이학석사학위논문

Individual differences in dynamic criterion shifts during perceptual decision-making

지각적 의사 결정 동안의 최적 기준 변경의 개인차

2013년 2월

서울대학교 대학원

뇌인지과학 전공

Issac Rhim

Individual differences in dynamic criterion shifts during perceptual decision-making

지도교수 이 상 훈

이 논문을 이학석사 학위논문으로 제출함.

2012년 12월

서울대학교 대학원

뇌인지과학 전공

Issac Rhim

Issac Rhim의 이학석사 학위논문을 인준함.

2013년 2월

위 원 장 이 인 아 (인)

부 위 원 장 이 상 훈 (인)

위 원 Randolph Blake (인)

Abstract

Individual differences in dynamic criterion shifts during perceptual decision-making

Issac Rhim
Brain and Cognitive Sciences
College of Natural Sciences
Seoul National University

Perceptual decision-making involves placing an optimal criterion on the axis of encoded sensory evidence to maximize outcomes for choices. Optimal criterion setting becomes critical particularly when neural representations of sensory inputs are noisy and feedbacks for perceptual choices vary over time in an unpredictable manner. Here, we first induced shifts in criteria by adopting stochastically generated reverse feedbacks, so that subjects sometimes received false feedbacks even for responses they should have gotten correct. Using a reinforcement-learning model, we captured subjects' behavior and we investigated rationale for optimal criteria placement, or lack of, in individuals. Utilizing the estimates of model parameters, specifically the shift rate for criteria and subjects' sensitivity, we identified the optimal regime. We demonstrate that optimal decision-making is closely correlated with the subject's sensitivity and shift rate. There

was a systematic trade-off between the two parameters indicating that sensitivity and shift rate reciprocate, allowing for optimal decision-making in complex environments.

Keywords: individual differences, perceptual decision-making, criterion

Student Number: 2010-24027

Contents

Figures

Introduction

In value-based decision tasks, subjects make choices based on their payoffs (Sugrue et al, 2004; Rangel et al., 2008). In sensory decision tasks, subjects are rewarded equally for correct choices regardless of which option was chosen (Romo and Slinas, 2003; Gold and Shadlen, 2007). In dynamic decision-making environments, however, sensory evidences are uncertain and rewards vary. Thus, observers must integrate sensory evidence and reward to optimize payoff. An ideal observer must not only have accurate sensory system, but also have a highly plastic reward integration system for optimal rewards possible (Jazayeri and Movshon, 2006; Navalpakkam et al., 2010).

Optimal decision-making in dynamic foraging environments have been studied extensively (Diederich and Busemeyer, 2006; Diederich, 2008; Bogacz et al., 2011; Teichert and Ferrera, 2010), however the question still remains why subjects do not fully optimize the magnitude of their reward bias, why species behavior differ in optimal decision-making tasks (Stuttgen et al., 2011; Feng et al., 2009; Gao et al., 2011), and what mechanisms are involved in choosing a decision-making strategy (Wozny et al., 2010).

In this paper, we investigated the source of optimality. In our experiment, we induced criteria shifts in subjects by adopting stochastically generated reverse feedbacks. This often led to false incorrect feedbacks for responses that subjects answered correctly. Using a reinforcement-learning model (Kahnt et al., 2009; Law and Gold, 2009), we captured subjects' behavior, and utilizing model parameters we found that optimal

decision-making is closely correlated with subjects' sensory sensitivity. In parallel, we found optimal decision-making is also closely correlated with subject's shift rate, the level at which how fast a subject can adapt for ideal criteria. We did not find a significant relationship between subjects' sensitivity and shift rate. However, data from individual differences suggest there is a robust, systematic pattern between sensitivity and shift rate, thus indicating that two parameters reciprocate to contribute in optimal perceptual decision-making.

Materials and Methods

Subjects. Thirty participants (13 females) between eighteen and thirty years of age (mean ± SD, 23.23 ± 2.93) with normal or corrected-to-normal vision participated in the study. Participants were recruited from the Seoul National University (SNU) community. Each subject gave written informed consent per Institutional Review Board requirements at SNU and were compensated approximately $10/h.

Stimuli and task. Visual stimuli were generated using the MGL tools implemented in MATLAB (version 7.1; MathWorks), presented at a frame rate of 60 Hz. Button press responses were made on standard keyboard numpad using the index and middle fingers of the right hand in a psychophysics darkroom.

Subjects performed a 2-alternative forced choice (2AFC) perceptual discriminating task (Figure 1A). Subjects were shown a thin (.07°) black-and-white ring stimulus at Gaussian-noise filtered contrast which flickered at 20 Hz on a gray luminance background. Subjects' task was to classify the stimulus ring into small or large rings. On each trial, a fixation appeared for 500 milliseconds (time jittered from 300 to 700 milliseconds) before the ring stimulus onset. The ring size was pseudorandomly selected with equal probability from one of five possible sizes varying in visual angle degrees (3.83, 3.92, 4.00, 4.08, 4.17°). Subjects were allowed to make a response any time after the stimulus onset; the stimulus was presented for 300 milliseconds with response time window closure at 1200 milliseconds from stimulus onset, followed by a feedback period of 500 milliseconds. Subjects were instructed to maximize correct trials. Here, we

adopted stochastically generated reverse feedback by presetting answers prior to the experiment run to determine which trials are labeled large and small. The feedback schedule for a given run was pseudorandomly selected with equal probability from three possible schedules; two schedules were biased in number of either small or large-correct trials, and one schedule was unbiased with equal number of small and large-correct trials. A single experiment session consisted of six block runs with hundred-eighty trials per run. Each run began with forty-five trials (jittered from forty to fifty trials) of unbiased, neutral schedule. All subjects performed five sessions. Before each run, we showed the subjects the mean ring size (4.00°) for duration of 15 seconds. Given the possibility of carry-over effect of bias from previous runs (Aberg& Herzog, 2012), we primed the subjects with the mean ring size in addition to initial neutral schedule to minimize the effect. At the end of each run, subjects were also shown block feedback of their performance and number of incorrect trials due to no response. To control for handedness bias, subjects were randomly assigned to small-large or large-small button configurations. Subjects were not paid additionally based on performance.

Perceptual learning. Perceptual learning was measured by computing the percent correct for each run across sessions averaged for all subjects. In addition, psychometric function for each session was estimated using cumulative Gaussian function to assess improvements in sensitivity, indicated by its slope, sigma(σ).

Induced criterion shift analysis. We analyzed the behavioral data to observe patterns for induced criterion shifts. Sessions were sorted into three schedules and averaged across

subjects. Given the jitter used in each run, 10-trials were discarded for each run in order to align the feedback schedule sequences. Responses were binned in sizes of fifteen trials with sliding bin at interval of one trial. For each stimulus type, the number of large ring responses were gathered and each bin was psychometrically fit with cumulative Gaussian function to give mu (μ) estimates. The μ estimates were plotted to show shifts in criterion across trials.

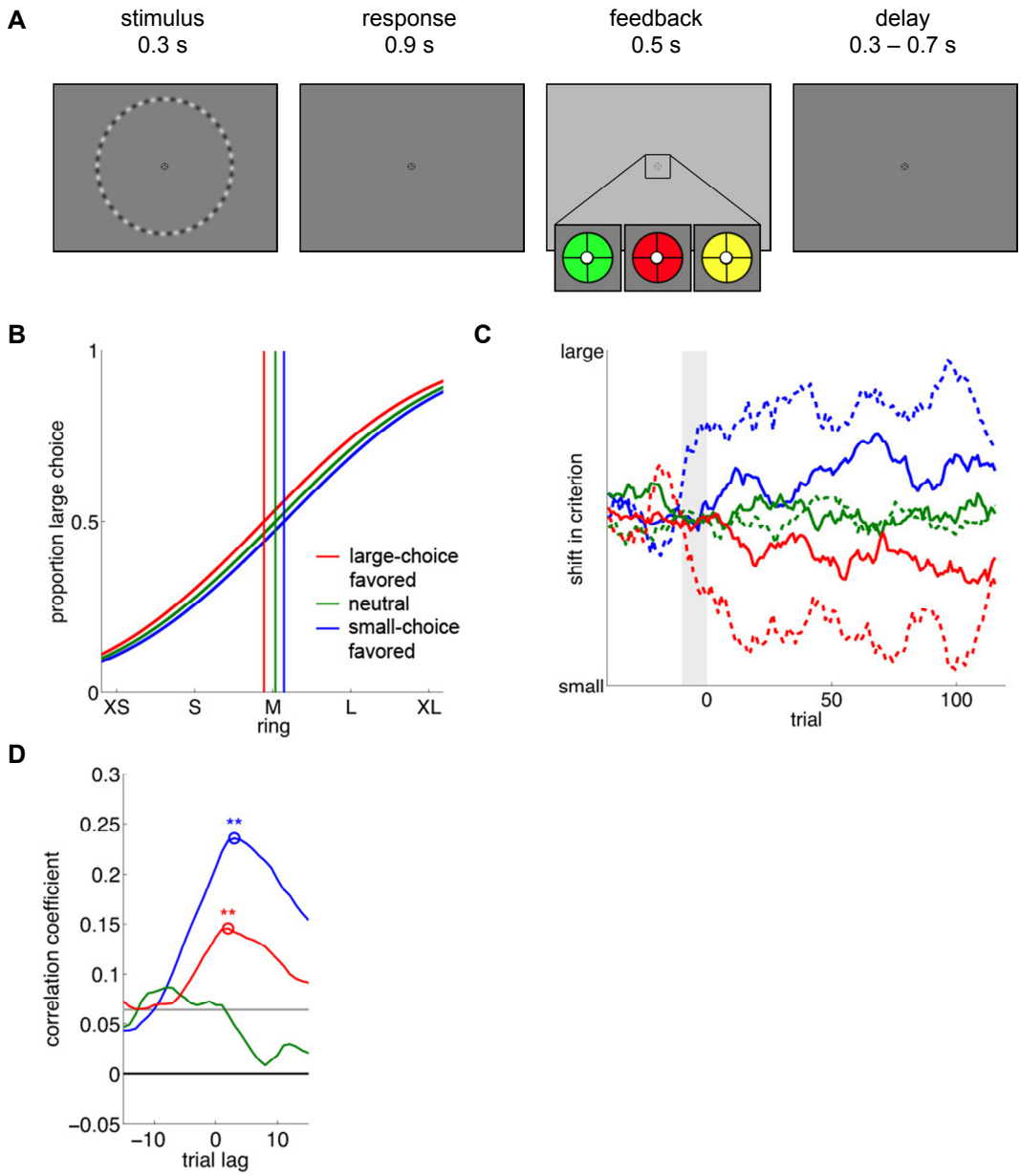


Figure 1. Experimental design and shifts in criteria. (A) Sketch of the ring-size discrimination task. Perceptual decisions and motor responses were dissociated by using randomly assigned buttons for small and large responses for each subject. Feedback was provided by changing the color of the fixation for 500 ms to green for correct choices, red for incorrect choices, and yellow for late choices or no choices made. Late trials were processed incorrect. (B) Psychometric curve fit across subjects for 3 experimental conditions. Percent large choices are labeled for each condition: blue for small choice favored condition, green for no choice bias favored condition, and red for

large choice favored condition. Vertical lines indicate criteria, or its 50th percentile. (C) Behavioral shifts in criteria for 3 conditions. Criteria shifts in psychometric functions across trials are shown with ideal (dashed) and observed (solid) lines for each condition. Gray highlight indicates trial intervals where two bias conditions (blue and red) were introduced. (D) Averaged cross-correlation from individual subjects. 2 biased conditions show significant positive correlation with trial lag between 3 and 5 trials, above chance level indicated by gray line. Neutral condition did not show any significant correlation.

Optimality indexing. To quantitatively demonstrate the undershooting, overshooting, and optimal behaviors, we developed the optimality index. We computed the local optimal shift value and computed its ratios for observed behavioral shifts for individual subjects. Initial and alternative method was to assume uniform bias feedback effect throughout each run, and assume global performance criteria to compute subjects' criteria placement. But, given the experiment design and observed dynamic shifts in criteria we wished to minimize under-estimation of behavioral shifts observed which may be due to the lag, or time, required for subjects to successfully chase the optimal criteria placements, especially when conditions were introduced. Similar to the bin-sized and sliding-window method for μ estimates, we computed the μ of preset answers from stochastically generated feedback schedules. The bin-size was 15-trials and each bin-window advanced incrementally by 5 trials. For each stimulus type, the number of large ring responses were gathered and psychometrically fit for optimal μ. These local optimal μ's were compared with observed μ's to determine local undershooting or overshooting behaviors. Undershooting occurred when observed μ fell short of its local optimal μ; overshooting occurred when observed μ exceeded its local optimal μ. We then normalized the index by computing the ratio of observed shift in criteria to optimal placement in criteria.

Reinforcement learning model. To account for the changes in criterion we utilized a modified reinforcement learning model for perceptual decision-making (Figure 2). Similar models have recently been applied to human behavioral, and monkey behavioral and electrophysiological data (Kahnt et al, 2011; Law and Gold, 2009). In each trial the

model makes a perceptual choice p(large). P(large) is computed based on the decision variable, analogous to stimulus:

$$p(large)_t = \lambda + (1 - \lambda) * erfc(-\frac{x_t - c_t}{2\beta})$$

where c is a bias term accounting for the placement in criterion. β is the slope of the complementary error function accounting for individual levels of sensitivity. λ notes the lapse rate. The EV, or expected value, is derived from the complementary error function with the exception of the stimulus distance from criterion in absolute. This determines the probability of expected reward:

$$EV_t = \lambda + (1 - \lambda) * erfc(-\frac{|x_t - c_t|}{2\beta})$$

Upon feedback, r_t, error is calculated, δ_t:

$$\delta t = r_t - EV_t$$

Error is then used to update the criterion:

$$c_{t+1} = c_t + \alpha * \delta_t$$

α is the shift rate, which accounts for the magnitude of shift in criteria. For each participant the free model parameters (α, β, and c_1) were estimated by fitting the vector of trial-by-trial model predictions p(large) against a vector participants' actual trial-by-trial perceptual choices (coded as small = 0 and large = 1). The data was fit for each run per individual subject, assuming one set of parameters global across all trials of runs in that session. The model decision, p(large), was fit against the observed behavior responses on trial-to-trial basis. The differences between the model behavior and the observed behavior were minimized and the best fit was used.

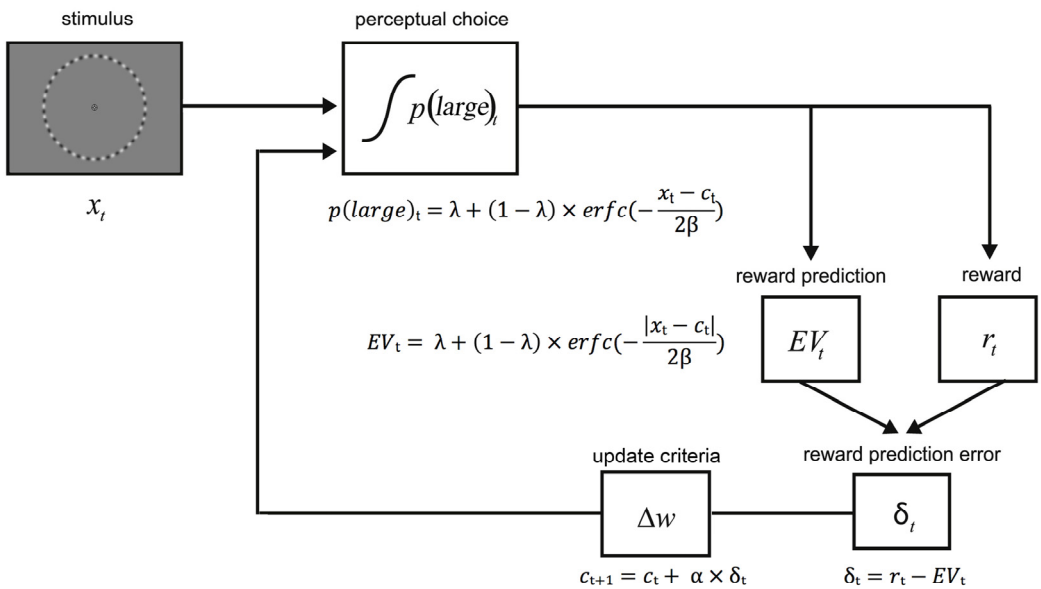


Figure 2. Reinforcement-learning model for dynamic criterion shift in perceptual decision-making. Perceptual decisions are based probabilistically on sensory information based on stimulus. The more positive x_t the more likely is a large-ring decision, and the more negative x_t the more likely is a small-ring decision.

Model and behavioral data comparison. To assess how well the model characterized behavioral data, the model data were derived using the estimated model parameters from each session per subject and compared with behavioral data. We reused the exact sequence of stimuli and feedback schedules implemented in experiment session to assess model's perceptual choices. The performance of the model was computed by using the probability of a correct decision, p(correct).

$$p(correct)_t = p(large)_t * k + (1 - p(large)_t)) * (1 - k)$$

k = 1 if the preset correct answer for the trial was large and k = 0 if the answer was small. The percentage correct was computed for each run and compared with behavioral data.

Additionally, the probability of perceptual choice for large response, p(large),was computed for each stimulus type in a given run and compared with behavioral data.

Sensitivity (σ) derived from behavioral data for individual subjects was compared with one of the model's parameters, β. To derive σ, we fit cumulative Gaussian function for all neutral feedback schedule conditions across all sessions for each subject.

Optimality regime analysis. To demonstrate optimality in relation to subjects' criteria shift rate, α, and sensitivity, β, the parameter regime was plotted using the weighted average of individual subject's estimated model parameters for each session. First, we computed the goodness of fit for each session set of parameters (G_s) by measuring its sum of squared errors (SSE):

$$G_s = \frac{(SSE_m - SSE_0)}{SSE_m}$$

SSE_m is the maximum possible SSE and SSE_0 is the measured sum of squared errors. Weights were applied for each set of parameters. W_s indicates the weight for each session:

$$W_s = \frac{G_s}{\sum G_s}$$

Alpha estimates for each subject were then multiplied by its respective weights. The beta parameter was weighted, also:

$$\alpha_{subject} = \sum W_s * \alpha_s, \qquad \beta_{subject} = \sum W_s * \beta_s$$

Once subjects had been indexed, we divided them categorically into overshootersand undershooters. Given that the majority of subjects were undershooting (n=28), the undershooting group was further divided into three groups. The undershooting subjects' index values were evenly divided among its range to avoid uneven representation of index values. The parameters were averaged within groups and were plotted to demonstrate the regime transition from undershooting to overshooting behaviors.

Effects of learning rate and sensitivity on index. We assessed the relationship between α and β with optimality to see if there was any correlation between learning rate, sensitivity, and optimality. We used the weighted α and β parameters.

Simulation analysis. Using the reinforcement-learning model, we simulated experiment runs with sets of fixed parameters of α and β as inputs. We used the exact set of stimuli and feedback schedules from experiment sessions; simulation runs randomly selected a set of stimuli sequence and its corresponding feedback schedule. α and β parameters were set in incremental increases within a set range, and all possible combinations within the range were simulated. Each simulation consisted of one thousand iterations, and upon completion we computed the μ using the same bin-size and sliding-window values from criterion shift analysis. The local optimal μ was measured using the same method when computing optimality index, and the index values were computed subsequently. The values were then transiently color-coded in the same, previously used color palette to demonstrate the transition from undershooting to overshooting groups.

Results

Perceptual learning. Our task was difficult for most subjects. Average performance of subjects was relatively low (mean ± SD, 69.20% ± 4.32%) in session 5 (Supplementary Fig. 2). However, despite its difficulty, subjects showed perceptual learning across sessions ($F(4,895)=15.9$, $p<0.001$), indicated by increase in performance and increase in sensitivity ($sigma^{-1}$). Even in highly uncertain settings, subjects learned to associate between stimuli and correct responses.

Induced criterion shift analysis. Observed shifts in criteria in subjects closely match shifts in criteria for optimal performance, though suboptimally (Figure 1B). For each condition, subjects shifted criteria in the corresponding direction that would yield more favorable reward outcome, indicating the effect of induced criterion shifts due to the bias of reverse feedbacks given. Subjects performed suboptimally in conditions where bias in feedbacks was present (Figure 1C). In the condition with no bias in feedback present, subjects maintained their criteria; subjects performed most optimally for the neutral condition. Furthermore, to confirm the induced criterion shifts, we computed cross-correlation between the shifts in bias for optimal performance set by each subject's schedule with shifts in bias in behavior (Figure 1D). We found correlations in both of the two biased conditions. In the small-choice favored condition (blue curve in Figure 1D), subjects' shift in criteria closely matched the schedule given ($r=0.24$, $p<0.01$), highest at lag of 3 trials. In large-choice favored condition, subjects' shift in criteria matched the schedule also ($r=0.15$, $p<0.01$), highest at lag of 2 trials. The small sizes of these trial

lags indicate that subjects were quick to follow the feedback schedules. Additionally, cross-correlation for averaged data was consistent with averaged individual cross-correlation result (Supplementary Fig. 1). Small-choice favored condition cross-correlation ($r=0.75$, $p<0.001$) and large-choice favored condition cross-correlation ($r=0.71$,$p<0.001$) both confirm shift in criterion due to feedback schedules.

Optimality indexing. Majority of subjects were classified undershooters (n=28), 3 of whom were further classified as anti-shooters, described as shifts in criteria opposite to the direction of favorable choices (Figure 4B). Additionally, we found a significant relationship between optimality and sensitivity ($r=0.48$, $p<0.01$). This indicates that optimal behaviors are correlated with sensitivity, and subjects with higher sensitivity are likely to perform more optimally.

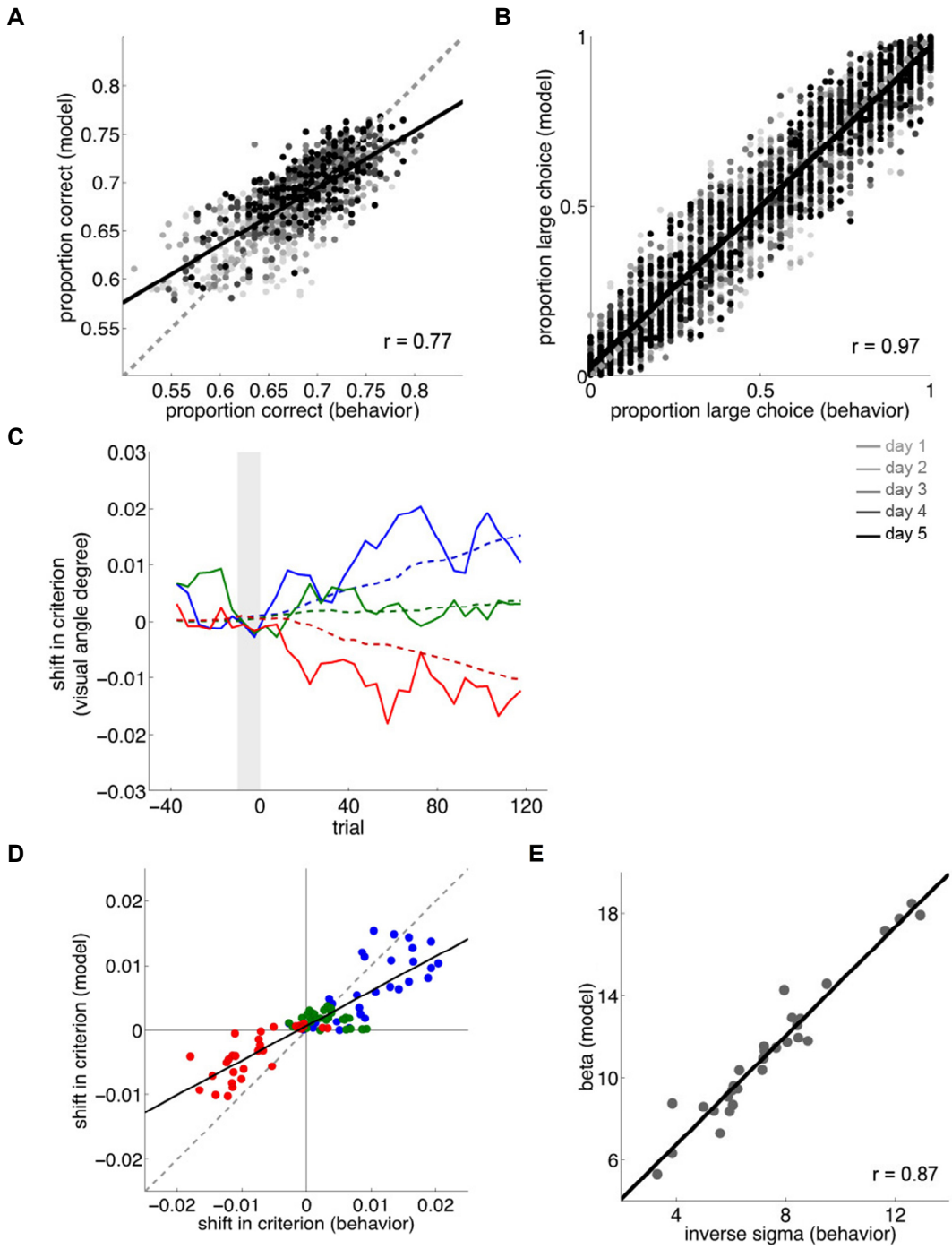


Figure 3. Comparison of the observed behavior and model. (A) Run-wise performance (percentage correct) for individual subjects' observed behaviors and model behaviors. Dashed gray line indicates identity line. (B) Stimulus-specific probability of large-ring

choice (p(large)) for individual subjects' observed behaviors and model behaviors. (C) Averaged subjects' shift in criteria for observed behaviors (solid) and model behaviors (dashed) across conditions. (D) Comparison between shifts in criteria for observed behaviors and model behaviors. (E) Sensitivity comparison between observed behaviors (sigma from psychometric fit curve) and model behaviors (β, model parameter).

Model fit analysis. The model predictions closely matched the behavioral data. Subjects' performance across runs was captured by the model ($r=0.77, p<0.001$; Figure 3A). Percentage for large choices made, (p(large)), from both model and observed data are also closely matched across training days and stimuli ($r=0.97$, $p<0.001$; Figure 3B). Behavioral sensitivity ($sigma^{-1}$) and model sensitivity parameter (beta) are closely correlated, indicating the consistency of model ($r=0.87$, $p<0.001$; Figure 3E).

Analysis on shifts in criteria between model and behavioral data revealed consistent result, indicating the model's behavior correctly reflects the shifts in criteria observed in subjects' behavior (Figure 3C). To compute the degree to which the model effectively captures the behavioral data, we plotted the shifts observed at each point correspondingly. Figure 3D illustrates the model fit to actual data ($r=0.87$, $p<0.001$).

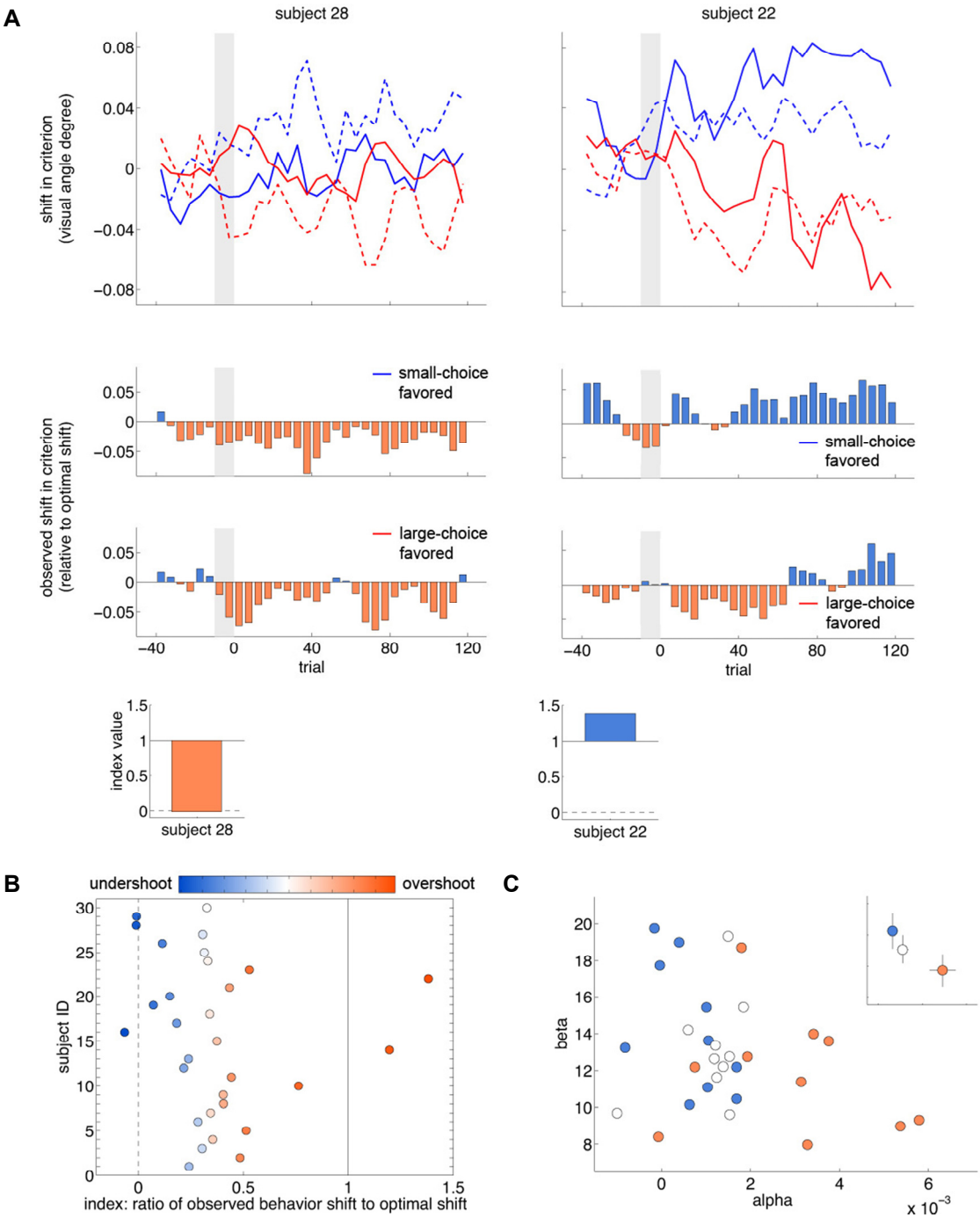


Figure 4. Individual differences in optimality and optimality regime analysis. (A) Optimality indexing method. Example of an undershooter, subject 28, and an example of an overshooter, subject 22, are shown in each column for comparison. (Upper figures) Observed shifts in criteria (solid), and optimal shifts (dashed) for each of 2 biased

conditions averaged across trials (See Fig. 1C). (Middle figures) Bar graphs indicate the direction and amount of shift across trials for each condition. (Lower figures) Sum of shifts across trials for all conditions for individual subjects. Values are normalized to depict ratio of shifts in criteria observed to optimal shifts in criteria, where 1 is optimal shift and 0 is no-shift observed. (B) Index for all subjects. Color map depicts transition from the most undershooter to the most overshooter. Dashed and solid lines depict categorical group differences between undershooters and overshooters. Subjects with negative index values shift in opposite direction (left of dashed line). Subjects within 0 to 1 index values (area between dashed and solid lines) are undershooters. Subjects with higher index values than 1 (right of solid line) are overshooters. (C) Optimality regime map. Scatterplot shows subjects divided into 3 groups in order of index, and model parameter estimates of α (shift rate) and β (sensitivity) for each subject plotted to show clusters of regimes. Upper right plot depicts group averages.

Optimality regime. Figure 4C illustrates the optimality regime given individuals' shift rate, α, and sensitivity, β. Dividing the subjects into 3 groups based on their optimality index, we plotted the group means to desmonstrate a systematic continuation diagonally across the map. Additionally we found significant correlation between alpha and beta under one-tailed Pearson test (Figure 5A; $p=-0.32$, $p<0.05$; two-tailed Pearson: $p=-0.31$, $p<0.10$). This pattern is consistent regardless of the bin sizes (Supplementary Fig. 3). Transition from undershooting behavior to overshooting behavior indicates subjects with higher sensitivity tend to shift less. This is consistent with previous research (Gao et al., 2011). On the other hand, each parameter is significantly correlated to optimality index: alpha (Figure 5B and C; Pearson: $r=0.84$, $p<0.001$; Spearman: $rho=0.73$, $p<0.001$), indicating dependence of optimality on criteria shift rate, as well as sensitivity (Pearson: $r=-0.44$, $p<0.05$, Spearman: $rho=-0.37$, $p<0.05$). This pattern is not trivial, and to demonstrate the variability of index on the parameter regime map, the index was randomized and same correlation analysis was applied (Supplementary Fig. 4). We found that the pattern is robust and this pattern continues throughout various bin sizes and

methods. There is a systematic continuation of optimality with significant correlation to parameters α, shift rate, and β, sensitivity.

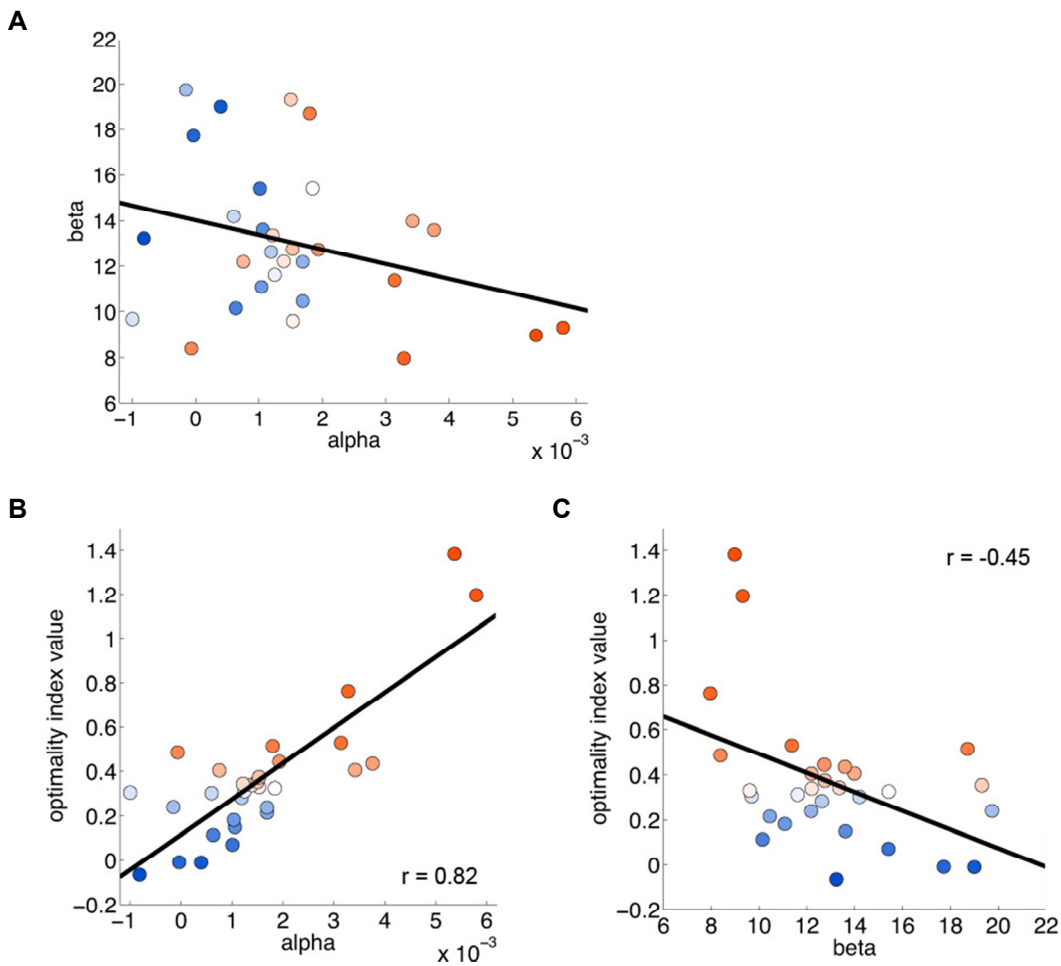


Figure 5. Correlation analysis of model parameters and index. (A) Correlation analysis between model parameters α and β (Pearson: r = -0.31, p = 0.10, One-tail Pearson: r = -0.32, p<0.05). (B) Model parameter α is positively correlated with optimality index (r = 0.82, p < 0.001). (C) Model parameter β is negatively correlated with optimality index (r = -0.45, p < 0.05).

Simulation. The simulation run demonstrates the categorical behavioral groups of undershooters and overshooters, and it demonstrates the continuum of behaviors between two categories (Figure 6). The alpha and beta values used for simulation were in ranges greater than observed from individual model fits (highest alpha value resulted from fit was 0.006; here used was 0.02). The transition is visible from upper-left corner to lower-right corner from extreme undershooters to overshooters. The optimal regime with ideal yield of reward is plotted (dotted black line) across the map to indicate threshold between undershooting and overshooting behaviors. As demonstrated from individual plot (Figure 4C), the color map is consistent with undershooting to overshooting behavior found in subjects.

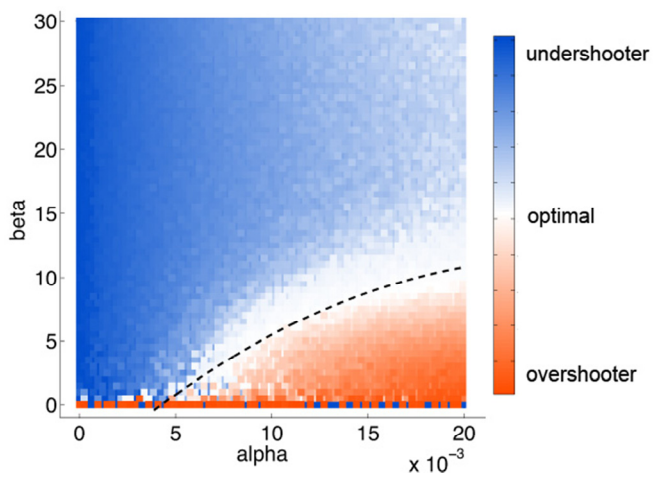


Figure 6. Simulation of reinforcement-learning model. The simulation runs for each corresponding alpha and beta pair note continuum from undershooters to overshooters (colored blue to orange). Dashed black line indicates the approximate region for combinations of alpha and beta in which the highest rewards are possible.

Discussion

Humans have been previously shown to optimally harvest reward in complex perceptual environments where stimuli and values change (Navalpakkam et al., 2009; Diederich and Busemeyer, 2006; Diederich, 2008; Bogacz et al., 2011). In our study, we induced criteria shifts in subjects by adopting stochastically generated reverse feedbacks. Sometimes, incorrect feedbacks were given for responses subjects should have gotten correct. We, then, captured subjects' behavior using the reinforcement-learning model where in which the model estimates reliably corresponded to observed data. Next, we created an optimality index to demonstrate variability in individual subjects' behaviors and we demonstrated optimality as a result of model parameter values, specifically the shift rate and sensitivity. Simulation result provided consistent results with behavioral analysis. We further assessed parameter characteristics by analyzing estimated parameters for each subject to their optimal decision-making index. We found that optimal decision-making is closely correlated with subject's shift rate and sensitivity. We further demonstrated that these relationships were not spurious by conducting the analysis in various bin-sizes with randomized index values. We showed that the correlation increases as bin sizes increased and all of the correlation between the optimality index and each parameter were significant. This indicates that typical undershooters tend to have higher visual sensitivity with lower shift rate values, while typical overshooters tend to have lower visual sensitivity with higher shift rate. Subsequently, this raises the possibility that two parameters reciprocate while contributing to perceptual decision-making and suggests there may be a constraint

exerted for both values. Altogether, with the help of reinforcement-learning model, our findings revealed intriguing relationship between sensory ability and criteria shift in determining individual differences in optimal perceptual decision-making under varying response-payoff contingency.

Subjects in our experiment exhibited perceptual learning but consistently performed suboptimally. This finding is interesting for two reasons. Firstly, it contradicts previous findings of human behavior in optimal decision-making where optimal behaviors in humans have been observed during complex, dynamic tasks (Navalpakkam et al., 2009; Diederich and Busemeyer, 2006; Diederich, 2008; Bogacz et al., 2011). Similarly, animals have often demonstrated optimal performance in dynamic foraging tasks (Sugrue et al., 2004; Teichert and Ferrara, 2010), although differences found between human and animal morphologies engender more complications for precise interpretation. Secondly, majority of subjects were categorized as undershooters (n=28). These may be due to task difficulty but they can also be explained by the slow learning rate in reinforcement learning system. We conjecture our rationale here.

We modified Kahnt et al. (2011)'s model to describe the temporal dynamics of shifts in decision criterion. Their model was based on a model-free reinforcement learning, which means that perceptual choices are made upon computing for expected reward. Depending on a reward outcome in a given trial, a value assigned to a particular stimulus is updated for a following trial. Model-free learning does not inherently account for plasticity in decision-making processes. However, we utilized reward prediction errors to update not the stimulus values but decision criteria. Arguably, this change in

criteria may be closer in reflecting the brain's decision-making plasticity in real life. This process, accounted by model-based learning, is slow and involves neocortical areas (van der Meer et al., 2012), whereas model-free learning system is highly flexible and primarily originates from the striatal areas. When subjects are first introduced to the perceptual learning task, their sensitivity may primarily increase while they are slowly adjusting for shift rate. It may be that this slow adjustment for shift rate is not possible for subjects within 5 training sessions to peak optimality. Hence, subjects are initially prone to undershooting and thus, performing suboptimally and failing for optimal integration of two systems.

Individual differences have been reported frequently in decision-making studies. Previous studies on individual differences in decision-making have noted progressive results in the dorsal anterior cingulate cortex (dACC; Santesso et al., 2008) and basal ganglia (BA; Santesso et al., 2008; Schonberg et al., 2007) activation levels between perceptual learners versus non- learners. Learners in probabilistic reinforcement task were characterized by stronger dACC and BG responses to reward outcomes. This is consistent with findings in which ACC is essential for learning action values (Kennerley et al., 2006; Shima and Tanji, 1998; Holroyd and Coles, 2002; Brown and Braver, 2005). Genetic accounts for individual differences in decision-making also exist (Krugel et al., 2009; Doll et al, 2011). While our study is limited to behavioral analysis and cannot extend the current understanding of the neural mechanisms involved in reinforcement-learning and perceptual decision-making, it raises a few important questions. First, what are the neural mechanisms involved in criteria shifts during perceptual learning? Second,

given how well the model explains shifts in criteria utilizing reward prediction errors, which is known to be a model-free learning component, could criteria shifts be solely explained by model-free system? Third, if criteria shifts are indicative of state changes in the brain, which points to model-based systems, how could we dissociate the two systems and understand the interactions between the two systems?

Recent studies made substantial progresses in identifying model-free versus model-based systems in our brain (Glascher et al., 2010). However, there are much-expected complexities in studying neural mechanisms involved in optimal decision-making. For example, recent studies propose that explicit representation can directly bias an underlying model-free system, thereby biasing action-selection process for optimal decision-making (Doll et al., 2011). In light of challenges, our studies provide a novel approach in experimental design for further pursuing learning systems and mechanisms involved in decision-making. We provide a behavioral framework and demonstrate that criteria shifts can be successfully captured by reinforcing-learning model. We account for the individual differences and note that subject's sensitivity and shift rate for criteria facilitate in optimal decision-making.

References

Aberg KC, & Herzog MH (2012) Different types of feedback change decision criterion and sensitivity differently in perceptual learning. *Journal of Vision*, 12(3): 3, 1-11.

Bayer HM, Lau B, & Glimcher PW (2007) Statistics of midbrain dopamine neuron spike trains in the awake primate. *J Neurophysiol*, 98: 1428-1439.

Bogacz RLT (2011) Integration of Reinforcment Learning and Optimal Decision-Making Theories of the Basal Ganglia. *Neural Computation.*

Brown JW, & Braver TS (2005) Learned predictions of error likelihood in the anterior cingulate cortex. *Science* 307: 1118-1121.

D'Ardenne K, McClure SM, Nystrom LE, & Cohen JD (2008) Bold responses reflecting dopaminergic signals in the human ventral tegmental area. *Science*, 319: 1264-1267.

Daw ND, Niv Y, and Dayan P (2005) Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nat Neurosci*. *8*, 1704-1711.

Daw ND, O'Doherty JP, Dayan P, Seymour B, & Dolan RJ (2006) Corticalsubstrates for exploratory decisions in humans. *Nature*, 441(7095), 876-879.

de Wit S, Corlett PR, Aitken MR, Dickinson A, & Fletcher PC (2009) Differential engagement of the ventromedial prefrontal cortex by goal-directed and habitual behavior toward food pictures in humans. *J Neurosci*, 29:11330-11338.

Delgado MR, Nystrom LE, Fissell C, Noll DC, & Fiez JA (2000) Tracking the hemodynamic responses to reward and punishment in the striatum. *J Neurophysiol*, 84:3072-3077.

Diederich A, & Busemeyer JR (2006) Modeling the effects on response bias in a perceptual discrimination task: Bound-change, drift-rate-change, or two-stage processing hypothesis. *Perception & Psychophysics*, 68, 194-207.

Diederich A (2008) A further test of sequential-sampling models that account for payoff effects on response bias in perceptual decision tasks. *Perception &*

Psychophysics 70(2): 229-256.

Doll BB, Simon DA, & Daw ND (2012) The ubiquity of model-based reinforcement learning. *Curr Opin Neurobiol* 22(6): 1075-1081.

Feng S, Holmes, P, Rorie, A, & Newsome, WT. (2009) Can monkeys choose optimally when faced with noisy stimuli and unequal rewards? *PLoS Computational Biology*, 5(2): e1000284.

Gao J, Tortell R, & McClelland JL (2011) Dynamic integration of reward and stimulus information in perceptual decision-making. *PLoS ONE*, 6(3): e16749.

Glascher J, Daw N, Dayan P, O'Doherty JP (2010) States versus rewards: dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. *Neuron*, 66: 585-595.

Gold JI, & Shadlen MN (2007) The neural basis of decision making. *Annu Rev Neurosci*, 30: 535-574.

Holroyd CB, & Coles MGH (2002) The neural basis of human error processing: Reinforcement learning, dopamine, and the error-related negativity. *Psychol. Rev.* 109, 679-709.

Haruno M, Kawato M (2006) Different neural correlates of reward expectation and reward expectation error in the putamen and caudate nucleus during stimulus-action-reward association learning. *J Neurophysiol*, 95: 948-959.

Jazayeri, M. &Movshon, JA. (2006) Optimal representation of sensory information by neural populations. *Nat Neurosci.* 9(5): 690-6.

Kahnt T, Grueschow M, Speck O, & Haynes J. (2011) Perceptual learning and decision-making in human medial frontal cortex. *Neuron*, 70, 549-559.

Kennerley SW, Walton ME, Behrens TEJ, Buckley MJ, & Rushworth MFS (2006) Optimal decision making and the anterior cingulate cortex. *Nat. Neurosci*. *9*, 940-947.

Knutson B, Taylor J, Kaufman M, Peterson R, Glover G (2005) Distributed neural representation of expected value. *J Neurosci*, 25: 4806-4812.

Krugel LK, Biele G, Mohr PNC, Li SC, & Heekeren HR (2009) Genetic variation in dopaminergic neuromodulation influences the ability to rapidly and flexibly

adapt decisions. *Proc Natl Acad Sci USA*, Oct 20; 106(42): 17951-6.

Law CT, & Gold JI (2009). Reinforcement learning can account for associative and perceptual learning on a visual-decision task. *Nat Neurosci* 12(5): 655-663.

Lee D (2006) Neural basis of quasi-rational decision making. *Curr Opin Neurobiol,* 16:191-198.

Navalpakkam V, Koch C, Rangel A & Perona P. (2010) Optimal reward harvesting in complex perceptual environments. *Proc Natl Acad Sci USA*, 107(11): 5232-5237.

O'Doherty JP, Dayan P, Friston K, Critchley H, and Dolan RJ (2003). Temporal difference models and reward-related learning in the human brain. *Neuron 38*, 329-337.

Rangel, A, Camerer, C, & Montague, PR. (2008) A framework for studying the neurobiology of value-based decision making. *Nat. Rev. Neurosci.* 9, 545-556.

Romo, R, & Salinas, E. (2003). Flutter discrimination: neural codes, perception, memory and decision making. *Nat. Rev. Neurosci.* 4, 203-218.

Santesso DL, Dillon DG, Birk JL, Holmes AJ, Goetz E, Bogdan R, & Pizzagalli DA (2008) Individual differences in reinforcement learning: behavioral, electrophysiological, and neuroimaging correlates. *NeuroImage*, 42(2):807-16.

Schönberg T, Daw ND, Joel D, & O'Doherty JP (2007) Reinforcement learning signals in the human striatum distinguish learners from nonlearners during reward-based decision making. *J Neurosci*, 27(47), 12860-12867.

Shima K, & Tanji J (1998) Role for cingulate motor area cells in voluntary movement selection based on reward. *Science* 282, 1335-1338.

Schultz W, Dayan P. & Montague PR (1997) A neural substrate of prediction and reward. *Science,* 275, 1593-1599.

Stuttgen, MC, Yildiz, A, &Gunturkun, O. (2011) Adaptive criterion setting in perceptual decision making. *Journal of the Experimental Analysis of Behavior*, 96, 155-176.

Sugrue LP, Corrado GS, & Newsome WT (2004) Matching behavior and the representation of value in the parietal cortex. *Science*, 304, 1782-1787.

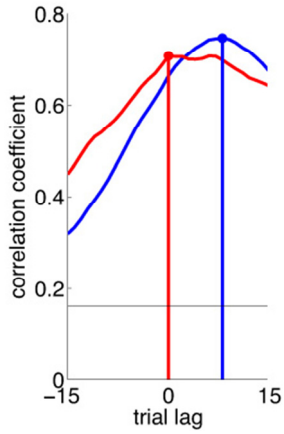
Teichert T, & Ferrara VP (2010) Suboptimal integration of reward magnitude and prior

reward likelihood in categorical decisions by monkeys. *Frontiers in Neursocience*, 4, 1-13.

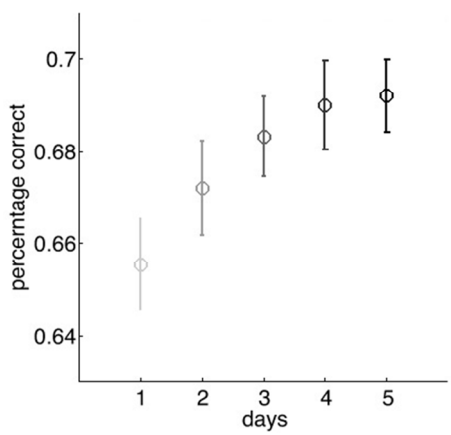
Valentin VV, Dickinson A, & O'Doherty JP (2007) Determining the neural substrates of goal-directed learning in the human brain. *J Neurosci*, 27(15), 4019-4026.

van der Meer MAA, Kurth-Nelson Z, & Redish AD (2012) Information processing in decision-making systems. *The Neuroscientist* 18(4): 342-359.

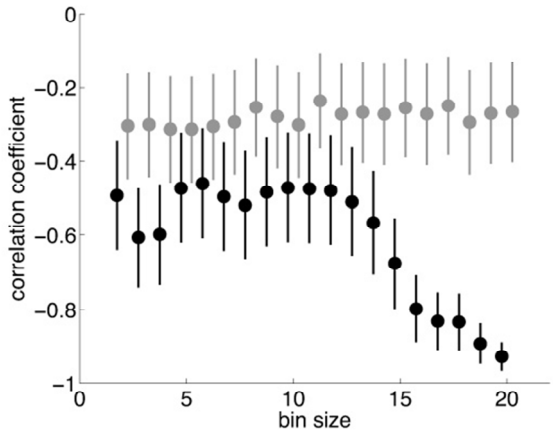
Wozny DR, Beierholm UR, &Shams L (2010) Probability Matching as a Computational Strategy Used in Perception. PLoS Comput Biol 6(8).



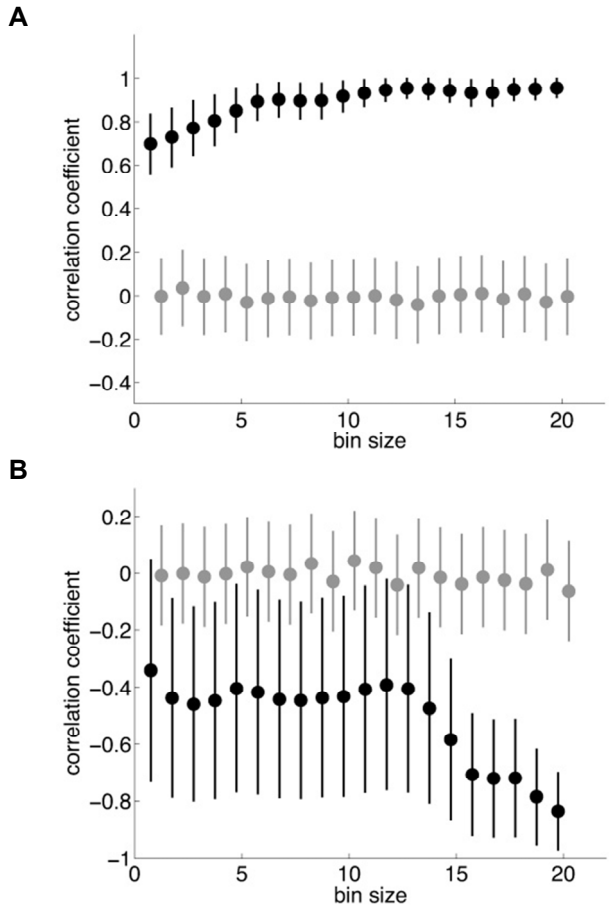
Supplementary Figure 1. Cross-correlation between ideal and observed shifts in criteria. Points at which vertices intersect indicate highest correlation coefficient and its location of trial lag. Gray line indicates 95% confidence bound at chance.



Supplementary Figure 2. Perceptual learning during decision-making. Averaged subject performance indicate perceptual learning across training days ($F(4,895)=15.9$, $p<0.001$). Errorbars indicate standard error.



Supplementary Figure 3. Robustness of optimality index and regime map. Correlation coefficient across different bin size applied to population. ‘Bin size’ indicates number of subjects per bin, lowest at 2 and highest at 20 subjects within a single bin. Black bars indicate negative correlation between model parameters α and β, with 90% confidence intervals. Gray bars indicate randomized index and its correlation with 90% confidence intervals. As bin size grows correlation between α and β become stronger, indicating the effect of optimality index.



Supplementary Figure 4. Model parameters strongly correlate with optimality index. (A) α correlation to optimality index. Identical to method applied in Supplementary Fig. 3B, subjects were binned into groups, lowest at 1 and highest at 20 subjects per bin. (B) β correlation to optimality index. There is a strong correlation between each of model parameters, α and β, to optimality index. All black bars shown are significant (p-value range: < 0.001 ~ < 0.5).

국문 초록

지각적 의사 결정 동안의 최적 기준 변경의 개인차

Issac Rhim (임동균)

뇌인지과학 전공

자연과학대학

서울대학교 대학원

지각적 의사결정은 선택의 결과를 최대화하기 위한 감각 증거가 입력되는 곳에 최적의 기준을 적용하는 것이다. 최적의 기준을 설정하는 것은 감각 정보의 신경 전달에 노이즈가 있거나 지각 선택에 따른 피드백이 시간이 지남에 따라 예측할 수 없는 방식으로 제시될 때 의사결정에서 특히 중요해 진다. 우리는 일정 확률로 실제의 피드백과는 반대인 피드백을 제시하는 지각 실험을 하였다. 예를 들어 피험자가 올바른 지각 판단을 하여도 가끔은 틀렸다는 잘못된 피드백을 받게 하는 것이다. 우리는 강화학습 모델을 이용하여 피험자의 실험 행동패턴을 잡아내고 개인이 설정한 최적의 기준 위치를 찾았다. 모델의 매개 변수들 가운데 기준 위치의 변화속도와 피험자의 민감도가 개인의 최적 기준 위치를 잘 반영하였다. 또한 피험자의 민감도와 변화속도는 최적의 의사결정과

밀접하게 연관되어 있었다. 이는 복잡한 환경에서 최적의 의사결정을 내리기 위해서 민감도와 변화속도, 이 두 변수의 상호보완적이고 체계적인 균형이 존재함을 보여주는 것이다.

주요어: 개인차, 의사 결정, 기준

학번: 2010-24027