Journal of Robotics and Autonomous Systems

Dynamic alternation of primate response properties during trial-and-error knowledge updating

Atsushi Fujimoto [1], Satoshi Nishida[1] and Tadashi Ogawa [1]

[1]Dept of Integrative Brain Science
Graduate School of Medicine, Kyoto University
Sakyo-ku 606-8501, Kyoto, Japan

Abbreviated title: Trial-and-error knowledge updating in primates

Corresponding author: Tadashi Ogawa, PhD
Dept of Integrative Brain Science
Graduate School of Medicine, Kyoto University
Sakyo-ku 606-8501, Kyoto, Japan
Phone: +81-75-753-4678
FAX: +81-75-753-4486

e-mail: togawa@brain.med.kyoto-u.ac.jp

Abstract

Humans and animals seek appropriate solutions to novel problems through trial-and-error (TE) actions and observation of their outcomes. Once an individual has obtained the knowledge (rule) to solve a problem, knowledge-based (KB) actions may be applied in a stereotypical manner. Solutions can thus be based on TE or KB actions. To characterize this learning process at the behavioral level, we developed a new cognitive task for a laboratory monkey (*Macaca fuscata*) to perform. In this task, a search array consisting of six elements of different colors was presented, one of which was the behaviorally relevant target. The target color was changed unpredictably with no instruction or signal, requiring the monkey to use a TE search strategy to find the target color. We found that once the monkey identified the relevant color by chance after a color change, correct performance increased in a step-like manner and at the same time, other response properties (reaction time and color-choice tendency) also changed discontinuously. These step-like alternations in behavioral performance may be attributed to the subject's switching between TE and KB search strategies in the two phases. The present study has therefore provided behavioral evidence for the timing and manner of switching between search strategies during the process of updating knowledge.

Key words: Problem solution, Trial and error, Learning, Knowledge, Monkey

1. Introduction

Humans and animals can flexibly adapt to a changing environment, typically by utilizing two types of search strategies. When encountering a novel problem, the individual attempts to discover appropriate solutions by testing the effects of potentially relevant options. If one option is not successful, the individual can test another option [trial-and-error (TE) search strategy]. Through such TE searches, the individual constructs rule-based knowledge that identifies task-relevant sensory stimuli and associates them with arbitrary actions to solve the problem. Once such problem-solving knowledge is formulated, the individual repeats a stereotypical action that effectively solves a given problem [knowledge-based (KB) search strategy]. Thus, the ability to switch between two search strategies is a crucial aspect of problem solving in dynamically changing environments. These two types of search strategies are conceptually analogous to the exploration and exploitation strategies in the field of machine learning [1]. To understand the brain's control of the switching process, it may be essential to develop primate models that incorporate switching between the two types of search strategies during problem solving. An animal model allows us to conduct invasive studies in which single-unit activity in the brain is directly measured using an electrophysiological method (usually not allowed in human subjects) and to repeat an experiment with the same subjects to obtain highly reliable data. In this study, we conducted behavioral studies to examine the flexible adaptability of laboratory animals in the changing environment, which would provide significant insights to aid the development of intelligent behaviors in robotics and autonomous systems.

The behavioral tasks used in previous studies have not been appropriately designed for the examination of brain functions during flexible switching between TE and KB search strategies. Although many primate studies have examined the ability of flexible switching between multiple rules, in most studies monkeys did not engage in TE search behaviors because explicit instruction about the task rules was provided in advance [2-6]. In only a few studies, monkeys were trained to perform flexible switching of task rules without explicit instruction, using a simplified version of the Wisconsin Card Sorting Task (WCST) [7-11], which has been frequently used to test flexibility in human subjects. However, the simplified version of WCST entails only two task rules (shape- and color-match rules). So, though on the first trial after a rule change the subjects regularly selected a stimulus based on the old rule and performed incorrectly, this error feedback signaled a rule change, and the subjects could change their cognitive set to the new rule within only one or two subsequent trials [8]. This process inhibits the direct comparison of response properties between trials under the control of the TE search strategy and those under the control of the KB search strategy.

Other studies have trained primate or human subjects to perform reward-based free-choice tasks [12-15]. These tasks assigned differently sized rewards to alternatives, and then changed these assignments over time. The subjects were thus required to continuously adapt their strategy to track the changes in choice-reward assignments. Indeed, this task design could evoke both the TE and KB search strategies and provide a good analytical example of the reinforcement learning theory known as the exploration and exploitation dilemma [1]. However, both search processes may largely intermingle, and hence it is difficult to clearly classify a given trial as being under the control of a TE or KB search and to examine the neural mechanisms underlying the flexible switching between them.

To overcome these issues in previous studies, we developed a new cognitive task, the *target-tracking visual search task*, in which trials could be clearly classified as employing either TE or KB strategies. In the task, a search array consisting of six different-color elements was presented, and one of the elements was associated with the reward. The target color was changed unpredictably with no instruction after multiple successful trials. This task design required the subject to find the target color by itself through trial and error on some trials after a color change, whereas once the subject found the target color, it could select the target in the subsequent trials without trial and error by using its knowledge about the target color. Thus, the present task allows us to directly compare the differences in behavioral response properties between the TE and KB search strategies to provide crucial inferences about the strategy-switching process in the brain. Our primate subject could execute the present task by switching between the two search strategies, providing behavioral evidence for the timing and manner of strategy switching during the process of updating knowledge.

2. Materials and methods

2.1 Subject and surgery

Data were collected from one female Japanese monkey (*Macaca fusucata*) weighing 7.0 kg. A head holder and a recording chamber were secured to the cranium with implanted cortical screws and dental acrylic using sterile techniques while the animal was under deep anesthesia (intravenously administered sodium pentobarbital) and the head was immobilized. A scleral search coil was also implanted around one eye for the purpose of monitoring eye movements through electromagnetic induction [16,17]. All animal care procedures and experimental protocols were in accordance with the National

Institutes of Health Guidelines for the Care and Use of Laboratory Animals (1996) and were approved by the Animal Care and Use Committee of Kyoto University.

2.2 Visual stimuli and behavioral tasks

The experiments were controlled by two personal computers using custom-written software. One computer controlled the task and the other collected data. Visual stimuli were generated using a video signal generator (ViSaGe; Cambridge Research Systems, Cambridge, UK) and displayed on a cathode ray tube (CRT) monitor (CPD-G420; Sony, Tokyo, Japan). During the experiment, the macaque sat in a primate chair with its head secured in place to face the CRT monitor at a distance of 48 cm. The stimulus field was 38 × 30 deg in size, and the screen refresh rate was 100 Hz. The monitor resolution was 800 × 600 pixels.

The monkey was trained to perform a target-tracking visual search task (Fig. 1A). Each trial began with the appearance of a fixation point (small white circle) at the center of the monitor. The monkey was required to fixate on that spot within a square window of ± 1.5–1.7 deg (not visible in the figure). After fixation for 1200–1500 ms, an array of six equidistant elements appeared in a circle around the fixation point. The elements were differently colored circles (red, orange, yellow, green, blue, and magenta) of identical size (2 deg^2) and luminance (7.0 cd/m^2). The radius (6.3–7.7 deg), orientation, and color order of an array were randomly changed from trial to trial. After fixation for an additional 800–1000 ms, the fixation point disappeared (go signal) and the monkey was required to make a saccade to one of the elements. One color was defined in advance as the behaviorally relevant target. Immediately after saccade onset, the array stimulus disappeared. If the monkey made a saccade landing inside a square window of ±2 deg centered on the target, another fixation point appeared at the target position. After fixation to this point for 600 ms, the monkey received a juice reward accompanied by a high-pitched tone, indicating a successful trial. If the monkey made a saccade to a non-target-color element or to another location outside the target window, the trial was terminated and only a low-pitched tone was delivered, indicating an erroneous trial. If the gaze position deviated from the fixation window before the go signal, the trial was immediately aborted without reward or sound feedback. The trial was also aborted if the saccade reaction time was too short (<130 ms) because that saccade may have been initiated before the go signal. The target color remained the same within a block of 20–40 successful trials, and then changed (Fig. 1B). Because the subject was not explicitly informed of the target color or the timing of a color change, it had to determine each sequential target color through TE searches and assessment of their outcomes.

2.3 Eye movement recording and data analyses

The monkey was seated in a primate chair during the experiment. Eye movements were measured using the electromagnetic search coil technique (MEL-250UD; Enzanshi Kogyo, Tokyo, Japan). Voltage signals that separately encoded the horizontal and vertical components of eye position were analyzed at a resolution of 16 bits and 1 kHz. All data were stored and transferred to another personal computer for analysis using software based on MATLAB (The MathWorks, Tokyo, Japan). In this study, we only analyzed successful and erroneous trials. Aborted trials were excluded from the current analyses because the subject did not fulfill the required behaviors in those trials.

3. Results

3.1 An example of a target-tracking visual search task session

Figure 2 shows a case session that involved 12 trial blocks. The target color was set to yellow in the first block of the session (Fig. 2, top row). The subject was required to conduct a TE search because it was not explicitly informed of the target color. The monkey selected a green element in the first trial, yielding no reward (first erroneous trial, white arrow). This trial was followed by four consecutive erroneous trials; the subject selected the yellow element by chance in the sixth trial and received a reward (first successful trial, black arrow). The monkey continued to select the yellow target in a stereotypical manner and to receive a reward until the end of the first trial block, suggesting that the subject explicitly recognized the target color after the first successful trial. After multiple successful trials (in this case, 20 trials), the target color was changed to blue (Fig. 2, second row). The monkey continued to select the yellow element based on the target color in the previous trial block because it was not informed of the timing of a color change. However, the feedback indicating the erroneous trials in the second block signaled the color change, and the subject again initiated a TE search by choosing differently colored elements.

The subject thus regularly performed incorrectly after a color change by continuing to select the previous target color. The error feedback signals of one or two erroneous trials, however, alerted the subject to a color change and prompted a TE search strategy. Once the subject found the new target color (first successful trial), it tended to use a KB search strategy to select that color in subsequent trials within the block. We defined two phases to illustrate this transition process. The TE phase was defined as the interval from the first trial after a color change in which the subject selected a color other than the previous target color to the trial preceding the first successful trial (Fig. 2, shaded rectangles). The KB phase was defined as the interval from the trial following the first successful trial after a color change to the end of a block (Fig. 2, unshaded rectangles).

3.2 Step-like elevation of the success rate

The example data shown in Figure 2 suggest that correct performance should increase markedly and discontinuously following the first successful trial after a color change. To confirm this trend, we calculated a success rate using behavioral data from 71 sessions (928 blocks). The same data set was used for all analyses described below. Figure 3A shows the average success rate (black curve) as a function of the number of trials after a color change (vertical dashed line). The mean success rate at each trial count is defined as a ratio of the number of successful trials to the summed number of successful and erroneous trials and is computed by collapsing across trial blocks and recording sessions.

While the success rate increased gradually with the number of trials, this trend does not indicate that the subject's task performance developed gradually within each trial block. When the success rate was recalculated as a function of the number of trials after the first successful trial, the increase was rapid and the rate stabilized around the terminal success rate (mean success rate after first successful trial = 86.3%; Fig. 3B). Erroneous trials were rarely repeated in the KB phase. The average number of consecutive erroneous trials was significantly larger in the TE phase (mean ± SD = 4.0 ± 1.3 trials) than in the KB phase (1.1 ± 0.1 trials; rank-sum test, $p < 0.001$). The observed step-like elevation in the success rate suggests that changes in task performance were not based on skill learning or Pavlovian learning, but instead was based on updating of the target color knowledge.

To determine the way in which the subject chose a color element during the TE phase, we considered two hypothetical color-choice models using computer simulations of 10,000 blocks. The first model assumed that the subject remembered the perfect history of color choices. Thus, one of the five colors that were not the target color in the previous block was randomly selected in the first TE trial, and the color pool was reduced to the remaining four colors for the second TE trial. In this manner, the size of the color pool decreased as the trial number increased during the TE phase until the subject found the target color. The subject was assumed to perform at the actual terminal success rate (86.3%) after finding the target color. The simulated performance curve generated by this model (Fig. 3A, gray solid curve) was obviously dissociated from the actual performance curve (Fig. 3A, black curve).

The second model assumed that the subject had an imperfect history of color choices. We hypothesized three possible levels of memory capacity: the number of color choices that the subject could remember was zero, one, or two colors, and hence it randomly selected a color from the six-color pool (Fig. 3A, right dashed curve), from the color pool reduced to five colors (middle dashed curve), or from that reduced to four colors (left dashed curve). As illustrated in the figure, the actual performance curve was highly correlated with the simulated performance curve in which the subject remembered only one color choice. These analyses suggest that the subject did not have perfect history of color choices, but had imperfect history during a TE search phase after a color change.

3.3 Reaction time differences between the TE and KB search phases

To explore potential differences in response times between TE- and KB-phase trials, we analyzed the reaction time to start the fixation on the central spot (RTFP) [18, 19], which was defined as the interval between the fixation point appearance and the time at which the eye entered the computer-generated window around the element (± 1.5–1.7

deg). The maximum RTFP value was 800 ms (if the RTFP was larger than 800 ms, it was set to 800 ms), and RTFP was recorded as 0 ms when eye fixation occurred before the fixation point appeared. Figure 4A shows the median RTFP values separately calculated from the erroneous TE-phase trials (black cross) and the successful and erroneous KB-phase trials (gray circle and cross, respectively). We found that RTFP in the erroneous trials of the TE phase (230 ms) was significantly longer than RTFPs for both the successful (157 ms) and erroneous trials (154 ms) of the KB phase (rank-sum test, $p < 0.01$). These results indicate that the RTFP value was dependent on the phase of the currently experienced trial.

To understand the dynamics of RTFP changes at the transition from the TE to the KB phase, we calculated the median RTFP value as a function of the trial count at the time of a color change (Fig. 4B). For simplicity, we plotted only the data from the erroneous TE-phase trials and the successful KB-phase trials because the number of the erroneous trials during KB phase was small, and hence their median RTFP values varied largely across trials. The RTFP of the first erroneous trial immediately after a color change (Fig. 4B, left black cross) remained similar to RTFP values before the color change. However, RTFP markedly and significantly increased in the following trial (Fig. 4B, second cross from the left; rank-sum test, $p < 0.01$). Figure 4C illustrates the median RTFP value as a function of the number of trials from the first successful trial after a color change. RTFP exhibited a significant step-like decrease in the trial following the first successful trial (rank-sum test, $p < 0.01$), indicating that RTFP depended on whether the upcoming response was a TE or KB search, but did not depend on whether that response would be successful or erroneous (successful *vs.* erroneous KB-phase trials, rank-sum test, $p > 0.01$). In other words, RTFP was more dependent on across-trial internal criteria that determined the search strategy than on either the ongoing trial performance or the expected outcome.

We next conducted the same analyses for the saccade reaction time (SRT), which was defined as the interval between the fixation point disappearance and saccade onset (Fig. 4D). The median SRT value for erroneous TE-phase trials (169 ms; Fig. 4D black cross) was significantly longer than that for erroneous KB-phase trials (164 ms; rank-sum test, $p < 0.01$; Fig. 4D gray cross) but did not differ from successful KB-phase trials (169 ms; rank-sum test, $p > 0.01$; Fig. 4D, circle). These properties remained true across trials (Fig. 4E and F; rank-sum test for all successive trial pairs, $p > 0.01$). Thus, SRT in the KB phase was dependent on whether the upcoming response was successful or erroneous, indicating that SRT was dependent on the performance of the individual trials. On the other hand, SRT from erroneous trials significantly differed between the KB and TE phases, although sensory stimuli, motor responses, and outcomes were identical between the two trial conditions, indicating that SRT depended not only on trial-base performance but also on whether the upcoming response was a TE or KB search. Thus, results obtained from the analyses of both RTFP and SRT suggest that the search strategy altered between the TE and KB phases. This view is further confirmed by the following analysis on color-choice tendency.

If three or more consecutive trials in the KB phase were erroneous, the data from those and all subsequent trials within the block were excluded from the present analysis, because there was the possibility that the monkey mistakenly started TE search behaviors. Indeed, if we calculated RTFP and SRT for those consecutive erroneous trials in the KB phase, the median values (RTFP =239 ms, SRT = 166 ms) were comparable to those (RTFP = 230 ms, SRT = 169 ms) in the TE-phase erroneous trials rather than those (RTFP = 154 ms, SRT = 164 ms) in the KB-phase erroneous trials.

3.4 Differential color-choice strategies in the TE and KB search phases

After a color change, the subject used a TE strategy to seek the new target color. In contrast, the subject needed only to remember and repeatedly select the target color within a trial block once it was found. TE- and KB-phase search strategies thus appeared to differ. We sought to confirm this difference by comparing color-choice tendencies after erroneous TE- and KB-phase trials. Example data in Figure 2 show that the subject tended to select the target color of the current block after erroneous trials during the KB phase, whereas a variety of colors tended to be selected after erroneous trials during the TE phase. As mentioned in Section 3.3, given the identical conditions of erroneous TE- and KB-phase trials, the difference in color-choice tendency presumably depended on internal search strategy criteria.

To confirm this tendency, we calculated the probability of previously rewarded color selection after all erroneous trials in the data set (Fig. 5). The previously rewarded color of TE-phase trials was defined as the target color of the previous block, and that of KB-phase trials was defined as the target color of the currently experienced block. Figure 5A illustrates the probability of previously rewarded color selection after erroneous TE- and KB-phase trials. The difference between TE- and KB-phase trials (rank-sum test, $p < 0.01$) was statistically significant. It was higher in the KB phase (mean ± SEM = 88.9 ± 0.4%), suggesting that the subject chose a color element based on knowledge about the relevant target color. In contrast, this probability was markedly lower in the TE phase (19.5 ± 1.1%) and did not significantly differ from that attributable to chance (16.6%; rank-sum test, $p > 0.05$). Figure 5B and C show the dynamics of the probability of previously rewarded color selection at the transition from the KB to the TE phase or vice versa. The probability value significantly changed in a step-like manner in the trial following the first successful or erroneous trial (rank-sum test, $p < 0.01$), indicating that color-choice tendency based on a distinct search mode was flexibly alternated. This result is consistent with the computational simulation shown in Figure 3A, suggesting that the subject tended to randomly select color elements during the TE phase.

3.5 A hypothetical model for knowledge updating while solving an unknown problem

In the present visual search task, the target color was changed unpredictably without instruction. Due to this task design, our monkey was required to seek the target color by itself through trial and error and to update the knowledge to select the relevant color after each of the color changes. To illustrate underlying neural mechanisms that guided the monkey to execute the task described herein, we propose a hypothetical model using a state-transition diagram (Fig. 6).

Our monkey exhibited a step-like increase in its behavioral performance within one or two trials immediately after the first successful trial following a color change and maintained a high success rate in the subsequent trials (Fig. 3). In contrast, the behavioral performance of our monkey decreased in a stepwise fashion after the first erroneous trial followed by a low success rate in the subsequent trials. These results suggest that changes in behavioral performance were not due to stimulus–reward association learning [20] but were due to the updating of the target-color knowledge. Based on the emergence of these two distinctively different trial phases, individual trials could be classified into those that were under the control of the TE search strategy or those under the control of the KB search strategy. Corresponding to changes in the success rate, response properties (reaction times and color-choice tendency) substantially altered between the two search phases (Figs. 4 and 5). These search-strategy-dependent changes in behavioral performance were maintained stably within each of search phases (Figs. 4 and 5), even though sensory stimuli and required motor responses were essentially identical across trials, indicating that the brain could internally maintain the ongoing search strategy across trials (circles in Fig. 6).

The observed step-like changes in behavioral performance (success rate, reaction times, and color-choice tendency) indicate that search-strategy switching might be completed very transiently within one or two trials after the first successful or erroneous trial. Because the timing of strategy switching was not explicitly informed in the task, the monkey could detect the switch based on visual, sound, and reward feedback (post-saccade fixation spot, high-pitched tone, and reward were delivered in the successful trials, whereas only a low-pitched tone was delivered in the erroneous trials). However, the timing of strategy switching could not be determined solely by feedback because success and error feedback on the first successful and erroneous trials were the same as those on other successful and erroneous trials. The monkey therefore should determine the timing of strategy switching based not only on the state of feedback but also on that of the ongoing search strategy. That is, the first success should be derived from the combination of success feedback and TE search, whereas the first error should be derived from the combination of error feedback and KB search. Thus, the brain should monitor both the exogenous state of feedback and the endogenous state of the ongoing search strategy and then determine whether search strategy should be switched (solid arrows in Fig. 6).

Humans and animals make erroneous responses even when they know what the correct responses are. In the present experiment, our monkey also exhibited erroneous color choices on a small number of trials during the KB search phase, although it knew the relevant target color, as demonstrated by the strong tendency of the correct color choice on the post-error trials (Fig. 5). The important point was that the feedback and the ongoing search strategy (error feedback and KB search) were the same in the erroneous trials during the KB search phase and the first erroneous trials following a color switch. Nevertheless, interestingly, the monkey tended to select the color that differed from the pre-rewarded color on the trials following the first erroneous trials (i.e., switching of search strategy from KB to TE search), whereas it tended to select the pre-rewarded color on the trials following the erroneous trials during the KB search phase (i.e., non-switching of search strategy) (Fig. 5), indicating that the monkey's brain discriminated between these two types of erroneous trials. This view might be supported by the behavioral evidence that the saccadic reaction times were significantly different between these two types of erroneous trials (Fig. 4D), suggesting that errors in the KB phase occurred due to impulsive reactions to the go signal and yielded a premature saccade response. Thus, the brain could determine whether the erroneous trials were derived from the subject (endogenously induced error) or due to changes in the relevant color (exogenously induced error, dashed arrow in Fig. 6) and then determine whether the search strategy should be switched based on the factor that led to erroneous trials.

Taken together, hypothetical brain mechanisms underlying knowledge updating through TE during the target-tracking visual search task can be summarized by the state-transition diagram shown in Figure 6. Circles indicate the states of the search strategies, and arrows represent the transitions between states or within a state as caused by the input. Further work must be done to test the credibility of this state-transition model in the actual brain system. For this purpose, we will investigate how the brain controls currently observed actions by conducting electrophysiological recording from the prefrontal cortex of monkeys performing the present task. This may provide significant insight about the brain systems that allow individuals to flexibly adapt to novel and changing environments.

4. Conclusion

We developed a new animal model to investigate the nature of flexible switching between TE and KB searches in situations with an unknown task-relevant stimulus feature (color). This model maintained the same visual stimuli and required behavioral responses across trials, eliminating the effects of factors related to sensory stimuli and motor responses. The distinct differences in response profiles (reaction times and color-choice tendencies) between the TE and KB phases were thus due to internal factors such as search strategies. Our results therefore provide behavioral evidence that the subject utilized and switched between two different search strategies when knowledge updating was required to solve an unknown problem.

Acknowledgements

The authors gratefully acknowledge the continuous support and encouragement of Prof. Kenji Kawano. We also acknowledge Nihon University and National Bio Resource Center for supplying the monkey used in the present study. This work was supported by grants from Ministry of Education, Culture, Sports, Science and Technology of Japan (MEXT): 20500283, 20033010, and 22120507 to T.O.

References

1. R.S. Sutton, A.G. Barto, Reinforcement learning. MIT press MA, (1998).
2. E. Hoshi, K. Shima, J. Tanji, Neuronal activity in the primate prefrontal cortex in the process of motor selection based on two behavioral rules. J Neurophysiol 83, (2000) 2355-73.
3. J.D. Wallis , K.C. Anderson, E.K. Miller, Single neurons in prefrontal cortex encode abstract rules. Nature 411, (2001) 953-6.
4. M. Sakagami, H. Niki, Spatial selectivity of go/no-go neurons in monkey prefrontal cortex. Experimental Brain Research 100, (1994) 165-169.
5. I.M. White, S.P. Wise, Rule-dependent neuronal activity in the prefrontal cortex. Exp Brain Res 126, (1999) 315-35.
6. T. Ogawa, H. Komatsu, Target selection in area V4 during a multidimensional visual search task. J Neurosci 24, (2004) 6371-82.
7. K. Nakahara, T. Hayashi, S. Konishi, Y. Miyashita, Functional MRI of macaque monkeys performing a cognitive set-shifting task. Science 295, (2002)1532-6.
8. T. Kamigaki, T. Fukushima, Y. Miyashita, Cognitive set reconfiguration signaled by macaque posterior parietal neurons. Neuron 61, (2009) 941-51.
9. F.A. Mansouri, K. Matsumoto, K. Tanaka, Prefrontal cell activities related to monkeys' success and failure in adapting to rule changes in a Wisconsin Card Sorting Test analog. J Neurosci 26, (2006) 2745-56.
10. F.A. Mansouri, M.J. Buckley, K. Tanaka, Mnemonic function of the dorsolateral prefrontal cortex in conflict-induced behavioral adjustment. Science 318, (2007) 987-90.
11. F.A. Mansouri, K. Tanaka, Behavioral evidence for working memory of sensory dimension in macaque monkeys. Behav Brain Res 136, (2002) 415-26.
12. J. Pearson, M.L. Platt, Confidence and corrections: how we make and un-make up our minds. Neuron 63, (2009) 724-6.
13. D.J. Barraclough, M.L. Conroy, D. Lee, Prefrontal cortex and decision making in a mixed-strategy game. Nat Neurosci 7, (2004) 404-10.
14. L.P. Sugrue, G.S. Corrado, W.T. Newsome, Matching behavior and the representation of value in the parietal cortex. Science 304, (2004) 1782-7.
15. N.D. Daw, J.P. O'Doherty, P. Dayan, B. Seymour, R.J. Dolan, Cortical substrates for exploratory decisions in humans. Nature 441, (2006) 876-9.
16. S.J. Judge, B.J. Richmond, F.C. Chu, Implantation of magnetic search coils for

measurement of eye position: an improved method. Vision Res 20, (1980) 535-8.
17. A.F. Fuchs, D.A. Robinson. A method for measuring horizontal and vertical eye movement chronically in the monkey. J Appl Physiol 21, (1966) 1068-70.
18. Y. Kobayashi, Y. Inoue, M. Yamamoto, T. Isa, H. Aizawa. Contribution of pedunculopontine tegmental nucleus neurons to performance of visually guided saccade tasks in monkeys. J Neurophysiol 88, (2002) 715-31.
19. K. Okada, K. Toyama, Y. Inoue, T. Isa, Y. Kobayashi. Different pedunculopontine tegmental neurons signal predicted and actual task rewards. J Neurosci 29, (2009) 4858-70.
20. A. Pasupathy, E.K. Miller. Different time courses of learning-related activity in the prefrontal cortex and striatum. Nature 433, (2005) 873-6.

Figure legends

Figure 1. Visual stimuli and behavioral task. (A) Target-tracking visual search task. A trial was initiated with the appearance of a fixation point (FP) at the center of the display. Once the subject fixated on the FP for a random fixation period (1200–1500 ms), a search array consisting of six differently colored elements was displayed around the FP. One color was defined in advance as the behaviorally relevant target, but the monkey received no instruction about the target color. After a delay period (600–1000 ms), the FP disappeared and the subject was required to make a saccade toward one of the elements within 2000 ms. The array stimulus was extinguished when the subject initiated a saccade (small arrow). If the subject correctly selected the target color element, a liquid reward and a high-tone sound were provided after a 600-ms post-saccade fixation at the target position. Otherwise, the array stimulus was extinguished and a low-tone sound was given without reward. (B) Block design paradigm of color changes. Trials surrounding target color changes are illustrated. An array stimulus and a choice response (small arrow) for each trial are displayed in each panel. Circles and crosses indicate successful and erroneous trials, respectively. The target color was changed using a pseudo-randomized block design. A block consisted of 20–40 successful trials. Because the subject was not

informed of the target color or the timing of a color change, it had to search for the target color through trial and error during the initial trials of each block.

Figure 2. A case session depicting target-tracking visual search task and trial classification. The correct/incorrect state of an individual trial is shown as a function of the number of trials after a target color change. Colored numbers on the left side of the figure indicate block number and target color. Circles and crosses indicate successful and erroneous trials, respectively. The color of each trial symbol indicates the color that the subject selected in that trial. Gray crosses indicate erroneous trials in which no element was selected by saccade eye movements after the go signal. Aborted trials (fixation breaks before the go signal) are not shown. The TE phase was defined as the interval from the first trial after a color change in which the subject selected a color other than the previous target color to the trial preceding the first successful trial (shaded rectangles). The KB phase was defined as the interval from the trial following the first successful trial after a color change to the end of a block (unshaded rectangles).

Figure 3. Correct performance across sessions and blocks. (A) Success rate (mean ± SEM) as a function of the number of trials after a target color change (black curve).

Negative trial counts indicate trials before a target color change in the previous block. The vertical dashed line indicates the timing of a target color change. The horizontal dotted line indicates the mean success rate after the first successful trials (86.3%). To determine how the subject chose a color element during the TE phase, computer-simulated performance curves obtained from models in which the subject had perfect memory of past color choices (gray solid curve) and three levels of imperfect memory (gray dashed curves) are superimposed. (B) The same data set was replotted as a function of the number of trials after the first successful trial in a block.

Figure 4. Reaction time to the fixation point (RTFP) and saccade reaction time (SRT). (A) Median RTFP values for erroneous trials during the TE phase (black cross), successful trials during the KB phase (circle), and erroneous trials during the KB phase (gray cross). Asterisks indicate significant differences (rank-sum test, $p < 0.01$). (B) Median RTFP values as a function of the number of trials after a color change (dashed line). Asterisks indicate significant differences between two consecutive trials (rank-sum test, $p < 0.01$). (C) Median RTFP values as a function of the number of trials after the first successful trial following a color change. (D) Median SRT values for erroneous TE-phase trials (black cross), successful KB-phase trials (circle), and erroneous KB-phase trials

(gray cross). (E, F) Median SRT values as a function of the number of trials after a color change (E) and after the first successful trial (F).

Figure 5. Color-choice tendencies after erroneous trials. (A) Mean probability of previously rewarded color selection after erroneous trials during the TE (cross) and KB (circle) phases. The asterisk indicates significant difference between two phases (rank-sum test, $p < 0.01$). (B, C) Mean probabilities (mean ± SEM) of pre-rewarded color selection as a function of the number of trials after a target color change (B) and after the first successful trial (C). The vertical dashed line indicates the timing of a target color change. Note that data points for trials following a target color change (white arrow) and the first successful trials (black arrow) are not shown because all of the preceding trials were successful (100% correct performance).

Figure 6. A state-transition diagram for possible neural mechanisms underlying flexible search-strategy switching. Two circles indicate states of search strategies (TE and KB searches). Edges represent "transitions" either between two states or within a state caused by inputs labeled on each edge. Only exogenously induced error during KB search (dashed arrow) and success during TE search (thick black arrow) elicit the transition

between search strategies.

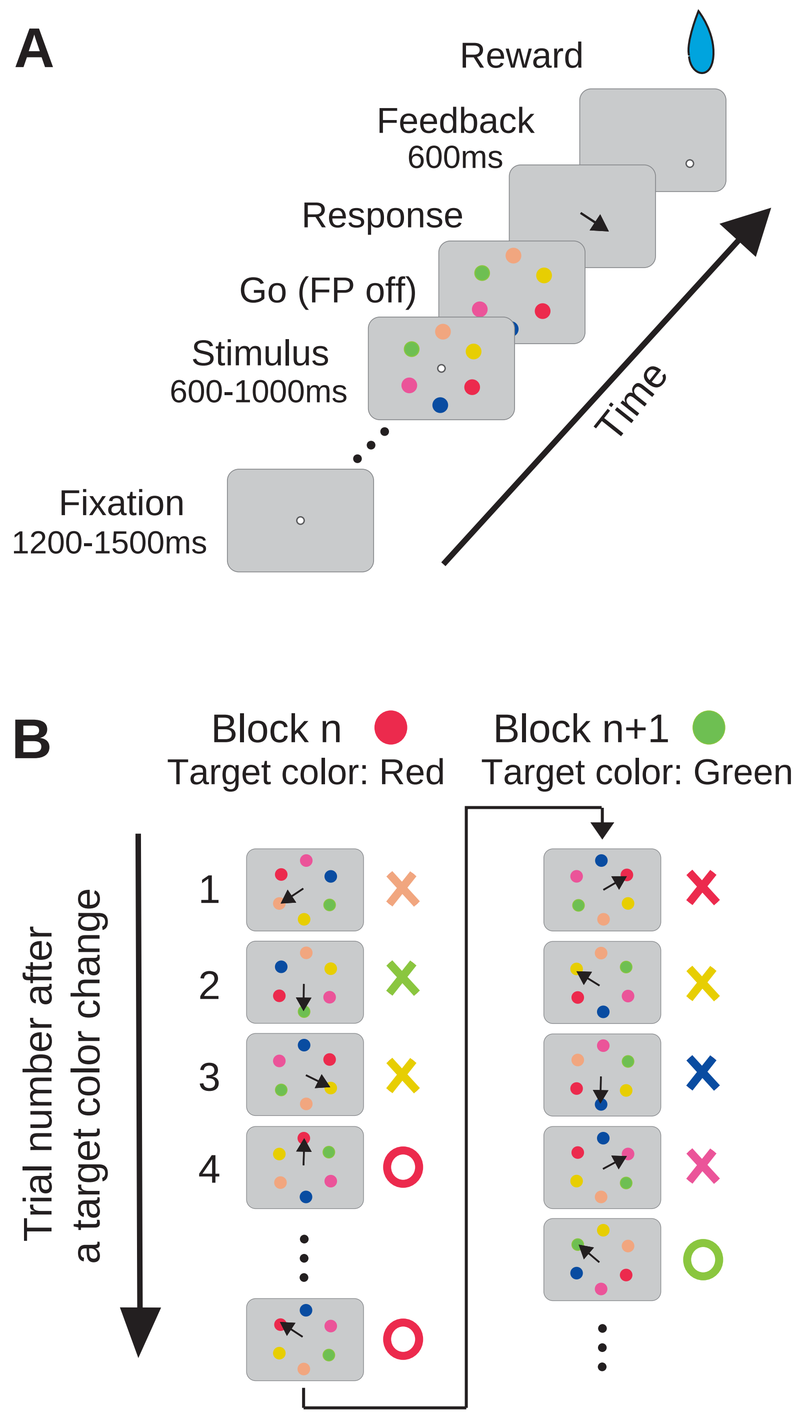

Figure 1 Color for Online

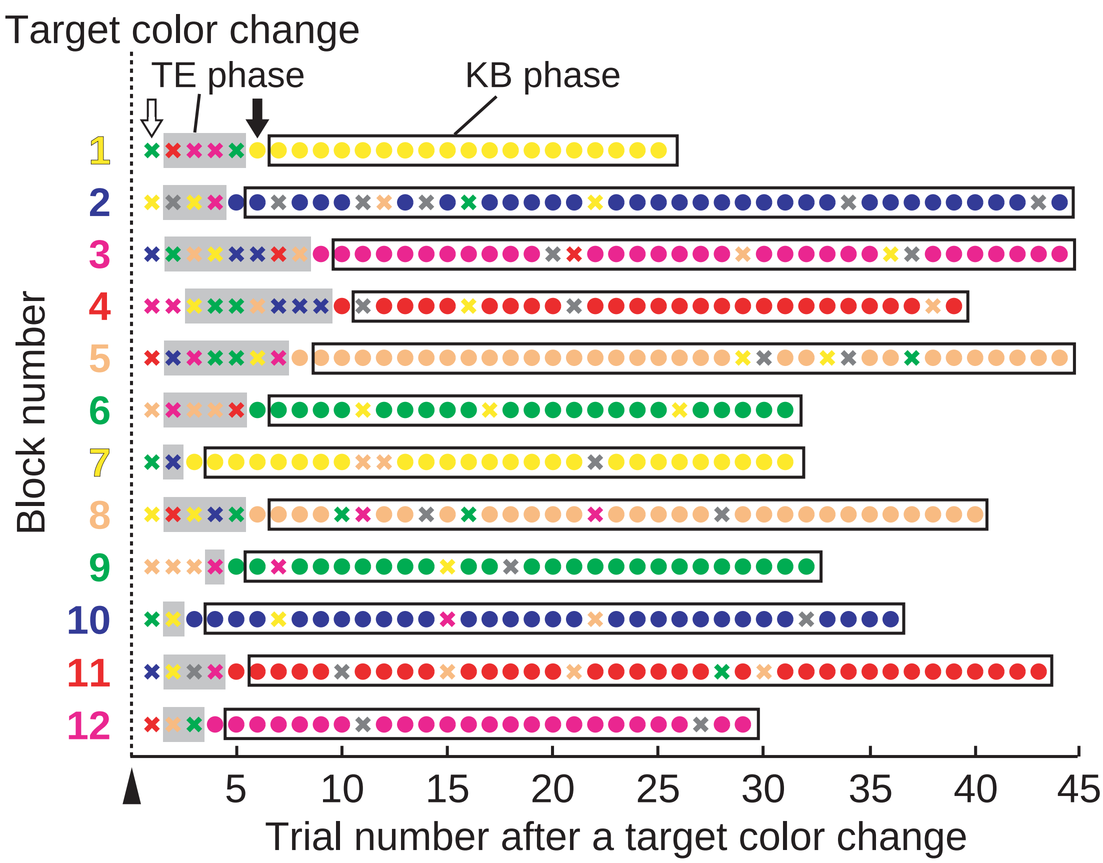

Figure 2 Color for Online

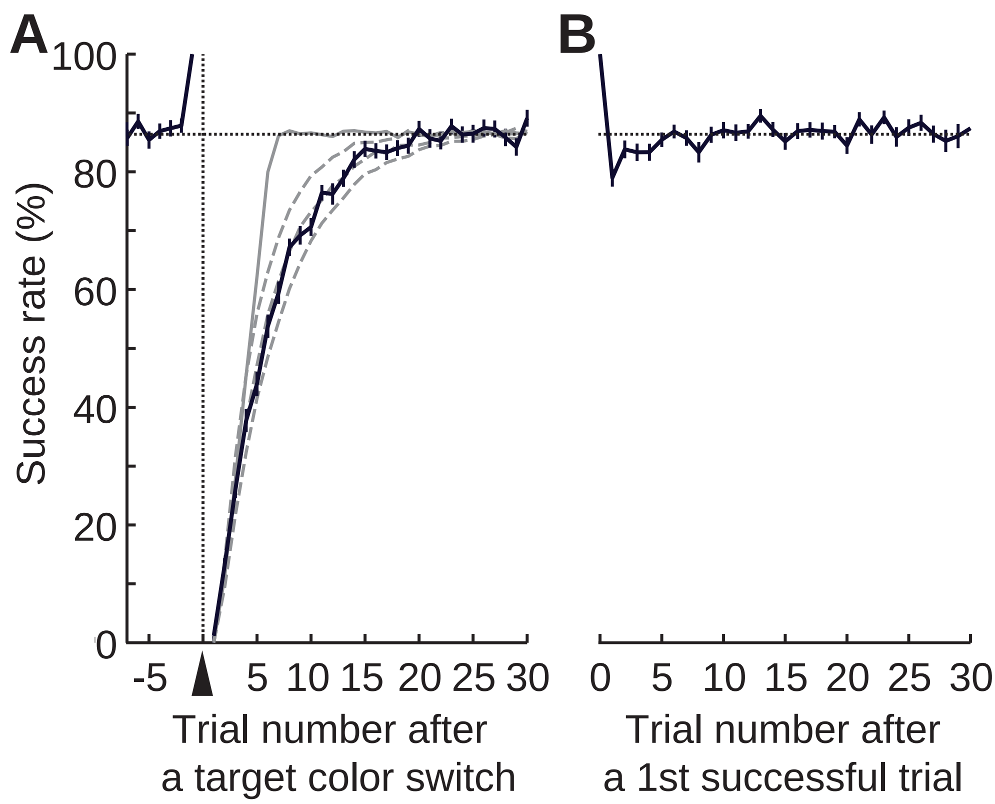

Figure 3 Black & White

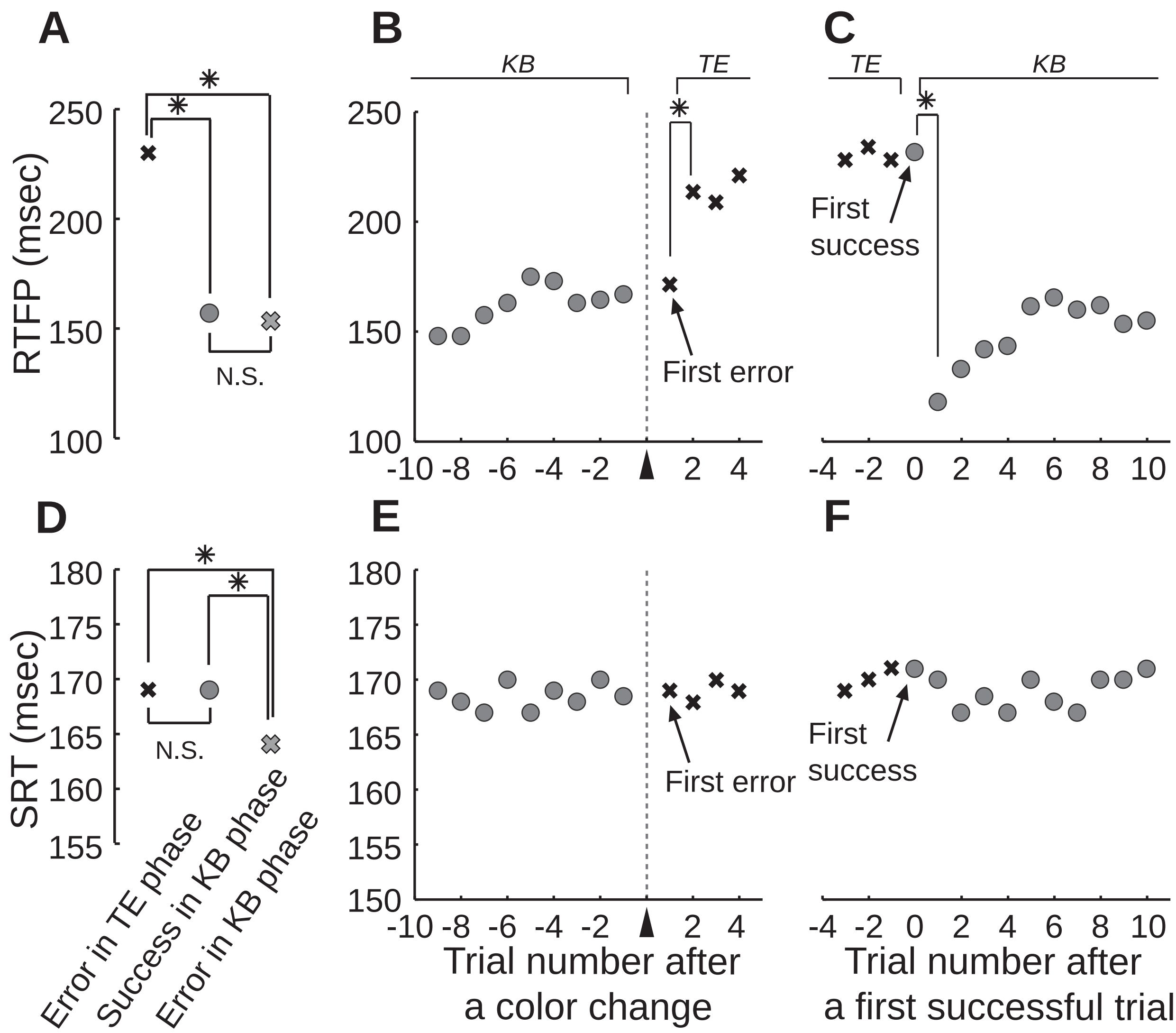

Figure 4 Black & White

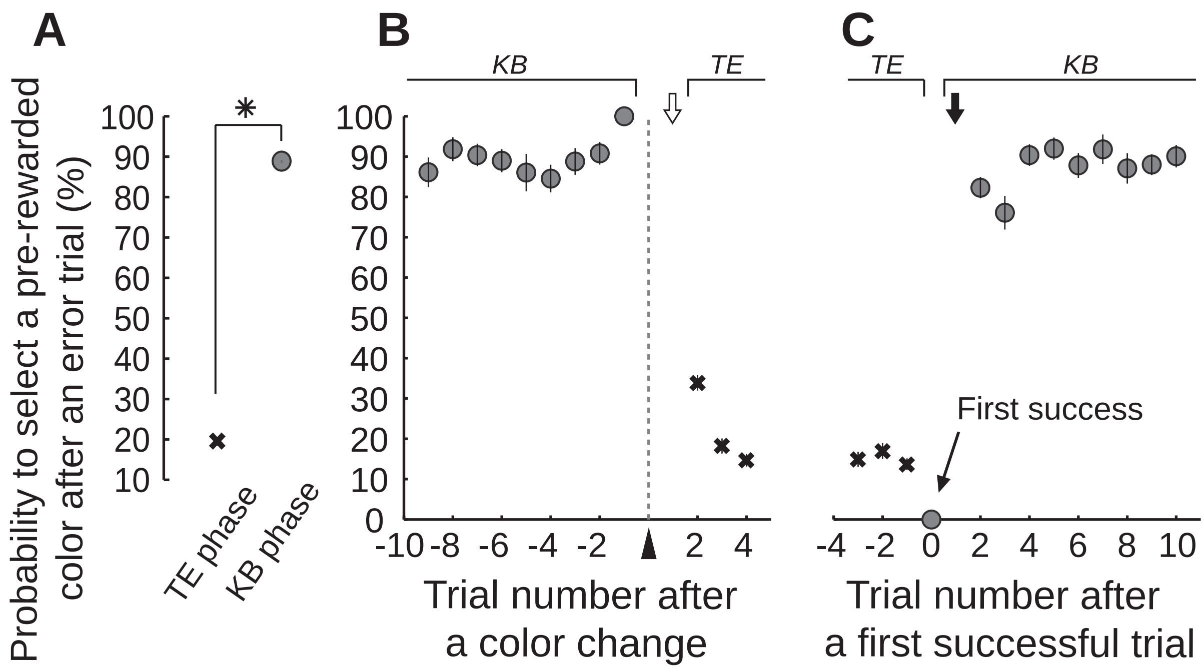

Figure 5 Black & White

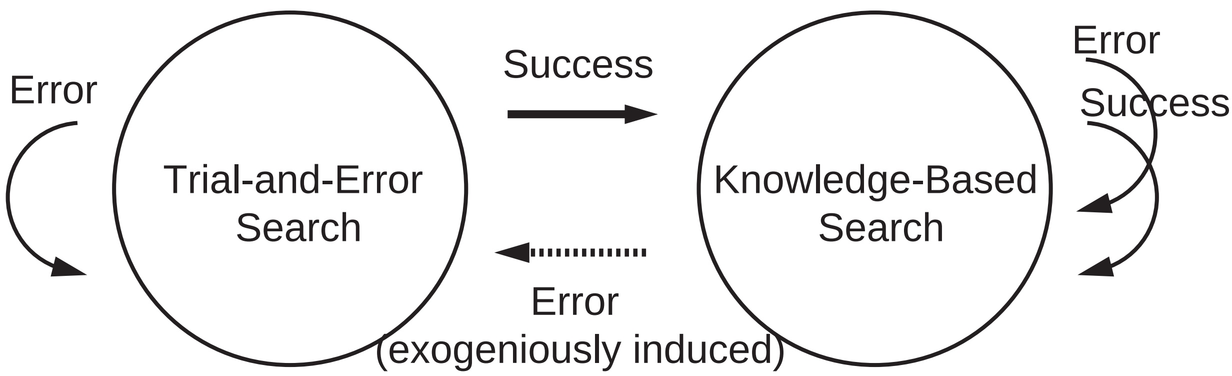

Figure 6 Black & White