Cognitive control and binding in context-based decision-making : normal and dopamine deviant populations

Wouwe, N.C. van

Chapter 3

Goal- versus stimulus-driven components of biased decision-making
A functional neuroimaging study of the AX-CPT task

Abstract

Working-memory maintenance of rules or intentions is crucial in acting adaptively, but decisions can be also biased by more stimulus-driven factors such as the automatic reactivation of features that previously accompanied the current event. By means of an event-related fMRI study, we aimed to distinguish the contribution of rule-based control processes and automatically learned cue-probe associations (bindings) in an adapted *AX*-continuous performance task, using words (cues) followed after a 2-8s interval by pictures of faces and houses (probes). The subjects' task was to respond to a target probe, given that it was preceded by a specific cue. The current study shows that top-down control and bindings both explain part of the variance in performance during such biased decision-making. The data demonstrated enhanced ACC and DLPFC activation and impaired performance on trials that call for increased top-down control (*AY* and *BX*) compared to target trials (*AX*). Furthermore, we provided insight into the way the additional binding-related performance changes may be represented in the brain. That is, if a specific cue stimulus was followed by a face on previous occasions, subsequent presentation of face probes with these cue stimuli showed increased reactivation in the 'face area' in the brain (FFA) with an improvement in performance, compared to face probes presented subsequent to cues previously associated with a house. We found similar results for cue stimuli followed by houses, which reactivated the 'house area' (PPA) upon probe presentation, also with an improvement in

performance. This latter pattern of activation and performance costs was predicted exclusively by the binding account.

Introduction

Top-down guidance of behavior in a complex and dynamically changing world is often based on information held in working memory. Such guidance serves to bias decision-making processes in directions consistent with externally set rules or internally maintained intentions. Occasionally, the information held in working memory may be inadequate for the task at hand, leading to performance costs. For instance, failures to update goals or intentions in WM lead to perseverative behaviour with tasks that require switching from one task rule to another (Braver et al., 2001). Orthogonal to this goal-driven guidance, decisions may be biased also by stimulus-driven factors, such as the automatic reactivation of episodic associations that accompanied the current stimulus in a previous instance. For instance, if a particular stimulus was accompanied by a particular action in the recent past, then the re-occurrence of that stimulus may trigger the same action again (Hommel, 2004; Logan, 1988). Such stimulus-driven biases typically benefit rapid and adequate decision-making. Occasionally, however, biases derived from episodic retrieval can be detrimental -- especially when some, but not all of the features of the current stimulus event coincide with features of a recent episodic memory trace, such as when a stimulus was associated with one action in a previous instance but currently designates a different action (Hommel, 1998).

Goal-driven and stimulus-driven factors in biased decision-making are typically investigated in separate studies, but may well apply simultaneously, sometimes converging on similar decisions but oftentimes yielding competing outcomes. The present study used functional magnetic resonance imaging (fMRI) to assess and disentangle the contributions of goal-driven and stimulus-driven biases in decision-making, using an *AX* version of the continuous performance task (*AX*-CPT; Rosvold, Mirsky, Sarason, Bransome, & Beck, 1956).

Goal-driven accounts

Behavioral flexibility refers to behavior that is adaptively tailored to changing situational demands, while at the same time abiding by goals and intentions. Such flexible agency is thought to call for some mechanism of cognitive control. Cognitive control relies on processes that are involved in adequate decision-making, such as biasing decisions towards task-relevant stimuli and actions, and updating and maintaining this bias when facing irrelevant incoming information or being challenged by prepotent but inappropriate action tendencies (Miller & Cohen, 2001). Another aspect of cognitive control involves the on-line evaluation of decisions and performance, and signaling the need for increased control to overcome decision conflict or to prevent erroneous performance on future trials (Botvinick, Braver, Barch, Carter, & Cohen, 2001; Botvinick, Cohen, & Carter, 2004; Ridderinkhof, Ullsperger, Crone, & Nieuwenhuis, 2004).

Failures to implement and maintain a goal-driven decision bias may lead to distractibility, such as seen when task-irrelevant information captures attention and elicits inappropriate responses in conflict situations such as in the Stroop color-word task (Stroop, 1935) or the Eriksen flanker task (Eriksen & Eriksen, 1974). Failures to update the current decision bias are reflected in perseverative behaviour under conditions that actually require a switch from one set of task rules to another, as in set shifting experiments (Jersild, 1927) or in the Wisconsin Card Sorting Task (Grant & Berg, 1948; Milner, 1963). These two types of decision-making failures occur frequently in the *AX*-CPT (Rosvold, Mirsky, Sarason, Bransome, & Beck, 1956); a particularly wellsuited task to investigate cognitive control, which will also be focus of the current study.

On each *AX*-CPT trial participants are presented with a cue stimulus and a subsequent probe stimulus, and are instructed to respond if a target probe (*X*) is immediately preceded by a specific cue (*A*) but to refrain from responding in all other sequences (*AY, BX* or *BY*). Target trials (*AX*) typically occur on the majority of trials in the *AX*-CPT task; this frequency induces a strong bias to issue a target response, even when either the cue (*BX*) or the probe (*AY*) designates that a response be withheld.

Braver and colleagues (Braver & Barch, 2002; Braver, Satpute, Rush, Racine, & Barch, 2005) described goal-driven cognitive control in the *AX*-CPT in terms of a Context

Processing Model. They used the term context representation to refer to the goal representations that influence planning, behavior and attentional processes. Top-down control can be exerted because the context information biases or primes the activation of a response or goal, as previously connected with that particular context information, in subsequent trials. For instance, successful performance on *BX* trials is often interpreted as a result of top-down control. According to their model, performance costs occur if a context primes an incorrect response, for example in *AY* trials, or if subjects fail to maintain the relevant context information (Braver et al., 2001). However, given the frequent observation that control can be confounded with episodic effects (e.g., Mayr, Awh, & Laurey, 2003; Nieuwenhuis et al., 2006) it seems relevant to consider whether they may play a role in the *AX*-CPT task as well. Episodic bindings in the *AX*-CPT may arise between specific cue and probe stimuli and reactivation of competing information as a result of these episodic bindings may bias performance on subsequent trials. Therefore the present study sought to disentangle the contribution of top-down control and episodic bindings in an *AX*-CPT task.

A number of studies have related top-down control processes to activity in prefrontal brain regions, including the dorsolateral prefrontal cortex (DLPFC) and anterior cingulate cortex (ACC) (Miller & Cohen, 2001). DLPFC activity increases with task related working-memory demands (Cohen et al., 1997), but also during other control operations, such as selecting between competing responses (Bunge, Hazeltine, Scanlone, Rosen, & Gabrieli, 2002; Rowe, Toni, Josephs, Frackowiak, & Passingham, 2000), allocating attention to task-appropriate behavior (MacDonald, Cohen, Stenger, & Carter, 2000), and with the reorganization of items in working memory (Blumenfeld & Ranganath, 2006). The ACC is assumed to play an evaluative role in decision-making by monitoring performance for unfavorable outcomes, response conflict, or errors and by signaling the need for adjustments in control (Botvinick et al., 2001; Ridderinkhof et al., 2004). Trials with competing stimulus or response information in Stroop color-word and Eriksen flanker tasks are associated with increased performance costs and activity in ACC, which then triggers the DLPFC to resolve the conflict (Durston et al., 2003; Kerns et al., 2004; MacDonald et al., 2000) by biasing posterior brain areas (i.e. parietal cortex) (Desimone & Duncan, 1995; Dove, Pollman, Schubert, Wiggins, & von Cramon, 2000; Gruber, Karch, Schlueter, Falkai, & Goschke, 2006) and enhance control on subsequent trials or the current trial (DePisapia & Braver, 2004).

Stimulus-driven accounts

Instead of via proactive preparation, a decision can also be biased by currently available stimulus information that reactivates previously associated information from episodic memory. The feature integration account advocated by Hommel (2004; Hommel, Proctor & Vu, 2004) assumes that the features of the stimuli encountered and the responses carried out in a situation are bound together into an episodic memory trace that is reactivated (retrieved) by the repetition of any of the features it contains. If this reactivation includes stimulus or action features that are incongruent with the stimuli or responses to be processed in the current trial, performance is impaired, either because previous associations need to be "undone" or because the conflict they induce needs to be overcome. In contrast, in situations where stimulus and action features are all repeated (so that the reactivation does not induce conflict) or all changed (so that no reactivation occurs), performance costs are absent (Hommel, 1998, 2004). A neuroimaging study by O'Craven, Downing, and Kanwisher (1999) contributed to the understanding of how these episodic bindings, or objects files as originally posed by Kahneman, Treisman, and Gibbs (1992), may be represented in the brain. They showed that attending to an event results in the cross-referencing and co-activation of the relevant and irrelevant features of this event, as reflected by brain activation in the corresponding brain areas. Follow-up work of Keizer et al. (2008) showed that this co-activation of object features indeed creates bindings between them, so that repeating one feature increases the activation of brain areas coding for the previously bound fellow-feature.

Performance costs previously attributed exclusively to top-down control operations might be explained at least in part, in terms of this binding account. Studies on negative priming (Huang, Holcombe, & Pashler, 2004; Tipper, 2001), inhibition (Verbruggen, Logan, Liefooghe, & Vandierendonck, 2008), task switching (Waszak, Hommel, & Allport, 2004), and spatial incompatibility (Hommel et al., 2004) indicated that the reactivation of competing information as a result of a retrieved episodic binding created on a previous trial bias performance on the current trial. Likewise, the conflict-adaptation effect in the Eriksen flanker task can also be explained in part in terms of repetition of specific stimulus episodes and may thus not necessarily involve

monitoring response conflict. While some authors find that conflict-adaptation effects are eliminated if stimulus repetitions are excluded (e.g., Mayr et al., 2003; Nieuwenhuis et al., 2006), others find that conflict-adaptation effects remain (e.g., Notebaert, Gevers, Verbruggen, & Liefooghe, 2006; Ullsperger, Bylsma, & Botvinick, 2005). Neurophysiological studies investigating conflict-driven adjustments of cognitive control in prefrontal regions are consistent with predictions from the conflict-adaptation account of control (Egner & Hirsch, 2005; Kerns et al., 2004, Stürmer, Leuthold, Soetens, Schröter, & Sommer, 2002), but they do not rule out contributions from episodic retrieval.

With respect to the *AX*-CPT, similar findings have been observed. In a previous behavioral study, we found initial evidence indicating that performance costs in the *AX*-CPT, typically attributed to top-down control operations, might be also explained in part in terms of binding between specific cue and probe stimuli (van Wouwe, Band & Ridderinkhof, 2008). Since both top-down goal-driven accounts and stimulus-driven accounts explain behavioral performance costs in several experimental paradigms (Verguts & Notebaert, 2008; for a review see Egner, 2007) it seems important to assess and try disentangling their contribution in terms of their underlying neural mechanisms by means of an *AX*-CPT task.

Present study

The current fMRI study investigates neural mechanisms of both top-down control and binding in an fMRI-adapted *AX*-CPT task, using words (cues) and pictures of faces and houses (probes). The *AX*-CPT paradigm is a modified version of the classical Continuous Performance Test (CPT; Rosvold et al., 1956) and performance costs have traditionally been explained strictly in terms of top-down control processes.

On each *AX*-CPT trial participants are presented with cues and probes (word and picture stimuli, respectively) on a computer screen. Subjects are instructed to respond to every word with a nontarget response, and to every picture with either a target or nontarget response by manually pressing a button. A target response (for example a left index finger press) is required if the target probe picture (*X*) is immediately preceded by a certain context cue-word (*A*). In every other case, for example in *AY*, *BX* or *BY*

sequences, participants are to respond to the probe with a nontarget response button (e.g., left middle finger). As pointed out earlier, target sequences (*AX*) typically occur on the majority of trials in the *AX*-CPT task, so to induce a strong bias to issue a target response even on trials other than *AX* (*BX* and *AY*).

Functional MRI was used to verify whether stimulus-driven performance costs can indeed be explained by reactivation of features previously associated with the imperative stimulus. Previous studies have shown that tasks involving face and house stimuli activate distinguishable brain areas in the occipitotemporal cortex; the fusiform face area (FFA) and parahippocampal place area (PPA), respectively (Epstein & Kanwisher, 1998; Kanwisher, McDermott, & Chun, 1997). We used this selective activity in FFA and PPA as an index of whether cue stimuli reactivate the face or house probes that they were associated with on previous events (Keizer et al., 2008). To investigate the episodic binding effects and to be able to measure this activity in the FFA and PPA, the traditional use of characters (*A, B, X* and *Y*) was replaced by words and pictures of faces and houses. This choice allows participants to distinguish between occurrences of specific context cues and probes on individual trials and to create unique episodic bindings for various combinations of words and pictures. Previous research by Colzato, Raffone and Hommel (2006) showed that the bindings between stimulus features are most powerful when using real-life pictures as compared to arbitrary feature conjunctions, presumably because real objects trigger top-down attentional processes (Hommel & Colzato, 2009).

Episodic bindings in the *AX*-CPT may arise between specific cue-words and probe-pictures, affecting performance on subsequent trials in the task. During a behavioural training session, specific word-cues were presented exclusively with face-probes, while other word-cues were presented exclusively with house-probes, to induce unique episodic bindings. During the actual experiment, the same cue-probe combinations were used again, in addition to novel cue-probe sequences in which the probe categories were changed, that is, words bound to houses were now presented with faces and vice versa.

Predictions

According to the goal-driven account, the biasing influence of the cue held in working memory (WM) will affect performance irrespective of the precise episodic binding (cue-probe combination of events) as instigated by the cue word of previous trials. With

intact top-down control (i.e., adequate use of cue information to perform the task) *AX*-CPT performance should be faster and more accurate on *AX* than on *AY* and *BX* trials. On *AY* trials, subjects are likely to falsely expect the appearance of a target probe after an *A*-cue and are thus inclined to respond with an incorrect (target) response (false alarm). In *BX* trials, the efficient use of the *B*-cue may result in correct rejection; however, performance in *BX* trials is often hampered by the *X*-probe, which is strongly associated with an *A*-cue and a target response.

A pure binding account of CPT performance, on the other hand, would suggest performance costs in cue-probe pairs if, and only if, the specific probe in the current cue-probe pair is non-identical to the probe that was previously associated with that specific cue. That is, if the cue was previously bound in an episodic memory trace with a specific probe, costs should occur if that cue now has to be unbound from this probe or because the probe-induced conflict needs to be resolved. Performance on *AY* and *BX* pairs should be less hampered if the probe (cue) is identical to the probe (cue) that was previously associated with the same cue (probe); a prediction that fits well with the additive effects of binding and top-down control observed in other paradigms (Notebaert et al., 2006).

According to predictions derived from the binding account, presentation of a cue-word (irrespective of the nature of the subsequent probe) will reactivate the brain area related to the probe that previously accompanied it (either house or face). This prediction presumes that the cue stimulus directly primes the previously associated probe information; however, the cue could instead lower the threshold for processing the probe which may affect probe based reactivation of associated information and improve performance (Huber, 2008). In addition, upon presentation of a probe-picture, FFA or PPA will be activated more if cue and probe features (house or face) match according to their previous association (complete repetition), than if they do not match (partial repetition). Note that during the experiment, these word-cues are randomly presented with different probes (houses and faces) independent of their previous association. Thus, based on frequency, word-cues do not selectively prime a face or house probe. Furthermore, FFA and PPA reactivation are expected to co-vary with performance costs, in that individuals who show greater differences in FFA/PPA reactivation between partial and complete repetitions should also show greater performance costs.

Additionally, fMRI was used to test the predictions derived from the context-processing model of top-down control with respect to activity in prefrontal regions. On this perspective, DLPFC activation is thought to reflect cue-induced goal-driven processes that bias decision-making about probe identity, whereas ACC activation is thought to reflect probe-induced conflict between the response as biased by the cue and the response actually required by the probe (on probes in *AY* and *EX* sequences). Based on the predictions from the binding account, however, we expect DLPFC and ACC to be recruited only if cue-based reactivation of associated features mismatches probe features of the current event (partial repetition of cue-probe sequences), but not if information reactivated by the cue matches the currently presented probe (complete repetition of stimuli). This prediction fits with evidence from fMRI studies demonstrating that DLPFC (Liston, Matalon, Hare, Davidson, & Casey, 2006) and ACC are indeed activated by competing stimulus information (Milham, Banich, & Barad, 2003; Van Veen & Carter, 2005). The goal-driven account, however, predicts that ACC and DLPFC will be exclusively recruited in trials that are biased by context information probing a currently incorrect response (*AY* and *BX*).

Method

Participants

Sixteen right-handed adults (all female), average age 22.25 years, participated in this fMRI study. Two participants had to be excluded from the fMRI analysis, because they were not able to hold their heads steady enough to avoid noticeable head motion throughout the fMRI recordings. All participants reported being in good health, with normal (or corrected-to-normal) vision and no history of psychiatric or primary degenerative neurological disorder; none were taking psychotropic medications. The experiment was conducted in accordance with relevant laws and institutional guidelines and was approved by the local ethics committee from the Faculty of Social Sciences and the Medical Ethical Committee of Leiden University Medical Center. The participants

gave written informed consent and received either course credits or €20 remuneration for their participation.

Figure 1

Overview of Trialtypes in the adapted AX-CPT. Fourteen word-cues (1) were presented with faces and 14 word-cues were presented with houses during a behavioral training (complete-repetition trials) with the features of each of the AX-CPT conditions, i.e. words in uppercase (A) or lowercase (B) and pictures rotated clockwise (X) or counterclockwise (Y). These trained pairs were presented again during the experiment, but also succeeded by probes from the category not trained with (partial-repetition trials).

AX-CPT-Condition	***Bindings***			
	Complete-repetition		***Partial-repetition***	
AX	Word-cues (1) & face-probe	**Word-cues (2) & house-probe**	Word-cues (1) & house -probe	Word-cues (2) & face -probe
AY	Word-cues (1) & face-probe	Word-cues (2) & house-probe	Word-cues (1) & house -probe	Word-cues (2) & face -probe
BX	Word-cues (1) & face-probe	Word-cues (2) & house-probe	Word-cues (1) & house -probe	Word-cues (2) & face -probe
BY	**Word-cues (1) & face-probe**	Word-cues (2) & house-probe	Word-cues (1) & house -probe	Word-cues (2) & face -probe

Example of the trialtypes presented in ***bold****. Subjects were instructed to respond with presentation of the probe-picture. They had to press a target response button if a clockwise rotated picture followed an uppercase word (AX) and in any other case press nontarget response button (AY, BX, BY).*

Design

The current study used a word-picture version of the *AX*-CPT task to measure reaction time (RT) and errors in each condition. Figure 1 presents an overview of the current design and two example trials. During each *AX*-CPT trial participants were presented with a sequence of stimuli on a computer screen, containing a cue word in uppercase (*A*) or lowercase (*B*) and a probe picture (a face or a house) rotated clockwise (*X*) or counterclockwise (*Y*).

Participants had to respond to the target *X*-probe (a face or a house that is rotated clockwise) with their left index finger, but only if this *X*-probe was immediately preceded by an *A*-cue (a word in uppercase). In every other case (*AY*, *BX* or *BY* sequences), subjects had to respond with a nontarget button-press using their left middle finger. *AX* trials (uppercase words followed by a picture rotated clockwise) occurred very often during the experiment (50%), in order to induce a strong tendency to make a target response to the *X*-probe. The remaining sequences (*AY, BX* and *BY*) conditions occurred equally frequently (14.1 % each); the remaining 7.7% of the trials contained No-go probes, which required participants to refrain from responding and were included to ascertain that attention would be sustained following the *B* cue. The assignment of response keys, word case, and the probe characteristics was counterbalanced across subjects.

Additionally, the *AX*-CPT included complete repetitions and partial repetitions of the subtypes *AX*, *AY*, *BX* and *BY* trials to investigate the effect of episodic bindings. Specific pairs of word-face and word-house bindings (counterbalanced for rotation direction and capitalization across *AX*-CPT conditions) were presented during a behavioral training session preceding the fMRI experiment. These complete repetitions of cue and probe each recurred eight times during training. In the experimental session, the same complete-repetitions pairs were presented again on 50% of the trials. The other 50% of the trials consisted of the same cues, but now randomly paired with a probe picture from the stimulus category opposite to that included in the complete repetition of cue and probe (for example, a novel house picture instead of the face picture that was previously paired with the cue). These trials are referred to as partial repetitions of cue and probe.

Stimuli and apparatus

The cues and probes that represent analogues of the *A, B, X, Y* letter stimuli from the classic *AX*-CPT consisted of words (cues) and equiluminant grayscale frontview photographs of houses (14) and faces (14) (probes), cropped to a square size of 10° by 10°. A no-go probe consisted of a red octagon with the word "stop" printed on it. The Dutch words (28) were selected from the Celex database (Burnage, 1990). They contained three to six characters and were conceptually as unambiguous as possible. Selected words were comparable in frequency (the average frequency of occurrence of the words chosen was 128 times per million words).

Each trial started with a cue word (500 ms), followed by a jittered interstimulus interval (2000-8500 ms, average jitter 2731 ms). Then the imperative stimulus appeared (a picture of a face or house; 500 ms) followed by a jittered intertrial interval (2000-8500 ms, average jitter 2731 ms).

Participants responded to the probe stimuli by pressing either a target or non-target response button on an fMRI-compatible response-button box. The experiment was programmed in E-prime (Version 1.2; Psychology Software Tools Inc.). Stimuli were projected on a screen at the back of the scanner, which participants viewed through a mirror mounted on the head coil.

Imaging details

Images were recorded with a Philips Achieva 3-T MR scanner (Philips Medical Systems, Best, the Netherlands). Functional images were acquired using a SENSE parallel imaging Gradient Echo EPI sequence of 38 axial slices (resolution = 2.75 mm^3 isotropic; repetition time [TR] = 2.211 ms; echo time [TE] = 30 ms; flip angle = 80°; field of view = 220 mm; matrix = 80 x 80). For each experimental run scanning was started after the subject had read the instructions for that run. Image acquisition was stopped when the subject finished that particular run, resulting in an average number of 280 volumes for the localizer scans run (see below) and 465 volumes for each of the three experimental blocks (experimental run). A T1-weighted structural image (1.2 mm^3 isotropic) and a high-resolution EPI scan (2 mm^3 isotropic) were obtained for registration purposes.

Procedure

Before the start of the fMRI experiment, subjects were familiarized with the *AX*-CPT in a 10-minute behavioral training session outside the scanner, which also served to implicitly induce episodic bindings between specific words and faces or houses (complete repetitions of cue and probe). The actual fMRI experimental session consisted of two runs; a localizer scans run and an experimental scans run. During the localizer run, subjects were instructed to passively view words, faces and houses, which enabled us to identify stimulus-category-selective Regions Of Interest (ROIs). Houses, faces and words (similar to the experimental stimuli) were shown in separate blocks for 700 ms followed by a fixation cross of 300 ms. We presented subjects with three blocks of 28 faces (14 unique faces were presented in both rotation directions), three blocks of 28 houses (14 unique houses were presented in both rotation directions) and four blocks of 28 words (28 words were presented randomly in uppercase or lowercase) in mixed order, each with a duration of 30 s. In between, there were blocks of 30 s rest, eleven in total. The localizer run lasted for about 10 minutes.

The experimental run consisted of three *AX*-CPT blocks, each block contained two mini-blocks, one mini-block containing 78 trials (total of 468 trials) with a duration of 8.5 minutes. Subjects were informed that they would receive a break after each mini-block (during which they also would obtain feedback about their average performance

(e.g., "try to perform more accurately") to encourage them to perform as accurately and quickly as possible. Subjects viewed sequences containing pairs of cues (words) and probes (pictures) on the computer screen. Participants had to respond to the clockwise-rotated picture (*X* target probe) with their left index finger, but only if this picture was immediately preceded by a word in uppercase (*A* cue). In all other cases (*AY, BX* or *BY* sequences; picture rotated counterclockwise or word in lowercase), subjects had to respond to the probe by pressing a non-target button with their left middle finger. Complete repetitions and partial repetition cue-probe combinations were presented in a randomized order.

Data analysis

Behavioral data

Trials with RTs shorter than 150 ms or longer than 2000 ms were removed from the analysis. Performance on each trial type (*AX, AY, BX,* and *BY*) was measured by mean error percentage and mean RT (for correct responses). *AX*-CPT trials were separated according to their trained association because we will use face and house behavioral performance in the fMRI analysis; cue-stimuli originally trained with faces and during the experiment also presented with faces (complete repetition) or with houses (partial repetitions) and cue stimuli trained with houses and during the experiment presented with houses (complete repetition) or with faces (partial repetitions).

The data were analyzed by means of a three-factor repeated-measures ANOVA, with the factors 'Feature Repetition' (complete, partial), 'Cue Association' (face, house) and 'CPT Condition' (*AX, AY* and *BX*), separately for RTs and error rates. Specific predictions were tested by using simple contrasts to compare *AY* and *BX* with the *AX* condition. Subsequently, a two-factor repeated-measures ANOVA analysis, with the factors 'Feature Repetition' and 'CPT Condition' was performed on face-trained and house-trained trials separately. Similar to the *AX-AY-BX* analysis, *BY*, *AY* and *BX* trials were compared by means of a repeated-measures ANOVA, with the factors 'Feature Repetition' (complete, partial), 'Cue Association' (face, house) and 'CPT Condition' (*BY, AY* and *BX*), separately for RTs and error rates. Again, face-trained and house-trained

trials were analyzed separately with a two-factor repeated-measures ANOVA, with the factors 'Feature Repetition' and 'CPT Condition'.

fMRI Data analyses

The pre-processing of the images and the statistical analyses were done using FSL (Smith et al., 2004; Woolrich et al., 2009). Before pre-processing the data was down-sampled to voxelsize 3.3 mm^3. The following preprocessing statistics were applied to the functional data: motion correction (Jenkinson, Bannister, Brady, & Smith, 2002), slice-time correction, spatial realignment to the first volume of that run, spatial smoothing using a fullwidth at half maximum Gaussian kernel of 8 mm, removal of non-brain tissue (Smith, 2002) and mean-based intensity normalisation of all volumes. The localizer scans run was temporally high-pass filtered with a cut-off of 180 seconds and the experimental run with a cut-off of 100 seconds to remove low-frequency artefacts using Gaussian-weighted least-squares straight line fitting. Time-series statistical analysis was carried out using a linear model with local autocorrelation correction (FMRIB's Improved Linear Model, Woolrich, Brady, & Smith, 2001).

Event-related regressors (EV's) of the presented word, house, and face stimuli were computed for the localizer scans run, related to stimulus presentation of 700 ms. For the three experimental runs, we computed regressors for every cue-probe combination of the *AX*-CPT, separately for complete repetitions and partial-repetition trials (correct trials), related to stimulus presentation of 500 ms. Error trials, non-responses, pauses, feedback screens, and stop trials were modeled separately and not included in the analysis. FSL FEAT (FMRI Expert Analysis Tool, version 5.9) was used to calculate contrasts based on cue- and probe-related activity from different experimental conditions, for each of the experimental blocks. The contrasts of interest (with *t*-values) and regions of interest (ROIs; see below) used are reviewed in Table 1. To analyze the data from the three experimental blocks, we first calculated the mean effect of the contrasts within subjects and then calculated the mean contrast effects across subjects (FMRIB's Local Analysis of Mixed Effects, Beckman, Jenkinson & Smith, 2003) with *z*-statistic images (Gaussianised *t*-statistics), thresholded at $p = 0.001$ (uncorrected).

We created five masks that were used as ROIs in the analyses of the experimental run; three for pre-masking the analysis testing binding reactivation effects in right FFA and left and right PPA, and two for pre-masking the analysis that examines the binding

account and goal-driven account with respect to activity in DLPFC and ACC. In order to create the FFA and PPA masks, whole-brain group level *z*–statistic images were used to compute stimulus-selective face- and house-area activation during the localizer scans run. Z-statistic images showing greater activity during house blocks than during face blocks (PPA), and regions showing the opposite pattern (FFA), were thresholded at $z > 3.1$ and a (corrected) cluster significance threshold of $p = 0.05$ (corrected for multiple comparisons using Gaussian Random Field Theory, GRFT, Worsley, 2001) to create left and right PPA and right FFA masks.

We intended to mask probe-related contrasts with DLPFC and ACC masks; therefore DLPFC and ACC masks were created based on a mean contrast effect across subjects (whole-brain group-level effect); the average probe-related activity from the experimental conditions was contrasted with fixation-related activity, thresholded at $z > 3.1$ and a (corrected) cluster significance threshold of $p = 0.05$ (corrected for multiple comparisons using GRFT, Worsley, 2001). Next, the resulting *z*-statistic image was masked with Brodmann area (BA) 46 and BA 9 to determine a DLPFC ROI. BA 24 and BA 32 were used to identify an ACC ROI.

The first set of ROI analyses was run to examine binding costs related to neural activity in FFA and PPA. FEAT was used to pre-mask four contrasts of interest with right FFA and left and right PPA ROIs, thresholded using clusters determined by $z > 1.9$ and a (corrected) cluster significance threshold of $p = 0.05$ (Worsley, 2001). Activation and local maxima are only reported for $p < 0.05$, cluster-corrected for multiple comparisons.

In a first contrast of *cue*-related activation, FFA and PPA activation was examined in *A*-cues previously accompanied with face probes versus *A*-cues previously accompanied with house probes. This contrast allowed us to test whether cues significantly reactivate the probe-related information they were associated with during training. In a second contrast of *probe*-related activation, FFA and PPA activation is examined in complete repetitions compared to partial repetitions. This contrast enables us to test whether a face or house probe triggers more pronounced FFA or PPA activation, respectively, if that probe matches the probe as reactivated by the specific cue (complete repetition) compared to when the cue reactivated a probe from the opposite category (partial repetition). FFA and PPA activation in this second contrast may be correlated (tested separately for face and house stimuli) with performance costs associated with partial

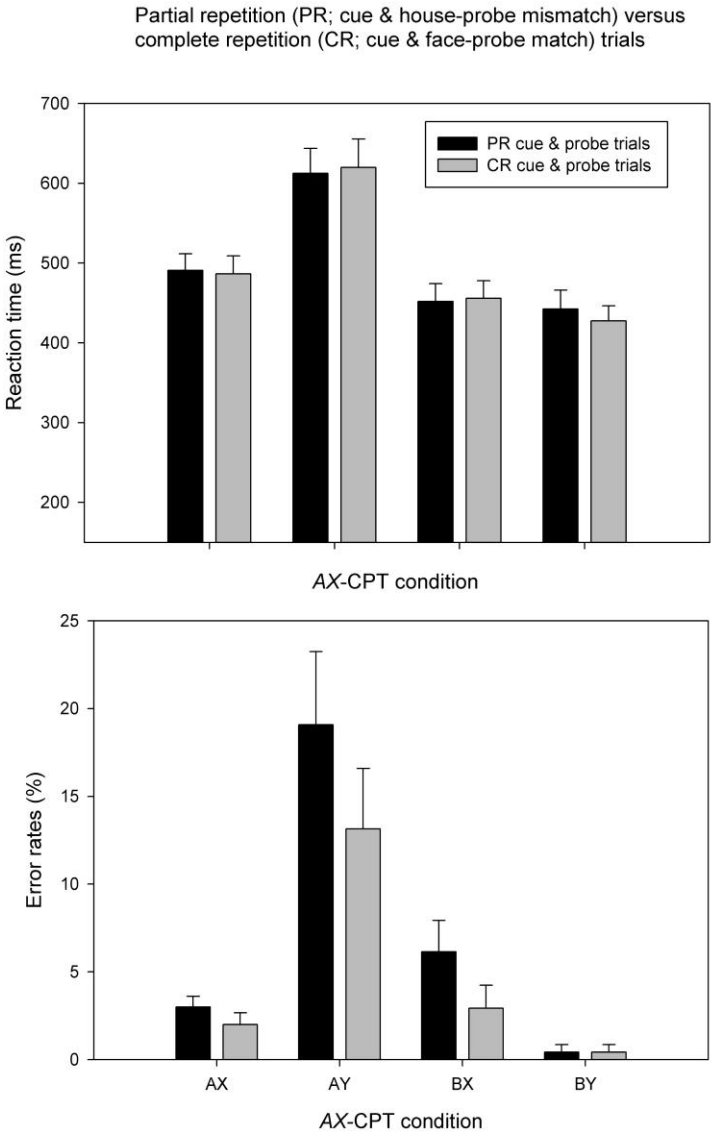
compared to complete repetitions. A third prediction derived from the binding account is that individuals who show a greater increase in cue-induced reactivation of FFA/PPA with complete than with partial repetitions show a greater increase in performance costs with partial than with complete repetitions. Thus, in a third contrast of *probe*-related activation, the increase in FFA/PPA activation with complete repetitions compared to partial repetitions is expected to be negatively correlated with performance costs associated with partial as compared to complete repetitions. Therefore, we included the error rates (complete versus partial repetitions) as an additional regressor with the probe-related contrasts in FEAT to correlate the probe-related hemodynamic response with performance costs.

The second set of ROI analyses was used to identify binding-induced vis-à-vis top-down control-induced activation in DLPFC and ACC. FEAT was used to pre-mask the contrasts of interest with ACC and DLPFC ROIs, thresholded using clusters determined by $z > 1.9$ and a (corrected) cluster significance threshold of $p = 0.05$ (Worsley, 2001). Furthermore, we included the error rates as an additional regressor with each contrast in FEAT to correlate the probe-related hemodynamic response with performance costs. Activation and local maxima are only reported for $p < 0.05$, cluster-corrected for multiple comparisons.

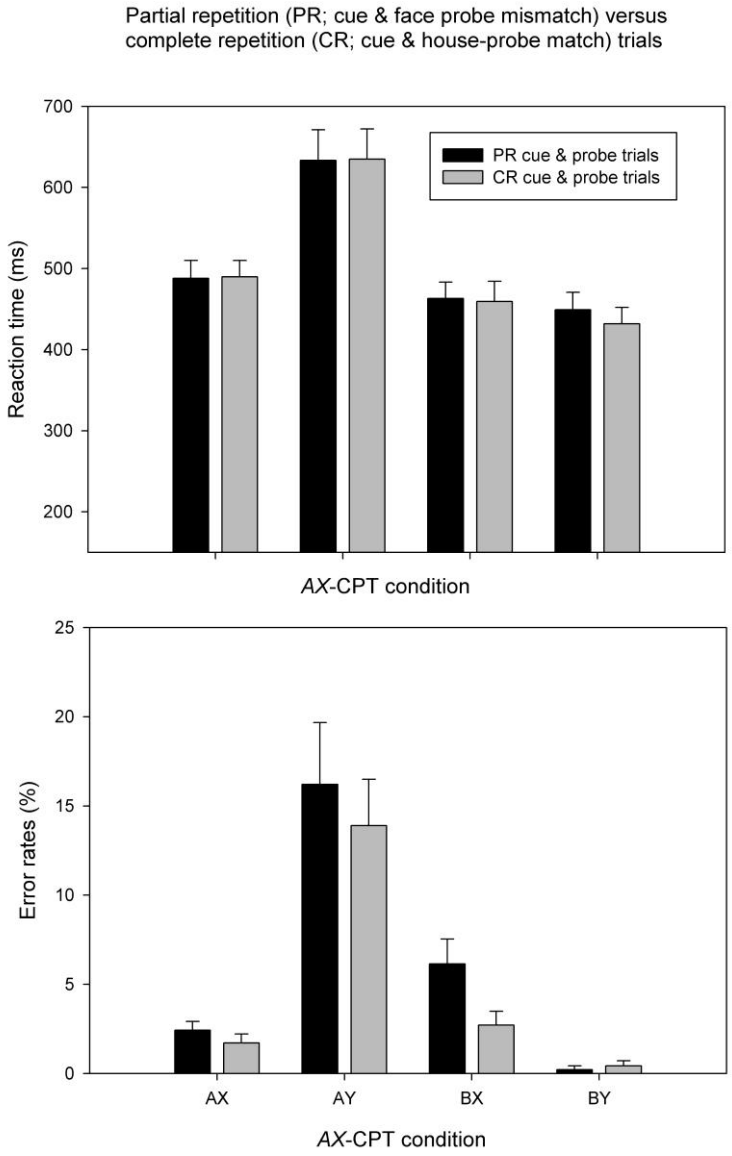
Evidence uniquely supporting the top-down control account is found if ACC and DLPFC are activated on complete repetition *AY* and *BX* pairs, whereas evidence uniquely supporting the binding account is found if ACC and DLPFC are activated *more* on partial- repetition than on complete-repetition *AY* and *BX* pairs. First, to examine activation in DLPFC and ACC that results from increased top-down control, the probe-induced activation on *AY* and *EX* pairs was contrasted with *AX* pairs and with *BY* pairs in complete-repetition sequences. Second, to examine activation in DLPFC and ACC that results from episodic binding, the probe-induced activation on *AY* pairs was contrasted with *AX* pairs and *BX* pairs with *BY* pairs in partial-repetition sequences (on a behavioral level, these contrasts showed binding and top down effects); we tested whether activation in the partial-repetition contrasts was greater than the corresponding activation in the complete-repetition contrasts. Third, to test a specific prediction derived from the binding account, probe-related brain activation on *AX* sequences was contrasted directly for partial repetitions against complete repetitions.

Figure 2

a. Mean reaction time (ms) and mean error rates (%) of complete repetitions and partial repetitions by AX-CPT condition for cues previously bound with faces. Error bars represent standard errors.



b. Mean reaction time (ms) and mean error rates (%) of complete repetitions and partial repetitions by AX-CPT condition for cues bound with houses. Error bars represent standard errors.



Results

Behavioral data

Mean RT and error rates are displayed in Figure 2. The first set of analyses tests binding (complete versus partial repetitions) and top-down control effects (*AX*-CPT Conditions) in *AY* and *BX* compared to *AX* trials. The second set of analyses investigates binding and top-down control processes in *AY* and *BX* compared to *BY* trials.

AX-AY-BX analyses

Reaction time

The ANOVAs showed that responses on *AX* trials were significantly faster than on *AY* trials, $F_{AX\text{-}AY}$ (1, 13) = 67.76, $p < 0.001$ and nearly significantly faster than *BX* trials, $F_{AX\text{-}BX}$ (1, 13) = 4.57, $p = 0.05$. Additionally, responses on face trials tended to be faster than house trials, F (1, 13) = 4.24, $p = 0.06$. However, we found no significant effects involving Feature Repetition, Fs < 1.

Separate ANOVAs for face- and house-trained trials revealed that responses on *AX* trials were significantly faster than on *AY* trials, $F_{AX\text{-}AYface}$ (1, 13) = 74.43, $p < 0.001$, M_{AXface} = 489 ms, M_{AYface} = 616 ms; $F_{AX\text{-}AYhouse}$ (1, 13) = 44.51 , $p < 0.001$, $M_{AXhouse}$ = 489 ms, $M_{AYhouse}$ = 634 ms, but slower than on BX_{face}-trained trials, $F_{AX\text{-}BXface}$ (1, 13) = 6, $p < 0.05$, M_{BXface} = 453 ms, and not significantly different from BX_{house} trials, $F_{AX\text{-}BXhouse}$ (1, 13) = 2.78, $M_{BXhouse}$ = 461 ms. No significant effects involving Feature Repetition were found in separate analyses of face and house trials, Fs < 1.

Error rates

The analogous ANOVAs on error rates revealed that performance on *AX* trials was more accurate than on *AY* trials, $F_{AX\text{-}AY}$ (1, 13) = 22.54, $p < 0.005$, and than *BX* trials, $F_{AX\text{-}BX}$ (1, 13) = 9.07, $p < 0.01$. Furthermore, the difference between *AX* and *BX* was nearly significantly larger in partial repetitions than in complete repetitions $F_{AX\text{-}BX}$ (1, 13) = 4.66, $p = 0.05$, but there was not a similar interaction of Feature Repetition x CPT Condition on *AX* compared to *AY* trials, F (1, 13) = 0.26. No other effects involving Feature Repetition or Cue Association were found, all Fs < 2.96, $ps > 0.1$

The separate ANOVAs for face-trained and house-trained trials both indicated that performance on *AY* trials (M_{AYface} = 16 %, $M_{AYhouse}$ = 14 %) was significantly more erroneous than performance on *AX* trials, M_{AXface} = 2 %, $F_{AX\text{-}AYface}$ (1, 13) = 14.66, p < 0.005; $M_{AXhouse}$ = 2 %, $F_{AX\text{-}AYhouse}$ (1, 13) = 29.06, p < 0.001. Furthermore, error rates on BX_{house} trials ($M_{BXhouse}$ = 4 %) were significantly larger than on AX_{house} trials, F (1, 13) = 10.59, p < 0.01. Performance on BX_{face} trials (M_{BXface} = 5 %) was not significantly less accurate than on AX_{face} trials, $F_{AX\text{-}BXface}$ (1, 13) = 2.84. Moreover, there was a significant main effect of Feature Repetition, exclusively in face-trained trials, F_{face} (1, 13) = 5.4, p < 0.05, and not in house-trained trials, F_{house} (1, 13) = 0.29. Performance on partial repetitions (M_{PR} = 9 %) was more impaired than on complete repetitions (M_{CR} = 6 %) in face-trained trials. No interaction effect of Feature Repetition and CPT Condition was found, $F_{face*CPT}$ (2, 26) = 1.5, $F_{house*CPT}$ (2, 26) = 0.99.

In sum, *AX*-CPT performance was as predicted: faster and smaller error rates on *AX* than on *AY* and *BX* trials. Feature Repetition affected performance in face trials only, where performance on complete repetitions was more accurate than with partial repetitions.

BY-AY-BX analyses

Reaction time

The ANOVAs demonstrated faster responses on *BY* trials than on *AY* trials, $F_{BY\text{-}AY}$ (1, 13) = 83.18, p < 0.001 and faster responses than on *BX* trials, $F_{BY\text{-}BX}$ (1, 13) = 5.42, p < 0.05. However, effects involving Feature Repetition and Cue Association were not significant, all Fs < 3.0, ps > 0.1.

The separate ANOVAs for face-trained and house-trained trials contrasting *AY* and *BX* against *BY* generated similar results as the ANOVAs including *AX* trials: performance on *BY* trials (M_{BYface} = 442 ms, $M_{BYhouse}$ = 449 ms) was significantly faster than on *AY* trials, $F_{BY\text{-}AYface}$ (1, 13) = 107.76, p < 0.001; $F_{BY\text{-}AYhouse}$ (1, 13) = 57.57, p < 0.001 and nearly significantly faster than on *BX* trials, $F_{BY\text{-}BXface}$ (1, 13) = 4.26, p = 0.06 , $F_{BY\text{-}BXhouse}$ (1, 13) = 4.30, p = 0.06 No significant effects involving Feature Repetition were found, Fs < 1.23, ps > 0.3.

Error rates

ANOVAs showed that performance on *BY* trials was more accurate than on *AY* trials, $F_{BY\text{-}AY}$ (1, 13) = 30.77, $p < 0.005$, and than on *BX* trials, $F_{BY\text{-}BX}$ (1, 13) = 20.32, $p < 0.005$. Furthermore, the difference between *BY* and *BX* was significantly larger for partial than for ccmplete repetitions $F_{BY\text{-}BX}$ (1, 13) = 6.82, $p < 0.05$. However, this interaction between Feature Repetition and CPT Condition was not found for *BY* compared to *AY* trials, *F* (1, 13) = 0.65. No significant effects involving Feature Repetition or Cue Association were found, *F*s < 2.25, *ps* > 0.1.

Likewise, the separate ANOVAs of face-trained and house-trained trials indicated that the number of errors on *AY* trials was significantly larger than on *BY* trials, M_{BYface} = 0.4 % errors, $F_{BY\text{-}AYface}$ (1, 13) = 20.10, $p < 0.01$; $M_{BYhouse}$ = 0.2 % errors, $F_{BY\text{-}AYhouse}$ (1, 13) = 38.90, $p < 0.001$. Error rates on *BX* trials were significantly increased compare to *BY* trials, $F_{BY\text{-}BXface}$ (1, 13) = 11.52, $p < 0.01$; $F_{BY\text{-}BXhouse}$ (1, 13) = 19.73, $p < 0.01$. Additionally, the error rate analysis yielded a significant main effect of Feature Repetition in face-trained trials, F_{face} (1, 13) = 5.1, $p < 0.05$, but not in house-trained trials, F_{house} (1, 13) = 0.14. Performance costs were larger in partial repetitions (M_{PR} = 9 %) than in complete-repetition trials (M_{CR} = 5%) for face-trained trials. No interaction effect of Feature Repetition and CPT Condition was found, $F_{face*CPT}$ (2, 26) = 1.94; $F_{house*CPT}$ (2, 26) = 1.13.

In sum, *AX*-CPT performance was as predicted: smaller error rates and faster performance on *BY* than on *AY* and *BX* trials. Feature Repetition enhanced the difference between *BY* and *BX* trials: error rates increased on partial-repetition *BX* trials. Additionally, we found a main effect of Feature Repetition in face trials: performance was better with complete than with partial repetitions.

Table 1

Results of contrasts testing the episodic binding theory masked by group base PPA and FFA ROIs. Error percentage was modeled as a covariate in the analysis and correlated negatively with contrasts of interest.

Contrast	**ROI**		
	PPA Main effect		**FFA Main effect**
	Left	**Right**	**Right**
	Cue related activity		
A-cue face > A-cue house	n.s.	n.s.	n.s.
	Probe related activity		
Complete repetitions (CR)> Partial repetitions (PR)	3.27*	3.59*	3.15*
	PPA covariate errors	**PPA covariate errors**	**FFA covariate errors**
	Left	**Right**	**Right**
Cue face -probe face (CR) > Cue house - probe face (PR)	n.s.	n.s.	2.34*
Cue house- probe house (CR)> Cue face - probe house (PR)	n.s.	2.22*	n.s.

**p < 0.05. N.s. = z values < 2.0 and p > 0.05. Z values are reported for 13 subjects (ROIs LPPA, RPPA and RFFA). Cue face = word-cues during training session presented with faces, cue house = word-cues during training session presented with houses*

Table 2

Results of binding and top-down related contrasts masked by group based ACC and DLPFC ROIs. Error percentage was modeled as a covariate in the analysis and correlated positively with contrasts of interest.

Theory	**Contrast**	**ACC**		**DLPFC**	
		Main effect	**Covariate errors**	**Main effect**	**Covariate errors**
Binding	*AX partial repetitions (PR) > AX complete repetitions (CR)*	2.14	n.s	2.23	2.68
Top-down	*AY > AX*	2.70*	2.52	2.24	2.25
control	*AY > BY*	2.24	n.s.	n.s.	2.45
	BX > AX	n.s.	n.s.	2.73*	n.s.
	BX > BY	n.s.	2.17	n.s.	2.37
Interaction	*AY (PR) > AY (CR) versus AX (PR) > AX (CR)*	n.s.	n.s.	n.s.	n.s.
	BX (PR) > BX (CR) versus BY (PR) > BY (CR)	n.s.	n.s.	2.60	n.s.

**p < 0.05. N.s. = z values < 2.0 and p > 0.05. Z values are reported for 14 subjects (ROI ACC and DLPFC).*

fMRI Results

We first verified whether each individual subject showed stimulus-selective activity in FFA and PPA based on the localizer scans run. One subject showed no activation in PPA and FFA and was therefore excluded from the group-based ROI analyses with PPA and FFA. Moreover, when contrasting face blocks with house blocks (group level) the right FFA was significantly activated but the left FFA was not, which is a common finding (Kanwisher, McDermott, & Chun, 1997; Puce et al., 1996). Figure 3 displays probe-related reactivation in FFA (a) and PPA (b) contrasting complete repetitions and partial repetitions. Figure 4 presents probe-related reactivation in FFA and PPA correlated with behavior.

Figure 3

Probe-related reactivation in FFA (a) and PPA (b) contrasting complete repetitions and partial repetitions (face and house bindings grouped, all AX-CPT trials included). FFA activation in blue, MNI_{RFFA} peak activation at x = 36 y = -52 z = -24, Z_{max} = 3.15. PPA activation in red; MNI_{LPPA} peak activation at x = -26 y = -50 z = -20, Z_{max} =3.27; MNI_{RPPA} peak activation at x = 30 y = -56 z = -14, Z_{max} =3.59.

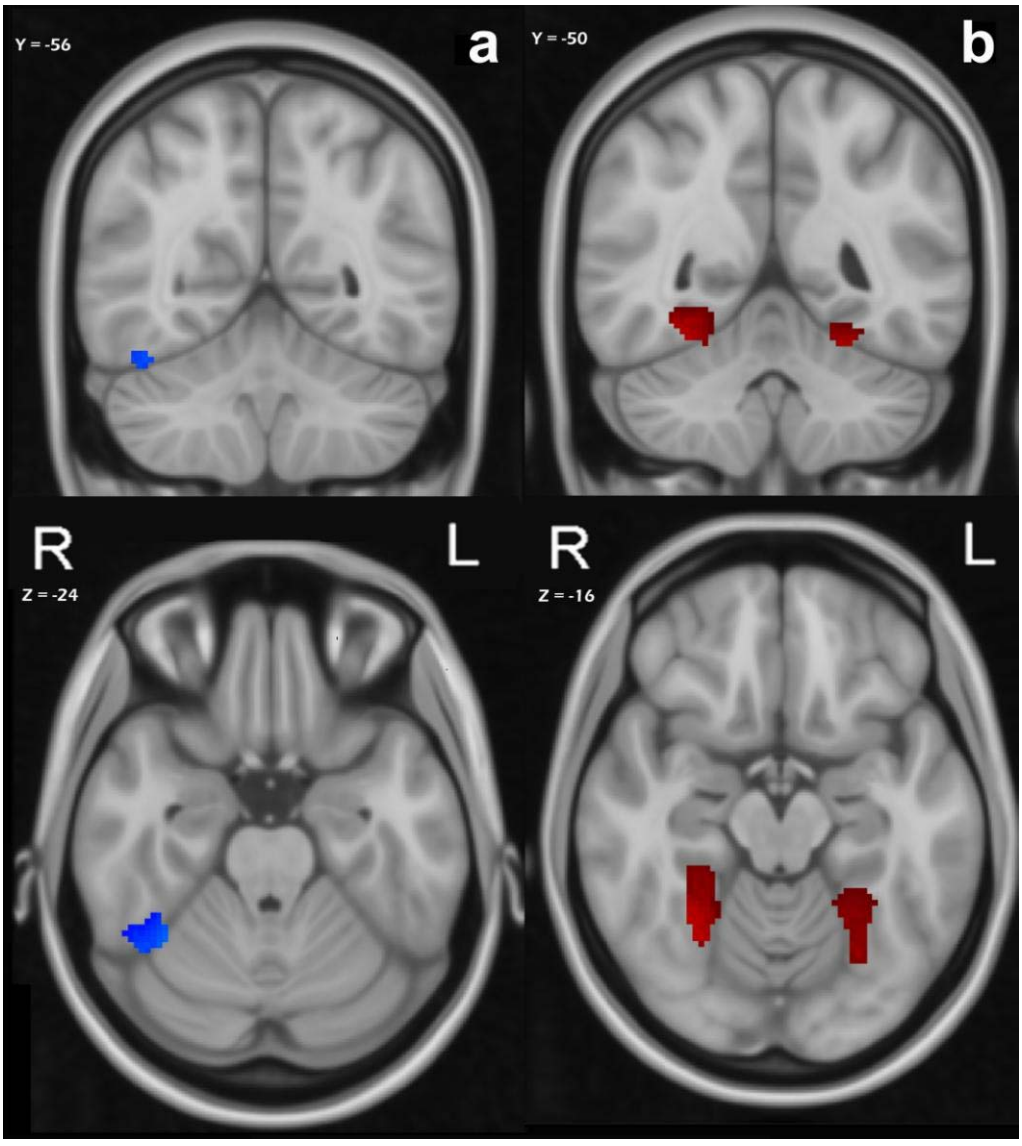
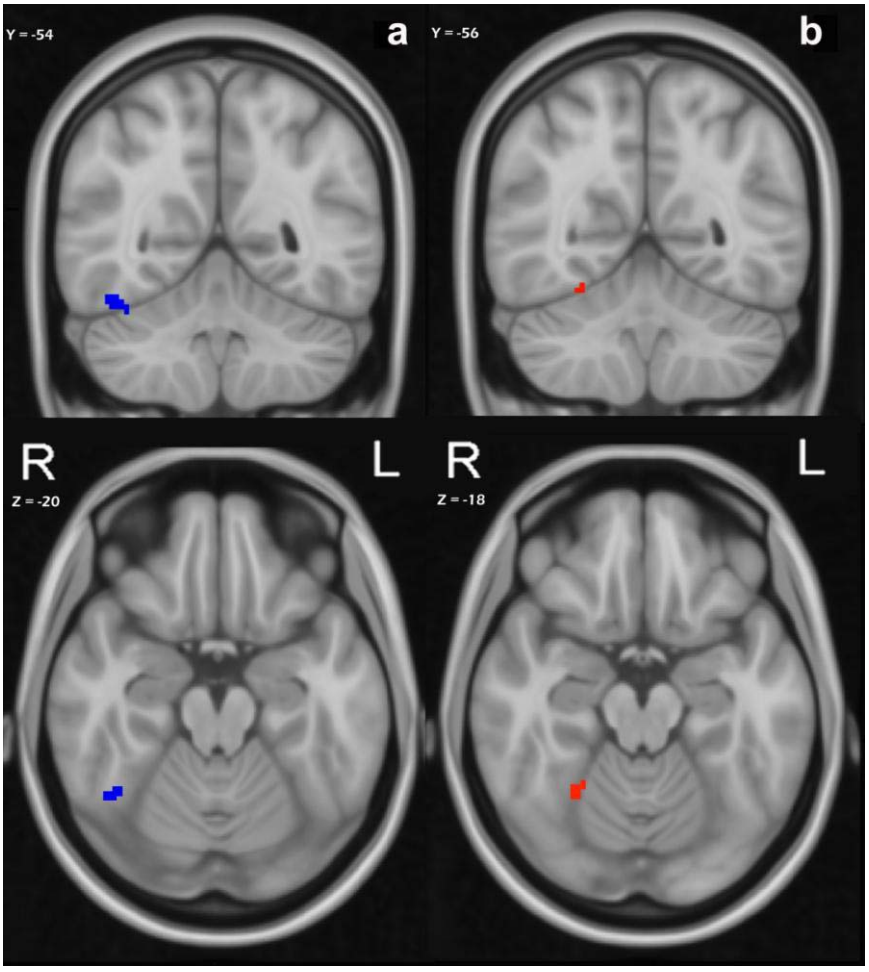


Figure 4

a. Reactivation in RFFA measured after face probes, contrasting complete repetitions compared to partial repetitions in negative correlation with error percentage (on complete compared to partial face repetitions) (MNI_{FFA} peak activation at x = 42 y = -54 z = -20, Z_{max} = 2.34).

b. Reactivation in RPPA measured after house probes, contrasting complete repetitions compared to partial repetitions in negative correlation with error percentage (on complete compared to partial house repetitions (MNI_{PPA} peak activation at x = 4 y = -50 z = -18, Z_{max} = 2.22).



Binding costs and neural activity in occipitotemporal areas

One direct prediction derived from the binding account is that if a specific cue was previously associated with a face probe, a new presentation of the same cue should reactivate the representation of the same face probe and hence should activate the FFA even in the absence of an actual face probe (and likewise for reactivation of house probes and the PPA). Thus, in a first contrast of *cue*-related activation, the FFA is expected to be activated with *A*-cues previously accompanied by face-probes versus *A*-cues previously accompanied by house-probes, whereas the PPA is expected to light up in the reverse contrast. Neither of the ROIs showed significant activation in either the face>house or house>face contrast (see Table 1 for *z*-values).

Another prediction derived from the binding account is that a face or house probe triggers more pronounced FFA or PPA activation, respectively, if that probe matches the probe that is reactivated by the specific cue (complete repetition) than if the cue reactivated a probe from the opposite category (partial repetition). Thus, in a second contrast of *probe*-related activation, FFA and PPA are expected to be activated more with complete repetitions than with partial repetitions. This prediction was confirmed: PPA and rFFA were significantly more active after complete-repetition trials relative to partial-repetition trials, $p < 0.05$, cluster-corrected (see Figure 3).

A third prediction derived from the binding account, and perhaps the most important prediction for the present purposes, is that individuals who show stronger cue-induced reactivation of FFA/PPA with probe presentation, show greater behavioral performance costs induced by cue-probe associations. More specifically, individuals who show a greater increase in cue-induced reactivation of FFA/PPA for complete than for partial repetitions are expected to show better performance as well; i.e. reduced error rates for complete as compared to partial repetitions. Note that significant differences in complete compared to partial repetitions were present in error rates but not in RT; hence, correlations will be based exclusively on error rates (on complete versus partial repetitions). Thus, in a third contrast of *face-probe*-related activation, the increase in FFA activation in complete repetitions compared to partial repetitions is expected to be negatively correlated with an increase in error rate associated with face probes in complete compared to partial repetitions. This prediction was confirmed for the right-hemisphere FFA, $p < 0.05$, cluster-corrected (see Figure 4a): the stronger the activation in rFFA with complete over partial repetitions, the smaller the amount of errors on this

contrast Likewise, in a contrast of *house-probe*-related activation, the increase in PPA activation in complete versus partial repetitions is expected to be negatively correlated with an increase in error rate for house probes in complete versus partial repetitions. This prediction was confirmed for both left and right PPA, $p < 0.05$, cluster-corrected (see Figure 4b): the stronger the activation in PPA for complete over partial repetitions, the smaller the amount of errors for complete versus partial repetitions.

Goal-driven vis-à-vis binding accounts of neural activity in prefrontal areas

The second set of analyses tested the activity following the probe, pre-masked by DLPFC and ACC and correlated with performance costs (percentage correct on the contrast of interest). Activity and local maxima are only reported for $p < 0.05$, cluster-corrected for multiple comparisons (Table 2). Figure 5 displays activity in ACC and DLPFC of *AY* and *BX* compared to *AX* trials. Figure 6 shows the correlation of DLPFC activation during *AX* partial-repetition trials compared to *AX* complete-repetition trials with error rates on the same contrast.

Figure 5

Activation in prefrontal regions associated with top-down control, (a) contrasting AY with AX trials (ACC in red; MNI_{ACC} peak activation at x = 10 y = 18 z = 34, Z_{max} = 2.7) and (b) BX with AX trials (DLPFC in blue; MNI_{DLPFC} peak activation at x = -42 y = 50 z = 4, Z_{max} = 2.73).

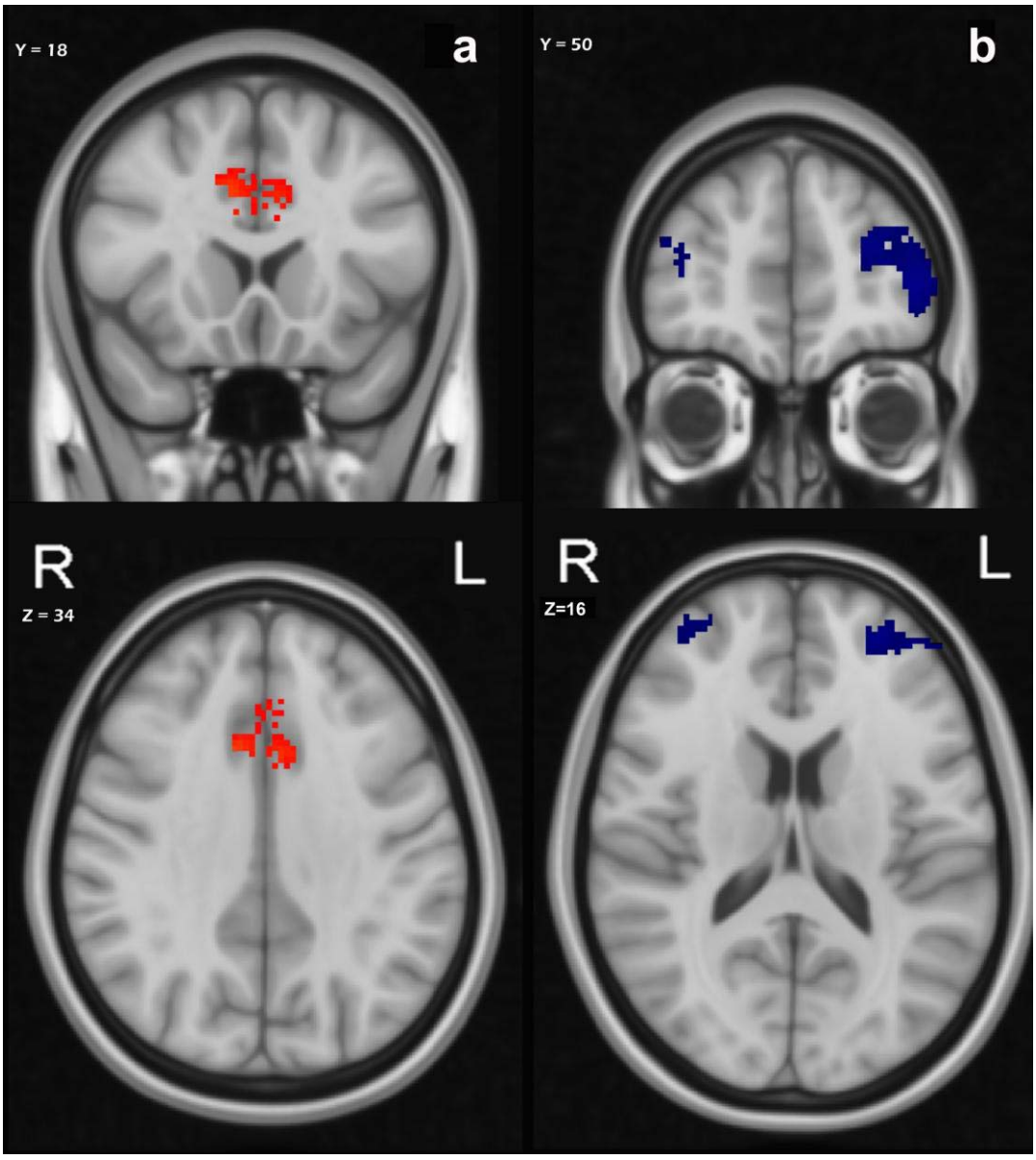
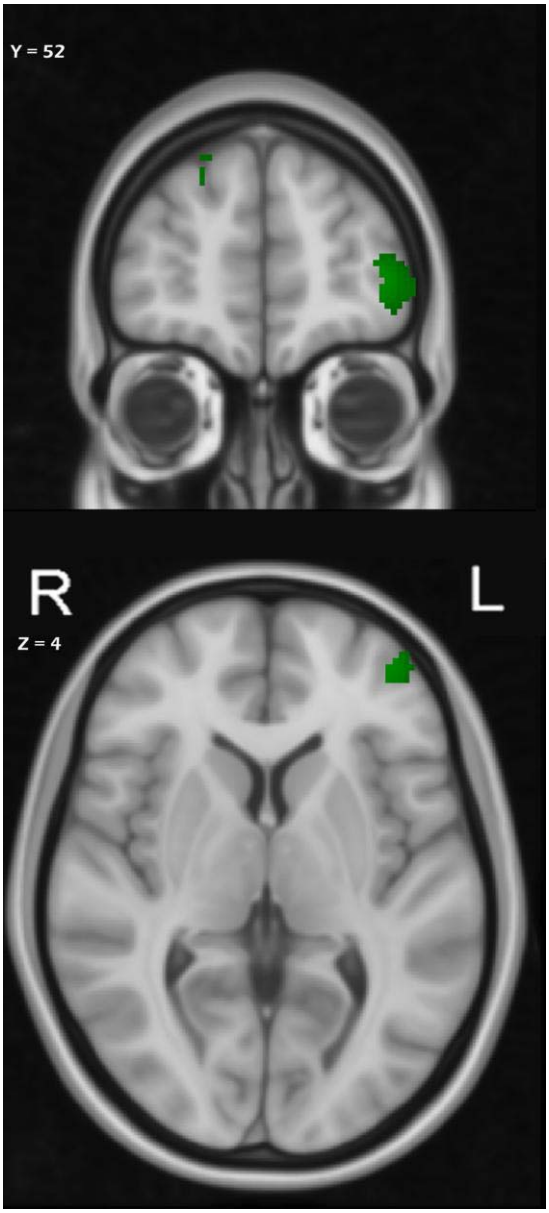


Figure 6

Activation in prefrontal regions implicated in unbinding previous associations, contrasting AX partial-repetition with AX complete-repetition trials in correlation with error percentage on AX partial-repetition with AX complete-repetition trials (MNI_{DLPFC}peak activation at x = -22 y = 58 z = 22, Z_{max}=2.68).



Top-down control accounts such as the context-processing model predict ACC and DLPFC to be activated on *AY* and *BX* pairs irrespective of whether the sequence constitutes a complete or partial repetition. The binding account, by contrast, would predict a need for DLPFC and ACC activation only on partial-repetition sequences. Thus, evidence for the top-down control account is found if ACC and DLPFC are activated on complete-repetition *AY* and *BX* pairs, whereas evidence for the binding account is found if ACC and DLPFC are activated *more* on partial-repetition than on complete-repetition *AY* and *BX* pairs. Note that these accounts are orthogonal, such that in principle predictions may be confirmed independently.

First, to examine activation in DLPFC and ACC that results from increased top-down control, the probe-induced activation on *AY* pairs was contrasted with *AX* pairs and, separately, with *BY* pairs in complete-repetition sequences. Likewise, probe-induced brain activation on *BX* pairs was contrasted with *BY* pairs and, separately, with *AX* pairs in complete-repetition sequences. The context-processing model predictions were confirmed, at least in part: the ACC was more active on *AY* than on *AX* trials and DLPFC was more active on *BX* than on *AX* trials, $p < 0.05$, cluster-corrected for both tests. No effects were found when comparing *AY* or *BX* with *BY* trials or correlating any of the contrasts with performance costs, $p > 0.05$ cluster-corrected for all tests.

Second, to examine activation in DLPFC and ACC that results from episodic binding, the probe-induced activation on *AY* pairs was contrasted with *AX* pairs in partial-repetition versus complete-repetition sequences. Likewise, activation on *BX* pairs was contrasted with *BY* pairs in partial-repetition versus complete-repetition sequences. Crucially, activation in the partial-repetition contrasts (for *AY* and for *BX*) was *not* greater than the corresponding activation in the complete-repetition contrasts nor did this contrast correlate with performance costs, thus disconfirming the binding hypothesis, $p > 0.05$, cluster-corrected for both tests.

Third, a specific prediction was unique to the binding account: even on *AX* sequences a partial repetition creates conflict and/or requires unbinding a recent episodic memory trace, and hence may induce greater ACC and DLPFC activation than complete-repetition *AX* sequences. Thus, probe-related brain activation on *AX* sequences was contrasted directly for partial repetitions against complete repetitions. The hypothesis was not confirmed: neither the ACC nor the DLPFC revealed significant activation, $p > 0.05$, cluster-corrected. However, an additional analysis suggested that those

individuals who showed increased performance costs for partial-repetition than for complete-repetition *AX* sequences showed marginally greater activation in DLPFC for partial-repetition than for complete-repetition *AX* sequences, $z = 2.68$, $p = 0.07$. Note that the actual error trials were not included in the fMRI analysis.

Discussion

The present fMRI study investigated the neural activity related to both goal-driven and stimulus-driven bias in an adapted *AX*-CPT. Both accounts have been shown recently to explain behavioral performance costs in several experimental paradigms (Egner, 2007). Performance costs in the *AX*-CPT, however, are typically attributed to a goal-driven bias. Our behavioral and fMRI data support this interpretation: the majority of the effects can indeed be explained in terms of top-down control, although cue-probe bindings also explain unique variance in performance during decision making. The fMRI data shed light on the neural mechanisms underlying these processes.

Performance costs were greatest on partial-repetition trials; trials in which the cue or the probe reactivates previously associated features that mismatch the current event (regardless of whether these trials required top-down control). In complete-repetition trials, when cue and probe match according to their previous association, however, this gives rise to reduced performance costs, which coincides with a reactivation of the relevant binding information in occipitotemporal areas. That is, if a specific cue stimulus was followed by a face on previous occasions, subsequent presentation of face probes with these cue stimuli increasingly reactivated the 'face area' in the brain (FFA) with an reduction in performance costs, compared to face probes presented subsequent to cues previously associated with a house. We found similar results for cue stimuli followed by houses, which reactivated the 'house area' (PPA) upon probe presentation, also with an improvement in performance. This pattern of activation and performance costs was predicted exclusively by the binding account (Hommel, 2004). Based on the binding account we also predicted that cue stimuli would directly activate the associated face or house area; this was not confirmed however.

Together with a recent study that investigated the neural mechanism underlying feature integration for visual objects (Keizer et al., 2008) and behavioral studies showing performance costs as a result of stimulus-response, stimulus-task or stimulus-stop signal bindings (Huang et al., 2004; Mayr et al., 2003; Nieuwenhuis et al., 2006; Tipper, 2001; Verbruggen et al., 2008; Waszak et al., 2004), the current study provides some additional neurophysiological support for the episodic binding account. The effects of feature integration have been mainly investigated in stimulus-task or stimulus-response bindings. The present study complements these findings in that it provides neurophysiological evidence for binding costs as a result of episodic stimulus-stimulus (cue-probe) bindings. Additionally, our findings shed some light on the neural mechanisms underlying a stimulus-driven bias in the AX-CPT. Cue-probe associations in the AX-CPT were not directly primed by the cue stimulus, that is the cue stimulus did not reactivate the previously probe information. Instead, the cue may have prepared the system for processing the associated probe, for example by lowering the threshold for the associated probe information (cf., Huber, 2008), which may have increased probe-based activation in occipitotemporal areas and improved performance on trials that match this primed information (complete repetitions).

Consistent with the predictions derived from theories on top-down control in the *AX*-CPT (Braver et al., 2001) and with other fMRI studies that reported ACC and DLPFC activity in trials with competing stimulus or response information (Durston et al., 2003; Kerns et al., 2004; MacDonald et al., 2000), we found enhanced ACC and DLPFC activity and impaired performance on trials demanding increased top-down control (*AY* and *BX*) compared to target trials (ACC with *AY*>*AX* and DLPFC with *BX*>*AX*). This pattern was not predicted by the binding account.

We did not find increased activation in ACC or DLPFC as induced by a partial repetition of cue and probe in *AY* or *BX* trials compared to a partial repetition in *AX* trials. This is contrary to what would be predicted from the binding account and findings from Notebaert et al., (2006). In contrast to the study of Notebaert et al., (2006), we investigated stimulus-stimulus bindings instead of stimulus-response bindings. Stimulus-stimulus bindings may have been less influential in generating performance costs beyond the strong *A*-cue and *X*-probe induced response bias.

A further and more specific prediction derived from the binding account was that, even in *AX* pairs, complete-repetition probes (matching the probe that was reactivated

by the specific cue) should trigger greater ACC and/or DLPFC activation than partial-repetition probes (not matching the probe as reactivated by the cue). This hypothesis also was not confirmed. Marginally significant activation in DLPFC was observed only if co-variation with individual differences in performance costs was taken into account. DLPFC tended to be increasingly recruited with an increase in the number of errors on *AX* trials when cue and probe information present a mismatch according to their previous association. This may seem somewhat contradictory, especially since DLPFC activity is often correlated with an improvement in performance (MacDonald et al., 2000), although it may be noted that some authors have reported similar patterns (Boettinger & D'Esposito, 2005; Sharp, Scott, & Wise, 2004). Boettinger and D'Esposito observed that DLPFC activity was inversely correlated with improved performance on an S-R learning task and concluded that DLPFC was involved in recalling, and organizing previously encountered S-R associations. In the current experiment, the enhanced DLPFC activation and reduced performance cannot be explained by increased task difficulty or a response bias on *AX* partial compared to *AX* complete repetitions because the associated response and task remain equal in these trial types; it is exclusively the stimulus information that changes. Sharp et al. (2004) suggested that DLPFC monitors whether current stimulus information is sufficient to allow a response; they found DLPFC activation, in correlation with a reduction in accuracy, with decisions that had to be based on degraded stimulus information. In line with these studies, we tentatively suggest that in our study DLPFC activation was associated with the need to re-organize and monitor information in WM; that is, with decisions based on stimulus information incongruent with previous events (partial cue-probe repetitions). This might increasingly tax WM and distract attentional resources from accurate task performance.

To sum up, our findings suggest that top-down control theories explain most of the performance costs and neural activity in an AX-CPT, but stimulus-driven effects play a role as well. Recently, Braver, Gray, and Burgess (2007) posed a Dual Mechanism of Control (DMC) theory, differentiating the importance of both proactive and reactive control in flexible adaptive behavior. DMC theory states that in addition to proactive preparation for adequate performance, reactive control may be required to resolve the conflict between overlearned action tendencies or stimulus-associations, and actions or associated stimulus features reactivated by the latest information. Similarly, reactive

control may be required with reactivation of competing information as a result of a retrieved episodic binding (created on a previous trial) which induces conflict cn the current trial. Future studies should aim at increasing our understanding of the relative importance of proactive and reactive control in adaptive behavior, for example whether and in what way trial-to-trial changes in top-down control affects attention and episodic binding.

Egner and Hirsch (2005) performed an interesting fMRI study with a word-face version of the Stroop task in which participants were presented with pictures of political figures or actors and a congruent or incongruent name (actor/pclitical figure) presented on top of it. Participants had to discriminate between political figures or actors based on the picture or the word. This study indicated that increased control, resulting from a preceding incongruent trial, changed attention on the current trial by amplifying the neural activation to relevant target information in an occipitotemporal area (FFA). The same might be observed with the reactivation of binding-related information and trial-to-trial changes of control; increased control due to incongruent information on a previous event might increasingly reactivate binding-related information on the current trial.

Limitations

Although the current fMRI study suggests support for a stimulus-driven account as well as a goal-driven account, some limitations have to be taken into account when interpreting the results. First, the behavioral error rate data mainly showed an effect for face-trained trials (i.e. trials in which cues were trained with faces and succeeded by a face (complete-repetition) contrasted with the same cues followed by a house probe (partial repetition)). Faces are highly salient and biologically significant stimuli that are easily recognized and memorized (Yin, 1969; Bahrick, Bahrick, & Wittilinger, 1975) which may explain why their features are more likely to be integrated with other aspects of an event and why their storage in WM is possibly stronger (Jackson & Raymond, 2008) compared to objects like houses.

Furthermore, the fMRI data revealed reactivation of binding-related information as triggered by probe stimuli, but not with cue presentation. This may be the result of the

experimental procedure: during the experiment face and house probes appeared equally frequently subsequent to the cues; thus, participants may have anticipated equally for either a house or a face, based on frequency of presentation, which may have deferred binding-related reactivation. Furthermore, considerable evidence from task switching studies (Karayanidis, Coltheart, Michie, & Murphy, 2003; Rogers & Monsell, 1995) indicate that both advanced cue preparation and stimulus-based activation of task-appropriate information are relevant for flexible alternation between tasks; task preparation may even remain incomplete until presentation of the target stimulus, especially if the imperative stimulus contains crucial task-relevant information (Wylie, Javitt, & Foxe, 2003). Apparently, probe information is necessary to activate the final appropriate response in an *AX*-CPT task, at least in a large proportion of the trials (trials starting with an *A*-cue).

Conclusion

The present study supports the notion that behavior in a complex and dynamically changing world is based on top-down guidance. Adaptive behavior also involves processing cue-probe associations that may appear in a different context than usual, which becomes crucial to performance upon probe presentation but not yet with cue based preparation. More importantly, our fMRI data provided insight into the way these additional binding-related performance changes may be represented in the brain.