Modifications of Reward Expectation-Related Neuronal Activity During Learning in Primate Striatum

LÉON TREMBLAY, JEFFREY R. HOLLERMAN, AND WOLFRAM SCHULTZ
Institute of Physiology, University of Fribourg, CH-1700 Fribourg, Switzerland

Tremblay, Léon; Jeffrey R. Hollerman; and Wolfram Schultz.
Modifications of reward expectation-related neuronal activity during learning in primate striatum. J. Neurophysiol. 80: 964–977, 1998. This study investigated neuronal activity in the anterior striatum while monkeys repeatedly learned to associate new instruction stimuli with known behavioral reactions and reinforcers. In a delayed go-nogo task with several trial types, an initial picture instructed the animal to execute or withhold a reaching movement and to expect a liquid reward or not. During learning, new instruction pictures were presented, and animals guessed and performed one of the trial types according to a trial-and-error strategy. Learning of a large number of pictures resulted in a learning set in which learning took place in a few trials and correct performance exceeded 80% in the first 60–90 trials. About 200 task-related striatal neurons studied in both familiar and learning conditions showed three forms of changes during learning. Activations related to the preparation and execution of behavioral reactions and the expectation of reward were maintained in many neurons but occurred in inappropriate trial types when behavioral errors were made. The activations became appropriate for individual trial types when the animals’ behavior adjusted to the new task contingencies. In particular, reward expectation-related activations occurred initially in both rewarded and unrewarded movement trials and became subsequently restricted to rewarded trials. These changes occurred in parallel with the visible adaptation of reward expectations by the animals. The second learning change consisted in decreases of task-related activations that were either restricted to initial trials or new learning problems persisted during the subsequent consolidation phase. They probably reflected reductions in the expectation and preparation of upcoming task events, including reward. The third learning change consisted in transient or sustained increases of activations. These might reflect the increased attention accompanying learning and serve to induce synaptic changes underlying the behavioral adaptations. Both decreases and increases often induced changes in the trial selective occurrence of activations. In conclusion, neurons in anterior striatum showed changes related to adaptations or reductions of expectations in new task situations and displayed activations that might serve to induce structural changes during learning.

INTRODUCTION

Several lines of evidence postulate a role of the basal ganglia in various forms of learning. Deficits in Parkinson and Huntington patients suggest an involvement in motor learning, habit formation, and procedural memory (Butters et al. 1985; Canavan et al. 1989; Harrington et al. 1990; Knopman and Nissen 1991; Saint-Cyr et al. 1988; Vriezen and Moscovitch 1990). This view is also supported by increases in striatal blood flow during motor learning (Seitz and Roland 1992) and, by exclusion, by preserved procedural learning after temporal lobe lesions leading to declarative memory deficits (Mishkin and Appenzeller 1987; Phillips and Carr 1987). Although a specific involvement in procedural memory has been questioned (Gaffan 1996; Wise 1996), a role in motor learning would be compatible with the known motor functions of basal ganglia, including the formation of movement sequences (Graybiel 1995; Hikosaka et al. 1995). Psychopharmacological experiments suggest a role for the basal ganglia in incentive or reward-directed learning, in particular the ventral striatum and the dopamine systems (Beninger 1983; Fibiger and Phillips 1986; Robbins and Everitt 1992; Wise 1982). Further cognitive learning functions are suggested by the effects of lesions of the monkey anterior striatum, which induce deficits in spatial delayed response and alternation learning (Bättig et al. 1960; Divac et al. 1967). Striatal lesions in rats lead to learning deficits in spatial navigation (Whishaw et al. 1987) and radial arm maze tasks (Packard et al. 1989). Finally, the basal ganglia may mediate the influence of declarative memory functions of the temporal lobe on behavioral output controlled by the frontal lobe, as visual discrimination learning remains unimpaired after section of all temporal lobe projections to prefrontal cortex and diencephalon except those via the striatum (Gaffan 1996). Taken together, the basal ganglia appear to be involved in a considerable number of learning functions, rather than subserving a single learning mechanism. This appears to be in accordance with the multiple behavioral functions attributed to these nuclei.

Recent neurophysiological experiments investigated neuronal mechanisms underlying some of these learning functions. Tonically active striatal neurons (Aosaki et al. 1994) and midbrain dopamine neurons (Ljungberg et al. 1992) acquired responses during appetitive learning and may be involved in reward detection, acquisition of stimulus-response associations, and reward prediction. Neurons in the tail of caudate failed to display major changes during discrimination learning (Brown et al. 1995). However, learning-related changes were found in frontal cortical areas projecting to the head and body of striatum. Neurons in dorsolateral prefrontal cortex transiently lost their behavioral selectivity during the initial learning of new instruction stimuli in delayed response tasks (Niki et al. 1990; Watanabe 1990). Selectivity reappeared when task performance reached 85–90% correct trials. Neurons in premotor cortex and supplementary and frontal eye fields showed decreased, increased, or even new task-related activations, as well as changes in behavioral selectivity, when new instruction stimuli were introduced in conditional motor tasks (Chen and Wise 1995a,b; Mitz et al. 1991).

In the present study, we used a learning set situation in
which only a single task component changed. Animals acquired the capacity to learn the new component within a few trials, and learning could be tested repeatedly during experimental sessions (Gaffan et al. 1988; Harlow 1949). Individual neurons were studied during a whole learning episode and their activity compared with familiar performance. This appeared appropriate in view of the heterogeneity of task relationships of striatal neurons, rather than studying different neurons before, during, and after the learning of entirely new tasks. The learning set was based on a delayed go-nogo paradigm comparable with the tasks used in learning studies on cortical neurons. Depending on an initial instruction picture, animals executed an arm movement reinforced either by liquid or a conditioned sound, or they withheld the movement and were rewarded by liquid. Neurons in the anterior striatum showed several forms of activations during the initial period of recording, the 2nd 39 at the midpoint, and the 3rd 39 at the end of neuronal recordings. In A and B, trial numbers start with introduction of a new learning problem, and % indicates level of correct performance. C: average learning during neuronal recordings for familiar (top) and learning trials (bottom) in monkey A. Tick marks below curves indicate learning problems, consisting of 3 new instruction stimuli. Percent correct performance was calculated from the 1st 15 trials of each type for each block.

METHODS

The study was performed on the same two Macaca fascicularis monkeys (A and B) using the same experimental procedures with the same delayed go-nogo task as described in the preceding report (Hollerman et al. 1998). One of three colored instruction pictures was presented on a computer monitor in front of the animal for 1.0 s (13 x 13") and specifically indicated one of three trial types (rewarded movement, rewarded nonmovement, and unrewarded movement). A red trigger stimulus presented at a random 2.5–3.5 s after instruction onset required the animal to execute or withhold a reaching movement according to each trial type (13 x 13", same position as instruction). The trigger was the same in each trial type. In rewarded movement trials, the animal released a resting key and touched a small lever below the trigger to receive a small quantity of apple juice (0.15–0.20 ml) after a delay of 1.5 s. In rewarded nonmovement trials, the animal remained motionless on the resting key for 1.5 s and received the same liquid reward after a further 1.5 s. In unrewarded movement trials, the animal reacted as in rewarded movement trials, but correct performance was followed not by liquid reward but by a 1-kHz sound. The sound constituted a conditioned auditory reinforcer, because it visibly helped the animal to perform the task, but it was not an explicit reward, hence the simplifying term “unrewarded” movements. Thus each instruction was the unique stimulus in each trial indicating the behavioral reaction to be performed after the trigger (execution or withholding of movement) and predicting the type of reinforcer (liquid or sound). Correctly performed unrewarded movements were followed by one of the rewarded trials. Any incorrectly performed trial was repeated. Apart from that, the three trial types alternated semirandomly, with the number of consecutive trials of the same type restricted to three rewarded movement trials, one or two nonmovement trials, and a single unrewarded movement trial. Trials lasted 11–13 s, and intertrial intervals were 4–7 s.

In the learning situation, a new instruction picture was presented in each of the three trial types, whereas all other task events and the global task structure remained unchanged. Learning consisted in associating each new instruction picture with the execution or withholding of movement and expecting liquid reward versus the conditioned auditory reinforcer. Whereas each familiar instruction consisted of a single fractal image (Fig. 1, top), each learning instruction was composed of two partly superimposed, simple geometric shapes that were drawn randomly from 256 images (64 shapes having one of the colors red, yellow, green, or blue), resulting in a total of 65,536 possible stimuli (Fig. 1, middle and bottom). A color subtraction mode produced superimposed com-

---

**Fig. 1.** Examples of instruction pictures for the 3 trial types. *Top row:* familiar stimuli. *Middle and bottom:* stimuli for 2 learning problems. From left to right, instructions indicate rewarded movement trials, rewarded nonmovement trials, and unrewarded movement trials, respectively.
posite pictures with up to five colors on a black background of 13 × 13" on the computer monitor (CopyBits procedure of the Macintosh Operating System, transfer mode 38).

Animals were first conditioned during 4 mo to perform all three trial types with the familiar pictures at >90% correct. Then each of the three instruction stimuli was successively replaced by a novel (learning) stimulus, the other task components being unchanged. When >90% correct performance was reached with a problem of three learning stimuli, those stimuli were discarded, and three new stimuli were successively introduced until a total of ~15 new images had been learned. This stage lasted ~1.5 mo. Subsequently, new learning stimuli for all three trial types were introduced at once as a new problem, and monkeys A and B were trained for a further 3 and 1.5 mo, respectively. The learning rate increased progressively with each new problem until an asymptote was reached and a learning set established. Errors in behavioral performance led to cancellation of all further signals in a given trial, including reward. Learning was facilitated by repeating erroneous trials until correct performance was obtained.

The activity of single neurons was recorded with moveable microelectrodes from histologically reconstructed positions in anterior parts of the left caudate, putamen, and ventral striatum during contralateral task performance, together with activity from the right forearm extensor digitorum communis and biceps muscles. Every neuron was tested in 40–80 trials with the 3 familiar instructions and, in a separate block, with >60 learning trials using 3 new instruction pictures that were discarded afterward. The sequence of testing varied randomly between familiar and learning blocks. Task-related neuronal activations were assessed with the sliding window procedure based on the one-tailed Wilcoxon test (P < 0.01). Neurons not activated in familiar or learning trials are not reported. Differences in magnitudes and latencies of task-related changes between familiar and learning trials were assessed in individual neurons and groups of neurons with the two-tailed Mann-Whitney U test (P < 0.05). Movement parameters were evaluated in terms of reaction time (from trigger onset to key release), movement time (from key release to lever touch), and return time (from lever touch back to touch of resting key). They were compared on a trial-by-trial basis with the Kolmogorov-Smirnov test (P < 0.001).

RESULTS
Learning behavior

The repeated learning of new stimuli within the same task structure resulted in a learning set in which each problem of three new stimuli was learned rapidly. At the onset of neuronal recordings, monkeys A and B had learned 275 and 78 problems. Thereafter, learning was relatively stable, with an overall >80% correct performance in blocks of 60–90 learning trials. Learning occurred largely within the 1st trials and approached an asymptote within 5–10 trials of each trial type, although animals occasionally made errors during the subsequent consolidation period (Fig. 2A; see also Figs. 13, 16, and 17). Learning curves remained stable during the course of experimentation in all three trial types (Fig. 2B), although reduced learning was occasionally observed (Fig. 2C). Medians of correct performance in the 1st 15 trials in each of the 3 trial types were, respectively, 87, 100, and 93% (monkey A) and 73, 80, and 93% (monkey B). Performance in familiar trials exceeded 95% throughout the period of neuronal recording.

Trial-by-trial analysis revealed that both animals adopted a similar learning strategy. They withheld the movement in 70–90% of the first trials with new stimuli (Fig. 2A), which was well above the probability of 0.33 for this trial type. This strategy yielded a high correct performance on initial...
### Table 1. Numbers of neuronal activations influenced by learning

<table>
<thead>
<tr>
<th></th>
<th>Instruction, Following</th>
<th>Trigger Preceding</th>
<th>Trigger Following</th>
<th>Reinforcer Preceding</th>
<th>Reinforcer Following</th>
<th>Instruction, Preceding</th>
<th>Total Neurons</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Maintained activations</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Event relationships</td>
<td>40</td>
<td>40</td>
<td>52</td>
<td>61</td>
<td>51</td>
<td>20</td>
<td>150 (73)</td>
</tr>
<tr>
<td>Maintained trial selectivity</td>
<td>35</td>
<td>38</td>
<td>45</td>
<td>50</td>
<td>41</td>
<td>20</td>
<td>141 (69)</td>
</tr>
<tr>
<td><strong>Transiently decreased activations</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Event relationships</td>
<td>8</td>
<td>9</td>
<td>8</td>
<td>19</td>
<td>5</td>
<td>5</td>
<td>46 (22)</td>
</tr>
<tr>
<td>Modified trial selectivity</td>
<td>5</td>
<td>9</td>
<td>6</td>
<td>8</td>
<td>3</td>
<td>4</td>
<td>40 (20)</td>
</tr>
<tr>
<td><strong>Sustained decreased activations</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Event relationships</td>
<td>14</td>
<td>10</td>
<td>14</td>
<td>22</td>
<td>15</td>
<td>6</td>
<td>48 (23)</td>
</tr>
<tr>
<td>Modified trial selectivity</td>
<td>9</td>
<td>4</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>4</td>
<td>33 (16)</td>
</tr>
<tr>
<td><strong>Transiently increased activations</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Event relationships</td>
<td>7</td>
<td>14</td>
<td>14</td>
<td>8</td>
<td>7</td>
<td>0</td>
<td>38 (19)</td>
</tr>
<tr>
<td>Modified trial selectivity</td>
<td>5</td>
<td>6</td>
<td>10</td>
<td>4</td>
<td>4</td>
<td>0</td>
<td>28 (14)</td>
</tr>
<tr>
<td><strong>Sustained increased activations</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Event relationships</td>
<td>22</td>
<td>14</td>
<td>29</td>
<td>20</td>
<td>11</td>
<td>6</td>
<td>72 (35)</td>
</tr>
<tr>
<td>Modified trial selectivity</td>
<td>9</td>
<td>6</td>
<td>14</td>
<td>8</td>
<td>5</td>
<td>4</td>
<td>38 (19)</td>
</tr>
<tr>
<td>Total tested neurons (100%)</td>
<td>63</td>
<td>63</td>
<td>82</td>
<td>78</td>
<td>62</td>
<td>32</td>
<td>205 (100)</td>
</tr>
</tbody>
</table>

Numbers in parentheses are percentages. Activations are listed horizontally according to their temporal relationships to the three task events in familiar trials. Activations preceding instructions are listed in reference to the preceding trial type. Total numbers of neurons exceeded the numbers recorded because of multiple event relationships with differential trial selectivities, and because the same neurons could show different changes in different trial types. Transient decreases and increases are limited to the learning period until more consistent task performance was obtained, whereas sustained changes outlasted this period and often remained present throughout the entire testing period of 60–90 trials.

Nonmovement trials that remained above the chance level in subsequent trials. Both animals showed the steepest learning curves for unrewarded movement trials. Rewarded movement trials were learned slowest. This learning strategy remained rather stable throughout the course of neuronal recording.

Rewarded and unrewarded movements in familiar trials differed in both reaction time and return time, often in monkey A and always in monkey B (Fig. 3, left). This allowed identification of typical rewarded versus unrewarded movements. With new stimuli, mean reaction times in the first movement trial were intermediate between the two movement trial types. By contrast, mean return times were initially typical for rewarded movements. Both parameters became significantly different and typical for the two types of movement trial after the first few correct trials of each type (P < 0.001; Fig. 3, right). Movement times developed similar to return times in monkey B but varied inconsistently.

![FIG. 5. Adaptation of reward expectation-related sustained instruction response during learning. During familiar performance, this caudate neuron showed a sustained response in rewarded movement trials (top) and a transient response in nonmovement trials (not shown). Typically, the hand returned later to the resting key in rewarded as compared with unrewarded movement trials (return times of 958–2,539 ms vs. 404–735 ms). During learning, the sustained response occurred initially also in unrewarded movement trials, which were performed with parameters of rewarded movements (return times of 1,606–2,971 ms in trials 1–9 and 1,141–2,814 ms in trials 13–16). The response disappeared when movement parameters became typical for unrewarded movements (return times of 687–888 ms in trials 10–12 and 457–700 ms in trials 17-end; brackets to the right). Rewarded movement, nonmovement, and unrewarded movement trials alternated semirandomly during the experiment and were separated for analysis. Familiar and learning trials were performed in separate blocks. Dots in rasters denote the time of occurrence of neuronal impulses, referenced to instruction onset. Each line of dots represents 1 trial. In this and the following figures, the sequence of trials is plotted chronologically from top to bottom, learning rasters beginning with the 1st presentations of new instructions. Data in familiar and learning trials are only shown from correctly performed trials, unless mentioned otherwise.](http://jn.physiology.org/ by 10.220.33.4 on October 28, 2016)
FIG. 6. Adaptation of reward expectation-related activation following the trigger stimulus during learning. This caudate neuron was activated when a movement for liquid reward was performed, as opposed to sound reinforcement, and showed no activation in non-movement trials. During learning, activations occurred in correct rewarded movement trials and in initial unrewarded movement trials performed with parameters of rewarded movements, as judged by the return to the resting key.

FIG. 7. Adaptation of reward expectation-related activation during learning. This ventral striatum neuron was activated during familiar performance and learning in both rewarded trials irrespective of movement (A and B), but not in trials in which only sound reinforcement was given (C). During learning, activations occurred in addition in initial unrewarded movement trials that were performed with parameters of rewarded movements (F). Neuronal activations in unrewarded movement trials disappeared 3 trials before the animal performed with parameters of unrewarded movements, as judged by return of the hand to the resting key (arrows at right in F). In agreement with activations in familiar nonmovement trials (B), erroneous nonmovement reactions in rewarded movement trials were also accompanied by activations (G). Here, impulses were referenced to trigger offset as the last signal before potential reward delivery, preceding reward delivery by a constant 1.5 s. In erroneous nonmovement trials in which a movement was performed (H), the trigger disappeared already on erroneous key release, and the movement was aborted by the animal before lever touch (return to key in H). This suggests an absence of reward expectation and may explain the missing activation in H. In addition, activations in correctly performed rewarded movement trials during learning (D) were increased compared with familiar trials (A).

Overview of neuronal changes

A total of 205 slowly discharging, task-related neurons was tested in both familiar and learning trials in the anterior striatum between the anterior commissure and 7 mm rostral to it. Their behavioral relationships were characterized in two respects. First, neuronal activations were event-related in temporally preceding or following the instructions, the trigger or the reinforcers. According to the preceding report (Hollerman et al. 1998), activations preceding individual task events were probably related to the expectation of the future event or the preparation of the behavioral reaction. Activations following an event constituted a response to the event or, when following the trigger stimulus, may be related to the execution or withholding of movement. Second, neuronal activations were trial selective in occurring only in single trial types or combinations of two trial types. Activations were more frequent in trials reinforced by liquid as
opposed to sound. They further discriminated between movement and nonmovement trials. Several neurons showed multiple event relationships, each one often with different trial selectivity.

Both event relationships and trial selectivities were influenced by learning (Table 1). In many cases, neuronal activations maintained their basic event relationships and trial selectivities and adapted to new task contingencies in parallel with the behavior. In other forms, event-related activations were decreased or increased, often leading to changes in trial selectivity. Neurons with multiple task relationships occasionally showed combinations of these learning changes with different task events and trial types, resulting in considerably modified task relationships. Learning-related changes were qualitatively comparable in all 25 neurons tested in 2 blocks of learning trials, and in 3 neurons tested in 3 blocks. This was seen with all six forms of activation related to the three task events, including instruction responses in eight neurons.

**Adaptation of maintained neuronal activations**

Activations were maintained in 150 of the 205 tested neurons (73%), with at least one event relationship in at least one trial type failing to show significant differences in magnitude between familiar and learning trials (Table 1). However, activations in only 39 neurons (19%) showed unchanged magnitudes in every event relationship in every trial type. Activations in 141 of the 205 tested neurons (69%) remained selective for the same trial types as with familiar performance. Activations reflected the type of trial actually performed by the animal, which was not necessarily the trial type indicated by the instruction. They occurred in inappropriate trial types when animals performed a different type of trial than indicated by the instruction. After a few learning trials, animals adapted their behavior to the type of trial indicated, and neuronal activations adapted accordingly. Maintained trial selectivities involved activations with unchanged, decreased, or increased magnitudes. Activations following the three task events showed mean latencies and durations of 112–380 ms and 250–2,400 ms, respectively, which varied statistically insignificantly during learning.

**RELATIONSHIPS TO REWARD.** As described in the preceding report, many of the task-related neuronal activations in the anterior striatum showed pronounced relationships to the reinforcers during familiar performance (Hollerman et al. 1998). Most activations related to the instructions, trigger, or reinforcers occurred only in rewarded movement trials, in both rewarded trials irrespective of movement, or occasionally only in unrewarded movement trials. Many preinstruction activations occurred selectively after rewarded trials and probably reflected an upcoming unrewarded trial.

Many activations selective for liquid-rewarded trials during familiar performance were maintained during learning but occurred also in unrewarded movement trials. Animals indiscriminately performed most movements in initial learning trials with parameters of rewarded movements. Activations in unrewarded movement trials disappeared once the animals learned to differentiate between rewarded and unrewarded movements, or even a few trials earlier. This was observed with instruction responses (Fig. 5), pretrigger activations, activations following the trigger (Fig. 6), prereward activations (Fig. 7), and preinstruction activations (Table 2). Activations occurred also in initial unrewarded move-
ment trials when rewarded and unrewarded movements were performed with similar parameters (Fig. 8). Activations in these neurons occurred also with erroneous movements in nonmovement trials, consistent with the frequent treatment of all initial movements as rewarded. The transient occurrence of inappropriate reward-related activations in unrewarded movement trials is schematically shown in Fig. 9. Responses following the reward remained mostly restricted to trials with liquid reward.

Adaptations in the opposite direction were observed with activations occurring only in unrewarded and not in rewarded movement trials (Fig. 10). Activations were absent in initial unrewarded movement trials during learning and reappeared after a few trials, consistent with the animal’s habitual treatment of initial movement trials as rewarded. This occurred also when movement parameters were occasionally similar in rewarded and unrewarded trials. Activations were also absent with erroneous movements in nonmovement trials.

RELATIONSHIP TO BEHAVIORAL REACTIONS. Several task-related neuronal activations in anterior striatum differentiated between the execution and withholding of movement during familiar performance (Hollerman et al. 1998). This was seen in relation to all three task events by activations occurring selectively or preferentially in rewarded, unrewarded or both movement trials, or in nonmovement trials. Some preinstruction activations occurred selectively after nonmovement trials and probably reflected an upcoming movement trial.

Many activations selective for movement trials during familiar performance were maintained during learning and occurred also with erroneous behavioral reactions in inappropriate trial types. Activations occurred in correctly performed movement trials and in erroneous nonmovement trials in which a movement was made but were absent in erroneous movement trials in which no movement was made. This was observed with instruction responses (Fig. 11), pretrigger activations, activations following the trigger, and preinstruction activations (Table 1). Conversely, activations related to withholding of movement occurred in correct nonmovement trials and in erroneous movement trials in which no movement was made, but were absent in erroneous nonmovement trials in which a movement was made. All inappropriate activations disappeared after a few trials when animals learned the new instructions and reacted appropriately.

![Figure 10](http://jn.physiology.org/)) Adaptation of sustained instruction response related to the expectation of the conditioned auditory reinforcer during learning. With familiar instructions, this caudate neuron was only activated in unrewarded movement trials (top, nonmovement trials not shown). During learning, the response was initially very weak in unrewarded movement trials and became substantial only after the 3rd trial of this type (bottom right). The initial absence of response was consistent with the animal’s habitual treatment of initial movements as rewarded. However, in this instance, movements in rewarded and unrewarded trials were performed with similar parameters, in both familiar and learning situations. Only data from correctly performed trials are shown.

![Figure 11](http://jn.physiology.org/)) Adaptation of movement preparatory activity during learning. With familiar instructions, this caudate neuron showed a sustained instruction response only in rewarded movement trials and a transient response in nonmovement trials. During learning, the sustained response was slightly increased in movement trials (bottom left) and occurred also in erroneous nonmovement trials in which a movement was performed (right). The response in correct nonmovement trials resembled that in familiar trials. Same neuron as Fig. 5.

---

**Fig. 9.** Schematic overview of the development of reward expectation-related neuronal activations during learning. A: activations during the instruction-trigger interval related conjointly to movement preparation and expectation of liquid reward. Activations occurred inappropriately also in unrewarded movement trials during early learning but became restricted again to rewarded movement trials in later learning stages. B: activations during the trigger-reward interval related to expectation of liquid reward irrespective of movement. Activations occurred inappropriately also in unrewarded movement trials during early learning. In A and B, inappropriate neuronal activations during early learning occurred in parallel with inappropriate behavioral reactions, indicating that animals had not yet acquired a differential reward expectation from the new instructions. Filled bars indicate activations occurring appropriately according to trial types; open bars show inappropriate activations. Drawings are based on time courses of activations in individual striatal neurons during the instruction-trigger and trigger-reinforcement intervals, original examples of which are shown in Figs. 5 and 6–8, respectively. Below graphs, i, t, and r indicate occurrence of the instruction, trigger, and reinforcer, respectively.

FIG. 10. Adaptation of sustained instruction response related to the expectation of the conditioned auditory reinforcer during learning. With familiar instructions, this caudate neuron was only activated in unrewarded movement trials (top, nonmovement trials not shown). During learning, the response was initially very weak in unrewarded movement trials and became substantial only after the 3rd trial of this type (bottom right). The initial absence of response was consistent with the animal’s habitual treatment of initial movements as rewarded. However, in this instance, movements in rewarded and unrewarded trials were performed with similar parameters, in both familiar and learning situations. Only data from correctly performed trials are shown.

FIG. 11. Adaptation of movement preparatory activity during learning. With familiar instructions, this caudate neuron showed a sustained instruction response only in rewarded movement trials and a transient response in nonmovement trials. During learning, the sustained response was slightly increased in movement trials (bottom left) and occurred also in erroneous nonmovement trials in which a movement was performed (right). The response in correct nonmovement trials resembled that in familiar trials. Same neuron as Fig. 5.
STRIATAL NEURONS DURING LEARNING

FIG. 12. Anatomic positions of neurons in both monkeys showing maintained relationships to behavior and reinforcers during learning. Positions of neurons are labeled according to the 6 possible task relationships, namely responses to and activations preceding instruction, trigger, and reinforcers. Interrupted lines within sections indicate approximate borders of striatal regions (Cd, caudate; Put, putamen; V St, ventral striatum; AC, anterior commissure). Data from both monkeys are plotted on coronal sections from the left hemisphere of one monkey, which are labeled according to the distances from the interaural line (A18–A25).

RECORDING POSITIONS. Neurons with maintained relationships during learning were distributed throughout the entire recording area in caudate nucleus (n = 71), putamen (n = 46), and ventral striatum (n = 24), without local preferences for any form of task relationship (Fig. 12).

Decreased neuronal activations

A total of 90 of the 205 tested neurons (44%) showed significantly decreased magnitudes of at least one event-related activation in at least one trial type. In 46 of these neurons, the decreases occurred transiently during initial learning trials until more consistent task performance was obtained. In 48 neurons, decreases were sustained in outlasting this period and often remaining present throughout the entire testing period of 60–90 trials (Table 1). Combinations between transient and sustained decreases occurred occasionally with different task events or trial types. Transient or sustained decreases resulted in the complete absence of activation in some trial types in 40 and 33 of these neurons, respectively. This abolished entirely the responsiveness in neurons activated in a single familiar trial type and increased the trial selectivity of activations in neurons activated in several familiar trial types. Neurons with decreased activations during learning were unpreferentially distributed over the entire recording area in caudate nucleus (n = 48), putamen (n = 22), and ventral striatum (n = 20).

Transient decreases during the initial two or three learning trials were seen with instruction responses, pretrigger activations (Fig. 13A), activations following the trigger (Fig. 13B), prereward activations (Fig. 14), reinforcer responses, and preinstruction activations. Sustained decreases outlasting the initial learning trials are shown in Fig. 15. During familiar trials, this neuron showed a combination of activations following the instructions, trigger stimulus, and liquid reward in both rewarded trials irrespective of movement. During learning, the instruction

FIG. 13. Transiently decreased activations in 2 caudate neurons during learning. A: during initial learning, decrease of the selective pretrigger activation in rewarded movement trials. B: decrease of postrigger activation during initial learning trials with gradual buildup during subsequent trials. The postrigger activation occurred in all 3 trial types during both familiar performance and learning. Correctly and incorrectly performed trials are indicated by plus sign and minus sign, respectively. Only rewarded movement trials are shown in A and B.
FIG. 14. Transiently decreased reward expectation-related activations during learning. In familiar trials, this putamen neuron was activated preceding liquid reward in rewarded movement and nonmovement trials, but not in unrewarded movement trials. During learning, the activations were initially reduced in both trial types and subsequently attained the level of familiar performance. Only data from correctly performed trials are shown.

Examples of newly appearing activations are shown in Fig. 16. The posttrigger activation of the neuron in A was restricted to the initial learning trials during which the animal made several errors. The posttrigger activation in B decreased gradually after the initial learning trials and disappeared completely before the end of the habitual testing period, although the animal happened to perform every trial correctly during learning. The newly appearing instruction response in C lasted during the entire testing period during learning, again without error performance. Figure 17 shows an increased instruction response in rewarded movement trials together with a newly appearing response in nonmovement trials that decreased the trial selectivity. The animal made several errors in both trial types that decreased only gradually during learning. Examples of increases of reward-related activations are shown in Fig. 18. Prereward activations showed transient (A) or sustained (B) increases. Similar increases were seen with reward responses (C).

DISCUSSION

These data show that task-related neuronal activations in the anterior striatum undergo changes when animals rapidly adapt their behavioral reactions to new instruction stimuli. In the first form of neuronal change during learning, the behavioral adaptations were accompanied by adapting neu-
FIG. 16. Appearance of activations with different durations in 3 caudate neurons during learning. A: transient appearance of activation following the trigger stimulus during learning in rewarded movement trials. A postrigger activation occurred also in unrewarded movement trials and was maintained during learning (not shown). B: sustained appearance of selective postrigger activation in initial nonmovement trials during learning, with gradual disappearance during subsequent trials. C: sustained appearance of instruction response in rewarded movement trials during learning. This neuron also showed a new instruction response in unrewarded movement trials during learning. Correctly and incorrectly performed trials are indicated by plus sign and minus sign, respectively.

neuronal activations that otherwise maintained their relationships to behavioral reactions and upcoming reinforcers. In the second form, neuronal activations were decreased or abolished. In the third form, neurons were more strongly or exclusively activated during learning. Together, these data suggest an involvement of the striatum in associating new environmental signals with known behavioral reactions and reinforcers.

FIG. 17. Sustained increase and new appearance of instruction response in a caudate neuron during learning. With familiar instruction stimuli, the neuron responded almost exclusively in rewarded movement trials. During learning, the response was increased in rewarded movement trials. In nonmovement trials, a new response appeared and became gradually weaker during the course of learning. In unrewarded movement trials, a response appeared also but remained present during the testing period (not shown). The appearance of substantial responses in other than rewarded movement trials constituted a reduction in trial selectivity. Correctly and incorrectly performed trials are indicated by plus sign and minus sign, respectively.
FIG. 18. Increases and new appearances of reward-related activations in 3 caudate neurons during learning. A: transient appearance of reward expectation-related activation in rewarded movement trials. This neuron also showed reward expectation-related activations in nonmovement trials during familiar performance and learning (not shown). B: sustained, massive increase of reward expectation-related activation in nonmovement trials. This neuron showed a similar activation increase in rewarded movement trials, and in initial unrewarded movement trials performed with parameters of rewarded movements (not shown). C: sustained appearance of reward response in nonmovement trials. This neuron failed to respond to liquid reward or sound reinforcer in any other trial type during familiar performance or learning. Only data from correctly performed trials are shown.

Learning behavior

In most learning situations, only a part of the environment changes, and learning is confined to the modified contingencies. Repeated learning in an unchanged context leads to a learning set in which new stimuli are fully learned in a few trials (Gaffan et al. 1988; Harlow 1949). The rapid learning is particularly appropriate for the limited periods of recordings from single neurons. The present learning paradigm was based on a conditional, delayed go-no-go task involving associations between visual stimuli and behavioral reactions without explicit sensorimotor guidance. Delay tasks typically test preparatory and expectation-related functions of the striatum (Alexander and Crutcher 1990; Apicella et al. 1992; Divac et al. 1967; Hikosaka et al. 1989). Similar conditional delay tasks were used in learning set situations for investigating neurons in prefrontal and premotor cortex (Mitz et al. 1991; Niki et al. 1990; Watanabe 1990) and supplementary and frontal eye fields (Chen and Wise 1995a,b, 1996).

The present experiment comprised three learning stages. Animals first learned the basic task, discriminating on the basis of different instructions between execution and withholding of movement, and between liquid reward and conditioned sound reinforcement. The second stage involved the acquisition of a learning set. New instruction stimuli were repeatedly introduced while keeping the remaining task components unchanged, and animals acquired the capacity to associate new instructions with known behavioral reactions and reinforcers within a few trials. The third stage comprised stable performance in the learning set, and neuronal recordings were done during early learning and subsequent task consolidation. Animals applied a trial-and-error learning strategy with a win-stay lose-shift rule. They usually showed a nonmovement reaction to the first instruction of a new problem, apparently economizing the effort. They subsequently learned unrewarded movements, thus avoiding to rerun error trials and further postpone the reward. In this way, learning occurred rapidly, involved only changes in the instruction as crucial task component, was based on a previously learned, unchanged background situation, and was stationary over successive learning episodes.

Initial unrewarded movements were often performed with parameters and muscle activations typical for rewarded movements, as judged by the return movement from the touch lever to the resting key. This suggests that animals initially expected a reward with all movements and only after a few trials distinguished between rewarded and unrewarded movements. The well-organized behavioral strategy suggests that animals guessed the significance of novel instructions and accordingly performed the three trial types with nearly the same parameters as familiar trials of corresponding types. Apparently, animals had an existing set of expectations corresponding to their previous experience in the familiar task in which only one of three trial types was possible. When confronted with a novel instruction, they could draw on these expectations to test the novel stimulus and later adapt their expectations and behavior to the new situation. Inappropriate or “erroneous” behavioral performance would thus reflect inappropriate but otherwise unchanged expectations evoked by the new instruction. The simplicity of evoking existing expectations and matching them to the experienced meaning of new instructions may explain the remarkable speed of learning.

Adaptation of maintained activations

Many neurons in the anterior striatum are activated in relation to the preparation of movements and the expectation of individual task events, including reward (Alexander and Crutcher 1990; Apicella et al. 1992; Hikosaka et al. 1989; Hollerman et al. 1998; Schultz et al. 1992). Apparently, these neurons have access to predictive information stored during previous task experience. In the present experiments, neuronal activity related to movement preparation during the instruction-trigger interval maintained the behavioral relationship and reflected the actually executed behavioral reaction, which did not necessarily correspond to the trial type indicated by the instruction. Thus erroneously performed movements in nonmovement trials were accompanied by inappropriate but otherwise maintained movement-related neuronal activations. Also, the initial “default” expectation of reward with all movements was frequently paralleled by
reward expectation-related neuronal activity, which in later trials became restricted to rewarded movements. This occurred in all reward expectation-related activations preceding or following the instruction, trigger, and reward. Correspondingly, activations reflecting the expectation of the auditory reinforcer were rare in initial learning trials and reappeared subsequently. It appears that initially inappropriate neuronal activations reflected inappropriate expectations evoked by instructions of guessed type. These activations adapted and became appropriate during the course of learning when expectations were matched to the new task contingencies and evoked appropriately. Interestingly, neuronal activations adapted a few trials before the behavioral reactions, which was also observed in prefrontal neurons (Watanabe 1990). Taken together, these data suggest a mechanism for adaptive learning in which existing expectation-related neuronal activity is simply matched to the new conditions, rather than acquiring all task contingencies from scratch.

Comparable learning changes were found in several regions of frontal cortex projecting to anterior striatum. In a similar go-nogo learning set paradigm as employed presently, some dorsolateral prefrontal neurons maintained their movement preparatory activity in close correspondence with the actually executed behavioral response (Niki et al. 1990; Watanabe 1990). Neurons in premotor cortex, supplementary eye field, and orbitofrontal cortex also showed maintained task relationships, albeit with lower incidence than in anterior striatum (Chen and Wise 1995a,b, 1996; Mitz et al. 1991; Tremblay and Schultz 1996). Orbitofrontal cortical neurons changed their activity in parallel with behavioral changes during reversal of visual stimuli (Rolls et al. 1996; Thorpe et al. 1983). Interestingly, the inappropriate coding of sample stimuli by hippocampal place cells was suggested to play a role in bringing about inappropriate behavioral reactions (Deadwyler et al. 1996). By contrast, task-specific expectations would not exist when a new task without preconceived structure is learned by a naive animal. Prefrontal neurons began to show expectation- and preparation-related activations during the learning of spatial delayed response tasks as soon as animals attained substantial performance (Germain and Lamarre 1993; Kubota and Komatsu 1985). Similarly, most tonically active neurons in striatum acquired responses to reward-predicting stimuli only after conditioning (Aosaki et al. 1994). Thus inputs from frontal cortex could mediate some of the maintained activations presently observed during learning set performance in anterior striatum, whereas the contributions of cortical changes during the acquisition of new tasks are difficult to evaluate.

**Decreased neuronal activations**

A sizable fraction of the investigated striatal neurons showed decreased event-related activations during learning compared with familiar performance. Decreases occurred in relation to any of the events and often resulted in the complete disappearance of activations in particular trial types, thus making the neurons either entirely unresponsive or increasingly task selective. The shorter lasting decreases resembled the “learning-dependent” changes that occurred transiently in premotor cortex and supplementary eye field during learning and subsequently recovered to levels of familiar performance (Chen and Wise 1995a; Mitz et al. 1991). Similar changes were also seen in prefrontal (Niki et al. 1990) and hippocampal neurons (Cahusac et al. 1993). The current longer lasting decreases resemble the “learning-static” decreases outlasting the learning phase in frontal eye field and supplementary eye field neurons (Chen and Wise 1995b).

About one-half of the presently observed decreases lasted only during the few learning trials. Apparently, some striatal neurons lost their relationship to the preparation of movement and expectation of reward during a transient period of uncertainty with the incompletely established task. The activations regained their magnitudes when events became better predicted with firmly established stimulus-response associations and reestablished behavioral performance. Other decreases outlasted the learning phase. Whereas the familiar instructions consisted of the same, highly structured, fractal images that were used for many months, the learning pictures were simple geometric forms that were usually presented for <100 trials and then replaced by new stimuli. The animals conceivably differentiated between the class of familiar stimuli and the class of novel, “disposable” stimuli. Thus the long-lasting decreases during learning may have reflected the transient nature of frequently changing associations.

**Increased neuronal activations**

Several striatal neurons showed transient or sustained increases of existing event-related activations in learning trials. Activations appearing in additional trial types during learning induced decreases in trial selectivity. These striatal neurons apparently processed unidentified instructions during learning with less selectivity and thus broadened their potential coding of task events. The transient nature of increases in several striatal neurons suggests that a comparable trial selectivity was regained after the significances of instructions were acquired.

The quantitative increases resembled augmented visual responses in the tail of caudate during learning (Brown et al. 1995) and might correspond to rapidly habituating responses to new visual stimuli in the same structure (Caan et al. 1984). The activations occurring exclusively during learning were similar to “learning-selective” activations in the supplementary eye field (Chen and Wise 1995a) and comparable changes in hippocampus (Cahusac et al. 1993). Such pronounced increases during learning are compatible with decreases in movement-related activations in the supplementary motor cortex with overtrained lever pressing (Aizawa et al. 1991). The present increases outlasting the learning phase resembled the learning-static increases in frontal and supplementary eye fields (Chen and Wise 1995b). Decreases in trial selectivity were also observed in prefrontal and orbitofrontal cortex (Tremblay and Schultz 1996; Watanabe 1990) and were compatible with the modification or breakdown of directional selectivity in the supplementary eye field (Chen and Wise 1996). Taken together, frontal cortical areas projecting to the striatum show learning-related changes that may contribute to the increased activations found in the anterior striatum.

It is conceivable that the activation increases during learn-
ing might simply reflect differences in visual stimulus features, in particular with instructions. However, many learning changes occurred only during initial learning trials, whereas the instruction pictures remained the same during a whole block of learning trials. Although some activation increases and changes in trial selectivity lasted during whole learning blocks, they remained comparable when different sets of learning pictures with different visual features were tested. Increases outlasting initial learning trials were also seen with activations following the trigger, although this stimulus was the same as in familiar trials. Taken together, many of the observed learning changes appeared to be unrelated to differences in visual features. By contrast, neuronal responses related to visual features were found in the tail of caudate (Brown et al. 1995).

In contrast to visual stimulus features, attention-related mechanisms may well have contributed to increased activations. Most task events were probably processed with increased levels of attention during learning. The new instruction stimuli were novel and had future behavioral significance, and the other task events were processed in a less automatic manner. The occurrence of environmental stimuli or behavioral reactions during heightened foveal or extrafoveal selective attention is accompanied by enhanced neuronal activations in cortical areas, in particular the parietal cortex (Bushnell et al. 1981; Mountcastle et al. 1981; Sakata et al. 1983). Inputs from these areas to the striatum might account, at least in part, for the observed learning-related increases in activation.

The increased task-related activations during learning might influence the storage of new stimulus-response associations in memory. Associating new instructions with known behavioral reactions and reinforcers would require modifications of neuronal processing, most probably involving changes at the synaptic level. These changes should take place between the introduction and acquisition of new instructions, and during the subsequent consolidation phase. Neuronal activations were presently found to be increased during both periods. According to the notion of Hebbian learning, synaptic transmission may undergo long-lasting changes following repeated and conjoint pre- and postsynaptic activations. The presently reported increased activations of striatal neurons most likely reflected increased cortical input and thus increased use of corticostriatal synapses. This may induce synaptic changes at corticostriatal synapses, which are known to undergo long-term potentiation (Boeijinga et al. 1993; Calabresi et al. 1992; Pennartz et al. 1993; Wickens et al. 1996). It could be speculated that transient increases induce synaptic changes at corticostriatal synapses and that more sustained increases participate in the consolidation of these synaptic modifications.

**Comparison with responses of dopamine neurons**

Midbrain dopamine neurons respond to primary rewards and to conditioned, reward-predicting stimuli in a manner compatible with the concept of an error in the prediction of reward (Ljungberg et al. 1992; Mirenowicz and Schultz 1994; Schultz et al. 1993, 1997). They are activated by unpredicted rewards, are not influenced by fully predicted rewards, and are depressed when predicted rewards are omitted. Prediction errors crucially determine the rate of associative learning (Dickinson 1980; Rescorla and Wagner 1972). It appears that the rather homogeneous and simultaneous responses of dopamine neurons may constitute a teaching signal that is propagated as a global reinforcement message along diverging axons to neurons in the striatum and frontal cortex. By contrast, many of the presently found changes of striatal neurons were related to adaptations of previously acquired expectations to new task situations, or of reductions of such expectations. This would correspond to the striatal involvement in expectation- and preparation-related cognitive processes. These adaptive changes would not constitute a reward prediction signal suitable for modifying synaptic weights. However, the increased striatal activations during learning might contribute to synaptic modifications, although their heterogeneous nature and the different architecture of axonal projections would not allow them to serve as a global reinforcement message.

We thank D. Ballard, A. Dickinson, and M. Watanabe for discussions and comments and B. Aebischer, J. Corpataux, A. Gaillard, A. Pisani, A. Schwarz, and F. Tinguely for expert technical assistance.

This study was supported by Swiss National Science Foundation Grants 31-21499.90, 31.4331.95, and NFP38.4038-43997, the Roche Research Foundation, Switzerland, and by postdoctoral fellowships from the Fondation pour la Recherche Scientifique de Quebec to L. Tremblay and National Institute of Mental Health Grant MH-10282 to J. R. Hollerman.

Present addresses: L. Tremblay, INSERM Unit 289, Hôpital de la Salpetrière, 47 Boulevard de l’Hôpital, F-75651 Paris, France; and J. Hollerman, Dept. of Psychology, Allegheny College, Meadville, PA 16335.

Received 11 February; accepted in final form 9 April, 1998.

**REFERENCES**


