

# Neural Correlates of Specific and General Pavlovian-to-Instrumental Transfer within Human Amygdalar Subregions: A High-Resolution fMRI Study

Charlotte Prévost,<sup>1,4</sup> Mimi Liljeholm,<sup>1,2</sup> Julian M. Tyszka,<sup>3</sup> and John P. O'Doherty<sup>1,2,4</sup>

<sup>1</sup>Division of Humanities and Social Sciences, <sup>2</sup>Computation and Neural Systems Program, and <sup>3</sup>Division of Biology, California Institute of Technology, Pasadena, California 91125, and <sup>4</sup>Trinity College Institute of Neuroscience and School of Psychology, Trinity College Dublin, Dublin 2, Ireland

It is widely held that the interaction between instrumental and Pavlovian conditioning induces powerful motivational biases. Pavlovian-Instrumental Transfer (PIT) is one of the key paradigms demonstrating this effect, which can further be decomposed into a general and specific component. Although these two forms of PIT have been studied at the level of amygdalar subregions in rodents, it is still unknown whether they involve different areas of the human amygdala. Using a high-resolution fMRI (hr-fMRI) protocol optimized for the amygdala in combination with a novel free operant task designed to elicit effects of both general and specific PIT, we demonstrate that a region of ventral amygdala within the boundaries of the basolateral complex and the ventrolateral putamen are involved in specific PIT, while a region of dorsal amygdala within the boundaries of the centromedial complex is involved in general PIT. These results add to a burgeoning literature indicating different functional contributions for these different amygdalar subregions in reward-processing and motivation.

## Introduction

Pavlovian and instrumental processes interact in at least two ways. In specific Pavlovian-to-instrumental transfer (specific PIT), a Pavlovian stimulus associated with a particular outcome selectively enhances the performance of responses associated with that specific outcome. In general PIT, a Pavlovian cue engenders a general increase in the vigor of instrumental responding, independently of the specific outcome involved (Dickinson and Balleine, 2002; Holland, 2004; Balleine and Killcross, 2006). For instance, imagine while on your way to a French restaurant, you notice a candy bar in a store window. A manifestation of specific PIT would be to go and buy the candy bar while general PIT would occur if the sight of the candy provided a stronger drive for you to run to the restaurant. These two different forms of PIT are proposed to reflect distinct forms of incentive processing: a sensory-specific system encoding events based on their specific sensory properties and a general system whereby the affective significance of an outcome is encoded independently of its specific sensory properties (Balleine and Killcross, 2006).

Animal studies have focused on the role of distinct subregions within the amygdala and ventral striatum in PIT, particularly the

basolateral complex (BLA) and the centromedial complex (CeN) for the former and the nucleus accumbens core and shell for the latter. Some studies have implicated the BLA (Blundell et al., 2001), while others found that lesions of the CeN, but not the BLA abolished this phenomenon (Blundell et al., 2001; Holland and Gallagher, 2003). Differences in the findings between these studies likely arise from the type of PIT investigated. Corbit and Balleine (2005) showed that lesions of the BLA spared general but abolished specific PIT whereas CeN lesions spared specific but abolished general PIT. Less is known about the neural mechanisms underpinning PIT in humans. Talmi et al. (2008) reported activity in the striatum and the amygdala in response to global PIT (the task did not discriminate specific or general PIT), while Bray et al. (2008) reported a contribution of the ventrolateral putamen but not the amygdala in specific PIT. However, in both studies the use of a conventional imaging protocol precluded segregation of function within the amygdala. Further, Talmi et al.'s study was not designed to simultaneously measure specific and general PIT while Bray et al.'s study reported specific PIT but did not find behavioral evidence of general PIT.

The goal of the present study was to investigate the role of human amygdalar subregions in specific and general PIT with fMRI. We used a novel free operant paradigm adapted to the fMRI scanner to enable us to measure general as well as specific PIT within the same participants in a manner directly analogous to rodent studies. A further innovation in the present study is the use of a high-resolution fMRI protocol (hr-fMRI) to enable segregation of activity within different subregions in the amygdala. We hypothesized that the BLA would contribute to specific PIT and that the CeN would contribute to general PIT.

Received Dec. 14, 2011; revised March 29, 2012; accepted April 30, 2012.

Author contributions: C.P., M.L., and J.P.O. designed research; C.P. performed research; J.M.T. contributed unpublished reagents/analytic tools; C.P. analyzed data; C.P., M.L., J.M.T., and J.P.O. wrote the paper.

This work was funded by Science Foundation Ireland Grant 08/IN.1/B1844 and a grant from the Gordon and Betty Moore Foundation to J.P.O. We thank Signe Bray for useful advice and Soyoung Park for helpful assistance with scanning.

Correspondence should be addressed to Charlotte Prévost, 1200 East California Boulevard, MC228-77, Pasadena, CA 91125. E-mail: prevost@caltech.edu.

DOI:10.1523/JNEUROSCI.6237-11.2012

Copyright © 2012 the authors 0270-6474/12/328383-08\$15.00/0

## Materials and Methods

### Participants

Twenty-six healthy (5 females), right-handed participants with a mean age of  $27.04 \pm 6.09$  took part in this study. All participants were free of neurological or psychiatric disorders and had normal or correct-to-normal vision. As the experiment involved food consumption, the eating attitudes test (EAT-26; see Garner et al., 1982) was administered before the experiment, which indicated no eating disorders in any of the participants (score:  $6.19 \pm 5.49$ ). Participants were asked to fast for at least 4 h before the experiment to ensure motivation for food reward, although they were allowed to drink water. Written informed consent was obtained from all participants, and the study was approved by the Caltech Institutional Review Board.

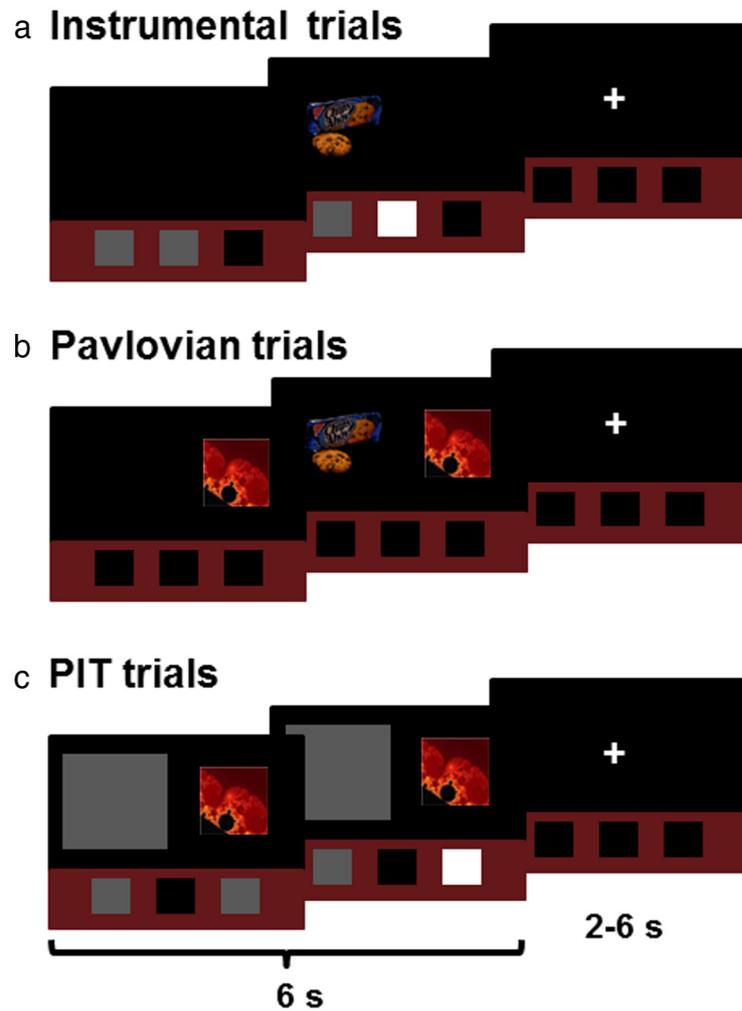
### Stimuli

Visual stimuli were presented via a projector positioned at the back of the room. Participants viewed a reflection of the projected image in a mirror attached to the scanner head coil. The food rewards consisted of sweet and salty snack foods (Chips Ahoy! cookies, Doritos ranch chips, Famous Amos cookies, Ghirardelli milk chocolate bar, Hostess Ho Hos, Kit Kat candy bar, Lindt Swiss chocolate bar, Mrs. Fields chocolate chip cookies, Oreo cookies, Rice Krispies Treats, Pepperidge Farm Milano cookies, Keebler Fudge Stripes cookies, Keebler Rainbow cookies, Toblerone chocolate bar, Twix caramel and chocolate bar). Participants were asked to provide subjective pleasantness ratings for each of the 15 different snack foods at the time of recruiting, and the experimenter selected three highly and equally valued food items for each participant based on their subjective ratings. Stimulus presentation and behavioral data acquisition were implemented in Matlab (The Mathworks) with the Cogent 2000 toolbox (Wellcome Department of Imaging Neuroscience).

### Training

Before being scanned, to familiarize participants with the set-up, they underwent a brief training on the instrumental and Pavlovian phases.

**Instrumental training.** During instrumental training trials, participants were instructed that two gray squares would light up signaling which two of three possible responses were available on that trial (Fig. 1*a*). They were asked to press any given response multiple times to get rewarded with food. Participants were explicitly told to press only one of the two available responses at a time. Every time one of these two keys was pressed, it flashed in white for 50 ms. In this free-operant task, each response was rewarded on a random ratio schedule of 0.1 as in Corbit and Balleine (2005), so that a food picture was displayed once every time a key was pressed a random number of times between 5 and 15 times (i.e., hence once every 10 times on average). Each of the three responses earned one of the three distinct food rewards (Table 1, Ins). As described in Table 1, there were three types of instrumental trials. Given that participants were told that rewards were contingent on their presses, key presses were also recorded during outcome presentation. Depending on their number of presses per response, participants could get more than one outcome per trial in this instrumental training. Each of the three responses earned the three distinct food rewards (Table 1, Ins). When-



**Figure 1.** PIT task illustration. Participants performed an instrumental (*a*) and a Pavlovian (*b*) task before performing the PIT task (*c*). During the instrumental trials, two of the three squares at the bottom of the screen changed color from black to gray for 6 s during which time participants were instructed to press any given button multiple times to get rewarded with food. The food picture corresponding to their response was displayed on one side of the screen for 1 s every 10 presses on average. In the Pavlovian trials, a conditioned stimulus was presented on one side of the screen for 6 s, and the food picture corresponding to the stimulus was displayed on the other side of the screen during the last second of the trial. During the PIT trials, the conditioned stimulus was presented simultaneously with two squares changing color for 6 s. Participants were instructed to press any given key multiple times to get rewarded with food, as in the instrumental trials. In this phase, the area where the food picture was usually displayed was hidden behind a gray square.

ever a certain food was depicted on the screen, the experimenter placed a small piece of food corresponding to the picture in the participant's left hand to be consumed immediately. Each trial lasted for 6 s and was followed by a jittered 2–6 s intertrial interval (ITI). A trial duration of 6 s was achieved to allow enough time for participants to generate a large number of presses, without making it effortful. This duration was fixed to allow reliable comparisons in the number of presses across the different trial types. Food pictures were presented for 1 s. The location of the food pictures was the same across all phases for a given participant but was counterbalanced across subjects so that it appeared on the left side of the screen for half of the subjects and on the right side of the screen for the other half of the subjects. This short instrumental training was comprised of 6 trials (2 of each type). The trial order was randomized within this phase as well as across participants.

**Pavlovian training.** Pavlovian training trials consisted of the presentation of associations between fractal images (conditioned stimuli, CSs) and outcomes, three of which were rewarding (food rewards) and one of which was neutral (no reward) (Fig. 1*b*; Table 1, Pav). There were four types of Pavlovian trials: stimulus 1 (S1) leading to outcome 1 (O1), S2 leading to O2, S3 leading to O3 and S4 leading to no outcome. Therefore,

**Table 1. Trial composition for the instrumental, Pavlovian, and PIT phases**

Phase	Number of trials	Pavlovian	Instrumental		Outcome	
Ins	10	—	R1	R2	O1	O2
		—	R1	R3	O1	O3
		—	R2	R3	O2	O3
Pav	10	S1	—	—	O1	—
		S2	—	—	O2	—
		S3	—	—	O3	—
		S4	—	—	—	—
PIT	5	S1	R1	R2	—	—
		S1	R1	R3	—	—
		S2	R1	R2	—	—
		S2	R2	R3	—	—
		S3	R1	R3	—	—
	10	S3	R2	R3	—	—
		S1	R2	R3	—	—
		S2	R1	R3	—	—
	10	S3	R1	R2	—	—
		S4	R1	R2	—	—
		S4	R1	R3	—	—
		S4	R2	R3	—	—

Ins, instrumental; Pav, Pavlovian, S1–S4, conditioned stimuli; R1–R3, three-button response actions; O1–O3, food rewards. Dashes represent absent events.

participants experienced four different outcomes during Pavlovian training, while they experienced only three in the instrumental trials. CSs were presented on one side of the screen for 6 s and rewards were delivered concomitantly during the last second of the trial (between 5 and 6 s) in a deterministic fashion. Again, every time a food picture was depicted on the screen, the experimenter placed a piece of food corresponding to the picture in the participant's left hand to be consumed immediately. The trial ended with an ITI of 2–6 s. The location of the fractal cues and food pictures was counterbalanced across participants. The assignment of the associations between fractal images and outcomes was randomized across participants. This short Pavlovian training was comprised of 4 trials (one of each type). The trial order was randomized within this phase as well as across participants.

### Task description

The scanned portion of the experiment consisted of an instrumental, a Pavlovian, and a PIT phase (Table 1).

**Instrumental phase.** This phase was exactly the same as during the training except that participants were told that whenever they saw a food picture, the experimenter would place the corresponding food in a cup which they could retrieve at the end of the experiment. The three black squares at the bottom of the screen corresponded to the three buttons on a response box (Current Designs) that the participants held in their right hand. Each trial type was presented 10 times, for a total of 30 trials and a total duration of 5 min (Table 1, Ins).

**Pavlovian phase.** Again, this phase was exactly the same as during the training except that participants were told that whenever they saw a food picture, the experimenter would place the corresponding food in a cup which they could retrieve at the end of the experiment. Each trial type was presented 10 times, for a total of 40 trials and a total duration of ~7 min (Table 1, Pav).

**PIT phase.** After the instrumental and Pavlovian phases, participants performed a series of transfer trials. During these trials, one of the Pavlovian cues was presented simultaneously with the two instrumental response options (Fig. 1c) and, as in instrumental trials, participants were given the opportunity to freely respond to either of the two available options. They were explicitly told that they would not be able to see any food picture during this phase, because the area where they were usually displayed would be hidden, but that they should assume that all reward deliveries would be as they had been during the previous phases. The reason for performing this phase in extinction (i.e., no outcomes delivered) was to allow assessment of the influence of the Pavlovian cues on instrumental responding without the confounding effects of the outcomes themselves (Bray et al., 2008). As previously described, PIT effects

can be further decomposed into outcome-specific and outcome-general PIT (Dickinson and Balleine, 2002; Holland, 2004; Corbit and Balleine, 2005). In specific PIT, the Pavlovian cue associated with a particular outcome enhances only an instrumental response associated with the same appetitive outcome. However, in general PIT, the Pavlovian cue enhances any instrumental response, even if the instrumental response is paired with a different appetitive outcome. Because this study was aimed at comparing the neural structures involved in these two distinct processes, we designed a task allowing us to behaviorally distinguish these phenomena. Thus, there were three different types of trials: specific, general and neutral trials (Table 1, PIT). To test for specific PIT, participants were given the opportunity to press two available keys previously paired with two particular reward outcomes, one of which also previously paired with the concurrently presented Pavlovian cue. For instance, response 1 (R1) previously paired with outcome 1 (O1) and response 2 (R2) previously paired with outcome 2 (O2) were made available in presence of stimulus 1 (S1) previously paired with O1 or stimulus 2 (S2), paired with O2. There were six subtypes of specific trials: R1 and R2 presented with S1, R1 and R3 presented with S1, R1 and R2 presented with S2, R2 and R3 presented with S2, R1 and R3 presented with S3 and R2 and R3 presented with S3. Evidence for specific PIT would be seen if the presence of the Pavlovian cue induced a higher response rate for the action associated with the same outcome as that Pavlovian cue. In the subsequent analysis, we pooled over all six specific trial subtypes, but differentiated between trials in which participants responded more to actions compatible with the Pavlovian outcome and those trials in which participants responded more to the action that was not compatible (as in Bray et al., 2008). There were 5 trials of each subtype, for a total of 30 specific trials. To test for general PIT, we also included general and neutral trials. In general trials, participants were given the opportunity to respond to two available responses, but this time the Pavlovian cue was previously associated with a different outcome that was not compatible with either response option. There were three subtypes of general trials: R2 and R3 presented with S1, R1 and R3 presented with S2 and R1 and R2 presented with S3. There were 10 trials of each subtype, for a total of 30 general trials. In neutral trials, participants were again given the opportunity to respond to two available responses but the Pavlovian cue was that associated with the neutral outcome. There were three subtypes of neutral trials: R1 and R2 presented with S4, R1 and R3 presented with S4 and R2 and R3 presented with S4. There were 10 trials of each subtype, for a total of 30 neutral trials. Evidence for general PIT would be seen if participants responded more to general than neutral trials, indicating that reward cues exert some general motivation over any instrumental response, acquired through their pairing with a reward outcome (as opposed to the neutral cue which has not been paired with a reward outcome and as such, has not acquired any motivational value). Each type of trial was presented 30 times for a total of 90 trials and a total duration of 15 min. The trial order was randomized within this phase as well as across participants.

### Behavioral measures

**Physiological fluctuations.** Although both heart rates and breathing rates were recorded during scanning to be further removed from time-series (see fMRI data analysis), only the time course derived from estimates of heart rate was used as a physiological index of conditioning.

**Response rates.** Response rates were recorded both during the instrumental and PIT phases.

**Affective evaluations of food pictures.** Before the start of the training and before and after the PIT phase, participants were asked to rate the pleasantness of the food pictures by moving a cursor along a scale from –5 to +5.

### Data acquisition

All magnetic resonance imaging was performed on a 3 tesla Magnetom Tim Trio scanner (Siemens Medical Solutions) equipped with a 32-channel receive-only phased array head coil. Since the focus of our study was on the amygdala, we acquired oblique axial T2\*-weighted echo planar images (EPI) with partial brain coverage centered on the medial temporal cortex while participants were performing the task. These im-

ages also encompassed the ventral striatum, ventral part of the prefrontal cortex, the insula, the hippocampus, the ventral part of the occipital lobe and the upper part of the cerebellum (among other regions). Twenty-two EPI slices with ascending interleaved ordering were acquired, with an isotropic voxel size of 1.8 mm (no slice gap); thereby providing ~4.6 times the volume resolution of a typical standard 3 mm isotropic voxel size acquisition. Other imaging parameters included: repetition time (TR) = 2000 ms; echo time (TE) = 41 ms; field of view = 180 × 180 × 39.6 mm; sampling matrix = 100 × 100 voxels. We discarded the first 3 EPI volumes before data processing and statistical analysis to minimize magnetization equilibration effects on the BOLD timeseries. Whole-brain high-resolution T1-weighted structural images (two repetitions, isotropic voxel size = 1 mm) and whole-brain T2\*-weighted images (three repetitions) were acquired for each participant. To address the problem of spatial EPI distortions which are particularly prominent in the medial temporal lobe (MTL) and especially in the amygdala, we also acquired dual gradient echo B<sub>0</sub> field maps. Finally, to account for the effects of physiological noise in the fMRI data, participants' cardiac and respiratory signals were recorded using the MRI system's pulse oximeter and respiratory bellows.

### Preprocessing

All EPI volumes (partial scans acquired while participants were performing the task and the three whole-brain functional scans acquired before the experiment) were corrected for differences in slice acquisition and spatially realigned. The mean whole-brain EPI was coregistered with the T1-weighted structural image, and subsequently, all the "partial" volumes were coregistered with the registered mean whole-brain EPI image. Partial volumes were then unwrapped using the gradient field maps. After the structural scan was normalized to a standard T1 template, the same transformation was applied to all the partial volumes with a resampled voxel size of 1 × 1 × 1 mm. To maximize the spatial resolution of our data, no spatial smoothing kernel was applied to the data. These preprocessing steps were performed using the statistical parametric mapping software SPM5 (Wellcome Department of Imaging Neuroscience).

**Amygdalae segmentation.** Amygdalae regions of interest (ROIs) were manually segmented for each participant by a single observer using a pen tablet (Wacom Intuos3 Graphics Tablet) in FSL View (FSL 4.1.2). This program allows magnification and the simultaneous viewing of volumes in coronal, sagittal and horizontal orientations. Amygdalae were manually outlined on each coronal image containing these regions using detailed tracing guidelines based on the Atlas of the Human Brain (Mai et al., 2008). Outlines were checked in horizontal and sagittal planes when they proved more valuable for the identification of structure boundaries. The anterior limit of the amygdala was defined using the horizontal and sagittal planes. The following guidelines were used: In its rostral part, the amygdala is bordered ventromedially by the entorhinal cortex, ventrally by the temporal horn of the lateral ventricle and subamygdaloid white matter and laterally by white matter of the temporal lobe. Midrostrocaudally, the amygdala increases in size and is bordered ventromedially by a thin tract of white matter separating the amygdala and the entorhinal cortex, laterally by the white matter of the temporal lobe and medially by the semiannular sulcus. Caudally, the amygdala is bordered dorsally by the substantia innominata and fibers of the anterior commissure, laterally by the putamen, ventrally by the temporal horn of the lateral ventricle and the alveus of the hippocampus and medially by the optic tract.

**Amygdalae normalization.** We normalized our functional data as opposed to extracting signal from individually segmented ROIs because performing voxelwise inferences at the group level offers a considerably greater statistical power than individual subject ROI analyses.

Because structures in the MTL exhibit significant interindividual anatomic variability, the signal-to-noise ratio in group analyses is substantially limited in these areas (Insausti et al., 1998). Atlas-based approaches used to register whole-brain EPI images across participants (such as SPM) look for a global optimum alignment which is achieved under the limitations imposed by the available degrees of freedom, and which is at the expense of regional accuracy. Consequently, BOLD signals may be underestimated or possibly missed (Miller et al., 2005). Alignment of

subcortical structures is substantially improved by a ROI-alignment (ROI-AL) approach, where segmentations of regions of interest (ROIs) are drawn on structural images and aligned directly, resulting in an increased statistical power (Yassa and Stark, 2009). The last iteration of this alignment tool is ROI-Demons, which has proven to be exceptionally accurate in the alignment of hippocampal subfields for instance (<http://darwin.bio.uci.edu/~cestark/roial/roial.html>). Thirion's original demons algorithm has been implemented by Vercauteren and enforces smooth deformations by operating on a diffeomorphic space of displacement fields (Thirion, 1998; Vercauteren et al., 2007). Here, we used the implementation of ROI-Demons in the DemonsRegistration command-line tool (<http://www.insight-journal.org/browse/publication/154>). Our segmented amygdalae ROIs were registered with a "clean" participant (Prévost et al., 2011) to serve as an initial model and to align all ROIs using DemonsRegistration. The resulting registered amygdalae were then averaged in SPM5 (using ImCalc) to create a first model (in this section, a model refers to an amygdalae template resulting from the averaged amygdalae ROIs). Subsequently, the initial nonregistered masks were registered with this first model and the newly registered masks were averaged to create a second model. We repeated the last two steps three more times to generate a more accurate model. We finally registered our initial ROIs with the fifth model to generate the resulting displacement fields (or transformation calculations). These individual displacement fields were then applied to each participant's normalized EPI scans to specifically normalize their amygdalae to our template mask. We applied the same transformation to each participant's structural scan before averaging all the aligned structural scans, to create an ROI-aligned average structural brain of our 26 participants. Finally, amygdalar subdivisions regions of interest (ROIs) were hand-drawn on our template amygdalae using the Atlas of the Human Brain (Mai et al., 2008). We delineated three subareas within the amygdala: the basolateral complex comprised of the basomedial, basolateral and lateral nuclei; the centromedial complex comprised of the central and medial nuclei; and the cortical complex (or cortical nucleus). In its most rostral part, the amygdala is exclusively composed of the basolateral complex. The cortical nucleus appears in the dorsomedial part of midrostral amygdala. The centromedial complex appears slightly more caudally than the cortical nucleus in the most dorsal part of the amygdala. The basolateral complex increases in size as one moves caudally from the anterior amygdala, has its maximal size midrostrocaudally and then decreases as one moves further back toward the caudal amygdala, whereas the cortical nucleus and centromedial complex slightly enlarge midrostrocaudally, but do not decrease in size as one moves further caudally within the amygdala. The cortical nucleus ends midcaudally, the basolateral complex ends in caudal amygdala while the centromedial complex ends in the most caudal part of amygdala.

### fMRI data analysis

The event-related fMRI data were analyzed by constructing sets of  $\delta$  (stick) functions. All three phases of the task were modeled separately and here we report results only from the PIT phase. The GLM included regressors at the time of trial onset and outcome delivery for both the instrumental and Pavlovian phases. For the PIT phase, we included regressors at the time of cue onset for 4 conditions: specific PIT when the option compatible with the Pavlovian cue elicited higher response rates, specific PIT when the incompatible option elicited higher response rates, general trials and neutral trials. All of these regressors were convolved with a canonical hemodynamic response function (HRF). The six scan-to-scan motion parameters derived from the affine part of the realignment procedure were included as regressors of no interest to account for residual motion effects. We also included 13 additional regressors to account for physiological fluctuations (4 related to heart rate, 9 related to respiration) which were estimated using the RETROICOR algorithm (Glover et al., 2000). Sixteen of the 78 (3 sessions × 26 participants) log files could not be used to estimate these regressors due to a technical problem during data collection, and the missing physiological regressors were simply omitted for those sessions. Finally, to account for variance induced by differences in response rates across conditions in the instrumental and PIT phases, we added a regressor of no interest correspond-

ing to the time course derived from this measure. This time course corresponded to the normalized number of response rates for each volume (period of 2 s) of each session of each participant. All of these regressors were entered into a general linear model and fitted to each participant individually using SPM5. The resulting parameter estimates for regressors of interest were then entered into second-level one sample *t* tests to generate random-effects level statistics. We then used an index of specific and general PIT for each participant as covariates at the second-level analysis on the contrast 'specific compatible > specific incompatible' for specific PIT and 'general trials > neutral trials' for general PIT. Separate covariates were used for specific and general PIT for each participant in a single second level analysis. All reported fMRI statistics and *p* values arise from group random-effects analyses. Two participants who never responded more to the incompatible option were excluded from this analysis. Note that among the 24 participants included, the minimum number of incompatible trials was 5. We present our statistical maps at a threshold of  $p < 0.005$ , corrected for multiple comparisons at  $p < 0.05$ . To correct for multiple comparisons in the striatum, we used small volume corrections (SVC) based on a 6 mm sphere centered on the coordinates reported by Bray et al. (2008) for specific PIT ( $[x, y, z]$  [27, -3, -3]). In the amygdala, we first used the 3dFWHMx function in AFNI to estimate the intrinsic smoothness of our data, within the area defined by a mask corresponding to our amygdala template. We then used the AlphaSim function in AFNI to estimate via Monte Carlo simulation an extent threshold for statistical significance that was corrected for multiple comparisons at  $p < 0.05$  for a height threshold of  $p < 0.005$  within the amygdala ROI.

#### Plotting of parameter estimates

Plots of parameter estimates were extracted using the MarsBaR toolbox (<http://marsbar.sourceforge.net/>). For each participant, average parameter estimates were extracted within a 2 mm sphere centered on the peak voxel of the amygdalar activity correlating with either specific or general PIT using the leave-one-participant-out method (Kriegeskorte et al., 2009), thereby avoiding a nonindependence bias in the voxel selection in Figure 4*b,c*, and without using this procedure in Figure 4*d,e*. These  $\beta$  estimates were correlated against the strength of behavioral specific and general PIT as illustrated in Figure 4*b–e*.

## Results

### Behavioral results

#### Subjective ratings of food rewards

Each of the three food rewards was reported to be subjectively pleasant by the participants in ratings taken before starting the experiment, and after the test phase (two-tailed *t* tests comparing the mean to zero, all  $p < 0.001$ , before: food 1: mean = 1.75, SD = 1.46; food 2: mean = 1.72, SD = 1.68; food 3: mean = 1.49, SD = 1.24; after: food 1: mean = 1.76, SD = 1.78; food 2: mean = 1.85, SD = 1.61; food 3: mean = 1.49, SD = 1.86). Furthermore, we performed a 2-way repeated-measures ANOVA (2 time points  $\times$  3 different foods) to test whether there was a significant difference between the ratings for the different foods after compared with before the experiment. We found no main effect for our factors time and food (all *F* values  $< 0.7$ , *p* values  $> 0.5$ ) as well as no interaction between these factors ( $F_{(2,50)} = 0.09$ ,  $p = 0.9$ ), indicating that the foods remained equally subjectively pleasant for participants after the experiment. This was further confirmed by hunger ratings which showed that participants were not significantly more or less hungry after compared with before starting the experiment (two-tailed paired *t* test,  $p = 0.12$ ).

#### Instrumental conditioning

In the instrumental trials, both response rates and response times (i.e., the number of presses and the duration between trial onset and the first key press respectively) were not significantly different across the three response-outcome pairs (repeated-measures ANOVAs, response rate:  $F_{(1,25)} = 1.31$ ,  $p = 0.26$ ; response time:

$F_{(1,25)} = 0.62$ ,  $p = 0.44$ ), suggesting that participants were on average equally motivated to obtain each of the available outcomes. Note that the number of presses was normalized for each participant by subtracting the mean number of presses for this participant and dividing it by the SE. At the end of the experiment, participants had to indicate which instrumental response was associated with which food reward. On average, they were correct 91% of the time (SD = 14.64), which is significantly above chance (*t* test comparing the mean to 50%,  $p < 0.001$ ), suggesting that participants learned the specific action-outcome associations.

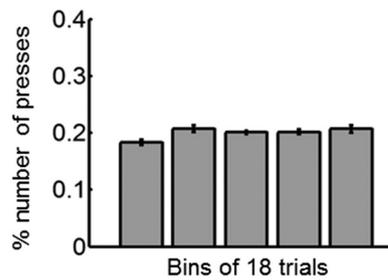
#### Pavlovian conditioning

Participants' heart rate was monitored using a pulse oximeter for the duration of the experiment. Existing research on heart rate responses to significant stimuli has identified an initial bradycardia associated with more aversive stimuli (Libby et al., 1973). This deceleration is thought to express attentional orienting to salient events through parasympathetic activity (Bradley, 2009). During the Pavlovian phase, nonrewarded trials were associated with a more pronounced cardiac deceleration (as assessed by the number of beats) compared with rewarded trials during anticipation, in a time window of 1.5–3.5 s following stimulus onset, as reported previously (Young et al., 2006) (two-tailed paired *t* test,  $p < 0.05$ ). Such physiological changes signal a difference in the elicited emotional state for cues presented on nonrewarded compared with rewarded trials, thereby reflecting a differential conditioned response to the nonrewarded versus rewarded cues. Furthermore, at the end of the experiment, participants had to indicate which Pavlovian cue was associated with which food reward. On average, they were correct 93% of the time (SD = 20.28), which is significantly above chance (two-tailed *t* test comparing the mean to 50%,  $p < 0.001$ ), providing further evidence that participants learned the specific stimulus-outcome associations.

#### Pavlovian-to-instrumental transfer effects

Each trial consisted of a 6 s duration where participants were performing a free-operant task, followed by a variable intertrial interval of 2–6 s (1). Behavioral evidence for specific PIT would be seen if the presence of the Pavlovian cue induced a higher response rate for the action associated with the same outcome as the Pavlovian cue. Conversely, evidence for general PIT would be seen if the presence of an appetitive Pavlovian cue induced a higher response rate for any action (despite being associated with a different outcome) than the presence of a cue that has not been paired with reward. This is exactly what we observed for both specific (two-tailed paired *t* test,  $p < 0.001$ ) and general (two-tailed paired *t* test,  $p < 0.05$ ) PIT.

Dividing the number of trials of the PIT phase into 5 blocks of 18 trials (as in Bray et al., 2008), we investigated whether the percentage in the number of presses significantly varied across these 5 blocks. A two-way repeated-measures ANOVA with the factors block and trial type was performed to investigate whether there was an effect of block during the extinction test. We found a main effect of trial type ( $F_{(2,46)} = 99.21$ ,  $p < 0.001$ ) but no main effect of block ( $F_{(4,92)} = 0.58$ ,  $p = 0.68$ ), and no interaction between these two factors ( $F_{(8,184)} = 1.13$ ,  $p = 0.36$ ), indicating that the biasing effect of the Pavlovian cues on response rates persisted for the duration of the extinction test and did not show any evidence of habituation (Fig. 2). Furthermore, when investigating whether response rates for the three different keys differed in the instrumental and PIT phases using a 2-way repeated-



**Figure 2.** Percentage of number of presses for five 18-trial blocks across all trials. There is no significant linear trend across the session. Error bars indicate SEM.

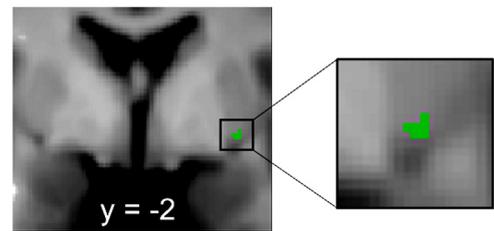
**Table 2.** Mean response rate (RR)  $\pm$  SEM for each participant during the instrumental training and the PIT phase

Participant	Mean RR during instrumental training	Mean RR during PIT phase
1	19.83 $\pm$ 1.46	15.18 $\pm$ 1.16
2	18.53 $\pm$ 1.58	13.16 $\pm$ 0.98
3	21.63 $\pm$ 1.36	18.32 $\pm$ 0.59
4	18.46 $\pm$ 0.81	15.44 $\pm$ 0.46
5	19.66 $\pm$ 1.05	22.05 $\pm$ 0.38
6	12.63 $\pm$ 0.66	18.17 $\pm$ 0.33
7	10.56 $\pm$ 0.67	22.17 $\pm$ 0.43
8	23.7 $\pm$ 0.58	22.06 $\pm$ 0.75
9	16 $\pm$ 0.60	16.88 $\pm$ 0.45
10	10.26 $\pm$ 0.62	15.92 $\pm$ 0.34
11	28.7 $\pm$ 0.49	27.68 $\pm$ 0.23
12	23.9 $\pm$ 1.88	26.4 $\pm$ 0.63
13	13.8 $\pm$ 0.68	23.03 $\pm$ 0.44
14	12.73 $\pm$ 0.68	12.63 $\pm$ 0.90
15	17.2 $\pm$ 1.03	23.62 $\pm$ 0.36
16	15.03 $\pm$ 0.71	21.12 $\pm$ 0.54
17	26.66 $\pm$ 1.28	24.96 $\pm$ 0.36
18	15.9 $\pm$ 0.52	22.56 $\pm$ 0.49
19	25.53 $\pm$ 1.36	26.12 $\pm$ 0.23
20	10.16 $\pm$ 0.57	9.32 $\pm$ 1.16
21	10.8 $\pm$ 0.78	18.68 $\pm$ 0.80
22	25.4 $\pm$ 0.62	22.58 $\pm$ 0.45
23	28.43 $\pm$ 1.21	30 $\pm$ 0.47
24	12.53 $\pm$ 0.77	17.91 $\pm$ 0.44
25	18.66 $\pm$ 0.96	10.56 $\pm$ 1.04
26	10.26 $\pm$ 0.52	3.33 $\pm$ 0.53

measures ANOVA (two phases  $\times$  three response keys), we found no main effect of the key or the phase (all  $F$  values  $< 0.5$ ,  $p$  values  $> 0.3$ ) and no interaction of these factors ( $F_{(2,50)} = 1.8$ ,  $p = 0.18$ ). Similarly, when performing the same analysis using response times, there was no main effect of either factor (all  $F$  values  $< 0.6$ ,  $p$  values  $> 0.5$ ) and no interaction ( $F_{(2,50)} = 0.62$ ,  $p = 0.54$ ). This indicates that even though participants did not receive any reward in the PIT phase, they were equally motivated as when they were getting food rewards in the instrumental phase (Table 2).

### fMRI results

We report results from our analyses within the striatum and amygdala using a height threshold of  $p < 0.005$ , with an extent threshold significant at  $p < 0.05$  corrected for multiple comparisons. We first examined the main effect of each type of PIT. To identify areas mediating specific PIT, we compared activity when participants were performing specific trials with activity when they were performing any other type of trial (i.e., general or neutral trials). As shown in Figure 3, we found significant activity in an area of right ventrolateral putamen (MNI  $[x, y, z]$  [29, -2, -6],  $T = 3.61$ ,  $k = 20$  voxels) overlapping with that reported by



**Figure 3.** Blood oxygen level-dependent (BOLD) signal for the main effect of specific PIT (specific trials  $>$  general and neutral trials) in right ventrolateral putamen, using a height threshold of  $p < 0.005$ , and a small volume correction significant at  $p < 0.05$  corrected for multiple comparisons.

Bray et al. (2008), but no significant activity was found in the amygdala in this contrast.

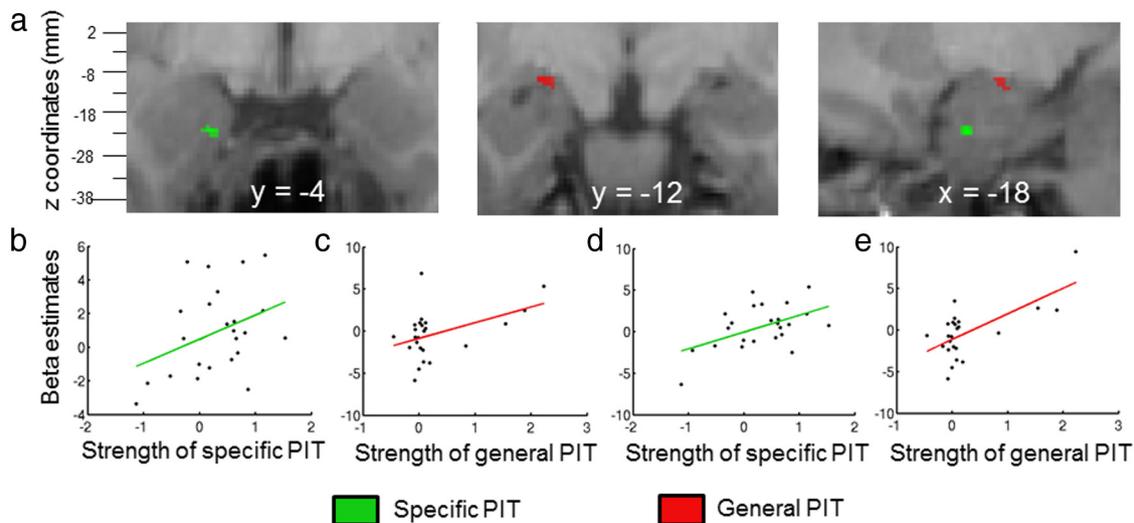
Next, we compared activity when participants responded more to the instrumental action compatible with the Pavlovian cue to trials when they responded more to the action associated with a different incompatible outcome. To detect regions mediating general PIT, we compared activity between general and neutral CS trials, that is when participants were presented with a cue that had been paired with a reward outcome that was different from the reward outcomes associated with the available actions, compared with when the Pavlovian cue presented was associated with no outcome. We did not find significant activity in the striatum or amygdala in this contrast at our statistical threshold.

Following the between-participant analysis performed in Talmi et al. (2008), we next examined the extent to which activity within the amygdala in the relevant contrasts was modulated as a function of the variation across participants in the extent to which they exhibited behavioral evidence of general and specific transfer, by including our behavioral measures of specific vs general PIT for each participant as a covariate in these contrasts (Talmi et al., 2008). Activity in a part of ventral amygdala consistent with the location of the basolateral complex (Mai et al., 2008) showed significant correlations with the magnitude of specific PIT exhibited by participants (Fig. 4a, in green, MNI  $[x, y, z]$  [-18, -3, -22],  $T = 5.16$ ,  $k = 24$  voxels). Conversely, a region of dorsal amygdala consistent with the location of the centromedial complex was found to be correlated with the degree of general PIT exhibited by participants (4a, in red, MNI  $[x, y, z]$  [-15, -10, -11],  $T = 5.83$ ,  $k = 38$  voxels).

Finally, we tested for a differential involvement of BLA and CeN in specific and general PIT. When correlating the mean difference in the  $\beta$  estimates from our general PIT contrast between our CeN ROI and our BLA ROI against participants' index of general PIT, we found a significant positive correlation ( $r = 0.48$ ,  $p < 0.05$ ), indicating that CeN is significantly more involved in general PIT than BLA. However, we did not find evidence suggesting that BLA is significantly more involved in specific PIT than CeN when correlating the mean difference in the  $\beta$  estimates from our specific PIT contrast between our BLA ROI and our CeN ROI against participants' index of specific PIT ( $r = 0.24$ ,  $p = 0.26$ ).

### Discussion

In the current study, we show the first behavioral evidence of general appetitive Pavlovian instrumental transfer in humans, replicate findings of ventral striatum involvement in specific PIT (Bray et al., 2008) and further provide evidence for the involvement of amygdala in mediating specific and general PIT effects by



**Figure 4.** Specific and General PIT in the amygdala. *a*, BOLD signal positively correlating with the magnitude of specific (in green) and general (in red) PIT across participants in the ventral and dorsal amygdalae, respectively, using a height threshold of  $p < 0.005$ , with an extent threshold significant at  $p < 0.05$  corrected for multiple comparisons. *b*, Scatter plot showing the  $\beta$  estimates extracted from a 2 mm sphere centered on the peak voxel within the ventral amygdala correlating with specific PIT against the strength of behavioral specific PIT for each participant. *c*, Scatter plot showing the  $\beta$  estimates extracted from a 2 mm sphere centered on the peak voxel correlating with general PIT in the dorsal amygdala against the strength of behavioral general PIT for each participant. The plots in *c* and *d* were produced using a leave-one-participant-out procedure, thereby avoiding nonindependence issues (see Materials and Methods, Plotting of parameter estimates, for details). *d*, *e*, Similar scatter plots as displayed in *b*, *c* except that no leave-one-participant-out procedure was used.

making use of a high-resolution imaging protocol. More specifically, we found a main effect of specific PIT in the ventrolateral putamen whereas we found between-participant differences for specific PIT in a region of human ventral amygdala within the boundaries of the BLA, and for general PIT in the dorsal amygdala within the boundaries of the CeN.

The ventral striatum has long been shown to be implicated in mediating PIT effects in rodents (Corbit et al., 2001; Hall et al., 2001; Corbit and Balleine, 2011) with recent evidence for a functional dissociation between the shell and core part of the nucleus accumbens in the mediation of specific and general PIT respectively (Corbit and Balleine, 2011). In humans, ventrolateral putamen activity for specific PIT has previously been reported (Bray et al., 2008), resonating with our findings.

Although we found a main effect of specific PIT in the ventral striatum, we did not find a main effect of specific or general PIT in the amygdala. Instead, we found an interaction between BOLD signal in the amygdala and behavior. It is interesting to note that Talmi et al. (2008) found a similar correlation between amygdala activity and the magnitude of their PIT effect, although the design of their study did not allow them to separately assess the effects of general and specific PIT. The results of these two studies are in agreement with the idea that amygdala involvement in PIT appears to depend on individual differences in behavioral expression of these effects, indicating that amygdala activity in PIT appears to be closely coupled to the degree of behavioral expression of these phenomena. Here, we expand on this earlier finding by showing that different types of PIT (specific versus general) seem to correlate with activity in different areas of the amygdala (BLA versus CeN). These results also dovetail with the results of a rodent lesion study in which it was found that lesions to the BLA impaired specific PIT while lesions of the CeN impaired general PIT (Corbit and Balleine, 2005). Therefore, the present findings indicate that a similar differentiation of function may be present within the human amygdala, suggesting that the neural underpinnings of the influence of Pavlovian cues on decision making within the amygdalae are highly preserved across mammalian species.

These findings are also important for furthering understanding of the different functional contributions of amygdalar subregions in incentive processing in both humans and other animals. Killcross and Balleine have proposed a model of amygdala function on the basis of extant rodent lesion data in which the BLA is suggested to be more involved in mediating processing of specific sensory features of an outcome, while the CeN is proposed to be more involved in mediating processing of the general affective value of an outcome in a manner analogous to the distinction between consummatory and preparatory conditioning first espoused by Konorski (Konorski, 1967). In preparatory conditioning, a conditioned cue elicits generalized conditioned responses to appetitive or aversive outcomes such as approach and avoidance, while in consummatory conditioning, conditioned responses which are specific to the particular outcome predicted are elicited. Specific and general PIT effects have further been suggested to operate via these distinct incentive-processing mechanisms. Whereas in specific PIT, a conditioned stimulus is suggested to elicit sensory-specific features of an outcome which in turn biases instrumental actions toward that outcome, in general PIT, the conditioned stimulus instead is suggested to elicit a generalized representation of the incentive value of that outcome in a manner independent of outcome identity (Dickinson and Balleine, 2002; Holland, 2004; Balleine and Killcross, 2006). Thus, the present findings are broadly consistent with the proposed differential contribution of the BLA and CeN in specific vs general incentive processing as proposed in the Killcross and Balleine model (Balleine and Killcross, 2006).

An important issue to note in the present study (and indeed any study involved in measuring specific and general PIT) is that because these phenomena manifest themselves as very different behaviors, such that one is manifested through an overall increase in response rates while the other is measured in the current study by assessing bias in choice behavior, these two phenomena may be scaled differently and therefore a direct contrast between the slopes of the degree of expression of specific versus general PIT is not meaningful. As a consequence, the appropriate test is to compare the slopes of activity in particular ROIs with respect to each

of these phenomena separately (as opposed to directly comparing the slopes for general and specific PIT). When testing for a significant difference in the slopes of general PIT between the two regions, we found that CeN was significantly more involved in general PIT than was the BLA. In a separate analysis, we tested whether our BLA ROI was significantly more correlated with specific PIT than the CeN was correlated with specific PIT, and did not find statistical support for this hypothesis. Accordingly, we must acknowledge the limitation that we can only claim a partial but not a double dissociation between CeN and BLA with regard to specific versus general PIT. A likely explanation for the absence of a double dissociation is that unlike the CeN ROI which is circumscribed due to its small spatial size, the BLA ROI lacks sensitivity because it encompasses a considerable territorial expanse of the amygdala and therefore includes many voxels that do not have any involvement with specific PIT. Thus it is likely that specific PIT effects are being drowned out in the average across the BLA ROI.

This caveat notwithstanding, our results still enable us to make the claim that the human BLA is involved in specific PIT while the human CeN is involved in general PIT, and furthermore that the CeN is significantly more involved in general than specific PIT. Our findings have important implications for addictive and economic behavior. Indeed, continued drug-seeking behavior, particularly in the presence of environmental drug-associated cues, has been considered to reflect PIT influence (O'Brien et al., 1998), and so has the impact of marketing strategies on consumer actions (Smeets and Barnes-Holmes, 2003). More generally, these results show how hr-fMRI can, when combined with appropriate behavioral paradigms, be deployed to address circuit-level questions about the functions of subregions within heterogeneous subcortical structures.

## References

- Balleine BW, Killcross S (2006) Parallel incentive processing: an integrated view of amygdala function. *Trends Neurosci* 29:272–279.
- Blundell P, Hall G, Killcross S (2001) Lesions of the basolateral amygdala disrupt selective aspects of reinforcer representation in rats. *J Neurosci* 21:9018–9026.
- Bradley MM (2009) Natural selective attention: orienting and emotion. *Psychophysiology* 46:1–11.
- Bray S, Rangel A, Shimojo S, Balleine B, O'Doherty JP (2008) The neural mechanisms underlying the influence of pavlovian cues on human decision making. *J Neurosci* 28:5861–5866.
- Corbit LH, Balleine BW (2005) Double dissociation of basolateral and central amygdala lesions on the general and outcome-specific forms of pavlovian-instrumental transfer. *J Neurosci* 25:962–970.
- Corbit LH, Balleine BW (2011) The general and outcome-specific forms of Pavlovian-instrumental transfer are differentially mediated by the nucleus accumbens core and shell. *J Neurosci* 31:11786–11794.
- Corbit LH, Muir JL, Balleine BW (2001) The role of the nucleus accumbens in instrumental conditioning: evidence of a functional dissociation between accumbens core and shell. *J Neurosci* 21:3251–3260.
- Dickinson A, Balleine BW (2002) The role of learning in the operation of motivational systems. In: Stevens' handbook of experimental psychology (Pashler H, ed), pp 497–533. New York: Wiley.
- Garner DM, Olmsted MP, Bohr Y, Garfinkel PE (1982) The eating attitudes test: psychometric features and clinical correlates. *Psychol Med* 12:871–878.
- Glover GH, Li TQ, Ress D (2000) Image-based method for retrospective correction of physiological motion effects in fMRI: RETROICOR. *Magn Reson Med* 44:162–167.
- Hall J, Parkinson JA, Connor TM, Dickinson A, Everitt BJ (2001) Involvement of the central nucleus of the amygdala and nucleus accumbens core in mediating Pavlovian influences on instrumental behaviour. *Eur J Neurosci* 13:1984–1992.
- Holland PC (2004) Relations between Pavlovian-instrumental transfer and reinforcer devaluation. *J Exp Psychol Anim Behav Process* 30:104–117.
- Holland PC, Gallagher M (2003) Double dissociation of the effects of lesions of basolateral and central amygdala on conditioned stimulus-potentiated feeding and Pavlovian-instrumental transfer. *Eur J Neurosci* 17:1680–1694.
- Insausti R, Juottonen K, Soinen H, Insausti AM, Partanen K, Vainio P, Laakso MP, Pitkänen A (1998) MR volumetric analysis of the human entorhinal, perirhinal, and temporopolar cortices. *AJNR Am J Neuroradiol* 19:659–671.
- Konorski J (1967) Integrative activity of the brain. Chicago: University of Chicago Press.
- Kriegeskorte N, Simmons WK, Bellgowan PS, Baker CI (2009) Circular analysis in systems neuroscience: the dangers of double dipping. *Nat Neurosci* 12:535–540.
- Libby WL Jr, Lacey BC, Lacey JI (1973) Pupillary and cardiac activity during visual attention. *Psychophysiology* 10:270–294.
- Mai JK, Paxinos G, Voss T (2008) Atlas of the human brain, Ed 3. San Diego: Academic.
- Miller MI, Beg MF, Ceritoglu C, Stark C (2005) Increasing the power of functional maps of the medial temporal lobe by using large deformation diffeomorphic metric mapping. *Proc Natl Acad Sci U S A* 102:9685–9690.
- O'Brien CP, Childress AR, Ehrman R, Robbins SJ (1998) Conditioning factors in drug abuse: can they explain compulsion? *J Psychopharmacol* 12:15–22.
- Prévost C, McCabe JA, Jessup RK, Bossaerts P, O'Doherty JP (2011) Differentiable contributions of human amygdala subregions in the computations underlying reward and avoidance learning. *Eur J Neurosci* 34:134–145.
- Smeets PM, Barnes-Holmes D (2003) Children's emergent preferences for soft drinks: stimulus-equivalence and transfer. *J Econ Psychol* 603–618.
- Talmi D, Seymour B, Dayan P, Dolan RJ (2008) Human pavlovian-instrumental transfer. *J Neurosci* 28:360–368.
- Thirion JP (1998) Image matching as a diffusion process: an analogy with Maxwell's demons. *Med Image Anal* 2:243–260.
- Vercauteren T, Pennec X, Perchant A, Ayache N (2007) Non-parametric diffeomorphic image registration with the demons algorithm. *Med Image Comput Comput Assist Interv* 10:319–326.
- Yassa MA, Stark CE (2009) A quantitative evaluation of cross-participant registration techniques for MRI studies of the medial temporal lobe. *Neuroimage* 44:319–327.
- Young TM, Asahina M, Nicotra A, Mathias CJ (2006) Skin vasomotor reflex responses in two contrasting groups of autonomic failure: multiple system atrophy and pure autonomic failure. *J Neurol* 253:846–850.