"Virus and Epidemic": Causal Knowledge Activates Prediction Error Circuitry

Daniela B. Fenker[1], Mircea A. Schoenfeld[1], Michael R. Waldmann[2], Hartmut Schuetze[1], Hans-Jochen Heinze[1], and Emrah Duezel[3,4]

Abstract

■ Knowledge about cause and effect relationships (e.g., virus–epidemic) is essential for predicting changes in the environment and for anticipating the consequences of events and one's own actions. Although there is evidence that predictions and learning from prediction errors are instrumental in acquiring causal knowledge, it is unclear whether prediction error circuitry remains involved in the mental representation and evaluation of causal knowledge already stored in semantic memory. In an fMRI study, participants assessed whether pairs of words were causally related (e.g., virus–epidemic) or noncausally associated (e.g., emerald–ring). In a second fMRI study, a task cue prompted the participants to evaluate either the causal or the noncausal associative relationship between pairs of words. Causally related pairs elicited higher activity in OFC, amygdala, striatum, and substantia nigra/ventral tegmental area than noncausally associated pairs. These regions were also more activated by the causal than by the associative task cue. This network overlaps with the mesolimbic and mesocortical dopaminergic network known to code prediction errors, suggesting that prediction error processing might participate in assessments of causality even under conditions when it is not explicitly required to make predictions. ■

INTRODUCTION

The knowledge of the specific relationship between a cause and an effect (e.g., virus–epidemic) and its dissociation from noncausally associated events (e.g., virus–bacteria) enables predicting consequences of events and actions. Since the pioneering work of Pavlov (1927), an extensive range of theoretical and experimental work has addressed the question how animals and humans learn to associate a cue with an outcome (Cobos, López, Caño, Almaraz, & Shanks, 2002; Shanks & López, 1996) or link a cause to an effect (Blaisdell, Sawa, Leising, & Waldmann, 2006; Waldmann, 1996, 2000, 2001; Waldmann, Holyoak, & Fratianne, 1995; Waldmann & Holyoak, 1992), enabling them to guide their behavior accordingly. There is converging evidence that processing prediction errors is one candidate mechanism driving such learning of causality (Pearce & Hall, 1980; Rescorla & Wagner, 1972). Learning is enabled because expectations regarding outcomes (e.g., reward or punishment) are updated following prediction errors until expectations and outcomes eventually converge.

Evidence from animal studies indicates that a designated neural network is critical for the ability to predict reward or punishment, to compute prediction errors, and to adjust response selection and goal-directed behavior on the basis of such errors. This network includes the medial pFC (Ostlund & Balleine, 2005; Matsumoto & Tanaka, 2004; Matsumoto, Suzuki, & Tanaka, 2003), the OFC (Wallis, 2007; Izquierdo, Suda, & Murray, 2004), the amygdala (Balleine, Killcross, & Dickinson, 2003), the striatum/nucleus accumbens (Balleine, Delgado, & Hikosaka, 2007; Cromwell, Hassani, & Schultz, 2005; de Borchgrave, Rawlins, Dickinson, & Balleine, 2002; Schultz, Tremblay, & Holland, 1998), and the thalamus (Corbit, Muir, & Balleine, 2003) and is closely linked to dopaminergic (DA) neuromodulation (Schultz, 2002, 2006; Andrzejewski, Spencer, & Kelley, 2005). In humans, components of this network have been implicated in making predictions for desired outcomes and computing prediction errors (Knutson & Wimmer, 2007; Knutson, Taylor, Kaufman, Peterson, & Glover, 2005; Breiter, Aharon, Kahneman, Dale, & Shizgal, 2001; Berns, McClure, Pagnoni, & Montague, 2001), in classification learning (Rodriguez, Aron, & Poldrack, 2006), in associative causal learning mechanisms (Corlett et al., 2004; Turner et al., 2004; Fletcher et al., 2001), and in encoding causal effects of actions (Tanaka, Balleine, & O'Doherty, 2008). Moreover, mesencephalic midbrain structures known to harbor DA neurons (substantia nigra/ventral tegmental area, SN/VTA) are activated during the encoding of a stimulus predicting monetary reward. Such an activation pattern has recently been shown to facilitate long-term memory for that stimulus (Wittmann et al., 2005).

The functional anatomical circuitry underlying prediction error learning has been most intensively studied for reward prediction errors (Schultz, 2004). During unexpected

[1]Otto-von-Guericke University, Magdeburg, Germany, [2]University of Göttingen, Germany, [3]Institute of Cognitive Neuroscience University College London, [4]Institute of Cognitive Neurology and Dementia Research, Magdeburg, Germany

primary (e.g., food) or secondary (e.g., money) rewards, the nucleus accumbens receives DA signals from the SN/VTA of the midbrain. Once a reward becomes predictable by a conditioned (i.e., reward-predicting) stimulus (CS+), the DA signal in the nucleus accumbens shifts forward from the time of reward outcome to the presentation of the CS+ (O'Doherty, Hampton, & Hackjin, 2007; O'Doherty, Dayan, Friston, Critchley, & Dolan, 2003). Conversely, omission of a reward after presentation of a CS+ leads to a decrease of DA signaling below baseline. It has thus been hypothesized that the DA signal in the SN/VTA and in the nucleus accumbens code both reward prediction and prediction error.

However, not only reinforcement learning paradigms provide valuable insights into the mechanisms underlying the acquisition of causal knowledge; associative causal learning studies have shown that prediction error mechanisms are also at play during the acquisition of cause–effect relationships even in the absence of apparent reward manipulation (Corlett et al., 2004).

Both human and nonhuman animals store a rich database of causal relationships. Knowledge of such causal relationships is instrumental in selecting appropriate goals and actions and, in a much broader perspective, in guiding social interactions as well as in shaping our understanding of natural phenomena. Given that the computation of prediction error seems to be an important component in the learning of causal contingencies, our study addresses the questions whether brain regions associated with prediction error processing participate in the representation of causal knowledge in semantic memory.

To date, only a few studies have examined how already acquired causal knowledge is represented and how it can be dissociated from noncausal associative relationships in human semantic memory (Fenker, Waldmann, & Holyoak, 2005; Satpute et al., 2005). These studies have shown that causal relations are stored and accessed differently from noncausal associative relations. Given the fact that we establish causal knowledge through life experience and that prediction error processing is involved during the acquisition of causal knowledge (Corlett et al., 2004; Turner et al., 2004), it is plausible that cause–effect relationships stored in semantic memory are represented differently from noncausal associations, in a way that allows updating cause–effect contingency changes, hence allowing modification by experience. According to this possibility, every time we encounter a causal relationship, prediction error circuitry is activated by default, allowing the modification of cause–effect relationships stored in semantic memory if a prediction error occurs due to changed contingencies. In contrast, noncausal associative relations (e.g., emerald–ring) should not engage prediction error circuitry.

The key question addressed here is whether prediction error circuitry is a part of the representation of causal semantic memories. An alternative possibility is that causal relationships, once acquired and part of our semantic knowledge, are stored and represented just like any other associative relationship (hence in a more "static" fashion), thereby reflecting the fact that we have learned extensively that certain causes determine specific effects. Updating this knowledge may thus not be necessary, and hence it may not be the case that prediction error circuitry is engaged whenever we encounter virus and epidemic together, thus making the representations of well-learned causal knowledge undistinguishable from well-learned noncausal associative relationships.

The dissociation of causal and noncausal associative relationships stored in semantic memory was investigated in two event-related fMRI experiments. We examined the potential involvement of prediction error circuitry in the retrieval of already existing causal in contrast to noncausal associative semantic knowledge. In our design, therefore, we deliberately avoided standard learning tasks that are already known to recruit prediction error circuitry. Hence, participants were not explicitly required to make predictions.

In the first experiment, word pairs denoting a cause–effect relationship were simply contrasted with pairs denoting a noncausal associative relationship. In the second experiment, we compared the task-related preparation for the retrieval of causal and noncausal associative relationships. The words of each pair were shown one after the other and they could either be unrelated (e.g., door–pinball), noncausal associatively related (related in the absence of a causal relationship e.g., emerald–ring), or causally related (e.g., virus–epidemic). In Experiment 1, participants had to evaluate the presented word pairs for a causal relationship. In the Experiment 2, the word pairs had to be evaluated either for a causal relationship or for a noncausal associative relationship indicated by a verbal cue presented before each pair.

If causal relationships are represented in a similar fashion as noncausal associative relationships, the retrieval of a cause–effect relationship from semantic memory should only differ in terms of the semantic meaning and would be reflected mainly in semantic memory regions, such as prefrontal (Noppeney, Phillips, & Price, 2004; Nyberg et al., 2003), temporal (Patterson, Nestor, & Rogers, 2007; Rogers et al., 2006; Noppeney et al., 2004), and parietal cortices (Rogers et al., 2006; Wiggs, Weisberg, & Martin, 1999). In contrast, if cause–effect relationships are represented as contingency-based relationships, their retrieval and evaluation could involve also updating processes that would be reflected in neural activation going beyond semantic processing of noncausal associations by involving mesolimbic and mesocortical circuitry known to compute prediction errors during learning.

METHODS

Experiment 1

Participants

Fifteen participants (age range between 19 and 30 years old, 12 women) with normal or corrected-to-normal vision participated in the study. They were paid €12 and gave

their written consent. All subjects were right-handed according to self-report and had no history of neurological illness. The study was approved by the ethics committee of the Otto-von-Guericke University, Magdeburg.

Stimuli

The stimuli were 210 German word pairs of 4- to 13-letter words. These word pairs were translated into German from the English word pairs used by Fenker et al. (2005). The original English causal and noncausal associative word pairs were selected from the University of South Florida (USF) Word Association Norm list (Nelson, McEvoy, & Schreiber, 1998) with a forward and backward strength <0.01. In addition, a norming study was used to select causally related item pairs equated in both directions in terms of the strength of statistical relations (e.g., frequency of occurrence of epidemic, given a virus versus frequency of occurrence of a virus given epidemic; Fenker et al., 2005). The German words were presented in white Arial 28-point font on black background via a custom projection system. Eighty-four (84) word pairs were causally related (e.g., virus–epidemic), 42 noncausal associatively related (e.g., ring–emerald) and 84 unrelated (e.g., door–pinball).

Procedure

Participants were given written instructions before scanning. A causal relation was defined as follows: "the event described by the first word causes or is caused by the event described by the second word." A noncausal associative relation was defined as follows: "meaningful relationship between the two events, but not a causal relationship." After ensuring that participants understood the instructions, they were placed in the scanner. Participants were asked to fixate on a central dot. In each trial, the first word replaced the fixation dot for 1 sec and then was replaced by the second word for 1 sec followed by fixation. The ISIs between trials were jittered between 4 and 8 sec in 2-sec steps (mean ISI = 4.67 sec). If a word pair was causally related, participants pressed a button with the right index finger, and if a word pair was noncausal associatively related or unrelated, they responded with the left index finger. The response hands were counterbalanced over participants (Figure 1, left panel).

fMRI Image Acquisition

Images were acquired on a neurooptimized General Electric Signa LX 1.5T system. Whole-head fMRI data were acquired, with 23 slices (matrix = 64 × 64; field of view = 22 cm; slice thickness = 5 mm; 1 mm gap; orientation AC–PC), using an echo-planar gradient-echo sequence (repetition time/echo time/flip angle = 2000 msec/35 msec/80°, ramp sampling off). The data were collected over two runs each lasting approximately 8.5 min (253 volumes).

fMRI Image Analysis

Data were time sliced, realigned, normalized to the Montreal Neurological Institute (MNI) template, and spatially smoothed (Gaussian kernel, 8 mm). Statistical analysis used the standard hemodynamic response function and the movement parameters serving as regressors in the event-related design for each subject (SPM99, Wellcome Department of Imaging Neuroscience, London). For group

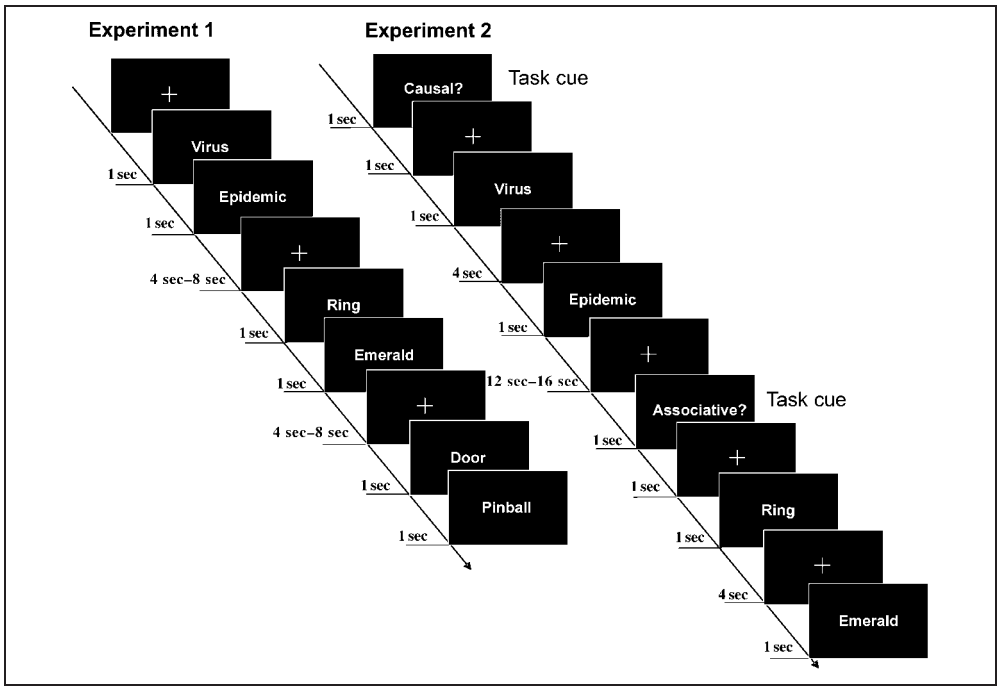

Figure 1. Examples of trials and timing. Left: Experiment 1, causal trial followed by a noncausal associative trial. Right: Experiment 2, causal task cue followed by a causally related word pair and associative task cue followed by a noncausal associative word pair.

analyses, we entered contrast images into one-sample *t* tests, treating subjects as a random variable. SPM contrasts were calculated using a threshold of $p < .001$ and a cluster size criterion of >10 voxels. All reported coordinates refer to MNI space. SPM group contrast images are depicted on single-subject T1-weighted slices from MRIcro (www.cla.sc.edu/psyc/faculty/rorden/mricro.html). In addition, to assess differences between the three conditions (i.e., type of relationship), we performed an ROI analysis for the left amygdala, left SN/VTA, and bilateral nucleus accumbens. The ROIs were functionally defined on the basis of the comparison causal versus noncausal associative word pairs. The beta values of the peak voxel were extracted for each participant and the three conditions (causally related, noncausal associatively related, and unrelated word pairs; Figure 3B).

Experiment 2

Participants

Eighteen healthy right-handed participants (14 women), ranging in age from 19 to 27 years, took part in the experiment and were paid €24. They all had to fulfill the same requirements as the participants in the Experiment 1. The samples of Experiments 1 and 2 were independent.

Stimuli

The stimuli were 252 German word pairs of length between 4 and 13 letters and were presented in white Arial 28-point font on black background via a custom projection system. The word pairs were the same as in Experiment 1, but 42 noncausal associatively related word pairs were added to account for stimulus balancing.

Procedure

Participants were given written instructions before scanning. In the scanner, each trial began with the presentation of the cue, which was either the word "Associative?" or the word "Causal?." The cue was presented for 1 sec, followed by a 1-sec fixation and the first word at fixation for 1 sec. Between the presentation of the first and the second word, fixation was shown for 4 sec. The ISI between the cues ranged from 12 to 16 sec (in steps of 2 sec) with a mean ISI of 12.9 sec (Figure 1, right panel). Participants were instructed to determine if the relationship between the two words match the cue, that is, if the word pairs describe a noncausal associative relationship (associative cue) or a causal relationship (causal cue). The definition of the relationships was identical to Experiment 1. Half of the pairs were presented with the associative cue and the other half with the causal cue. This combination was counterbalanced across participants; that is, word pairs presented with the associative cue for a subject were presented with the causal cue for another subject and vice versa. Participants responded by pressing a button with the left or right index finger for matching and nonmatching, respectively. The response fingers were counterbalanced across participants. Response classes for correct responses were "causal–causal" for causal cue and causal relation (*yes* response), "associative–associative" for associative cue and noncausal associative relation (*yes* response), "causal–associative" for causal cue and noncausal associative relation (*no* response), "associative–causal" for associative cue and causal relation (*yes* response), "causal–unrelated" for causal cue and unrelated word pairs (*no* response), and "associative–unrelated" for associative cue (*no* response).

fMRI Image Acquisition

The parameters of the image acquisition were the same as in Experiment 1. The data were collected over six separate runs each lasting approximately 9.5 min, resulting in the acquisition of 285 volumes.

fMRI Image Analysis

The functional data of Experiment 2 were analyzed using SPM99 (Wellcome Department of Imaging Neuroscience, London) using the same parameters as in Experiment 1. To assess brain activation separately for "type of cue" and "type of relationship," we performed a multiple regression analysis for each single subject (Postle, 2005; Schon, Hasselmo, Lopresti, Tricarico, & Stern, 2004; Postle, Berger, Taich, & D'Esposito, 2000). We created the following three regressors (Figure 2): Regressor 1 assessed preparatory difference induced by the cues, Regressor 2 differentiated the influence of the causal cue and the associative cue on relationship processing, and Regressor 3 assessed the difference between the causal relationship and the noncausal associative relationship, regardless of the previous cue. Given the fact that the results of a regressor cannot be interpreted independently from another regressor if they share a significant amount of variance, we orthogonalized our three regressors to minimize the shared variance. First, we assigned weight coefficients for each regressor such that the internal sum of the products of the weights was zero. Second, we made sure that the sum of the products of the weight coefficients (for cue and word) of each regressor pair was also zero (Bortz, 2005). For group analyses, we entered the regressor images into one-sample *t* tests, treating subjects as a random variable. SPM group contrast images are depicted on single-subject T1-weighted slices from MRIcro (www.cla.sc.edu/psyc/faculty/rorden/mricro.html) with a threshold of $p < .001$, >10 voxels and $p < .005$, >10 voxels, respectively.

RESULTS

Experiment 1: Behavioral Results

We computed two separate ANOVAs over RTs and response accuracy using the type of relationship (causal,

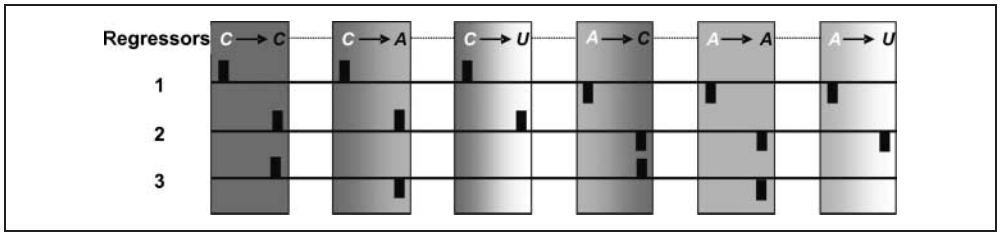

Figure 2. Regressors of Experiment 2. Regressor 1 assessed preparatory difference induced by the cues. Regressor 2 differentiated the influence of the causal and the associative cue at the time of relationship evaluation, and Regressor 3 assessed the relationship evaluation between the causal and the noncausal associative word pairs regardless of the cue. The white letters refer to the type of cue (c = causal; a = associative), and the black letters refer to the type of relationship (c = causal; a = noncausal associative; u = unrelated). The three regressors were orthogonalized with respect to each other; that is, the sum of the products of the weight coefficient (for cue and word) of each regressor pair was zero.

noncausal associative, and unrelated) as within-subjects factor. The RTs for the three conditions differed significantly, $F(2, 28) = 13.05, p < .01$. A post hoc least significant difference (LSD) test showed that participants responded significantly slower to noncausal associative word pairs than to causal word pairs and unrelated word pairs (all ps $< .01$). Their response times did not differ between the causal and the unrelated word pairs ($p > .05$; Table 1). Response accuracy did also differ significantly across conditions, $F(2, 28) = 18.67, p < .01$. Participants gave significantly more correct responses to the unrelated word pairs than to the causal and to the noncausally associated word pairs (all ps $< .01$). There was no significant difference in response accuracy between the causally related and the noncausally associated word pairs ($p > .05$; Table 1).

fMRI Results

During the evaluation of noncausally associated word pairs in contrast to unrelated word pairs, a network of areas predominantly in the left hemisphere was activated. Significant activations were observed in bilateral frontal (left Brodmann's area [BA] 8, 46, 47 and right BA 9, 47), left temporal (BA 37) and parietal areas (BA 40), right occipital regions (BA 18, 19), left thalamus, and left and right caudate nucleus (Table 2). The presence of a causal relationship between the word pairs compared with unrelated word pairs also elicited activity in a widespread network of regions including bilateral frontal (left BA 8, 46 and right BA 9, 47), left temporal (BA 37), left parietal (BA 40), left posterior cingulate gyrus (BA 31), right occipital (BA 18) and left and right caudate nucleus, and left amygdala (Table 2). The differences of both evaluation processes were substantiated by exclusively masking ($p = .05$, for the mask) the activity elicited by the comparison causal versus unrelated with the comparison noncausal associative versus unrelated word pairs. During this masking procedure, all voxels that reach the default level of significance in the masking contrast will be removed. The remaining activation (causal vs. unrelated word pairs) is devoid of the activation found in the masking contrast (noncausal associative vs. unrelated word pairs) and therefore demonstrates the distinct processing of causal relations. Therefore, significant activations reflected those regions that were exclusively activated by the presence of a causal relationship resulting in a network of areas predominantly located in left hemisphere including in middle frontal (BA 6, 46), superior/medial frontal (BA 9) regions, superior temporal regions (BA 22), left and right nucleus accumbens, posterior cingulate gyrus (BA 31), amygdala, and SN/VTA (Figure 3A and B).

To differentiate the evaluation of cause–effect relationships from the evaluation of noncausal associative relationships, we directly contrasted the activation for causally

Table 1. Behavioral Data of the Two Experiments

Experiment 1	*Causal Relationship*	*Noncausal Associative Relationship*	*Unrelated Word Pairs*
RTs	1359 (73)	1510 (93)	1293 (83)
Accuracy	87 (2)	83 (3)	99 (1)

Experiment 2	*Causal–Causal*	*Causal–Noncausal Associative*	*Causal–Unrelated*	*Associative–Causal*	*Associative–Noncausal Associative*	*Associative–Unrelated*
RTs	1949 (112)	2272 (122)	1560 (72)	1751 (92)	1945 (112)	1708 (81)
Accuracy	86 (2)	66 (4)	97 (1)	93 (1)	80 (3)	95 (1)

Mean RTs in milliseconds and accuracy in percent for all participants. Data in parentheses refer to *SEM*s. For Experiment 2, the top word refers to the type of cue and the bottom word refers to the type of relationship.

Table 2. Significant Activation Found in Experiments 1 and 2

			MNI coordinates		
	BA	*z Value*	*x*	*y*	*z*
Experiment 1					
Causal vs. unrelated					
Left caudate nucleus		5.18	−9	9	6
Left inferior frontal	46	4.47	−39	39	6
Left inferior parietal	40	5.46	−39	−66	45
Left insula		4.43	−51	15	0
Left middle temporal	37	4.21	−57	−51	−9
Left orbito-frontal		4.06	−12	6	−18
Left posterior cingulate	31	4.48	−6	−48	30
Left superior frontal/medial part	8	5.21	−3	30	54
Left cerebellum		3.81	−30	−69	−39
Right caudate		4.42	12	12	3
Right inferior frontal	47	4.28	36	24	0
Right lingual gyrus	18	5.07	9	−81	−18
Right middle frontal	9	3.46	51	24	42
Noncausal associative vs. unrelated					
Left caudate		5.03	−9	6	6
Left inferior frontal	46	5.47	−45	42	9
	47	3.8	−30	24	0
Left inferior parietal	40	5.03	−45	−57	45
Left middle temporal	37	3.58	−60	−51	−12
Left superior frontal/medial part	8	5.45	−3	39	48
Left thalamus		4.65	−3	−27	3
Right caudate		4.62	12	12	3
Right inferior frontal	47	4.5	54	21	−6
Right lingual gyrus	18	5.38	6	−78	−12
	19	4.1	21	−51	−3
Right middle frontal	9	3.47	51	24	42
Causal vs. unrelated exclusively masked (p = .05) by noncausal associative vs. unrelated					
Left amygdala		5.44	−18	0	−18
Left caudate/nucleus accumbens		6.2	−6	−12	−3
Left middle frontal	6	5.72	−39	3	36
	46	4.83	−45	36	24
Left posterior cingulate	31	5.71	−6	−45	30
Left SN/VTA		*3.57*	*−9*	*−15*	*−12*
Left superior frontal/medial part	9	5.06	−6	54	36
Left superior temporal	22	6.33	−54	−60	15
Right putamen/nucleus accumbens		4.85	15	12	−3

Table 2. (*continued*)

	BA	z Value	MNI coordinates		
			x	*y*	*z*
Causal vs. noncausal associative					
Left amygdala		5.06	−18	−3	−15
Left inferior frontal	44	3.83	−54	6	27
	45	3.69	−51	30	12
Left inferior parietal lobe	40	4.01	−60	−36	42
Left nucleus accumbens		4.53	−6	12	−6
Left posterior cingulate gyrus	31	4.33	−3	−15	45
Left SN/VTA		4.86	−6	−27	−15
Left superior frontal gyrus/medial part	9	4.67	−6	51	39
Right inferior frontal	44	3.92	63	12	15
Right nucleus accumbens		4.05	12	12	−6
Right orbito-frontal	11	3.53	3	45	−12
Experiment 2					
Cue: causal vs. associative, at cue presentation (Regressor 1)					
Left amygdala/posterior orbito-frontal		4.43	−18	0	−12
Left anterior cingulate	32	3.83	−6	39	15
Left caudate nucleus		3.73	−15	12	3
Left cerebellum		3.61	−15	−45	−24
Left insula/putamen		3.53	−27	18	0
Left parahippocampal gyrus	35/36	3.71	−27	−30	−18
Left thalamus		4.13	−9	−6	0
Right inferior frontal/posterior orbito-frontal	47/25	3.65	24	12	−18
Right insula		3.61	45	6	0
Right parahippocampal gyrus	28/35	3.56	24	−21	−21
Right precuneus		3.94	15	−66	24
Right putamen/globus pallidus		3.38	15	3	6
Right thalamus		3.91	6	−12	12
Left SN/VTA (cluster size >5 voxels)		4.06	−9	−15	−18
Cue: causal vs. associative, at relational evaluation (Regressor 2)					
Left amygdala/posterior orbito-frontal		4.02	−21	3	−15
Left inferior frontal	47	3.75	−39	39	−6
	44	3.62	−54	12	24
	46	3.39	−42	42	15
	10	3.29	−33	51	12
Left middle frontal	10	3.64	−42	51	−6
	6/8	3.73	−42	15	48

Table 2. (*continued*)

	BA	z Value	MNI coordinates		
			x	*y*	*z*
Relational evaluation: causal vs. noncausal associative (Regressor 3)					
Cingulate gyrus	31	4.95	0	−33	45
Left caudate nucleus		3.85	−9	3	6
Left inferior parietal	40	5.26	−48	−66	42
Left middle frontal	9	5	−42	18	42
	46	3.67	−45	36	18
	10	3.58	−42	51	9
Left middle temporal	21	4.38	−63	−51	0
Left precuneus	7	4.06	−6	−78	45
Left superior frontal	8	4.14	−21	33	54
	8	3.92	−15	45	45
Left thalamus		3.94	−3	−27	3
Right cerebellum		4.07	36	−72	−45
Left orbito-frontal (p = .005)		3.64	−15	6	−18

If not stated otherwise in the table, the significance level was p = .001, uncorrected at a cluster size of at least 10 voxels. BA = Brodmann's area.

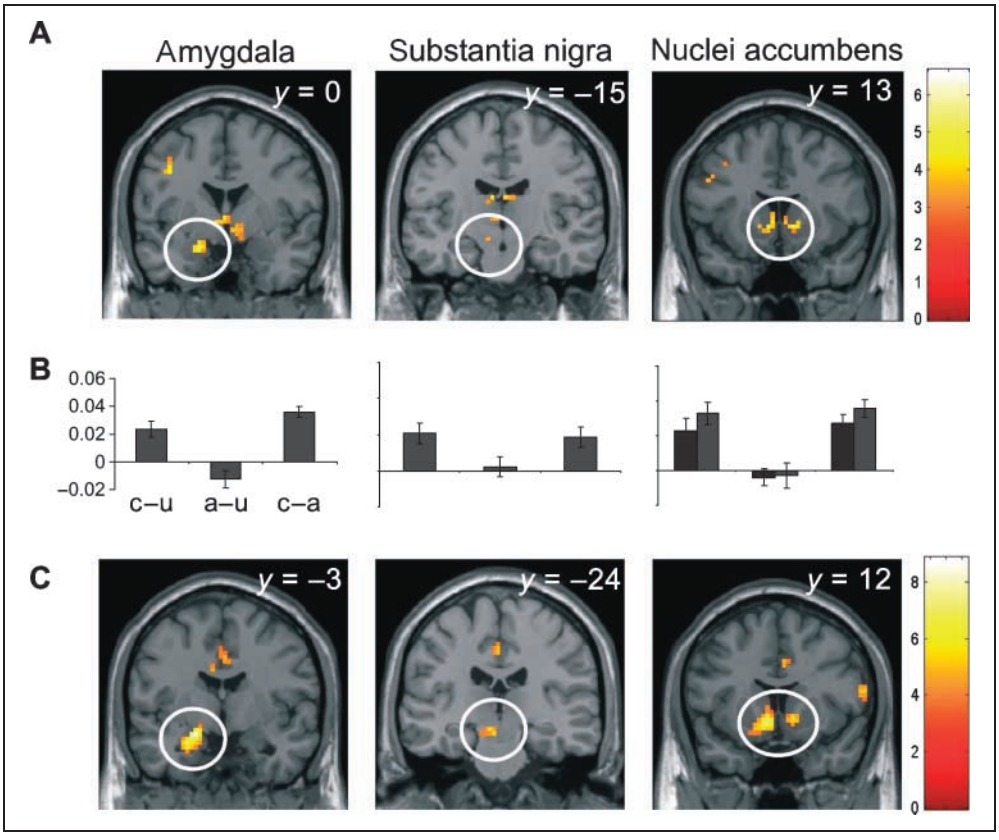

Figure 3. Activation found during relationship evaluation for Experiment 1. (A) SPM contrast images of the contrast causal versus unrelated word pairs exclusively masked with the contrast of noncausal associative word pairs versus unrelated word pairs. (B) Results of the ROI analysis. The bars depict the difference of the mean beta values, that is, causal minus unrelated (c − u), noncausal associative minus unrelated (a − u), and causal minus noncausal associative word pairs (c − a). Error bars refer to the *SEM*. The gray bars below the bilateral nucleus accumbens activation refer to the right and the black bars refer to the left nucleus accumbens activation. (C) SPM contrast images for causal versus noncausal associative word pairs. The scales refer to SPM *t* values.

versus noncausally associated word pairs. This comparison revealed a higher activation for the presence of a causal relationship in the left medial (BA 9) and inferior frontal gyri (BA 44, BA 45), right inferior frontal gyrus (BA 44), right orbito-frontal (BA 11), left amygdala, left and right nucleus accumbens, left posterior part of the SN/VTA, left parietal lobe (BA 40), and left cingulate gyrus (BA 31) (Table 2, Figure 3C).

Experiment 2: Behavioral Results

Behavioral data were analyzed with two separate repeated measures ANOVAs applying type of cue (causal vs. associative) and type of relationship (causal, noncausal associative, and unrelated) to measures of RTs and response accuracy. For the RTs, there was a significant main effect of factor Type of Cue, $F(1, 17) = 10.58, p < .01$. The responses were faster for the associative cue. Factor Type of Relationship also showed a significant main effect, $F(2, 34) = 42.12, p < .01$. Participants gave the fastest response to unrelated word pairs and slowest response to noncausal associative word pairs. The interaction of the two factors was also significant, $F(2, 34) = 12.42, p < .01$. Post hoc (LSD) testing revealed that for the causal cue, the responses to unrelated word pairs were the fastest and the responses to noncausal associative word pairs were the slowest (ps $< .05$). For the associative cue, the responses to unrelated word pairs and causal word pairs were significantly faster than the responses to noncausal associated word pairs, ps $< .05$ (Table 1). The ANOVA for the accuracy data revealed a significant main effect of Type of Cue, $F(1, 17) = 32.14, p < .01$. Participants gave more correct responses for the associative cue than for the causal cue. The factor Type of Relationship was also significant, $F(2, 34) = 50.66, p < .01$. Participants were most accurate if the word pairs were unrelated, and their accuracy was worst for noncausal associative word pairs. Finally, the Type of Cue × Type of Relationship interaction was significant as well, $F(2, 34) = 6.81, p < .01$. Post hoc testing (LSD) showed for the causal cue that participants had the highest accuracy for unrelated word pairs and the lowest for noncausal associative word pairs, ps $< .05$. They also gave significantly more correct response for causal and unrelated word pairs presented after the associative cue in comparison to noncausal associative word pairs, ps $< .05$ (Table 1).

fMRI Results

In Experiment 2, statistical analyses for Regressors 2 and 3 were calculated to separate task-related and meaning-related aspects of activity elicited by the word pairs themselves. With Regressor 3, we contrasted causal with noncausal associative word pairs (Figure 2) compatible with the fMRI analysis of Experiment 1. This showed a higher BOLD response for causal word pairs in the left inferior parietal lobe (BA 40), left posterior cingulate gyrus (BA 31), left middle frontal (BA 9), left inferior frontal (BA 46), left middle temporal (BA 21), left superior frontal (BA 8), left caudate nucleus, left thalamus, and at a more lenient threshold ($p = .005$) left OFC, regardless of the preceding cue (Table 2, Figure 4A). To differentiate task-related aspects of causal versus noncausal associative processing of word pairs independent of type of word pair (collapsed over causal, noncausal associative, and unrelated word pairs),

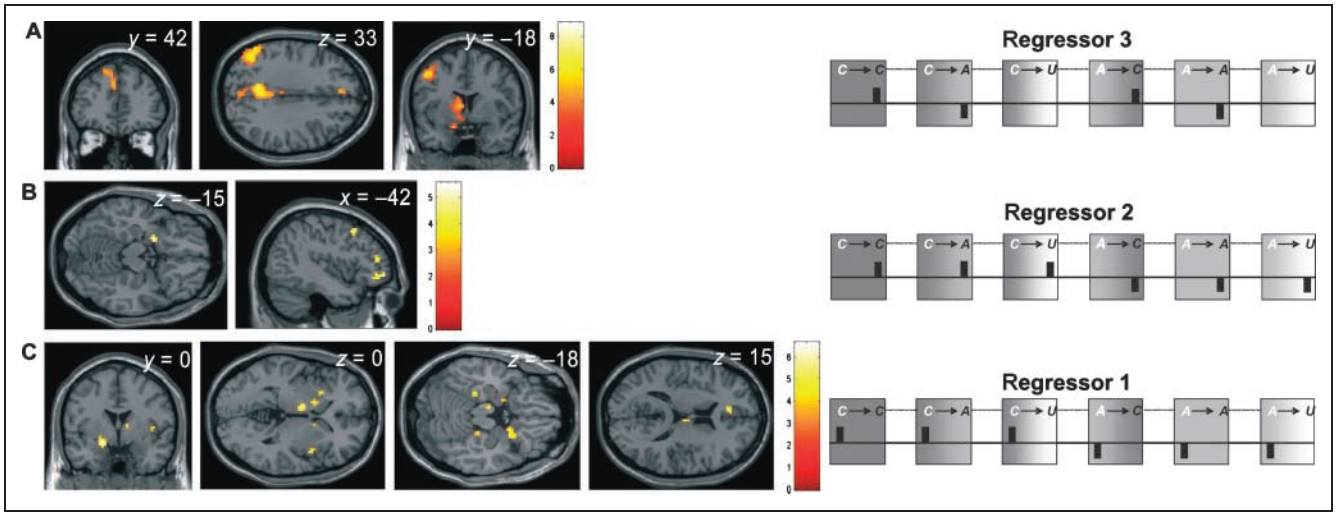

Figure 4. Significant activations of Experiment 2. Left side, activations found; right side, corresponding regressor. (A) SPM images for Regressor 3 (causal vs. noncausal associative word pairs) show a significant activation in the left superior/medial/frontal cortex, left inferior parietal cortex, posterior cingulate gyrus, left middle frontal cortex, left caudate nucleus, and left OFC ($p < .005$). (B) Regressor 2 (influence of causal vs. associative cue on relationship evaluation) yielded significant activation in the left amygdala/orbito-frontal region and left inferior and middle frontal gyri. The scales refer to SPM t values. (C) Higher preparatory activation (Regressor 1: causal vs. associative cue) was found in the left amygdala, bilateral insula, bilateral thalamus, bilateral caudate nucleus, bilateral parahippocampal gyrus, bilateral orbito-frontal, and left anterior cingulate. The scales refer to SPM t values.

we compared the activation following the causal cue with activation following the associative cue at the time when the second word was presented (Figure 2, Regressor 2). There was a stronger BOLD response following the causal cue in the left amygdala, left inferior (BA 10, 44, 46, 47), and middle frontal gyrus (BA 6, 8, 10), right middle frontal (BA 8, 10), and left caudate nucleus (Table 2, Figure 4B).

Preparatory Activity

Experiment 1 investigated the dissociation between the evaluation of a causal relationship and a noncausal associative relationship. Experiment 2 was designed to assess to what extent activity patterns observed in Experiment 1 were related to the task demands of assessing causal relationships (Regressor 1). Therefore, activation elicited by the causal cue was compared with the activation elicited by the associative cue (Figure 2, Regressor 1) to investigate different activations in terms of preparing for a causal or a noncausal associative judgment. The preparation for a causal judgment revealed higher activation in the left amygdala and OFC, left and right caudate nucleus, left and right insula, left thalamus, right OFC (BA 25/47), left SN/VTA, left and right parahippocampal gyrus (left BA 35, 36 and right BA 28/35), left anterior cingulate gyrus (BA 32), right precuneus, and the cerebellum (Table 2, Figure 4C).

DISCUSSION

In two event-related fMRI experiments, we sought to determine whether representations of cause–effect relationships are distinct from noncausal associative relational representations. We focused on the retrieval of preexisting, well-learned semantic causal knowledge. We show that the evaluation of cause–effect relationships engages a mesolimbic and mesocortical circuitry known to mediate prediction error learning. This indicates that prediction error circuitry is engaged when well-known causal associations are retrieved from semantic memory even under conditions when prediction error learning is not explicitly required. Such default engagement of prediction error circuitry in the representation of well-known cause–effect relationships that are already stored in semantic memory may allow for efficient updating to keep causal knowledge accurate.

Common Networks of Causal and Noncausal Semantic Processing

Experiment 1 gave us the opportunity to investigate commonalities of semantic processing of noncausal associative and causal relationships compared with semantic processing of unrelated word pairs. These common networks included areas, mainly left lateralized, in inferior parietal, inferior and superior frontal, middle temporal, thalamus, caudate nuclei, and occipital regions. These results are compatible with previous studies which have found that several of these areas are activated during semantic memory retrieval (Patterson et al., 2007; McDermott, Petersen, Watson, & Ojemann, 2003; Nyberg et al., 2003; Cabeza & Nyberg, 2000; Wiggs et al., 1999), selection of semantic information among competing alternatives (Thompson-Schill, D'Esposito, Aguirre, & Farah, 1997), and verbal working memory maintenance or manipulation (Veltman, Rombouts, & Dolan, 2003). It is unlikely that these activation differences between related (causal and noncausal associative) and unrelated word pairs were driven solely by difficulty because there was no systematic RT or response accuracy difference between related and unrelated pairs across both experiments.

Dissociating Causal Relationships from Noncausal Associative Relationships beyond Semantic Processing

To derive a clear distinction between causal and noncausal associative knowledge processing, we directly compared the evaluation of causal and noncausal associative word pairs. Activation in bilateral inferior, right orbitofrontal, left superior frontal regions, left parietal regions, left amygdala, and bilateral nucleus accumbens exclusively differentiated causal relationships from noncausal associative relationships. Activations in the OFC and the left amygdala as well as bilateral nucleus accumbens considerably overlap with regions from studies investigating prediction error learning for rewards (Schultz, 2006; Berns et al., 2001), contingency learning (Balleine et al., 2003), and decision making processes (Hampton, Adolphs, Tyszka, & O'Doherty, 2007; Wallis, 2007; Yang & Shadlen, 2007; Bechara, Damasio, & Damasio, 2003; Krawczyk, 2002). As in Experiment 1, causal versus noncausal associative relationships activated OFC and caudate (Regressor 3 in Experiment 2; however, unlike in Experiment 1, we observed no amygdala activation for this contrast in Experiment 2). This suggests that the activity difference between causal and noncausal associative relationships in both the orbito-frontal and the caudate holds independent of task cues. The analysis of task cue effect, on the other hand, showed that word pair processing following a causal cue activated posterior left OFC and amygdala irrespective of whether the word pairs denoted causal, noncausal associative, or unrelated relationships. Therefore, the data from Regressors 2 and 3 in Experiment 2 together with the data from Experiment 1 suggest that word pairs activate OFC and caudate more strongly when they denote causal as opposed to noncausal associative relationships. Irrespective of their meaning, word pairs activate posterior OFC and amygdala more strongly when word processing is directed toward causality rather than noncausal associative relationships.

Thus, the contribution of Experiment 2 was that it allowed us to investigate the pure task-related, preparatory aspects of retrieving a cause–effect relationship by

comparing activation specifically found at the time of the task cue. Cues indicating to prepare to retrieve a causal relationship as opposed to a noncausal associative relationship (Regressor 1) again elicited activation of a network including mesolimbic and mesocortical structures (SN/VTA and amygdala), the caudate nucleus, and the OFC. This cue-related finding effectively rules out the possibility that the differences between causal and noncausal associative relationships were driven solely by stimulus differences on a semantic level.

Causal Knowledge and Prediction Error Circuitry

Our data show that the retrieval of cause–effect relationships is distinguished from noncausal associative relationships by engaging activity of brain regions such as the SN/VTA and ventral striatum, which are known to code prediction errors during stimulus-reward learning (Schultz, 2007) and associative causal learning (Corlett et al., 2004). The amygdala and OFC, key players in contingency learning and decision making (Hampton et al., 2007), can also cooperatively contribute to learning from prediction errors. The OFC can signal the value of expected outcomes (Wallis, 2007), and these outcome expectancies are held to permit the rapid recognition of unexpected outcomes and prediction errors, thereby driving new learning through facilitation of associative flexibility in downstream regions, such as the amygdala. Hence, this set of structures together permits the representation of outcomes, prediction errors and allows associative flexibility for the updating of existing contingencies.

Involving this prediction error circuitry by default, that is, in the absence of an explicit requirement to make predictions or learn from outcomes, during retrieval of well-known causal relationships is a plausible mechanism that would allow the updating of stored causal knowledge on the basis of potential alterations of the cause–effect connections. Such a mechanism seems suitable for keeping causal knowledge accurate and maintaining knowledge adaptable in changing environments. The results of Experiment 2 also support this possibility and furthermore suggest that even a potential encounter with a cause–effect relationship involves activation of prediction error circuitry.

A striking aspect of our data is that both experiments did not actually allow for differences in explicit predictions to contribute to our findings. That is in both experiments, we did not explicitly manipulate prediction error processing. In Experiment 1, the first word of each pair was unpredictive as to whether the second word was related causally, noncausal associatively, or unrelated, and it is therefore implausible to explain the findings here by a differential engagement of participants in making explicit predictions about the second word of each pair. Likewise, in Experiment 2, the cue itself did not allow for any explicit predictions regarding the specific content of the upcoming word pair. Hence, the finding that only causal relationships in contrast to noncausal associative relationships engage prediction error circuitry was unrelated to explicit differential engagement in making predictions regarding cause and effect. Rather, the data suggest that, independent of actual explicit predictions, the prediction error network is activated by default when causality is an actual (Experiment 1) or an expected (Experiment 2) part of stored semantic relationships. One interesting difference between the two is the involvement of the ventral striatum with actual causality and the dorsal striatum with expected causality.

DA neuromodulation is critical for the coding of prediction errors of reward (Schultz, 2002, 2006; Schultz & Dickinson, 2000; Schultz et al., 1998; Schultz, Dayan, & Montague, 1997; Schultz, Apicella, Scarnati, & Ljungberg, 1992) but has not yet been implicated in the ability to update already acquired causal knowledge in the absence of any apparent reward-related reinforcement. Our data, particularly the finding that SN/VTA is activated by causality, raise the possibility that DA circuitry may indeed play a role also in updating existing causal knowledge. This leads to the interesting possibility that neurodegenerative epidemics that are characterized by dysfunction of DA circuitry, such as schizophrenia and Parkinson's epidemic, might be associated with impaired updating of causal knowledge leaving these patients less adaptive in changing environments. Hence, the reported reasoning impairments in schizotypy (Sellen, Oaksford, & Gray, 2005) as well as schizophrenia (Moore & Sellen, 2006), but also prediction error-dependent delusion formation in psychosis (Murray et al., 2008; Corlett et al., 2007), and the recently demonstrated problems of patients with Parkinson's epidemic to learn from prediction errors of rewards (Schott et al., 2007) and their problems in counterfactual reasoning (McNamara, Durso, Brown, & Lynch, 2003) may extend to a more general impairment of modifying existing causal relationships within semantic memory.

To summarize, we have shown that the representation of causal knowledge is different from noncausal associative knowledge and involves mesolimbic and mesocortical structures that are part of a prediction error circuitry. Thus, prediction error circuits are not only recruited during learning but also play a role in the representation of knowledge that has already been acquired earlier. These findings shed light on mechanisms that allow for a flexible updating of already acquired causal knowledge. In more general terms, they redefine the functional architecture of brain regions that are likely to contribute to updating semantic memories.

Acknowledgments

This study was supported by grants from the Deutsche Forschungsgemeinschaft (Klinische Forschergruppe "Kognitive Kontrolle," TP4) and the Volkswagen Foundation from the University of Magdeburg.

Reprint requests should be sent to Daniela B. Fenker, Department of Neurology II, Otto-von-Guericke University, Leipziger Strasse 44, 39120 Magdeburg, Germany, or via e-mail: daniela.fenker@med.ovgu.de.

REFERENCES

Andrzejewski, M. E., Spencer, R. C., & Kelley, A. E. (2005). Instrumental learning, but not performance, requires dopamine D1-receptor activation in the amygdala. *Neuroscience, 135,* 335–345.

Balleine, B. W., Delgado, M. R., & Hikosaka, O. (2007). The role of the dorsal striatum in reward and decision-making. *Journal of Neuroscience, 27,* 8161–8165.

Balleine, B. W., Killcross, A. S., & Dickinson, A. (2003). The effect of lesions of the basolateral amygdala on instrumental conditioning. *Journal of Neuroscience, 23,* 666–675.

Bechara, A., Damasio, H., & Damasio, A. R. (2003). Role of the amygdala in decision-making. *Annals of the New York Academy of Sciences, 985,* 356–369.

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.

Blaisdell, A. P., Sawa, K., Leising, K. J., & Waldmann, M. R. (2006). Causal reasoning in rats. *Science, 311,* 1020–1022.

Bortz, J. (2005). *Statistik für Human- und Sozialwissenschaftler* (6th ed.). Heidelberg: Springer Medizin Verlag.

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.

Cabeza, R., & Nyberg, L. (2000). Imaging cognition II: An empirical review of 275 PET and fMRI studies. *Journal of Cognitive Neuroscience, 12,* 1–47.

Cobos, P. L., López, F. J., Caño, A., Almaraz, J., & Shanks, D. R. (2002). Mechanisms of predictive and diagnostic causal induction. *Journal of Experimental Psychology: Animal Behavior Processes, 28,* 331–346.

Corbit, L. H., Muir, J. L., & Balleine, B. W. (2003). Lesions of mediodorsal thalamus and anterior thalamic nuclei produce dissociable effects on instrumental conditioning in rats. *European Journal of Neuroscience, 18,* 1286–1294.

Corlett, P. R., Aitken, M. R. F., Dickinson, A., Shanks, D. R., Honey, G. D., Honey, R. A. E., et al. (2004). Prediction error during retrospective revaluation of causal associations in humans: fMRI evidence in favor of an associative model of learning. *Neuron, 44,* 877–888.

Corlett, P. R., Murray, G. K., Honey, G. D., Aitken, M. R. F., Shanks, D. R., Robbins, T. W., et al. (2007). Disrupted prediction-error signal in psychosis: Evidence for an associative account of delusions. *Brain, 130,* 2387–2400.

Cromwell, H. C., Hassani, O. K., & Schultz, W. (2005). Relative reward processing in primate striatum. *Experimental Brain Research, 162,* 520–525.

de Borchgrave, R., Rawlins, J., Dickinson, A., & Balleine, B. (2002). Effects of cytotoxic nucleus accumbens lesions on instrumental conditioning in rats. *Experimental Brain Research, 144,* 50–68.

Fenker, D. B., Waldmann, M. R., & Holyoak, K. J. (2005). Accessing causal relations in semantic memory. *Memory & Cognition, 33,* 1036–1046.

Fletcher, P. C., Anderson, J. M., Shanks, D. R., Honey, R., Carpenter, T. A., Donovan, T., et al. (2001). Responses of the human frontal cortex to surprising events are predicted by formal associative learning theory. *Nature Neuroscience, 4,* 1043–1048.

Hampton, A. N., Adolphs, R., Tyszka, M. J., & O'Doherty, J. P. (2007). Contributions of the amygdala to reward expectancy and choice signals in human prefrontal cortex. *Neuron, 55,* 545–555.

Izquierdo, A., Suda, R. K., & Murray, E. A. (2004). Bilateral orbital prefrontal cortex lesions in rhesus monkeys disrupt choices guided by both reward value and reward contingency. *Journal of Neuroscience, 24,* 7540–7548.

Knutson, B., Taylor, J., Kaufman, M., Peterson, R., & Glover, G. (2005). Distributed neural representation of expected value. *Journal of Neuroscience, 25,* 4806–4812.

Knutson, B., & Wimmer, G. E. (2007). Splitting the difference: How does the brain code reward episodes? *Annals of the New York Academy of Sciences, 1104,* 54–69.

Krawczyk, D. C. (2002). Contributions of the prefrontal cortex to the neural basis of human decision making. *Neuroscience & Biobehavioral Reviews, 26,* 631–664.

Matsumoto, K., Suzuki, W., & Tanaka, K. (2003). Neuronal correlates of goal-based motor selection in the prefrontal cortex. *Science, 301,* 229.

Matsumoto, K., & Tanaka, K. (2004). The role of the medial prefrontal cortex in achieving goals. *Current Opinion Neurobiology, 14,* 178–185.

McDermott, K. B., Petersen, S. E., Watson, J. M., & Ojemann, J. G. (2003). A procedure for identifying regions preferentially activated by attention to semantic and phonological relations using functional magnetic resonance imaging. *Neuropsychologia, 41,* 293–303.

McNamara, P., Durso, R., Brown, A., & Lynch, A. (2003). Counterfactual cognitive deficit in persons with Parkinson's epidemic. *Journal of Neurology, Neurosurgery and Psychiatry, 74,* 1065–1070.

Moore, S. C., & Sellen, J. L. (2006). Jumping to conclusions: A network model predicts schizophrenic patients' performance on a probabilistic reasoning task. *Cognitive, Affective, & Behavioral Neuroscience, 6,* 261–269.

Murray, G. K., Corlett, P. R., Clark, L., Pessiglione, M., Blackwell, A. D., Honey, G., et al. (2008). Substantia nigra//ventral tegmental reward prediction error disruption in psychosis. *Molecular Psychiatry, 13,* 267–276.

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (1998). *The University of South Florida word association, rhyme, and word fragment norms*. Retrieved from http://w3.usf.edu/FreeAssociation/.

Noppeney, U., Phillips, J., & Price, C. (2004). The neural areas that control the retrieval and selection of semantics. *Neuropsychologia, 42,* 1269–1280.

Nyberg, L., Marklund, P., Persson, J., Cabeza, R., Forkstam, C., Petersson, K. M., et al. (2003). Common prefrontal activations during working memory, episodic memory, and semantic memory. *Neuropsychologia, 41,* 371–377.

O'Doherty, J. P., Dayan, P., Friston, K., Critchley, H., & Dolan, R. J. (2003). Temporal difference models and reward-related learning in the human brain. *Neuron, 38,* 329–337.

O'Doherty, J. P., Hampton, A., & Hackjin, K. (2007). Model-based fMRI and its application to reward learning and decision making. *Annals of the New York Academy of Sciences, 1104,* 35–53.

Ostlund, S. B., & Balleine, B. W. (2005). Lesions of medial prefrontal cortex disrupt the acquisition but not the expression of goal-directed learning. *Journal of Neuroscience, 25,* 7763–7770.

Patterson, K., Nestor, P. J., & Rogers, T. T. (2007). Where do you know what you know? The representation of semantic knowledge in the human brain. *Nature Reviews Neuroscience, 8,* 976–987.

Pavlov, I. P. (1927). *Conditioned reflexes: An investigation of the physiological activity of the cerebral cortex*. Translated and Edited by G. V. Anrep. London: Oxford University Press.

Pearce, J. M., & Hall, G. (1980). A model for Pavlovian conditioning: Variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychological Review, 87,* 532.

Postle, B. R. (2005). Delay-period activity in the prefrontal cortex: One function is sensory gating. *Journal of Cognitive Neuroscience, 17,* 1679–1690.

Postle, B. R., Berger, J. S., Taich, A. M., & D'Esposito, M. (2000). Activity in human frontal cortex associated with spatial

working memory and saccadic behaviour. *Journal of Cognitive Neuroscience, 12(Suppl. 2),* 2–14.
Rescorla, R. A., & Wagner, A. R. (1972). *A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and non-reinforcement.* New York: Appleton-Century-Crofts.
Rodriguez, P. F., Aron, A. R., & Poldrack, R. A. (2006). Ventral-striatal/nucleus-accumbens sensitivity to prediction errors during classification learning. *Human Brain Mapping, 27,* 306–313.
Rogers, T. T., Hocking, J., Noppeney, U., Mechelli, A., Gorno-tempini, M., Patterson, K., et al. (2006). Anterior temporal cortex and semantic memory: Reconciling findings from neuropsychology and functional imaging. *Cognitive, Affective, & Behavioral Neuroscience, 6,* 201–213.
Satpute, A. B., Fenker, D. B., Waldmann, M. R., Tabibnia, G., Holyoak, K. J., & Lieberman, M. D. (2005). An fMRI study of causal judgments. *European Journal of Neuroscience, 22,* 1233–1238.
Schon, K., Hasselmo, M. E., Lopresti, M. L., Tricarico, M. D., & Stern, C. E. (2004). Persistence of parahippocampal representation in the absence of stimulus input enhances long-term encoding: A functional magnetic resonance imaging study of subsequent memory after a delayed match-to-sample task. *Journal of Neuroscience, 24,* 11088–11097.
Schott, B. H., Niehaus, L., Wittmann, B. C., Schutze, H., Seidenbecher, C. I., Heinze, H.-J., et al. (2007). Ageing and early-stage Parkinson's epidemic affect separable neural mechanisms of mesolimbic reward processing. *Brain, 130,* 2412–2424.
Schultz, W. (2002). Getting formal with dopamine and reward. *Neuron, 36,* 241–263.
Schultz, W. (2004). Neural coding of basic reward terms of animal learning theory, game theory, microeconomics and behavioural ecology. *Current Opinion in Neurobiology, 14,* 139–147.
Schultz, W. (2006). Behavioral theories and neuophysiology of reward. *Annual Review of Psychology, 57,* 87–115.
Schultz, W. (2007). Behavioral dopamine signals. *Trends in Neurosciences, 30,* 203–210.
Schultz, W., Apicella, P., Scarnati, E., & Ljungberg, T. (1992). Neuronal activity in monkey ventral striatum related to the expectation of reward. *Journal of Neuroscience, 12,* 4595.
Schultz, W., Dayan, P., & Montague, R. R. (1997). A neural substrate of prediction and reward. *Science, 275,* 1593.
Schultz, W., & Dickinson, A. (2000). Neuronal coding of prediction errors. *Annual Review of Neuroscience, 23,* 473.
Schultz, W., Tremblay, L., & Holland, A. J. (1998). Reward prediction in primate basal ganglia and frontal cortex. *Neuropharmacology, 37,* 421–429.
Sellen, J. L., Oaksford, M., & Gray, N. S. (2005). Schizotypy and conditional reasoning. *Schizophrenia Bulletin, 31,* 105–116.
Shanks, D. R., & López, F. J. (1996). Causal order does not affect cue selection in human associative learning. *Memory and Cognition, 25,* 125–134.
Tanaka, S. C., Balleine, B. W., & O'Doherty, J. P. (2008). Calculating consequences: Brain systems that encode the causal effects of actions. *Journal of Neuroscience, 28,* 6750–6755.
Thompson-Schill, S. L., D'Esposito, M., Aguirre, G. K., & Farah, M. J. (1997). Role of left inferior prefrontal cortex in retrieval of semantic knowledge: A reevaluation. *Proceedings of the National Academy of Sciences, U.S.A., 94,* 14792–14797.
Turner, D. C., Aitken, M. R., Shanks, D. R., Sahakian, B. J., Robbins, T. W., Schwarzbauer, C., et al. (2004). The role of the lateral frontal cortex in causal associative learning: Exploring preventative and super-learning. *Cerebral Cortex, 14,* 872–880.
Veltman, D. J., Rombouts, S. A., & Dolan, R. J. (2003). Maintenance versus manipulation in verbal working memory revisited: An fMRI study. *Neuroimage, 18,* 247–256.
Waldmann, M. (1996). *Knowledge-based causal induction: Vol. 34. Causal learning*. San Diego, CA: Academic Press.
Waldmann, M., Holyoak, K. J., & Fratianne, A. (1995). Causal models and the acquisition of category structure. *Journal of Experimental Psychology: General, 124,* 181–206.
Waldmann, M. R. (2000). Competition among causes but not effects in predictive and diagnostic learning. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 26,* 53–76.
Waldmann, M. R. (2001). Predictive versus diagnostic causal learning: Evidence from an overshadowing paradigm. *Psychonomic Bulletin & Review, 8,* 600–608.
Waldmann, M. R., & Holyoak, K. J. (1992). Predictive and diagnostic learning within causal models: Asymmetries in cue competition. *Journal of Experimental Psychology: General, 121,* 222–236.
Wallis, J. D. (2007). Orbitofrontal cortex and its contribution to decision-making. *Annual Review of Neuroscience, 30,* 31–56.
Wiggs, C. L., Weisberg, J., & Martin, A. (1999). Neural correlates of semantic and episodic memory retrieval. *Neuropsychologia, 37,* 103–118.
Wittmann, B. C., Schott, B. H., Guderian, S., Frey, J. U., Heinze, H. J., & Duzel, E. (2005). Reward-related fMRI activation of dopaminergic midbrain is associated with enhanced hippocampus-dependent long-term memory formation. *Neuron, 45,* 459–467.
Yang, T., & Shadlen, M. N. (2007). Probabilistic reasoning by neurons. *Nature, 447,* 1075–1080.