The effect of self-choice on behavioral performance in reward seeking

博士論文

The effect of self-choice on behavioral performance in reward seeking

(報酬探索における行動成績に対する自己選択の効果)

平成 25 年度

筑波大学大学院　人間総合科学研究科　感性認知脳科学専攻

瀬戸川　剛

筑波大学

Doctor Thesis

The effect of self-choice on behavioral performance in reward seeking

Tsuyoshi Setogawa

Graduate school of Comprehensive Human Sciences

University of Tsukuba

University of Tsukuba

March, 2014

Contents

Abstract ····· 4

Introduction ····· 5

A. Estimated reward value underlies decision-making ····· 5

B. Value estimation in reward-seeking behavior ····· 7

C. Self-choice and motivation ····· 10

D. The objective and the outline of my study ····· 11

Materials and Methods ····· 13

Subjects ····· 13

Experimental conditions ····· 13

Task procedures ····· 14

Computer assigned reward schedule task ····· 14

Decision-making reward schedule task ····· 15

Probability matching ····· 16

Data analysis and model fitting ····· 17

Results ····· 21

Discussion ····· 26

Model selection and interpretation of the model fitting ····· 26

Interpretation of my results and comparison with other studies ····· 29

Speculations and future directions ····· 32

References ····· 34

Acknowledgements ····· 46

Figure Legends ····· 47

Table Legends ····· 52

Figures ····· 53

Tables ····· 65

Abstract

When an individual chooses one item from two or more alternatives, they compare the values of the expected outcomes. The outcome value can be determined by the associated reward amount, the probability of reward, and the workload required to earn the reward. Rational choice theory states that choices are made to maximize rewards over time, and that the same outcome values lead to an equal likelihood of choices. However, the theory does not distinguish between conditions with the same reward value, even when acquired under different circumstances, and does not always accurately describe real behavior. Recent psychological studies in humans have revealed that the performance of subjects during cognitive and behavioral tasks improves when they choose the task conditions themselves. Here, I examined whether the same is true with reward-seeking behavior. I hypothesized that self-choice could increase the value of the chosen material, and the increased value would then lead to improved performance. To investigate this, I trained monkeys with two kinds of reward schedule task: a decision-making reward schedule task (RSd) and a computer-assigned reward schedule task with matched probability (RSm). In RSd, the monkeys could choose one of two alternatives associated with a different workload and a different reward amount. In contrast, in RSm, a computer assigned reward schedules randomly. Task performances (error rate and reaction time) improved significantly in RSd compared to RSm. Theoretical analysis using a modified temporal-difference learning model showed an enhanced schedule state value in RSd. These results suggest that an increased reward value underlies the improved performances by self-choice during reward-seeking behavior.

Introduction

A. Estimated reward value underlies decision-making

To survive a battle of existence, an organism needs to estimate the values of outcomes, compare them, and then choose the one of alternatives by which the gain can be maximized. Thus, it has been generally acknowledged that the animals and the human beings are endowed with the ability to estimate optimal values and to make rational decisions for their survivals.

When an individual chooses one item from two or more alternatives, they make a choice based on their preference. This suggests that the item selected is the one with the greatest value to the individual, regardless of its "objective value". This is referred to as the "subjective value". Many researchers report that the subjective values are determined by the future reward amount, the probability of reward, the delay to receiving the reward, and the workload required to earn the reward. Indeed, the animals consistently preferred larger rewards to smaller ones (Boysen et al., 2001; Watanabe et al., 2001). Though the animals sometimes preferred immediate smaller rewards to later larger rewards (Ainslie, 1974; Kalenscher and Pennartz, 2008; Richards et al., 1997; Rodriguez and Logue, 1988), a theoretical model of the temporal discounting of reward, which will be described in detail in the next chapter, could explain this. In the studies based on the framework of operant conditioning, the monkeys also tended to choose the alternative associated with smaller workload rather than the one with larger workload (Kennerley and Walton, 2011; Hosokawa et al., 2013). Furthermore the animals preferred the alternatives associated with the higher probability of reward than with the lower one (Samejima et al., 2005).

In recent studies using multi-trial reward schedule tasks, my colleagues showed that the monkey's behavioral performance was influenced by the workload necessary to earn reward (Mizuhiki et al., 2012; Ravel and Richmond, 2006; Shidara and Richmond, 2002; Simmons et al., 2007). The task required the monkeys to complete the repeat of simple visual-discrimination trials. Those studies revealed that the error rate in visual-discriminations decreased as the remaining workload reduced. Again the error rate also decreased as the reward amount increased (Inaba et al., 2013; Toda et al., 2012). In another study using the deferent reward that was dispensed after a visual-discrimination, the animal showed the higher behavioral performances as the delay became shorter (Minamimoto et al., 2009). These results suggest that there is a close relationship between animal's motivation to perform visual-discriminations and expected reward value that is calculated from amount, workload or delay of the reward.

When describing the choices that animals including humans make, a starting point is how closely the behavior matches theory. Rational choice theory, a standard theory in economics that is too simplistic to describe human and animal behavior, postulates that behavior is organized to maximize rewards over time. The primary axioms of rational choice theory are completeness (where potential outcomes are ranked according to their values), transitivity (where the outcome of non-adjacent items in the list of alternatives respects the order of the values) (Kreps, 1990; Allingham, 2002), and independence of irrelevant alternatives (where adding an item to the list of potential alternatives must not interfere with the order of the items already in the list) (Ray, 1973). A corollary of transitivity is that outcomes of equal value should elicit equal choices.

However, in some previous reports, there is considerable criticism for an economic model '*Homo economicus*', which states that the humans have ability to

estimate unbiased values and to judge rationally (Kickert, 1979; Persky, 1995). A research suggests that people rely on a limited number of heuristic principles to make simple judgment rather than to calculate optimal values in actual environment (Kahneman et al., 1982). They provided the evidences against the optimal and rational computation that are believed as inherent in our value estimation (Cox et al., 1982; Delgado et al., 2008). According to them, the probability weighting, framing, loss aversion and other heuristics underlies the decisions of animals and humans, and these often give rise to suboptimal value estimation and irrational decision-making (Kahneman et al., 1982). However, it is unknown whether these suboptimal value estimations occur in the reward-seeking behavior. In this thesis, I tried to show whether the value estimation in reward-seeking behavior was always optimal. For this purpose, I developed the theoretical model to account for the behavioral performances during the reward-seeking behavior. I explain this point in the next chapter.

B. Value estimation in reward-seeking behavior

In an actual environment, an outcome is often delayed after its associated action is completed. As described in the preceding chapter, the error rates during the multi-trial reward schedule task decreased as the remaining workload to the reward reduced. This can be interpreted generally as follows: the outcome value becomes larger as the delay becomes shorter. Grounded in this interpretation, it could be considered that many decisions are involved in trade-off between the amount of an outcome and the time of acquisition of an outcome. When the two alternatives, which are constructed from different amount of reward and different time to reward, are chosen with equal probability, these value of alternatives could be almost same. The temporal discounting

rule (Frederick et al., 2002; Green and Myerson, 2004) can be derived from the experiments using an 'inter-temporal choice' paradigm. This paradigm is defined as a decisions to choose the one of alternatives that associated with the different onsets of reward. Indeed an experiment based on the temporal discounting rule clearly showed that the human subjects chose the option with the maximum of temporally discounted value (Rao, 2010). Other studies using animals showed that the animal's decision-making is also well accounted by this discounting function (Kable and Glimcher, 2010; Kim et al., 2008; Louie and Glimcher, 2010; Minamimoto et al., 2009).

Here I introduce the three mathematical models of temporal discounting, which are widely accepted by researchers. Since the 1930s, the exponential function has been the one of canonical models to explain the discounted value of future rewards (Samuelson, 1937). This can be described as follows:

$$V = \frac{R}{e^{kD}} \qquad (1)$$

where V is a subjective reward value. R is a reward amount. D is a delay to reward. k is a discount factor that determines the rate of discounting the value. It has been believed that this model can secure the subject's preference throughout the time course (Fig. 1A).

The hyperbolic function has been another major function to describe the discounting. Several researchers suggested that the hyperbolic function could explain the actual inter-temporal choice more correctly (Bickel and Marsch, 2001; Mazur, 1987; Schultz, 2010). This function can be described as follows:

$$V = \frac{R}{1+kD} \qquad (2)$$

where V is the subjective reward value. R is the reward amount. D is the delay to reward. k is the discount factor. Interestingly, this model predicts the "preference reversals". Let

us assume that there are two alternatives associated with different amount of reward with different delay. Let us further assume that the value of the larger reward with the larger delay is larger than that of the smaller reward with the smaller delay. The notion "impulsive preference reversals" refers that the order of the values in those outcomes often reverses if the equal length of time is added to the delay until both outcomes (Fig. 1B). For example, the one prefers the $100 right now to the $120 in one month though the one prefers the $120 in 12+1 months to the $100 in 12 months under the impulsive preference reversals. Note the differences in delay between two alternatives are equal in either combination.

Finally, a model I introduce here is the temporal-difference (TD) learning model. This model is implemented in a reinforcement learning procedure by which the decision-maker maximizes the long-term returns (Potjans et al., 2011; Samejima et al., 2003; Sutton and Barto, 1998). In the reinforcement learning, an agent continuously interacts with an environment by receiving the information about the current state S_t and feedback reward r_t for the previous action at S_{t-1} (how good or bad it was) and by choosing a next action. These interactions occur in a sequence of every discrete state at time step t (Fig 1C). The agent makes a useful update using the observed information of reward r_{t+1} and estimates environmental state S_{t+1}. The goal of this learning is to generate the optimal actions leading to maximal reward. The TD learning rule can be generalized as follows:

$$V(S_t) \leftarrow V(S_t)+\alpha[r_{t+1}+\gamma V(S_{t+1})- V(S_t)] \qquad (3)$$

where $V(S_t)$ is a value of environmental state S at time t. α is learning rate. r_{t+1} is the observed reward of S_{t+1}. Parameter γ is a temporal discount rate corresponding to the discount factor, k, in the temporal discounting (γ is equal to the inverse of k). The

learning will be finished when the difference between the value of reward obtained by action $r_{t+1}+\gamma V(S_{t+1})$ and the estimated current environmental state value $V(S_t)$ becomes 0. The TD learning method approximates its current estimate based on previously learned estimates and is utilized for understanding value-based decision-making (Nakahara and Kaveri, 2010).

For better understanding of mathematical substrate for animals' reward-seeking behavior, I propose these models to provide a clue to analyze the behavior. In the next chapter, I will discuss the relationship between an individual's choice and their motivation.

C. Self-choice and motivation

When we face the case that we can take the one from several items, there are two ways of doing it. We choose the item by ourselves, or we let the others choose the items for us. The psychological studies using word memory task showed that the recall performance of participants was better when they chose the items by themselves than when the experimenters assigned items (Perlmuter et al., 1971; Takahashi, 1991). Recently these behavioral changes have been referred to as a 'self-choice effect' in word memory. Consistent with this, another study reported that the participants performing anagram tasks were more efficient in learning when they chose the task items by themselves (Iyengar and Lepper, 1999).

There are two putative mechanisms to account for self-choice effect. 1. Motivational hypothesis: Perlmuter and Monty (1973) proposed that a subject's level of motivation increased when the participant chose the words that should be remembered by their own. They suggest that the increased motivation might elicit the improvement

in task performances. 2. Meta-memory hypothesis: In this hypothesis, Takahashi (1991) suggest that the participants chose items that were easily remembered. This might enhance the memory performances.

Besides the word memory in above, there is possibility that the self-choice enhances the wide range of cognitive and behavioral performances. Zuckerman and his colleagues (1978) reported that the participants showed the higher performances in playing a puzzle game that was chosen by participants. They hypothesized that the increased motivation promoted to achieve higher performances in the game. According to a Self-Determination Theory introduced by Deci and Ryan (1985, 2002), the motivation could be classified into two functional divisions: an intrinsic and an extrinsic motivation (Lepper et al., 1973). The intrinsic motivation was defined as the natural, inherent drive to seek out challenges and new possibilities associated with cognitive and social development. On the other hand, they stated that the extrinsic motivation came from external sources. Zuckerman et al. (1978) suggested that subjects' intrinsic motivation might play a key role in enhancing the puzzle game performances. This was called an 'enhancing effect'.

D. The objective and the outline of my study

These findings raise a possibility that the behavioral performance improvement by self-choice generally take place in reward-based decision-making. However, if, as the economists assume, the subject can rationally estimate the values of items, the performance should not be improved by self-choice since the value of item remains to be insusceptible to whether the items are chosen by oneself or others. To address this, I introduced two kinds of reward schedule task. In the decision-making reward schedule

task (RSd), a subject was able to choose one of two schedules associated with different amounts of reward and workload. In the computer-assigned reward schedule task, the amount of reward and workload were determined by computer. I trained rhesus monkeys to perform these two tasks, and compared their behavioral performances (error rate and reaction time during the visual-discrimination trials). Then I examined whether behavioral performance could be estimated by the theoretical model of the modified reinforcement learning rule.

Materials and Methods

Subjects

Data were obtained from three adult male rhesus monkeys (*Macaca mulatta*; monkey P, ~7.1 kg; monkey H, ~8.4 kg; monkey K, ~6.0 kg). There was an advantage to use the monkeys because they were free from prejudice against / in favor of performing the task. Monkeys P and H were naïve monkeys whereas monkey K was used after other experiment of recording from dorsal raphe nucleus during reward schedules. All monkeys learned the all tasks within 12 months of training. The experiments were approved by the Animal Care and Use Committee of University of Tsukuba, and were all conducted in strict accordance with the guidelines for the Care and Use of Laboratory Animals of University of Tsukuba. The guideline is based on the recommendations of the National Research Council (USA) as published in the ILAR "Guide for the Care and Use of Laboratory Animals", and all research procedures followed the recommendations of the ILAR Guide.

Experimental conditions

Monkeys sat in a primate chair facing a 22-inch cathode-ray tube (CRT) monitor (CV921X; TOTOKU, Japan) placed 1.0 m from their eyes. Three touch-sensitive bars were attached to the front panel of the primate chair at the level of the monkeys' hand. These bars were referred to as the center bar, and right and left choice bars. A water reward was dispensed from a stainless tube that was positioned at the monkey's lips. Experiments were conducted in a sound-isolated dark room, and sound was masked further using white noise. Experimental control and data acquisition were performed

using the real-time experimental system “REX” adapted for the QNX operating system (Hays et al., 1982). Visual stimuli were presented by “Presentation” (Neurobehavioral Systems, Inc., Albany, CA) running on a Windows computer.

Task procedures

Computer assigned reward schedule task

I designed behavioral paradigms and visual stimuli based on previous studies (Shidara and Richmond, 2002; Mizuhiki et al., 2012; Toda et al., 2012; Inaba et al., 2013). These monkeys were initially trained to perform simple visual discrimination trials (Fig. 2A). The monkey had to touch the center bar to initiate each trial. Immediately thereafter, a white rectangle visual cue, which I explain later, was presented at the top of the monitor. Then, 800 ms from the onset of the visual cue, a fixation spot (a small white square, 0.17 × 0.17°) was presented at the center of the monitor. The fixation spot was replaced after 400 ms with a red square (WAIT signal, 0.40 × 0.40°). When the red square was present, the monkey had to keep touching the center bar. After a randomly chosen period (400, 600, 800, 1000, or 1200 ms), the color of the square changed to green (the GO signal). To receive a reward, the monkey had to release the center bar 150–1000 ms after the GO signal. If the monkey released the center bar successfully, the color of the square changed to blue (OK signal), which indicated that the trial had been completed correctly. The visual cue and the square were extinguished after 250–350 ms from the onset of the OK signal, and a liquid reward was delivered. An error occurred when the monkey released the center bar too early (while the square was red or earlier than 150 ms after the appearance of the GO signal), or did not release the center bar within 1 s of the onset of the GO signal. When the monkey made an error, the visual cue and square

were extinguished immediately and the trial was terminated. The inter-trial interval (ITI) was 1 s after a rewarded trial or error.

When the percentage of correct trials for simple visual discriminations exceeded 80%, the computer assigned reward schedule task was introduced (Fig. 2B). In this task, the monkey was required to perform 1, 2, 3, or 4 repeats of a visual discrimination trial (schedule) successfully to earn 1, 2, 3, or 4 drops of liquid reward (0.15, 0.30, 0.45, or 0.60 mL water). During the trials, the visual cue was presented at the top of the monitor. The brightness and length of the visual cue indicated the reward amount and the number of remaining trials, respectively (Fig. 2C). The brightness of the visual cue was proportional to the reward amount: 25% brightness, 1 drop of water; 50% brightness, 2 drops; 75% brightness, 3 drops; and 100% brightness (white, 30.19 lux) 4 drops. The length of the visual cue was extended in proportion to the schedule progress. The schedule states were abbreviated as trial number / schedule length: 1/4, 25% of full length (6.06 × 0.60°); 1/3, 33.3% of full length (8.08 × 0.60°); 1/2 and 2/4, 50% of full length (12.12 × 0.60°); 2/3: 66.7% of full length (16.16 × 0.60°); 3/4, 75% of full length (18.18 × 0.60°); 1/1, 2/2, 3/3 and 4/4, 100% of full length (24.24 × 0.60°). The trials with the longest cues were reward trials, whereas those with shorter cues were no-reward trials. When the monkey made an error, the same schedule state was repeated.

Decision-making reward schedule task

The decision-making reward schedule task (RSd, Fig. 3) consisted of self-choice and reward schedule parts. When the monkey touched the center bar, the self-choice part began. At 500 ms from the onset of the fixation spot (a small white square of 0.17 ×

0.17 deg), the choice targets appeared on either side of the fixation spot for 3 s. The choice targets indicated the alternatives that the monkey could choose. The brightness and length of the choice target were proportional to the reward amount (25% brightness, 1 drop of water; 50% brightness, 2 drops; 75% brightness, 3 drops; and 100% brightness [white] 4 drops) and schedule length: 1 schedule, 25% length (1.50 × 0.60°); 2 schedules, 50% length (3.00 × 0.60°); 3 schedules, 75% length (4.50 × 0.60°); 4 schedules, 100% length (6.00 × 0.60°), respectively (Fig. 3C). Pairs of choice targets were picked randomly from the choice target set. There were ${}_{16}P_2$=240 combinations of left and right choice targets. To make a decision, the monkey had to touch either the right or the left bar that was on the same side as the chosen target 150–3000 ms after the onset of the choice targets. If the monkey kept touching the chosen bar for 500 ms, the unchosen target and the fixation spot were extinguished. The chosen target was also extinguished after an additional 500 ms, and the chosen reward schedule began 1 s after a successful self-choice. If the monkey did not touch either choice bar within 150–3000 ms, touched the choice bar too early (within 150 ms of the onset of the choice targets), or touched the center or unchosen bars within 500 ms of making a choice, the trial was scored as a choice error. After a choice error, the fixation spot and the choice targets were extinguished and the trial was terminated. An ITI of 1 s occurred after a choice error. Then the self-choice part of the trial began again with the same options as the preceding trial.

Probability matching

In RSd, the number of the chosen schedules was biased by the monkeys' decision preference, because they chose the shorter schedules with the larger rewards

preferentially (Fig. 4). The percentage of each schedule in computer assigned reward schedule task was matched with the percentage from the reward schedule parts of RSd. I abbreviated the computer assigned reward schedule task with matched probability as 'RSm'. First, behavioral data were collected from a 5-day RSd experiment. Then the probabilities of each schedule (Table 1) were calculated from these data, and a 5-day RSm was conducted with the same probability as RSd. Three sessions (weeks) of RSd and RSm were conducted in an alternating fashion (total 6 weeks) for monkeys P and H, and two sessions for monkey K.

Data analysis and model fitting

The "R" statistical programming language (R Foundation for Statistical Computing, Team RDC, 2004) was used for statistical analyses. For model fitting, I developed software using C++ in the Integrated Development Environment of Visual Studio 2010 (Microsoft).

To investigate whether behavioral performances improved during the schedules chosen by the monkeys, the error rates and reaction times of the reward schedule part were analyzed in all sessions. The error rate was defined as the ratio of the number of failed trials to the number of whole correct and failed trials in each schedule state. The reaction time was defined as the time to release the center bar after GO signal appeared. The χ^2 test and t-test were used to perform a statistical test for the error rates and reaction times, respectively.

Then model fitting was carried out to estimate the error rates in the reward schedule parts of RSd and RSm using the context-sensitive model described by La Camera and Richmond (2008). The context-sensitive model incorporated the sunk cost

(Sutton, 1991) into a conventional temporal difference learning rule (Sutton and Barto, 1998), formulated as follows:

$$V(\tau,s)= r+\gamma V(\tau+1,s)+\sigma V(\tau-1,s) \quad (4)$$

where $V(\tau,s)$ is the current schedule state value; r denotes the reward amount; τ = 1, 2, … , s denotes the trial number; s = 1, 2, 3, 4 denotes the schedule length.; γ ($0\leq\gamma<1$) is a temporal discount rate; and σ ($0\leq\sigma<1$) quantifies the fraction of the value carried forward to the next trial. When τ = s, the trial terminated. La Camera and Richmond employed this model to describe every state value in 1, 2, and 3 trial schedules with a fixed amount of reward. I expanded this model to estimate the state value for up to 4 trial schedules with 1–4 drops of reward and incorporated the nonlinear effect of reward amount. I called this model an Extended Context-Sensitive model (ECS model). In the ECS model, the current state value $V(\tau,s)$ was expressed as:

$$V(\tau,s) = r^m + \gamma V(\tau+1,s) + \sigma V(\tau-1,s) \quad (5)$$

where the exponent m ($0<m<1$) controls the effect of the nonlinear reward amount because the relationship between reward amount and error rates was nonlinear (Cochran–Armitage test, Table 2). Following this equation, each state value can be written as:

$$\begin{aligned}
&V_{11}=r^m \\
&V_{22}=r^m(1-\gamma\sigma)^{-1},\ V_{12}=\gamma V_{22} \\
&V_{33}=r^m(1-\gamma\sigma)(1-2\gamma\sigma)^{-1},\ V_{23}=\gamma(1-\gamma\sigma)^{-1}V_{33},\ V_{13}=\gamma V_{23} \\
&V_{44}=r^m(1-2\gamma\sigma)(1-3\gamma\sigma+\gamma^2\sigma^2)^{-1},\ V_{34}=\gamma(1-\gamma\sigma)(1-2\gamma\sigma)^{-1}V_{44},\ V_{24}=\gamma(1-\gamma\sigma)^{-1}V_{34},\ V_{14}=\gamma V_{24}
\end{aligned} \quad (6)$$

Finally, the error rate $E(\tau,s)$ in trial τ in schedule s was calculated from the state value $V(\tau,s)$ using a three-parameter logistic function:

$$E(\tau,s) = C + \frac{1-C}{1+e^{(\beta V(\tau,s)-\delta)}} \quad (7)$$

where the parameter C ($0 \leq C < 1$) determines the lower asymptote, the inverse temperature parameter β controls the steepness of the sigmoidal curve, and δ determines the degree of horizontal translation. To find the optimal values of the parameters, all combinations of parameter values that were changed in 1/50 step were searched. Then the square errors between the actual and estimated error rates were calculated to identify the optimal parameter values that minimized the square error. For model fitting, the error rates from the schedule states were excluded when the observations occurred in less than 10 trials, for example, a four-trial schedule with one drop in monkey K. This state was observed only two and nine times in RSd and in RSm, respectively. The schedule state values were calculated by applying the optimal parameter values to equation 6, and then were compared between RSd and RSm.

Also I examined the correlation between the differential values of two choice targets and the time for choosing one of these targets in self-choice part. By this procedure, I tested whether the decisions were based on the differential values of two choice targets (differential values were defined as the difference between first schedule state values calculated by equation 6). The dexterity of hand movement seemed to be different between when choosing right and left targets. Thus all of 240 combinations of choice targets were further divided in the cases of choosing right target and left target.

Then I hypothesized that the monkey might be able to perform the schedules that were longer than 4 trials if the values were increased after self-choice. To test this, I preliminarily introduced longer schedule version of RSd and RSm for one monkey. Both of these tasks contained 1, 2, 4 or 8 trial schedules with 1, 2, 4 or 8 drops of reward.

Finally I removed the trials from the end of each session of RSd and RSm in

order to adjust the numbers of cumulative reward drops in any sessions to be equal. Using these truncated data I compared the error rates between two tasks and fit them by ECS model. As shown later, the number of performed trials was different day by day. Furthermore the monkeys could perform the larger number of trials in RSm than RSd probably because there was no self-choice part in RSm (see results). If the monkeys intended to achieve the lower error rates in the fewer number of trials, they should avoid further loss by making errors. Using the truncated data, I tried to balance the degree of motivation related with the different level of loss aversion among all sessions.

Results

The percentage of chosen schedules in all combinations of choice targets during the self-choice part of RSd (Fig. 4) was analyzed. Monkeys preferentially chose shorter schedules with a larger reward. Even if the ratio of workload to reward amount was the same for two alternatives (e.g., four trials with two drops vs. two trials with one drop), the shorter one was chosen preferentially. These results suggest that the monkeys could discriminate between choice targets and correctly identify the workload and reward amount.

To investigate whether behavioral performances improved when subjects chose the schedules, the error rates and reaction times of the reward schedules were compared between RSd and RSm. In either task, I counted whole number of errors throughout all sessions then calculated the percentage of errors by dividing the number of errors by that of all trials. The percentage of errors in the RSd (Monkey P: 674/12682, 5.54% [total number of error trials / total number of trials started]; Monkey H: 1875/16586, 11.30%; Monkey K: 122/3377, 3.63%) was lower than that in the RSm (Monkey P: 1612/16602, 10.59%; Monkey H, 2975/19785, 15.04%; Monkey K: 310/5888, 5.47%) in all animals (Fig. 5A) (χ^2-test, Monkey P: χ^2=221.43, df=1, p<0.01; Monkey H: χ^2=108.41, df=1, p<0.01; Monkey K: χ^2=14.88, df=1, p<0.01). Furthermore all monkeys showed shorter reaction time in RSd (Monkey P: 554.6 ms, Monkey H: 529.6 ms, Monkey K: 405.6 ms) than in RSm (Monkey P: 581.1 ms, Monkey H: 564.2 ms, Monkey K: 441.3 ms) (Fig. 5B) (t-test, Monkey P: t=16.71, df=24683.09, p<0.01; Monkey H: t=20.90, df=31411.06, p<0.01; Monkey K: t=21.52, df=8267.58, p<0.01). When the error rates and reaction times in each schedule state were examined, most

were smaller and shorter in RSd than in RSm, respectively (Fig. 6). These results are consistent with my hypothesis that improved performance with self-choice also occurs in reward-related behavior.

However, the lack of predictability could have limited the performances in RSm, because the future workload and reward amount in RSd were guaranteed when the monkeys chose schedules but were unpredictable in RSm until the schedules began. To determine whether the improvement of the behavioral performances in RSd were due to control over predictability, I next compared the error rates in the non-first schedule states (2/2, 2/3, 3/3, 2/4, 3/4, 4/4) between RSd and RSm because there was no difference in the predictability in these states, as the visual cue in the first trial had already indicated the remaining workload and reward amount. However, the error rates and reaction times in the non-first schedule states in RSd were significantly lower and shorter than those in RSm, respectively (Fig. 7) (error rate: χ^2-test, Monkey P: χ^2=131.54, df=1, p<0.01; Monkey H: χ^2=18.90, df=1, p<0.01; Monkey K: χ^2=6.38, df=1, p<0.01) (reaction time: t-test, Monkey P: t=12.17, df=12486.14, p<0.01; Monkey H: t=20.83, df=15843.02, p<0.01; Monkey K: t=16.93, df=4055.63, p<0.01), though there was no difference in schedule predictability in the non-first schedule states of both tasks. These results suggest that the better performances in RSd might depend on the context whether the monkeys chose the schedules that should be performed, but not on the future predictability.

To analyze the information processing of reward value, I performed model fitting of the error rates with the ECS model, which incorporates the sunk cost into a computational temporal difference learning rule. Figure 8 shows the error rates

estimated by the ECS model, which are consistent with the actual error rates in RSd and RSm for all three monkeys. The optimal parameters are summarized in Table 3. In each monkey, parameter γ (which controlled the discounting of future outcomes) was larger in RSm than RSd, whereas *m* (which controlled the nonlinear reward amount effect) was smaller in RSm than in RSd. There were no specific trends for σ, which accounted for the sunk cost effect, β, C, and δ, which determined the shape of logistic curve.

Next, I calculated the schedule state values V(τ,s) using these optimal parameters with equation 6 (Fig. 9). The values of the first schedule states (1/1, 1/2, 1/3, and 1/4) were similar between the two tasks, whereas the values of the non-first schedule states exhibited a greater increase in RSd than in RSm as the schedule progressed. These tendencies were consistent among the monkeys though the parameters in logistic functions (β, C and δ) were very different among three monkeys (Table 3). These results suggest that the values of schedules chosen by monkeys were higher than those chosen by computers, which led to better behavioral performances in RSd.

I investigated whether the choice behavior in the self-choice part could be accounted for by the values of two choice targets. In the self-choice part, the monkeys might estimate the values of two choice targets, compare them and choose the one with higher value from two targets. If both of the choice targets had close values, the comparison of these values might be more difficult for monkeys. Consequently, the reaction time, that is defined as the time from the onset of choice targets to touching either left or right bar, might be longer as the difference of the values between two targets become smaller. Here I defined the difference in value as the difference of the values between two targets that were simultaneously presented in the self-choice part. I

found that there were significant inverse correlations between difference in value and the reaction time (Fig. 10) (Monkey P: when left target was chosen; $p<0.001$, $r^2=0.25$, when right target was chosen; $p<0.001$, $r^2=0.33$, Monkey H: when left target was chosen; $p<0.001$, $r^2=0.28$, when right target was chosen; $p<0.001$, $r^2=0.26$, Monkey K: when left target was chosen; $p<0.001$, $r^2=0.18$, when right target was chosen; $p<0.001$, $r^2=0.26$).

Next, I examined the possibility that each monkey might perform schedules that were longer than four trials after self-choice, because most monkeys did not perform RSm schedules with more than four trials (unpublished observation in the past). Therefore, for one monkey, I repeated the task using revised schedules (1, 2, 4, or 8 trials) and rewards (1, 2, 4, or 8 drops). The monkey stopped performing RSm (100% error) immediately and irritably shook the primate chair when given an eight-trial schedule, even with eight drops of reward. In contrast, the monkey performed RSd even when using eight trial schedules with one drop of reward (Fig. 11). This highlights how powerful the effect self-choice may be.

Finally, I compared the error rates between two tasks using the truncated data. Although the monkeys were allowed to perform the task until they stopped each day, the number of completed trials was lower in RSd than in RSm, possibly because of the addition of the self-choice part of the RSd. Because the total reward amount earned in each session was smaller in RSd than in RSm, it is possible that the monkeys in RSm became more satiated during the extra trials, leading to more errors. To assess this possibility, I analyzed the trials from the beginning of each session to the time when the cumulative drops of reward reached a specific number (defined as the smallest number of drops for each monkey in any day: 234 drops for monkey P, 403 drops for monkey H,

and 109 drops for monkey K). Nevertheless, the monkeys made significantly fewer errors during RSd (Fig. 12A) (χ^2-test, Monkey P: χ^2=176.31, df=1, p<0.01; Monkey H: χ^2=28.44, df=1, p<0.01; Monkey K: χ^2=5.50, df=1, p=0.02). Then I collected the error rates in each schedule state and tried to fit them using the ECS model. The results were consistent with those by using the error rates collected from whole trials (Fig. 12B, 12C), except for the data from monkey K because the number of error trials was not enough to perform the model fitting.

Discussion

Although improved behavioral performance with self-choice has been described in psychology studies, it has remained unknown whether self-choice affects reward-seeking behaviors. Here, I used RSd and RSm to investigate this possibility, and found that error rates and reaction times were significantly lower and shorter in RSd than in RSm, respectively. Furthermore, the schedule state values (calculated using the ECS model that incorporated the sunk cost and effect of reward amount into a conventional reinforcement learning rule) were higher in RSd than in RSm. These results suggest that the item values could be influenced by the action carried out by the decision-maker, which leads to better performances. The underlying mechanisms of my findings might be consistent with those described previously in human psychological studies, where cognitive skills such as memory and learning improve for a self-chosen item.

Model selection and interpretation of the model fitting

The value of future reward is inversely correlated with the time taken to achieve it. In standard behavioral models, the widely used functions that account for the temporal discounting of future reward are exponential or hyperbolic functions (Glimcher et al., 2007; Kim et al., 2008; Schweighofer et al., 2006; Schultz, 2010; Louie and Glimcher, 2010). In studies using deferred reward, either function can be used to weigh rewards received at different points in time and applied to fit the behavioral performances of operant actions (Minamimoto et al., 2009). However, it was inappropriate to apply these temporal discounting functions to predict the error rates in my experiment, because my

task consisted of a sequence of discrete trials with varying times required to complete a schedule; the number of errors was different during each schedule, and the time to release the bar varied. The main factors that are likely to have influenced the reward value in my schedule task are remaining workload (Mizuhiki et al., 2012; Shidara and Richmond, 2002), the completed no-reward trials (La Camera and Richmond, 2008), and the reward amount (Toda et al., 2012; Inaba et al., 2013). The completed no-reward trials can be interpreted as an incurred cost that is irretrievable, which is known generally as the sunk cost (Sutton, 1991). To calculate the schedule state value using these parameters, I developed the ECS model by combining the influence of reward amount with a context-sensitive reinforcement model which was used to estimate the error rates in the schedule task by considering the remaining workload and sunk cost. Then I calculated the error rates from the schedule state value using a three-parameter logistic function. This function is often used in studies based on item-response theory (Harris, 1989). An advantage for using this function in my task is to standardize the diversity in behavior sensitivity to reward value, kinetic skill of hand movement and frequency of failure in rational decision. Indeed there were the cases that the error rates in reward trial never reached 0% possibly due to low ability to perform bar-release (see Fig. 6A, 6C, 6E). As for such cases, the behavioral response at the high end of the ability continuum was accounted by the parameter C. Besides, the parameter β and δ determine the steepness and horizontal translation of the curve, respectively. In the item response theory in which the 3 parameter logistic model has been often used, β and δ can approximate the discriminability for response items (Hambleton and Jones, 1993). Based on this, I interpreted that these parameters corresponded to the behavioral sensitivity toward the reward value in my task. Despite the differences in a manner to

transform reward value to behavioral response among individuals, reward value could be well accounted for with error rates by regulating C, β and δ in the 3-parameter logistic function.

The ECS model analysis could provide a clue to examine the underlying mechanisms of my findings. According to the normative view, the value of reward can computationally be determined by associated quantities. At one extreme, we can say that the reward values are computed in the environment rather than in the decision-maker (Sutton and Barto, 1998). However, I found that the value estimation is susceptible to the context as to whether the subjects chose the items. Thus, this study now forges an idea that the choice itself is involved with the estimation of outcome value.

Two major parameters, the discount rate (γ) and the nonlinear effect of reward amount (*m*) showed consistent trends between three monkeys (Table 3). I initially expected that γ would be larger in RSd than in RSm, because a smaller γ discounts the value of future reward more. However, γ in RSd was smaller in all monkeys. The larger discounting is thought to be the substrate of impulsiveness (Onoda et al., 2011); therefore, the monkeys could select an impulsive strategy while performing RSd. In contrast, a higher *m* contributed to higher state values in RSd, as shown in Figure 9. If the self-choice part is considered the sunk cost, it elevates the values of the subsequent reward schedule part of RSd. This leads to larger γ values in the first schedule states of RSd compared to RSm. However, the values of the first schedule states were similar between the two tasks. Therefore, the rate of ramping in the schedule state values along the schedule progress appeared to depend on the difference in context between the self-chosen and enforced schedules. Finally, using the best estimates from reward schedule part, I could successfully predict the behavior in the self-choice part. I found

significant correlation between differential values of two choice targets and time to choose one of them (Fig. 10). This result supports the legitimacy of my model.

Interpretation of my results and comparison with other studies

Previous studies have reported that the animals preferred free-choice to forced-choice even if the alternatives were same in either choice. (Catania, 1975, 1980; Cerutti and Catania, 1997; Suzuki, 1999). The human psychological studies also showed that people preferred the cases in which there were opportunity to make choices and to control their own outcomes (Condry, 1977; Lepper and Malone, 1987). Consistently, it has been shown that the absence of choice produced a variety of detrimental effects on motivation and performance (Deci et al., 1982; Schulz and Hanusa, 1978). The words memory tasks in the earlier studies also showed that the words were better remembered when they were chosen by participants rather than provided by the experimenter (Perlmuter et al., 1971; Monty et al., 1973; Takahashi et al, 1991; Watanabe, 2004). Taken together, I hypothesized that the performances of cognitive and behavioral skills related to self-chosen item might improve. This hypothesis has been supported by the results that the monkeys exhibited better performances in bar-releases during the schedules chosen by the monkeys (Fig. 6). However, there were three other possibilities that give rise to the performance improvement.

Apparently, the monkeys chose shorter schedules with larger reward in RSd. The percentages of such preferred schedules were disproportionally large in RSd (see Fig. 4 and table 1). Consequently, the whole amount of reward in RSd session would become larger than that in the reward schedules with equal probability of each schedule. The larger amount of reward might reinforce the behavior in RSd. I could solve this

discrepancy by introducing RSm in which the probability of every schedule state was matched to that of RSd. Rather, the amount of reward per unit time was smaller in RSd than RSm because of the necessary time for the self-choice part. Nonetheless, the performance in bar-release during RSd yet remained as better.

The schedule predictability might also improve the performances in RSd. The future workload and reward amount in RSd were promised before reward schedule part began. This might allow the monkeys to alter their internal state in order to facilitate correct and fast response in the following schedules. The premovement neural activity might provide the foundation of this (Lebedev et al., 1994). On the other hand the monkeys did not have enough time for motor preparation before the beginning of schedules in RSm. However I could not agree with this hypothesis because the error rates in the non-first schedule states in RSd were better than those in RSm (Fig. 7).

The number of performed reward schedules was larger in RSm than in RSd probably because of the necessary time for the self-choice part (see results). I further hypothesized that the better performance in RSd was achieved because the monkeys intended to avoid further loss by making errors. Thus I analyzed monkeys' performances using truncated data in which the trials were extracted so as to adjust the amount of cumulative reward to be equal. As a result, the error rates in RSd remained better than RSm. Thus I could not agree with this hypothesis (Fig. 12).

Taken together, the self-choice itself seemed to affect the task performances as well as amount (Inaba et al., 2013; Kobayashi et al., 2010), probability (Abler et al., 2006) and delay (Minamimoto et al., 2009) of reward. Furthermore, none of the hypotheses referring the higher frequency of reward dispensing, future predictability or avoiding further loss could explain the improved performance under self-choice.

Rational choice theory has been used to approximate animal and human behavior, and it remains a useful theoretic framework because it describes these behaviors well. However, there are many examples where one or more of the axioms of the theory is broken. Observations by Kahneman and Tversky (1979) have led to a whole school of economic work now called “prospect theory”, which attempts to demonstrate how reward-seeking behavior can be biased by the context in which options are presented. The present study also describes an example that does not fit into rational choice theory, because there were differences in the subjective value between a self-chosen work schedule and a single required work schedule. The increased value occurred even when the monkey had to do a little more work that took more time in RSd than in RSm. Rational choice theory would have predicted that the value of these two conditions should be equal; however, they were not. This suggests that self-choice is one of the heuristics of bounded rationality, where an economic agent restricts rationality but to the limitations of cognitive ability and psychological biases (Simon, 1955). What this implies for human behavior is intriguing. Previous psychological studies on humans have reported that behavioral performances such as memory and learning are improved when a subject chooses an item used in the task compared to when it is assigned (Perlmuter et al., 1971; Takahashi, 1991; Iyengar and Lepper, 1999). My results suggest that these phenomena can also be explained by an increase in the subjective value of items when chosen rather than assigned. Such value enhancement by self-choice might substantially affect our daily decision behavior.

Speculations and future directions

The beneficial aspects of value changing due to the self-choice, one can easily conceptualize the frameworks of system to enhance the performances and outcomes in industrial, educational and economical fields. To obtain excellent results, educators might think that they should not restrain the students from independence actions, and leaders might dare to leave the management of system to the discretion of people. However, the self-choice biased value sometimes prevents us from making optimal decisions. For instance, people often choose lottery with number selection by themselves although its odds of winning might be equal to lottery without number selection. We have to consider that the reliance on the value changing due to the self-choice could sometimes lead to systematic errors.

In the present study, I offered only two alternatives to the monkeys at a time. However, in our daily life, we often face situations in which we have to choose one from three or more alternatives. When there are more alternatives, do the behavioral performances become increasingly better? A modern psychological research shows that people increasingly feel unhappy even if they experience greater material abundance and freedom of choice (Schwartz et al, 2002). One explanation is that the opportunity cost, that is, the potential loss by not making the next-best choices exceeds the benefits from a given choice (Keeney and Raiffa, 1993). This effect was referred to as 'the paradox of choice' (Schwartz, 2004). However the relation between the effect of self-choice and the number of alternatives has not been clarified. Further study is needed to determine how many alternatives are best to maximize reward value.

I have not examined neuronal mechanism underlying the effect of self-choice in this study. However, it might be an important issue which brain regions are related to

this effect. Previous studies showed that choice behavior is regulated by several brain regions such as prefrontal cortex. Some neuroimaging studies pointed out that the fronto-cortico-striatal network might play important roles in calculating temporal discounting (Ballard and Knutson, 2009; Kable and Glimcher, 2007; Tanaka et al., 2004). Single unit recordings in non-human primates also suggested that the neuronal activities in prefrontal cortex are involved in decision-making (Kennerley and Wallis, 2009; Kim et al., 2009; Lee et al., 2007). Neurons in the dorsolateral prefrontal cortex (DLPFC) were reported to be implicated in encoding of information about magnitude and timing of an upcoming reward (Leon and Shadlen, 1999; Tsujimoto and Sawaguchi, 2005). Moreover, Kim et al. (2008) demonstrated that some DLPFC neurons encode the difference in the temporally discounted values of alternative. This finding suggests that the DLPFC neurons might play a key role in inter-temporal choice. On the other hand, lesions of the orbitofrontal cortex (OFC) in monkey impair the ability to modify behavior when the expected outcomes of decisions dynamically changed (Izquierdo et al., 2004). In addition, Padoa-Schioppa and Assad (2008) reported that OFC neurons in monkey encode the subjective reward value. Taken together, these findings suggested important roles of DLPFC and OFC in reward-based decision-making. Thus, increase of reward value in self-choice might possibly be regulated by neural activity of these brain regions. In future studies, I am planning to record single unit activities from DLPFC and OFC to investigate neuronal mechanisms of self-choice effect.

References

Abler B, Walter H, Erk S, Kammerer H, Spitzer M (2006) Prediction error as a linear function of reward probability is coded in human nucleus accumbens. *Neuroimage* **31**: 790-795.

Ainslie GW (1974) Impulse control in pigeons. *Journal of the Experimental Analysis of Behavior*, **21**: 485-489.

Allingham M (2002) Choice Theory: A Very Short Introduction. *Oxford University Press.*

Ballard K, Knutson B (2009) Dissociable neural representations of future reward magnitude and delay during temporal discounting. *NeuroImage*, **45**: 143-150.

Bickel WK, Marsch LA (2001) Toward a behavioral economic understanding of drug dependence: delay discounting processes. *Addiction*, **96**: 73-86.

Bowman EM, Aigner TG, Richmond BJ (1996) Neural signals in the monkey ventral striatum related to motivation for juice and cocaine rewards. *Journal of Neurophysiology,* **75**: 31061-31073.

Boysen ST, Berntson GG, Mukobi KL (2001) Size matters: impact of item size and quantity on array choice by chimpanzees (Pan troglodytes). *Journal of Comparative Psychology*, **115**: 106-110.

Catania AC (1975) Freedom and knowledge: an experimental analysis of preference in pigeons. *Journal of the Experimental Analysis of Behavior*, **24**: 89-106.

Catania AC (1980) Freedom of choice: a behavioral analysis. *The Psychology of Learning and Motivation*, **14**: 97-145.

Cerutti D, Catania AC (1997) Pigeons' preference for free choice: number of keys versus key area. *Journal of the Experimental Analysis of Behavior*, **68**: 349-356.

Condry J (1977) Enemies of exploration: Self-initiated versus other-initiated learning. *Journal of Personality and Social Psychology*, **35**: 459-477.

Cox JC, Roberson B, Smith V (1982) Theory and Behavior of Single Object Auctions. *Research in Experimental Economics, vol. 2, JAI Press*: 1–43.

Croner LJ, Albright TD (1994) Segmentation by color improves performance on a visual motion task. *Investigative Ophthalmology & Visual Science*, **35**: S 1643.

Deci EL, Speigel NH, Ryan RM, Koestner R, Kaufman M (1982) Effects of performance standards on teaching styles: Behavior of controlling teachers. *Journal of Educational Psychology*, **74**: 852-859.

Deci EL, Ryan RM (1985) The general causality orientations scale: Self-determination in personality. *Journal of Research in Personality*, **19**: 109-134.

Deci EL, Ryan RM (2002) Handbook of self-determination research. *University of Rochester Press.*

Delgado MR, Schotter A, Ozbay EY, Phelps EA (2008) Understanding overbidding: using the neural circuitry of reward to design economic auctions. *Science*, **321**: 1849-1852.

Frederick S, Loewenstein G, O'Donoghue T (2002) Time discounting and time preference: A critical review. *Journal of Economic Literature*, **40**: 351-401.

Glimcher PW, Kable J, Louie K (2007) Neuroeconomic studies of impulsivity: Now or just as soon as possible? *American Economic Review*, **97**: 142-147.

Green L, Myerson J (2004) A discounting framework for choice with delayed and probabilistic rewards. *Psychological Bulletin*, **130**: 769-792.

Hambleton RK, Jones RW (1993) Comparison of classical test theory and item response theory and their applications to test development. *ITRMS*, 253-262.

Harris D (1989) Comparison of 1-, 2-, and 3-parameter IRT models. *ITEMS*, **8**: 35-41.

Hays AV, Richmond BJ, Optican LM (1982) Unix-based multiple-process system, for real-time data acquisition and control. *WESCON Conference Proceedings*. 1-10.

Hosokawa T, Kennerley SW, Sloan J, Wallis JD (2013) Single-neuron mechanisms underlying cost-benefit analysis in frontal cortex. *Journal of Neuroscience*, **33**: 17385-17397

Inaba K, Mizuhiki T, Setogawa T, Toda K, Richmond BJ, Shidara M (2013) Neurons in monkey dorsal raphe nucleus code beginning and progress of step-by-step schedule, reward expectation, and amount of reward outcome in the reward schedule task. *Journal of Neuroscience*, **33**: 3477-3491.

Iyengar SS, Lepper MR (1999) Rethinking the value of choice: A cultural perspective on intrinsic motivation. *Journal of Personality and Social Psychology*, **76**: 349-366.

Izquierdo A, Suda RK, Murray EA (2004) Bilateral orbital prefrontal cortex lesions in rhesus monkeys disrupt choices guided by both reward value and reward contingency. *The Journal of Neuroscience*, **24**: 7540-7548.

Kable JW, Glimcher PW (2007) The neural correlates of subjective value during intertemporal choice. *Nature Neuroscience*, **10**: 1625-1633.

Kable JW, Glimcher PW (2010) An "as soon as possible" effect in human intertemporal decision making: behavioral evidence and neural mechanisms. *Journal of Neurophysiology*, **103**: 2513-2531.

Kahneman D, Slovic P, Tversky A (1982) Judgment under uncertainty: Heuristics and biases. *Cambridge University Press.*

Kahneman D, Tversky A (1979) Prospect theory: An analysis of decision under risk. *Econometrica* **47**: 263–291.

Kalenscher T, Pennartz CM (2008) Is a bird in the hand worth two in the future? The neuroeconomics of intertemporal decision-making. *Progress in Neurobiology*, **84**: 284-315.

Keeney RL, Raiffa H (1993) Decisions with multiple objectives: Preferences and value tradeoffs. *Cambridge University Press.*

Kennerley SW, Wallis JD (2009) Evaluating choice by single neurons in the frontal lobe: outcome value encoded across multiple decision variables. *European Journal of Neuroscience*, **29**: 2061-2073.

Kennerley SW, Walton ME (2011) Decision making and reward in frontal cortex: complementary evidence from neurophysiological and neuropsychological studies. *Behavioral Neuroscience*, **125**: 297-317.

Kickert WJM (1979) Fuzzy theories on decision making. *Kluwer Boston Inc.*

Kim S, Hwang J, Lee D (2008) Prefrontal coding of temporally discounted value during intertemporal choice. *Neuron*, **59**: 161-172.

Kim S, Hwang J, Seo H, Lee D (2009) Valuation of uncertain and delayed rewards in primate prefrontal cortex. *Neural Networks*, **22**: 294-304.

Kobayashi S, Pinto de Carvalho O, Schultz W (2010) Adaptation of reward sensitivity in orbitofrontal neurons. *Journal of Neuroscience*, **30**: 534-544.

Kreps DM (1990) A Course in Microeconomic Theory. *Princeton University Press.*

La Camera G, Richmond BJ (2008) Modeling the violation of reward maximization and invariance in reinforcement schedules. *PLoS Computational Biology*, **4**: e1000131.

Lebedev MA, Denton JM, Nelson RJ (1994) Vibration-entrained and premovement activity in monkey primary somatosensory cortex. *Journal of Neurophysiology*, **72**: 1654-1673.

Lee D, Rushworth MF, Walton ME, Watanabe M, Sakagami M (2007) Functional specialization of the primate frontal cortex during decision making. *The Journal of Neuroscience*, **27**: 8170-8173.

Leon MI, Shadlen MN (1999) Effect of expected reward magnitude on the response of neurons in the dorsolateral prefrontal cortex of the macaque. *Neuron*, **24**: 415-425.

Lepper MK, Greene D, Nisbett R (1973) Undermining children's intrinsic interest with extrinsic reward: A test of the "overjustification" hypothesis. *Journal of Personality and Social Psychology*, **28**: 129–137.

Lepper MR, Malone TW (1987) Intrinsic motivation and instructional effectiveness in computer-based education. *Aptitude, learning and instruction: Vol. 3 Conative and affective process analysis*, Erlbaum: 255-286.

Louie K, Glimcher PW (2010) Separating value from choice: Delay discounting activity in the lateral intraparietal area. *Journal of Neuroscience*, **30**: 5498-5507.

Mazur JE (1984) Tests of an equivalence rule for fixed and variable reinforcer delays. *Journal of Experimental Psychology: Animal Behavior Processes*, **10**: 426-436.

Minamimoto T, La Camera G, Richmond BJ (2009) Measuring and modeling the interaction among reward size, delay to reward, and satiation level on motivation in monkeys. *Journal of Neurophysiology*, **101**: 437-447.

Mizuhiki T, Richmond BJ, Shidara M (2012) Encoding of reward expectation by monkey anterior insular neurons. *Journal of Neurophysiology*, **107**: 2996-3007.

Monty RA, Rosenberger MA, Perlmuter LC (1973) Amount of locus of choice as sources of motivation in paired-associate learning. *Journal of Experimental Psychology*, **97**: 16-21.

Nakahara H, Kaveri S (2010) Internal-time temporal difference model for neural value-based decision making. *Neural Computation*, **22**: 3062-3106.

Newsome WT, Paré EB (1988) A selective impairment of motion perception following lesions of the middle temporal visual area (MT). *Journal of Neuroscience*, **8**: 2201-2211.

Onoda K, Okamoto Y, Kunisato Y, Aoyama S, Shishida K, Okada G, Tanaka SC, Schweighofer N, Yamaguchi S, Doya K, Yamawaki S (2011) Inter-individual discount factor differences in reward prediction are topographically associated with caudate activation. *Experimental Brain Research*, **212**: 593-601.

Padoa-Schioppa C, Assad JA (2008) The representation of economic value in the orbitofrontal cortex is invariant for changes of menu. *Nature Neuroscience*, **11**: 95-102.

Perlmuter LC, Monty RA, Kimble GA (1971) Effect of choice on paired-associate learning. *Journal of Experimental Psychology*, **91**: 41-53.

Perlmuter LC, Monty RA (1973) Effects of choice of stimulus on paired-associate learning. *Journal of Experimental Psychology*, **99**: 120-123.

Persky J (1995) The Ethology of Homo Economicus. *Journal of Economic Perspectives,* **9**: 221-231.

Potjans W, Diesmann M, Morrison A (2011) An imperfect dopaminergic error signal can drive temporal-difference learning. *PLoS Computational Biology,* **7**: e1001133.

Rao RP (2010) Decision making under uncertainty: a neural model based on partially observable markov decision processes. *Frontiers in Computational Neuroscience,* **4**: 146.

Ray P (1973) Independence of irrelevant alternatives. *Econometrica,* **41**: 987-991.

Ravel S, Richmond BJ (2006) Dopamine neuronal responses in monkeys performing visually cued reward schedules. *The European Journal of Neuroscience*, **24**: 277-290.

Richards JB, Mitchell SH, de Wit H, Seiden LS (1997) Determination of discount functions in rats with an adjusting-amount procedure. *Journal of the Experimental Analysis of Behavior,* **67**: 353-366.

Rodriguez ML, Logue AW (1988) Adjusting delay to reinforcement: Comparing choice in pigeons and humans. *Journal of Experimental Psychology. Animal Behavior Processes*, **14**: 105-117.

Samejima K, Doya K, Kawato M (2003) Inter-module credit assignment in modular reinforcement learning. *Neural Networks*, **16**: 985-994.

Samejima K, Ueda Y, Doya K, Kimura M (2005) Representation of action-specific reward values in the striatum. *Science*, **310**: 1337-1340.

Samuelson PA (1937) A note on measurement of utility. *The Review of Economic Studies*, **4**: 155-161.

Schultz W (2010) Subjective neuronal coding of reward: temporal value discounting and risk. *The European Journal of Neuroscience*, **31**: 2124-2135.

Schulz R, Hanusa BH (1978) Long-term effects of control and predictability-enhancing interventions: Findings and ethical issues. *Journal of Personality and Social Psychology*, **36**: 1194-1201.

Schwartz B, Ward A, Monterosso J, Lyubomirsky S, White K, Lehman DR (2002) Maximizing versus satisficing: Happiness is a matter of choice. *Journal of Personality and Social Psychology*, **83**: 1178-1197.

Schwartz B (2004) The Paradox of Choice: Why More Is Less. *HarperCollins Publishers Inc.*

Schweighofer N, Shishida K, Han CE, Okamoto Y, Tanaka SC, Yamawaki S, Doya K (2006) Humans can adopt optimal discounting strategy under real-time constraints. *PLoS Computational Biology*, **2**: e152.

Shidara M, Richmond BJ (2002) Anterior cingulated: Single neuronal signals related to degree of reward expectancy. *Science,* **296**: 1709-1711.

Simmons JM, Ravel S, Shidara M, Richmond BJ (2007) A comparison of reward-contingent neuronal activity in money orbitofrontal cortex and ventral striatum: Guiding actions toward rewards. *Annals of New York Academy of Sciences*, **1121**: 376-394.

Simon HA (1955) A Behavioral Model of Rational Choice. *The Quarterly Journal of Economics*, **69**: 99–118.

Sutton RS, Barto AG (1998) Reinforcement Learning. *MIT Press.*

Sutton J (1991) Sunk Costs and Market Structure. *MIT Press*.

Suzuki S (1999) Selection of forced- and free-choice by monkeys (Macaca fascicularis). *Perceptual and Motor Skills*, **88**: 242-250.

Takahashi M (1991) The role of choice in memory as a function of age: Support for a metamemory interpretation of the self-choice effect. *Psychologia*, **34**: 254-258.

Tanaka SC, Doya K, Okada G, Ueda K, Okamoto Y, Yamawaki S (2004) Prediction of immediate and future rewards differentially recruits cortico-basal ganglia loops. *Nature Neuroscience*, **7**: 887-893.

Toda K, Sugase-Miyamoto Y, Mizuhiki T, Inaba K, Richmond BJ, Shidara M (2012) Differential encoding of factors influencing predicted reward value in monkey rostral anterior cingulate cortex. *PLoS One*, **7**: e30190.

Tsujimoto S, Sawaguchi T (2005) Neuronal activity representing temporal prediction of reward in the primate prefrontal cortex. *Journal of Neurophysiology*, **93**: 3687-3692.

Watanabe M, Cromwell HC, Tremblay L, Hollerman JR, Hikosaka K, Schultz W (2001) Behavioral reactions reflecting differential reward expectations in monkeys. *Experimental Brain Research*, **140**: 511-518.

Watanabe T (2004) The self-choice effect from a multiple-cue perspective. *Psychonomic Bulletin & Review*, **11**: 168-172.

Zuckerman M, Porac J, Lathin D, Smith R, Deci EL (1978) On the Importance of Self-Determination for Intrinsically-Motivated Behavior. *Personality and Social Psychology Bulletin*, **4**: 443-446.

Acknowledements

I’m grateful to Dr. M. Shidara, Dr. T. Mizuhiki and members of physiological group. I also thank Dr. Narihisa Matsumoto in National Institute of Advanced Industrial Science and Technology (AIST) and I also received generous support for MRI examination from Dr. Keiji Matsuda and Toshiharu Takasu in AIST.

Figure legends

Figure 1. Value estimation models.

(**A**) Exponential discounting curves from a smaller-sooner (SS) and a larger-later (LL) reward. At every point their heights stay proportional to their subjective values at the time (e.g. T_1, T_2). The heights of the bars represent the actual reward amounts. (**B**) Hyperbolic discounting curves from a smaller-sooner (SS) and a larger-later (LL) reward. The curved lines represent change in subjective value as a function of time (if one were offered a choice between SS and LL rewards at T_1, one would choose LL reward, whereas if one were offered a choice between the same rewards at T_2, one would choose the SS reward). The heights of the bars represent the actual reward amounts. (**C**) The scheme of TD learning. The agent observes an input environmental state S_t and takes an action. Then, it receives a reward feedback r_{t+1} from the environment S_{t+1}. The agent makes a useful update using the observed reward r_{t+1} and estimate environmental state S_{t+1}.

Figure 2. Reward schedule task.

(**A**) Sequential red-green visual discrimination trial. When the monkey touched a center bar, the fixation spot came on in the center of the monitor in front of the monkey. The monkey must release the center bar within 1 sec after red target (WAIT signal) changed to green (GO signal). If the monkey successfully released the bar, the color of the square changed to blue (OK signal) and liquid reward was given. (**B**) Example of reward schedule task (with 4 drops reward). The reward schedule task was composed of 1, 2, 3 and 4 repeats of the visual discriminations to earn 1, 2, 3 or 4 drops of liquid reward

(0.15, 0.30, 0.45 or 0.60 ml of water). Throughout trials, the visual cue was presented at the top of monitor and its brightness and size indicated reward amount and number of remaining trials, respectively (see Fig. 2C). Schedule states were abbreviated as 'trial number / schedule length' (*e.g.* the second trial in 3 trial schedule was labeled as '2/3'). The number of required trials and the amount of reward were randomly picked up by computer. Blue and red arrows indicate the sequences of correct and error responses, respectively. (**C**) Visual cues in 4 trial schedules with different amount of reward. Brightness and length of the visual cue indicate reward amount and proximity, respectively.

Figure 3. Decision-making reward schedule task (RSd).

(**A**) The settings of monkey and experiment apparatus. The monkey was squatted in a primate chair equipped with three touch sensitive bars (right bar, center bar, and left bar). (**B**) The self-choice part in RSd. By touching the center bar, two different choice targets were randomly selected and presented on either side of the fixation spot for 3 sec. The brightness and length of these choice targets were proportional to the reward amount and schedule length, respectively (see Fig. 3C). After the monkey chose one of them, chosen schedule began. (**C**) The choice targets set. The brightness and length of choice target were proportional to the reward amount and schedule length, respectively.

Figure 4. Choice probability for all target combinations.

Choice probabilities in all combinations of two choice targets during RSd in (**A**) monkey P, (**B**) monkey H and (**C**) monkey K. In the labels of the columns (showing the right target) and rows (showing the left target), the reward amount and schedule length

of the choice targets are abbreviated as “drop–trial.” The labels are ordered according to the ratio of reward amount and schedule length. The color density shows the probability of choosing the left target.

Figure 5. Behavioral performances.

(**A**) Gray and black bars show the error rate in the reward schedule part of RSd and RSm, respectively. Error rates were averaged across whole trials (*$p < 0.01$, χ^2 test). (**B**) Gray and black bars show the reaction time in the reward schedule part of RSd and RSm, respectively (*$p < 0.01$, t-test). Error bars indicate SE.

Figure 6. Behavioral performances in every schedule state.

In the panel, the figures in left column show the error rate in every schedule state. Again the figures in right column show the reaction time in every schedule state. (**A**) (**B**) from monkey P, (**C**) (**D**) monkey H and (**E**) (**F**) monkey K. Gray and black bars show the behavioral performances in the reward schedule part of RSd and RSm, respectively. The horizontal axis represents the schedule states and reward amounts. In every schedule state, the error rates and the reaction times associated with 1, 2, 3, and 4 drops of reward are ordered from left to right. The error rates and the reaction times of the four-trial schedule with one drop of reward in monkey K were excluded because of the small number of trials (see Materials and Methods).

Figure 7. Behavioral performances in non-first schedule states.

Gray and black bars show the performances in the reward schedule part of RSd and RSm, respectively. (**A**) Error rate and (**B**) reaction time in all monkeys. Asterisks

indicate significant differences (*$p < 0.01$, χ^2 test). Error bars in (B) indicate SE.

Figure 8. Fitting of error rates.

The actual error rates (gray bars) are overlaid with the fitted error rates using the ECS model (solid lines with open diamonds). In the panel, the numbers in the left column show the results of fitting in RSd. The numbers in the right column show the results of fitting in RSc. (**A**) (**B**) monkey P, (**C**) (**D**) monkey H, and (**E**) (**F**) monkey K. The labels on the horizontal axis are the same as in Fig. 6. The model parameters that most accounted for actual error rates were sought using least-square minimization procedure. The error rates of the four-trial schedule with one drop of reward in monkey K were excluded because of the small number of trials (see Materials and Methods).

Figure 9. Estimated values of every schedule state.

Estimated values of every schedule state in (**A**) monkey P, (**B**) monkey H, and (**C**) monkey K. Red lines with red circles show the values in RSd, whereas blue lines with blue open squares show RSm. The vertical axis describes the value of the estimated schedule state. The labels on the horizontal axis are the same as in Fig. 6.

Figure 10. Relation between difference in values of two targets and the time to choose one of them.

In the panel, the figures in left column show the relation between difference in values of two choice targets and the reaction time to touch the left bar. Again the figures in right column show the relation between difference in values of two choice target and the reaction time to touch the right bar. (**A**) (**B**) monkey P, (**C**) (**D**) monkey H, and (**E**) (**F**)

monkey K. Circles indicate all of 120 combinations of choice targets. Linear regression revealed that there was significant correlation between difference in values of two targets and the time to touch the one of bars (A: $p<0.001$, $r^2=0.25$, B: $p<0.001$, $r^2=0.33$, C: $p<0.001$, $r^2=0.28$, D: $p<0.001$, $r^2=0.26$, E: $p<0.001$, $r^2=0.18$, F: $p<0.001$, $r^2=0.26$).

Figure 11. Error rates in 8 trial schedule versions.

Error rates in eight trial schedule versions of RSd (monkey P). The horizontal axis represents the schedule states and reward amounts. In every schedule state, the error rates associated with 1, 2, 4, and 8 drops of reward are ordered from left to right.

Figure 12. Error rate in the truncated data.

(**A**) Error rates from truncated data. Gray and black bars show the performances in the reward schedule part of RSd and RSm, respectively. Asterisks indicate significant differences (**$p < 0.01$, *$p < 0.05$, χ^2 test). (**B**) Fitted error rates. The actual error rates (gray bars) are overlaid with the predicted error rates by ECS model (solid lines with open diamonds). The labels in horizontal axis are same as in Fig. 6. (**B-1**) RSd in monkey P, (**B-2**) RSm in monkey P (**B-3**) RSd in monkey H, (**B-4**) RSm in monkey H. I could not fit the data from monkey K because the truncated data did not contain enough trials for fitting. (**C**) Estimated schedule state value. (**C-1**) monkey P (RSd: $\gamma = 0.34$, $\sigma = 0.96$, $m = 0.38$; RSm: $\gamma = 0.82$, $\sigma = 0.12$, $m = 0.2$) and (**C-2**) monkey H (RSd: $\gamma = 0.76$, $\sigma = 0$, $m = 0.72$; RSm: $\gamma = 0.90$, $\sigma = 0$, $m = 0.22$). Red lines with red circles indicate the values in RSd, while the blue lines with blue open squares indicate those in RSm. The vertical axis describes the value of the estimated schedule state. The labels on the horizontal axis are the same as in Fig. 6.

Table legends

Table 1. The probabilities of chosen schedules in RSd.

The horizontal row means the reward amount. The vertical column means the number of visual discrimination trials. Each color shows the range of choice probability (blue: 0%-5%, green: 5%-10%, red: 10%-15%).

Table 2. Results of Cochran–Armitage test of error rates for reward amount.

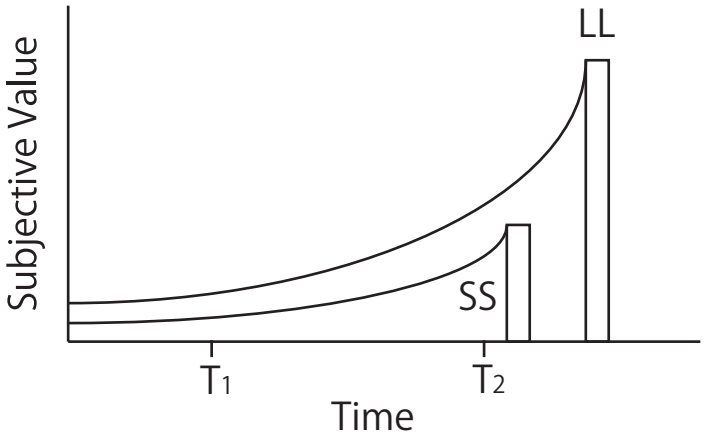
The upper table shows the test results of RSd and the lower table shows the test results of RSm. The horizontal row is the test result for each schedule state in 3 monkeys.

Table 3. The optimal value of all parameters by the ECS model.

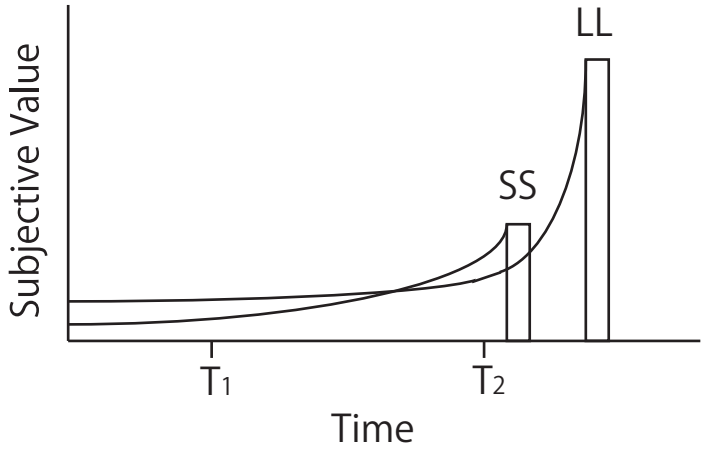
The estimated values in RSd and RSm for 3 monkeys are shown in each column. γ; the rate of discounting the reward, σ; the fraction of value carried forward to the next trial, β; the steepness of the sigmoidal curve, *m*; non-linear effect of reward amount, δ; the degree of horizontal translation, C; the lower asymptote.

Figure 1

A


LL
Subjective Value
SS
T1
T2
Time

B


LL
Subjective Value
SS
T1
T2
Time

C

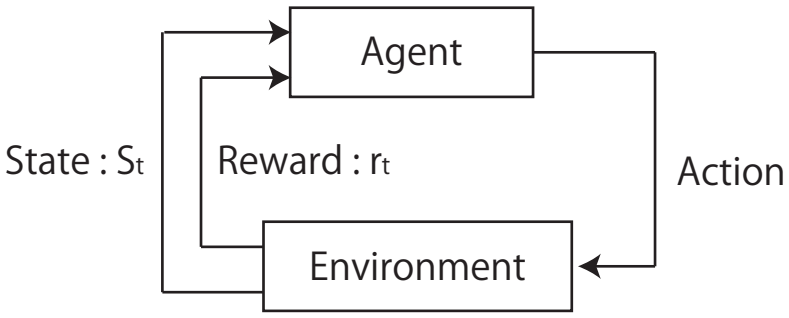
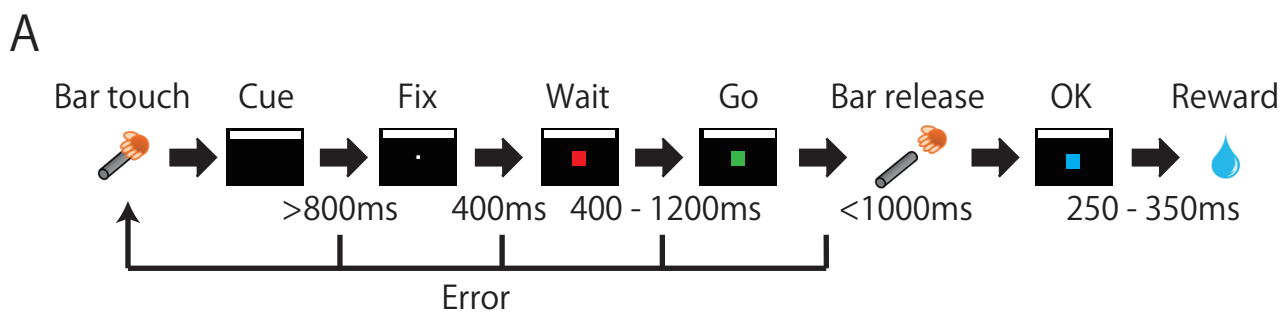
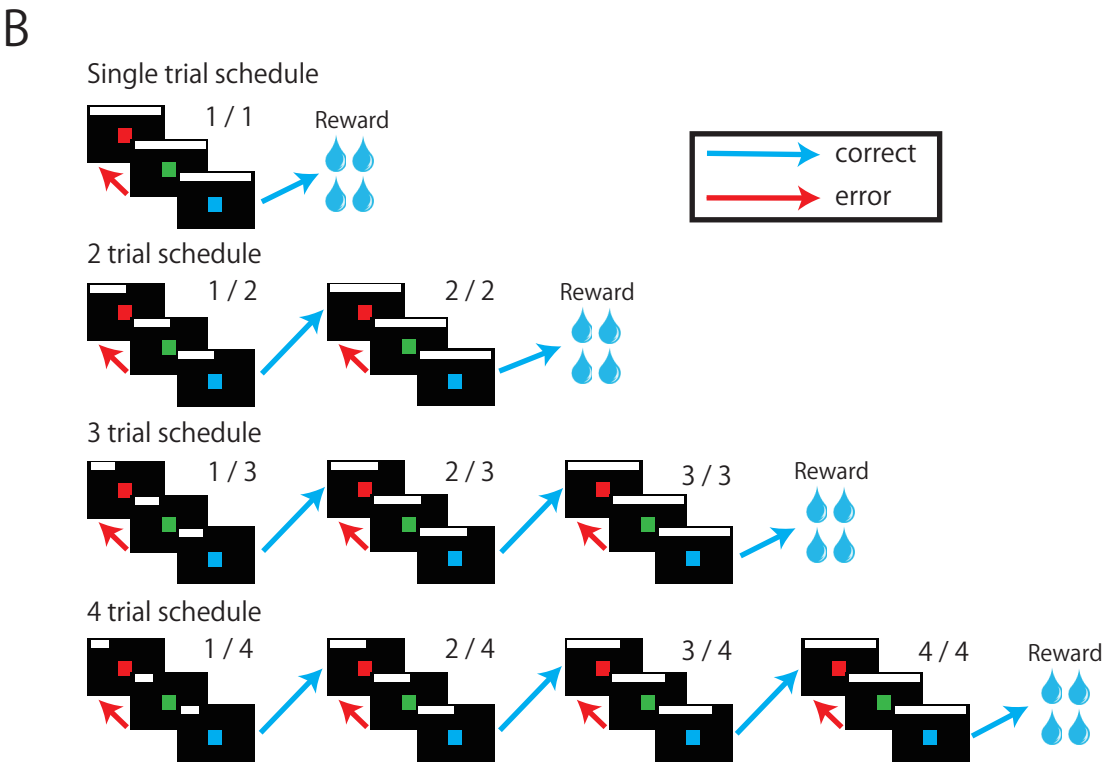

Agent
State : St
Reward : rt
Action
Environment

Figure 2


A
Bar touch
Cue
Fix
Wait
Go
Bar release
OK
Reward
>800ms
400ms
400 - 1200ms
<1000ms
250 - 350ms
Error


B
Single trial schedule
1 / 1
Reward
correct
error
2 trial schedule
1 / 2
2 / 2
Reward
3 trial schedule
1 / 3
2 / 3
3 / 3
Reward
4 trial schedule
1 / 4
2 / 4
3 / 4
4 / 4
Reward

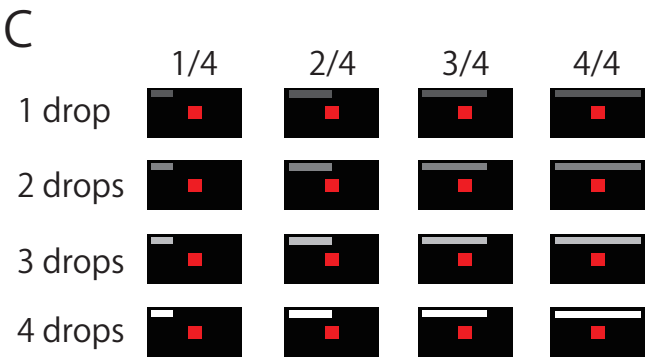

C
1/4
2/4
3/4
4/4
1 drop
2 drops
3 drops
4 drops

Figure 3

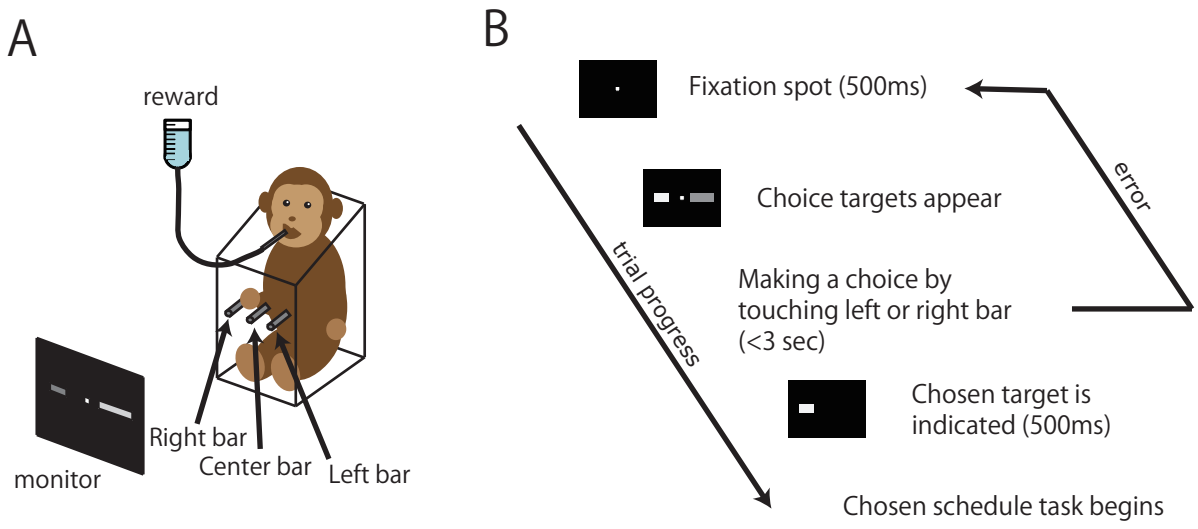

A
reward
monitor
Right bar
Center bar
Left bar
B
Fixation spot (500ms)
Choice targets appear
Making a choice by
touching left or right bar
(<3 sec)
Chosen target is
indicated (500ms)
Chosen schedule task begins
error
trial progress
C
Schedule length (workload)
single trial
2 trials
3 trials
4 trials
Reward
amount
1 drop
2 drops
3 drops
4 drops

Figure 4

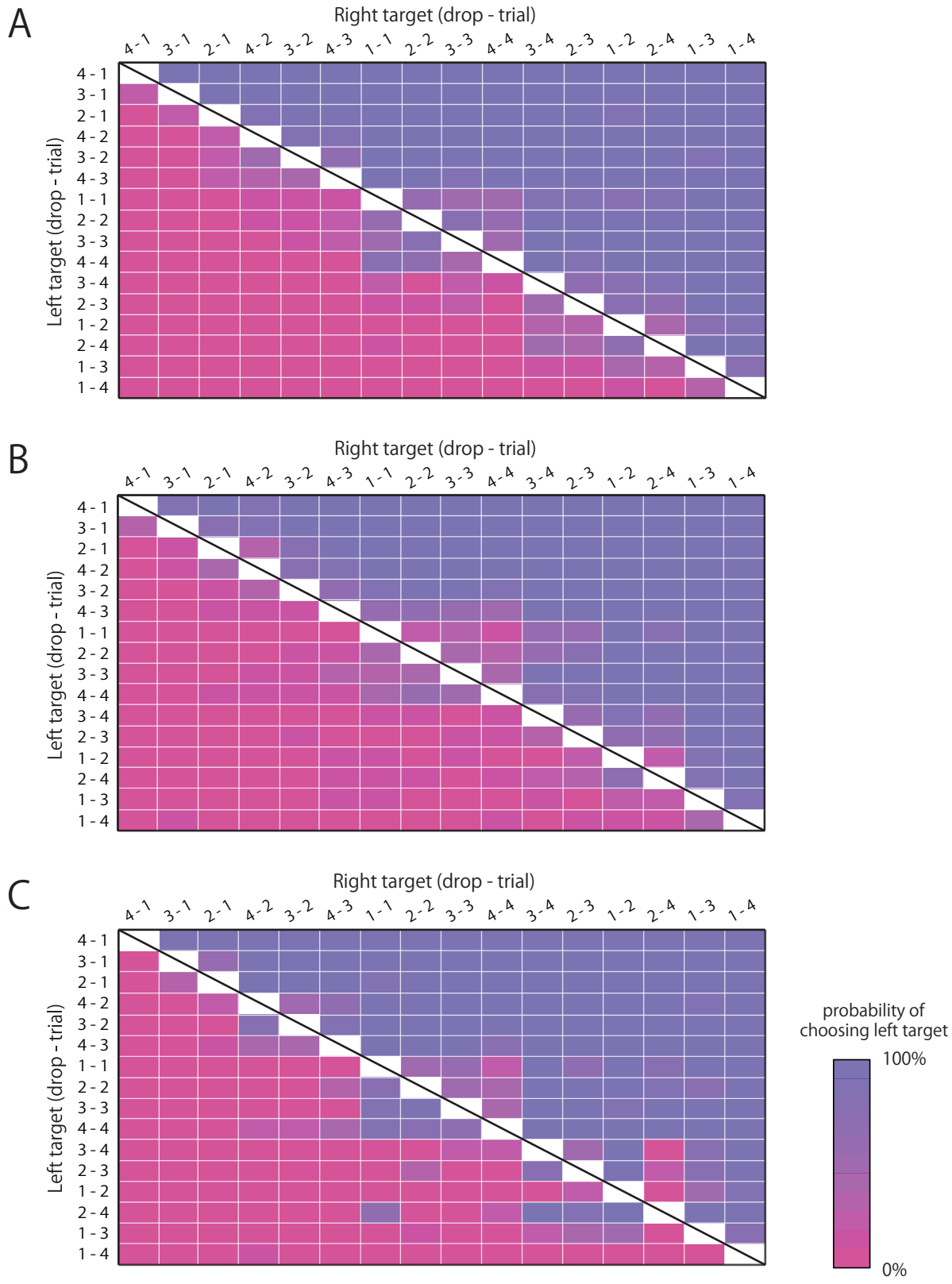

A
Right target (drop - trial)
4-1 3-1 2-1 4-2 3-2 4-3 1-1 2-2 3-3 4-4 3-4 2-3 1-2 2-4 1-3 1-4
Left target (drop - trial)
4 - 1
3 - 1
2 - 1
4 - 2
3 - 2
4 - 3
1 - 1
2 - 2
3 - 3
4 - 4
3 - 4
2 - 3
1 - 2
2 - 4
1 - 3
1 - 4
B
Right target (drop - trial)
4-1 3-1 2-1 4-2 3-2 4-3 1-1 2-2 3-3 4-4 3-4 2-3 1-2 2-4 1-3 1-4
Left target (drop - trial)
4 - 1
3 - 1
2 - 1
4 - 2
3 - 2
4 - 3
1 - 1
2 - 2
3 - 3
4 - 4
3 - 4
2 - 3
1 - 2
2 - 4
1 - 3
1 - 4
C
Right target (drop - trial)
4-1 3-1 2-1 4-2 3-2 4-3 1-1 2-2 3-3 4-4 3-4 2-3 1-2 2-4 1-3 1-4
Left target (drop - trial)
4 - 1
3 - 1
2 - 1
4 - 2
3 - 2
4 - 3
1 - 1
2 - 2
3 - 3
4 - 4
3 - 4
2 - 3
1 - 2
2 - 4
1 - 3
1 - 4
probability of choosing left target
100%
0%

Figure 5

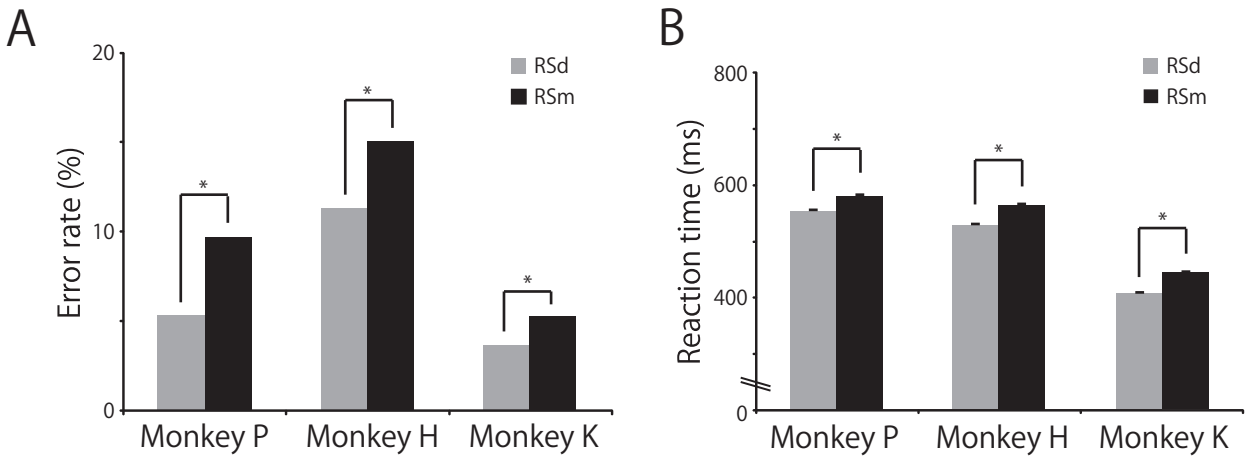

A
20
10
0
Error rate (%)
RSd
RSm
*
*
*
Monkey P
Monkey H
Monkey K
B
800
600
400
0
Reaction time (ms)
RSd
RSm
*
*
*
Monkey P
Monkey H
Monkey K

Figure 6

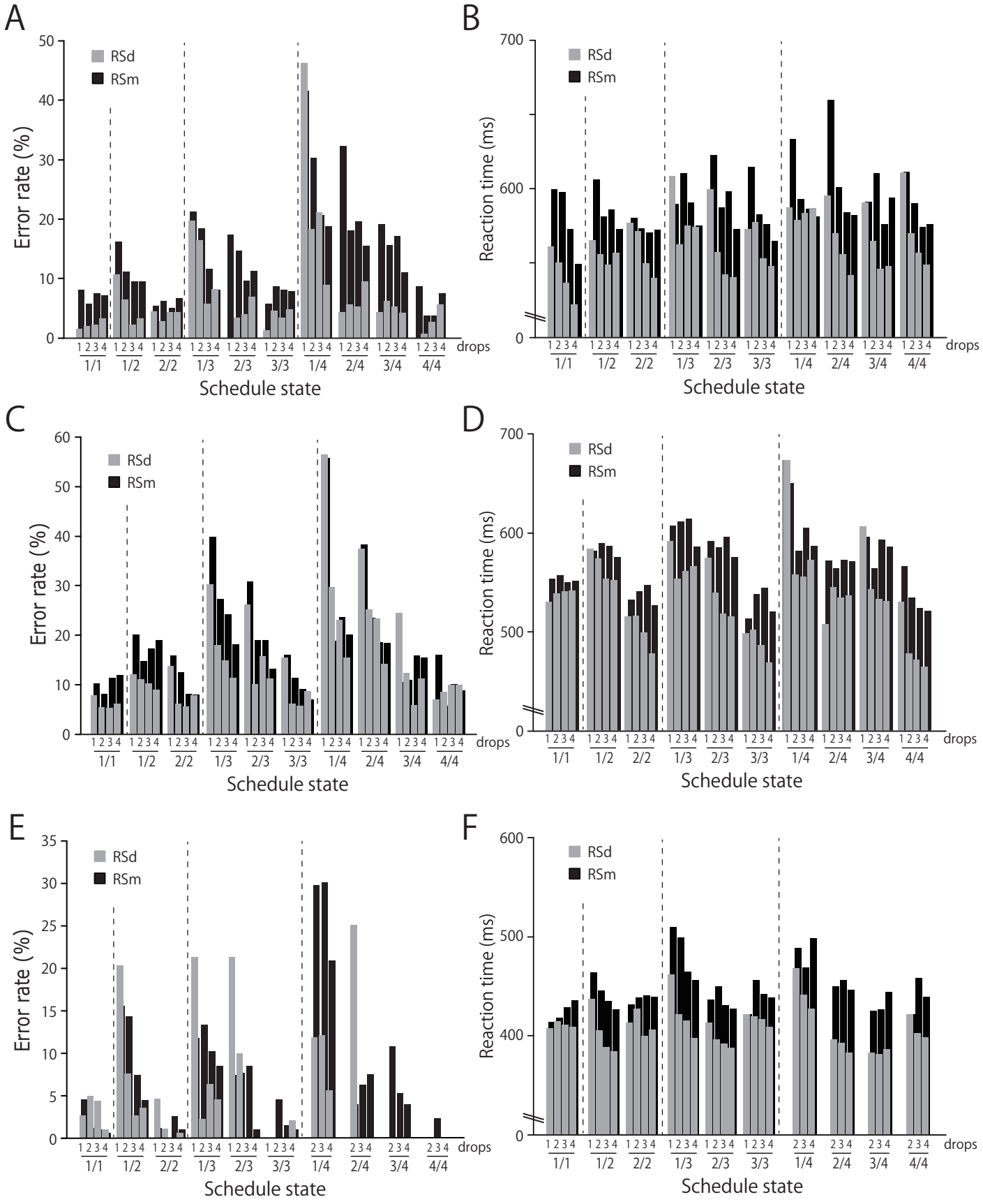

A
B
C
D
E
F
RSd
RSm
Error rate (%)
Reaction time (ms)
drops
1 2 3 4
1/1
1/2
2/2
1/3
2/3
3/3
1/4
2/4
3/4
4/4
Schedule state

Figure 7

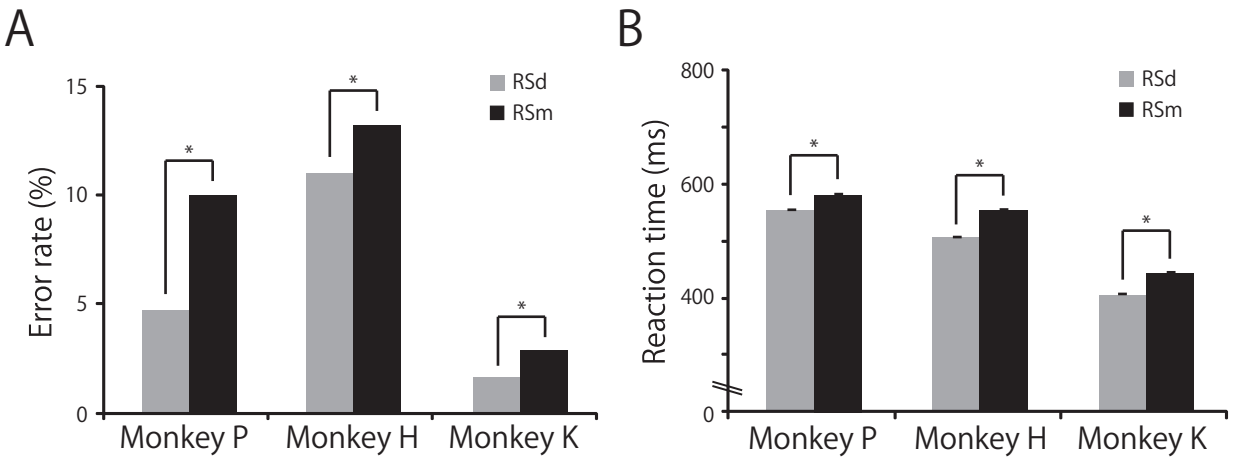

A
B
Error rate (%)
15
10
5
0
RSd
RSm
*
Monkey P
Monkey H
Monkey K
Reaction time (ms)
800
600
400
0

Figure 8

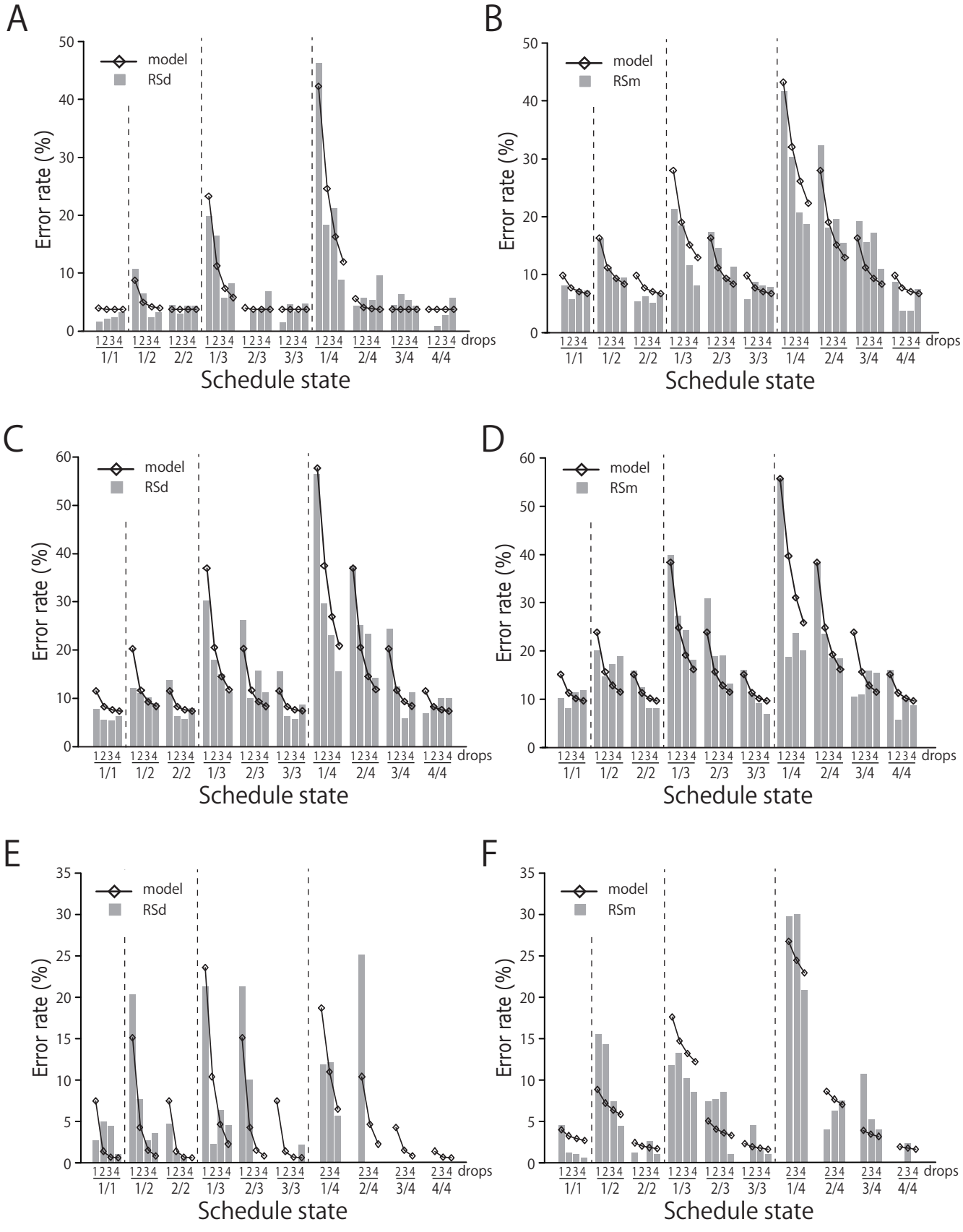

A
B
C
D
E
F
model
RSd
RSm
Error rate (%)
drops
1/1
1/2
2/2
1/3
2/3
3/3
1/4
2/4
3/4
4/4
Schedule state

Figure 9

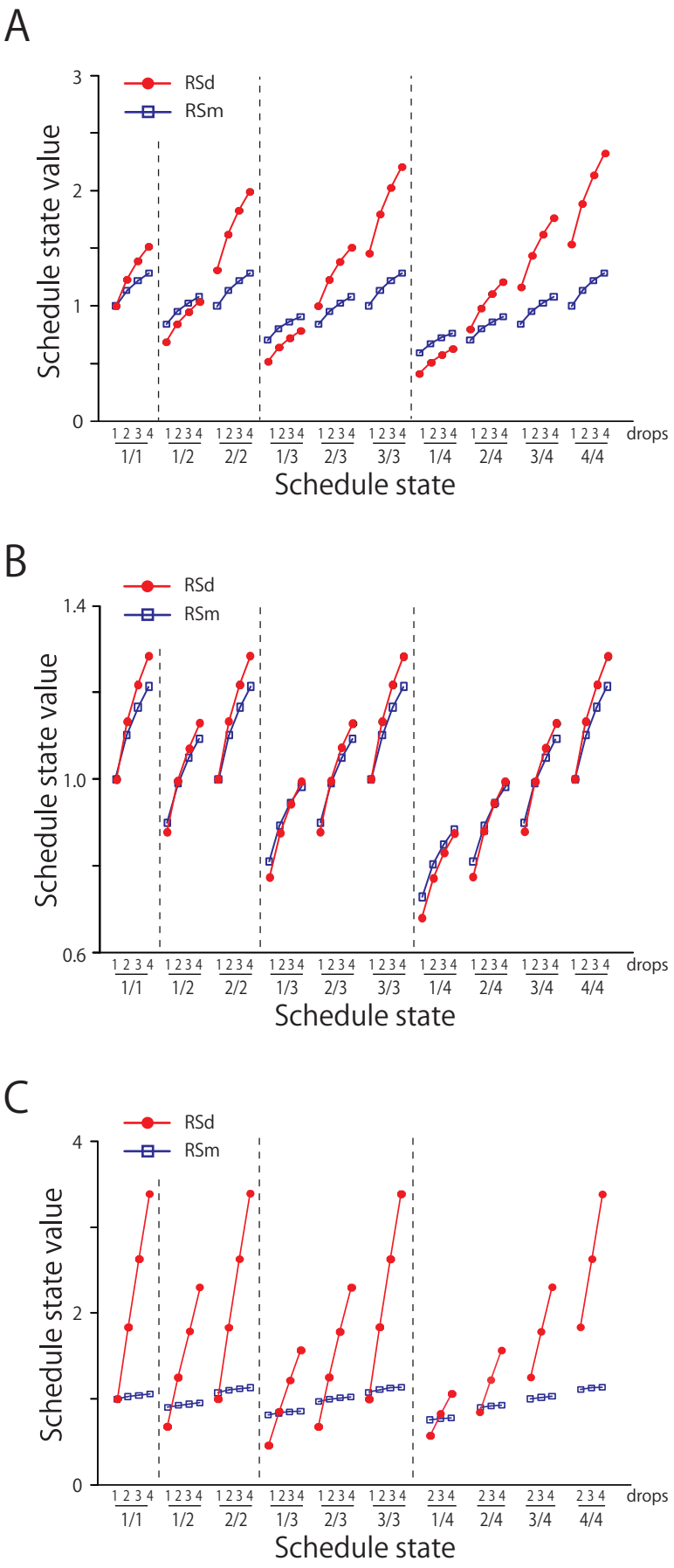

A
RSd
RSm
Schedule state value
3
2
1
0
1 2 3 4
1/1
1/2
2/2
1/3
2/3
3/3
1/4
2/4
3/4
4/4
drops
Schedule state
B
RSd
RSm
Schedule state value
1.4
1.0
0.6
drops
Schedule state
C
RSd
RSm
Schedule state value
4
2
0
1 2 3 4
2 3 4
drops
Schedule state

Figure 10

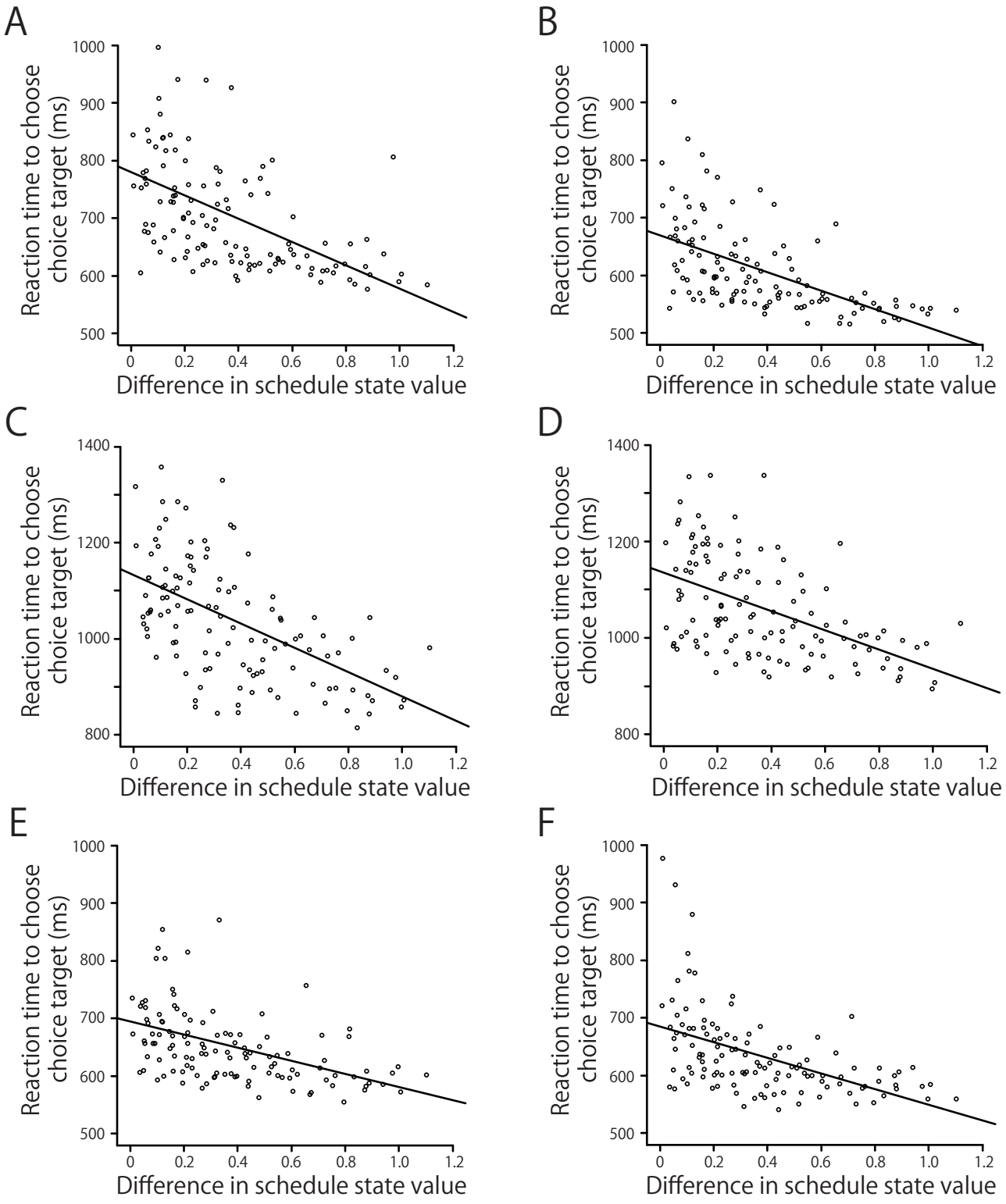

A
B
C
D
E
F
Reaction time to choose choice target (ms)
Difference in schedule state value
1000
900
800
700
600
500
1400
1200
1000
800
0
0.2
0.4
0.6
0.8
1.0
1.2

Figure 11

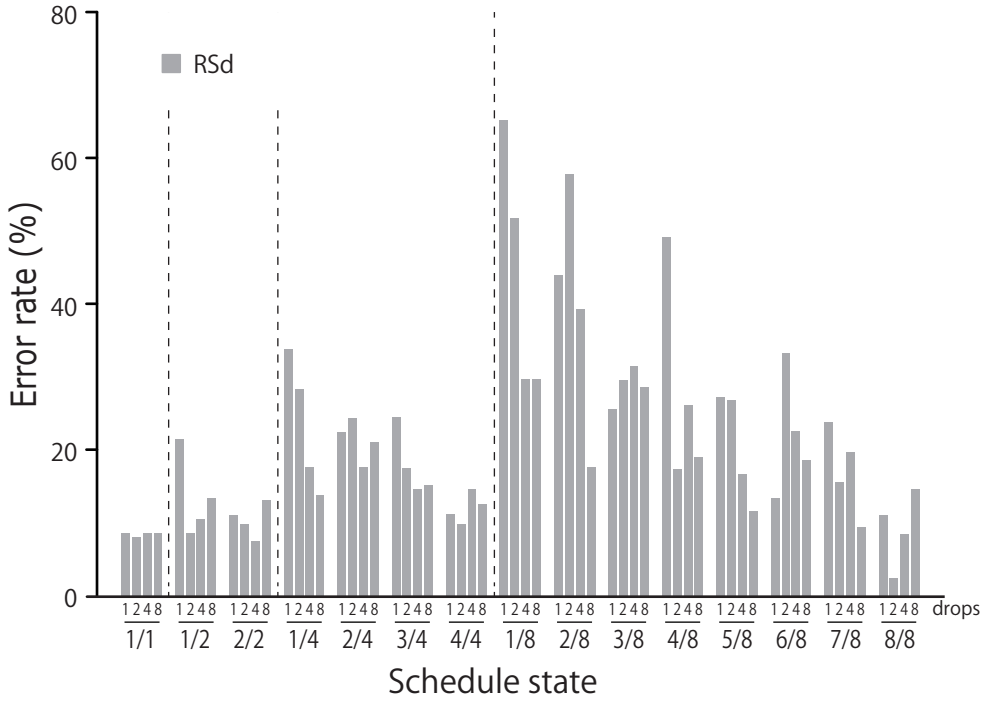

80
60
40
20
0
RSd
Error rate (%)
1 2 4 8
drops
1/1
1/2
2/2
1/4
2/4
3/4
4/4
1/8
2/8
3/8
4/8
5/8
6/8
7/8
8/8
Schedule state

Figure 12

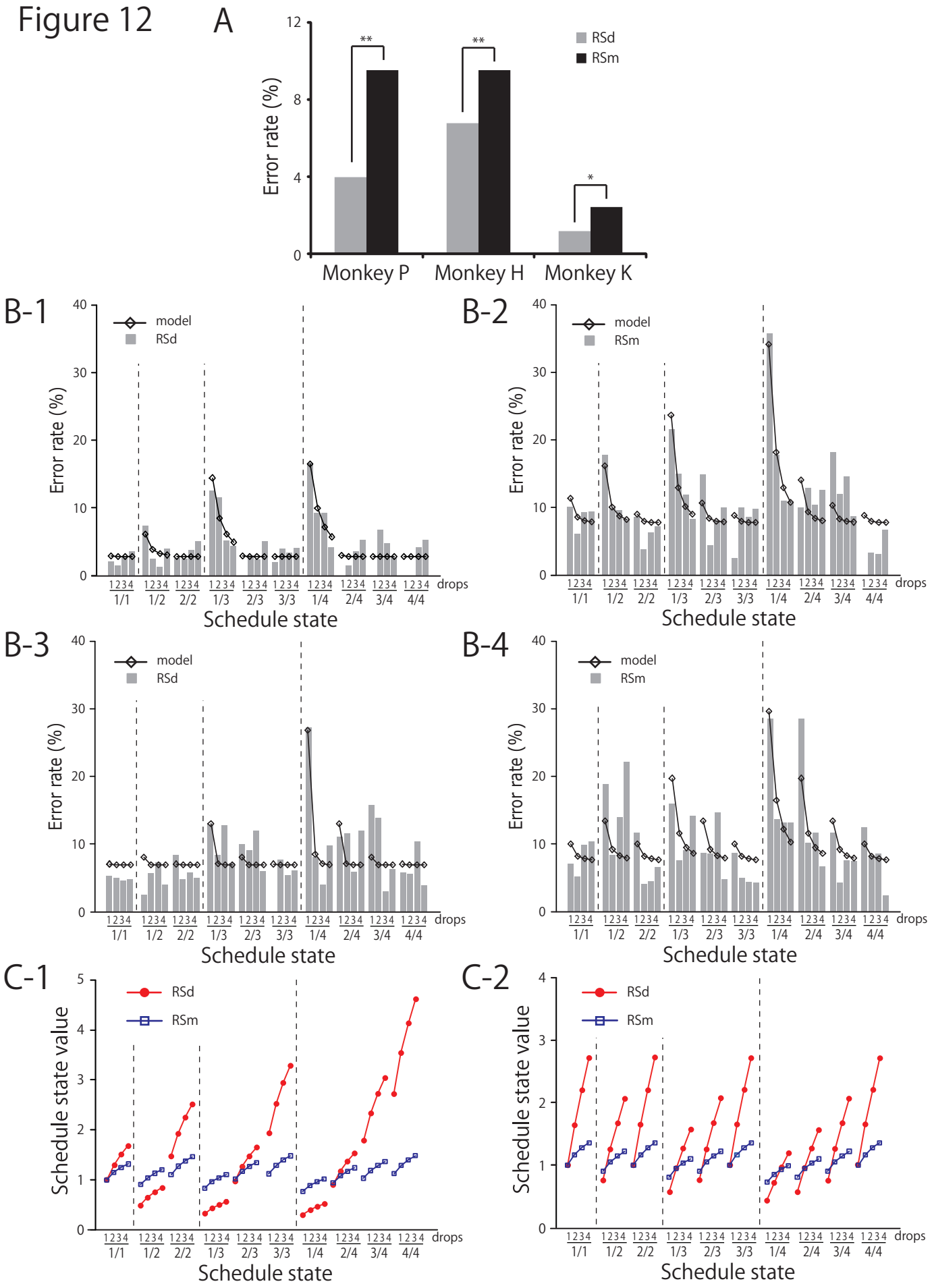

A
Error rate (%)
RSd
RSm
Monkey P
Monkey H
Monkey K
B-1
B-2
B-3
B-4
model
Schedule state
drops
C-1
C-2
Schedule state value

Table 1. The probabilities of chosen schedules in RSd

Monkey P	single trial	2 trials	3 trials	4 trials
1 drop	5.71 %	2.69 %	1.31 %	0.40 %
2 drops	9.29 %	5.98 %	3.00 %	2.39 %
3 drops	11.59 %	9.41 %	6.07 %	3.88 %
4 drops	13.10 %	11.28 %	8.25 %	5.66 %

Monkey H	single trial	2 trials	3 trials	4 trials
1 drop	7.00 %	2.69 %	1.05 %	0.55 %
2 drops	11.05 %	6.83 %	3.90 %	2.25 %
3 drops	11.29 %	8.80 %	6.40 %	3.73 %
4 drops	11.49 %	9.67 %	7.47 %	5.84 %

Monkey K	single trial	2 trials	3 trials	4 trials
1 drop	6.84 %	3.92 %	1.62 %	0.12 %
2 drops	9.58 %	6.03 %	2.80 %	0.93 %
3 drops	11.19 %	9.27 %	5.53 %	3.17 %
4 drops	13.50 %	12.13 %	9.20 %	4.17 %

Each color indicates the range of choice probability.

0 % ≦ ▭ < 5 %

5 % ≦ ▭ < 10 %

10 % ≦ ▭ < 15 %

Table 2. Results of Cochran–Armitage test

RSd Schedule state	*df*	Monkey P χ^2	*p*	Monkey H χ^2	*p*	Monkey K χ^2	*p*
1/1	3	3.81	= 0.28	4.01	= 0.26	6.26	= 0.10
1/2	3	27.97	< 0.05	2.51	= 0.47	31.69	< 0.05
2/2	3	1.55	= 0.67	17.64	< 0.05	10.54	< 0.05
1/3	3	28.13	< 0.05	27.83	< 0.05	14.41	< 0.05
2/3	3	9.29	< 0.05	22.13	< 0.05	42.52	< 0.05
3/3	3	2.49	= 0.48	12.27	< 0.05	3.21	= 0.36
1/4	3	44.97	< 0.05	77.92	< 0.01	14.89	< 0.05
2/4	3	4.70	= 0.20	28.34	< 0.05	31.11	< 0.05
3/4	3	0.98	= 0.81	18.36	< 0.05	—	—
4/4	3	8.56	< 0.05	0.74	= 0.86	—	—

RSm Schedule state	*df*	Monkey P χ^2	*p*	Monkey H χ^2	*p*	Monkey K χ^2	*p*
1/1	3	2.79	= 0.43	9.42	< 0.05	15.51	< 0.05
1/2	3	9.38	< 0.05	6.61	= 0.09	22.96	< 0.05
2/2	3	1.62	= 0.66	21.47	< 0.05	5.76	= 0.12
1/3	3	26.85	< 0.05	39.62	< 0.05	1.96	= 0.58
2/3	3	6.83	= 0.08	27.35	< 0.05	13.49	< 0.05
3/3	3	0.74	= 0.86	12.64	< 0.05	6.24	= 0.10
1/4	3	20.26	< 0.05	60.82	< 0.01	6.30	= 0.10
2/4	3	6.62	= 0.09	16.62	< 0.05	0.66	= 0.88
3/4	3	6.38	= 0.09	3.34	= 0.34	2.39	= 0.49
4/4	3	5.82	= 0.12	5.67	= 0.13	2.89	= 0.41

Table 3. The optimal values of all parameters estimated by ECS model

	Monkey P		Monkey H		Monkey K	
	RSd	RSm	RSd	RSm	RSd	RSm
γ	0.52	0.84	0.88	0.90	0.68	0.84
σ	0.46	0	0	0	0	0.08
β	9.2	6.8	10	9.8	2.6	9.80
m	0.30	0.18	0.18	0.14	0.88	0.04
δ	3.4	3.6	7.0	7.2	0	6.4
C	0.038	0.062	0.072	0.089	0.006	0.008