Published in final edited form as:

Exp Brain Res. 2008 May ; 187(2): 321–330. doi:10.1007/s00221-008-1306-z.

Saccade selection when reward probability is dynamically manipulated using Markov chains

Samuel U. Nummela,

Systems Neurobiology Laboratories, Salk Institute for Biological Studies, 10010 N. Torrey Pines Road, La Jolla, CA 92037, USA

Neuroscience Graduate Program, University of California San Diego, 9500 Gilman Drive 0662, La Jolla, CA 92093, USA

Lee P. Lovejoy, and

Systems Neurobiology Laboratories, Salk Institute for Biological Studies, 10010 N. Torrey Pines Road, La Jolla, CA 92037, USA

Neuroscience Graduate Program, University of California San Diego, 9500 Gilman Drive 0662, La Jolla, CA 92093, USA

Richard J. Krauzlis

Systems Neurobiology Laboratories, Salk Institute for Biological Studies, 10010 N. Torrey Pines Road, La Jolla, CA 92037, USA

Abstract

Markov chains (stochastic processes where probabilities are assigned based on the previous outcome) are commonly used to examine the transitions between behavioral states, such as those that occur during foraging or social interactions. However, relatively little is known about how well primates can incorporate knowledge about Markov chains into their behavior. Saccadic eye movements are an example of a simple behavior influenced by information about probability, and thus are good candidates for testing whether subjects can learn Markov chains. In addition, when investigating the influence of probability on saccade target selection, the use of Markov chains could provide an alternative method that avoids confounds present in other task designs. To investigate these possibilities, we evaluated human behavior on a task in which stimulus reward probabilities were assigned using a Markov chain. On each trial, the subject selected one of four identical stimuli by saccade; after selection, feedback indicated the rewarded stimulus. Each session consisted of 200–600 trials, and on some sessions, the reward magnitude varied. On sessions with a uniform reward, subjects ($n = 6$) learned to select stimuli at a frequency close to reward probability, which is similar to human behavior on matching or probability classification tasks. When informed that a Markov chain assigned reward probabilities, subjects ($n = 3$) learned to select the greatest reward probability more often, bringing them close to behavior that maximizes reward. On sessions where reward magnitude varied across stimuli, subjects ($n = 6$) demonstrated preferences for both greater reward probability and greater reward magnitude, resulting in a preference for greater expected value (the product of reward probability and magnitude). These results demonstrate that Markov chains can be used to dynamically assign probabilities that are rapidly exploited by human subjects during saccade target selection.

Keywords

Eye movement; Saccade; Selection; Markov; Probability

Introduction

Markov chains are stochastic processes in which probabilities are assigned based on the previous outcome. Because of their ability to describe complex sequences of events, Markov chains have many applications, ranging from machine learning to modeling population growth. Of particular interest, Markov chain analysis is one of the key methods used to understand transitions between bouts of behavior in individuals as well as transitions in social dynamics (MacDonald et al. 2000). For example, a simple Markov chain model, in which foraging behavior depends upon the level of satiation, has been used to explain why many animals, including primates, spend more time resting than foraging (Herbers 1981). Learning based on Markov chains may be particularly important to social primates because it would allow them to infer the likely behavior of other animals based on observation of their current status. A series of experiments (Kornblum 1968, 1969) used Markov chains to manipulate target probabilities for manual responses, in which humans were explicitly informed about the underlying statistical structure. However, little is known about how well humans infer Markovian structure from a sequence of outcomes.

A strong candidate behavior for testing how well primates learn from Markov chains are voluntary saccadic eye movements. A variety of factors influence voluntary saccades. Basic visual properties like spatial contrast can make parts of a scene more likely to be the target of a saccade (Reinagel and Zador 1999; Peters et al. 2005). Nevertheless, top-down or cognitive factors also influence voluntary eye movements. For instance, in the well-known studies of Yarbus (1967), subjects' scan paths were very different when assessing disparate features of the painting such as the wealth of the figures versus their age.

Recent studies have identified several top-down factors that can influence saccade preparation, including estimates of target probability (Carpenter and Williams 1995; Dorris and Munoz 1998), reward magnitude (Kawagoe et al. 1998; Ikeda and Hikosaka 2003), and expected value (Milstein and Dorris 2007). Moreover, neural activity correlated with target probability has been observed in brain regions involved in generating saccades (e.g. Basso and Wurtz 1998; Dorris and Munoz 1998; Platt and Glimcher 1999).

In most cases, manipulation of target probability is related to some other variable that could be responsible for effects on saccade reaction time or physiological recordings. One confounding factor is that grouping trials into lengthy blocks with an identical probability distribution (e.g. Carpenter and Williams 1995; Dorris and Munoz 1998) makes target probability directly related to the frequency of target repetitions. This is a problem because target repetitions influence saccade preparation even when there is no accompanying change in target probability (Dorris et al. 2000; Fecteau and Munoz 2003). Although it is possible that repetition effects are driven entirely by empirical estimates of local target probability by the subject, other factors, like motor priming (Hackley and Valle-Inclan 2003), might be responsible for repetition effects on saccade preparation. Another confounding factor is that changing the number of potential targets in a visual scene may alter saccade preparation because of bottom-up changes in visual processing rather than the top-down influence of target probability. These sensory effects can be controlled by presenting blocks of trials with the same stimuli and probability distribution, but this solution reintroduces the problems caused by repetition effects.

Thus, although there is abundant evidence for top-down influences in saccade preparation, it is not clear from previous studies whether saccade selection can be influenced by probabilities governed by stochastic processes like Markov chains. Furthermore, Markov chains provide a potential technique to eliminate the confounds present in other task designs, because they can be used to control probability dynamically on a trial-by-trial basis instead of using an identical probability distribution for each trial.

In the following experiments, we assessed how well humans learn stimulus reward probabilities from a saccade selection task in which the probabilities were assigned by a Markov chain. On some sessions, we also examined the interactions between reward probabilities and reward magnitudes by varying the size of the reward depending on the location of the stimulus.

Methods

Subjects

Six naïve adult human subjects participated in these experiments (three women and three men, 22–36 years of age). Subjects gave their written informed consent, and all procedures were reviewed and approved by the Institutional Review Board. Each subject participated in seven to eight sessions, the first of which was a training session, and most sessions were close to 600 trials long.

General procedures

Subjects sat in a dark room 41 cm from a computer monitor displaying stimuli at a 75-Hz frame rate. Head movements were minimized using a bite-bar, and eye movements were sampled at 240 Hz using a video-based eye tracker (Iscan, Burlington, MA, USA). Experimental control and data acquisition were performed by Tempo software package (Reflective Computing, St Louis, MO, USA), and stimuli were presented using the Psychophysics Toolbox (Brainard 1997; Pelli 1997) extensions for Matlab (MathWorks, Natick, MA, USA).

Markov tasks

Subjects performed a saccade selection task among four identical stimuli. The temporal sequence of stimuli presentation during two experimental trials is illustrated in Fig. 1a. Subjects began each trial by fixating a central $0.5 \times 0.5^\circ$ spot for a randomized 200–400 ms interval. Eye position was monitored to confirm that subjects maintained fixation within 3° of the central stimulus. After fixation, the central spot disappeared with simultaneous appearance of four identical choice stimuli (discs of 0.4° radius evenly spaced at an eccentricity of 7° from fixation). After onset of the four stimuli, subjects made a saccade directly to the stimulus location of their choice and were required to fixate the stimulus to an accuracy of 3° for 200 ms. Following selection, the central fixation spot reappeared, and upon returning to the fixation spot, subjects were given visual and auditory feedback about their choice. Visual feedback consisted of an image of a coin presented at the rewarded location with a number indicating the monetary award (in cents) for that trial; auditory feedback consisted of a 1-s tone that indicated whether the choice for that trial was correct (100 Hz incorrect, 200 Hz correct). A correct choice meant that the subject selected the rewarded location, and only one stimulus location was rewarded on each trial. This location was drawn from a uniform probability distribution, and subsequent rewarded locations were assigned by a Markov chain. On each trial, the rewarded location was drawn according to one of four probability distributions, each of which depended on the rewarded location of the previous trial (i.e., the state of the Markov chain). Thus, the rewarded location on each

trial transitioned between trials, and the transition probabilities formed a four by four contingency table such as the one shown in Fig. 1b.

Subjects received specific instructions before participating in the task. First, subjects were shown a schematic of the task, similar to Fig. 1a, to generally indicate how they should orient their gaze during each epoch of the task. Second, subjects read the following instructions about what to do during the selection epoch: “On each trial, only one stimulus will have a reward. Shift your gaze to the stimulus that you think is most likely to be rewarded.”

Subjects were run under two conditions. In the “uniform-reward” condition, correct selections to any of the choice stimuli yielded a uniform 4-cent reward. In the “varying-reward” condition, correct selections of the choice stimuli yielded a reward magnitude of 1, 2, 6, or 8 cents, dependent only on the location of the choice stimulus. The assigned magnitudes remained stationary for entire sessions and were revealed before the start of each session by displaying coins with numbers indicating the monetary award at each location; the same coins were also used to indicate the rewarded location during the visual feedback of each trial. Subjects were paid the greater of their accumulated reward total or ten dollars for each session.

Contingency tables for each session, each describing the transition probabilities of a Markov chain, were constructed in Matlab with several constraints. For all uniform-reward experiments, the four stimulus locations were assigned probabilities of 0.0136, 0.0915, 0.2791, and 0.6157, regardless of the Markov state. However, each Markov state assigned each probability to a different location (e.g. see Fig. 1b). For all varying-reward experiments, the four stimulus locations were assigned probabilities of 0.0444, 0.1494, 0.3037, and 0.5025. The smallest reward probability was always assigned to the previously rewarded stimulus to prevent long strings of the same rewarded stimulus. In addition, the probabilities were assigned so that each location would be rewarded with a probability of 0.25. For each subject, by shuffling the transition probabilities, we ensured that no two sessions shared identical contingency tables.

After the third session of the uniform-reward task, five of six subjects were asked to describe, by written statement or diagram, their strategy for making selections during the task.

Analyses

Saccades were detected using eye velocity and acceleration criteria (velocity $>35^\circ$ per s, acceleration $>700^\circ$ per s²) and confirmed by visual inspection (Krauzlis and Miles 1996). Saccadic reaction times for each trial were defined as the difference between the time of fixation spot offset and the time of saccade initiation. Saccade selections for analyses were determined by calculating the closest stimulus to the saccade end point.

To determine how quickly performance, on average, improved to better than chance, binomial distributions were fit to the fraction of correct selections binned across sessions for each trial, and we determined the first trial which was significantly greater than performance expected by chance. To determine if subject performance improved across sessions, binomial distributions were also used to describe the fraction of correct selections binned across the first and final 100 trials of each session; no significant differences in performance were found between these epochs, so full data sets were used for analyses. As described in the results, in an additional set of experiments we informed naïve subjects that a Markov chain assigned reward probabilities; significant improvements in the fraction of correct

selections between the first and final 100 trials were found for the “informed” subjects, so only the last 100 trials were used for further analyses.

For each session, correlation between reward probabilities and the fraction of subject selections of each reward probability were calculated, and a bootstrap procedure using 1,000 replicates was used to test significance (Efron and Tibshirani 1998). A Wilcoxon rank sum test was used to determine if subjects made significantly more selections of the greatest reward probability in sessions after they were informed than in sessions when they were naïve. A chi-square goodness of fit was used to test if selections were significantly different from a uniform distribution.

To test for significant effects of reward probability on saccade reaction times, one-way ANOVAs were performed separately for each subject with saccade reaction times grouped by reward probability. The data set for each subject included all uniform-reward session where the subject was naïve about the Markov task. Correlation coefficients between reward probability and mean saccade reaction times were also calculated, and a bootstrap procedure, as described above, was used to test for significance.

Statistical dependence between subject selections and Markov states was assessed by estimating mutual information (Cover and Thomas 1991), and confidence intervals for mutual information were constructed by a bootstrap procedure. Statistical dependence expected by chance was calculated by estimating mutual information after scrambling the relationship between Markov states and subject selections; this was done 1,000 times for each session, and the resulting distributions were used to calculate a one-tailed 95% confidence interval. A Wilcoxon rank sum test was used to compare the mutual information content of all sessions with naïve subjects to the mutual information content of all sessions with informed subjects.

To test if subjects improved their reward accumulation over the course of varying-reward sessions, we compared the accumulated reward in the first and final 100 trials of each session, again using a bootstrap procedure. Full data-sets were used for further analyses when no significant change in performance within a session was found. If a significant difference in performance was observed, only the latter half of the session was used for calculating the fraction of trials selected by reward probability, reward magnitude, and expected value (the product of reward probability and magnitude). A Wilcoxon rank sum test was used to assess whether the numbers of choices related to reward probability, reward magnitude and expected value were significantly different from each other.

Results

The reward probabilities assigned by the Markov chains consistently and significantly influenced the distribution of saccade selections with no consistent effect on saccade reaction time (see Table 1). Therefore the rest of the “Results” section focuses on examining the subjects’ choice behavior.

The time-course of learning from a Markov chain

For the uniform-reward experiment, in which reward probabilities were assigned by the Markov state but reward magnitudes were uniform across all stimuli, individual sessions indicated that subjects made use of the underlying Markov statistics when selecting targets. Figure 2a, plots the locations of the first 200 selected stimuli of a typical session for two subjects. The Markov chains were designed so that each stimulus location would be rewarded with a probability of 0.25. Accordingly, subjects exhibited no obvious pattern of preference for particular stimulus locations. However, sorting the subject selections by

reward probability, shown in Fig. 2b, reveals a preference for higher reward probabilities. Figure 2c illustrates how well subjects accumulated rewards over the course of the sample session by plotting the fraction of correct (rewarded) selections in a running boxcar of 20 trials. Within the first 20 trials, the fraction correct rises above the 0.25 value, expected by chance, and remains elevated throughout most of the session.

To examine the course of learning at the timescale of a few trials, we calculated the fraction of correct selections across all uniform-reward sessions for all subjects. Figure 3a shows the fraction of correct selections in bins of four trials, with 95% one-tailed confidence intervals, and Fig. 3b provides a closer view of the first 30 trials. By the second bin of four trials, subject performance was significantly above the 0.25 fraction correct expected by chance, and remained significant for most of the subsequent trials. Thus, learning from the Markov chains in our task occurred very quickly.

In contrast to the rapid learning observed at the start of a session, there was no significant improvement in subject performance over the rest of the session. Figure 4 quantifies performance for each of the six subjects by plotting the fraction of correct selections in a running boxcar over the course of each session (three sessions per subject). As expected by the rapid improvement of performance shown in Fig. 3, the fraction of correct selections for each session was significantly greater than the fraction correct expected by chance ($P < 0.05$). However, subject performance remained fairly steady over the course of individual sessions. After the initial rapid learning, there were no significant changes in the fraction of correct selections between the beginning and the end of a session for any of the subjects (comparison of binomial distributions of the first 100 trials and last 100 trials, $P > 0.05$ for all sessions). Due to the absence of significant changes in performance beyond the rapid improvement at the start of each session, all trials from each session were pooled for subsequent analyses.

How the Markov chains influenced saccade selections

The clearest evidence that the Markov structure influenced saccade selections is the consistent preference for stimuli assigned high reward probabilities. Figure 5a shows the mean fraction of subject selections by location for each subject. Although subjects did not select each location an equal number of times ($P < 0.05$ for 14/18 sessions), their preference for particular locations did not improve their fraction of rewarded selections. In contrast, Fig. 5b demonstrates that, on average, subjects had a monotonically increasing preference for increasing probabilities of reward. Analysis of individual sessions revealed significant correlations between reward probability and fraction of subject selections ($P < 0.05$ for all sessions, smallest correlation 0.52, mean correlation 0.90). Subject preference for stimuli with greater reward probabilities can only be explained by subjects conditioning their selections on the Markov state.

Another clear piece of evidence that the Markov chains influenced saccades selections is that, after performing the task, subjects were able to report the contingencies in the task. Upon completing their third session of the uniform-reward task, five of the six subjects were asked to describe their strategy in making selections. Subjects tended to describe their strategies in terms of the reward transitioning from one location to another on a trial-by-trial basis, indicating knowledge that the previously rewarded location was important for determining the next rewarded location. For example, one of the subjects drew a diagram of the four choice stimuli with arrows indicating their preferred selection following a reward at each stimulus. Four of the five subjects identified the stimulus most likely to be rewarded following each Markov state; the other subject identified three of the four. All subjects also described idiosyncratic strategies. For example, one subject reported, "I noted that occasionally it would go across, so I occasionally went across, but I could not predict

when.” In general, the self-reports indicate that subjects employed strategies that drew on knowledge of the Markov chains as well as incidental patterns observed during the session.

Despite the use of the underlying Markov statistics, most subjects did not approach a strategy that maximized reward. One possible reason for this outcome is that the stochastic draws from the Markov chain did not provide enough evidence for an observer to infer the greatest reward probability over the course of a session. However this is contradicted by the subjects’ ability to verbally report which stimulus was most likely to be rewarded following each Markov state. Moreover, by 100 trials, for every Markov state, the stimulus with greatest reward probability was indeed the most rewarded stimulus. Another possible reason is that subjects never learned to construct a completely Markovian model for guiding their selections; this explanation is consistent with subject descriptions of idiosyncratic strategies. To further test for this possibility, three subjects performed another uniform-reward session after being informed about Markov models and that rewards in the task were assigned by a Markov chain (without revealing the particular contingency table used). In contrast to their performance when they were naïve about the underlying Markov structure, informed subjects made significantly more correct selections over the entire course of each session (comparison of binomial distributions of the first 100 trials and last 100 trials, $P < 0.05$ for all sessions). Complimentary to improvements in performance, Fig. 5c shows that informed subjects selected the stimulus with greatest reward probability more often than when naïve (Wilcoxon, $P < 0.01$).

We quantified the degree to which subjects used the underlying Markov chains to guide saccade selections by estimating mutual information. If subjects were guiding their decisions by the Markov state, as subject performance suggests, then there should be a significant amount of mutual information between the Markov states and subject selections. Figure 6 shows the mutual information for each session. In this four-alternative forced-choice task, 2 bits of information would completely account for subject selections, and less than 0.05 bit of information was expected by chance for any given session. We found that the Markov state provided a mean of 0.55 bits of information about subject selections. A statistical dependence of 0.55 bits is consistent with a subject preference for higher reward probabilities, based on the Markov state, while sometimes still selecting smaller reward probabilities. After being informed that reward probabilities were assigned by a Markov chain (Fig. 6, dark gray bars), subjects had significantly more mutual information than their best session while naïve (comparison of distributions from bootstrap, $P < 0.05$). Informed subjects also had significantly more mutual information than the population of naïve subjects (Wilcoxon, $P < 0.05$) with a mean of 1.1 bits across all three sessions, indicating a stronger relationship between Markov state and saccade selection.

Interactions between reward probability and magnitude

Subjects accumulated more reward than expected by chance during varying-reward sessions, in which each stimulus location had a different reward magnitude of 1, 2, 6, or 8 cents. Figure 7 displays a running boxcar of the reward accumulated over the previous 50 trials for each session, sorted by subject. Once again, every subject on every session accumulated more reward than expected by chance ($P < 0.05$), whereas generally they did not show significant improvement in performance past the first 100 trials of a session (comparison of bootstrapped distributions, 14/18 sessions, $P > 0.05$). The collection of more reward than expected by chance could be from a selection preference for greater reward magnitude, greater reward probability, or greater expected value (the product of reward probability and magnitude).

Saccade selections from varying-reward sessions show that subjects improved reward accumulation by combining reward probability and magnitude. Figure 8a–c summarizes all

subject selections for varying reward sessions sorted by reward probability (a), reward magnitude (b), and expected value (c). Whereas subjects favored greater reward probabilities, reward magnitudes, and expected values, on most trials, these factors are confounded because the stimulus with the greatest expected value usually also had the greatest reward probability or magnitude (or both). However, there was a subset of trials in which the greatest reward probability, reward magnitude, and expected value were assigned to different stimuli. This subset of trials, referred to as “conflicting trials” and summarized in Fig. 8d–f, demonstrates that subjects selected the greatest expected value (f) significantly more often than either the greatest reward probability (d) (Wilcoxon, $P < 0.0001$) or the greatest reward magnitude (e) ($P < 0.0001$). Thus subjects were willing to sacrifice selection of the greatest reward probability and the greatest reward magnitude in order to accumulate more reward by selecting the greatest expected value most often.

Discussion

Our results demonstrate that human subjects can take advantage of the statistical structure of a Markov chain during a saccade selection task, even when naïve to the existence of the underlying Markov chain. The clearest evidence is provided by our observation that distributions of saccade selections were significantly correlated with stimulus reward probabilities. We also measured statistical dependence between Markov state (or previous outcome) and the subsequent saccade selection, indicating the extent to which subjects incorporated the Markov structure into their decisions. Whereas naïve subjects take partial advantage of Markovian structure in order to improve performance, they benefited more after being informed that the probabilities comprised a Markov chain. Informed subjects selected the greatest reward probability significantly more often than naïve subjects, and had a stronger statistical dependence between Markov state and saccade selection.

Naïve subjects tended to adopt strategies that did not maximize reward accumulation. This result is consistent with an extensive literature of suboptimal human performance on manual free choice tasks in which probabilities are assigned independently between trials (reviewed by Vulkan 2000), and probability classification tasks (Gluck and Bower 1998; Knowlton et al. 1994; Gluck et al. 2002). One possible reason subjects did not reach optimal performance is that they may have used a poor model to guide their saccade selections. We tested this possibility by having several subjects perform the same task after they were informed that a Markov chain assigned reward probabilities. The informed subjects improved their performance over hundreds of trials and approached a strategy that maximized reward accumulation. This outcome provides evidence that naïve subjects used incorrect models to guide their selections. Nevertheless, even informed subjects did not always select the greatest reward probability, indicating that some other factor, such as motivation or uncertainty of probability, also limited subject performance.

Although naïve subjects did not tend to adopt maximal strategies, they did manage to take advantage of the Markov chain very quickly (within ten trials) in order to improve performance. Such learning is feasible within very few trials so long as subjects immediately viewed the location of previous rewards as a salient cue to guide selections, and they largely based their selections from the small number of observed transitions in reward location. It is well documented, and perhaps unsurprising, that humans would treat a small sample of transitions as representative of future transitions (Tversky and Kahneman 1971). It is more surprising that subjects would so quickly view the transitions in reward locations as an important factor in performing the task. The fast learning indicates that Markov structures are salient environmental features for shaping human behavior.

The post-experiment reports from the naïve subjects indicate that there was an explicit component to their learning of the Markov chains in our task. All of the five subjects questioned about their strategy identified most likely transitions from one reward location to another, indicating knowledge that they were guiding selections based on the previous outcome. However, it does not appear any of the subjects realized that only the reward location on the previous trial mattered. Indeed, some subjects indicated variations in their strategy based on the previous two trials (for example one subject would select unlikely reward probability on the next trial, having just viewed an unlikely transition in reward locations).

Results from the varying-reward sessions indicate that humans combine reward probabilities from a Markov chain with reward magnitude in order to increase their accumulated reward. This was most clearly demonstrated on trials in which subjects sacrificed the greatest reward probability and the greatest reward magnitude in order to select a stimulus that had the greatest expected value. The ability to combine reward probability and magnitude to calculate expected value requires a quantitative, rather than just qualitative (for example, rank ordering), representation of these values. Whereas subjects were given an accurate quantitative description of the reward magnitudes at the start of each session, subjects were not informed about reward probability—these quantities must have been learned from the Markov chain.

In our Markov task, manipulating reward probability did not have a consistent effect on saccade reaction times, which are typically used as an indication of the degree of saccade preparation (e.g. Carpenter and Williams 1995; Dorris and Munoz 1998). This is presumably because our human subjects were able to plan for their next selection as soon as feedback was provided at the end of the previous trial. We expect that manipulating target probabilities conditioned on a Markov state would have influenced saccade reaction time if presented as a reaction time task, as has been done for manual selections (Kornblum 1968, 1969).

The use of Markov chains provides an additional tool for manipulating probabilities for the purpose of understanding target selections. Although there is an explicit component to the learning that could presumably be used to guide selections by any type of response (e.g. manual selections), we have focused on saccadic eye movements, which have been extensively studied with regard to target selection (Schall and Thompson 1999). Our current Markov task, used to demonstrate learning, is most similar to free choice experiments used to demonstrate neural correlates of subjective valuation or reward probability in the lateral intraparietal area (Platt and Glimcher 1999; Sugrue et al. 2004; Dorris and Glimcher 2004; Yang and Shadlen 2007) and prefrontal cortex (Barraclough et al. 2004). However, Markov chains could be easily adapted to a reaction time task in which the Markov state would determine target probability. The use of a Markov chain would avoid confounds with target repetitions or changes in the visual scene. Some recent studies have dynamically manipulated probabilities by using symbolic cues that indicate the probability of an outcome, such as the weather prediction task (Gluck et al. 2002), which has recently been applied to non-human primates (Yang and Shadlen 2007). The use of probabilistic cues is an effective way to manipulate target probability, although the learning of the cues can take time, especially in non-human primates. In contrast, Markov chains provide a rich statistical structure that has relevance in modeling animal behavior and is quickly utilized by naïve human subjects.

In conclusion, we have demonstrated that human subjects can learn about the statistics of a Markov chain from a saccade selection task. Naïve subjects select stimuli at a frequency close to reward probability and do so within ten trials. If informed of the Markovian

structure, subjects approach a strategy that maximizes reward over hundreds of trials. In addition, humans incorporate reward magnitude into their selection process in order to select stimuli with greater expected values. Using a Markov chain to manipulate probabilities could provide a powerful tool for further study into how the brain makes use of probabilities and expected value in order to influence saccade target selection.

Acknowledgments

This study is supported by a NASA grant NAG 9-1465. We thank Garth Fowler for helpful feedback regarding the implementation of the task.

References

- Barraclough DJ, Conroy ML, Lee D. Prefrontal cortex and decision making in a mixed-strategy game. *Nat Neurosci.* 2004; 7:404–410. [PubMed: 15004564]
- Basso MA, Wurtz RH. Modulation of neuronal activity in superior colliculus by changes in target probability. *J Neurosci.* 1998; 18:7519–7534. [PubMed: 9736670]
- Brainard DH. The psychophysics toolbox. *Spat Vis.* 1997; 10:433–436. [PubMed: 9176952]
- Carpenter RHS, Williams MLL. Neural computation of log likelihood in control of saccadic eye movements. *Nature.* 1995; 377:59–62. [PubMed: 7659161]
- Cover, TM.; Thomas, JA. *Elements of information theory.* 2nd edn.. Wiley-Interscience; New York: 1991. p. 12-19.
- Dorris MC, Glimcher PW. Activity in posterior parietal cortex is correlated with the relative subjective desirability of action. *Neuron.* 2004; 44:365–378. [PubMed: 15473973]
- Dorris MC, Munoz DP. Saccadic probability influences motor preparation signals and time to saccadic initiation. *J Neurosci.* 1998; 18:7015–7026. [PubMed: 9712670]
- Dorris MC, Pare M, Munoz DP. Immediate neural plasticity shapes motor performance. *J Neurosci.* 2000; 20:RC52. [PubMed: 10627629]
- Efron, B.; Tibshirani, RJ. *An introduction to the bootstrap.* CRC Press LLC; Boca Raton: 1998. p. 168-176.
- Fecteau JH, Munoz DP. Exploring the consequences of the previous trial. *Nat Rev.* 2003; 4:1–9.
- Gluck MA, Bower GH. From conditioning to category learning: an adaptive network model. *J Exp Psychol Gen.* 1998; 117:227–247. [PubMed: 2971760]
- Gluck MA, Shohamy D, Myers C. How do people solve the “weather prediction” task?: individual variability in strategies for probabilistic category learning. *Learn Mem.* 2002; 9:408–418. [PubMed: 12464701]
- Hackley SA, Valle-Inclan F. Which stages of processing are speeded by a warning signal? *Biol Psychol.* 2003; 64:27–45. [PubMed: 14602354]
- Herbers JM. Time resources and laziness in animals. *Oecologia.* 1981; 49:252–262.
- Ikeda T, Hikosaka O. Reward-dependent gain and bias of visual responses in primate superior colliculus. *Neuron.* 2003; 39:693–700. [PubMed: 12925282]
- Kawagoe R, Takikawa Y, Hikosaka O. Expectation of reward modulates cognitive signal in the basal ganglia. *Nat Neurosci.* 1998; 1:411–416. [PubMed: 10196532]
- Knowlton BJ, Squire LR, Gluck MA. Probabilistic classification learning in amnesia. *Learn Mem.* 1994; 1:106–120. [PubMed: 10467589]
- Kornblum S. Serial-choice reaction time: inadequacies of the information hypothesis. *Science.* 1968; 159:432–434. [PubMed: 5634661]
- Kornblum S. Sequential determinants of information processing in serial and discrete choice reaction time. *Psychol Rev.* 1969; 76:113–131.
- Krauzlis RJ, Miles FA. Decreases in the latency of smooth pursuit and saccadic eye movements produced by the “gap paradigm” in the monkey. *Vis Res.* 1996; 36:1973–1985. [PubMed: 8759437]

- Macdonald, DW.; Stewart, PD.; Stopka, P.; Yamaguchi, N. Measuring the dynamics of mammalian societies: an ecologist's guide to ethological methods.. In: Boitani, L.; Fuller, TK., editors. Research techniques in animal ecology: controversies and consequences. Columbia Press; New York: 2000. p. 332-388.
- Milstein DM, Dorris MC. The influence of expected value on saccadic preparation. *J Neurosci*. 2007; 27:4810–4818. [PubMed: 17475788]
- Peters RJ, Iyer A, Itti L, Koch C. Components of bottom-up gaze allocation in natural images. *Vis Res*. 2005; 45:2397–2416. [PubMed: 15935435]
- Pelli DG. The Video Toolbox software for visual psychophysics: Transforming numbers into movies. *Spat Vis*. 1997; 10:437–442. [PubMed: 9176953]
- Platt ML, Glimcher PW. Neural correlates of decision variables in parietal cortex. *Nature*. 1999; 400:233–238. [PubMed: 10421364]
- Reinagel P, Zador AM. Natural scene statistics at the centre of gaze. *Network*. 1999; 10:341–350. [PubMed: 10695763]
- Schall JD, Thompson KG. Neural selection and control of visually guided eye movements. *Annu Rev Neurosci*. 1999; 22:241–259. [PubMed: 10202539]
- Sugrue LP, Corrado GS, Newsome WT. Matching behavior and the representation of value in the parietal cortex. *Science*. 2004; 304:1782–1787. [PubMed: 15205529]
- Tversky A, Kahneman D. Belief in the law of small numbers. *Psychol Bull*. 1971; 2:105–110.
- Vulkan N. An economist's perspective on probability matching. *J Econ Surv*. 2000; 14:101–118.
- Yang T, Shadlen MN. Probabilistic reasoning by neurons. *Nature*. 2007; 447:1075–1080. [PubMed: 17546027]
- Yarbus, AL. Eye movements and vision. Plenum Press; New York: 1967.

Fig. 1.

Description of the behavioral task. **a** A schematic for two sample trials of the saccade task. Following fixation of a white spot, subjects were given, up to 1 s, to initiate a saccade directly to the stimulus of their choice. Upon offset of the choice stimuli, subjects returned to fixation, which triggered visual feedback of the reward location and magnitude and auditory feedback indicating whether they selected the rewarded stimulus. A blank screen marked transition to the next trial. Sessions were comprised of at least 200 trials and typically lasted from 500 to 600 trials. Because stimuli were identical, subjects could only determine the next stimulus by using estimates of reward probability. **b** A sample probability distribution and Markov chain. The stimulus diagram on the *right* shows the reward probabilities of each stimulus location for a trial following a reward at location 1. The full structure of the Markov chain is given by the accompanying contingency table, which provides the reward probabilities for trials following a reward at any of the four locations. The probabilities were assigned such that each stimulus location had a reward probability of 0.25 over the course of a session

Fig. 2. Sample behavior on the Markov task. The first 200 saccade selections from two sample sessions plotted by **a** the location of the stimulus and **b** the reward probability. **c** Fraction correct was also calculated for each session by taking the fraction of rewarded selections over the previous 20 trials. The *upper dashed line* indicates the greatest reward probability, and the *lower dashed line* indicates the fraction correct expected by chance

Fig. 3.

The average time course of learning. **a** The *black line* shows the mean fraction of correct selections, averaged over all uniform-reward sessions, and binned into four trial samples. One-tailed 95% confidence intervals for each bin are provided by the *gray line*. **b** A separate plot is dedicated to the first 30 trials of the task. The fraction of correct selections is significantly greater than chance by the second bin of four trials, and remains above chance for most of the following trials. *Reference lines* are same format as Fig. 2c

Fig. 4.

Fraction of correct selections for each uniform-reward session. Sliding averages of the fraction of correct (rewarded) selections over the course of individual session sorted by subject. *Light gray traces* indicate the first session, *dark gray traces* indicate the second session, and *black traces* indicate the final session of the uniform-reward task. *Reference lines* are same format as Fig. 2c

Fig. 5.

Summary of saccade selections for uniform-reward sessions. The mean distribution of saccade selections for each naïve subject (three sessions per subject) by **a** stimulus location and **b** stimulus reward probability. The *bar plot* indicates the mean of all sessions combined. **c** The distribution of saccade selections for three subjects after being informed that reward probabilities were assigned by a Markov chain

Fig. 6.

Mutual information between selections and Markov states. The estimate for mutual information between subject selections and Markov states is given for each session, sorted by subject. Sessions are plotted in chronological order, and *dark gray bars* are given for subjects that completed a fourth session after being informed that reward probabilities were assigned by a Markov chain. *Error bars* show the 95% confidence intervals calculated by bootstrap (see “Methods”). Maximum information for a four alternative forced choice task is 2 bits. The *dashed black line* indicates the largest 95% confidence interval of the mutual information expected by chance from all of the sessions shown

Fig. 7.

Reward collected for each varying-reward session. Sliding totals of rewarded collected (in cents) over the course of individual sessions sorted by subject. Colors for each session trace were assigned chronologically like in Fig. 4. *Reference lines* are given for the expected reward accumulation by chance (*dark gray, solid*), and by selection of the greatest reward probabilities (*dark gray, long dashed*), greatest reward magnitudes (*dark gray, short dashed*), and greatest expected values (*dark gray, dotted*)

Fig. 8.

Summary of saccade selections for varying-reward sessions. Mean fraction of selections by **a** reward probability, **b** reward magnitude, and **c** rank-ordered expected value, for all Markov states. Mean fraction of selections by **d** reward probability, **e** reward magnitude, and **f** rank-ordered expected value for only the Markov state in which the stimulus with greatest expected value had neither the greatest reward probability nor the greatest reward magnitude. All plots are from a sample of 18 sessions, three for each of six subjects, *error bars* are standard error of the mean

Table 1

Saccade reaction times by reward probability

Subject	Reward probability	Mean SRT (ms)	n	ANOVA P value	Correlation
A	0.01	284	94	0.03 *	-0.64
	0.09	264	421		
	0.28	273	519		
	0.62	262	670		
B	0.01	311	52	0.002 *	-0.79 *
	0.09	295	260		
	0.28	274	552		
	0.62	278	826		
D	0.01	297	17	0.20	-0.55
	0.09	303	390		
	0.28	289	552		
	0.62	293	826		
J	0.01	326	12	0.13	-0.70
	0.09	281	284		
	0.28	268	478		
	0.62	269	923		
K	0.01	117	1	0.22	+0.52
	0.09	246	76		
	0.28	233	102		
	0.62	234	385		
T	0.01	154	5	0.49	+0.54
	0.09	184	271		
	0.28	184	604		
	0.62	182	667		

Summary of saccade reaction time (*SRT*) data. One-way ANOVAs of saccade reaction times grouped by reward probability indicate that reward probability accounts for a significant amount of variance for two subjects. Only one subject, B, had a significant correlation between reward probability and mean saccade reaction time

* Statistical significance at the 0.05 level