

Evidence of Action Sequence Chunking in Goal-Directed Instrumental Conditioning and Its Dependence on the Dorsomedial Prefrontal Cortex

Sean B. Ostlund,¹ Neil E. Winterbauer,³ and Bernard W. Balleine^{2,4}

¹Semel Institute for Neuroscience and Human Behavior and ²Department of Psychology and Brain Research Institute, University of California, Los Angeles, Los Angeles, California 90095, ³Department of Psychology, University of Vermont, Burlington, Vermont 05405, and ⁴Brain & Mind Research Institute, University of Sydney, Sydney, New South Wales 2006, Australia

The current study investigated the contribution of the dorsomedial prefrontal cortex (dmPFC) to instrumental action selection. We found that cell body lesions of the dmPFC, centered on the medial agranular area, spared rats' ability to choose between actions based on either the value or the discriminative stimulus properties of an outcome. We next examined the effects of these lesions on action sequence learning using a concurrent bidirectional heterogeneous chain task in which the identity of the reward delivered was determined by the order in which the two lever press actions were performed. Although both lesioned rats and sham controls learned to perform the task, we found that they relied on different behavioral strategies to do so. In subsequent tests, rats in the sham group were able to withhold their performance of a sequence when either its associated outcome was devalued or the contingency between that sequence and its outcome was degraded by delivering the outcome noncontingently. Interestingly, lesioned rats failed to reorganize their performance at the action sequence level and, rather, were found to withhold their performance of the terminal response in the sequence that had earned the devalued outcome relative to the more distal response, suggesting that they represented the elements of the sequence as distinct behavioral units. These findings demonstrate that rats can use sequence-level representations, or action chunks, to organize their behavior in a goal-directed manner and indicate that the dmPFC plays a critical role in this process.

Introduction

Rats are capable of choosing instrumental actions in a goal-directed manner. This capacity is demonstrated by the sensitivity of instrumental performance to manipulations of outcome value and action–outcome contingency (Dickinson and Balleine, 1994). Although such manipulations are commonly used to distinguish between goal-directed and habitual performance, they may also be applied more generally to assay action encoding and discrimination, which may be particularly useful for determining how actions are being represented in situations in which this is ambiguous. For example, obtaining a goal often involves performing a sequence of two or more discrete actions. Although rats and other animals can be trained to perform action sequences of this kind, there is considerable debate about what is learned in such tasks (Terrace, 2005). One possibility is that the entire sequence becomes represented (perhaps through an action chunking process) as an integrated unit of behavior capable of entering into associations with discriminative cues and/or rewards (Lashley, 1951; Miller et al., 1960; Rosenbaum et al., 1983). Alternately,

the elements of the sequence may be represented separately, allowing individual responses to become associated with only the most proximal events. This account seems particularly plausible for situations in which the elements of the sequence are signaled by unique discriminative stimuli, allowing subjects to learn about the various steps needed to perform the sequence without having to encode the sequence as a whole.

We assessed the content of action sequence learning in an unsignaled, free-operant task designed to encourage the development of sequence-level action representations. Rats were reinforced for performing a sequence of two different lever press responses, with response order determining which of two outcomes would be delivered (R1→R2→sucrose pellets and R2→R1→Polycose solution). Therefore, the content of action–outcome encoding for this task depends on how subjects represent their behavior: the use of a response-based strategy should result in each lever press response becoming differentially associated with the most proximal outcome (R2→sucrose and R1→Polycose), whereas the use of a sequence-based strategy predicts that each two-action sequence will become associated with the outcome that it delivers ([R1–R2]→sucrose and [R2–R1]→Polycose). To evaluate these accounts, rats were administered outcome devaluation and contingency degradation tests after training to determine whether these treatments would cause subjects to reorganize their performance at the level of individual responses or entire action sequences.

Action sequence performance in primates appears to be me-

Received March 10, 2009; revised April 29, 2009; accepted May 8, 2009.

This work was supported by Grant HD59257 from the National Institute of Child Health and Human Development to B.W.B. and National Institute of Mental Health Training Fellowship T32 MH17140 to S.B.O.

Correspondence should be addressed to Sean B. Ostlund, Department of Psychology and Brain Research Institute, Franz Hall, Box 951563, University of California, Los Angeles, Los Angeles, CA 90095-1563. E-mail: sostlund@ucla.edu.

DOI:10.1523/JNEUROSCI.1176-09.2009

Copyright © 2009 Society for Neuroscience 0270-6474/09/298280-08\$15.00/0

diated by the supplementary motor area (SMA) (Tanji, 2001; Hoshi and Tanji, 2004; Kennerley et al., 2004; Rushworth et al., 2004). The rodent medial agranular cortex has anatomical and functional similarities with the primate SMA (Donoghue and Wise, 1982; Passingham et al., 1988), and has recently been implicated in action sequence performance (Bailey and Mair, 2007). The current study investigated whether dorsomedial prefrontal cortex (dmPFC) lesions targeting the medial agranular cortex would disrupt rats' ability to use sequence-level action representations to modify their instrumental performance in a goal-directed manner. An initial experiment assessed the effects of dmPFC lesions on action selection using a nonsequential task to characterize the involvement of this structure in instrumental learning and performance more generally.

Materials and Methods

Subjects and apparatus. Twenty-nine adult female Long–Evans rats (Harlan) served as subjects. Rats were group housed in transparent plastic tubs in a temperature-controlled vivarium. Training and testing were conducted during the light phase of the 12 h light/dark cycle. Rats were food deprived throughout training and testing by restricting their daily food allotment to 10–12 g of home chow, sufficient to maintain them at ~85% of their free-feeding weight. Tap water was continuously available when rats were in their home cages. Behavioral training and testing were performed in 16 identical Med Associates operant chambers enclosed in sound- and light-attenuating shells. Each chamber had two retractable levers that could be inserted to the left and right of a recessed food magazine. Four distinctive outcomes (grain pellets, sucrose pellets, sucrose solution, and Polycose solution) could be delivered into the food magazine. The pellet outcomes (45 mg; Bioserv) were delivered via separate dispensers. The two liquid outcomes (20% sucrose or 20% Polycose with 0.9% NaCl; mixed in tap water) were delivered in 0.1 ml aliquots via separate syringe pump systems. An infrared photobeam crossed the magazine opening, allowing for the detection of head entries. Illumination was provided by a house light (3 W, 24 V) located on the wall opposite the magazine. A set of two microcomputers running the Med-PC program (Med Associates) controlled all experimental events and recorded responses. For specific satiety sessions, rats were singly housed in plastic tubs identical to their home cages. Graduated glass drinking bottles were used to provide access to the fluid outcomes (at least 50 ml) and small glass bowls were used for pellet outcomes (at least 50 g).

Surgery. Rats undergoing surgery were given *ad libitum* access to food for at least 1 d before and 7 d after surgery. Immediately before surgery, rats were anesthetized with sodium pentobarbital (Nembutal, 50 mg/kg) and administered atropine sulfate (0.1 mg). During surgery, rats were placed in a stereotaxic frame (Stoelting) and had their scalp incised and retracted to expose the skull surface. The incisor bar was then adjusted to horizontally align bregma and lambda. Bilateral burr holes were made above the eight target sites: anteroposterior, +3.5, +2.2, +0.9, –0.4; lateral, ±1.3, ±1.0, ±1.0, ±0.7; ventral, –4.0, –3.7, –3.7, –3.2 (all coordinates in mm relative to bregma). For rats receiving excitotoxic lesions, 0.25 μ l of NMDA (20 μ g/ μ l in PBS) was infused at a rate of 0.1 μ l per minute into each site using a 1 μ l Hamilton syringe. The needle was left in place for an additional 2 min to allow for diffusion of the drug. Sham lesions were made using the same basic procedure except that the needle was not lowered and no infusions were made. Rats were given at least 10 d to recover before undergoing behavioral training/testing. Fifteen rats underwent surgery before initial training (pre-sham, $n = 7$; pre-dmPFC, $n = 8$) and 14 underwent surgery between training and testing (post-sham, $n = 6$; post-dmPFC, $n = 8$).

Initial instrumental training. The rats were first given two sessions of magazine training. Each session consisted of 20 deliveries of grain pellets and 20 deliveries of the sucrose solution presented in random order according to a random time (RT), 60 s schedule. They were then given 11 d of instrumental training. Each rat received two training sessions per day, one with the left lever and one with the right lever. Each session ended after 20 outcomes were earned or 30 min had elapsed, whichever

came first. The sessions were separated by 10–15 min and session order was alternated over days. Half of the rats in each group earned grain pellets by performing the left response and sucrose solution by performing the right response, whereas the remaining rats were trained with the opposite response–outcome relationships. A continuous reinforcement schedule was used for days 1–2. The schedule was then shifted to random ratio (RR)-5 for days 3–5, then to RR-10 for days 6–8, and finally to RR-20 for days 9–11. Post-training surgeries were conducted on the following day.

Outcome devaluation testing. We then conducted two separate outcome devaluation tests, one with the grain pellet devalued and one with the sucrose solution devalued. For each test, rats were given 1 h of unlimited access to one of the two training outcomes. Half of the rats in each group were sated on grain pellets and half were sated on sucrose solution. Immediately after the prefeeding period they were placed in the operant chambers for a 5 min choice extinction session in which both levers were available but were inactive. Forty-eight hours later, rats were given a second devaluation test using the same procedure, except that they were sated on the opposite outcome.

Reinstatement testing. An outcome-selective instrumental reinstatement test was conducted 48 h after the second outcome devaluation test. Both levers were available throughout the session but were inactive. The session began with a 15 min period of extinction to lower the rats' rate of responding on both levers. They then received four reinstatement trials separated by 7 min each. Each reinstatement trial consisted of a single delivery of either the sucrose solution or the grain pellet. All rats received the same trial order: sucrose, pellet, pellet, sucrose. Responding was measured during 2 min periods immediately before (Pre) and after (Post) each outcome delivery.

R1–R2 training. Rats were then trained to perform a two-action sequence (R1→R2; a left lever press followed by a right lever press) for sucrose pellets. Each training session lasted until 20 outcomes had been earned or 1 h had elapsed, whichever came first. Both levers were continuously available throughout the session and no discriminative cues were used to guide action selection. Because of their previous random ratio training, the rats acquired the tendency to make a burst of presses on one of the two levers before checking the food magazine, and made very few direct transitions between the left and right lever without first checking the food magazine. Therefore, we began sequence training with a relaxed response requirement. A sucrose pellet was delivered whenever a rat performed a new left-right sequence, regardless of whether or not they checked the magazine between presses. We then shifted to a stricter response requirement for the next eight sessions; the outcome was delivered only after the rat performed a new left-right sequence that was uninterrupted by a magazine approach response. Other action sequences (right-right, left-left, right-left) were not reinforced.

R2–R1 training. Rats were then given six sessions of training to perform the reversed two-action sequence (R2→R1; a right lever press followed by a left lever press) for Polycose solution. Reversal training sessions were otherwise identical to initial sequence training, except that the relaxed response requirement was not used; i.e., rats were only reinforced for performing right-left sequences that were uninterrupted by magazine checking.

Concurrent sequence training. Rats were then given four sessions of training with both sequence contingencies in place. Both levers were continuously available throughout the session and no discriminative cues were presented to guide action selection. As before, the left-right sequence was reinforced with sucrose pellets and the right-left sequence was reinforced with Polycose. The session lasted until 20 sucrose pellets and 20 Polycose aliquots were earned or 1 h had elapsed, whichever came first. This training procedure allowed rats to control the order of outcome deliveries. Therefore, after reaching the limit for a given outcome type, the corresponding action sequence was no longer reinforced. However, the time that remained in the session could be used to earn the other, nondepleted outcome.

Outcome devaluation test for action sequences. Two separate outcome devaluation tests were conducted using the same general procedure described above. Before the first test, rats were prefed for 1 h on either sucrose pellets or Polycose solution. They were then immediately placed in the chamber for a 5 min choice extinction test. Forty-eight hours later,

the rats were given a second test using the same procedure except that they were sated on the other outcome.

Instrumental contingency degradation with action sequences. Rats were then given a session of instrumental contingency degradation training in which one of the two sequence–outcome relationships was weakened by delivering its corresponding outcome noncontingently. Both levers were continuously available for the entire 20 min session. As during concurrent sequence training, the left-right sequence earned sucrose pellets and the right-left sequence earned Polycose. At test, however, these sequences were reinforced according to a common random interval-30 s schedule, such that the first sequence (either R1–R2 or R2–R1) completed after the interval expired earned the corresponding outcome. This interval schedule was the only constraint on the number of outcomes a subject could earn in the session. One of the two training outcomes was also delivered noncontingently according to an RT-30 s schedule. Half the rats in each group received noncontingent deliveries of sucrose pellets and the other half received noncontingent Polycose.

Histological analysis. Rats were given an overdose of Nembutal (1.5 ml per rat), and their brains were extracted and postfixed in a 30% sucrose-formalin solution for 48 h. The brains were then frozen and 50 μ m coronal sections were collected from the frontal cortex, mounted on glass slides, and stained with thionin. A light microscope was used to evaluate lesion placement and extent of damage through comparison with sections from sham-lesioned rats and a rat brain atlas (Paxinos and Watson, 1998).

Dependent measures. In experiment 1, the data were plotted and analyzed as the number of lever presses performed per minute. In experiment 2, two measures of action sequence performance were used: (1) the relative frequency of each sequence, calculated as a percentage of all two-action sequences performed (e.g., $[R1-R2/(R1-R2 + R1-R1 + R2-R1 + R2-R2)] \times 100$), and (2) the number of presses performed per minute on the individual levers (R1 and R2).

Results

Histology

The NMDA infusions produced substantial cell loss throughout the medial agranular cortex (M2 using the nomenclature of Paxinos and Watson, 1998) and the dorsal region of the anterior cingulate (Cg1) in all lesioned rats ($n = 16$), although the most rostral and caudal aspects of these structures were spared (Fig. 1). For some rats, moderate cell loss was also observed in surrounding cortical areas, including the dorsomedial region of the prelimbic cortex ($n = 3$), the ventral region of the anterior cingulate (Cg2; $n = 4$), and the most medial region of the lateral agranular cortex (M1; $n = 4$), although such damage was often unilateral. Pretraining and post-training lesion groups had similarly sized and placed lesions.

Experiment 1: effects of dmPFC lesions on nonsequential instrumental conditioning

Acquisition of lever pressing

Rats with dmPFC lesions responded at a significantly lower rate during training than either sham-lesioned rats or unoperated rats (i.e., the rats that would undergo surgery after training) (Fig. 2B). A mixed factors group by day ANOVA found a significant main effect of group ($F_{(2,26)} = 3.65$; $p < 0.05$) and day ($F_{(10,260)} = 174.19$; $p < 0.001$), as well as a significant group by day interaction ($F_{(20,260)} = 2.71$; $p < 0.001$). Further analysis revealed that the effect of group only reached significance for sessions 10

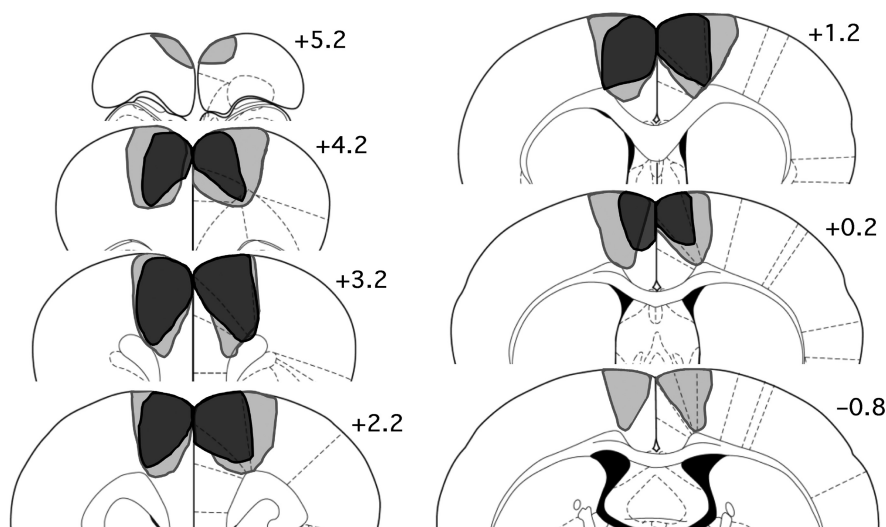


Figure 1. Schematic representations of the smallest (light gray) and largest (dark gray) dmPFC lesions. Numbers refer to distance (mm) relative to bregma.

($F_{(2,26)} = 3.63$; $p < 0.05$) and 11 ($F_{(2,26)} = 6.11$; $p < 0.01$), indicating that the effect of dmPFC lesions on response rate was restricted to the final sessions instrumental training, when responding was reinforced according to a RR-20 schedule, which required subjects to exert more effort to obtain reward, relative to schedules used early in training. Although this finding could be attributable to increased sensitivity to response demand in the dmPFC group, consistent with previous reports that this general area is important for effort-based decision making (Walton et al., 2002; Floresco and Ghods-Sharifi, 2007), this interpretation must be taken cautiously since a performance effect of this kind could also reflect an impairment in motivational processing, motor control, or associative learning (Yin et al., 2008).

Action selection based on anticipated outcome value

Neither pretraining nor post-training lesions of the dmPFC had any detectable effect on outcome devaluation performance (Fig. 2C). A group by time of surgery by outcome value ANOVA found no effect of lesion type ($F < 1$), but did find a main effect of time of surgery ($F_{(1,25)} = 8.52$; $p < 0.01$), indicating that the rats that received pretraining surgery exhibited a generally lower overall rate of responding than rats that received post-training surgery. This effect did not significantly interact with lesion type ($F_{(1,25)} = 2.13$; $p > 0.15$). The ANOVA found a significant main effect of outcome value ($F_{(1,25)} = 54.03$; $p < 0.001$), and this effect did not interact with lesion type or time of surgery (F values < 1), nor was there a significant three-way interaction among these factors ($F_{(1,25)} = 1.13$; $p > 0.20$), suggesting that the groups did not differ in their ability to select actions based on outcome value.

Action selection based on the discriminative stimulus properties of the outcome

Both pretraining and post-training lesioned rats exhibited normal reinstatement performance (Fig. 2D), indicating that the ability to select actions using the discriminative stimulus properties of an associated outcome does not depend on this structure (Colwill, 1994; Ostlund and Balleine, 2007). A group by time of surgery by reinstatement period ANOVA found no main effect of lesion type ($F < 1$) or time of surgery ($F_{(1,25)} = 1.91$; $p > 0.15$), nor was there a significant interaction between these two factors ($F < 1$). There was, however, a significant main effect of reinstatement period ($F_{(3,75)} = 39.54$; $p < 0.001$). This effect did not significantly interact with lesion type

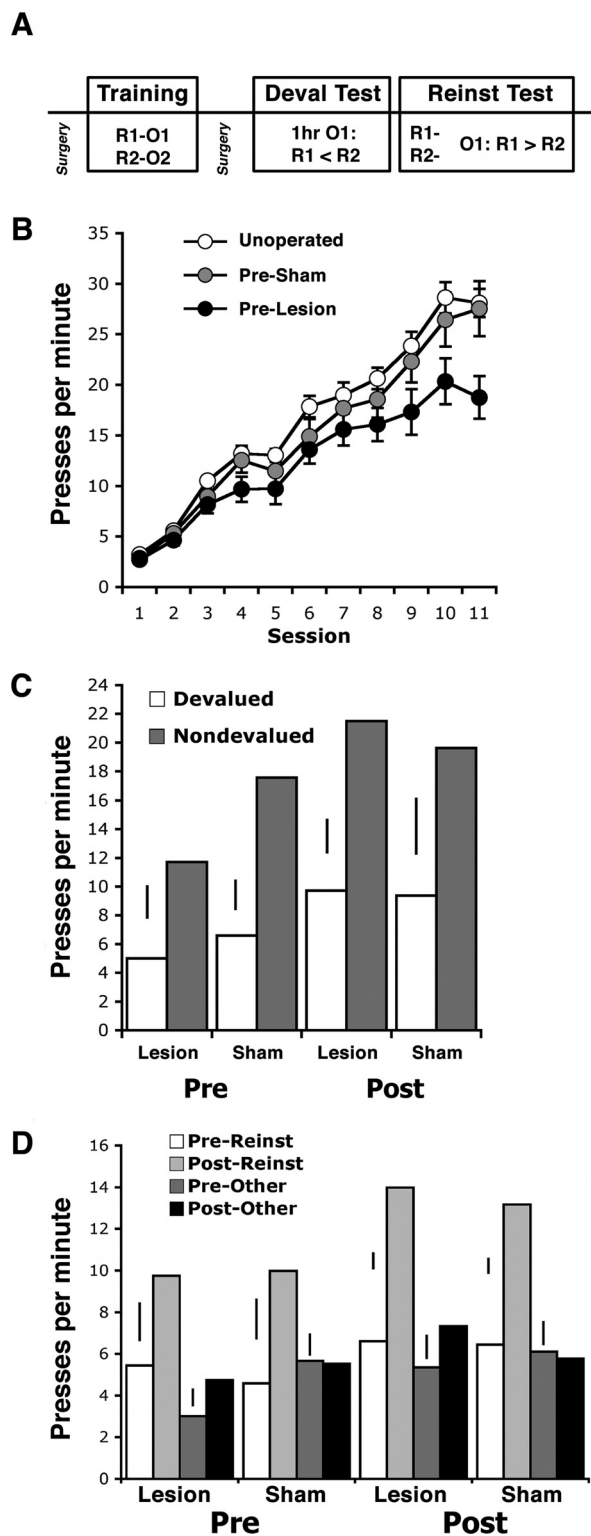


Figure 2. Effects of dmPFC lesions on acquisition and performance of the single-lever (nonsequence-based) instrumental conditioning task in experiment 1. **A**, Experimental design (see Materials and Methods for details). **B**, Acquisition of lever pressing, plotted as the number of responses performed per minute over successive sessions (\pm SEM). **C**, Results from outcome devaluation testing, plotted as the mean number of responses performed per minute for the response that had earned the devalued outcome and the response that had earned the other, nondevalued outcome [error bars represent 1 SE of the difference (SED) between means for the effect of outcome value]. **D**, Results from instrumental reinstatement testing, plotted as the mean rate of responding (presses per minute) during the 2 min periods before (Pre) and after (Post) each outcome delivery for the response trained with that outcome (Reinst) and the response trained with the other outcome (Other) (error bars represent 1 SED between means for the effect of period, presented separately for each response).

($F_{(3,75)} = 1.12$; $p > 0.30$) or time of surgery ($F_{(3,75)} = 1.32$; $p > 0.20$), nor was there a significant three-way interaction among these variables ($F < 1$). Further analysis found that, across groups, rats responded more in period Post-Reinst (reinstatement) than in periods Pre-Reinst ($F_{(1,28)} = 62.16$; $p < 0.001$) or Post-Other ($F_{(1,28)} = 52.84$; $p < 0.001$), confirming the outcome selectivity of the reinstatement effect.

Experiment 2: effects of dmPFC lesions on sequence-based instrumental conditioning

Acquisition of heterogeneous action sequences

The data were collapsed across pretraining and post-training surgery conditions because all rats were given sequence training after surgery (sham: $n = 13$; dmPFC: $n = 16$). During phase 1, when the relaxed response requirement was in place, rats increased their performance of sequences R1–R2 and R2–R1 (Fig. 3A). A group by sequence by session ANOVA found an effect of session ($F_{(3,81)} = 29.26$; $p < 0.001$), but found no effect of group ($F < 1$) or sequence ($F_{(1,27)} = 2.65$; $p > 0.10$), and found no significant interactions between these factors ($F_{(3,81)} \leq 1.76$; $p > 0.15$). Although the reinforced sequence (R1–R2) was not performed at a significantly higher rate than the nonreinforced sequence (R2–R1) during phase 1, both groups showed a selective increase in their rate of responding on the lever that was proximal to reward delivery (R2), relative to the distal lever press response (R1), consistent with the use of a response-based action selection strategy (Fig. 3B). A group by lever by session ANOVA found a significant main effect of lever ($F_{(1,27)} = 30.73$; $p < 0.001$) and session ($F_{(3,81)} = 13.74$; $p < 0.001$), as well as a significant lever by session interaction ($F_{(3,81)} = 6.90$; $p < 0.001$). Although there was no main effect of group ($F_{(1,27)} = 2.13$; $p > 0.10$), the ANOVA did find a marginal three-way interaction among group, lever, and session ($F_{(3,81)} = 2.46$; $p = 0.07$), suggesting that the lesion group may have had difficulty overcoming their preference for the proximal response. Further analysis supported this interpretation; whereas the sham group showed a significant lever by session interaction ($F_{(3,36)} = 6.25$; $p < 0.01$), the lesion group did not ($F < 1$).

Performance of sequence R1–R2 increased substantially in both groups during phase 2, when subjects were reinforced only after performing uninterrupted transitions from R1 to R2 (Fig. 3A). A group by sequence by session ANOVA found significant effects of sequence ($F_{(1,27)} = 46.44$; $p < 0.001$) and session ($F_{(7,189)} = 24.30$; $p < 0.001$), but found no effect of group ($F < 1$). The analysis also resulted in a significant sequence by session interaction ($F_{(7,189)} = 21.87$; $p < 0.001$) and a marginally significant group by sequence by session interaction ($F_{(7,189)} = 1.98$; $p = 0.06$), perhaps indicating a slight disruption of action sequence learning in the lesioned group. During phase 2, both groups continued to suppress their overall rate of lever pressing (Fig. 3B). A group by lever by session ANOVA found a significant effect of session ($F_{(5,189)} = 4.38$; $p < 0.001$), but found no effect of group ($F_{(1,27)} = 1.58$; $p > 0.20$) or lever ($F < 1$). There was a significant lever by session interaction ($F_{(7,189)} = 2.70$; $p < 0.05$), but no lever by group ($F < 1$) or lever by session by group interactions ($F_{(7,189)} = 1.26$; $p > 0.20$).

In phase 3, when rats were reinforced for performing the two actions in the reverse order, the dmPFC group displayed considerably less flexibility than the sham group in adapting their sequence performance to the current contingencies. A group by sequence by session ANOVA found significant effects of sequence ($F_{(1,27)} = 8.78$; $p < 0.01$) and session ($F_{(5,135)} = 4.54$; $p < 0.01$), but found no effect of group ($F < 1$) and no interaction between group and sequence ($F < 1$) or group and session ($F < 1$). There was, however, a significant sequence by session interaction

tion ($F_{(5,135)} = 27.67$; $p < 0.001$) and, more importantly, a significant group by sequence by session interaction ($F_{(5,135)} = 6.24$; $p < 0.001$), confirming that the groups were differentially sensitive to the sequence reversal. Further analysis restricted to data from the last session of this phase (session 18) found that lesioned rats were more likely to perform the previously reinforced sequence (R1–R2) than shams ($F_{(1,28)} = 5.98$; $p < 0.05$), but did not significantly differ from shams in their performance of the currently reinforced sequence (R2–R1; $F_{(1,28)} = 2.39$; $p > 0.10$). Both groups acquired a (somewhat transient) preference for the lever press response that was proximal (R1) to reward during phase 3 (Fig. 3B), relative to the distal response (R2). A group by lever by session ANOVA found significant main effects of lever ($F_{(1,27)} = 9.40$; $p < 0.01$) and session ($F_{(5,135)} = 38.23$; $p < 0.001$), as well as a significant lever by session interaction ($F_{(5,135)} = 18.69$; $p < 0.001$). There was also a marginal main effect of group ($F_{(1,27)} = 3.57$; $p = 0.07$). No other main effect or interaction reached significance ($F_{(5,135)} \leq 1.85$; $p > 0.10$).

During phase 4, when the two sequence–outcome relationships were concurrently active, both groups came to perform these sequences at comparable levels, reacquiring the performance of sequence R1–R2 and maintaining their performance of sequence R2–R1 (Fig. 3A). A group by sequence by session ANOVA found a significant effect of session ($F_{(3,81)} = 24.93$; $p < 0.001$) and a significant sequence by session interaction ($F_{(3,81)} = 14.42$; $p < 0.001$). No other effect or interaction reached significance ($F_{(3,81)} \leq 1.48$; $p > 0.20$). Both groups also came to distribute their responses evenly across the two levers during phase 4 (Fig. 3B), which should not be surprising since each lever press response was now both proximal and distal to reward. A group by lever by session ANOVA found a significant main effect of session ($F_{(3,81)} = 17.58$; $p < 0.001$) and a marginally significant group by lever interaction ($F_{(1,27)} = 3.28$; $p = 0.08$). No other main effect or interaction reached significance ($F_{(3,81)} \leq 1.74$; $p > 0.15$).

Effect of outcome devaluation on sequence performance

At this point we began to evaluate the rats' ability to use sequence-level action representations to organize their performance in a goal-directed manner. By the end of training, both groups of rats had learned to efficiently perform the two heterogeneous action sequences. However, this by itself does not demonstrate that they had encoded the action sequences as integrated units of behavior. Instead, it is possible that the rats were relying on a more elemental strategy, representing each lever press response as a distinct behavioral unit. These two accounts make very different predictions about the sensitivity of sequence performance to outcome devaluation. The chunking account holds that each two-action sequence should be selectively associated with its outcome (i.e., [R1→R2]→O3 and [R2→R1]→O4). If one of the outcomes is

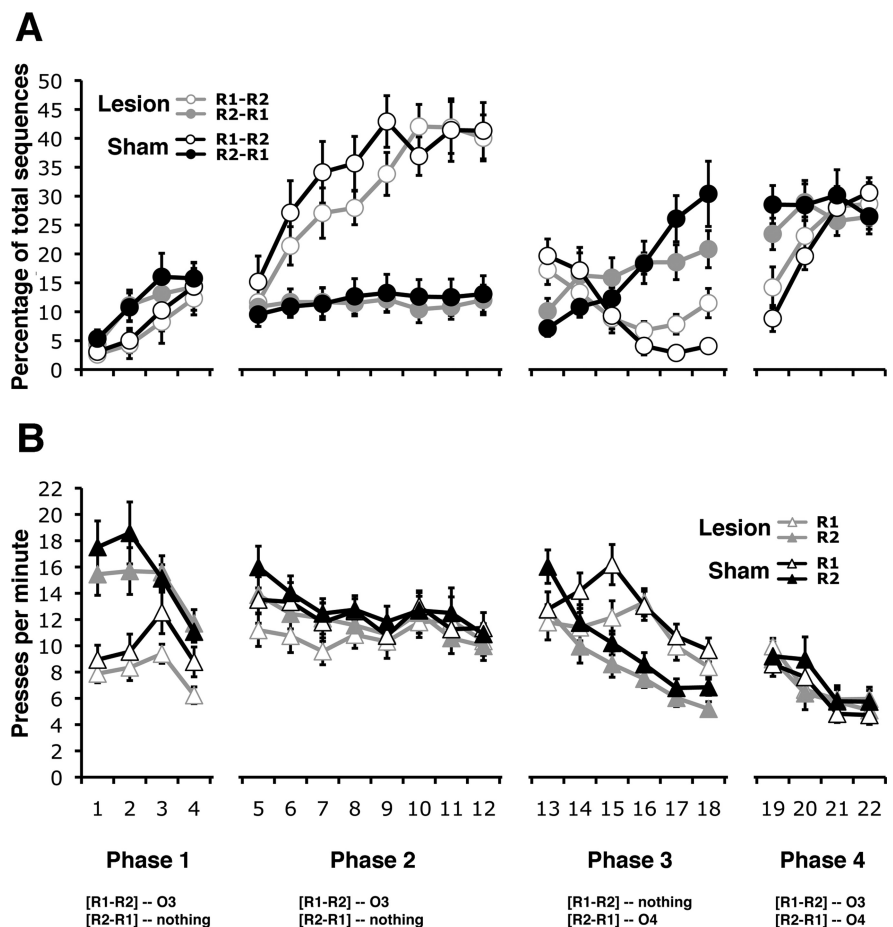


Figure 3. Acquisition of heterogeneous chain performance across training phases 1–4 in dmPFC- and sham-lesioned rats. **A**, Performance of sequence R1–R2 and sequence R2–R1 as a percentage of all two-lever press sequences performed (see Materials and Methods for details; \pm SEM). **B**, The mean number of lever presses performed per minute (\pm SEM), plotted separately for actions R1 and R2. The contingencies that were in place during each of the four training phases are outlined at the bottom of the figure (see Materials and Methods for details).

devalued, according to this account rats should suppress their performance of the corresponding sequence as a whole. In contrast, the elemental account holds that, because of their relative proximity to reward during sequence training, the individual lever press responses should have become differentially associated with the two outcomes (i.e., R2→O3 and R1→O4). Consequently, a selective reduction in outcome value should decrease the rate with which the associated lever press response is performed but should have no effect on the rats' preference between the two sequences.

Although both groups displayed sensitivity to outcome devaluation at test, the way this effect manifested itself in performance differed across groups. Whereas the lesioned group's sequence performance was insensitive to outcome devaluation, the sham group showed a clear bias in responding, performing the nondevalued sequence more often than the devalued sequence (Fig. 4A). This interpretation was confirmed by a group by sequence ANOVA, which found a significant main effect of sequence ($F_{(1,27)} = 9.15$; $p < 0.01$) and a significant sequence by group interaction ($F_{(1,27)} = 4.59$; $p < 0.05$). The main effect of group was not significant ($F < 1$). Further analysis found a significant effect of sequence for the sham group ($F_{(1,12)} = 11.20$; $p < 0.01$) but not for the lesioned group ($F < 1$). Interestingly, whereas the sham group responded at approximately the same rate on the two levers, the dmPFC group responded more on the lever that was

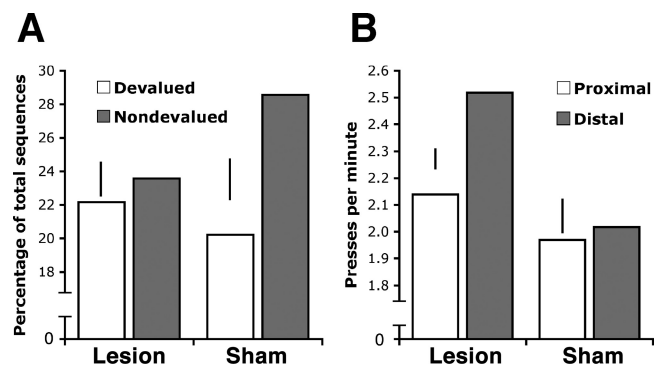


Figure 4. Sensitivity of chain performance to outcome devaluation in dmPFC- and sham-lesioned rats. **A**, The percentage of all two-lever press sequences performed (see Materials and Methods), plotted separately for the sequence that earned the devalued outcome and for the sequence that earned the other, nondevalued outcome [error bars represent 1 SE of the difference (SED) between means for the effect of sequence]. **B**, The mean number of lever presses performed per minute, plotted separately for the two lever press actions based on whether the action was proximal or distal to the devalued outcome during training (error bars represent 1 SED between means for the effect of lever).

distal to the devalued outcome than on the more proximal lever (Fig. 4B). A group by lever ANOVA found a significant effect of lever ($F_{(1,27)} = 8.64$; $p < 0.01$) and a significant lever by group interaction ($F_{(1,27)} = 5.20$; $p < 0.05$). Further analysis revealed the source of this interaction: there was a significant effect of lever in the dmPFC group ($F_{(1,15)} = 23.37$; $p < 0.001$) but not in the sham group ($F < 1$). This pattern of results would seem to indicate that, without the ability to use sequence-level action representations to guide their performance, the dmPFC group came to rely on individual action–outcome relations when choosing between actions. It is also worth considering the possibility that dmPFC lesions render rats relatively myopic in their evaluation of action–outcome relationships; i.e., rather than being unable to acquire and/or use action sequence representations per se, the lesioned rats may have been biased toward associating outcomes with only those actions directly involved in their delivery. This hypothesis deserves additional study.

Effect of contingency degradation on sequence performance

Finally, as a further test of the action sequence encoding, we assessed the sensitivity of sequence performance to instrumental contingency degradation. If the sham group used sequence-level action representations to guide their goal-directed behavior, then we should expect these rats to selectively decrease their performance of a sequence if the predictive relationship between that specific sequence and its outcome is degraded by delivering that outcome regardless of sequence performance.

Consistent with this prediction, we found that the sham group adapted to the contingency manipulation by selectively withholding their performance of the degraded sequence (Fig. 5A). In contrast, the lesioned group failed to show this effect and maintained comparable levels of performance for both sequences throughout the session. A group by sequence by block ANOVA found no effects of group ($F_{(1,27)} = 2.54$; $p > 0.10$), sequence ($F < 1$), or block ($F < 1$). Although none of the interactions reached the conventional level of significance, there was a marginally significant group by sequence interaction ($F_{(1,27)} = 3.45$; $p = 0.07$), suggesting that the groups differed in their choice between the two sequences. Further analysis of the sham group's data found a significant effect of sequence ($F_{(1,12)} = 5.77$; $p < 0.05$) and a significant sequence by block interaction ($F_{(4,48)} = 3.77$; $p < 0.01$), confirming that their choice between the two sequences

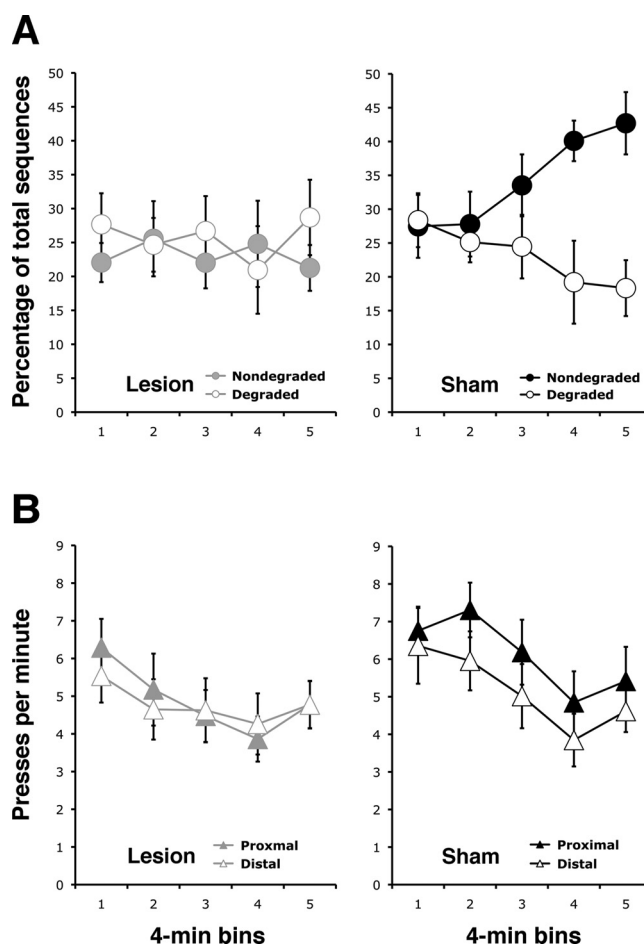


Figure 5. Sensitivity of chain performance to instrumental contingency degradation in dmPFC- and sham-lesioned rats. **A**, The percentage of all two-lever press sequences performed (\pm SEM) (see Materials and Methods), plotted separately for the sequence that earned the freely delivered outcome (Degraded) and the sequence that earned the other outcome (Nondegraded). **B**, The mean number of lever presses performed per minute (\pm SEM), plotted separately for the two-lever press actions based on whether the action was proximal or distal to the freely delivered outcome.

changed over the session as they experienced the change in contingency. In contrast, analysis of the lesion group's data found no effect of either sequence or block (F values < 1), and found no sequence by block interaction ($F < 1$).

To confirm this apparent group difference in sensitivity to contingency degradation, we also conducted a separate group by sequence by block ANOVA using only the first and last 5-min block of the session. The analysis found no effects of group ($F_{(1,27)} = 2.19$; $p > 0.05$), sequence ($F < 1$) or block ($F < 1$), nor did it detect a significant group by bin interaction ($F < 1$). However, the group by sequence ($F_{(1,27)} = 6.30$; $p < 0.05$), sequence by block ($F_{(1,27)} = 5.18$; $p < 0.05$), and group by sequence by block ($F_{(1,27)} = 6.90$; $p < 0.05$) interactions all reached significance. Further analysis revealed the source of this three-way interaction: whereas the groups responded similarly during the first block of the session (group by sequence interaction: $F < 1$), choice between sequences differed across groups during the last block (group by sequence interaction; $F_{(1,27)} = 10.15$; $p < 0.01$), with the sham group ($F_{(1,27)} = 9.64$; $p < 0.01$), but not the lesioned group ($F < 1$), showing a preference for the nondegraded sequence.

Over blocks, both groups showed a nonselective decrease in

their rate of responding on the two levers, regardless of their relative position in the degraded sequence (Fig. 5B). A group by lever by block ANOVA found a main effect of block ($F_{(4,108)} = 5.50$; $p < 0.001$) but found no other effect or interaction ($F_{(1,27)} \leq 1.55$; $p > 0.20$).

Discussion

We found that neither pretraining nor post-training dmPFC lesions reliably affected rats' ability to select actions using the motivational value or discriminative stimulus properties of an associated reward, at least for actions that involved the performance of a single response type. Furthermore, although these lesions did not prevent rats from learning to perform a heterogeneous action sequence, they were effective in biasing the way this task was performed; whereas the shams used sequence-level action representations to adjust their performance, lesioned rats did not and instead appeared to apply a more elemental, response-based strategy.

Action sequencing in primates is known to depend on the SMA (Tanji, 2001; Hoshi and Tanji, 2004; Kennerley et al., 2004; Rushworth et al., 2004). There is some indication that the rodent dmPFC plays a similar role, as lesions of this area have been shown to disrupt performance on a serial reaction time task similar to those used to study action sequencing in primates (Bailey and Mair, 2007). However, such tasks reveal little about the strategy subjects use to perform an action sequence. For instance, because subjects are provided with discriminative cues that signal the identity of the response to be performed next, they may come to solve the task by learning a set of simple stimulus–response associations rather than by encoding the entire sequence of actions (Terrace, 2005). Alternatively, action sequence learning may be aided by the capacity to construct, or chunk, hierarchical action representations out of more elementary units of behavior (Lashley, 1951; Miller et al., 1960; Rosenbaum et al., 1983). Consistent with the chunking hypothesis, studies investigating the microstructure of action sequence performance in humans have shown that subjects tend to complete longer sequences by making two or more uninterrupted chunks of responses, each preceded by a brief pause in performance (Rosenbaum et al., 1983; Sakai et al., 2002; Kennerley et al., 2004). Interestingly, Kennerley et al. (2004) have recently reported that repetitive transcranial magnetic stimulation delivered over the rostral SMA when subjects are about to perform an action chunk can disrupt the initiation of that chunk.

There is, however, little evidence that rats engage in action chunking (but see Reed and Morgan, 2006; Bachá-Méndez et al., 2007). To assess this possibility, we used a free-operant heterogeneous chain task in which rats could earn reward by performing two different lever press actions in the appropriate order. Such tasks pose a problem (cf. credit assignment in machine learning; Minsky, 1961) for models of instrumental conditioning since they tend to assume that the acquired strength of a response is determined by its temporal proximity to reward during training (Staddon and Zhang, 1991). Consequently, these models predict that the most distal element of the chain should be performed less frequently than more proximal elements, although the distal response must logically be performed first. Our results confirm a previous report (Bachá-Méndez et al., 2007) that heterogeneous chain performance is initially dominated by a response-based approach of this kind, in that rats showed a transient preference for the proximal action over the distal action during the first few sessions with a new chain (Fig. 3A). This response bias waned as rats learned to perform the reinforced sequence of lever presses more efficiently.

Although such findings are consistent with a shift in strategy, they do not necessarily indicate the use of an action chunking strategy. As

noted above, even a relatively simple stimulus–response learning strategy could mediate action sequence performance. Although the unsignaled free-operant choice procedure used in the current study should have discouraged this approach, it is possible that our rats were able to rely on local cues embedded within the task (e.g., the sight of the lever), perhaps in combination with proprioceptive cues generated by previous responses. Regardless of which specific associations are formed, however, this general strategy is based on the assumption that individual responses serve as the fundamental units of behavior on which operations like action selection and inhibition are performed. In contrast, the action chunking account assumes that an entire action sequence can serve as a behavioral unit, allowing subjects to make sequence-level changes in their behavior. Therefore, action selection phenomena like the outcome devaluation and instrumental contingency degradation effects should be expressed at the level of either individual responses or action sequences, depending on which strategy is being used. Using these tests, we found evidence that rats are indeed capable of action chunking; sham-lesioned rats were able to selectively suppress their performance of a sequence whose outcome was devalued at test (Fig. 4A) and adjust their choice between the two sequences in response to a selective contingency degradation manipulation in which one of the two outcomes was delivered regardless of whether or not its corresponding sequence was performed (Fig. 5A).

These findings have implications for theories of action sequence learning, which do not agree about the role of goal representations in sequence performance. For example, some accounts (Botvinick and Plaut, 2004; Graybiel, 2008) hold that stimulus–response (i.e., habit or procedural) learning is responsible for the acquisition of routine sequential behavior, and therefore predict that sequence performance should be impervious to manipulations of outcome value or sequence–outcome contingency. Our results would therefore appear to be more compatible with a goal-directed analysis of sequence performance that assumes that action sequences are selected through an executive (or supervisory) process based on their likely outcomes (Cooper and Shallice, 2006). Of course, it is also possible that the action chunking process responsible for constructing higher-order action representations is mediated by stimulus–response learning, and that the action chunks generated by this process then become associated with reward and are selected in a goal-directed manner.

Our results also indicate a role for the dmPFC in action chunking. Unlike shams, lesioned rats were unable to adjust their choice between the two sequences in response to a selective shift in either reward value (Fig. 4A) or instrumental contingency (Fig. 5A). Importantly, rather than using sequence-level action representations to organize their performance in a goal-directed manner, lesioned rats appeared to apply a response-level strategy, withholding their performance of the lever press response that had been proximal to the devalued outcome, relative to the other, distal action (Fig. 4B). This finding, together with the results of outcome devaluation testing in experiment 1 (Fig. 2C), indicates that, rather than playing a fundamental role in goal-directed action selection, the dmPFC plays a more limited role restricted to action chunking.

Lesioned rats also displayed a distinctive pattern of responding during action sequence training (Fig. 3). When first learning to perform sequence R1–R2, for instance, group dmPFC showed a slight delay in inhibiting their preference for the lever press action that was proximal to reward (R2), relative to the distal action (R1). Moreover, during phase 3 of sequence training, when rats were reinforced for performing these actions in the reverse order (R2→R1), lesioned rats perseverated in their per-

formance of the previously reinforced sequence. Although such findings could indicate a general deficit in response inhibition, we found no effect of these lesions on rats' capacity to selectively withhold an action whose outcome had been devalued (Fig. 2C). Alternatively, given their inability to use action sequence representations in outcome devaluation and contingency degradation testing, this response persistence may reflect the use of a strategy based on chaining discrete stimulus–response associations in lieu of the presumably more flexible action chunking strategy used by the sham group.

The region targeted in the current study, the medial agranular cortex, shares rich connections (Ray and Price, 1992; Reep and Corwin, 1999) with several areas implicated in instrumental action–outcome encoding, including the dorsal striatum, prelimbic cortex, and mediodorsal thalamus (Balleine and Dickinson, 1998; Corbit et al., 2003; Yin et al., 2005). It is notable, therefore, that neither pretraining nor post-training lesions of this region had any effect on the sensitivity of lever press performance to outcome devaluation (Fig. 2C). These results may also be somewhat surprising given the supposed homology between this area and the primate SMA, which is widely regarded as playing a central role in the control of voluntary behavior (Eccles, 1982; Goldberg, 1985; Passingham, 1993). For instance, there have been a number of reports that the SMA is involved in transitioning from prompted (or externally guided) performance to self-paced (or internally guided) performance (Thaler et al., 1995; Deiber et al., 1999; Jenkins et al., 2000), as might be expected if it were involved in the voluntary selection of actions. However, there is also some indication that, rather than being directly involved in goal-directed action selection (i.e., evaluating action–outcome contingency and outcome value), the SMA plays a downstream role in the preparation and execution of actions that have already been selected (Roesch and Olson, 2004). Regardless of the role of the primate SMA in action selection, our data suggest that the homologous region in rats is critical, not for goal-directed action selection per se, but for using action sequence representations in a goal-directed manner. How the dmPFC interacts with the basal ganglia to support action chunking in instrumental conditioning is a matter for future research.

References

- Bachá-Méndez G, Reid AK, Mendoza-Soylovna A (2007) Resurgence of integrated behavioral units. *J Exp Anal Behav* 87:5–24.
- Bailey KR, Mair RG (2007) Effects of frontal cortex lesions on action sequence learning in the rat. *Eur J Neurosci* 25:2905–2915.
- Balleine BW, Dickinson A (1998) Goal-directed instrumental action: contingency and incentive learning and their cortical substrates. *Neuropharmacology* 37:407–419.
- Botvinick M, Plaut DC (2004) Doing without schema hierarchies: a recurrent connectionist approach to normal and impaired routine sequential action. *Psychol Rev* 111:395–429.
- Colwill RM (1994) Associative representations of instrumental contingencies. *Psychol Learn Motiv* 31:1–72.
- Cooper RP, Shallice T (2006) Hierarchical schemas and goals in the control of sequential behavior. *Psychol Rev* 113:887–916.
- Corbit LH, Muir JL, Balleine BW (2003) Lesions of mediodorsal thalamus and anterior thalamic nuclei produce dissociable effects on instrumental conditioning in rats. *Eur J Neurosci* 18:1286–1294.
- Deiber MP, Honda M, Ibañez V, Sadato N, Hallett M (1999) Mesial motor areas in self-initiated versus externally triggered movements examined with fMRI: effect of movement type and rate. *J Neurophysiol* 81:3065–3077.
- Dickinson A, Balleine BW (1994) Motivational control of goal-directed action. *Learn Behav* 22:1–18.
- Donoghue JP, Wise SP (1982) The motor cortex of the rat: cytoarchitecture and microstimulation mapping. *J Comp Neurol* 212:76–88.
- Eccles JC (1982) The initiation of voluntary movements by the supplementary motor area. *Arch Psychiatr Nervenkr* 231:423–441.
- Floresco SB, Ghods-Sharifi S (2007) Amygdala–prefrontal cortical circuitry regulates effort-based decision making. *Cereb Cortex* 17:251–260.
- Goldberg G (1985) Supplementary motor area structure and function: review and hypotheses. *Behav Brain Sci* 8:567–615.
- Graybiel AM (2008) Habits, rituals, and the evaluative brain. *Annu Rev Neurosci* 31:359–387.
- Hoshi E, Tanji J (2004) Differential roles of neuronal activity in the supplementary and presupplementary motor areas: from information retrieval to motor planning and execution. *J Neurophysiol* 92:3482–3499.
- Jenkins IH, Jahanshahi M, Jueptner M, Passingham RE, Brooks DJ (2000) Self-initiated versus externally triggered movements. II. The effect of movement predictability on regional cerebral blood flow. *Brain* 123:1216–1228.
- Kennerley SW, Sakai K, Rushworth MFS (2004) Organization of action sequences and the role of the pre-SMA. *J Neurophysiol* 91:978–993.
- Lashley KS (1951) The problem of serial order in behavior. In: *Cerebral mechanisms in behavior* (Jefferies LA, ed), pp 112–136. New York: Wiley.
- Miller GA, Galanter E, Pribram KH (1960) Plans and the structure of behavior. New York: Holt, Rinehart and Winston.
- Minsky M (1961) Steps toward artificial intelligence. *Proc IRE* 49:8–30.
- Ostlund SB, Balleine BW (2007) Selective reinstatement of instrumental performance depends on the discriminative stimulus properties of the mediating outcome. *Learn Behav* 35:43–52.
- Passingham RE (1993) The frontal lobes and voluntary action. Oxford, UK: Oxford UP.
- Passingham RE, Myers C, Rawlins N, Lightfoot V, Fearn S (1988) Premotor cortex in the rat. *Behav Neurosci* 102:101–109.
- Paxinos G, Watson C (1998) The rat brain in stereotaxic coordinates, Ed 4. San Diego: Academic.
- Ray JP, Price JL (1992) The organization of the thalamocortical connections of the mediodorsal thalamic nucleus in the rat, related to the ventral forebrain–prefrontal cortex topography. *J Comp Neurol* 323:167–197.
- Reed P, Morgan TA (2006) Resurgence of response sequences during extinction in rats shows a primacy effect. *J Exp Anal Behav* 86:307–315.
- Reep RL, Corwin JV (1999) Topographic organization of the striatal and thalamic connections of rat medial agranular cortex. *Brain Res* 841:43–52.
- Roesch MR, Olson CR (2004) Neuronal activity related to reward value and motivation in primate frontal cortex. *Science* 304:307–310.
- Rosenbaum DA, Kenny SB, Derr MA (1983) Hierarchical control of rapid movement sequences. *J Exp Psychol Hum Percept Perform* 9:86–102.
- Rushworth MF, Walton ME, Kennerley SW, Bannerman DM (2004) Action sets and decisions in the medial frontal cortex. *Trends Cogn Sci* 8:410–417.
- Sakai K, Ramnani N, Passingham RE (2002) Learning of sequences of finger movements and timing: frontal lobe and action-oriented representation. *J Neurophysiol* 88:2035–2046.
- Staddon JER, Zhang Y (1991) On the assignment-of-credit problem in operant learning. In: *Quantitative analyses of behavior: neural network models of conditioning and action* (Commons ML, Grossberg S, and Staddon JER, eds), pp 279–293. Hillsdale, NJ: Erlbaum.
- Tanji J (2001) Sequential organization of multiple movements: involvement of cortical motor areas. *Annu Rev Neurosci* 24:631–651.
- Terrace HS (2005) The simultaneous chain: a new approach to serial learning. *Trends Cogn Sci* 9:202–210.
- Thaler D, Chen Nixon PD, Stern CE, Passingham RE (1995) The functions of the medial premotor cortex. I. Simple learned movements. *Exp Brain Res* 102:445–460.
- Walton ME, Bannerman DM, Rushworth MFS (2002) The role of rat medial frontal cortex in effort-based decision making. *J Neurosci* 22:10996–11003.
- Yin HH, Ostlund SB, Knowlton BJ, Balleine BW (2005) The role of the dorsomedial striatum in instrumental conditioning. *Eur J Neurosci* 22:513–523.
- Yin HH, Ostlund SB, Balleine BW (2008) Reward-guided learning beyond dopamine in the nucleus accumbens: the integrative functions of corticobasal ganglia networks. *Eur J Neurosci* 28:1437–1448.