Cognitive and Neural Mechanisms of Decision Biases in Recognition Memory

In recognition memory tasks, stimuli can be classified as ‘old’ either on the basis of accurate memory or a bias to respond ‘old’, yet bias has received little attention in the cognitive neuroscience literature. Here we examined the pattern and timing of bias-related effects in event-related brain potentials (ERPs) to determine whether the bias is linked more to memory retrieval or to response verification processes. Participants were divided into a High Bias and a Low Bias group according to their bias to respond ‘old’. These groups did not differ in recognition accuracy or in the ERP pattern to items that actually were old versus new (Objective Old/New Effect). However, when the old/new distinction was based on each subject’s perspective, i.e. when items judged ‘old’ were compared with those judged ‘new’ (Subjective Old/New Effect), significant group differences were observed over prefrontal sites with a timing (300–500 ms poststimulus) more consistent with bias acting early on memory retrieval processes than on post-retrieval response verification processes. In the standard old/new effect (Hits vs Correct Rejections), these group differences were intermediate to those for the Objective and the Subjective comparisons, indicating that such comparisons are confounded by response bias. We propose that these biases are top-down controlled processes mediated by prefrontal cortex areas.

Introduction

Whenever a judgement cannot rely on perfect knowledge, it must be based, at least partially, on guessing. This is true not only in laboratory settings but also in many everyday situations. Consider, for instance, a not so uncommon situation in which you are not sure whether a given item was or was not a part of some past event, and yet must nonetheless decide and act accordingly. Moreover, even when that judgement could, in principle, rely on memory, you may have processing preferences based on various non-sensory, non-memory- and non-motor-related factors that affect which decision you are most likely to make. What are the cognitive and neural mechanisms that enable such recognition decisions?

This scenario is exemplified by old/new recognition memory tasks in which participants are asked to indicate whether or not they had experienced some test item in a previous study phase. If their memory fails, they are obliged to guess. Under these circumstances, some participants are disposed to give one response alternative (‘old’) while others are disposed towards the other (‘new’), even when their actual memory for the test item is the same. How are these different behavioral tendencies mediated in the brain, i.e. what processes support them, and when during recognition decisions are they engaged?

While considerable research within cognitive neuroscience has been devoted to unearthing the processes involved in accurate recognition memory, very little effort has been directed at answering questions of this kind. A significant number of cognitive models are based on the assumption that individuals render recognition memory decisions by comparing the strength of each test item in memory to some internal threshold or decision criterion (Green and Swets, 1966; Mandler, 1981; Hockley and Murdock, 1987; Snodgrass and Corwin, 1988). When this criterion is ‘liberal’ (as opposed to ‘strict’ or ‘conservative’), participants are more likely to respond ‘old’ than ‘new’, whether or not the item actually is old. Within signal-detection theory, this likelihood is referred to as an individual’s ‘response set’ or ‘response bias’.

Empirical evidence shows that at a behavioral level the response bias is largely independent of accurate memory (Snodgrass and Corwin, 1988). This independence must be accounted for within any neuropsychological and/or neuro-computational model of recognition memory. One possibility is that accuracy-related and bias-related processes are mediated by functionally and/or neuroanatomically distinct brain regions.

It is well known that intact medial temporal lobes are essential for episodic memory (Milner, 1966; Rolls, 1996), and that the prefrontal cortex is crucially involved in decision-making processes (Bechara et al., 1997, 2000; Kim and Shadlen, 1999; Tremblay and Schultz, 1999, Elliott et al., 2000). Nonetheless, there has not yet been any systematic investigation of how directly these two neural systems map onto accuracy- and bias-related functions within recognition memory. This is somewhat surprising given the widely held belief that any given recognition judgement can result from accurate memory, bias, or some combination thereof. A hit rate of 1 (i.e. 100% correct ‘old’ responses), for example, can result solely from bias, partially from bias, or not at all from bias (depending on the number of incorrect responses). Thus from behavioral data alone we can never be sure that a ‘correct’ response indicates accurate recognition and not a correct guess. These alternative possibilities matter in theory when it comes to understanding the psychological and neural mechanisms involved as well as in practice when the results of techniques involving signal averaging of correct versus incorrect response trials are interpreted. It is clearly important for cognitive (neuro)scientists to get a better grasp on the factors that influence response bias, and of how and when these contribute to any decision, be it in a recognition memory task or elsewhere.

Even within the domain of behavioral research, where signal-detection theory has seen wider application, researchers have nonetheless focused more on accuracy measures than on the response bias (Hirshman, 1995). The bias has often been treated as little more than a nuisance variable that has to be corrected for in determining accuracy rates. However, mounting evidence suggests that this variable carries important information. While it has long been known that subjects routinely adjust their bias to task characteristics such as stimulus probabilities and pay-off matrices (Buchner et al., 1995; Hirshman, 1995), recent studies indicate that there are persistent differences in bias among individuals as well. For example, bias measures have...
been found to be correlated with age and personality trait variables in healthy individuals as well as with clinical symptoms in psychiatric populations (Berch and Evans, 1973; Brébion et al., 1997a,b; Jacoby, 1999; McBrien and Dagenbach, 1998; Windmann and Krüger, 1998; Merckelbach et al., 2000). While the origin and functional significance of these inter-individual differences are poorly understood at present, they appear to be more systematic than originally thought and thus require systematic exploration and ultimately explanation.

Various studies also have shown that the response bias fluctuates systematically on a trial-to-trial basis even within a given individual. For example, items presented for a second or third time seem to induce an ‘accessibility’ or ‘retrieval’ bias resulting in more liberal decision criteria, even when they are not explicitly or consciously recollected as repeated (Ratcliff and McKoon, 1996; Hintzman and Curran, 1997; Elliott and Dolan, 1998; Jacoby, 1999; Jacoby et al., 1999). Measures of bias also have been shown to vary with the explicitness of the recognition poststimulus is typically referred to as the ‘early’ old/new effect. However, the relative timing of memory- versus detection do not specify at what stage of processing the decision biases in memory tasks are not invariably under the strategic control of participants and may fluctuate at times in lockstep with certain stimulus-specific variables. This naturally raises the question of the appropriateness of the term ‘response bias’, a label that seems to ascribe the source of the bias to motor output processes, i.e. to processes following rather than preceding or coincident with perceptual and recognition processes (Light and Kennison, 1996; Miller and Wolford, 1999; Schacter et al., 1996, 1998; Swick and Knight, 1999). Processing-tree models likewise seem to encompass the view that memory biases materialize only after individuals realize that accurate memory has failed, although it should be noted that these models illustrate transitional probabilities rather than the temporal dynamics of the process (Riefer et al., 1994; Buchner et al., 1995; Endfield and Buchner, 1998; Windmann and Krüger, 1998; Wallsten et al., 1999). Other recognition memory models based on signal detection do not specify at what stage of processing the decision criterion is invoked (being, in fact, mathematically indifferent to this issue). However, the relative timing of memory- versus bias-related processes is clearly very important if we are to understand the dynamics of recognition memory processes as implemented in the brain.

Event-related brain potential (ERP) measures should prove useful in examining these issues as they have in the area of human memory more generally (Rugg, 1995; Allan et al., 1998; Windmann and Kutas, 1998; Merckelbach et al., 2000). While the origin and functional significance of these inter-individual differences are poorly understood at present, they appear to be more systematic than originally thought and thus require systematic exploration and ultimately explanation.

To address this question, we chose a straightforward approach based on a between-subjects design for the present study. Specifically, we divided 30 subjects into two groups on the basis of the median value of their bias to respond ‘old’ to words with an emotionally neutral meaning. We then compared the ERP old/new effects of the High Bias group with those of the Low Bias group. [Note that we deliberately chose not to manipulate pay-off matrices (Miller et al., 2001) or stimulus probabilities in a within-subjects design as these manipulations would have changed the emotional value or salience of the stimuli, factors which are known to affect P3 and other ERP components.] The experimental procedures were the same as in two previous studies of ours investigating the effects of emotional word valence on decision biases (Windmann and Kutas, 2001; Windmann et al., 2002). However, only words with an emotionally neutral meaning were considered in the present analyses as we were interested in ‘spontaneous’ or ‘unprovoked’
variations of the response bias, and not in those related to emotionality or any other variable. Fortunately, old/new recognition accuracy was virtually identical for the High Bias and the Low Bias groups (see Fig. 1 and the Results section below). Hence, all variables other than response bias can be regarded as matched between the two groups, including the random intermingling of emotionally negative items with the neutral items during the recognition test phase as described in Windmann and Kutas (Windmann and Kutas, 2001).

We were especially interested in ERP differences between the High and the Low Bias groups for items they considered ‘old’ as opposed to ‘new’, because we expected this comparison to provide insight into the cognitive and neural processes underlying their different response criteria. In addition, we aimed to demonstrate that these group differences emerged solely as a function of decision bias, and not as a function of memory strength or the actual old/new status of the items. To these ends, we conducted the following three analyses:

- **Subjective Old/New ERP Effect.** ERPs to items subjects considered ‘old’ were compared with those they considered ‘new’, whether or not these classifications were correct. By definition, participants in the High Bias group were more willing than those in the Low Bias group to render an ‘old’ decision, while the reverse was true for ‘new’ decisions. If ERPs are sensitive to the processes underlying these different response preferences, then this comparison should yield a significant group difference.

- **Objective Old/New ERP Effect.** ERPs to items that had been presented during the study phase were compared with those that had not been presented. This difference should reflect the memory for the actual study status of the items, regardless of participants’ response criteria. As the two groups were equally capable of discriminating old from new items, this analysis should minimize group differences.

- **Correct classification Old/New ERP Effect.** ERPs to old and new items that were correctly classified, i.e. Hits and Correct Rejections, were compared. This is the canonical comparison in ERP studies of recognition memory. It typically yields the largest old/new effect since, unlike in the two other analyses, the subjective memory experience converges with the actual study status of the items. However, even though this analysis includes only correct response trials, it may nonetheless be sensitive to the effects of bias since a certain percentage of the items are likely to be classified correctly solely on the basis of guessing.

It is important to note that as a result of the virtually identical old/new discrimination performance in the two bias groups, the proportion of trials due to accurate recognition memory was the same for both groups in all three analyses (~20%, see Fig. 1, center), as was the proportion of actually old and actually new items. Since other variables (gender, age, handedness, etc.) were constant as well, any group effects emerging in any of the three analyses could be attributed to group differences in response bias.

The only thing that varied for the two bias groups across the three analyses is the proportion of trials in the ‘old’ ERP average as compared to those in the ‘new’ ERP average that reflected guessing as opposed to accurate memory. This proportion is maximally different for the two groups in the Subjective Old/New comparison. Here, participants in the High Bias group based their ‘old’ responses on guessing more often than on accurate recollection while this was much less true for ‘new’ responses (5 times more often vs 2 times more often, as it turned out). The opposite held for the Low Bias participants. Hence, although both groups responded ‘old’ in one case and ‘new’ in the other, they did so for different reasons because of their different response biases. We expected this difference to be reflected in the associated ERPs.

Similar effects should be evident in the Correct Classification Old/New Analysis, albeit to a lesser extent since approximately half of the trials in which participants guessed were eliminated (namely, the ones in which they guessed incorrectly, i.e. False Alarms and Misses). As a result, the relative contribution of accurate memory processes to the ERPs will be larger in this comparison than in the Subjective Old/New analysis, and the contribution of response bias will be smaller.

Finally, the group differences should be minimal in the Objective Old/New comparison when the old/new distinction was based solely on the actual study status of the items, regardless of subjects’ responses (i.e. collapsed across ‘old’ and ‘new’ responses). Although the High Bias subjects still classified items as ‘old’ more often than the Low Bias subjects, they did so in response to both old and new items (by definition of bias). Hence, while this may affect their overall ERP amplitudes, it should not affect the ERP difference between old and new items in this analysis.

A crucial aspect of these three comparisons will be the point in time by which the ERPs of the two groups differ. The time course of this effect is most informative with respect to the temporal dynamics of the cognitive and neural mechanisms involved, and is not available from either behavioral indices or blood-flow measures. If the relevant group ERP differences occur within the first 500 ms post stimulus onset, they will be viewed as coincident with automatic memory effects and incidental retrieval-related processes. If, however, they occur ~500 ms post stimulus onset or beyond, they will be more closely tied to explicit memory and controlled response selection processes.

**Materials and Methods**

**Subjects**

Participants were 30 healthy, right-handed subjects (seven men). Data from 17 of these participants were also included in Windmann and Kutas (Windmann and Kutas, 2001). Mean age of the 15 participants in the High Bias group (three men) was 21.0 years (range 18–30); mean age of the 15 participants in the Low Bias group (four men) was 22.9 years (range 18–32). Since the distribution of the response bias was approximately uniform for all 30 participants, the distributions within the High and the Low Bias groups were approximately symmetric and did not contain any outliers.

**Stimuli**

One hundred and fifty-eight verbs with a maximum frequency of 75 per million (mean = 8.59, SD = 11.89) as given by Kucera and Francis (Kucera and Francis, 1967) were used [the complete list is provided in the Appendix of Windmann and Kutas (Windmann and Kutas, 2001)]. Seventy of these were presented at study. Target and distracter item lists were counterbalanced across subjects.

**Procedures**

Participants were seated on a comfortable chair in a light- and sound-attenuated room facing a 21° computer monitor ~1.5 m away. They were instructed to memorize the words for a subsequent memory test. Emotionally negative and neutral words were randomly presented one at a time on the screen for 400 ms every 2600 ms with an interstimulus interval of 2200 ms. Only trials in which emotionally neutral stimuli were presented were analyzed for present purposes. Recognition memory was tested following a retention interval of ~30 min, during which subjects performed a lexical decision task with different materials. Participants
were instructed to indicate whether each word was old or new by pressing a button with their right or left hand as quickly as possible, and to guess if they were unable to recognize it. The hands used for 'old' and 'new' responses were counterbalanced across subjects. Each word appeared 1600 ms after a response to the previous word was given.

**ERP Recordings**

The electroencephalogram was recorded from 26 tin electrodes embedded in an elastic cap arranged in four equally spaced concentric rings; see Figure 7 in Windmann and Kutas (Windmann and Kutas, 2001). Two additional electrodes (LVP and RVP) were attached at left and right 'ventral' prefrontal sites (5% of the sagital midline dorsal to the nasion, and 10% of the interaural distance lateral to the sagittal midline). Online recordings were referenced to the left mastoid, and re-referenced offline to the mathematical average of the left and right mastoid. The electro-oculogram (EOG) was recorded to monitor eye movements and blinks. The horizontal EOG was recorded via electrodes placed at the outer canthus of each eye, referenced to each other; the vertical EOG was recorded with an electrode placed below the right eye, referenced to RVP.

Signals were amplified using a Nicolet SM2000 amplifier with a bandpass of 0.016–100 Hz (12 dB/octave), and digitized at 250 Hz for electronic storage. Eyeblinks were corrected in 26 of the subjects using an adaptive spatial filter procedure developed by A. Dale. Other artifact-contaminated trials (~14%) were excluded from further analyses. After artifact rejection, average trial counts for the ERP averages in the three analyses ranged between 57 and 72. Data were averaged and filtered using a digital bandpass filter of 0.2–15 Hz. ERP averages were digitally filtered with a lowpass filter of 10 Hz for the purposes of visual presentation only. For statistical analyses, mean amplitude measures were taken in early (300–500 ms), late (500–700 ms) and very late (1000–1500 ms) time-windows.

**Data Analysis**

Old/new discrimination accuracy \( Pr = \text{Hit} – \text{FA} \) and the response bias \( Br = \text{FA}/(1 – Pr) \) were computed according to two-high-threshold theory (Snodgrass and Corwin, 1988), where \( Pr \) is the probability of 'old' response to an old item, and \( Br \) is the probability of an 'old' response to a new item. Mean amplitudes for ERPs were taken from 24 electrode sites (the four midline sites were not included). Repeated measures ANOVAs were performed involving the two factors Hemisphere (left/right) and Anteriority (anterior/posterior), with mean ERP amplitudes collapsed across six sites for each factor level (Windmann and Kutas, 2001). Group Condition (Old/New) Site interaction effects turned out to be not associated with overall mean amplitude differences between the two groups in any of the three analyses; hence, normalization of the ERP data (McCarthy and Wood, 1985) would have yielded practically the same results.

**Results**

The median bias used for dividing the groups was 0.483, reflecting an almost perfectly neutral response criterion (which would be 0.5). This means that subjects in the High Bias group tended to guess ‘old’ when accurate memory failed while subjects in the Low Bias group tended to guess ‘new’.

Figure 1 shows the behavioral results. As previously mentioned, accurate old/new recognition was almost identical in the Low Bias and High Bias groups \( F(1,28) = 0.4 \). (The same was true for the emotionally negative items; \( F = 0.00 \), data not shown.) By contrast, there was a significant group difference in response bias \( F(1,28) = 72.83, P < 0.0001 \).

There was virtually no correlation between response bias and old/new discrimination performance \( r = –0.03 \), consistent with the assumption of statistical independence between the two measures (Snodgrass and Corwin, 1988).

An ANOVA of the reaction times with one between-subjects factor of Group and two within-subjects factors of Response Type (‘old’/‘new’) and Correctness of Response (correct/incorrect) showed no significant main effect of Group \((F < 0.1)\) or Response Type \((F < 1)\). However, the main effect for Correctness of Response was significant \( F(1,28) = 12.24, P < 0.005 \) with correct responses being overall faster than incorrect responses (see Fig. 1). The interaction between Response Type and Correctness of Response also was significant \( F(1,28) = 11.47, P < 0.005 \). Bonferroni-corrected post hoc tests showed that this interaction was due to the fact that within ‘old’ responses, correct responses (Hits) were made faster than incorrect responses [False Alarms, FA; correct ‘new’ responses to new items (Correct Rejections, CR), and incorrect ‘new’ responses to old items (Misses)].

![Figure 1. Behavioral results. Top: Hit rates and False Alarm rates for the High Bias and the Low Bias groups. Center: Old/new recognition accuracy \( Pr \) and the response bias \( Br \) of the High Bias and the Low Bias groups. Bottom: Reaction times of the High Bias and the Low Bias groups for correct ‘old’ responses to old items (Hits), incorrect ‘old’ responses to new items (False Alarms, FA), correct ‘new’ responses to new items (Correct Rejections, CR), and incorrect ‘new’ responses to old items (Misses).](image-url)
The ERP waveforms are shown in Figure 2. Results of the statistical analyses of mean ERP amplitudes taken in the early (300–500 ms) and late (500–700 ms) time-windows are presented in Table 1 (no significant differences involving the Hemisphere factor were found in any of the analyses).

All three comparisons yielded significant old/new ERP effects as old items elicited more positive waveforms than new items. Between 300 and 500 ms, analyses of the Subjective Old/New ERP Effect also yielded significant Group × Old/New and Group × Old/New × Anteriority interactions. These effects were due to the fact that the difference between ERPs to items judged ‘old’ versus those judged ‘new’ was greater in the Low Bias group than in the High Bias group, mainly at prefrontal sites (see Fig. 2, left column) where the old/new difference also significantly correlated with bias (r = –0.453, P < 0.013). Figure 3 shows these effects of bias by means of ERP difference potentials (old minus new) over all electrode sites. There was no such bias effect in the analysis of the Objective × Old/New × ERP Comparison. In the Correct Classification Old/New ERP analysis, low as compared to High Bias was associated with marginally greater old/new effects at anterior sites (see Fig. 2, right column).

When the sizes of the bias-related effects in the three analyses were compared by means of ANOVA, they were found to be significantly different, as revealed by a four-way interaction of Comparison Type × Group × Old/New × Anteriority [F(2,56) = 3.16, P < 0.05]. This finding directly confirms our presumption that the three comparisons would be differentially sensitive to the effects of bias, namely at anterior sites.

It is important to note that the reduced old/new difference in the High Bias group in the Subjective Comparison as compared to the other two comparisons is not solely due to a reduction of the positivity associated with ‘old’ items (as might be attributable to the higher proportion of False Alarms). This can be seen clearly from the ERPs at the ventral and medial prefrontal electrode sites depicted in Figure 2. The positivity to ‘old’ items is reduced at least to the same extent as the positivity associated with ‘new’ items is increased. Hence, the effects of bias on ERPs to ‘old’ as compared to ‘new’ items cannot be described in uniform terms (i.e. as either a reduction or an increase in positivity), but go into opposite directions.

This conclusion is confirmed by additional analyses of the data. In Figure 4, we plotted the ERPs of the two bias groups.
separately for Hits, Correct Rejections, False Alarms, and Misses. The central finding of note here is that the two bias groups show qualitatively different ERP old/new patterns. In the Low Bias group, False Alarms tend to pattern with Hits while Misses pattern with Correct Rejections. The reverse is true for the high Bias group where Hits pattern with Misses and False Alarms pattern with Correct Rejections. In other words, the ERP difference between False Alarms and Misses is positive for the Low Bias group [2.54 vs 1.44 µV; F(1,14) = 4.51, P < 0.05] but negative for the High Bias group [1.91 vs 2.91 µV; F(1,14) = 4.62, P < 0.05].

From these data patterns, we can infer what impact the inclusion of incorrect response trials has on grand average ERPs in recognition memory studies. If the bias is high (i.e. more 'liberal' than neutral), False Alarms will tend to reduce the positivity associated with correct recognition of old items (Hits), and Misses will tend to reduce the relative negativity associated with correct recognition of new items (Correct Rejections), mainly at frontal sites between 300 and 500 ms. However, when the bias is low (i.e. 'stricter' than neutral), these effects will be much less pronounced since old/new effects associated with false recognition (False Alarms and Misses) go into the same direction as those associated with correct recognition (Hits and Correct rejections).

In the later time-window between 500 and 700 ms, the ERP differences between the two bias groups were relatively small and not significant in any of the three analyses (see Table 1 and Fig. 2). However, there were significant old/new main effects in the Objective and the Correct Classification analysis. The classical old/new effect (difference between Hits and Correct Rejections), which is typically largest in this time-window at left parietal sites, was clearly present in both groups (see Fig. 4).

Between 1000 and 1500 ms, no significant or marginally significant group differences were observed in any of the three analyses. Hence, the effects of bias were evident only in the early time-window (between 300 and 500 ms poststimulus) at anterior sites.

Discussion
The present study used ERP measures to investigate the cognitive and neural mechanisms associated with different guessing tendencies in an old/new recognition memory task. By comparing the ERPs of normal, young adults who were equally accurate but had a high versus a low response bias, we could examine the relative courses of the recognition memory processes when a 'strict' as opposed to a 'liberal' response criterion is actualized from the moment a test stimulus is presented to the moment the response is rendered.

We found that ERPs were sensitive to bias at frontal recording sites, particularly between 300 and 500 ms post stimulus, albeit to varying extents depending on which trials comprised the old/new comparison: as we had expected, bias effects in this time-window were largest when what counted as old and new was completely determined by each participant’s responses regardless of the actual study status of the items (Subjective Old/New comparison), smallest when what counted as old and new was determined by the actual study status of the items regardless of participants’ responses; i.e. collapsed across correctness (Objective Old/New comparison), and intermediate in size for the canonical old/new comparison based on correctly recognized items (i.e. Hits vs Correct Rejections).

In the Subjective Comparison, the ERPs of the Low Bias group exhibited a marked old/new effect at (pre)frontal sites that was virtually absent in the High Bias group: ERPs associated with

---

Table 1

<table>
<thead>
<tr>
<th>Time-window</th>
<th>300–500 ms</th>
<th>500–700 ms</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Subjective Old/New Effect</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Old/New</td>
<td>14.72*</td>
<td>1.43</td>
</tr>
<tr>
<td>Group Old/New</td>
<td>5.52*</td>
<td>3.08*</td>
</tr>
<tr>
<td>Group × Old/New Anteriority</td>
<td>5.58*</td>
<td>0.96</td>
</tr>
<tr>
<td><strong>Objective Old/New Effect</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Old/New</td>
<td>10.77*</td>
<td>11.70*</td>
</tr>
<tr>
<td>Group Old/New</td>
<td>1.65</td>
<td>1.46</td>
</tr>
<tr>
<td>Group Old/New Anteriority</td>
<td>0.00</td>
<td>0.83</td>
</tr>
<tr>
<td><strong>Correct Classification Old/New Effect</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Old/New</td>
<td>20.22*</td>
<td>5.99*</td>
</tr>
<tr>
<td>Group Old/New</td>
<td>0.23</td>
<td>0.33</td>
</tr>
<tr>
<td>Group Old/New Anteriority</td>
<td>3.73(*)</td>
<td>1.48</td>
</tr>
</tbody>
</table>

Significant and marginally significant effects are marked with an asterisk.

---

Figure 3. Topographic maps of ERP difference waves (old minus new) of the High Bias and the Low Bias groups in the early time-window (between 300 and 500 ms). Refer to the legend of Figure 2 for further information.
old' responses were more positive than those associated with 'new' responses. All four response conditions (i.e. Hits, Correct Rejections, False Alarms and Misses) were found to contribute to this interaction, indicating that high versus low bias was associated with qualitatively different ERP old/new effects, not just with a quantitative difference due to more hits and fewer misses or vice versa among the 'old' and 'new' responses.

By contrast, in the Objective Comparison, where the old/new distinction was based on the actual study status of the items, old items elicited more positive ERPs than new items at virtually all recording sites in both the low and high bias groups in this early time-window. This effect thus seems to reflect the actual strength of the memory traces for old and new items, i.e. the bottom-up signal, independent of bias. The absence of any group differences in behavioral accuracy supports this interpretation.

Finally, for the old/new comparison limited to correct responses, ERP differences between the two bias groups appeared to be a marginally significant remnant of that observed for the Subjective Old/New comparison with regards to both timing and scalp distribution. This result shows that old/new effects typically reported in ERP studies of recognition memory cannot be interpreted unambiguously in terms of accurate recognition unless bias effects have been accounted for. In other words, the standard comparison between Hits and Correct Rejections may include influences from response bias, and not just processes associated with accuracy, and thus must be interpreted with this possibility in mind. This concern also holds for the interpretation of any neuroimaging data in which the signals of interest are averaged across the trials of one or another response category (e.g. correct or incorrect), as these inevitably reflect some mixture of bias- and accuracy-related processes.

It is important to note here that the proportion of trials associated with accurate old/new recognition (hits versus correct rejections) was the same in all three of the comparisons (Subjective, Objective and Correct Classification). What differed was the extent to which the old/new distinction on which the ERP analysis was based corresponded with each participant's subjective perspective of what was old and what was new. So the crucial difference between the three analyses was the degree to which the response criteria of the High Bias and Low Bias groups (= the top-down process) as opposed to their accurate recognition memory (= the bottom-up signal) determined the old/new distinction. It is this difference that determined the size of the (pre)frontal old/new effects in the High Bias and the Low Bias groups.

We think that these early, prefrontal ERP effects reflect the criterion setting functions of the prefrontal cortex (Schacter et al., 1998; Swick and Knight, 1999; Miller et al., 2001). The prefrontal cortex is considered to be crucially involved in initiating, monitoring and controlling item-retrieval from memory (Buckner, 1996; Fletcher et al., 1998; Schacter et al., 1998; Wagner et al., 1998; Henson et al., 1999; Tomita et al., 1999). Dysfunctions of these regions are associated with more False Alarms, intrusion errors, false memories and confabulations (Schacter et al., 1996; Béliveau et al., 1997a; Moscovitch and Melo, 1997; Melo et al., 1999; Schindler and Patak, 1999; Swick and Knight, 1999), usually in conjunction with relatively high hit rates. Presumably, the role of the prefrontal cortex during memory retrieval is to maintain a description of the information being sought, e.g. in terms of familiarity, context or source (Koutstaal and Schacter, 1997; Moscovitch and Melo, 1997; Ranganath and Paller, 2000), and to actively inhibit
memory traces that do not match this description (Schnider and Ptak, 1999; Elliott et al., 2000; Schnider et al., 2000; Anderson and Green, 2001). These control functions may serve to prevent memories that are irrelevant to the task at hand from guiding ongoing thoughts and actions (Schnider and Ptak, 1999; Schnider et al., 2000). We think that both bias groups made use of these mechanisms during recognition. However, we suggest that they did so to differing degrees presumably because they differed in their appreciation of what was task (ir)relevant. Participants in the Low Bias group seem to have endowed information signaling ‘newness’ of the test items at the expense of information signaling familiarity with the consequence of facilitating ‘new’ responses while inhibiting ‘old’ ones. On the contrary, those in the High Bias group gave ‘old’ responses more often and more quickly than ‘new’ responses, suggesting greater reduction or ‘relaxation’ of the inhibitory control of currently irrelevant memory by prefrontal cortices than in participants with a stricter bias (Hirshman et al., 2000). Indeed, ‘old’ eventuated in a response rate pattern resembling that typically observed in patients with frontal lobe damage (Swick and Knight, 1999).

The qualitatively different response time patterns of the High and Low Bias groups, however, cannot account for their different ERP old/new effects because the overall average of the mean reaction times did not differ for the two groups. Response times varied only as a function of the direction of the bias, not independent of it (i.e. ‘new’ responses were facilitated by the bias to respond ‘new’ while ‘old’ responses were facilitated by the bias to respond ‘old’). In fact, this pattern suggests that response preferences serve to speed response times up, i.e. to allow for quick responding even if there is uncertainty as to what the correct reaction should be.

Having shown that the frontal ERP effects are related to bias, we can look to their timing to provide us with further constraints on the nature of the mechanisms involved. These effects are maximal between 300 and 500 ms, a time-window that has been found to be particularly sensitive to automatic (as opposed to controlled) memory processes (Paller et al., 1995; Düzel et al., 1997, 2001; Allan et al., 1998; Rugg et al., 1998; Curran, 1999, 2001; Mecklinger, 2000; Paller, 2000). This timing suggests that participants’ decision criteria were set even before they called upon controlled processes to recollect any specific item information. This conclusion then would be at odds with any theoretical view that links criterion setting functions to post item-retrieval or response-selection processes in recognition memory (Light and Kennison, 1996; Miller and Wolford, 1999; Schacter et al., 1996, 1998; Swick and Knight, 1999). Likewise, this conclusion does not jibe with the proposal that bias-related mechanisms become active only after participants realize that mechanisms supporting accurate memory have failed (Riefer et al., 1994; Buchner et al., 1995; Wallsten et al., 1999). Indeed, our findings suggest instead that the criteria underlying the bias to respond ‘old’ or ‘new’ are set prior to any attempts at controlled recollection and response selection.

In this sense, then, the recognition memory process seems predisposed towards either an ‘old’ or a ‘new’ decision from the moment at which the earliest memory retrieval processes are initiated. Naturally this bias can subsequently be overridden or modulated by information about the actual study status of each item. However, when such information is only weakly represented or cannot be accessed, participants respond according to their preset bias. On this view, criterion-setting mechanisms function much like an ‘old’ or ‘new’ gate (depending on the individual’s bias) during the recognition memory process. They seem to remain active for as long as attempts to retrieve information from memory are made. Finally, the retrieved memory signal is ‘compared’ with this criterion-related signal (i.e. the two processes interact in some way) to determine the final response decision that is passed on to the motor systems.

As previously mentioned, however, it is important to note that response biases can be influenced by a whole host of variables, including those inherent in the test stimuli. Accordingly, we would not expect all bias-related influences to be evidenced in the same ERP effects. The temporal characteristics of various bias-related effects are especially likely to vary with the nature of the task and the complexity of the criterion descriptor employed in memory retrieval. In the present study, the relatively high proportion of the simple, quick old/new classifications rendered was probably based mainly on familiarity and perceptual fluency (or the lack thereof). These effects of bias were, therefore, maximal before 500 ms post stimulus, although they were not significantly present in the subsequent time window as well.

Complex search criteria involving more specific requirements such as context or source might require more comparative iterations as sketched above and would thus be more likely to yield more prolonged ERP bias effects, perhaps with a later onset (Wilding and Rugg, 1996; Donaldson and Rugg, 1999; Ranagath and Paller, 2000). That said, it should be noted that the temporal and spatial topography of the ERP effects reported herein are similar to those observed for emotion-induced variations of decision bias (Windmann and Kutas, 2001). The only difference is that in our previous report, it was negative emotion that induced a greater bias to respond ‘old’. Hence in that case bias effects were only evidenced in the ERPs associated with ‘old’ responses (Hits and False Alarms), and not with ‘new’ responses. Additional analyses of the data (not presented) confirmed that the emotion-related bias effects and the subject-specific bias effects reported herein are independent of one another. That is, in the present sample, the Subjective Comparison (which is sensitive to the subject-specific effects of bias) is insensitive to the effects of emotion, whereas the Hits versus False Alarms comparison (which is sensitive to the effects of emotions on the bias) does not show any significant ERP difference between the High and the Low Bias groups.

Finally, we consider why it is that different participants might develop different response preferences in a recognition memory task under identical stimulus and task conditions. In other words, we examine what it might mean when an individual employs a ‘liberal’ as opposed to a ‘strict’ decision criterion during memory retrieval. We think that different decision biases reflect different predictions the brain makes about the probability and the nature of upcoming events — heuristics that can guide behavior and facilitate responding in the face of uncertainty (Elliott et al., 1999, 2000; Wallsten et al., 1999). Hirshman observed a systematic reduction in the response bias variability across participants in a recognition memory task when they were provided feedback about the correctness of their responses (Hirshman, 1995). Perhaps, subject-specific variability in decision biases occurs only or primarily when the circumstances and/or input are under-specified, leaving individuals with no means of verifying the appropriateness of their responses, as was the case in the present study. Elliott et al. (Elliott et al., 1999, 2000) suggested that these guessing functions depend crucially on the orbitofrontal brain regions, consonant with our interpretation. Other evidence suggests that individuals who feel very uncertain about past events also might relax their retrieval criteria as a natural means of compensating for poor recollection, thereby increasing their ‘hit rates’ (Jacoby, 1999; Miller and Wolford, 1999; Swick
and Knight, 1999). Shifts in retrieval criteria may serve social functions as well. For instance, when people recount past events with the intent of convincing, entertaining, teaching, amusing, impressing, pleasing or deceiving others, they bias their retellings as well as their actual memories (Schneider and Watkins, 1996; Tversky and Marsh, 2000). These sorts of findings clearly have important implications in forensic as well as clinical contexts (Brebion et al., 1997a; b; Windmann and Krüger, 1998; Carli, 1999), and are relevant in gaining a better understanding of cognitive alterations in normal aging (Nielsen-Bohlman and Knight, 1995; Chao and Knight, 1997; Jacoby, 1999).

**Notes**

This work was supported by grants MH52893, HD22614 and AG08313 to M.K. and a post-doctoral scholarship of the DAAD (Bonn, Germany) to S.W. within the ‘Hochschulsonderprogramm III von Bund und Ländern’. Corresponding author: Dr Sabine Windmann, Ruhr-University Bochum, Faculty of Psychology, AE Biopsychology, GAFO 05/623, 44780 Bochum, Germany. Email sabine.windmann@ruhr-uni-bochum.de.

**References**


