Understanding Addictive Behavior on the Iowa Gambling Task Using Reinforcement Learning Framework

Amir K. Dezfuli (akdezfuli@ece.ut.ac.ir)
Center of Excellence for Control and Intelligent Processing
Department of Electrical and Computer Engineering, University of Tehran, Iran

Mohammad Mahdi Keramati (mm_keramati@gsdm.sharif.edu)
School of Management and Economic, Sharif University of Technology, Iran

Hamed Ekhtiari (h_ekhtiari@razi.tums.ac.ir)
Cognitive Assessment Laboratory, Iranian National Center for Addiction Studies, Iran

Hooman Safaei (hooman.safaei@gmail.com)
Cognitive Assessment Laboratory, Iranian National Center for Addiction Studies, Iran

Caro Lucas (lucas@ut.ac.ir)
Center of Excellence for Control and Intelligent Processing
Department of Electrical and Computer Engineering, University of Tehran, Iran

Abstract

Neurocognitive decision-making disorders in Iowa Gambling Task (IGT) can be better understood in the light of computational modeling methods. In this study, we use Reinforcement Learning (RL) framework to decompose subjects behavior into its underlying factors. Both healthy subjects and Substance Dependent Individuals (SDIs) show poor performance in the task, with significant decline in SDIs (net score = -2.3) compared with control subjects (net score = 6.2). Fitting various models of RL family, the results show that for both groups, frequency-based learning model coupled with softmax exploration strategy for action selection is the best descriptor model for choices of subjects in the task (based on Bayesian Information Criterion). So, being under the influence of reinforcer frequency instead of its magnitude is the major factor behind poor performance of subjects. In addition, sensitivity analysis shows that the performance of the best fitted model is sensitive to the valence weight parameter in SDIs. The estimated value of the parameter reveals that higher deviation of SDIs to harm-avoidance characteristic (in relation to healthy subjects) causes performance difference between two groups. Neural and cultural discussions are also presented to explain the results.

Keywords: Iowa Gambling Task; Reinforcement Learning Addiction; Decision Making; Computational Modeling.

Intoduction

Addiction is characterized as compulsive drug use, despite awareness of the deleterious future consequences (Hyman & Malenka, 2001) The transition from regulated to compulsive drug use is rooted in actions of drugs of abuse on a vulnerable brain. Changing motivational circuitry is followed by associated alterations in several psychological functions, such as decision making. Such drug-induced decision making malfunctions are evidenced to be generalized to real-life circumstances. This provides researchers to investigate addicts brain disorders via tasks which simulate real-life decision making situations. The Iowa Gambling Task (IGT) (Bechara, Damasio, Damasio, & Anderson, 1994), which originally was introduced to shed light on decision making deficits in bilateral ventromedial prefrontal patients, is a widely used framework for measuring decision making ability. In ABCD version of the task, participants make a series of 100 choices from four decks of cards. Two of the decks are advantageous (decks C and D) and two of them are disadvantageous (decks A and B). The subjects goal is to maximize its net score across trials. The two disadvantageous (bad) decks lead to relatively high gains ($100) but also to occasional large losses ($125), which result in an average loss of -$25 per trial. The two advantageous (good) decks lead to lower gains each time ($50) but produce smaller losses, resulting in an average gain of +$25 per trial. Performance of a subject in the task is defined as difference of number of cards selected from good decks minus cards selected from back decks (net score: (C+D) (A+B)). In respect of healthy subject, Substance Dependent Individuals (SDIs) commonly show decision making deficit in gambling task(Grant, Contoreggi, & London, 2000; Bechara et al., 2001; Bechara & Damasio, 2002; Bechara & Martin, 2004) To understand processes behind decision making impairments in gambling task, Busemeyer and Stout (2002) in their seminal work, have utilized cognitive modeling method. This approach makes it possible to track the revealed behavior of the decision maker back to its underlying causes. The models are designed such that their parameters have meaningful cognitive interpretations, so the modeler can reduce observed behavior to parameters values (or model structure) and gains information about the cognitive properties of modeled subject(s). In the same motivational line with previous works (Kalidindi & Bowman, 2007; Stout, Busemeyer, Lin, Grant, & Bonson, 2004) in this study we are to identify the principal components which influence behavior of control group and SDIs. SDI subjects (n=217) are male treatment seeking opioid dependents (based on DMS-IV (American Psychiatric Association, 2000)) referred to Iranian National Center for Addictive Studies (INCAS). Age, sex and education matched

control subjects (n=130) are included from patients relatives with no history of drug abuse (except cigarette). Demographics of control and SDI subjects are presented in Table . As Table shows, both groups have poor IGT performance with reference to previous studies (net score < 10) (Bechara et al., 2001). Reinforcement Learning (RL) framework (Sutton & Barto, 1998) is used to answer why subjects have decision making deficit in IGT. Furthermore, SDIs have significant weaker performance compared with healthy subjects in IGT ($P \leq 0.002$); we also make use of modeling approach to see what lies beneath this difference. The following section is dedicated to introduction of RL models used in this paper. The method of estimating models' free parameters, using IGT data will be presented in the third part. Findings are discussed and concluded based on numerical results in the last section.

Model Description

RL addresses the problem of decision making in an uncertain environment. Alongside with rich mathematical foundation, it has been shown that RL has strong structural relevancy to underlying neural mechanism of value learning and action selection (Daw, 2003). In order to act optimally, a RL agent desires to know expected value of each action. So, it makes use of precepts and rewards to learn value of states and actions, and then utilizes the result of learning for decision making. There are various methods for value learning and action selection. Each method is governed by some free parameters which shape behavior of the model. In the following section we introduce some variants of RL family.

Learning Methods

After the decision maker makes a decision and experiences the result, it weights positive and negative rewards differently as in prospect theory (Kahneman & Tversky, 1979). Based on Expectancy-Valence Learning Model (Busemeyer & Stout, 2002), this property can be modeled with a weighted average between rewards and punishments:

$$r_t(a) = w.r_t^+(a) + (1-w).r_t^-(a) \qquad (1)$$

Where $r_t^+(a)$ and $r_t^-(a)$ are respectively reward and punishment received after execution of action a. Parameter w is named valence weight ($0 \leq w \leq 1$) (Kalidindi & Bowman, 2007). Values of w near 1 mean reward-seeking characteristic and values near 0 indicate harm avoidance behavior. The question of how the valence signal $r_t(a)$ can be used to update the value of each action is answered in the following.

Sample Averaging (Kalidindi & Bowman, 2007) The value of each action can be estimated as the average of all valences experienced before:

$$Q_T(a) = \frac{1}{K_a} \sum_{t=1}^{T} r_t(a) \qquad (2)$$

K_a is the total number of times action a has been taken

prior to time T.

Variance-Driven Learning (Kalidindi & Bowman, 2007) In this method, risk seeking behavior is aimed to be modeled:

$$Q_T(a) = \frac{1}{K_a} \sum_{t=1}^{T} (r_t(a) - \bar{r}_t(a))^2 \qquad (3)$$

$r_t(a)$ is the average of past received reinforcers.

Frequency-Driven Learning (Kalidindi & Bowman, 2007) Instead of using magnitude of received valences for value prediction, frequency of valences is used in this model:

$$Q_t(a) = \begin{cases} Q_{t-1}(a) + 1 & r_t(a) > 0 \\ Q_{t-1}(a) - 1 & r_t(a) < 0 \\ Q_t(a) & else \end{cases} \qquad (4)$$

Error-Driven Learning Learning in this method is done via calculating the error signal which is the difference between expected and observed value of an action. The error signal is used to update the value of the respective action:

$$Q_t(a) = Q_{t-1}(a) + \gamma(r_t(a) - Q_{t-1}(a)) \qquad (5)$$

In this model, recent experiences are weighted more than distant ones. In (5) the parameter γ is learning rate ($0 < \gamma < 1$). Large value of γ makes fast changes in estimated values and rapid forgetting of previous experiences, while small values produce slow changes and declines the effect of recent experiences.

Error-Frequency Learning (Kalidindi & Bowman, 2007) This method is similar to the previous one, but instead of using the magnitude of received valence to update the value of an action, the number of times that reward (punishment) has been received is used for value updating:

$$Q_t(a) = \begin{cases} Q_{t-1}(a) + \gamma(1 - Q_{t-1}(a)) & r_t(a) > 0 \\ Q_{t-1}(a) - \gamma(1 + Q_{t-1}(a)) & r_t(a) < 0 \\ (1-\gamma)Q_{t-1}(a) & else \end{cases} \qquad (6)$$

Reversal Learning (Kalidindi & Bowman, 2007) The logic behind reversal learning is that learning is slowed down if the expected value of executing an action is in the opposite sign of its experienced value:

$$
\begin{aligned}
if \quad & sign(Q_{t-1}(a)) = sign(Z) \quad then \\
& Q_t(a) = Z \\
else \\
& Q_t(a) = Q_{t-1}(a) + \lambda.\gamma(r_t(a) - Q_{t-1}(a)) \\
where \\
& Z = Q_{t-1}(a) + \gamma(r_t(a) - Q_{t-1}(a))
\end{aligned}
\qquad (7)
$$

	Variables	Age (year)	Education (year)	(C+D) – (A+B)	(B+D) –(A+C)
Groups	Healthy Subjects (n=130)	30.25 ± 8.77	12.03 ± 3.77	6.88 ± 28.47	19.65 ± 21.55
	Opioid Dependent Subjects (n=217)	29.87 ± 7.60	11.49 ± 3.00	-2.39 ± 26.10	17.52 ± 21.85

Parameter λ is reversal-deficit rate $(0 < \lambda < 1)$. If $\lambda = 1$, then reversal learning and error driven algorithms will be the same, which means that learning will not be slowed down in any case.

Amygdala-OFC Learning This method is inspired by computational model of amygdala introduced by (Balkenius & Morén, 2001). The model assumes the role of amygdala in value learning:

$$Q_t^A(a) = Q_{t-1}^A(a) + \gamma_A \left[Max\left(R - Q_{t-1}^A(a), 0\right) \right] \quad (8)$$

And the role of orbitofrontal cortex (OFC) for devaluation of estimations previously learned by amygdala:

$$
\begin{aligned}
if \quad & R \neq 0 \quad then \\
& Q_t^O(a) = \\
& Q_{t-1}^O(a) + \gamma_O \left[Max\left(Q_{t-1}^A(a) - R, 0\right) - Q_{t-1}^O \right] \\
else \\
& Q_t^O(a) = \\
& Q_{t-1}^O(a) + \gamma_O \left[Max\left(Q_{t-1}^A(a) - Q_{t-1}^O, 0\right) \right]
\end{aligned}
\quad (9)
$$

$Q_t(a)$ is final estimated value which is output of OFC subtracted from amygdalas output:

$$Q_t(a) = Q_t^A(a) - Q_t^O(a) \quad (10)$$

γ_A and γ_A are respectively learning rates in amygdala and OFC $(0 \leq \gamma_A \leq 1, 0 \leq \gamma_O \leq 1)$. Notice that, in this model, because of parallel role of γ_A and γ_O with w, positive and negative reinforcers are not weighted.

Action Selection Methods Values learned by the learning module are used for action selection. The main challenge is deciding when to exploit in previous knowledge and when to explore the environment. Exploration allows the agent to gain better estimations of expected values. Followings are two well-known methods for balancing exploration and exploitation strategies.

ε-Greedy Action Selection The parameter ε in the ε-greedy algorithm is the probability of exploration and $1 - \varepsilon$ is the probability of selecting the action with the highest expected value $(0 < \varepsilon < 1)$. Exploration here means each action is equiprobable to be selected, regardless of its expected value. Let $X(t)$ be a random number with uniform distribution that ranges form 0 to 1, G set of possible actions at time t and Z set of actions that have maximum Q-value at time t. Then probability of selecting each action is as follow:

$$
\begin{aligned}
if \quad & X(t) > \varepsilon \quad then \\
& P(a) = \begin{cases} 1/|Z| & a \in Z \\ 0 & a \notin Z \end{cases} \\
else \\
& P(a) = \begin{cases} 1/|G| & a \in G \\ 0 & a \notin G \end{cases}
\end{aligned}
\quad (11)
$$

Softmax Action Selection In this model, the probability of choosing action a at time t is exponentially proportional to its expected value. So unlike the previous method, probability of selecting an action is sensitive to its value:

$$P_t(a) = \frac{e^{\beta \cdot Q_{t-1}(a)}}{\sum_i e^{\beta \cdot Q_{t-1}(a_i)}} \quad (12)$$

Estimation Method

Reinforcement learning model is a nonlinear, stochastic and dynamic system with a set of free parameters (P). Obviously, behavior of a model is sensitive to its parameters' values. We desire to find parameter vector which makes model's output most similar to observed behavior from subjects in IGT For this purpose, we used maximum likelihood estimation method. To calculate the likelihood function, we need to have the system's (Control and SDI) outputs at each period of time $(T = 1 \ldots 100)$ and also probability distribution function of model output at each period of time for each parameter vector $(P_j = (\beta, w, \gamma))$. The former, denoted by $O_{C,i,t}(a_k)$ (for control group) and $O_{A,i,t}(a_k)$ (for SDIs), is total number of choosing action a_k at time t. The later, denoted by $Pr_t(P_j, a_k)$, is the probability of choosing action a_k at time t by the model governed by P_j. In order to estimate $Pr_t(P_j, a_k)$, due to model's stochastic character, the model has been simulated for 3000 times for each P_j and the average of outputs are used for estimation of $Pr_t(P_j, a_k)$. Therefore, the likelihood function can be formulated as:

$$f^{Control}(y \mid P_j) = \prod_{i=1}^{N} \prod_{t=1}^{100} \prod_{k=1}^{4} Pr_t(P_j, a_k)^{O_{C,i,t}(a_k)} \quad (13)$$

$$f^{SDI}(y \mid P_j) = \prod_{i=1}^{M} \prod_{t=1}^{100} \prod_{k=1}^{4} Pr_t(P_j, a_k)^{O_{A,i,t}(a_k)} \quad (14)$$

k stands for taking actions A, B, C or D by the agent $(1 \leq k \leq 4)$. N and M are the number of Control and SDI subjects respectively. The maximum likelihood rule implies:

$$P_{Control}^* = \arg\max_j f^{Control}(y|P_j) \quad (15)$$

$$P^*_{SDI} = \arg\max_j f^{SDI}(y|P_j) \qquad (16)$$

For models with less than four degree of freedom we used exhaustive search to find parameter vector which satisfies (15) and (16). For models with four degree of freedom because of intractable computational time of exhaustive search, genetic algorithm method was used.

Results and Conclusion

Numerical results of estimations are presented in Table 2. Bayesian Information Criterion (BIC) which considers both model fitness and complexity is used for model selection:

$$BIC = -2\ln f\left(y \mid \hat{P}_j\right) + k\ln n \qquad (17)$$

In (17), n represents the number of data points and k denotes degree of freedom (number of free parameters) in the model. Based on BIC criteria, the best-fitted model for both groups, SDI and control subjects, is error-frequency learning with softmax rule for action selection. Optimal parameters are $(P_j = (\beta, w, \gamma))$:

$$P^*_{Control} = (0.70,\ 0.35,\ 0.2) \qquad (18)$$

$$P^*_{SDI} = (0.50,\ 0.15,\ 0.35) \qquad (19)$$

The best-fitted models were validated for control group by chi-square goodness of fit test for each trial:

$$chi-square\,(trial_t) =$$

$$\sum_{k=1}^{4} \frac{\left(\sum_{i=1}^{N} O_{C,i,t}(a_k) - N*Pr_t(P^*_{Control}, a_k)\right)^2}{N*Pr_t(P^*_{Control},\ a_k)}$$

As like, for SDIs (20) was used with corresponding values. The best fitted model for control group satisfies fitness criteria $(df=3, p<0.01)$ in 96 trails (out of 100 trial) and in SDIs best fitted model satisfies 94 trails (out of 100 trial). These results indicate model validity for describing subjects' choices. Performance comparison between best-fitted model for control group and averaged control data is presented in Figure 1 (notice that goal of estimation was not fitting performances, estimations are done so that model's selections are best fitted on subjects' choices).

Influencing control subjects by reinforcer frequency misguides them to choose bad cards. This may justify poor performance of control group (net score = 6.2 < 10). In previous modeling of IGT, sample-averaging provides best match model for healthy subjects (Kalidindi & Bowman, 2007). Ambiguity of gambling concepts such as amount of facsimile monetary rewards and punishments among Iranian subjects, due to religious limitations for gambling in Islamic law, may play an important role in frequency-based valuation in control subjects (Ekhtiari, Behzadi, Jannati, & Mokri, 2002). Cultural aspects of risky decision making (Weber & Hsee, 1998) such as aversion from frequent punishment or priority of punishment times over its value in emotional process of expected choice out comes may act as another possible cause.
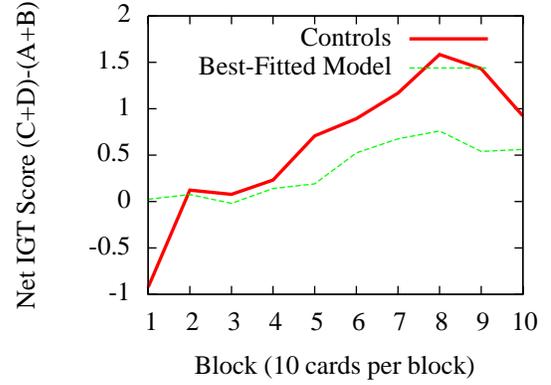
Figure 1: Performance of control subjects and best-fitted model
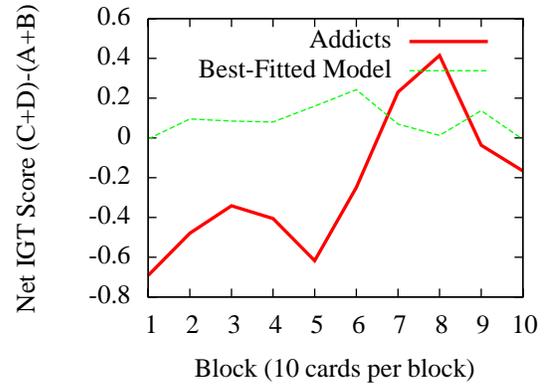
Figure 2: Performance of SDIs and best-fitted model

Figure 2 shows performance of SDIs and best-fitted model. Like control subjects, poor performance of SDIs (net score = -2.3 < 10) is partially down to ignorance of reward magnitude. It may be as a result of depression of neural processing in prefrontal parts after chronic exposure to drugs (Jentsch & Taylor, 1999). This decrease in activity of frontocortical parts may increase activity of sub cortical dopamine regions (Meyer-Lindenberg et al., 2002). Thus, decision making become more dependent on basal ganglia, which is known to have important role in representation of reward and punishment frequency (Frank & Claus, 2006). But, if both groups decide based on reinforcer frequency, what causes substantial difference between controls and SDIs performances?

Figure 3 shows performance of the softmax error-frequency model with respect to its parameters at point P^*_{SDI}. As it is evident, performance of the model is sensitive to parameter w (valence weight) and the two other parameters have no significant effect on performance of the model. (at $w = 0.35$ the SDIs model meets performance of best-fitted model for control group). So, it seems deviation to harm avoidance in SDIs, in addition to being under the influence of reward frequency, is the major factor of SDIs poor per-

Table 2: Parameter Estimation Results

Learning	ε -Greedy Action Selection							Softmax Action Selection						
	$\hat{w}$	$\hat{\varepsilon}$	$\hat{\gamma}$	$\hat{\lambda}$	$\hat{\gamma}_A$	$\hat{\gamma}_O$	BIC	$\hat{w}$	$\hat{\varepsilon}$	$\hat{\gamma}$	$\hat{\lambda}$	$\hat{\gamma}_A$	$\hat{\gamma}_O$	BIC
Sample Averaging														
Control	0.00	0.88	-	-	-	-	35672	0.36	0.00	-	-	-	-	35936
SDI	0.66	0.94	-	-	-	-	59937	0.78	0.00	-	-	-	-	60158
Variance Driven														
Control	0.68	0.00	-	-	-	-	35978	0.67	0.80	-	-	-	-	35821
SDI	0.89	0.95	-	-	-	-	59881	0.67	0.80	-	-	-	-	59872
Frequency Driven														
Control	0.00	0.77	-	-	-	-	35425	0.46	0.04	-	-	-	-	35481
SDI	0.24	0.81	-	-	-	-	59442	0.00	0.07	-	-	-	-	59411
Error Driven														
Control	0.00	0.55	1.00	-	-	-	35436	0.90	0.08	0.30	-	-	-	35921
SDI	0.00	0.60	1.00	-	-	-	59547	0.95	0.10	0.90	-	-	-	59948
Error Frequency														
Control	0.35	0.80	0.10	-	-	-	35423	0.35	0.70	0.20	-	-	-	35414
SDI	0.20	0.55	1.00	-	-	-	59439	0.15	0.50	0.35	-	-	-	59395
Reversal Learning														
Control	0.60	0.80	0.33	0.34	-	-	35534	0.36	0.00	0.12	0.14	-	-	35566
SDI	0.57	0.83	0.31	0.27	-	-	59503	0.45	0.00	0.25	0.16	-	-	59531
Amygdala-OFC														
Control	-	0.70	-	-	0.20	0.75	35570	-	0.00	-	-	0.10	0.05	35965
SDI	-	0.80	-	-	0.20	0.75	59542	-	0.00	-	-	0.05	0.00	59915
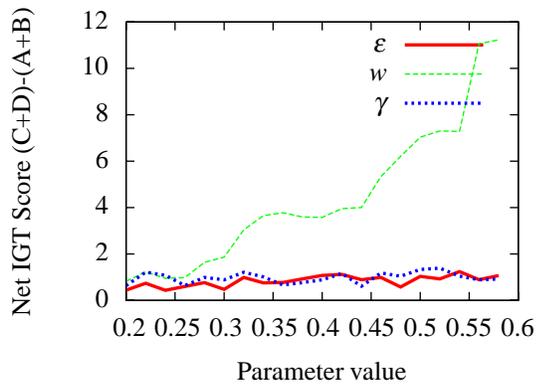
Figure 3: Performance of the model with respect to its parameters at point P^*_{SDI}

formance. These findings are compatible with various lines of investigations that suggest alteration of sensitivity to reinforcement in chronic substance users (Ersche et al., 2005). Also higher degree of punishment aversion and harm avoidance is reported in previous studies with self report measures in chronic abusers and treatment seeking subjects (Abbate-Daga, Amianto, Rogna, & Fassino, 2007). Fitness of softmax model over ε-Greedy is consistent with previous modeling of gambling task (Daw, O'Doherty, Dayan, Seymour, & Dolan, 2006). In ABCD version of IGT, first cards of bad decks are more rewarding in comparison with first cards of good decks. Such card arrangement causes ε-Greedy exploration strategy to choose more cards from bad decks in few first choices. This early behavioral convergence differs from human one that shows more exploratory pattern in first choices. Softmax rule does not suffer from this fallacy, (because of its value sensitive exploration strategy) and this property may underlie its superiority over ε-Greedy method. Designing new variant versions of IGT to evaluate error frequency learning in healthy subjects, providing some statistical opportunity to fit the models for each subject (that makes it possible to have some analysis to assess significance of differences among models and among groups), clustering opioid dependent subjects into different groups (based on addiction severity, possible comorbidities, history of imprisonment and anti-social behaviors and types of drug of usage, Heroin injection, Opium smoking or Heroin sniffing and smoking) and performing other studies on non-treatment seeker, abstinent and under substitution treatment (such as methadone) SDI groups can be the next steps toward better understanding of addictive behavior.

References

Abbate-Daga, G., Amianto, F., Rogna, L., & Fassino, S. (2007). Do anorectic men share personality traits with opiate dependent men? a case-control study. *Addictive Behaviors*, *32*, 170-174.

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders dsm-iv-tr (text revision)* (4 Sub ed.). American Psychiatric Publishing, Inc.

Balkenius, C., & Morén, J. (2001). Emotional learning: A computational model of the amygdala. *Cybernetics and Systems*, *32*, 611-636.

Bechara, A., Damasio, A., Damasio, H., & Anderson, S. (1994). Insensitivity to future consequences following damage to human prefrontal cortex. *Cognition*, *50*, 7-15.

Bechara, A., & Damasio, H. (2002). Decision-making and addiction (part i): impaired activation of somatic states in substance dependent individuals when pondering decisions with negative future consequences. *Neuropsychologia*, *40*, 1675-89. (PMID: 11992656)

Bechara, A., Dolan, S., Denburg, N., Hindes, A., Anderson, S. W., & Nathan, P. E. (2001). Decision-making decits, linked to a dysfunctional ventromedial prefrontal cortex, revealed in alcohol and stimulant abusers. *Neuropsychologia*, *39*, 376-89. (PMID: 11164876)

Bechara, A., & Martin, E. M. (2004). Impaired decision making related to working memory decits in individuals with substance addictions. *Neuropsychology*, *18*, 152-62. (PMID: 14744198)

Busemeyer, J. R., & Stout, J. C. (2002). A contribution of cognitive decision models to clinical assessment: decomposing performance on the bechara gambling task. *Psychological assessment*, *14*, 253-62. (PMID: 12214432)

Daw, N. D. (2003). *Reinforcement learning models of the dopamine system and their behavioral implications*. Unpublished doctoral dissertation, Carnegie Mellon University. (Chair-David S. Touretzky)

Daw, N. D., O'Doherty, J. P., Dayan, P., Seymour, B., & Dolan, R. J. (2006). Cortical substrates for exploratory decisions in humans. *Nature*, *441*, 876-879.

Ekhtiari, H., Behzadi, A., Jannati, A., & Mokri, A. (2002). Cultural characteristics in standard and variant persian versions of iowa gambling task. *New Advances in Cognitive Sciences*, *3*, 46-57.

Ersche, K. D., Roiser, J. P., Clark, L., London, M., Robbins, T. W., & Sahakian, B. J. (2005). Punishment induces risky decision-making in methadone-maintained opiate users but not in heroin users or healthy volunteers. *Neuropsychopharmacology : oicial publication of the American College of Neuropsychopharmacology*, *30*, 2115-24. (PMID: 15999147)

Frank, M. J., & Claus, E. D. (2006). Anatomy of a decision: striato-orbitofrontal interactions in reinforcement learning, decision making, and reversal. *Psychological review*, *113*, 300-26. (PMID: 16637763)

Grant, S., Contoreggi, C., & London, E. D. (2000). Drug abusers show impaired performance in a laboratory test of decision making. *Neuropsychologia*, *38*, 1180-7. (PMID: 10838152)

Hyman, S. E., & Malenka, R. C. (2001). Addiction and the brain: the neurobiology of compulsion and its persistence. *Nat Rev Neurosci*, *2*, 695703.

Jentsch, J. D., & Taylor, J. R. (1999). Impulsivity resulting from frontostriatal dysfunction in drug abuse: implications for the control of behavior by reward-related stimuli. *Psychopharmacology*, *146*, 373-90. (PMID: 10550488)

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, *47*, 263-291. (ArticleType: primary article / Full publication date: Mar., 1979 / Copyright 1979 The Econometric Society)

Kalidindi, K., & Bowman, H. (2007). Using -greedy reinforcement learning methods to further understand ventromedial prefrontal patients' decits on the iowa gambling task. *Neural Netw.*, *20*, 676-689.

Meyer-Lindenberg, A., Miletich, R. S., Kohn, P. D., Esposito, G., Carson, R. E., Quarantelli, M., et al. (2002). Reduced prefrontal activity predicts exaggerated striatal dopaminergic function in schizophrenia. *Nature neuroscience*, *5*, 267-71. (PMID: 11865311)

Stout, J. C., Busemeyer, J. R., Lin, A., Grant, S. J., & Bonson, K. R. (2004). Cognitive modeling analysis of decision-making processes in cocaine abusers. *Psychonomic bulletin & review*, *11*, 742-7. (PMID: 15581127)

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning: An introduction*. Cambridge, MA: MIT Press.

Weber, E. U., & Hsee, C. (1998). Cross-cultural dierences in risk perception, but cross-cultural similarities in attitudes towards perceived risk. *Management Science*, *44*, 1205-1217.