Performance Feedback Drives Caudate Activation in a Phonological Learning Task

Elizabeth Tricomi[1,2], Mauricio R. Delgado[3], Bruce D. McCandliss[4], James L. McClelland[2,5], and Julie A. Fiez[1,2]

Abstract

■ Adults have difficulty discriminating nonnative phonetic contrasts, but under certain circumstances training can lead to improvement in this ability. Despite the ubiquitous use of performance feedback in training paradigms in this and many other domains, the mechanisms by which feedback affects learning are not well understood. In this event-related functional magnetic resonance imaging study, we examined how performance feedback is processed during perceptual learning. Thirteen Japanese speakers for whom the English phonemes [r] and [l] were nondistinct performed an identification task of the words ''road'' and ''load'' that has been shown to be effective in inducing learning only when performance feedback is present. Each subject performed alternating runs of training with and without feedback, followed by performance of a card-guessing task with monetary reward and punishment out-comes. We found that the caudate nucleus was more robustly activated bilaterally when performing the perceptual identification task with feedback than without feedback, and the right caudate nucleus also showed a differential response to positive and negative feedback. Moreover, using a within-subjects design, we found that the caudate nucleus also showed a similar activation pattern to monetary reward and punishment outcomes in the card-guessing task. These results demonstrate that the caudate responds to positive and negative feedback during learning in a manner analogous to its processing of extrinsic affective reinforcers and indicate that this region may be a critical moderator of the influence of feedback on learning. These findings impact our broader understanding of the mechanisms underlying nondeclarative learning and language acquisition. ■

INTRODUCTION

Prelinguistic infants can discriminate among speech sounds according to adult phonetic categories in any language, yet adults have difficulty discriminating phonetic contrasts that are not present in their native language. For instance, native Japanese speakers tend to hear the English liquid consonants [r] and [l] as the same sound, and thus are greatly impaired relative to native English speakers at discriminating between these two phonemes (Miyawaki et al., 1975). This apparent loss of plasticity, however, is not necessarily the consequence of permanent neural loss (Werker & Tees, 1984). Indeed, studies aiming to train speakers to identify or discriminate between nonnative phonemes have shown that some improvement is possible; these include studies of the [r]–[l] contrast (Callan et al., 2003; Bradlow, Pisoni, Yamada, & Tohkura, 1997; Lively, Logan, & Pisoni, 1993; Logan, Lively, & Pisoni, 1991; Strange & Dittmann, 1984), as well as other nonnative contrasts (Golestani & Zatorre, 2004; Morosan & Jamieson, 1989; Jamieson & Morosan, 1986; Tees & Werker, 1984).

In each of these studies, performance feedback was provided during the training phase of the experiment, but the specific role that feedback played in affecting performance was not examined. In a study performed in our laboratory, however, the effect of different types of training paradigms, including the presence or absence of immediate performance feedback, was examined (McCandliss, Fiez, Protopapas, Conway, & McClelland, 2002). Japanese speakers were trained to identify the beginning sound in the words ''road'' and ''load'' or ''rock'' and ''lock.'' For two ''fixed'' stimuli conditions, the same two speech tokens were always used in training, one beginning with [r] and one with [l]. These tokens could easily be identified by native English speakers, but could not be reliably identified by the Japanese speakers prior to training. One of these groups was trained using immediate performance feedback, whereas the other group was trained using no feedback. For two other experimental conditions, an ''adaptive'' training paradigm was used. In this paradigm, training began with synthesized sound tokens with the initial sound exaggerated such that the participants could distinguish between the contrasting words. The amount by which the tokens were exaggerated was then gradually reduced based on subject performance. Again, one of

[1]University of Pittsburgh, [2]Center for the Neural Basis of Cognition, Pittsburgh, [3]Rutgers-Newark, [4]Weill Medical College of Cornell University, [5]Carnegie Mellon University

these groups trained with feedback, and the other group trained without feedback, forming a 2 × 2 experimental design with four experimental conditions.

Participants in the adaptive/no-feedback condition improved from pre- to posttraining tests significantly more than a control group that was not trained, whereas improvement of the participants in the fixed/no-feedback condition was minimal and not significantly different from the control group. However, participants in the fixed/feedback condition (as well as the adaptive/feedback condition) also improved their performance significantly more than the control group and the fixed/no-feedback group. A Hebbian account of learning was proposed to explain the improvement in the adaptive training condition with no feedback. However, the Hebbian account does not provide a complete explanation of the learning mechanisms underlying the development of perceptual representations, because with feedback, learning occurred in the fixed condition. Rather, a feedback-based learning mechanism must also exist. We hypothesized that this feedback-dependent learning in the fixed condition must be reflected in some critical difference in neural processing.

We examined this issue directly by using functional magnetic resonance imaging (fMRI) to scan native Japanese speakers while they performed an [r]–[l] identification task (Figure 1), similar to the fixed stimuli training condition utilized by McCandliss et al. (2002). Participants performed this task both with and without performance feedback, allowing for a within-subject comparison. In this way, we were able to contrast activation from a training paradigm that had been shown to be effective in producing learning to one that was not effective. This approach is similar to that of lesion-based experiments, in which patients' performance is compared on two similar tasks that vary in one critical respect thought to involve the lesion site. After the scanning session, the participants performed six 20-min sessions of identification training using the fixed stimuli; half the participants trained with feedback and half without. This allowed us to confirm the behavioral finding that the fixed training paradigm is effective in inducing learning only when performance feedback is present.

Although it has no extrinsic value, the intrinsic value of feedback may be quite potent to motivated learners. If a shared system is involved in processing intrinsic and extrinsic rewards and punishments, then similar neural mechanisms should be recruited during both a language learning task with feedback (i.e., intrinsic rewards) and a simple guessing task with monetary incentives (i.e., extrinsic rewards). We therefore hypothesized that the feedback condition would elicit activation in brain regions associated with processing reward-related information, whereas the no-feedback condition would not. Both electrophysiological and imaging studies have implicated the striatum in the processing of primary and secondary rewards
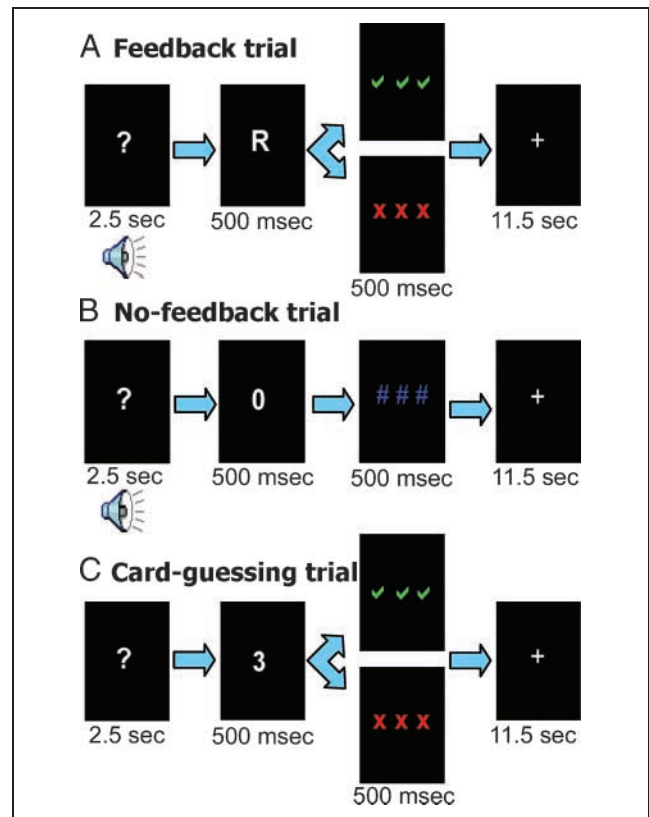

Figure 1. Experimental design. Feedback trials (A) in the [r]–[l] identification task began with the presentation of the sound token "road" or "load." After a 2.5-sec choice period, the first letter of the presented word was shown for 500 msec, followed by a 500-msec feedback display. Three green checkmarks indicated a correct answer; three red Xs indicated an incorrect answer. A delay period followed. No feedback trials (B) in the [r]–[l] identification task proceeded similarly to the feedback trials, except that a "0" was displayed instead of the first letter of the presented word, and three blue # signs replaced the feedback display on every trial. Trials in the card-guessing task (C) had the same temporal structure as [r]–[l] identification task trials. During the choice period, participants guessed whether the value on the back of the "card" was higher than or lower than 5. After the choice period, the number was presented for 500 msec, followed by a feedback display; in this task, the green checkmarks signified a $3.00 monetary reward, and the red Xs signified a $1.50 monetary punishment.

(Berns, McClure, Pagnoni, & Montague, 2001; Breiter, Aharon, Kahneman, Dale, & Shizgal, 2001; Delgado, Nystrom, Fissell, Noll, & Fiez, 2000; Elliott, Friston, & Dolan, 2000; Knutson, Westdorp, Kaiser, & Hommer, 2000; Schultz, 1998; Hikosaka, Sakamoto, & Usui, 1989). Human fMRI studies have shown that the caudate nucleus, a structure in the dorsal striatum, is activated by extrinsic monetary rewards (Delgado, Nystrom, et al., 2000; Elliott et al., 2000; Knutson, Westdrop, et al., 2000), especially when participants believe that their responses determine the outcome (Tricomi, Delgado, & Fiez, 2004). The caudate might therefore be activated by feedback, and could facilitate learning by linking

actions with their consequences. In order to directly compare brain regions involved in processing extrinsic monetary rewards and those involved in feedback processing during the [r]–[l] task, participants were scanned while performing a two-choice card-guessing task that involves monetary gain or loss, in addition to the [r]–[l] identification task.

METHODS

Participants

Nineteen healthy, right-handed volunteers took the pretest for this experiment; of these, 14 qualified to continue with the experiment (see Pretest section). Data from one subject were not used in the analyses due to excessive head motion, leaving 13 participants with analyzable data (8 women, 5 men; mean age ± SD, 30 ± 5.6 years). Participants were recruited through posted advertisements and earned $114 for their participation in the study (including winnings from the card-guessing task, which were actually fixed and equaled $60 for all participants). All participants were native Japanese speakers who were born in Japan and lived there until at least the age of 18. No participants reported hearing problems or conditions that would preclude them from having an MRI scan done (e.g., metal in body). Additionally, behavioral data were used from a control group ($n = 8$) from a previous study (McCandliss et al., 2002), who took a pre- and posttest but did not train or participate in an fMRI scanning session. All participants gave informed consent according to the Institutional Review Board at the University of Pittsburgh.

Procedure

Pretest

Participants were given a brief pretest to assess their ability to distinguish between the phonemes [r] and [l]. Stimuli were drawn from two synthetic speech continua used in previous studies, one ranging from "road" to "load" and the other ranging from "mode" to "node" (McCandliss et al., 2002). Ten stimuli from each continuum, spaced equally along it, were used.

The pretest consisted of six blocks. The first two blocks involved a road/load identification task, in which a sound token from the road/load continuum was presented and participants were asked to determine whether the word began with an "R" or "L" by pressing one of two buttons without feedback. There were 66 trials per block, for a total of 132 trials. The successive two blocks involved a mode/node identification task as a control because the [m] and [n] phonemes are present in the Japanese language. Categorization curves were plotted from each subject's identification data. Only participants who identified fewer than 75% of the sound tokens sampled at the 0.2 and 0.6 steps along the road/

load continuum as "load" and "road," respectively, participated further in the experiment. These two stimuli, which are reliably discriminated by native English speakers yet unreliably discriminated by many native Japanese speakers (McCandliss et al., 2002), were incorporated in the imaging experiment below. The pretest battery concluded with additional measures to characterize discrimination performance at the natural endpoints of the two continua.

Scan Session

Participants were scanned using a conventional 1.5-T GE (Milwaukee, WI) Signa whole-body scanner and standard radio frequency head coil. Structural images were collected using a standard T1-weighted pulse sequence, in 36 contiguous slices parallel to the AC–PC line (0.9375 × 0.9375 × 3.8 mm voxels). Oblique axial functional images were collected at the location of the middle 20 of the structural slices (3.75 × 3.75 × 3.8 mm voxels), which resulted in brain coverage from approximately $z = 51$ to $z = -25$, according to the atlas of Talairach and Tournoux (Talairach & Tournoux, 1988). A one-shot spiral pulse sequence was used (TR = 1500 msec, TE = 35 msec, FOV = 24 cm, flip angle = 70°). Stimulus presentation and behavioral data acquisition were controlled using E-prime software (Schneider, Eschman, & Zuccolotto, 2002) and the integrated functional imaging system (IFIS-SA, MRI Devices, Waukesha, WI).

Behavioral Paradigm

The experiment consisted of 12 5-min runs using three different paradigms: six runs of event-related road/load identification, two runs of blocked road/load identification, and four runs of an event-related gambling task.

Event-related Road/Load Identification

The experimental session began with six runs of an event-related road/load identification task (5 min per run). Participants alternated between performing the task with feedback (three runs) and without feedback (three runs). Whether the first run was performed with or without feedback was counterbalanced across participants.

Each run consisted of 20 trials, for a total of 120 trials (60 feedback, 60 no feedback), each 15 sec long (Figure 1). Each trial began with the visual presentation of a white question mark in the center of the black screen, surrounded by a white border, simultaneous with an auditory presentation of a sound token of the word "load" or the word "road" (0.2 and 0.6 along continuum, respectively). Order of sound token presentation was random with the constraint that there were 10 presentations of the "road" token and 10 of the "load" token. Participants had 2.5 sec to respond by pressing

the middle finger button of the response glove if they thought the word started with an "L" or the first finger button if they thought the word started with an "R." During feedback runs, the question mark was then replaced with a capital R if the word was "road" or a capital L if the word was "load," regardless of subject response. For no-feedback runs, a neutral symbol ("0") was shown for every trial.

After 500 msec, this display was replaced with a feedback or a neutral display, which was shown for 500 msec; for each trial within a feedback run, if the subject answered correctly, three green √s were shown in the center of the screen, whereas if the subject answered incorrectly, three red Xs were shown. If the subject made no response, three white hyphens were shown and the trial was excluded from analysis; this occurred on 0.9% of trials. For no-feedback runs, the subject's response was followed by three blue # symbols, regardless of accuracy, rather than the √/X feedback. If the subject did not respond during the allowed period, however, three white hyphens were shown as in the feedback condition runs, and again the trial was excluded from analysis; 1.4% of trials were consequently excluded. Each trial ended with an 11.5-sec delay period in which participants fixated on a white cross in the center of the screen.

Blocked Road/Load Identification

So that we could independently identify regions of activation that could then be applied to the event-related datasets from the [r]–[l] identification task and the card-guessing task, we also acquired data from two 5-min runs of the identification task utilizing a blocked design. Within each run, there were five cycles of trials, with each cycle containing a block of five 3-sec feedback trials with a subsequent 15-sec rest period and a block of five no-feedback trials and another subsequent rest period. In one of the runs, cycles began with a feedback block and in the other, cycles began with a no-feedback block; the order of these two runs was random. Stimulus order was also random.

Each trial began with the simultaneous presentation of a sound token and a visually displayed question mark, as in the event-related trials. There was a 2-sec response period, followed immediately by a feedback or a neutral display. This display was shown for 500 msec and was then replaced with a white fixation cross, which was displayed for 500 msec while the subject prepared for the next trial. For trials within feedback task blocks, the display consisted of three green √s following each correct answer and three red Xs following each incorrect answer. For trials within no-feedback task blocks, the display consisted of three neutral blue # signs, whether the response was correct or incorrect. For trials in either block type, if the subject did not respond, the feedback display was of three white hyphens. During the 15-sec

rest period, participants fixated on a + sign in the center of the screen; it was gray to distinguish it from the white fixation cross used in between trials within a block.

Event-related Card-guessing Task

The scanning session ended with four 5-min event-related runs of a card-guessing task like that used in previous work (Delgado, Nystrom, et al., 2000). For this task, participants were instructed that on each trial, they had to guess whether the number on a playing card was higher or lower than 5. It was explained that the correct number is always a number between 1 and 9, but never exactly 5, and that the probability of any of these eight numbers appearing is equal.

The structure of the card-guessing task approximated the structure of the event-related identification runs with feedback (Figure 1). Participants earned $3.00 for every correct guess and lost $1.50 for every incorrect guess. To ensure an equal number of correct and incorrect guesses, the outcome of guesses was actually determined by the experimenter without the participants' knowledge. Fifty percent of the trials were "reward" trials, in which the subject won money by guessing correctly, and 50% of the trials were "punishment" trials, in which the subject lost money by guessing incorrectly. No response errors occurred on 1.7% of trials, and these trials were excluded from analysis. Total winnings equaled $60.

Follow-up Training

To investigate the effects on performance of training with and without feedback over a longer period than the scanning session, each subject took home a laptop computer after the scanning session and used a training program involving a "road"–"load" identification task for 6 days, before returning to the laboratory for a posttest. The training program was the same as the "fixed stimuli" training program used in prior work (McCandliss et al., 2002). On each trial, participants heard the word "road" or "load" and responded as to whether they thought the word began with "R" or "L." Half the participants trained with feedback and half without feedback. Data from noncompliant participants with fewer than four sessions of usable training data were not included in the extended training behavioral analysis, leaving five participants in the feedback group and six in the no-feedback group. Participants completed 530 trials per day, which took approximately 20 min.

At the end of the study, participants returned to the laboratory and took a posttest, which was identical to the pretest. This allowed us to test whether our results replicated those of McCandliss et al. (2002), that is, whether the feedback training more effectively facilitated learning than the no-feedback training. To ensure

that participants were given the opportunity to improve their identification as much as possible, those in the no-feedback group who wished to do an additional week of training with feedback after the posttest were allowed to do so.

Data Analysis

Performance on the [r]–[l] identification task during the scanning session was assessed using two-tailed t tests to compare the participants' overall performance versus chance, accuracy on feedback trials versus no-feedback trials, and accuracy during the first and last run of the task in the scanner. Reaction times (RTs) were also analyzed. The effects of extended training were assessed through analysis of the participants' pre- and posttest categorization data. Additionally, performance of each training group was compared to a control group that took the pre-and posttest 3 days apart, but did not train.

The NeuroImaging Software package (NIS 3.5), developed at the University of Pittsburgh and Princeton University, was used to analyze the fMRI data, along with the graphical computing environment, Functional Imaging Software Widgets (fiswidgets; Fissell et al., 2003). Images were reconstructed and corrected for subject motion with Automated Image Registration (AIR 3.08; Woods, Cherry, & Mazziotta, 1992). Runs in which motion exceeded 4 mm or 4° in any direction were not used in analysis; this left analyzable data from 13 participants for the event-related [r]–[l] task, 11 for the blocked [r]–[l] task, and 9 for the card-guessing task. The images were detrended to adjust for scanner drift within runs. The structural images of each subject were stripped to remove the skull and coregistered to a common reference brain chosen from among the participants (Woods, Mazziotta, & Cherry, 1993). Functional images were transformed into the same common space, normalized by a mean scaling of each image to match global mean image intensities across participants, and smoothed using a three-dimensional Gaussian filter (8-mm FWHM) to account for anatomical differences between participants. This set of data was then analyzed statistically. To visualize the data, the AFNI software program was used (Cox, 1996); this program was also used to warp the data into Talairach space (Talairach & Tournoux, 1988).

A repeated measures two-way ANOVA was performed on the blocked design [r]–[l] task data with subject as a random factor and block type (active/rest) and feedback presence (feedback/no feedback) as within-subject factors. Activation clusters were defined as regions with three or more contiguous voxels showing a significant effect at a threshold of $p < .001$; this contiguity threshold serves as a precaution against type 1 errors (Forman et al., 1995). The Block Type × Feedback Presence activation clusters obtained from this analysis were then applied to the event-related identification task data and

the resulting time series data for each activation cluster were analyzed by plotting mean blood oxygen level dependent intensity for each condition for each 1.5-sec scan period. The same clusters were also applied to the event-related gambling data and the resulting time series plotted in a similar fashion. This enables a direct comparison, in independently identified voxels, of whether the areas activated by the feedback in the identification task produce similar hemodynamic responses when the subject is receiving feedback indicating monetary reward and punishment in the card-guessing task. Dissociations between positive and negative feedback were also assessed through this analysis.

Although this primary analysis affords a direct comparison of activation in identical regions across tasks, follow-up statistical analyses were performed to verify the primary analysis and to provide additional information. To verify that the activation clusters identified from the blocked design data are similar to the regions showing significant effects in the event-related version of the identification task, the event-related identification task data were analyzed using a voxelwise repeated measures two-way ANOVA with subject as a random factor and feedback presence (feedback/no feedback) and time (1.5-sec periods T1–T10) as within-subject factors. A second ANOVA was performed on the feedback data only, with subject as a random factor and valence (correct/incorrect) and time (1.5-sec periods T1–T10) as within-subject factors. Similarly, the regions showing significant effects in the gambling task were identified through a voxelwise repeated measures ANOVA with subject as the random factor and valence (reward/punishment) and time (1.5-sec periods T1–T10) as within-subject factors. Using the AFNI software program (Cox, 1996), the Valence × Time activation clusters from these two ANOVAs were overlayed to visualize their degree of overlap. Additionally, a conjunction analysis was performed by doing an ANOVA on the combined dataset from both the gambling and [r]–[l] tasks with subject as the random factor and valence (monetary rewards and positive feedback/monetary punishments and negative feedback) and time (1.5-sec periods T1–T10) as within-subject factors. This served as another indication of which brain areas are activated by positive and negative outcomes in the two tasks.

RESULTS

Behavioral Results from Scanning Session

Prior to training, participants were scanned while performing six runs of the event-related [r]–[l] identification task and two runs of the blocked version of the task, followed by four runs of the event-related card-guessing task (Figure 1). Overall, accuracy for the [r]–[l] task was 55%, which was significantly greater than chance, $t(12) = 2.82$, $p < .05$, two-tailed. However, there were

no significant differences in accuracy between the feedback and no-feedback trials, nor did accuracy significantly improve over the course of the scanning session. These results emphasize the fact that the scanning session occurred at the beginning of the learning process in a slowly learned task.

For the event-related [r]–[l] task, RTs were faster for feedback runs than for no-feedback runs: mean feedback RT = 1211 msec, mean no-feedback RT = 1271 msec, $t(12) = 2.62$, $p < .05$, two-tailed. This difference may be due to increased motivation during the feedback runs. However, during the blocked [r]–[l] task, which was faster-paced, there were no significant differences between feedback and no-feedback blocks (mean feedback RT = 967 msec, mean no-feedback RT = 942 msec), and thus simple RT differences cannot account for task differences observed across both the event-related and blocked paradigms. No significant differences were found for RT on correct versus incorrect trials.

Functional Imaging Results

Regions Identified from Blocked Design [r]–[l] Task

We performed a voxelwise ANOVA on the [r]–[l] task blocked-design data, with block type (active/fixation) and feedback presence (feedback/no feedback) as factors (fixation blocks were coded as feedback or no feedback based on the condition of the immediately preceding active task block). This enabled us to obtain activation clusters that could then be applied to the event-related [r]–[l] and card-guessing tasks, which allowed us to compare activation from these two tasks in independently identified regions. Activation clusters showing a main effect of block type are listed in Table 1. This article focuses on the activation clusters that show a Block Type × Feedback Presence interaction because these clusters are differentially active when performing the task versus a fixation rest period, with the level of activation further affected by whether the task is performed with or without feedback. The three clusters that exhibited such an interaction are listed in Table 1, and include bilateral activation clusters in the caudate nuclei, as well as a cluster in the right parahippocampal gyrus. Figure 2A shows the bilateral activation clusters in the caudate nuclei, which exhibited increased activation during feedback blocks relative to no-feedback blocks and fixation. The parahippocampal cluster, in contrast, showed decreased activation for feedback blocks relative to no-feedback blocks and fixation.

Blocked Design Regions Applied to the Event-related Datasets

To further investigate activation in these regions, we applied the activation clusters identified from the blocked design dataset to the event-related [r]–[l] task and card-guessing task datasets. This allowed us to compare time courses of activation across the two tasks in voxels identified from an independent dataset. A repeated measures ANOVA on the event-related data in the caudate clusters showed a significant Feedback Presence × Time effect, $F(9,108) = 8.15$, $p < .001$ for left caudate; $F(9,108) = 11.73$, $p < .001$ for right caudate. As shown in Figure 2, activation in the caudate nuclei bilaterally was greater for the feedback compared to the no-feedback trials of the [r]–[l] identification task. The feedback effect was somewhat stronger in the right caudate cluster than in the left caudate cluster, although in a further exploratory analysis, the Hemisphere × Feedback Presence × Time effect was not significant at the $p < .001$ threshold, $F(9,108) = 2.78$, $p = .006$. In contrast to the findings in the caudate nuclei, the time course of activation in the parahippocampal cluster was relatively flat with no discernible peak and did not show a significant Feedback Presence × Time effect in the event-related dataset. Therefore, it seems that any block-level effects of feedback presence in this area were not present at the trial level.

Time Courses of Activation

Figure 3 shows time courses for the caudate nuclei for the event-related feedback trials of the [r]–[l] task, broken down by whether the subject was correct or incorrect (i.e., whether the feedback was positive or negative). Prior studies using the card-guessing task found the largest and most reliable difference between reward and punishment responses to occur 6–9 sec after the outcome display, which corresponds to periods T7 and T8 in the current study. We examined these periods in the [r]–[l] identification task with two-tailed paired t tests and were surprised to find only a slightly greater response to correct than incorrect feedback, which was a nonsignificant difference, $t(12) < 1.78$, ns. Further exploratory analysis of the other time periods showed none to have a significant valence effect, $t(12) < 1.78$, ns.

In these same voxels in the caudate nuclei, there was robust activation during the event-related card-guessing task: main effect of time, $F(9,72) = 9.30$, $p < .001$ for left caudate; $F(9,72) = 14.65$, $p < .001$ for right caudate. The time courses in the caudate clusters for the card-guessing task, broken down by monetary gain and monetary loss (reward and punishment, respectively), are also shown in Figure 3. The reward and punishment responses replicated previous work done with the card-guessing task. After an anticipatory rise between the cue and the outcome, the response peaks and then rapidly decreases below baseline for punishment trials, whereas there is a more sustained response for reward trials (Delgado, Locke, Stenger, & Fiez, 2003; Delgado,

Table 1. Regions Identified from the Blocked Design [r]−[l] Task

Region of Activation	BA	Size (No. voxels)	Peak Talairach Coordinates (x, y, z)	Maximum F Value
Blocked design [r]–[l] task				
Main effect of block type				
Active > rest				
Middle frontal gyrus (L)	46	5	−40, 24, 23	25.4
Inferior frontal gyrus (R)	13	7	34, 24, 8	23.8
Insula (L)	13	3	−32, 15, 8	19.9
Insula/transverse temporal gyri (L)	13/41	29	−40, −22, 12	84.8
Insula/superior temporal gyrus (R)	13/22	19	53, −8, 4	37.1
Inferior occipital gyrus (R)	18	8	30, −88, −4	26.4
Thalamus (L)		3	−14, −8, 12	25.8
Midbrain/thalamus		23	−10, −19, −4	30.3
Rest > active				
Precentral gyrus (L)	4	4	−43, −15, 38	33.6
Precentral gyrus (R)	6	17	9, −11, 30	29.2
Superior frontal gyrus	9	68	−6, 53, 27	59.7
Inferior frontal gyrus (L)	47	8	−31, 20, −8	26.6
Anterior cingulate cortex (R)	32	6	8, 44, 0	27.3
Temporoparietal cortex (R)	19/39	132	45, −58, 30	79.3
Temporoparietal cortex (L)	19/39	196	−40, −83, 15	170.8
Middle temporal gyrus (L)	21	3	−58, −37, 0	28.4
Middle temporal gyrus (L)	21	30	−51, −3, −20	41.8
Middle temporal gyrus (R)	21	9	53, −11, −12	43.3
Precuneus/posterior cingulate cortex (L)	7/31	97	−7, −50, 30	93.5
Parahippocampal gyrus (R)		17	34, −14, −12	33.9
Amygdala (L)		4	−25, −7, −16	33.5
Blocked design [r]–[l] task				
Interaction of Block Type × Feedback Presence				
Caudate nucleus (L)		3	−6, 8, 8	37.4
Caudate nucleus (R)		13	8, 8, 8	109.6
Parahippocampal gyrus (R)		9	30, −18, −20	47.8

$p < .001$, cluster threshold of three voxels. BA = Brodmann's area; L = left; R = right.

Nystrom, et al., 2000). Again we examined periods T7 and T8. Two-tailed paired t tests revealed significant reward/punishment differences at these two periods for the left caudate, $t(8) = 4.03$, $p < .01$ at T7, $t(8) = 4.00$, $p < .01$ at T8, whereas in the right caudate, the contrast is significant at T8, $t(8) = 3.44$, $p < .01$, and there is a trend toward significance at T7, $t(8) = 2.28$, $p = .05$. Finally, significant activation in the parahippo-campal cluster, was not found for either the [r]−[l] or guessing task.

Regions Identified from the Event-related Datasets

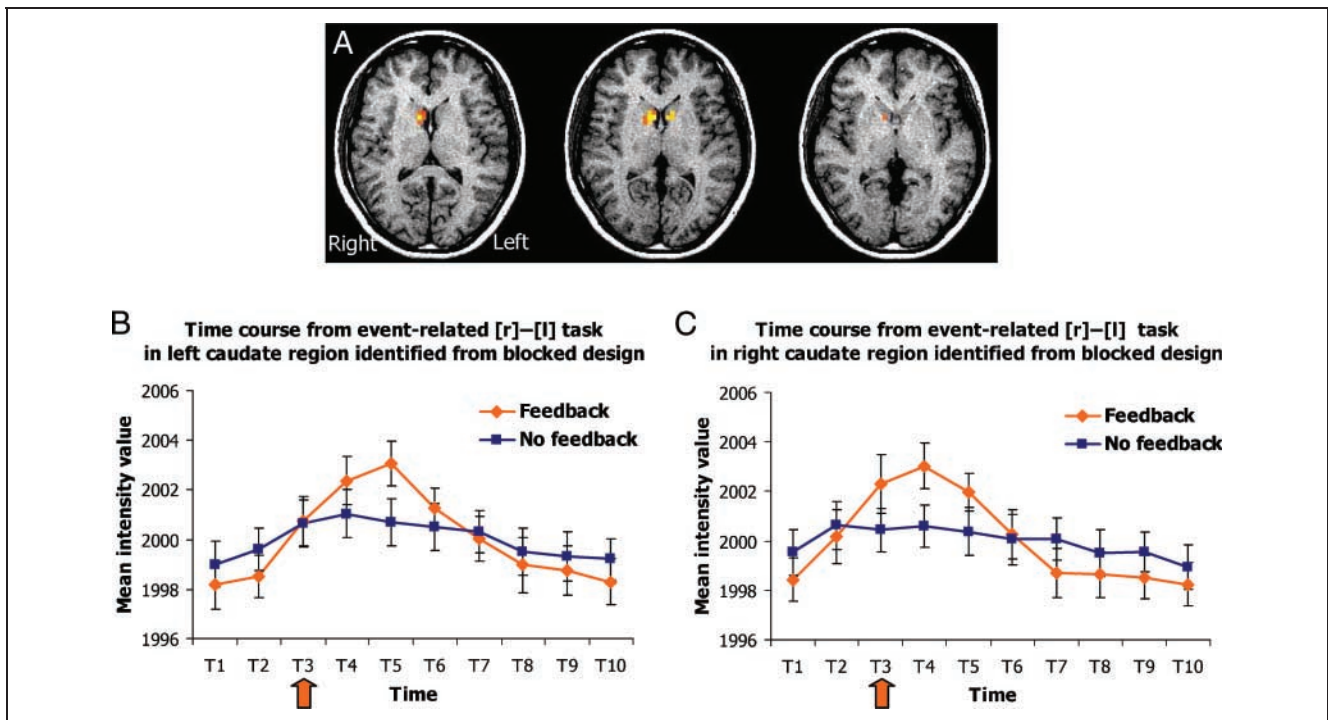Because the activation clusters from the blocked design data were identified based on their sensitivity to

Figure 2. Activation is modulated by feedback presence in the [r]–[l] identification task in the left and right caudate nuclei clusters identified from the blocked design analysis. Bilateral activation clusters in the caudate nuclei (A) show a significant Block Type × Feedback Presence interaction in the [r]–[l] identification task ($p < .001$; contiguity threshold of three voxels). Images are left–right reversed. Greater activation is observed in the left (B) and right (C) caudate nucleus during feedback trials (orange) in the [r]–[l] identification task than during the no-feedback trials (blue). Each period (T1, T2, etc.) represents a 1.5-sec image acquisition. The orange arrows indicate the time at which the outcome was revealed.

feedback presence rather than feedback valence, they were not ideally suited to revealing regions that might be differentially sensitive to positive versus negative feedback. Therefore, we performed voxelwise analyses of the event-related datasets to examine regions that might show such a valence sensitivity. A voxelwise ANOVA was performed on both the feedback trial data from the [r]–[l] identification task and the card-guessing task data, with valence (correct/incorrect or monetary reward/punishment) and time (1.5-sec periods T1–T10) as factors. The resulting activation clusters are listed in Table 2, $F > 3.41$ for [r]–[l] task, $F > 3.58$ for card-guessing task, $p < .001$, cluster threshold of three voxels. Figure 4 displays the activation clusters showing a significant Valence × Time interaction in this analysis. The clusters for the card-guessing task and the [r]–[l] task with feedback overlapped in two areas: the right caudate nucleus and the ventromedial prefrontal cortex (vmPFC). It is also interesting to note that the anterior cingulate cortex was activated more on incorrect than correct trials in the [r]–[l] task; the peak activation coordinates for this region are very similar to those found when trials with negative feedback are contrasted to trials with positive feedback in a motion prediction task (Ullsperger & von Cramon, 2003). A conjunction analysis comparing positive outcomes from both tasks (monetary rewards and positive feedback) with negative

outcomes from both tasks (monetary punishments and negative feedback) revealed activation clusters similar to those shown in Figure 4, including the right and left caudate nuclei and the vmPFC ($p < .001$, cluster threshold of three voxels).

Additionally, several activation clusters were identified in temporal regions as showing a Valence × Time effect in the [r]–[l] task, but no such regions were identified for the card-guessing task. A Valence × Time effect could arise in the [r]–[l] task either from differential recruitment of an area whose activation is predictive of whether a subject will make a correct versus an incorrect response, or from the area responding differentially to positive and negative feedback. If the former explanation is correct, then we would expect the area to show differential activation between correct and incorrect trials even when no feedback is present, whereas we would not expect such an effect for no-feedback trials if the latter explanation is correct. Therefore, we coded the no-feedback trials for valence (correct vs. incorrect responses) and performed an ANOVA on this data. We did not observe any temporal regions showing a Valence × Time effect in this analysis.

Given that the right caudate nucleus showed a Valence × Time effect for both tasks, it is surprising that the caudate regions identified from the blocked [r]–[l] task did not show significant differentiation between
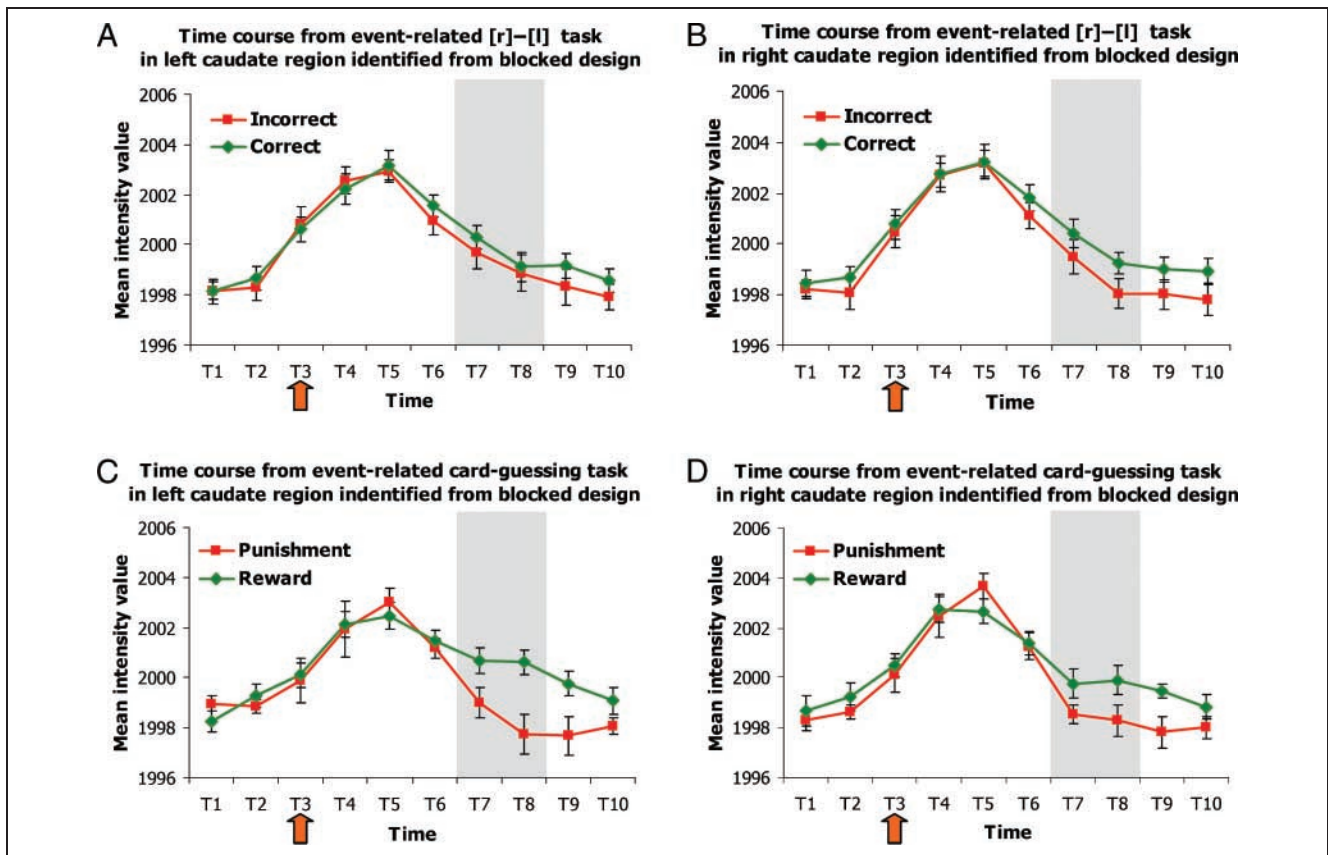
Figure 3. Valence effects in caudate activation clusters identified from blocked design analysis. Clusters showing a sensitivity to feedback presence in the blocked design data were examined for valence effects in the [r]–[l] task with feedback and in the card-guessing task. Trials with correct (green) and incorrect (red) feedback show only weak, nonsignificant differentiation for the [r]–[l] identification task in the left (A) and right (B) caudate regions. In the card-guessing task, feedback indicating monetary reward (green) produced more sustained activation than feedback indication monetary loss (red) in the left (C) and right (D) caudate regions. Throughout, each period (T1, T2, etc.) represents a 1.5-sec image acquisition. The orange arrows indicate the period at which the outcome was revealed. The gray area indicates the periods at which the largest and most reliable valence differences have been observed in previous work using the card-guessing task (Delgado, Locke, et al., 2003; Delgado, Nystrom, et al., 2000).

positive and negative feedback when applied to the event-related [r]–[l] dataset. Although the right caudate Valence × Time cluster is in a very similar location as the cluster identified from the blocked design analysis, the former cluster extends further in the ventral direction and the latter extends further in the dorsal direction. Thus, within the head of the caudate voxels may vary in their relative sensitivities to feedback presence and feedback valence.

Time Courses of Activation

In the caudate clusters identified as showing a significant Valence × Time effect for the [r]–[l] task with feedback and for the card-guessing task, similar time course patterns are observed. In both cases, positive outcomes result in more sustained activation, whereas negative outcomes show a sharper decrease in activation following the response peak (Figure 5). For the [r]–[l] task, a two-tailed paired t test reveals a significant difference at

T8 between correct and incorrect trials, $t(12) = 2.84$, $p < .05$. For the card-guessing task, reward and punishment trials differentiate significantly at both T7, $t(8) = 2.94$, $p < .05$, left caudate; $t(8) = 3.66$, $p < .01$, right caudate, and T8, $t(8) = 2.97$, $p < .05$, left caudate; $t(8) = 7.40$, $p < .0001$, right caudate.

Because we have only inconsistently observed a significant Valence × Time effect in the vmPFC in previous work using the card-guessing task (Delgado, Stenger, & Fiez, 2004; Delgado, Locke, et al., 2003; Delgado, Nystrom, et al., 2000), we did not have a priori hypotheses regarding activation in this region. This area is prone to susceptibility artifact, which can cause signal dropout. Indeed, the overall mean signal intensity in our vmPFC activation clusters was lower in comparison to our caudate activation clusters; the vmPFC/caudate signal ratio was 0.67 for the event-related [r]–[l] task and 0.76 for the card-guessing task. It is possible that varying degrees of susceptibility artifact among different studies has led to inconsistent findings in the region. Nevertheless, it is interesting that in our current study, the effect

Table 2. Regions Displaying an Interaction of Valence by Time

Region of Activation	BA	Size (No. Voxels)	Peak Talairach Coordinates (x, y, z)	Maximum F Value
Event-related [r]–[l] task, feedback trials only				
Interaction of Valence × Time				
Correct trials > incorrect trials				
Ventromedial prefrontal cortex	10/11	9	−3, 44, −12	4.1
Middle temporal gyrus (L)	21	3	−62, −37, −4	3.8
Middle temporal gyrus (R)	21	5	56, −21, −4	4.7
Middle temporal gyrus (R)	37	5	50, −63, 4	4.1
Inferior occipital gyrus (L)	19	6	−43, −75, −4	3.9
Caudate nucleus (R)		37	16, 16, 4	6.4
Incorrect trials > correct trials				
Precentral gyrus (R)	6	6	34, −1, 30	3.7
Anterior cingulate cortex	32	8	6, 23, 34	4.5
Superior temporal gyrus/insula (L)	22/13	8	−47, −14, 0	5.3
Event-related card-guessing task				
Interaction of Valence × Time				
Reward > punishment				
Ventromedial prefrontal cortex	10	56	1, 48, −4	6.2
Middle frontal gyrus (R)	46	3	45, 36, 20	4.0
Middle temporal gyrus (L)	37	5	−51, −50, −8	4.0
Posterior cingulate gyrus	31/23	16	4, −42, 28	4.6
Precuneus (L)	19	5	−28, −79, 42	4.8
Caudate nucleus (L)		32	−14, 16, 4	7.8
Caudate nucleus (R)		14	11, 16, 4	5.2
Thalamus		7	1, −11, 12	4.8
Punishment > reward				
Medial frontal gyrus/anterior cingulate cortex (L)	9/32	7	−3, 41, 16	4.2
Insula/postcentral gyrus (R)	40	3	50, −19, 16	3.9
Lingual gyrus/cuneus (L)	17	3	−14, −91, −4	4.0

$p < .001$, cluster threshold of three voxels. BA = Brodmann's area; L = left; R = right.

in the vmPFC was consistent between the event-related [r]–[l] and card-guessing tasks. Figure 6 shows that the activation time courses for this region are similar for the two tasks; the activation peaks at T3 decreases and then slowly increases, with the signal recovering more quickly for correct or reward trials than for incorrect or punishment trials. This pattern bears some resemblance to that shown in the caudate nucleus in that the peak is relatively early in the trial, and the greater signal for positively versus negatively valenced events emerges later in the trial rather than at the peak. The shape of the activation time courses in the vmPFC is less typical of a hemodynamic response, however, and future work will

be necessary before strong claims can be made about the role of this region in our tasks.

Behavioral Results from Extended Training

All subjects performed one 530-trial training session per day for the 6 days following their scanning session. Half trained with feedback and half trained with no feedback, allowing us to address the question of whether training with and without feedback eventually leads to different levels of performance. The results replicated those of McCandliss et al. (2002); that is, there was substantial improvement for the feedback group but not for the no-
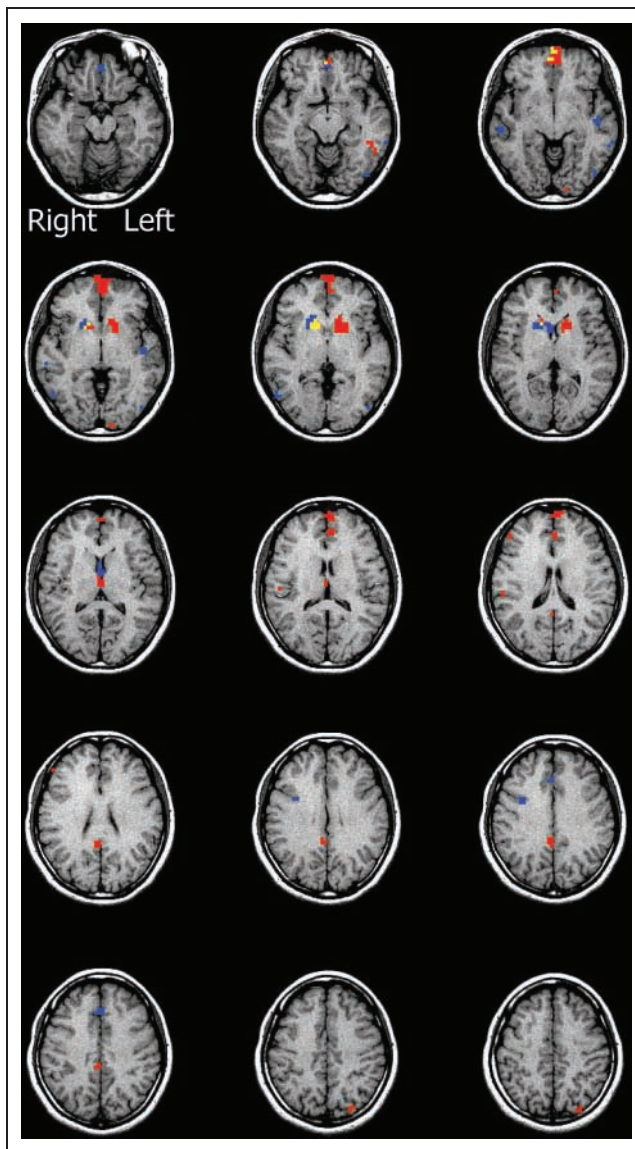
Figure 4. Brain regions showing a valence by time interaction in the [r]–[l] task with feedback and in the card-guessing task. Regions displaying a Valence × Time interaction are shown, color-coded by experiment ($p < .001$; contiguity threshold of three voxels). Red = significant effect for the card-guessing task. Blue = significant effect for the [r]–[l] task with feedback. Yellow = significant effect for both tasks. Images are left–right reversed.

feedback group. The no-feedback group's performance was similar to that of a control group that took the pre- and posttests but did not train, whereas the feedback group's performance was significantly greater than that of both the control group and the no-feedback group.

DISCUSSION

In our experimental paradigm, the only experimental difference between the two [r]–[l] identification conditions is that participants were given feedback about their performance in one condition but not in the other. Yet

our behavioral findings replicate previous work that has shown that participants' performance on this task increases dramatically after a few training sessions if they are given feedback, but relatively little learning occurs without that feedback (McCandliss et al., 2002). It is reasonable to expect that this difference in performance must be reflected in some critical difference in neural processing during training. The pattern of findings in our [r]–[l] identification task lends support to this notion by the fact that the caudate nucleus was more strongly activated in the feedback condition than in the
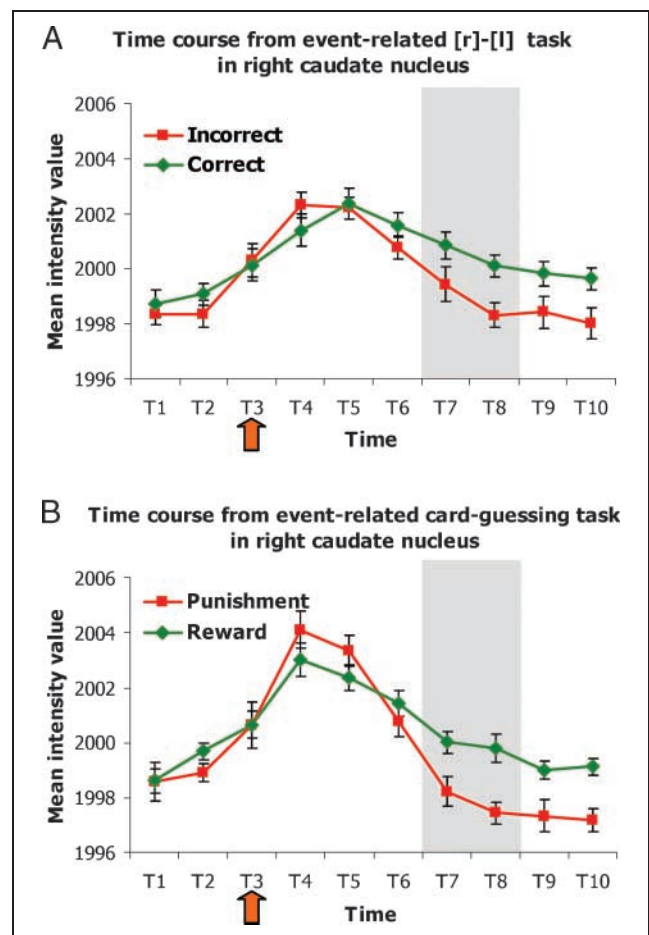

Figure 5. Activation time courses in caudate activation clusters showing a valence by time interaction. In the right caudate activation cluster showing a valence by time period interaction in the event-related [r]–[l] task with feedback (A), positive feedback indicating a correct response (green) produced more sustained activation than negative feedback indicating an incorrect response (red). Similarly, in the right caudate activation cluster showing a valence by time period interaction in the card guessing task (B), feedback indicating monetary reward (green) produced more sustained activation than feedback indication monetary loss (red). Throughout, each period (T1, T2, etc.) represents a 1.5-sec image acquisition. The orange arrows indicate the period at which the outcome was revealed. The gray area indicates the periods at which the largest and most reliable valence differences have been observed in previous work using the card guessing task (Delgado et al., 2003; Delgado et al., 2000).
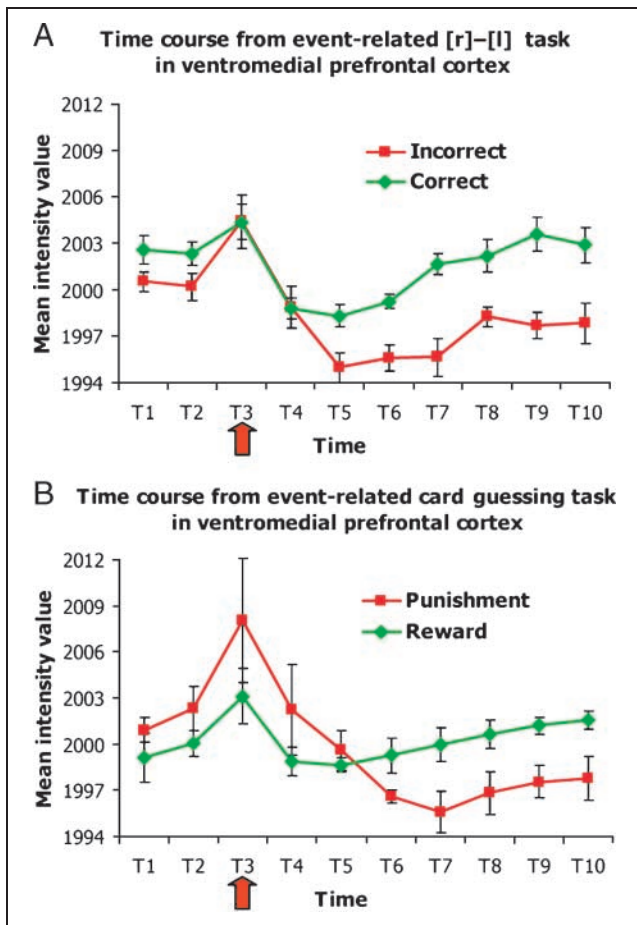
Figure 6. Activation time courses in ventromedial prefrontal activation clusters showing a valence by time interaction. In the ventromedial prefrontal activation cluster showing a valence by time interaction in the event-related [r]–[l] task with feedback (A), negative feedback indicating an incorrect response (red) produced a larger decrease in activation than positive feedback indicating a correct response (green). Similarly, in the ventromedial PFC cluster showing a valence by time period interaction in the card guessing task (B), activation decreased below baseline for feedback indicating monetary loss (red), while the signal recovered more quickly following feedback indicating monetary reward (green). Throughout, each period (T1, T2, etc.) represents a 1.5-sec image acquisition. The orange arrows indicate the time at which the outcome was revealed.

no-feedback condition, and, furthermore, that it differentiated between positive and negative feedback.

This finding provides a novel perspective on studies of language by suggesting possible mechanisms that may contribute to plasticity in language learning. Recruitment of the striatum may enhance other learning mechanisms, such as Hebbian learning, by acting as an informative signal that guides listeners to better attune to cues that differentiate the two sound tokens (Callan et al., 2003; McCandliss et al., 2002). Potentially, the Valence × Time effects that we observed in temporal areas may reflect the interface of the feedback processing system with the language system. For example, the left superior temporal gyrus, which we found to exhibit

more activation on incorrect trials than correct trials, has been implicated as a locus of plasticity in the brains of listeners who have improved their ability to make a nonnative phonetic distinction (Golestani & Zatorre, 2004; Callan et al., 2003). That these areas did not respond differentially on correct versus incorrect trials in the no-feedback condition of our experiment suggests that the effect was probably not driven solely by responding accurately or inaccurately. Although it is possible that this activation was not present during the no-feedback runs because of differences in overall approach or motivation between the feedback versus no-feedback runs, a more likely explanation is that the feedback was exerting a top-down influence on speech processing, perhaps supporting the changes in phonetic representations necessary for learning to take place.

Although our work was done in adults learning a second language, it warrants some speculation into influences on first language learning in infancy. Much research on first language learning focuses on how the language system self-organizes based on the linguistic environment. However, there is some evidence that feedback may play a role as well. Specifically, one study showed that positive social feedback facilitates complex vocal behavior in infants. When mothers contingently respond to their infants' vocalizations by smiling, moving closer to and touching their infants, the infants increase their production of mature vocalizations, whereas infants whose mothers' social behavior is equivalent but not contingently linked to vocalizations do not show similar changes in vocal behavior (Goldstein, King, & West, 2003). Several researchers have argued that birdsong learning occurs in a process that parallels human speech learning, and that cross-species comparisons between the two processes may be informative (Doupe, Perkel, Reiner, & Stern, 2005; Kuhl, 2004; Goldstein et al., 2003). Interestingly, social feedback has been shown to facilitate birdsong learning (Kuhl, 2004; Goldstein et al., 2003). What the neural basis of this effect in birds might be is still an open question; however, the anterior forebrain pathway of songbirds, which is a specialized cortical–basal ganglia circuit that is similar to that of mammals, is necessary for song learning (but not song production) to occur (Doupe et al., 2005). Although any conclusions drawn from these findings must be highly speculative, the cross-species similarities are intriguing and provide suggestive evidence that feedback may be an important element in natural language learning.

That the caudate responds to a wide array of affective reinforcers puts it in a good position to facilitate learning from the many types of feedback one may encounter (e.g., verbal, social, monetary, etc.). Beyond the specific domain of language, our finding fits in well with the hypothesis that dopamine released in the striatum may act as a reinforcement signal, allowing behavior to be adaptively modulated to maximize future reward

(O'Doherty et al., 2004; Holroyd & Coles, 2002; Montague & Berns, 2002; Barto, 1995). Indeed, the caudate has been found to be activated more on trials with positive than negative performance feedback in a motion prediction task and a time estimation task (Nieuwenhuis, Slagter, Geusau, Heslenfeld, & Holroyd, 2005; van Veen, Holroyd, Cohen, Stenger, & Carter, 2004; Ullsperger & von Cramon, 2003). By directly comparing two seemingly disparate tasks, we found that the neural response to feedback in the caudate nuclei is similar in a learning situation and a guessing task with monetary consequences. This shows that in the brains of highly motivated individuals, performance feedback has an effect analogous to that of winning or losing money. The caudate nuclei appear to be sensitive to outcomes with intrinsic, as well as extrinsic, value. Based on this and previous imaging work, it is possible to outline several task properties that are necessary to drive robust activation in the caudate. First, the task must include some form of goal-directed action; strong activation is found only when there is a perception that a positive or negative outcome is contingent upon one's behavior (Tricomi et al., 2004). Second, our study indicates that goal-directed action without knowledge of the outcome does not strongly activate the caudate; rather, the action must be performed in a context in which feedback about the action is expected. Third, there must be an incentive that makes an outcome important to the individual performing the task (Delgado, Stenger, & Fiez, 2004). This incentive may be a typical reward or something more abstract, such as signaling good performance on a task the individual would like to be able to perform well.

Additionally, several studies have found that reward-related activity in the striatum is greatest when the reward is unpredictable (McClure, Berns, & Montague, 2003; Berns et al., 2001). In our study, the outcome valence was unpredictable for both the card-guessing and [r]–[l] identification tasks. However, as the learners mastered the [r]–[l] task, the outcome would become more and more predictable and the feedback less and less informative; therefore the caudate might not continue to show a strong effect of feedback presence. Indeed, it should be noted that work from other investigators indicates that neural processing in the caudate nuclei appears to change over the course of learning (Delgado, Miller, Inati, & Phelps, 2005; Pasupathy & Miller, 2005; Seger & Cincotta, 2005; Haruno et al., 2004; Packard & Knowlton, 2002).

Taken together, these studies suggest a role for the caudate in flexibly guiding behavior based on response-dependent feedback that is valued by the individual, especially when the information provided by the feedback is most informative (i.e., unpredictable). This conclusion has implications for interpreting work on the role of the caudate in nondeclarative learning. Striatal activation has been found in fMRI studies using probabilistic learning tasks (Aron et al., 2004; Seger & Cincotta,

2002; Poldrack, Clark, et al., 2001; Poldrack, Prabhakaran, Seger, & Gabrieli, 1999), and this activation was found to occur only when such a probabilistic task was feedback-based (Poldrack, Clark, et al., 2001). Moreover, recent neuropsychological work indicates that patients with Parkinson's disease, which disrupts striatal function, are impaired on a feedback-based version of a probabilistic task, but not a nonfeedback version of the task (Shohamy et al., 2004). Patients with Parkinson's disease are not impaired, however, on artificial grammar learning and prototype learning tasks, which involve category learning but which are not learned via feedback (Reber & Squire, 1999). Our finding that the striatum is recruited in a nonprobabilistic, feedback-based perceptual learning task further supports the idea that it may not be the probabilistic nature of certain tasks that causes Parkinson's patients to be impaired on them, but rather the feedback processing they necessitate.

An open question is whether there are functional distinctions within the striatum in feedback processing. For example, much research has implicated the ventral striatum in processing reward-related information (e.g., Pagnoni, Zink, Montague, & Berns, 2002; Berns et al., 2001; Knutson, Adams, Fong, & Hommer, 2001). However, whereas activation in the caudate nucleus has been found to depend on the behavioral relevance of a stimulus, activation in the nucleus accumbens, a key region in the ventral striatum, does not show this effect (Zink, Pagnoni, Martin, Dhamala, & Berns, 2003). Therefore, it makes sense that the caudate should be more involved in processing performance-dependent feedback than the ventral striatum. It has also been suggested that the body of the caudate may be involved in classification learning independently of the role of the head of the caudate in feedback processing (Seger & Cincotta, 2005). In visual classification research, the role of the tail and sometimes body of the caudate is often emphasized (e.g., Seger & Cincotta, 2005; Ashby & Ell, 2002; Maddox, Molis, & Diehl, 2002). Similarly, Ashby and Ell (2002) posit that auditory stimulus-category associations should activate the body of the caudate. An investigation of our time courses of the activation in the body of the caudate showed that it did display hemodynamic responses, but this was the case for both the card-guessing task and the [r]–[l] task, and the responses did not differentiate based on condition. Therefore, it seems that the body of the caudate is playing a role in our tasks, but our data do not allow us to make strong inferences about the nature of this role.

In summary, our results complement previous work that has demonstrated that simple visual feedback may act as a reward or punishment and activate the caudate nucleus even when learning is not possible (Delgado, Stenger, & Fiez, 2004; Elliott, Frith, & Dolan, 1997). When feedback is meaningful, such caudate activation may serve to facilitate learning. This finding supports the notion that striatum-dependent nondeclarative learning

and the processing of reward-related information are related, in that feedback processing is fundamental to both of these roles. Furthermore, our study also impacts our understanding of language learning, indicating a way that feedback can have a top-down effect on speech perception. By using a paradigm that was motivated by different research domains, we were able to gain a more expansive understanding of how reward processing in the striatum may cut across seemingly different cognitive tasks.

Acknowledgments

REFERENCES

Aron, A. R., Shohamy, D., Clark, J., Myers, C., Gluck, M. A., & Poldrack, R. A. (2004). Human midbrain sensitivity to cognitive feedback and uncertainty during classification learning. *Journal of Neurophysiology, 92,* 1144–1152.

Ashby, F. G., & Ell, S. W. (2002). Single versus multiple systems of learning and memory. In J. Wixted & H. Pashler (Eds.), *Stevens' handbook of experimental psychology: Volume 4, Methodology in experimental psychology* (3rd ed. pp. 655–692). New York: Wiley.

Barto, A. G. (1995). Adaptive critics and the basal ganglia. In J. C. Houk, J. L. Davis, & D. G. Beiser (Eds.), *Models of information processing in the basal ganglia* (pp. 215–232). Cambridge: MIT Press.

Berns, G. S., McClure, S. M., Pagnoni, G., & Montague, P. R. (2001). Predictability modulates human brain response to reward. *Journal of Neuroscience, 21,* 2793–2798.

Bradlow, A. R., Pisoni, D. B., Yamada, R. A., & Tohkura, Y. (1997). Training the Japanese listener to identify English [r] and [l]. IV: Some effects of perceptual learning on speech production. *Journal of the Acoustical Society of America, 101,* 2299–2310.

Breiter, H. C., Aharon, I., Kahneman, D., Dale, A., & Shizgal, P. (2001). Functional imaging of neural responses to expectancy and experience of monetary gains and losses. *Neuron, 30,* 619–639.

Callan, D. E., Tajima, K., Callan, A. M., Kubo, R., Masaki, S., & Akahane Yamada, R. (2003). Learning-induced neural plasticity associated with improved identification performance after training of a difficult second-language phonetic contrast. *Neuroimage, 19,* 113–124.

Cox, R. W. (1996). AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages. *Computers and Biomedical Research, 29,* 162–173.

Delgado, M. R., Locke, H. M., Stenger, V. A., & Fiez, J. A. (2003). Dorsal striatum responses to reward and

punishment: Effects of valence and magnitude manipulations. *Cognitive, Affective & Behavioral Neuroscience, 3,* 27–38.

Delgado, M. R., Miller, M. M., Inati, S., & Phelps, E. A. (2005). An fMRI study of reward-related probability learning. *Neuroimage, 24,* 862–873.

Delgado, M. R., Nystrom, L. E., Fissell, C., Noll, D. C., & Fiez, J. A. (2000). Tracking the hemodynamic responses to reward and punishment in the striatum. *Journal of Neurophysiology, 84,* 3072–3077.

Delgado, M. R., Stenger, V. A., & Fiez, J. A. (2004). Motivation-dependent responses in the human caudate nucleus. *Cerebral Cortex, 14,* 1022–1030.

Doupe, A. J., Perkel, D. J., Reiner, A., & Stern, E. A. (2005). Birdbrains could teach basal ganglia research a new song. *Trends in Neurosciences, 28,* 353–363.

Elliott, R., Friston, K. J., & Dolan, R. J. (2000). Dissociable neural responses in human reward systems. *Journal of Neuroscience, 20,* 6159–6165.

Elliott, R., Frith, C. D., & Dolan, R. J. (1997). Differential neural response to positive and negative feedback in planning and guessing tasks. *Neuropsychologia, 35,* 1395–1404.

Fissell, K., Tseytlin, E., Cunningham, D., Iyer, K., Carter, C. S., Schneider, W., & Cohen, J. D. (2003). A graphical computing environment for neuroimaging analysis. *Neuroinformatics, 1,* 111–125.

Forman, S. D., Cohen, J. D., Fitzgerald, M., Eddy, W. F., Mintun, M. A., & Noll, D. C. (1995). Improved assessment of significant activation in functional magnetic resonance imaging (fMRI): Use of a cluster-size threshold. *Magnetic Resonance in Medicine, 33,* 636–647.

Goldstein, M. H., King, A. P., & West, M. J. (2003). Social interaction shapes babbling: Testing parallels between birdsong and speech. *Proceedings of the National Academy of Sciences, U.S.A., 100,* 8030–8035.

Golestani, N., & Zatorre, R. J. (2004). Learning new sounds of speech: Reallocation of neural substrates. *Neuroimage, 21,* 494–506.

Haruno, M., Kuroda, T., Doya, K., Toyama, K., Kimura, M., Samejima, K., Imamizu, H., & Kawato, M. (2004). A neural correlate of reward-based behavioral learning in caudate nucleus: A functional magnetic resonance imaging study of a stochastic decision task. *Journal of Neuroscience, 24,* 1660–1665.

Hikosaka, O., Sakamoto, M., & Usui, S. (1989). Functional properties of monkey caudate neurons. III. Activities related to expectation of target and reward. *Journal of Neurophysiology, 61,* 814–832.

Holroyd, C., & Coles, M. (2002). The neural basis of human error processing: Reinforcement learning, dopamine, and the error-related negativity. *Psychological Review, 109,* 679–709.

Jamieson, D. G., & Morosan, D. E. (1986). Training non-native speech contrasts in adults: Acquisition of English /ð/–/θ/ contrast by francophones. *Perception & Psychophysics, 40,* 205–215.

Knutson, B., Adams, C. M., Fong, G. W., & Hommer, D. (2001). Anticipation of increasing monetary reward selectively recruits nucleus accumbens. *Journal of Neuroscience, 21,* 1–5.

Knutson, B., Westdorp, A., Kaiser, E., & Hommer, D. (2000). fMRI visualization of brain activity during a monetary incentive delay task. *Neuroimage, 12,* 20–27.

Kuhl, P. K. (2004). Early language acquisition: Cracking the speech code. *Nature Reviews Neuroscience, 5,* 831–843.

Lively, S. E., Logan, J. S., & Pisoni, D. B. (1993). Training Japanese listeners to identify English [r] and [l]. II: The

role of phonetic environment and talker variability in learning new perceptual categories. *Journal of the Acoustical Society of America, 94,* 1242–1255.

Logan, J. S., Lively, S. E., & Pisoni, D. B. (1991). Training Japanese listeners to identify English [r] and [l]: A first report. *Journal of the Acoustical Society of America, 89,* 874–886.

Maddox, W. T., Molis, M. R., & Diehl, R. L. (2002). Generalizing a neuropsychological model of visual categorization to auditory categorization of vowels. *Perception & Psychophysics, 64,* 584–597.

McCandliss, B. D., Fiez, J. A., Protopapas, A., Conway, M., & McClelland, J. L. (2002). Success and failure in teaching the [r]–[l] contrast to Japanese adults: Predictions of a Hebbian model of plasticity and stabilization in spoken language perception. *Cognitive, Affective & Behavioral Neuroscience, 2,* 89–108.

McClure, S. M., Berns, G. S., & Montague, P. R. (2003). Temporal prediction errors in a passive learning task activate human striatum. *Neuron, 38,* 339–346.

Miyawaki, K., Strange, W., Verbrugge, R. R., Liberman, A. M., Jenkins, J. J., & Fujimura, O. (1975). An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English. *Perception & Psychophysics, 18,* 331–340.

Montague, P. R., & Berns, G. S. (2002). Neural economics and the biological substrates of valuation. *Neuron, 36,* 265–284.

Morosan, D. E., & Jamieson, D. G. (1989). Evaluation of a technique for training new speech contrasts: Generalization across voices, but not word-position or task. *Journal of Speech and Hearing Research, 32,* 501–511.

Nieuwenhuis, S., Slagter, H. A., von Geusau, N. J., Heslenfeld, D. J., & Holroyd, C. B. (2005). Knowing good from bad: Differential activation of human cortical areas by positive and negative outcomes. *European Journal of Neuroscience, 21,* 3161–3168.

O'Doherty, J., Dayan, P., Schultz, J., Deichmann, R., Friston, K., & Dolan, R. J. (2004). Dissociable roles of ventral and dorsal striatum in instrumental conditioning. *Science, 304,* 452–454.

Packard, M. G., & Knowlton, B. J. (2002). Learning and memory functions of the basal ganglia. *Annual Review of Neuroscience, 25,* 563–593.

Pagnoni, G., Zink, C. F., Montague, P. R., & Berns, G. S. (2002). Activity in human ventral striatum locked to errors of reward prediction. *Nature Neuroscience, 5,* 97–98.

Pasupathy, A., & Miller, E. K. (2005). Different time course of learning-related activity in the prefrontal cortex and striatum. *Nature, 433,* 873–876.

Poldrack, R. A., Clark, J., Pare-Blagoev, E. J., Shohamy, D., Moyano, J. C., Myers, C., & Gluck, M. A. (2001). Interactive memory systems in the human brain. *Nature, 414,* 546–550.

Poldrack, R. A., Prabhakaran, V., Seger, C. A., & Gabrieli, J. D. E. (1999). Striatal activation during acquisition of a cognitive skill. *Neuropsychology, 13,* 564–574.

Reber, P. J., & Squire, L. R. (1999). Intact learning of artificial grammars and intact category learning by patients with Parkinson's Disease. *Behavioral Neuroscience, 113,* 235–242.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-prime user's guide.* Pittsburgh: Psychological Software Tools.

Schultz, W. (1998). Predictive reward signal of dopamine neurons. *Journal of Neurophysiology, 80,* 1–27.

Seger, C. A., & Cincotta, C. M. (2002). Striatal activity in concept learning. *Cognitive, Affective & Behavioral Neuroscience, 2,* 149–161.

Seger, C. A., & Cincotta, C. M. (2005). The roles of the caudate nucleus in human classification learning. *Journal of Neuroscience, 25,* 2941–2951.

Shohamy, D., Myers, C. E., Grossman, S., Sage, J., Gluck, M. A., & Poldrack, R. A. (2004). Cortico-striatal contributions to feedback-based learning: Converging data from neuroimaging and neuropsychology. *Brain, 127,* 851–859.

Strange, W., & Dittmann, S. (1984). Effects of discrimination training on the perception of /r–l/ by Japanese adults learning English. *Perception & Psychophysics, 36,* 131–145.

Talairach, J., & Tournoux, P. (1988). *Co-planar stereotaxic atlas of the human brain: An approach to medical cerebral imaging.* Stuttgart, Germany: Thieme.

Tees, R. C., & Werker, J. F. (1984). Perceptual flexibility: Maintenance or recovery of the ability to discriminate non-native speech sounds. *Canadian Journal of Psychology, 38,* 579–590.

Tricomi, E. M., Delgado, M. R., & Fiez, J. A. (2004). Modulation of caudate activity by action contingency. *Neuron, 41,* 281–292.

Ullsperger, M., & von Cramon, D. Y. (2003). Error monitoring using external feedback: Specific roles of the habenular complex, the reward system, and the cingulate motor area revealed by functional magnetic resonance imaging. *Journal of Neuroscience, 23,* 4308–4314.

van Veen, V., Holroyd, C. B., Cohen, J. D., Stenger, V. A., & Carter, C. S. (2004). Errors without conflict: Implications for performance monitoring theories of anterior cingulate cortex. *Brain and Cognition, 56,* 267–276.

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior & Development, 7,* 49–63.

Woods, R. P., Cherry, S. R., & Mazziotta, J. C. (1992). Rapid automated algorithm for aligning and reslicing PET images. *Journal of Computer Assisted Tomography, 16,* 620–633.

Woods, R. P., Mazziotta, J. C., & Cherry, S. R. (1993). MRI–PET registration with automated algorithm. *Journal of Computer Assisted Tomography, 17,* 536–546.

Zink, C. F., Pagnoni, G., Martin, M. E., Dhamala, M., & Berns, G. S. (2003). Human striatal response to salient nonrewarding stimuli. *Journal of Neuroscience, 23,* 8092–8097.