Effects of Value Function in Prediction Performance of Human Behavioral Models in Iowa Gambling Task

Mohammad Rahmani Fadiheh, Farzad Towhidkhah, and Golnaz Baghdadi

Abstract**—**This study investigate different assumptions about valuation process in modeling human decision making in a standard gambling task. Our research considered three different assumptions about valuation of a particular action to find out which can predict human decisions with the highest accuracy. This investigation leads to a new model which is better than previous models according to its performance in predicting behavioral data. We benefit from markov chain monte-carlo (MCMC) methods to obtain the maximum likelihood and the best performance of the model in predicting human behavior. The results demonstrates that the model with error-frequency driven value function with reversal learning gain the best BIC score and is able to predict behavioral data with BIC score more than 200.

Keywords**—**Iowa gambling task, Bayesian information criterion, likelihood, Annealed importance sampling.

I. Introduction

THE modeling of human cognition with exploitation of both theoretical mathematics and simulation tools can lead to a profound knowledge of how people think and decide. This can be a key factor for interpreting human behavioral disorders and abnormalities. In this way, by applying functional models it is possible to approach a precise structural model of human brain that is critical for development of cognitive science.

One of the most important aspects of human behavior modeling is to determine how people value each decision over others. Usually the best choice in a set of possible actions is not so clear and a complex value estimation procedure is needed to decide advantageously. The aim of this paper is to study and investigate how people consider advantageous and disadvantageous decisions in uncertain situations. In this study model evaluation for human decision on a psychological decision making task, the Iowa Gambling Task will be performed [1].

In this article, we will use the Bayesian Information Criterion [2] (BIC) to achieve a numerical scale in which show that whether a model is able to mimic the human behavior better than a Bernoulli process or not. The main advantage of BIC is that it will consider the complexity of the model which it make it a powerful tool in order to select the best model of human decision making process. Three different factors in value function will be considered and the results will be compared by the BIC.

The employed model learning model for comparing different value functions is consist of three layers. The first layer is a utility function that will represent the human estimation of his choice immediately after seeing the payoff [3],[4]. Second, a learning rule will be applied in order to weight all of alternative actions based on the output previous layer and form what the person should expect from his choices [3],[5]. The last layer will produce a probability distribution over the action set which the person is choosing from [4]. This research will focus on the first layer and appraise effect of different value functions in modeling human behavior in IGT.

The Iowa Gambling Task has been used in several previous cognitive researches in order to validating behavioral models (e.g., [3],[4],[5],[6]) and they analysis different aspects of human decision making in IGT. However there wasn't enough heed on the role of value function in a decision making model. Kalidindi et al. [4] considered different value functions and this study aim to continue and redress the shortage by studying their results and assess their model using BIC.

Also this study will discuss other aspects such as, describing the experimental procedure and findings based on empirical data. Describe the model and the method used to fit it to the empirical data. Finally using the Bayesian Information criterion in order to compare models with different utility functions and discuss the role of value function in the learning.

II. Experiment Procedure And Empirical Data

The IGT is a laboratory probe to detect decision making deficits in patients with ventromedial prefrontal cortex (VMPC) lesion [1]. VMPC damaged patients express perfect performance in Wisconsin Card Sorting Test [7], paradigms requiring self-ordering [8] and cognitive estimations [9]. However their decisions lead to negative results in real life conditions and they are usually careless about future consequence [10]. The main advantage of the IGT is that it is

Mohammad Rahmani Fadiheh is with the faculty of Biomedical Engineering, Amirkabir university of Technology, Tehran, Iran

Farzad Towhidkhah is with the faculty of Biomedical Engineering, Amirkabir university of Technology, Tehran, Iran (corresponding author's phone: +982164542363 ; e-mail: towhidkhah@aut.ac.ir).

Golnaz Baghdadi is with the faculty of Biomedical Engineering, Amirkabir University of Technology, Tehran, Iran.

able to spot decision making abnormalities caused by VMPC damages which are obvious in real life situations but other psychological tasks are useless in detecting them.

In this task, the subject sits in front of a computer screen that shows four decks of cards which are identical in shape and size. He is given some amount of money as the loan of play money. The subject is asked to select cards from decks and keep on doing it unless he is told to stop. After choosing a card, he will be paid an amount of money and will be asked to pay another amount of money as penalty. The amount of payoff and penalty of the cards of each deck varies with others. The final goal of the task is to maximize the amount of money in the account of the subject (For further information on the IGT see [11].)

In the ABCD version of the task which is used in this study, there are two good decks and two bad decks. The good decks C and D provides a regular win of 1$ and loss of 5$ per 10 cards. The bad decks, A and B, give the subject amount of 2$ per cards and will take 25$ as penalty per 10 cards. The penalty in decks B and D is infrequent and large however decks A and C will take money more frequently in smaller amounts. Each deck consists of 40 decks [1] and each participant should choose 100 cards. Table I represents scores of first ten card of each deck in ABCD version of IGT.

TABLE I
WINS AND LOSSES FOR FIRST 10 CARD IN EACH DECK IN ABCD STANDARD VERSION OF IGT

Card no.	Deck A		Deck B		Deck C		Deck D	
	Win	Loss	Win	Loss	Win	Loss	Win	Loss
1	2$	0	2$	0	1$	0	1$	0
2	2$	5$	2$	0	1$	1$	1$	0
3	2$	0	2$	0	1$	0	1$	0
4	2$	5$	2$	0	1$	1$	1$	0
5	2$	0	2$	0	1$	0	1$	0
6	2$	5$	2$	0	1$	1$	1$	0
7	2$	0	2$	0	1$	0	1$	0
8	2$	5$	2$	0	1$	1$	1$	0
9	2$	0	2$	0	1$	0	1$	0
10	2$	5$	2$	25$	1$	1$	1$	5$

A. Participants

Fourteen individuals from Amirkabir University of Technology participated in this study voluntarily. There were 8 women and 6 men in the participants and almost all of them were bachelor and master students (ages 19-24 years; mean 20.9±1.4). All of them were paid 5$ as show up bonus and an amount of reward based on their performance in the task.

B. Procedure

The subjects were told to choose card repeatedly over four decks. It was clearly said that there were good decks and bad decks which they should find the good decks to maximize payoffs. Based on the instruction told to the participants, they are free to choose each of the decks they want and the only goal of the game is to maximize the payoff. Each subject starts the task with 10$ as an initial play money and they were informed that they will receive their wining at the end of the task. In the ABCD version that used in this study, decks A and B are bad decks and decks C and D are good decks. Decks B and D have infrequent comparably large losses and decks A and C have frequent smaller losses.

C. Results

We divide each trial by trial result into 5 blocks of 20 cards and for each block a score calculated by (1), that is difference of number of cards selected from good decks(C and D) and number of cards selected from bad decks(A and B).

$$NetScore = ((C+D)-(A+B)) \quad (1)$$

Average of Score over all of participants is plotted in fig. 1. Subjects were divided into two groups based on their score. The first group, learner subjects, expressed learning during the task and their score have been increased over 5 blocks. The second group, non-learner subjects, was not capable of finding the good decks and their score have been decreased over 5 blocks [11]. There are 12 learner subjects among 14 participants (85% of all) and just 2 participants failed to learn the task and increase their score. Although even the non-learner group gained a comparably good score at the end of the task but poor performance of these subjects may be the result of the uncertain structure of the task or the immature cortex of the young college students [12].

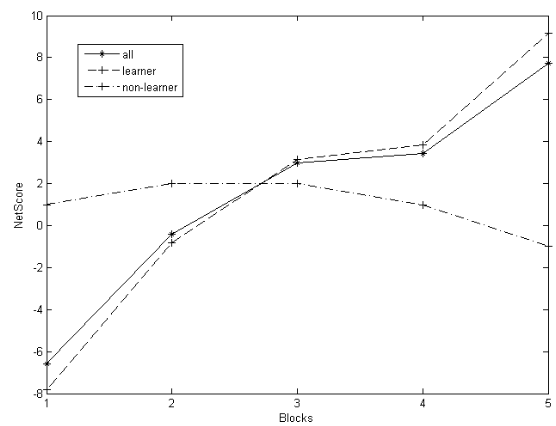

Fig. 1 average score over 5 blocks in learner and non-learner subjects

III. BEHAVIORAL MODELS

To examine which utility function mimic the way person evaluate the situation with the best accuracy, a computational model with three sections for evaluation, expectation and decision is considered. Three different kinds of utility function are described below for the first section of the model. Our novel model is based on previous studies that cover their shortages and perform much better in predicting results gained from the IGT according to its BIC score.

A. Error-driven value function

This function was used in the model proposed by Kalidindi et al. [4]. It is slightly inspired from the delta rule [13] and defined by (2).

$$U_t(a) = U_{t-1}(a) + \gamma(r_t(a) - U_{t-1}(a)) \quad (2)$$

In equation (2), $U_t(a)$ is the estimated value for the action a in the trial t. $r_t(a)$ denotes the payoff gained from action a in the trial t and γ represents the learning rate.

This function updates the estimation of an action with the difference between the estimated amount for the reward and the actual reward gained at the last time the action were selected.

B. Error-frequency Driven Value Function

Kalidindi et al. [4] used this function in their research and this is based on the idea that the orbitofrontal cortex biases the decision relying on the magnitude of the wining and losses, however the basal ganglia has more tendency towards the frequency of reward and punishment [14].

$$\begin{aligned} &if\,(r_t(a) > 0), \quad U_t(a) = U_{t-1}(a) + \gamma(1 - U_{t-1}(a)) \\ &if\,(r_t(a) < 0), \quad U_t(a) = U_{t-1}(a) - \gamma(1 + U_{t-1}(a)) \\ &if\,(r_t(a) = 0), \quad U_t(a) = (1-\gamma).U_{t-1}(a) \end{aligned} \quad (3)$$

For the first trial, the function will follow (4)

$$U_1(a) = sign(r_1(a)) \quad (4)$$

In aforementioned equations, γ represents the learning rate, and r(t) and $U_t(a)$ are same as described in (2). According to (3) and (4), the function will update the estimated value for each action with the difference between the actual value and the desired value which will be one if the action result in reward in all trials.

C. Reversal-learning Based Value Function

Reversal learning is not solely a way of estimating the value of an action but it can affect the way subject esteem an action. The reversal learning is mainly about how people adapt to stimuli that are against their previous beliefs [15]. In other words, if the subject has some positive beliefs about an action but this action lead to negative consequences in reality, the reversal learning can affect the decision in this kind of situations. It is also true if the relation become reverse.

Orbitofrontal cortices play a key role in reversal learning and previous studies proves that patients with lesion in orbitofrontal regions of brain failed in reversal learning tasks [30]. Also Fellows et al. [31] shows that patients with ventromedial prefrontal cortex lesion which failed to perform in IGT like normal healthy people, express a good performance in developed shuffled ABCD variant of the IGT. In the shuffled version of the task, the bad cards that contain large losses moved to the top of the bad decks so it eliminates the initial tendency toward the bad decks and the necessity of reversal learning during the task [31]. The good performance in shuffled IGT by VMPC damaged patients posits the idea that reversal learning plays a key role in human decision making procedure in IGT. This is why we add the reversal learning term to the value function in order to mimic human behavior.

This function adds the mathematical interpretation of reversal learning to the *Error-frequency Driven* value function. Equation proposed by [4] for implementation of reversal learning was used in our model.

$$\begin{aligned} &if\,(r(t) > 0) \quad Z = U_{t-1}(a) + \gamma(1 - U_{t-1}(a)) \\ &if\,(r(t) < 0) \quad Z = U_{t-1}(a) - \gamma(1 + U_{t-1}(a)) \\ &if\,(r(t) = 0) \quad Z = (1-\gamma).(1 - U_{t-1}(a)) \\ &if\,(sign(Z) = sign(U_{t-1}(a)) \vee U_{t-1}(a) = 0) \\ &\quad U_t(a) = Z \qquad (5) \\ &else \\ &\quad if\,(r(t) > 0) \quad U_t(a) = U_{t-1}(a) + \lambda.\gamma(1 - U_{t-1}(a)) \\ &\quad if\,(r(t) < 0) \quad U_t(a) = U_{t-1}(a) - \lambda.\gamma(1 + U_{t-1}(a)) \\ &\quad if\,(r(t) = 0) \quad U_t(a) = (1-\lambda).(1-\gamma).(1 - U_{t-1}(a)) \end{aligned}$$

In this function, γ represents the learning rate, λ denotes the reversal learning rate, Z is an auxiliary variable that help to determine necessity of reversal learning and $U_t(a)$ and r(t) are same as described for (2). According to (5), when the actual estimated value of the selected action considering the last reward or punishment is not same with the previous value, the function will decrease the learning rate with λ.

D. Experience Weighted Attraction (EWA) for Expectancy Function

EWA combines different assumptions to achieve a more general model for learning. This model considers how the subject expectation will be changed through new rewards and punishments. Moreover, it takes the experience into account for learning [16]. Experience of the subject affects the expectancy in the way that more experience will result in less sensitivity to new payoffs. EWA models have been used in various studies [3]. The expectation for each action will be calculated using the following equation.

$$\begin{aligned} &C_t = \rho.C_{t-1} + 1 \\ &E_t(a) = \frac{C_{t-1}.\phi.E_{t-1}(a) + U_t(a).\delta_t(a)}{C_t} \end{aligned} \quad (6)$$

In this function, $E_t(a)$ represents the expectation for action a in trial t. C_t accounts for the experience in trial t and ϕ is the discount factor for the previous attraction [16] and ρ denotes the weight for the experience. For ρ = 0, this function will be insensitive to the gained experience during the task. On the other hand, as the subject proceeds in the task, C_t will be increased and the weight for new rewards and punishments become less.

E. Quantal (Logit) Best Response for Choice Rule

This decision rule based on the prominent idea in behavioral economics that human decision making procedure is more probabilistic rather than deterministic [17]. In other words, the probability of committing a fault will be more as the cost of the fault decreases. So the best action will not be chosen with

certainty [17]. This idea formulized by the following equation in the literature ([17],[18],[19]).
There will be a mixed strategy that will follow (7).

$$P_t(a) = \frac{e^{\theta . E_t(a)}}{\sum_{a' \in A} e^{\theta . E_t(a')}} \qquad (7)$$

In the aforementioned equation, θ is the precision factor, $E_t(a)$ is the expectancy for action a calculated in the previous section and A is the action set which the subject is choosing from. At the end, this function provides $P_t(a)$ which determines the probability of choosing action a over the action set A in the trial t. The precision factor denotes the tendency to greedy decisions so for $\theta = 0$ the decision will be purely explorative and uniformly random and for $\theta = \infty$ the decision will be completely exploitive.

IV. Model verification Method

In this study, we have judged models using the maximum likelihood estimation and Bayesian Information Criterion that will be described below. The final model has 5 independent parameters so we should investigate a five-dimensional space to find the point with the maximum likelihood to compute BIC. So we use the Annealed Importance Sampling method for sampling the space and finding the proper point.

A. Likelihood Estimation

Likelihood is a powerful tool for evaluating validity of a model in predicting behavior of the phenomena in reality. In models which predict human behavior in IGT like this study, likelihood is obtained from the equation below.

$$L = \Pr(\text{Empirical data}|\text{Model}) = \prod_{t=1}^{N} P_t(a_t) \qquad (8)$$

a_t denotes the selected deck at the trial t and $P_t(a_t)$ represents the probability of choosing deck a_t at that trial.

B. Annealed Importance Sampling

This method was proposed by [20] to sample from complex distribution with an easy to sample distribution using markov chain monte carlo (MCMC) methods (for further details on this method see [20]).

If X is the vector that represents a point in the space of parameters of the model, so $\pi(x)$ will be the prior distribution of the parameters X and P(x) will be the posterior distribution or the distribution of interest. The final goal is to compute the expectation of function F(x) over the distribution of interest P(x). To this end, a markov chain ($T_j(x,x')$) should be calculated that converge to the P(x) and do not change the corresponding distribution. $T_j(x,x')$ denotes the probability that the chain goes from x to x'. The desired markov chain for this purpose could be calculated using Metropolis-Hasting ([21],[22]) or Gibbs sampling updates algorithms [23]. Each chain will start from a point x_n which is produced by the prior distribution $\pi(x)$ and will continue for certain number of samples (n+1) to achieve x_0. This procedure will provide a sequence of samples ($x_n,x_{n-1},\ldots, x_1,x_0$) and a corresponding weight ω for the last sample x_0 that can be derived from the following equations.

$$\begin{aligned} & p_j = {p_0}^{\beta_j} {p_n}^{1-\beta_j} \\ & \omega = \frac{p_{n-1}(x_{n-1})\,p_{n-2}(x_{n-2})\ldots p_1(x_1)\,p_0(x_0)}{p_n(x_{n-1})\,p_{n-1}(x_{n-2})\ldots p_2(x_1)\,p_1(x_0)} \\ & where\ p_0 = P(x)\ \&\ p_n = \pi(x) \end{aligned} \qquad (9)$$

$\pi(x)$ is the prior distribution and $P(x)$ is the posterior distribution. The sequence of distributions in the markov chain is denoted by $p_0,\ldots,p_n$.

The factor for the sequence of distributions (β) can be a geometric series with the following constraints [20]. This will lead to an easier distribution for sampling [20].

$$\beta_0 = 1 \ \ \& \ \ \beta_n = 0 \ \ \& \ \ \beta_0 > \beta_1 > \ldots > \beta_n \qquad (10)$$

This chain should be repeated N times and each will provide one sample and a corresponding weight. At the end, the expectation of the desired function can be computed as follows.

$$\overline{F} = \sum_{i=1}^{N} F(X_i)\omega_i \Bigg/ \sum_{i=1}^{N} \omega_i \qquad (11)$$

F(x) is the function which we want to calculate its expectation over the distribution of interest and ω_i is the weight of sample (X_i) computed during its markov chain.

For our application, the distribution of interest will be the likelihood and the prior distribution should be the bounding for each parameter. The function of interest should be considered for each parameter separately to compute their expectation to achieve the point with maximum likelihood.

C. Bayesian Information Criterion

After finding the proper point in parameter space that the model will work correctly, BIC should be calculated in order to make judgment between different models. This criterion was first proposed by [24] in order to choose the right dimension for a particular model. First, the estimated maximum likelihood of the desired model should be compared to a baseline Bernoulli model that assign a constant probability for each option [5]. So the BIC will be computed using (12) and (13).

$$G^2 = 2.(Ln(likelihood_{\text{model}}) - Ln(likelihood_{baseline})) \qquad (12)$$

$$BIC = G^2 - \Delta k . Ln(N) \qquad (13)$$

In this study, baseline model is a Bernoulli process which set a constant probability for selecting card from each deck. So this lead to 3 free parameters ($p_1,p_2,p_3,p_4=1-(p_1+p_2+p_3)$) that represents the probability of selecting from decks A, B, C and D. in equation (13) Δk denotes difference between number of free parameters in desired model and Bernoulli model and N is the number of observations or trials which is 100 for this study.

V. Model analysis

Three models with differences in the evaluation part were investigated in this article using the aforementioned methods. For considering the effect of a suitable value function in the model, three models using same functions as expectation and decision sections.

Maximum likelihood was calculated using the Annealed Importance Sampling method and then BIC score was computed in the maximum likelihood condition for each model. Average and standard deviation of BIC score over 14 participants for all models is mentioned in table II.

TABLE II
BIC Score for all participants

Model	Mean	Standard deviation
Error driven value function[a]	34.18	22.42
Error frequency driven value function[a]	181.43	13.61
Reversal learning based value function[b]	208.38	22.74

[a] These models have 4 free parameters and for the BIC score $\Delta k = 1$
[b] This model has 5 free parameters and for the BIC score $\Delta k = 2$

According to table I, error driven value function is not able to mimic human behavior in reality and valuation in human mind is strongly dependent on the frequency of losses. Furthermore, implementation of reversal learning in value functions of models improves BIC score significantly. That's because it is necessary for the subject to change his beliefs about decks during the task and overcome contrasting bias that tend to keeping previous plans.

Implementation of reversal learning adds one more free parameter to the model and increases complexity of the model, however this modification improve BIC score and compensate effect of complexity on that. It proves that reversal learning has much more effect on the maximum likelihood and the most resemblance to the valuation procedure in the human mind.

Figures 1 – 3 demonstrate information on parameters setting in three models. Each figure represents mean and standard deviation of free parameters of each model.

Figure 4 shows that there is high amount of deviation in every parameter in the reversal learning model. This may be result of complexity of the model. As complexity or number of free parameters increases, the model will fit to single set of data and the final model will have a wide range in parameters because of differences in data sets.

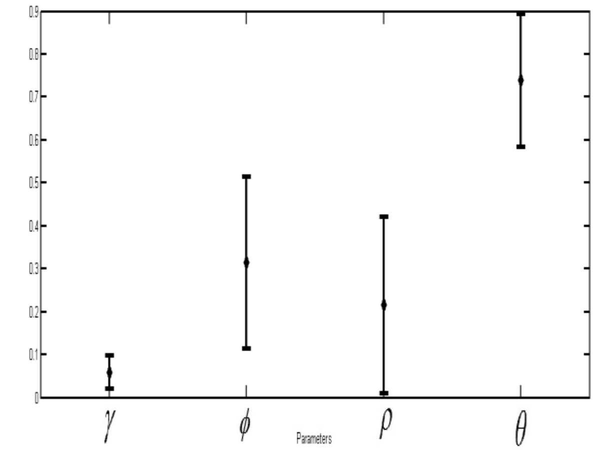

Fig. 2 Mean and standard deviation for parameters in model with *error driven value function,* γ : learning rate, ϕ : discount factor for previous attraction, ρ : experience weight, θ : precision factor.

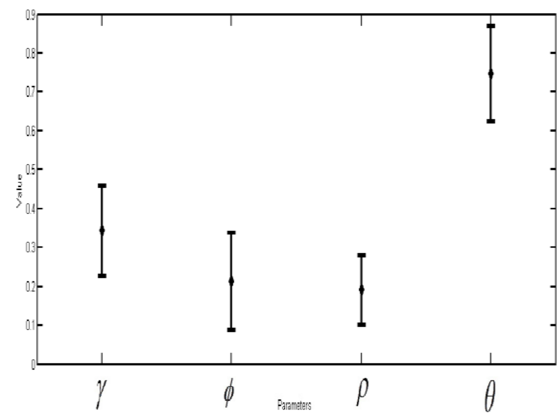

Fig. 3 Mean and standard deviation for parameters in model with *error-frequency driven value function,* γ : learning rate, ϕ : discount factor for previous attraction, ρ : experience weight, θ : precision factor

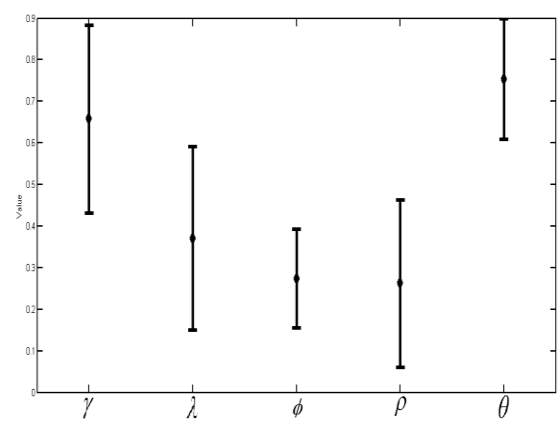

Fig. 4 Mean and standard deviation for parameters in model with *reversal learning based value function,* γ : learning rate, ϕ : discount factor for previous attraction, ρ : experience weight, θ : precision factor

http://dx.doi.org/10.15242/IIE.E0514052

VI. Discussion

In this study, we provide a novel computational model based on previous studies in gambling tasks which emphasizes importance of valuation process in human decision making and investigates its effect on the behavioral modeling. Comparing BIC scores obtained for each model, there is significant increment in BIC after implementing new value functions and it posits that a suitable value function has a drastic impact on the performance of the model.

The first model that uses error driven value function has poor performance in predicting human behavior according to its BIC score. It can be interpreted by analyzing human decision making anatomy. The two parts of the brain that contribute a key role in human emotional process of adaptive decision making repeated choice tasks are basal ganglia and orbitofrontal cortices [25]—[29]. Frank and Claus [14] shows that basal ganglia is sensitive to the frequency of wins and losses. Their basal ganglia model performed optimally in modified version of the IGT that good and bad decks are based on the frequency of losses. According to this evidence and the role of basal ganglia in decision making, the subject is sensitive to both frequency and magnitude of wins and losses concurrently. So models that consider just magnitude of error would not be able to predict human behavior and it is necessary to add terms of frequency based valuation to the model.

For the role of reversal learning in performing this task which is proved by previous studies [31],[30], we add the mathematical interpretation of this kind of learning to the model and the final result posits that this idea make the model more powerful in predicting human behavior in IGT.

Investigating relationship between each parameter in the model and a particular part of human brain may be the focus point of future works in the field of computational modeling human behavior in IGT.

VII. References

[1] Bechara a, Damasio a R, Damasio H, Anderson SW. Insensitivity to future consequences following damage to human prefrontal cortex. Cognition [Internet]. 1994;50(1-3):7–15. http://dx.doi.org/10.1016/0010-0277(94)90018-3

[2] Schwarz, Gideon. "Estimating the dimension of a model." *The annals of statistics* 6.2 (1978): 461-464. http://dx.doi.org/10.1214/aos/1176344136

[3] Yechiam, Eldad, and Jerome R. Busemeyer. "Comparison of basic assumptions embedded in learning models for experience-based decision making." Psychonomic bulletin & review 12.3 (2005): 387-402. http://dx.doi.org/10.3758/BF03193783

[4] Kalidindi, Kiran, and Howard Bowman. "Using ϵ-greedy reinforcement learning methods to further understand ventromedial prefrontal patients' deficits on the Iowa Gambling Task." Neural Networks 20.6 (2007): 676-689. http://dx.doi.org/10.1016/j.neunet.2007.04.026

[5] Ahn, Woo-Young, et al. "Comparison of decision learning models using the generalization criterion method." Cognitive Science 32.8 (2008): 1376-1402. http://dx.doi.org/10.1080/03640210802352992

[6] Biele, Guido, Jörg Rieskamp, and Richard Gonzalez. "Computational models for the combination of advice and individual learning." Cognitive science 33.2 (2009): 206-242. http://dx.doi.org/10.1111/j.1551-6709.2009.01010.x

[7] Milner, Brenda. "Effects of different brain lesions on card sorting: The role of the frontal lobes." Archives of Neurology 9.1 (1963): 90.

[8] Petrides, Michael, and Brenda Milner. "Deficits on subject-ordered tasks after frontal-and temporal-lobe lesions in man." Neuropsychologia 20.3 (1982): 249-262. http://dx.doi.org/10.1016/0028-3932(82)90100-2

[9] Shallice, Tim, and Margaret E. Evans. "The involvement of the frontal lobes in cognitive estimation." Cortex 14.2 (1978): 294-303. http://dx.doi.org/10.1016/S0010-9452(78)80055-0

[10] Damasio, Antonio R., Daniel Tranel, and Hanna Damasio. "Somatic markers and the guidance of behavior: Theory and preliminary testing." Frontal lobe function and dysfunction (1991): 217-229.

[11] Li, Xiangrui, et al. "The Iowa gambling task in fMRI images." Human brain mapping 31.3 (2010): 410-423.

[12] Bechara, A. "Iowa gambling task professional manual." Psychological Assessment Resources, Inc (2007).

[13] Gluck, Mark A., and Gordon H. Bower. "From conditioning to category learning: an adaptive network model." *Journal of Experimental Psychology: General* 117.3 (1988): 227.

[14] Frank, Michael J., and Eric D. Claus. "Anatomy of a decision: striato-orbitofrontal interactions in reinforcement learning, decision making, and reversal." *Psychological review* 113.2 (2006): 300.

[15] Rolls, Edmund T. "The functions of the orbitofrontal cortex." *Neurocase* 5.4 (1999): 301-312. http://dx.doi.org/10.1080/13554799908411984

[16] Camerer, Colin, and Teck Hua Ho. "Experience-weighted Attraction Learning in Normal Form Games." *Econometrica* 67.4 (1999): 827-874. http://dx.doi.org/10.1111/1468-0262.00054

[17] McKelvey, Richard D., and Thomas R. Palfrey. "Quantal response equilibria for normal form games." *Games and economic behavior* 10.1 (1995): 6-38. http://dx.doi.org/10.1006/game.1995.1023

[18] Luce, R. Duncan. "Individual choice behaviour: a theoretical analysis." *New York* 142 (1959).

[19] McFadden, Daniel. "Conditional logit analysis of qualitative choice behavior." (1973): 105.

[20] Neal, Radford M. "Annealed importance sampling." *Statistics and Computing* 11.2 (2001): 125-139. http://dx.doi.org/10.1023/A:1008923215028

[21] Metropolis, N., et al. "Calculation of equations of state by fast computing machines." *J. chem. Phys* 21 (1953): 1087-1091. http://dx.doi.org/10.1063/1.1699114

[22] Hastings, W. Keith. "Monte Carlo sampling methods using Markov chains and their applications." *Biometrika* 57.1 (1970): 97-109. http://dx.doi.org/10.1093/biomet/57.1.97

[23] Geman, Stuart, and Donald Geman. "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images." *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 6 (1984): 721-741. http://dx.doi.org/10.1109/TPAMI.1984.4767596

[24] Schwarz, Gideon. "Estimating the dimension of a model." *The annals of statistics* 6.2 (1978): 461-464. http://dx.doi.org/10.1214/aos/1176344136

[25] Beiser, David G., and James C. Houk. "Model of cortical-basal ganglionic processing: encoding the serial order of sensory events." *Journal of Neurophysiology* 79 (1998): 3168-3188.

[26] Brown, Joshua, Daniel Bullock, and Stephen Grossberg. "How the basal ganglia use parallel excitatory and inhibitory learning pathways to selectively respond to unexpected rewarding cues." *The journal of neuroscience* 19.23 (1999): 10502-10511.

[27] Mink, Jonathan W. "The basal ganglia: focused selection and inhibition of competing motor programs." *Progress in neurobiology* 50.4 (1996): 381-425. http://dx.doi.org/10.1016/S0301-0082(96)00042-1

[28] Tremblay, Léon, and Wolfram Schultz. "Modifications of reward expectation-related neuronal activity during learning in primate orbitofrontal cortex." *Journal of neurophysiology* 83.4 (2000): 1877-1885.

[29] Schoenbaum, Geoffrey, et al. "Encoding predicted outcome and acquired value in orbitofrontal cortex during cue sampling depends upon input from basolateral amygdala." *Neuron* 39.5 (2003): 855-867. http://dx.doi.org/10.1016/S0896-6273(03)00474-4

[30] Fellows, Lesley K., and Martha J. Farah. "Ventromedial frontal cortex mediates affective shifting in humans: evidence from a reversal learning paradigm." *Brain* 126.8 (2003): 1830-1837. http://dx.doi.org/10.1093/brain/awg180

[31] Fellows, Lesley K., and Martha J. Farah. "Different underlying impairments in decision-making following ventromedial and dorsolateral frontal lobe damage in humans." *Cerebral cortex* 15.1 (2005): 58-63. http://dx.doi.org/10.1093/cercor/bhh108