisions. Thus, considering the dynamics of learning delivers insights that equilibrium models did not.

Why does suboptimal behavior persist at the aggregate level 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 an prescription for intervention: increasing the number of outcomes subjects observe simultaneously should help them make better decisions.

To test this hypothesis, we recruited $N=273$ more participants. In experiment 2 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 capacity (Miller 1956). Thus one may expect that more information is only useful if it is in easily "digestible" form. To test this 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 section 4 for examples of feedback screens).

Results

Figure 11 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 26-30 mean offer in control = 5.12, mean offer in *more info* = 5.18). However, simple information in the form of pithy recaps does appear to be useful (rounds 26-30 mean offers = 4.12). Table 7 confirms the statistical significance of these results.

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 terms of efficiency (Figure 12) than in the control and more information treatments. (Table 7; col. 2).

In addition to comparing sample averages we can also see what effect the simple information condition has on the full distribution of behavior. Figure 13 shows a histogram of bids in the last 10 rounds by condition. As the figure suggests, the simple information condition doesn't just affect a subset of the population but rather seems to drive the whole distribution of bids towards the optimum as well as decreasing the variance (test for equal variance in last 10 rounds of baseline vs. simple info. condition rejects equal variance, $p < .001$).

Figure 11: Bid behavior by 5 round blocks in experiment 3. Error bars represent standard errors clustered at participant level.

Table 7: OLS regression studying the effect of condition on bid behavior and efficiency. Standard errors in parentheses clustered at participant level.

	bid	efficiency
1=Extra Info	0.213 (0.263)	-0.213 (0.171)
1=Simple Info	-0.812 (0.192)***	0.489 (0.111)***
constant	5.046 (0.162)***	0.130 (0.104)
R^2	0.05	0.06
N	5,730	5,730

Coefficients from OLS regressions.

Standard errors in parentheses clustered at participant level.

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

Figure 12: Efficiency of strategies chosen in recaps experiment (last 10 rounds). Error bars represent standard errors clustered at participant level.

Figure 13: Bids in the last 10 rounds of the recaps experiment, by condition.

We note that this last result stands in contrast to work on the ACG, which finds that many interventions (Tor & Bazerman 2003), including one similar to our *simple info* condition (Bereby-Myer & Grosskopf 2008), do not appear to help individuals converge to optimal behavior. We hypothesize this is partially due to an aversion to make offers that are on the extremes of the interval.

To test this hypothesis we recruit an additional $N=79$ individuals to participate in 30 rounds of our baseline ALP conditions with offers restricted to the integers 0-6. This restriction leaves the optimal offer in the $k=6$ condition on the edge of the interval and the optimal offer in the $k=3$ condition squarely in the middle. We see that this lowers offers in the final 5 rounds to 4.18 (clustered s.e. = .209) and 4.24 (clustered s.e. = .185) in the $k=3$ and $k=6$ conditions respectively. In particular, now average offers in the $k=6$ condition exhibit the underbidding that CE (and BE) predict.

Combined with other results showing that individuals are prone to ‘avoid the edges’ in experimental paradigms (Rubinstein et al. 1993) this suggests that a potential confound to learning in the ACG is, in fact, the extremeness of the optimal offers. With this confound removed proper “packaging” of information can greatly improve individual decision-making.

Experiment 4: Monty Hall

To test whether our recency-based intuitions extend beyond the ALP we recruited $N=150$ more participants to participate in 30 rounds of the Monty Hall problem. In the Monty Hall problem participants are faced with 3 doors with a prize randomly hidden (by Monty Hall) behind one of the doors. The participants choose a door and Monty opens one of the other doors that does not hide the prize. Participants are then asked if they want to stick to their original choice or switch to the other unopened door.

Note that because Monty always removes a non-winning door, the probability of winning the prize from switching is 66% while the probability of winning from your initial choice is 33%. As in the ALP to find the optimal strategy participants need to correctly compute conditional probabilities – thus we again expect mistakes. Moreover, past work on Monty Hall games has concluded that most individuals will not learn the

optimal strategy, even with experience and some sorts of nudges” (Friedman (1998), Kraus & Wang (2003), Tor & Bazerman (2003), Palacio-Huerta (2003)).

Our subjects play a modified version of this game. First, we added a small cost of switching choices (the prize was 10 points and the cost of switching was 1 point). Without this modification CE predicts that participants are indifferent between switching and keeping (both having a probability of $1/3$ of winning in CE). With the addition of this small cost the unique CE prediction is for participants to choose to keep their original choice.

Our experiment included two conditions. One control condition where participants played 30 rounds of the Monty Hall game with full feedback and one simple information condition where participants played 30 rounds against 30 computers per round and received feedback in a manner similar to the simple information condition of the ALP (see online Appendix for screenshots of feedback screens).

Because the Monty Hall Problem has received attention in popular culture, we framed it in terms of choosing one of three cards (see the online Appendix for screenshots of the experiment/full experimental instructions). At the end of the experiment, participants were asked whether they were familiar with this game or any game like it: approximately 9% of our participants responded affirmatively (and had a 98% switch rate during their trials). We restrict our analysis to the naïve subjects ($N=137$).

Results

Figure 14 shows the aggregate behavior by condition. As in the ALP we found that with only standard feedback subjects did quite poorly. In the last 5 rounds subjects played the optimal strategy only 35% of the time in the control condition. However, subjects in the simple information condition played the optimal strategy (always switch) 75% of the time in the last 5 rounds. This difference is statistically significant (Table 8 col. 1).

Figure 14: Monty Hall behavior by condition, by five round blocks. Error bars represent standard errors clustered at participant level.

There is an experience effect in the control condition. But this effect disappears in the last 10 rounds. To confirm this statistically, we use a linear probability model to predict choosing to switch or not. We include a round indicator, a condition dummy and an interaction term. The coefficient on round is significant and positive in the first 10 rounds (as is the interaction with condition; Table col. 2) but it is much smaller and not significant in the last 10 rounds (Table 8 col. 3). This is again consistent with high recency - the more recency a learning model displays, the faster it discounts initial conditions and the faster aggregate behavior converges to the ergodic distribution.

To make the finding even more stark figure 15 shows the number of switches individuals make in the last 10 rounds by condition. Since it is always optimal to switch, making 10 switches implies that an individual played optimally in each of the last 10 rounds. Many more individuals act in an optimal manner in the last 10 rounds of the simple information condition (57% vs 10%; Ranksum test $p < .001$).

Table 8: Linear probability model predicting play of optimal strategy in the Monty Hall experiment.

	All Rounds	Round < 10	Round > 20
1=Simple Info	0.421 (0.058)***	0.282 (0.062)***	0.514 (0.184)**
round		0.021 (0.005)***	0.003 (0.006)
SimpleInfoXround		0.020 (0.008)*	-0.003 (0.007)
constant	0.291 (0.038)***	0.123 (0.035)***	0.262 (0.157)
R^2	0.18	0.18	0.19
N	4,128	1,237	1,378

Coefficients from OLS regressions.

Standard errors in parentheses clustered at participant level.

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

Figure 15: Number of switches made in the last 10 rounds of the Monty Hall problem, by condition.

Conclusion

Although suboptimal behavior *can* persist in the experimental play of simple Bayesian decision problems, this depends crucially on the structure of the feedback that individuals receive. Importantly, we find that just giving participants more information does not aid convergence to optimal behavior, but recaps can be highly effective

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, requires only simple assumptions on individuals' computational ability and 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. Our results thus support incorporating more accurate representations of bounded rationality into existing learning models. For reinforcement-learning based models in particular there is the additional challenge of how to identify the set of strategies that are reinforced. In the ALP it is easy to think of a plausible specification but even in the Monty Hall problem it is no longer as clear.¹¹ Answering this question intelligently is important for the application of any learning model to economic settings.

In addition, there is a debate in the literature 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.

¹¹ Bayesian-based models face a related problem in the specification of the state space and prior.

On the other hand, existing research has shown 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 (Malmandier & Nagel 2011) and IPO investment (Kaustia & Knupfer 2008). These findings suggest that while personal recaps are available they may not always be utilized. Testing the effect of recaps in other field environments could have both scientific and social benefit.

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*, 488-500.
- Amir, O., Rand, D.G. (2012). Economic games on the internet: The effect of \$1 stakes. *PLoS One*, 7(2).
- 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.
- 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., & Hua 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.
- Crawford, V. P., & Iriberri, N. (2007). Level-k Auctions: Can a Nonequilibrium Model of Strategic Thinking Explain the Winner's Curse and Overbidding in Private-Value Auctions? *Econometrica*, 75, 1721-1770.
- 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.
- Erev, I., & Roth, A. E. (1998). Predicting how people play games: Reinforcement learning in experimental games with unique, mixed strategy equilibria. *American Economic Review*, (volume) 848-881.
- Eyster, E., & Rabin, M. (2005). Cursed equilibrium. *Econometrica*, 73, 1623-1672.
- Friedman, D. (1998). Monty Hall's three doors: Construction and deconstruction of a choice anomaly. *The American Economic Review*, 88), 933-946.
- Fudenberg, D., & Levine, D. K. (1993). Self-confirming equilibrium. *Econometrica* 61 523-545.
- Fudenberg, D., & Levine, D. K. (1993). Steady state learning and Nash equilibrium.

Econometrica 61:, 547-573.

Fudenberg, D. L., David. (1998). *The theory of learning in games* MIT press, Cambridge MA

Glimcher, P. W., Fehr, E., Camerer, C., & Poldrack, R. A. (2008). *Neuroeconomics: Decision making and the brain*: Access Online via Elsevier.

Guarnaschelli, S., McKelvey, R. D., & Palfrey, T. R. (2000). An experimental study of jury decision rules. *American Political Science Review*, volume 407-423.

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. *The American Economic Review*, 894-920.

Kaustia, M., & Knüpfer, S. (2008). Do investors overweight personal experience? Evidence from IPO subscriptions. *The Journal of Finance*, 63(6), 2679-2702.

Krauss, S., & Wang, X. (2003). The psychology of the Monty Hall problem: discovering psychological mechanisms for solving a tenacious brain teaser. *Journal of Experimental Psychology: General*, 132(1), 3.

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?*. *The quarterly journal of economics*, 126(1), 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(2), 81.

Ouss, A., and 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*.

Posada, D., & Buckley, T. R. (2004). Model selection and model averaging in phylogenetics: advantages of Akaike information criterion and Bayesian approaches over likelihood ratio tests. *Systematic biology*, 53, 793-808.

Rand, D. G., Joshua D. Greene, and Martin A. Nowak. (2012). Spontaneous giving and calculated greed. *Nature*, 489, 427-430.

Rand, D. G., Peysakhovich, Alexander, Kraft-Todd, Gordon T., Newman, George E., Wurzbacher, Owen, Nowak, Martin A. and Greene, Joshua D. (2013). Intuitive Cooperation and the Social Heuristics Hypothesis: Evidence from 15 Time Constraint Studies. *SSRN Working*

Paper Series.

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.

Schultz, W. (1998). Predictive reward signal of dopamine neurons. *Journal of Neurophysiology*, 80, 1-27.

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*, 7(, 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.