Epiphany Learning for Bayesian Updating:

Overcoming the Generalized Monty Hall Problem

Wei (James) Chen and Joseph Tao-yi Wang*

Abstract

We study how people learn the correct action in a probabilistic situation. In particular, we create a modified version of the Monty Hall problem, and conduct laboratory experiments to show how a 100-door variant of the problem helps people learn to play optimally (always switch). Experimental results show that after playing the 100-door variant, subjects obtain an average switching rate above 80% in the 3-door problem, higher than most of the previous work without subject communication and/or competition. Moreover, results from estimating structural learning models using subject-level data show that the individual learning process is more likely to be an epiphany rather than a gradual one.

Keywords: Bayesian Updating, Change Management, Eureka Learning, Problem of Three Prisoners, Three Door Problem, Laboratory Experiments

---

* December 7, 2010. Department of Economics, National Taiwan University, 21 Hsu-Chow Road, Taipei 100, Taiwan. Chen: r97323056@ntu.edu.tw; Wang: josephw@ntu.edu.tw. Yu-Tsong Tai, Yahan Chuang, Ally Wu, Sara Yi-Ping Bai and Vivian Tzu-Fan Own provided excellent research assistance. We thank comments from Daniel Friedman, Chen-Ying Huang, Colin F. Camerer, and Walter Yuan. All remaining errors are our own.

Epiphany Learning for Bayesian Updating:

Overcoming the Generalized Monty Hall Problem

Introduction

In 1991, British Petroleum (BP) announced a "no dry holes" policy regarding oil exploration, counting every drill that did not produce oil as a failure. Unlike traditional expected-value calculations, this new policy prevented BP's explorers from using probabilistic predictions as a cover of failure, such as, "if we drill ten of these 1-in-10 wells, we'll hit at least one of them and we'll all make a lot of money."[1] This policy also forced them to carefully evaluate all available information before going ahead, instead of learning from holes that did not hit. In consequence, explorers started to systematically aggregate various aspects that support an oil field, and only drilled locations where all geological tests are positive. By 2000, BP's hit rate increased to an industry-leading 2 in 3, which is three times to the success rate of 1989 (1 in 5).[2]

This is an example of people learning the correct action (to drill or not) in a probabilistic situation. In particular, investigations prior to the policy change showed that BP explorers were very accurate when they estimated a probability of hitting between 20 and 70 percent, but an "estimated probability of 10 percent" was more like 1 percent.[3] Hence, the company policy of no dry holes forced BP explorers to stop fudging with the low probabilities, and concentrate on accurately estimating the high probabilities that would eventually lead to an action. Though we

---

1 Heath and Heath (2010), p.89.

2 See p.87-93 of Heath and Heath (2010) for the full story of BP. For example, the authors also discuss another reason why the BP campaign was successful: the new policy forced management to stop drilling holes that they deem as "strategic" (usually to please a government or business partner relationship they want to maintain) against the evaluation of frontline technical teams.

3 Heath and Heath (2010), p.88-89.

cannot verify whether BP explorers actually learned to estimate various probabilities more accurately (since they no longer drilled holes that they perceived as low probability), BP's success in terms of drilling hit rate (2 out of 3), as well as the resulting decrease in exploration cost, is an important achievement at where it counts, namely the "economic value" (Camerer, Ho and Chong, 2004) of correct probability estimation.

Interestingly, such success is not a result of the conventional adaptive learning models usually discussed in economics, such as belief-based fictitious play (Boylan and El-Gamal, 1993), reinforcement learning (Erev and Roth, 1998), or hybrid models like experience-weighted attraction (EWA) learning (Camerer and Ho, 1999). Though these learning models have been successful in describing experimental choices in controlled environments that are encountered repeatedly (See Camerer, 2003, Chapter 6 for a review), they are usually too simplified to capture actually learning behavior in more realistic situations.[4] Instead, BP's success is an example of firms implementing changes by fine-tuning the situation at hand to create a "simpler" environment for its employees to learn and adopt the better practice. In fact, after adopting the no dry holes policy, BP employees could no longer blame their failures on bad luck, and had to take full responsibility for their wrong decisions. This policy made the environment "simpler" for people to learn to make better decisions. In this regard, they are more close to classroom experiments where instructors use simplified situations to teach students about more complicated settings (See for example, Ball, Eckel and Rojas, 2006; Dickie, 2006; Durham, Mckinnon and Schulman, 2007).

[4] For example, Erev, Roth, Slonim and Barron (2007) found that learning models based on small sampling updates (the Inertia, Sampling and Weighting model, I-SAW) outperform conventional models in randomized environments. Using eyetracking data, Knoepfle, Wang and Camerer (2009) foundthat the lookup patterns of experimental subjects reject conventional adaptive models in favor of more complicated models such as "anticipatory learning" (Selten, 1991).

In additional to classroom experiments, several recent studies in experimental economics have also dealt with similar issues. For example, Bednar, Chen, Liu and Page (2009) found that when playing two different repeated games simultaneously, self-interest behavior in the stage-game prisoner's dilemma "spilled over" to the other game played, while the presence of a "hard" game (where the Pareto dominant dynamic equilibrium is alternating between two outcomes) requires a larger cognitive load and induces the use of "simpler" strategies such as stage-game dominant strategies in the other game. Dufwenberg, Sundaram and Butler (2010) found that subjects learned to play the "game of 21" (a perfect information game similar to Nim that requires forward looking ability) if they played the simplified "game of 6" first.

In this paper, the BP success is replicated in the laboratory: We identify a particular environment where it is difficult to learn the optimal action under a probabilistic situation, and design a simplified environment where subjects can learn to make the correct decision, which will be carried back to the original environment. In particular, the difficult environment consists of a modified version of the Monty Hall problem, a situation commonly used to study biases in human decision making (Friedman, 1998; Kluger and Wyatt, 2004; Kluger and Friedman, 2010). Indeed, experimental subjects fail to learn the optimal action even after 30 rounds of repeated play. Then, we design a simplified version of the game (100-door instead of 3-door), and demonstrate how most subjects can learn to play the optimal action (always switch) within 15 rounds after playing first 15 rounds of the simplified game.

The Monty Hall problem is one of the strongest choice anomalies studied during the past decade. Originated from a famous game show, the problem involves the decision of a contestant choosing between three doors, in which one of them contains a big prize. After selecting one door, the game show host (Monty) opens one of the remaining two doors revealing that it is empty. Then, the contestant has to decide whether to switch to the other opaque door, or keep their initial choice. Assuming the host always randomly opens one empty door (among all empty doors not chosen), Bayes' rule suggests the contestant should always switch so the probability of winning the prize would be 2/3 compared to 1/3 (if one does not switch). However, most people simply cannot learn that switching is the best strategy, even after experiencing the same situation thirty or forty times (Friedman, 1998).

Researchers have used the Monty Hall problem as a tool to study judgment errors in finance (Kluger and Wyatt, 2004) and attempted to use various treatments to teach people how to respond optimally to this problem, i.e. to always switch. These treatments include group competition and communication (Slembeck and Tyran, 2004), providing past history (Friedman, 1998), introducing financial market (Kluger and Friedman, 2010), and so on. One striking result is that it is very difficult for subjects to learn to switch. To achieve a switching rate of 70% or higher, one would have to introduce sophisticated institutions, such as group competition or financial markets. This leaves one to wonder whether there exists a "simple" way to teach subjects to achieve a switching rate of 70% or higher.

In this paper, we take on the challenge to design a simple treatment to induce subjects to learn that switching is always optimal (when faced with two doors). Specifically, subjects play 15 rounds of a 100-door variant of the Monty Hall problem, and then play the original 3-door version for another 15 rounds. The 100-door variant consists of 100 doors. After choosing one door, the host opens 98 doors according to the same rules of the standard Monty Hall problem, randomizing with equal probability when he has a choice, and allows the subject to switch. The winning probability for switching now soars to 99/100, while not switching wins only one out of a hundred times. As Marilyn vos Savant initially discussed in her news paper column *Ask Marylin*, "*Suppose there are a million doors, and you pick door #1. Then the host, who knows what's behind the doors and will always avoid the one with the prize, opens them all except door #777,777. You'd switch to that door pretty fast, wouldn't you?*" (Savant, 1997)

This design is adopted for several reasons. First of all, in the 100-door variant, the probability of winning if one switches is 99%, matching the BP story where the prior probabilities of the new environment are close to one after they adopted the "no dry holes" policy. Secondly, Page (1998) showed that subjects do switch in the 100-door variant, but this knowledge does not carry over to the 3-door version when both games were played simultaneously, but only once. Hence, it is plausible that knowledge gained in the 100-door variant could spill over to the 3-door version after sufficient learning.

A third design issue we had to face is that there are some rules of thumb other than Bayesian updating that would also lead to switching in the Monty Hall Problem. For example, the heuristic of "Irrelevant Therefore Invariant (ITI)", documented by Shimojo (1989), states that since the host's action was irrelevant to the contestant's initial choice, the probability of winning with the initial choice is therefore unchanged by the host's action. Thus, the (posterior) chances of winning if one switches is one minus the (prior) winning probability of the initial choice. Since both ITI and Bayesian updating lead to the same choice of switching, we cannot distinguish the two reasoning in the standard game.

Therefore, in our experiments, subjects play a variant of the Monty Hall problem so that ITI and Bayesian updating would predict different behavior: One of the three doors is transparent, showing that it is empty, and choosing this door would end the game immediately. All other aspects of the game are the same. By introducing the transparent door, we shift the prior probability from (1/3, 1/3, 1/3) to (1/2, 1/2, 0),[5] making subjects who follow the ITI heuristic indifferent between switching and not. In contrast, if one carefully performs Bayesian updating, one would realize that there are two separate situations after the host opens the door: If the host opens the other non-transparent door, it is obvious one should not switch (since one loses for sure switching to a transparent door); if the host opens the transparent door, switching is the optimal strategy since the winning probability of switching is still 2/3 and that of not switching is 1/3 (same as in the standard game). Observing behavior in this new Monty Hall problem, we are able to separate those who adopt the ITI heuristic from those who truly follow Bayesian updating. Finally, this variant also makes the experiment closer to the BP example---subjects initially see only two possible choices, drill or not (take the outside option) each having the same chance of success, but after investigation (Monty's move), the posterior probability shifts to 2/3. As BP moves from the status quo to the "no dry holes" policy (3-door to 100-door), the posterior probability becomes 99/100.

Comparing the results of subject who underwent the 100-door treatment with those who played the same 3-door treatment, we address the following questions: Does the 100-door treatment "teach" subjects to always switch in the 3-door one? For those who learn the optimal strategy during the experiment, how did they learn?

Experimental Design

Participants and Procedure

Participants in the experiment were National Taiwan University (NTU) students recruited through the Taiwan Social Science Experimental Laboratory (TASSEL) website or from various intermediate-level economics courses in NTU. They were

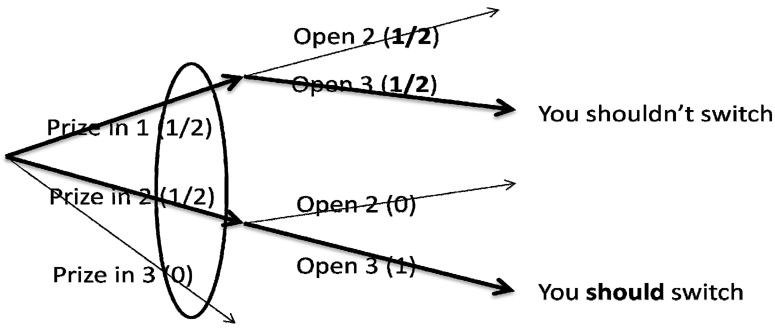
[5] Or (1/100, …, 1/100) to (1/2, 1/2, 0, …, 0) in the 100-door variant.

assigned randomly to one of the two groups: Control group (43 students) and Treatment group (39 students). They were paid a 100 NT dollar (approx. US$3.00) show up fee, and earned 10 NT dollars (approx. US$0.30) each time they chose the door containing the prize.

Each subject individually plays the generalized Monty Hall problem, either the 3-door version with prior probability (1/2, 1/2, 0), or the 100-door version with prior probability (1/2, 1/2, 0,…, 0). In the treatment group, subjects first play the 100-door game for 15 periods and then play the 3-door version for 15 periods, while in the control group, subjects first played the 3-door game for 15 periods and repeat another 15 periods of the same game.

Optimal Strategy

Although the modified game is more complex, according to Bayes' Rule, the optimal strategy is still to switch whenever one faces two opaque doors, and this would gives the subject a 2/3 (or 99/100 in the 100-door version) chance to win. Figure 1 illustrates this with an example of the 3-door game where door 3 is transparent, and the subject chooses door 1 initially. If door 3 is opened by the computer and the subject chooses to switch (to door 2), s/he will win 2/3 of the time.



$$P(\text{Winning if you choose to switch}) = \frac{\frac{1}{2}\times 1}{\frac{1}{2}\times 1 + \frac{1}{2}\times\frac{1}{2}}$$

Figure 1

Results

Aggregate Behavior

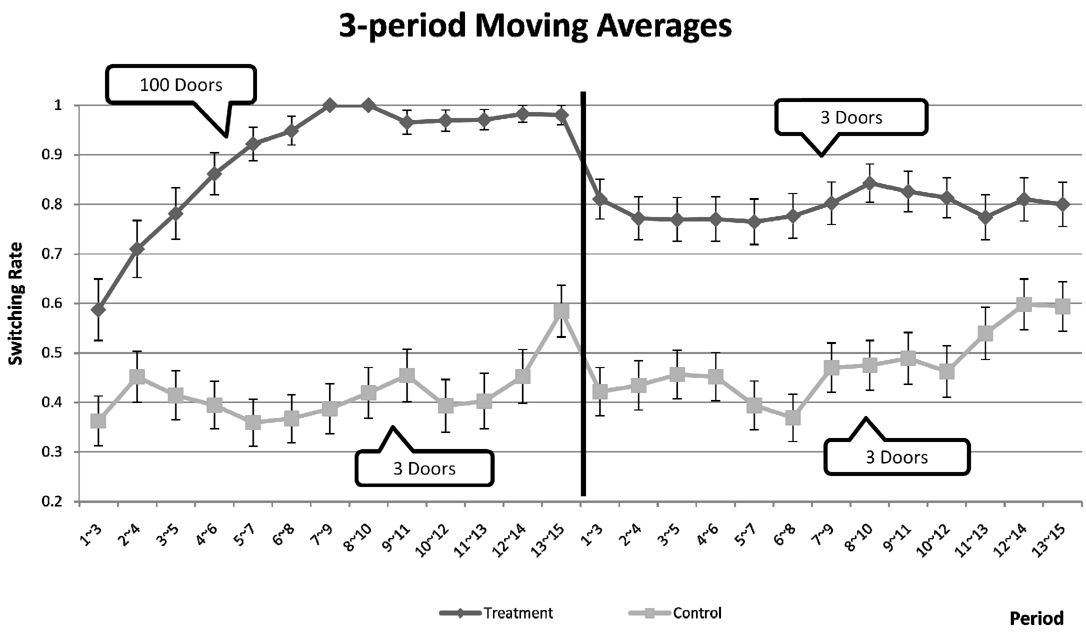


Figure 2

We first compare aggregate switching behavior between the treatment and control group. Note that there are chances that a subject would face one transparent door and one opaque door if the computer opens the other opaque door for him. Cases in which this particular situation happens are not interesting since it is common knowledge that one may win for sure by choosing to stay. As a result, we drop these cases and show in Figure 2 the 3-period moving average (with standard error bars) of the switching rate for cases in which subjects have a choice between two opaque doors. The treatment group starts with a switching rate close to 60%, and quickly learns to always switch (> 95%) in the 100-door game. Most subjects then carry this knowledge over to the 3-door game, resulting in a stable switching rate around 80%. In contrast, the control group starts with a switching rate below 40%, and only gradually increases to 60%.

Moreover, we conduct a *probit* regression with random effects, predicting switching behavior (the probability that the dummy variable *Switch* equals to 1) with a constant term, the period number (correlated with how many periods the subject have played the relevant case), and two variables that represent past experience: *Switch_bonus* (the cumulative earning difference between always switching and always remaining) and *Switch_won* (a dummy variable which equals to 1 if switching would have won the prize in the most recent period). A similar model was used by Friedman (1998) to support reinforcement learning. Consistent with the findings of Friedman (1998), results in Table 1 show that the variable *Switch_bonus*, is strongly correlated with *Switch* ($p < 0.05$), while *Switch_won* is not ($p > 0.1$).

	(1)	(2)	(3)	(4)
	Control_1	Control_2	Treatment_1	Treatment_2
VARIABLES	**Switch** (=1 if and only if subject switch, =0 otherwise)			
Switch_bonus	0.0234***	0.0119**	0.112**	0.0265***
	(0.00654)	(0.00475)	(0.0460)	(0.00674)
Switch_won	0.113	-0.0276	0.271	-0.225
	(0.180)	(0.164)	(0.391)	(0.210)
Period	-0.0318	0.0221	-0.495	-0.0942**
	(0.0347)	(0.0243)	(0.372)	(0.0374)
Constant	-0.533*	-1.022**	0.671	0.411
	(0.272)	(0.435)	(0.509)	(0.458)
Observations	461	495	294	439
Number of subject	43	43	39	39

Standard errors in parentheses

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

We drop all observations where there was only one opaque door left.

Table 1: Probit regression (with random effects) for switching behavior

Individual learning model analysis

In addition to the aggregate results, we (structurally) estimate individual learning patterns with two different learning models, epiphany and reinforcement, to calibrate individual switching data. The epiphany model is the one that the subject chooses not to switch in the first *n* periods until a certain "epiphany point" where the subject realizes the optimal strategy and starts to always switch from then on. Hence, the probabilistic model predicts a subject to switch with probability 1-ε before the given epiphany point and switch with probability ε after that. The parameters *n* and ε are estimated by maximizing empirical likelihood for each subject.

On the other hand, the reinforcement model we used is a special case of the Experienced-Weighted Attraction (EWA) model (Camerer & Ho 1999, 2002)[6]. In this model, the probability that a subject chooses (not) to switch increases if and only if he choses (not) to switch in the previous period and won the prize, but decays by a fixed proportion in all other cases.

In particular, we use a logit (exponential) reinforcement model to calibrate the data in both stages and both groups and the combined data of the two stages in the control group (30 periods in total). In this model, the subject chooses his strategy according to the following formula:

$$P^1(t+1) = \frac{e^{\lambda * R^1(t)}}{e^{\lambda * R^1(t)} + e^{\lambda * R^2(t)}}$$

Where P^1 (t+1) indicates subject's probability of choosing strategy 1 (switch) at period t+1, λ is a parameter that is estimated via maximum likelihood, and $R^i(t)$ is the attraction at period t if one chooses strategy *i* (= 1 if switch), modeled as:

$$R^i(t) = \begin{cases} \phi R^i(t-1) + \pi(t), & \text{if } s(t) = s^i \\ \phi R^i(t-1), & \text{if } s(t) \neq s^i \end{cases}$$

[6] *We do not estimate the full EWA model (with 3 free parameters) since we have at most 30 data points per subject, which was shown by Salmon (2001) to be inefficient to identify the parameter precisely.*

where ϕ is also a parameter that is estimated via maximum likelihood, $\pi(t)$ is the pay-off that the subject receive at period t, s(t) is the subject's strategy at period t, s^i indicates the subject's strategy choices i (= 1 for switch), and $R^i(0) = 50$. We again estimate the value of ϕ and λ by maximizing empirical likelihoods. The estimation results (see supplemental material) show that the epiphany model has higher mean log likelihood for more than 90 percent of the subjects in both the control and treatment group. In other words, though aggregate results from the probit regressions support some form of reinforcement learning, individual structural estimations indicate subjects seem to learn by "epiphany" instead of (purely) reinforcement.

Moreover, by estimating the epiphany point for each individual, we can measure the speed subjects learn to play the optimal strategy. In the control group it takes an average of 17.65 (standard deviation 12.25) out of 30 periods to learn, while in the 100-door game, it takes only 1.21(1.79) periods. After experiencing the 100-door game, it now takes subjects only 2.54 (4.86) periods to learn (to switch) in the 3-door game.

Discussion and Conclusion

Results from our experiment show that, even though we introduced a more difficult and counter-intuitive version of the Monty Hall Problem, our 100-door treatment effectively teaches subjects to learn the optimal strategy. However, it is not clear from the behavioral results alone whether subjects learned true Bayesian updating or not. To at least partially address this, we conducted post experimental questionnaires and asked subjects to report their belief about the probabilities of wining if they choose to switch/stay. The results show that more subjects in the treatment group report the exact probabilities (66.66%, 66.6%, 66%, or 67%) than those in the control group (23.08% versus 6.98%). Thus, in addition to learning to switch and reap the "economic value" (Camerer, Ho and Chong, 2004) of correct probability estimation, we also have some evidence that people do learn the correct probabilities themselves.[7]

Moreover, close to 40% of the subjects started with switching. This is much higher than what is documented in the literature since Friedman (1998) reported an initial switching rate around 10%. This could be the result of Asian students being more capable of mathematics. However, it could also be due to the fact that the Monty Hall problem is more well-known now (say due to the movie "21") than ten years ago. Initially, we did not ask subjects whether they had seen a similar game before, but we started to ask this question after seven of the first 30 subjects reported exact probabilities. Among the remaining 52 subjects, 18 of them claimed to have at least seen a similar game before, while 34 did not. Among those who have seen it before, four of them reported exact probabilities (22.2%), while among those who never seen it, only one reported exact probabilities (2.9%).

Thirdly, although aggregate regression results could indicate reinforcement learning, a comparison between the individual epiphany model and the reinforcement model suggests that at least 90% of the subjects are more likely to be classified as an epiphany learner. Hence, we believe that most of our subjects can learn the optimal strategy in the 100-door game because a "eureka" moment occurred after several periods, and this epiphany is transferred to the 3-door game.

Finally, although we now know that epiphany is a better model to calibrate subjects' learning behavior in this case, we still do not know the actual learning mechanism behind it. Is it like a real "epiphany" from God in which the subject suddenly realizes the optimal strategy at some point in the experiment? Or, is it like a

[7] If we allow an estimating error of 5%, the numbers become 30.77% (in the treatment group) and 11.63% (in the control group).

neuron stimulation process with a preset threshold where subjects aggregate stimuli when exposed to the games repeatedly and realize the optimal strategy after passing the threshold, such as the model estimated by Krajbich, Armel and Rangel (2010)? Answers to these questions would await further research.

Reference

S.B. Ball, Eckel, C., Rojas, C. 2006. Technology Improves Learning in Large Principles of Economics Classes: Using Our WITS. *American Economic Review*. **96**(2) 442-446.

J. Bednar, Chen, Y., Liu, X., Page, S. 2009. *Behavioral Spillovers in Multiple Games: An Experimental Study*. University of Michigan, City.

R.T. Boylan, El-Gamal, M.A. 1993. Fictitious Play: A Statistical Study of Multiple Economic Experiments. *Games and Economic Behavior*. **5**(2) 205-222.

C.F. Camerer. 2003. *Behavioral game theory: Experiments on strategic interaction*. Princeton University Press, Princeton.

C.F. Camerer, Ho, T.H. 1999. Experience-weighted attraction learning in normal form games. *Econometrica*. **67**(4) 827-874.

C.F. Camerer, Ho, T.H., Chong, J.-K. 2002. Sophisticated experience-weighted attraction learning and strategic teaching in repeated games. *Journal of Economic Theory*. **104**(1) 137-188.

C.F. Camerer, Ho, T.-H., Chong, J.-K. 2004. A cognitive hierarchy model of games. *Quarterly Journal of Economics*. **119**(3) 861-898.

M. Dickie. 2006. Do Classroom Experiments Increase Learning in Introductory Microeconomics? *Journal of Economic Education*. **37**(3) 267-288.

M. Dufwenberg, Sundaram, R., Butler, D.J. 2010. Epiphany in the Game of 21. *Journal of Economic Behavior & Organization*. **75**(2) 132-143.

Y. Durham, Mckinnon, T., Schulman, C. 2007. Classroom Experiments: Not Just Fun and Games. *Economic Inquiry*. **45**(1) 162-178.

I. Erev, Roth, A., Slonim, R., Barron, G. 2007. Learning and equilibrium as useful approximations: Accuracy of prediction on randomly selected constant sum games. *Economic Theory*. **33**(1) 29-51.

I. Erev, Roth, A.E. 1998. Predicting How People Play Games: Reinforcement Learning in Experimental Games with Unique, Mixed Strategy Equilibria. *American Economic Review*. **88**(4) 848-881.

D. Friedman. 1998. Monty Hall's Three Doors: Construction and Deconstruction of a Choice Anomaly. *American Economic Review*. **88**(4) 933-946.

D. Heath, Heath, C. 2010. *Switch: How to Change Things When Change Is Hard*. Random House of Canada.

B. Kluger, Friedman, D. 2010. Financial Engineering and Rationality: Experimental Evidence Based on the Monty Hall Problem. *Journal of Behavioral Finance*. **11**(1) 31 - 49.

B.D. Kluger, Wyatt, S.B. 2004. Are Judgment Errors Reflected in Market Prices and Allocations? Experimental Evidence Based on the Monty Hall Problem. *Journal of Finance*. **59**(3) 969-998.

D.T. Knoepfle, Wang, J.T.-y., Camerer, C.F. 2009. Studying Learning in Games Using Eye-Tracking. *Journal of the European Economic Association*. **7**(2-3) 388-398.

I. Krajbich, Armel, C., Rangel, A. 2010. Visual fixations and the computation and comparison of value in simple choice. *Nat Neurosci*. **13**(10) 1292-1298.

S.E. Page. 1998. Let's make a deal. *Economics Letters*. **61**(2) 175-180.

T.C. Salmon. 2001. An Evaluation of Econometric Models of Adaptive Learning. *Econometrica*. **69**(6) 1597-1628.

M.v. Savant. 1997. *The Power of Logical Thinking: Easy Lessons in the Art of Reasoning...and Hard Facts About Its Absence in Our Lives*. St. Martin's Griffin, New York.

R. Selten. 1991. *Anticipatory learning in two-person games*. Springer, City.

S. Shimojo, Ichikawa, S. 1989. Intuitive reasoning about probability: Theoretical and experimental analyses of the "problem of three prisoners". *Cognition*. **32**(1) 1-24.

T. Slembeck, Tyran, J.-R. 2004. Do institutions promote rationality?: An experimental study of the three-door anomaly. *Journal of Economic Behavior & Organization*. **54**(3) 337-350.

Supplemental Online Material

1st Stage Aggregate 3-period Moving Average

1st Stage

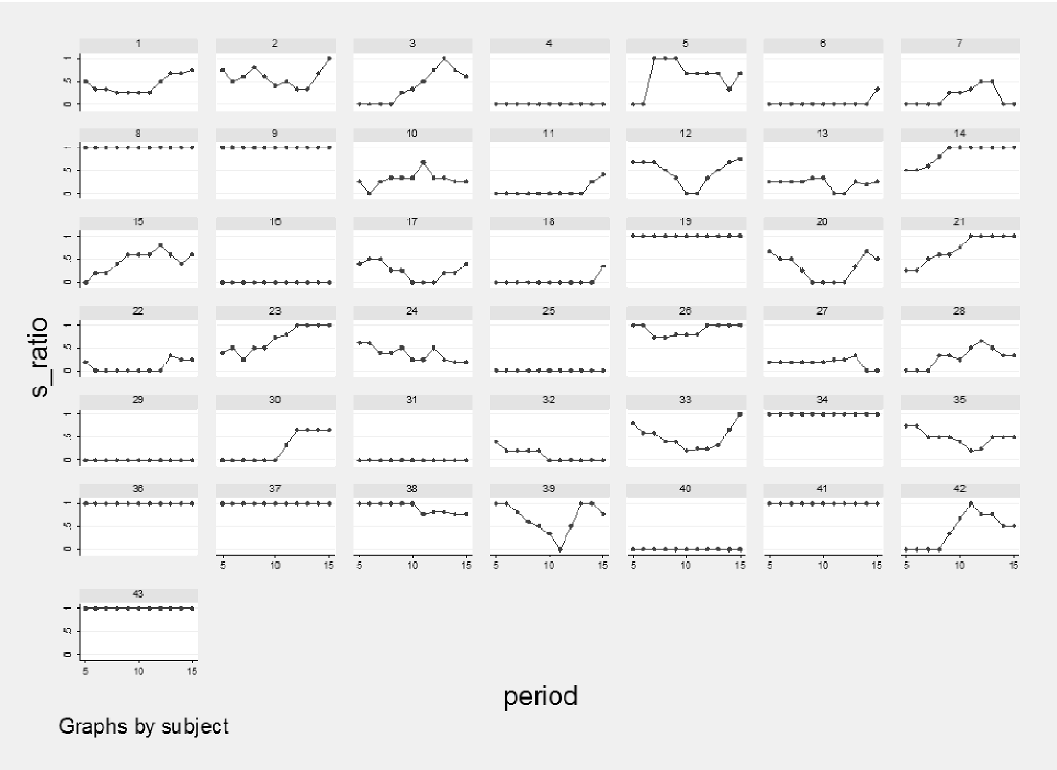
	Control			Treatment		
Period	Switch	Chances	Ratio	Switch	Chances	Ratio
1~3	33	91	36.26%	37	63	58.73%
2~4	42	93	45.16%	44	62	70.97%
3~5	41	99	41.41%	50	64	78.13%
4~6	41	104	39.42%	56	65	86.15%
5~7	37	103	35.92%	59	64	92.19%
6~8	36	98	36.73%	55	58	94.83%
7~9	36	93	38.71%	50	50	100.00%
8~10	39	93	41.94%	50	50	100.00%
9~11	40	88	45.45%	56	58	96.55%
10~12	33	84	39.29%	63	65	96.92%
11~13	31	77	40.26%	66	68	97.06%
12~14	38	84	45.24%	56	57	98.25%
13~15	52	89	58.43%	50	51	98.04%

2nd Stage Aggregate 3-period Moving Average

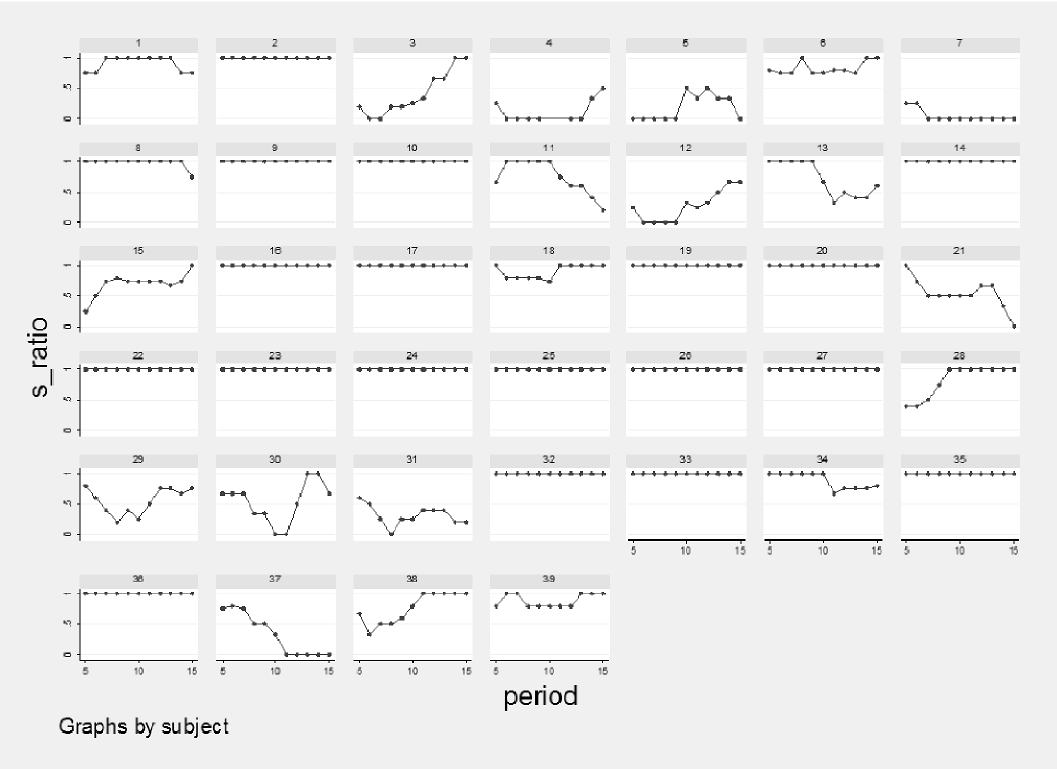
2nd Stage

	Control			Treatment		
Period	Switch	Chances	Ratio	Switch	Chances	Ratio
1~3	43	102	42.16%	77	95	81.05%
2~4	43	99	43.43%	71	92	77.17%
3~5	47	103	45.63%	70	91	76.92%
4~6	47	104	45.19%	67	87	77.01%
5~7	39	99	39.39%	65	85	76.47%
6~8	38	103	36.89%	66	85	77.65%
7~9	47	100	47.00%	69	86	80.23%
8~10	47	99	47.47%	75	89	84.27%
9~11	45	92	48.91%	71	86	82.56%
10~12	43	93	46.24%	74	91	81.32%
11~13	48	89	53.93%	65	84	77.38%
12~14	55	92	59.78%	64	79	81.01%
13~15	57	96	59.38%	64	80	80.00%

Personal Moving Average (Control Group)



Personal Moving Average (Treatment Group)



Maximum Likelihood Estimation Results for Epiphany and Reinforcement Learning

Control Group			**1st stage only**				**Both stages**			
Obs. with 2doors			**Epiphany**		**Reinforcement**		**Epiphany**		**Reinforcement**	
Subj. ID#	1st Stage	2nd Stage	mean log-L	Epiphany Point	mean log-L	mean log-L difference	mean log-L	Epiphany Point	mean log-L	mean log-L difference
1	12	10	0.00	0	-0.12	0.12	-0.54	0	-0.52	-0.02
2	13	12	-0.54	8	-0.68	0.14	-0.59	8	-0.69	0.10
3	10	13	-0.50	15	-0.69	0.19	-0.46	23	-0.63	0.17
4	6	13	0.00	14	-0.46	0.46	0.00	30	-0.15	0.15
5	12	8	0.00	14	-0.17	0.17	-0.42	17	-0.59	0.16
6	13	11	0.00	15	-0.05	0.05	0.00	27	-0.30	0.30
7	11	9	-0.59	5	-0.66	0.07	-0.56	30	-0.65	0.09
8	12	12	0.00	0	-0.12	0.12	0.00	0	-0.06	0.06
9	12	11	0.00	0	-0.06	0.06	0.00	0	-0.03	0.03
10	12	11	-0.67	13	-0.75	0.08	-0.63	30	-0.72	0.09
11	12	13	-0.56	15	-0.46	-0.10	-0.37	28	-0.47	0.10
12	9	9	-0.53	4	-0.67	0.14	-0.64	4	-0.66	0.03
13	11	11	-0.47	6	-0.68	0.21	-0.66	27	-0.69	0.03
14	11	11	-0.30	14	-0.60	0.30	-0.40	14	-0.53	0.13
15	13	15	-0.54	3	-0.64	0.10	-0.67	20	-0.69	0.02
16	12	12	0.00	15	-0.06	0.06	0.00	30	-0.09	0.09
17	13	14	-0.27	14	-0.59	0.32	-0.48	27	-0.63	0.15
18	13	8	-0.27	12	-0.58	0.31	-0.41	27	-0.57	0.16
19	12	11	0.00	12	-0.25	0.25	0.00	12	-0.13	0.13
20	9	10	0.00	8	-0.22	0.22	-0.58	26	-0.67	0.09
21	10	12	-0.33	15	-0.14	-0.19	-0.30	21	-0.32	0.01
22	11	12	-0.30	14	-0.56	0.26	-0.39	30	-0.56	0.18
23	10	12	-0.61	13	-0.59	-0.02	-0.59	22	-0.59	0.00
24	12	14	-0.45	14	-0.69	0.24	-0.62	30	-0.69	0.08
25	6	12	-0.64	12	-0.69	0.06	-0.35	30	-0.45	0.10
26	11	11	-0.66	1	-0.65	-0.01	-0.54	1	-0.52	-0.02
27	13	12	-0.43	15	-0.64	0.21	-0.44	29	-0.59	0.15
28	8	9	-0.56	8	-0.69	0.13	-0.65	30	-0.69	0.04
29	10	10	-0.50	15	-0.52	0.02	-0.33	30	-0.26	-0.06
30	9	11	-0.68	0	-0.76	0.08	-0.70	23	-0.69	0.00
31	12	13	0.00	13	-0.06	0.06	0.00	30	-0.03	0.03
32	11	15	-0.31	12	-0.54	0.23	-0.44	30	-0.53	0.09
33	10	13	-0.33	3	-0.52	0.20	-0.57	3	-0.62	0.05
34	12	13	0.00	0	-0.06	0.06	0.00	0	-0.03	0.03
35	8	13	0.00	10	-0.18	0.18	-0.60	10	-0.65	0.05
36	11	12	0.00	0	-0.06	0.06	0.00	0	-0.03	0.03
37	10	10	0.00	0	-0.07	0.07	0.00	0	-0.03	0.03
38	12	13	0.00	1	-0.36	0.36	-0.17	1	-0.32	0.15
39	9	11	-0.35	6	-0.67	0.33	-0.50	6	-0.67	0.17
40	8	10	0.00	14	-0.09	0.09	0.00	30	-0.04	0.04
41	7	11	-0.41	2	-0.55	0.14	-0.21	2	-0.22	0.01
42	13	11	-0.43	13	-0.59	0.16	-0.51	21	-0.66	0.15
43	9	12	0.00	0	-0.08	0.08	0.00	0	-0.03	0.03
Mean				8.56				17.65		
(std)				(5.93)				(12.25)		

Treatment Group			1st stage				2nd stage			
Obs. with 2doors			Epiphany		Reinforcement		Epiphany		Reinforcement	
Subj. ID#	1st Stage	2nd Stage	mean log-L	Epiphany Point	mean log-L	mean log-L difference	mean log-L	Epiphany Point	mean log-L	mean log-L difference
1	6	13	0.00	0	-0.12	0.12	-0.43	0	-0.56	0.13
2	8	14	0.00	1	-0.17	0.17	0.00	0	-0.10	0.10
3	8	13	0.00	3	-0.26	0.26	-0.43	7	-0.57	0.14
4	8	8	0.00	1	-0.17	0.17	-0.38	13	-0.36	-0.02
5	5	9	0.00	1	-0.28	0.28	-0.35	15	-0.56	0.21
6	7	13	-0.60	0	-0.58	-0.02	-0.43	0	-0.48	0.05
7	6	9	0.00	0	-0.12	0.12	-0.35	13	-0.39	0.39
8	9	10	0.00	2	-0.15	0.15	-0.33	0	-0.47	0.47
9	10	13	0.00	3	-0.28	0.28	0.00	0	-0.05	0.05
10	5	10	0.00	1	-0.28	0.28	0.00	0	-0.14	0.14
11	4	12	0.00	0	-0.17	0.17	-0.64	1	-0.57	0.12
12	9	10	-0.53	4	-0.62	0.09	-0.50	11	-0.53	0.03
13	5	11	0.00	0	-0.14	0.14	-0.59	0	-0.65	0.06
14	6	12	0.00	0	-0.12	0.12	0.00	0	-0.06	0.06
15	8	12	0.00	0	-0.09	0.09	-0.45	5	-0.68	0.23
16	9	10	0.00	2	-0.23	0.23	0.00	0	-0.07	0.07
17	12	10	0.00	1	-0.12	0.12	0.00	0	-0.07	0.07
18	9	12	0.00	0	-0.08	0.08	-0.29	0	-0.32	0.04
19	9	12	0.00	4	-0.31	0.31	0.00	0	-0.06	0.06
20	7	12	0.00	0	-0.10	0.10	0.00	0	-0.06	0.06
21	5	11	0.00	4	-0.55	0.55	-0.69	0	-0.43	-0.05
22	9	10	-0.35	0	-0.28	-0.07	0.00	0	-0.07	0.07
23	9	11	0.00	0	-0.08	0.08	0.00	0	-0.06	0.06
24	10	10	0.00	0	-0.07	0.07	0.00	0	-0.07	0.07
25	6	11	0.00	1	-0.23	0.23	0.00	0	-0.06	0.06
26	9	9	0.00	0	-0.08	0.08	0.00	0	-0.15	0.15
27	5	11	0.00	0	-0.14	0.14	0.00	0	-0.06	0.06
28	7	13	0.00	0	-0.10	0.10	-0.27	4	-0.27	0.00
29	2	13	0.00	0	-0.35	0.35	-0.67	0	-0.69	0.02
30	6	7	-0.70	6	-0.81	0.11	-0.76	2	-0.79	0.04
31	10	14	0.00	5	-0.28	0.28	-0.65	15	-0.69	0.28
32	7	9	0.00	1	-0.20	0.20	0.00	0	-0.08	0.08
33	7	13	0.00	1	-0.20	0.20	0.00	0	-0.05	0.05
34	7	12	-0.41	0	-0.36	-0.05	-0.29	0	-0.32	0.04
35	7	12	0.00	0	-0.10	0.10	0.00	0	-0.06	0.06
36	8	11	0.00	0	-0.09	0.09	0.00	0	-0.13	0.13
37	8	9	0.00	6	-0.26	0.26	-0.69	12	-0.69	0.34
38	7	12	0.00	0	-0.10	0.10	-0.45	0	-0.29	-0.16
39	7	14	0.00	0	-0.10	0.10	-0.26	1	-0.42	0.17
Mean				1.21				2.54		
(std)				(1.79)				(4.86)		

Note:

These tables report the MLE result of fitting individual choices into different learning models.
The last column reports the differences between two models' log-liklihood result.
Shaded row are those subjects who are better predict under the reinforcement model.

Questionnaire Result (Control)

Subject	Part1p1	Part1p2	Part2p1	Part2p2
1	75	25	75	25
2	60	40	50	50
3	50	50	33	75
4	50	50	60	40
5	80	20	80	20
6	50	50	70	30
7	60	40	40	60
8	66	34	66	34
9	66.6	33.3	50	50
10	20	80	70	30
11	50	50	50	50
12	50	50	80	20
13	50	50	50	50
14	75	25	75	25
15	50	50	50	50
16	50	50	50	50
17	50	50	50	50
18	30	50	50	50
19	20	80	80	20
20	70	30	50	50
21	10	90	90	10
22	40	60	50	50
23	50	50	75	25
24	40	60	50	50
25	20	80	40	60
26	70	30	90	10
27	40	60	50	50
28	40	60	40	60
29	18	82	5	95
30	0.5	0.5	0.5	0.5
31	40	60	60	40
32	50	50	50	50
33	50	50	50	50
34	66	33	66	33
35	50	50	50	50
36	66	33	66	33
37	60	40	60	40
38	67	33	50	50
39	60	40	50	50
40	20	80	20	80
41	30	70	90	10
42	60	40	60	40
43	50	33	50	33

Questionnaire Result (Treatment)

Subject	Part1p1	Part1p2	Part2p1	Part2p2
1	99	1	50	50
2	100	0	75	25
3	50	50	70	30
4	30	70	50	50
5	1	99	50	50
6	90	10	90	10
7	25	75	50	50
8	50	50	100	0
9	100	0	60	40
10	99.5	0.5	66.7	33.3
11	99	1	67	33
12	50	50	70	30
13	0	100	50	50
14	99	1	67	33
15	50	50	50	50
16	100	0	100	0
17	99	1	50	50
18	99	1	66	34
19	100	0	90	10
20	99	1	66	33
21	100	0	50	50
22	100	0	50	50
23	99	1	67	33
24	50	50	66.7	33.3
25	100	0	70	30
26	99	1	75	25
27	100	100	70	100
28	100	0	50	50
29	100	0	50	50
30	50	60	50	60
31	100	0	50	50
32	100	0	80	20
33	100	0	87.5	12.5
34	100	0	50	50
35	80	20	66	34
36	65	35	60	40
37	100	0	66	34
38	90	10	50	50
39	100	0	50	50

The second and fourth column report subject's belief about the wining probability of choosing to switch (part1 means in the 1st stage), and the third and fifth column report the opposite probability.