### Neural Correlates of False Memories



Study





Studied as picture or not?

Table 1. Behavioral dat	a in the memory t	est.	
	Resp	onse	
	Yes	No	
<b>Proportion of responses</b>	± s.e.m.		
Old word plus picture	0.75 ± 0.03	0.25 ± 0.03	-
Old word only	0.30 ± 0.03	0.70 ± 0.03	FALSE
New	0.09 ± 0.02	0.91 ± 0.02	MEMORIES
Response time (ms; mea	nn ± s.e.m.)		-
Old word plus picture	1133 ± 25	1308 ± 42	
Old word only	1279 ± 38	1305 ± 43	
New	1354 ± 54	1193 ± 42	

"Old word plus picture' trials with a 'yes' response are termed 'accurate picture memories,' and 'old word only' trials with a 'yes' response are termed 'faise memories.'

There are 30% false memories in word only condition.

Paller lab, Northwestern

#### Midline occipital site





Fig. I. Brain potentials associated with the formation of true and faise memories. (a) ERPs to objects were averaged according to whether those objects were subsequently remembered. (b) ERPs to words were averaged according to whether people later mistakenly thought they had seen the corresponding objects. Recordings were from the midline occipital scalp location. (c) Topographic maps were interpolated based on differences between pairs of waveforms at each electrode location (shown by small circles). (A dapted from Ref. a.)

The more vivid, detailed, or robust the visual imagery generated in response to a single word, the more likely the memory for that imagery will be mistakenly attributed to a memory resulting from actually viewing the corresponding object.



Fig. 3. Topographic maps of the primary studyphase and test-phase effects. Statistical comparisons of mean ERP amplitudes at each electrode site were used to generate t-maps with spherical spline interpolation on a schematic view of the head from above. (a) Study-phase topography showing t-values from the comparison of later false memories to later correct rejections from 600-900 ms. Positive t-values > 2.23 (uncorrected) indicate regions where ERPs in response to later false memories were significantly more positive than ERPs in response to later correct rejections. (b) Test-phase topography showing t-values from the comparison of accurate picture memories to false memories from 900-1200 ms, RT - matched. Positive t-values > 2.20 (uncorrected) indicate regions where ERPs to accurate picture memories were significantly more positive than ERPs to false memories.



Posterior distribution of false memory Dm is consistent with hypothesis that Dm is related to imagery

### Summary: Dm (difference due to memory) effect

Items that are subsequently remembered (recognized or recalled) are often although *not always* associated with more positive ERPs between 300-800 ms post-item onset.

- distribution of effect varies with nature material
- size of Dm effect varies with nature of task (incidental vs intentional, shallow vs elaborative); larger for deeper & more distinctive encoding
- overlaps several components including N400, P300/P600/LPC



Fig. 1. Grand mean ERPs elicited by words during a study phase averaged according to whether they were subsequently correctly recognized (i.e. hit) or not recognized (i.e. miss) during the subsequent test phase. Shading between the waveforms indicates the Dm or subsequent memory effect. The difference waveform (subsequently hit-subsequently missed) is depicted to the right of the unsubtracted waveforms. Arrows mark stimulus onset, with time lines every 200 ms.

### TYPICAL RETRIEVAL PARADIGM

Items are presented two times (or more) Compare ERPs at first vs second (or second+) presentations

The difference is memory-related, but what kind of memory depends on task

#### RETRIEVAL DISTINCTIONS FOR LONG TERM MEMORY

# *Indirect (implicit) memory task*: no mention of prior episode, or memory retrieval (*repetition paradigm*)

- lexical decision
- semantic judgment
- identification of visually degraded stimuli
- stem completion (sta \_\_\_)

# *Direct (explicit) memory task*: conscious attempt to retrieve memory (*recognition paradigm*)

- participants asked to recognize items, detect repetitions
- old vs new
- remember vs know vs new
- old vs new, plus source?

# Recognition and repetition paradigms often differ only in instructions at retrieval and therefore in the possible inferences allowed!

At some locations, recognition and repetition effects are indistinguishable

Repetition: repeated more positive than unrepeated from 300 ms+

Recognition: correctly recognized old items more positive than new items from 300 ms+



Negativity between 300-500 ms (N400 region) is reduced in amplitude for old versus new and for repeated versus unrepeated, and the following positivity (LPC/P600) between 500-800 ms is also larger for old than new, and repeated versus not repeated for recognition and repetition, respectively, in most cases.



### Old/New ERP recognition effect

ERPs to items correctly recognized as Old (Hits) are more positive from ~300-700 ms than ERPs to unstudied items correctly identified as New (CRs).

#### There are many different explicit memory old-new effects!



Fig. 4. Grand mean ERP waveforms elicited by correctly recognized old and correctly rejected new items from Johnson et al. (1998a). The left column depicts the old and new waveforms at the electrode site and hemiscalp where that subcomponent was largest. Repoduced from Johnson et al. (1998a) with permission of the publisher.

### EXPLICIT MEMORY EFFECTS

There is a family of old/new EM (explicit memory) effects that are distinguished by time course, scalp distribution, sensitivity to experimental variables

#### 1. Left parietal old-new effect

- overlaps N400 (300-500), P300 (sometimes called P600 or LPC, 500-800)
- largest L. temporo-parietal

- recollection

#### 2. Right (pre)frontal old-new effect

- late, lasts much longer than parietal old-new effect
- functional significance controversial (e.g., source memory? relational processing)
- 3. Left medial frontal old new effect (FN400)
  - 300-500 ms
  - functional significance also controversial (e.g., familiarity versus conceptual priming)



#### Left Parietal Old/New effect

Old/New Effect



Fig. 2. Top row. Grand mean ERPs elicited by previously studied (old) and unstudied (new) words recorded at the midline parietal (Pz) scalp site during the test phase of a study/test paradigm. Bottom row. The result of subtracting the ERPs elicited by new items from the ERPs elicited by old items, i.e. the old/new difference waveform. In the top row, vertical hash marks indicate the N400 region, and horizontal hashmarks, the P3b region of the old/new effect; in the bottom row, these are referred to as the 'early' and 'late' regions of the old/new effect. Arrow marks stimulus onset, with time lines every 200 ms.

#### OLD/NEW EFFECT

ITEM RECOGNITION TASK



Figure 1. Grand average event-related potentials (ERPs) from 24 participants during the item-recognition task of Experiment 1. ERPs elicited by correctly categorized old and new items were compared at all electrode sites. Sites over the left side of the head are shown in the left column from anterior (top) to posterior (bottom), midline sites in the middle column, and right scalp sites in the right column. Negative voltage is plotted upward. Pf corresponds to prefrontal, B to inferior prefrontal (Broca's), T to temporal, W to parietotemporal (Wernicke's), and O to occipital. Le denotes an electrode site below the right eye. Fz, Cz, and Pz are located at the frontal, central, and parietal midline, respectively.





A. Trott et al., 1999
+ 5μV
'-
B. Wilding, 1999
<ul> <li>Hit &amp; Source correct</li> <li>Hit &amp; Source incorrect</li> <li>Correct rejection</li> </ul>

Traditional recognition paradigm requires old/new comparison. But, old and new items require different responses. So, how do know that ERP difference seen is related to old/new rather than to the different responses? It's a confound that leads to two different explanations of the effect

**Two hypotheses**: old/new effect is reflection of memory old/new effect is due to different responses

Design: Continuous recognition paradigm across several blocks (Rugg)



Do these two differ? They have different overt responses but both are repetitions.

**Predicted outcomes**: if ERP due to memory, then *table* and *glass* will have similar ERPs; if due to response processes, then these 2 words will have different ERPs.



Old-new ERP effect reflects explicit memory for item not response!



L. Parietal old-new effect is not seen for misses (also old) or false alarms (thought to be old). It requires an accurate "OLD" judgement.

Its amplitude is related to retrieval success.

L. Parietal old/new effect indexes recognition based on recollective processes.

### Left Parietal Old-New effect

Often large over L. parietal sites, though widespread

Larger for consciously remember items (remember>know>new)

Sensitive to depth of encoding manipulations (larger for more deeply encoded)

Larger for items recognized and recalled than just recognized

Larger for items for which source is known

Presumed to index recollective processes and to depend on intact medial temporal lobe structures.

#### ITEM RECOGNITION TASK



Figure 1. Grand average event-related potentials (ERPs) from 24 participants during the item-recognition task of Experiment 1. ERPs elicited by correctly categorized old and new items were compared at all electrode sites. Sites over the left side of the head are shown in the left column from anterior (top) to positrior (bottom), midline sites in the middle column, and right scalp sites in the right column. Negative voltage is plotted upward. Pf corresponds to prefrontal, B to inferior prefrontal (Broca's), T to temporal, W to parietotemporal (Wernicke's), and O to occipital. Le denotes an electrode site below the right eye. Fz, Cz, and Fz are located at the frontal, central, and parietal midline, respectively.





Figure 3. Grand average event-related potentials from 24 participants during the sourcerecognition test of Experiment 1. "Fift" includes all trials with successfully recognized words, independent of the accuracy of the voice judgment (e.g., hit-hit plus hit-miss). Left scalp sites are displayed in the left column, midline sites in the middle column, and right scalp sites in the right column, from most anterior (top) to most posterior (bottom). Pf corresponds to prefrontal, B to inferior prefrontal (Broca's), T to temporal, W to parietotemporal (Wernicke's), and O to occipital. Le denotes an electrode site below the right eye. Fz, Cz, and Pz are located at the frontal, central, and parietal midline, respectively.

SENKFOR AND VAN PETTEN

#### ITEM RECOGNITION TASK



Figure 1. Grand average event-related potentials (ERPs) from 24 participants during the item-recognition task of Experiment 1. ERPs elicited by correctly categorized old and new items were compared at all electrode sites. Sites over the left side of the head are shown in the left column from anterior (top) to positrior (bottom), midline sites in the middle column, and right scalp sites in the right column. Negative voltage is plotted upward. Pf corresponds to prefrontal, B to inferior prefrontal (Broca's), T to temporal, W to parietotemporal (Wernicke's), and O to occipital. Le denotes an electrode site below the right eye. Fz, Cz, and Fz are located at the frontal, central, and parietal midline, respectively.

#### SENKFOR AND VAN PETTEN





Figure 3. Grand average event-related potentials from 24 participants during the sourcerecognition test of Experiment 1. "Hit" includes all trials with successfully recognized words, independent of the accuracy of the voice judgment (e.g., hit-hit plus hit-miss). Left scalp sites are displayed in the left column, midline sites in the middle column, and right scalp sites in the right column, from most anterior (top) to most posterior (bottom). Pf corresponds to prefrontal, B to inferior prefrontal (Broca's), T to temporal, W to parietotemporal (Wernicke's), and O to occipital. Le denotes an electrode site below the right eye. Fz, Cz, and Pz are located at the frontal, central, and parietal midline, respectively.



### (Right) (Pre)Frontal old-new ERP effect

- starts about same time as parietal old-new effect but frontal
- lasts longer than parietal old-new effect
- laterality is inconsistent/controversial!
   bilaterally symmetric vs right
- functional significance is controversial!
   processes that operate on products of retrieval vs source memory or related control processes
- especially prominent during source memory tasks, if so why?
- depends on intact prefrontal cortex



Fig. 6. ERP data recorded at left parietal and right prefrontal electrode sites from two different source memory paradigms. A: From Trott et al. (1999). Thick solid lines indicate the ERPs associated with correctly rejected new items; thin solid lines represent the ERPs associated with hit trials for which the list was correctly judged; dashed lines represent the ERPs associated with hit trials for which the list was incorrectly judged. B: Data from Wilding (1999). Thick solid lines represent the ERPs to correctly rejected new items; thin solid lines represent the ERPs to hit trials for which the gender of voice was correctly judged; dashed lines represent the ERPs to hit trials for which the gender of voice was incorrectly judged. Modified from Wilding, 1999, with permission of the publisher.



Fig. 7. Scalp distributions of the ERP EM effects depicted in Figure 5. A: Data from Trott et al. (1999) for hits associated with list correct and incorrect judgments. The isopotential lines are separated by 0.40 microvolts. B: Hits that attracted correct gender of voice judgments. Reproduced from Wilding, 1999, with permission of the publisher.

Prefrontal Engagement during Source Memory Retrieval Depends on the Prior Encoding Task Trudy Y. Kuo and Cyma Van Petten

The prefrontal cortex is strongly engaged by some, but not all, episodic memory tests. Prior work has shown that source recognition tests—those that require memory for conjunctions of studied attributes—yield deficient performance in patients with prefrontal damage and greater prefrontal activity in healthy subjects, as compared to simple recognition tests. Here, we tested the hypothesis that there is no intrinsic relationship between the prefrontal cortex and source memory, but that the prefrontal cortex is engaged by the demand to retrieve weakly encoded relationships. Subjects attempted to remember object/color conjunctions after an encoding task that focused on object identity alone, and an integrative encoding task that encouraged attention to object/color relationships. After the integrative encoding task, the late prefrontal brain electrical activity that typically occurs in source memory tests was eliminated. Earlier brain electrical activity related to successful recognition of the objects was unaffected by the nature of prior encoding. JCN



Kuo & Van Petten, JCN, 2006

#### **STUDY PHASE** (during encoding)



в.





RETRIEVAL

HYPOTHESIS: Here, we tested the hypothesis that there is no intrinsic relationship between the prefrontal cortex and source memory, but that the prefrontal cortex is engaged by the demand to retrieve weakly encoded relationships (vs alternative)

What pattern of ERPs would support this hypothesis? And which would not?

- what experimental conditions and what comparisons within a condition
- what electrodes should be focus
- what is expected direction of ERP differences

#### ERPS DURING SOURCE MEMORY TEST



Prefrontal engagement in source memory task is not mandatory!

Role of the prefrontal cortex is to aid in the recovery of weakly encoded relationships – retrieving attribute conjunctions. This burden for prefrontal cortex however can be alleviated with the right sort of encoding (e.g., integrative). There is no intrinsic relationship between prefrontal cortex and source memory. There are many different explicit memory old-new effects!



Fig. 4. Grand mean ERP waveforms elicited by correctly recognized old and correctly rejected new items from Johnson et al. (1998a). The left column depicts the old and new waveforms at the electrode site and hemiscalp where that subcomponent was largest. Repoduced from Johnson et al. (1998a) with permission of the publisher.

### Rugg et al. (1998)



### Controversy on functional significance of Left medial (pre)frontal EM component Negativity in N400 (300-500 ms) region or FN400

Anterior/Frontal N400 has been related to familiarity (as distinct from recollection)

versus

Anterior/Frontal N400 has been related to verbal/conceptual priming.

Relates to broader issue of whether familiarity and recollection are psychologically and neurally different processes or whether the two are cut from the same cloth – i.e., similar mechanisms but differing quantitatively.



### **Recollection and Familiarity**

Dual Process theories of recognition memory

*Familiarity* – unsubstantiated experience of having previously encountered a given item or event.

**Recollection** – conscious retrieval of specific information regarding studied items (physical attributes, associative/contextual information)

### Alternative Hypotheses

Familiarity is just the same as recollection. The same neural mechanisms are involved in both, but to a different degree.

Familiarity and recollection are distinct neural processes. Different neural mechanisms underly each of them.

### Typical recognition paradigm



One way to differentiate between "familiarity" and "recollection" is by comparing Know (familiarity) vs. Remember (recollection) responses but there are others.

### Dissociating familiarity from recollection (Curran, 2000)

	Study Phase		Test Phase		Exan old/new	nine effects	
Lis plu <i>Ex</i>	t of singular a Iral words <i>ample: car, p</i> o	and 1) ots 2)	Previously stu items - OLD Words with th	idied e	<u>Define Reco</u> Studied "yes	<u>llection as</u> " – Similar	: ^ "yes"
		3)	opposite plura <i>(cars, pot)</i> - N New words	ality EW	<u>Define Famil</u> Similar "yes"	<u>iarity as:</u> ' – New "ne	0"
			response	1	familiarity	recollec	tion
Studied: (S) car , (T) car		Yes - hit No - miss		high ?	high Iow		
Simi	l <u>ar:</u> (S) car , (	T) cars	Yes – false ala No – correct re	arm jection	high high	low high	
Nev	<u>w:</u> (S) car , (T	) lamp	Yes – false ala No – correct re	arm jection	? low	? low	

### Dissociating familiarity from recollection (Curran, 2000)



#### **Critical assumptions for comparisons and inferences:**

- 1. Correctly recognized studied (YES): high in recollection and familiarity
- 2. Incorrectly recognized similar (opposite number s,pl) words (YES): high in familiarity
- 3. Correctly recognized new words (NO): low in familiarity

### Predicted Outcomes for any ERP effect



The two hypotheses differ in their prediction for the <u>new similar items that participants</u> <u>mistakenly consider old.</u>

### Left Parietal old/new ERP effect (400-800ms)



**Result:** New [no] = Similar [yes] < Studied [yes] **Inference**: This potential reflects recollection and not familiarity

### Frontal 300-500ms (FN400)



**Result**: New [no] < Similar [yes] = Studied [yes] **Inference**: This potential reflects familiarity not recollection

### Curran studies conclusions

- Recollection and familiarity are two distinct processes manifested by different neural mechanisms
- The parietal LPC is associated with recollection
- The frontal N400 is associated with familiarity

### But, is it really familiarity that Curran is measuring?

Yovel & Paller question the experimental paradigms used to assess familiarity. They suggest that Remember/Know paradigms are contaminated by subjective introspection and suggest that when items are known (i.e., familiar) before the experiment then there is a bias to respond familiar. Rapid Communication

#### The neural basis of the butcher-on-the-bus phenomenon: when a face seems familiar but is not remembered

Galit Yovel1 and Ken A. Paller\*

Department of Psychology and Institute for Neuroscience, Northwestern University, Evanston, IL 60208-2710, USA

Received 28 February 2003; revised 9 September 2003; accepted 12 September 2003

A common distinction in contemporary research on episodic memory is between familiarity, an unsubstantiated impression that an event was experienced previously, and recollection, remembering some information plus the spatiotemporal context of the episode in which it was acquired. The epitome of pure familiarity-the butcher-on-the-bus phenomenon-occurs when one believes that a person is familiar (often upon seeing their face in an atypical context) while failing to recall any information about that person whatsoever. Prior research on familiarity and recollection has relied on verbal material. Whereas word meanings and pronunciations are well learned in advance, here we produced pure familiarity and recollection using photographs of faces never seen before the experiment. When participants recognized a face, recollection was inferred if they also remembered either the occupation associated with that face earlier in the experiment or any other episodic detail. Pure familiarity was inferred when recognition occurred in the absence of any such contextual retrieval. Analyses of brain potentials recorded during initial encoding showed that right-sided neural activity predicted subsequent face familiarity, whereas bilateral potentials predicted subsequent face recollection. Results during memory testing were inconsistent with the popular idea that familiarity is generically indexed by reduced frontal N400-like potentials. Instead, both memory experiences were associated with bilateral, parietalmaximum brain potentials, although with smaller amplitudes and for a shorter duration for familiarity. These similarities between electrophysiological correlates of pure familiarity and recollection suggest that familiarity with faces may arise by virtue of a subset of the neural processing responsible for recollection.

© 2003 Elsevier Inc. All rights reserved.

Keywords: Memory for faces; ERPs; Familiarity; Recollection



Fig. 1. Schematic representation of the memory paradigm. (A) In the study phase, participants viewed a series of faces, each paired with a unique spoken occupation. Participants were instructed to remember these face-occupation associations, and also to respond on each trial indicating whether the pairing was thought to be a good or bad fit. Pilot data showed that this procedure led to a high proportion of pure familiarity judgments. (B) In the test phase, faces were presented without occupations. The first response indicated whether the face was old or new (i.e., one from the study phase or not, respectively). A second response was made for faces endorsed as old to indicate whether (1) the occupation could be recalled, in which case the occupation was spoken as the third response; (2) only other specific details could be remembered from the study phase episode (such as an observation made about the expression or a noted resemblance to a friend); or (3) that no specific information from the study episode could be remembered (familiarity without recollection).

Face + spoken occupation; fit?

Old or new?

# "The Butcher on the Bus" (Yovel & Paller 2004)



Butcher





Test

**Phase** 

Response: New, Old + Occupation, Old + other specifics, Old no specifics



### **Recollection defined as:**

Old + Occupation/other specifics

Familiarity defined as:

Old with no specifics



Memory effects are larger amplitude for recollection than familiarity trials, but otherwise not different, even in topography.



Amplitude ~400ms systematically varies as a function of subsequent memory, and is larger for subsequent memory based on recollection than on familiarity.

### Yovel & Paller (2004) conclusions

The recollection and familiarity ERP effects seem to be *quantitatively but not qualitatively different*; both elicit a late parietal positivity, although is larger in amplitude under recollection

Familiarity seems to arise from a subset of the neural processing responsible for recollection

Whatever it is, the FN400 does not reflect familiarity!

#### IMPLICIT MEMORY

#### **REPETITION PARADIGM**

... light strand grape snow blint spring table light done rabbit...

...light strand grape snow lion spring table light done rabbit...

TASKS: lexical decision (respond to nonword) respond/count animals

#### **REPETITION EFFECT**



The brain's processing of a stimulus differs on 1<sup>st</sup> and 2<sup>nd</sup> presentation. Negativity between 300-500 ms (N400) reduced in amplitude and positivity (LPC/P600) between 500-800 ms is increased in amplitude for repeated versus unrepeated items (initial presentation). But, what does repetition effect reflect – repetition of what?

What does ERP repetition effect reflect?

Extraction of visual features? Extraction of word form? Extraction of meaning?

# Nonword:XXVTQ (checking visual level)Pseudoword:BLURN (checking word form)Word:TABLE (checking word & meaning)

Related Words: SOFA-COUCH (checking just meaning)

Nonword:	XXVTQ	(checking visual level)	NO Rep effect
Pseudoword:	BLURN	(checking word form)	Rep effect
Word:	TABLE	(checking word & meaning)	Rep effect

Related Words: SOFA-COUCH (checking just meaning) Rep effect



Fig. 1. Examples of the structurally possible and impossible figures employed in experiment 1 (A), and the non-target and target patterns employed in experiment 2 (B).

Rugg et al. 1995

#### STRUCTURALLY POSSIBLE OBJECTS

#### NON-TARGET PATTERNS



Fig. 2: Grand-swrage ERP waveforms elicited in experiment 1 by structurally possible objects on their first and second presentations. Fz, Cz, Pz sign frontal, central and parietal middline electrodes. EF, RF, LTa, RTa, LP, RP, LTP, RTP, LO, RO signify left and right frontal, anterior temporal, parietaposterior temporal, and occipital electrodes. Note that the gain of EOG channel is ×4 lower than that for the ERPs.



Fig. 3. Grand-average ERP waveforms efficited in experiment 2 by non-target patterns on their first and second presentations. Gain of the EOG channel is ×4 lower than that for the ERPs.

Rugg et al. 1995

#### Positive up

ERPs TO FIRST AND SECOND STIMULI



Figure 7. Event-related brain potentials (ERPs) elicited by first presentations (dashed line) and second presentations (solid line) of targets, and the ERP repetition effect in the five tasks.

Illegal nonwords, distorted pictures, and other items do not elicit this sort of ERP repetition effect, although pseudowords do. So, it seems that reasonable potential for meaning is important eliciting an ERP repetition effect in N400/LPC regions.

### Is the repetition effect a single effect?



Repetition influences amplitude of negativity in N400 region and the following positivity – N400 is reduced with repetition and LPC/P600 is increased with repetition. The two subcomponents are functionally distinguishable (e.g., frequency).