BIOLOGICAL PSYCHOLOGY

Biological Psychology 47 (1998) 121–135

Context effects in a category verification task as assessed by event-related brain potential (ERP) measures

Hans-Jochen Heinze [a,*], Thomas-Frank Muente [b], Marta Kutas [c]

[a] *Department of Clinical Neurophysiology, Otto-von-Guericke University, Leipziger Straße 44, 39120 Magdeburg, Germany*

[b] *Neurologische Klinik, Medizinische Hochschule, Hannover, Germany*

[c] *Department of Cognitive Science, University of California, San Diego, CA, USA*

Received 24 February 1995; received in revised form 2 June 1995; accepted 15 October 1995

Abstract

This study examined the extent to which decisions about category membership (between category structure) are dependent on or dissociable from the typicality structure (within category structure) of categories. To this end, behavioral and event-related brain potential (ERP) measurements were obtained in a category verification task that incorporated a context manipulation (list composition effect) that was previously shown to affect membership verification. In a German and an American English version of the experiment, pairs of words were presented and subjects were asked to decide whether or not the second word was an exemplar of the superordinate indicated by the first word. Both versions yielded similar results. The behavioral data showed that both typicality and context manipulation affected the verification times for the true exemplars. Furthermore, atypical true exemplars were affected more by context manipulation than mere typical ones. By contrast, the N400 component of the ERPs elicited by the true exemplars was influenced by typicality, but invariant in the face of the context manipulation. These results underscore the previously shown dissociation between reaction times and N400 changes and suggest that category membership decision (as reflected in verification times) and semantic relatedness or typicality

* Corresponding author.

PII S0301-0511(97)00024-0

of an exemplar (as reflected in N400 modulation) can vary independently from each other and, therefore, likely index the activity of different brain systems.

Keywords: Event related brain potentials; N 400; Category membership; Semantic relatedness

1. Introduction

The storage and retrieval of knowledge about the world in relation to perceived words or sentences is usually attributed to a 'semantic memory' system. Indeed, our understanding of an utterance is critically dependent on memory functions in that it relates a lexical entry to its meaning.

Any theory of semantic memory has to deal with at least two questions:

1. What type of information is stored in semantic memory?and
2. How is information stored in semantic memory?

For instance, a long-standing question within the field has been the implementation of categories, specifically category membership relations.

Earlier models of semantic memory viewed categories as being comprised of determinate sets of exemplars with well-defined boundaries for membership; exemplars either were or were not members of a particular category (for review see Mervis and Rosch, 1981). However, empirical evidence quickly revealed the inadequacy of this view in accounting for many categorization phenomena. Not all category members are equivalent; instances clearly differ in the extent to which they are considered representative or 'typical' of some target category and these differences are reflected in categorization times. Thus, for example, subjects take longer to decide that 'penguin' is an exemplar of the category 'bird' than to decide that 'robin' is; this difference in reaction times (RTs) has been referred to variously as the typicality or relatedness effect. Whether the typicality effect is a consequence of the relation between an instance and its category, or due to characteristics of the individual stimuli per se is still at issue. On the one hand, several reports document the persistence of the typicality effect after instances are redistributed over categories (McFarland et al., 1974; Ashcraft, 1978). On the other hand, there have been several observations of context-induced changes of typicality, as for example, after a category is embedded in different reference sentences (Roth and Shoben, 1983) or combined with other categories (Smith and Osherson, 1984, 1987; Medin and Shoben, 1988).

Equally problematic for the classical view has been the finding that the same item may or may not be categorized as an instance of a category by the same subject on different occasions or by different subjects given the same set of stimuli; such inconsistencies are especially likely for atypical exemplars (Berlin and Kay, 1969; Labov, 1973; Kempton, 1978; McCloskey and Glucksberg, 1978; Cantor et al., 1980). Likewise, it is the atypical instances which are most sensitive to a variety of context manipulations. The question thus arises as to whether the same information subserves both typicality effects and category membership judgements or if the relevant information and processes are to some degree dissociable.

Feature comparison (Smith et al., 1974; McCloskey and Glucksberg, 1979) and network models (Collins and Loftus, 1975; Glass and Holyoak, 1975) offer very different answers to this question. According to feature comparison models, categories consist of a set of attributes (or features) which do not explicitly store class inclusion relations. Category membership relations, then, result from comparisons of the features (or properties) of the category and the exemplar, with verification time being a function of some criterion for membership acceptance. Network models, on the other hand, represent categories as nodes in a quasineural network with direct storage and access of inclusion relations built in. In this case, membership verification time is presumed to reflect the order of access to membership information; such information is obtained more quickly for typical than atypical exemplars, thereby yielding the typicality effect.

Clearly, these two classes of models differ in their accounts of how context influences category membership decisions and interacts with the typicality structure of categories. Whereas network models assume context dependent changes of accessibility to prestored information, feature comparison models require a reweighting of the importance of the features that are to be compared. In other words, in a network model, context has its influence during encoding while in a feature comparison model, context effects are a consequence of criterion changes in decision processes subsequent to encoding. Both network and feature comparison models assume that category membership decisions are based on a variety of sources of evidence. According to network models, however, membership verification and typicality rating are closely related, both being determined by the order of access to prestored memory structures. Feature comparison models, on the other hand, do not necessarily predict such a strong relationship between these two category-related phenomena.

To assess context effects on category organization, it would be useful to have different indices of category structure and/or of the various processes involved in category membership decisions. Recordings of event-related brain potentials (ERPs) afford a real-time examination of at least some of these mental events. For example, the N400 component of the ERP can be used as an index of sematic relatedness or association (Bentin et al., 1985; Kutas, 1985; Bentin, 1987; Boddy, 1986; Holcomb, 1988; Holcomb and Neville, 1990; van Petten and Kutas, 1990; Besson et al., 1992; Kounios and Holcomb, 1992, 1994). The N400 is a negative wave that peaks at ≈ 400 ms after word onset in young adults. The N400 to a word preceded by a semantically related or associated word is reduced in amplitude relative to one preceded by a semantically unrelated one, whether the preceding word actually occurs or is merely expected (but does not occur) as a function of the preceding context (Kutas and Hillyard, 1984; Kutas et al., 1984). The N400 component is elicited by all words to a degree that reflects the word's predictability within its context (Kutas and Hillyard, 1984; Kutas et al., 1984). Several studies have shown that the amplitude of the N400 elicited by major lexical items (open class words including nouns, verbs, adjectives and adverbs) is reduced in direct proportion to the buildup of semantic constraints (van Petten and Kutas, 1990). In a typical category-exemplar verification task, the ERP to the exemplar is characterized by an N400 component whose amplitude is sensitive to the degree of semantic relatedness between exemplar and the category name. Stuss et al. (1983) found that the N400 amplitude elicited by the exemplar in this task was larger for atypical than typical items.

The primary aim of the present study is to examine the effects of context on typicality and category membership decisions as indexed by both ERP and RT measures so as to shed light on the mechanisms of category flexibility. Context effects were investigated by recording ERPs and RTs in a category verification task modelled after McCloskey and Glucksberg (1979). Subjects were asked to decide whether or not the second word of a pair of words was an exemplar of the category named by the first. For the 'true' category–exemplar pairs, half the exemplars were typical and half were atypical. Context was manipulated by interspersing different types of 'false' category–exemplar pairs with the 'true' pairs; in two different experimental runs, the 'false' pairs consisted of a category name followed by a word that was either semantically related to it (e.g. fish–whale) or not (e.g. fish–window). In the McCloskey and Glucksberg (1979) experiments, this manipulation produced a significant context effect, namely, a general slowing of the RTs for membership verification of all pairings in the related vs. unrelated 'false' condition. Provided that this RT effect would occur in the present experiment, the question of interest was whether or not the N400 to the true exemplars in the two conditions would mirror this sensitivity to the change in context. If context effects are a consequence of the differential encoding in the two conditions, the N400 elicited by the true exemplars in the two conditions should reflect this difference in encoding. On the other hand, if context effects are a consequence of a criterion change during the decision making process, then the N400 in the two conditions should be impervious to changes in the list composition (i.e. to context); based on previous studies, criterion changes are more likely to be reflected in the later positive component (Paul and Sutton, 1972; Squires et al., 1975; Parasuraman and Beatty, 1980). This latter outcome would suggest that typicality and membership verification are, at least in part, dissociable.

2. Methods

Experiment I was conducted at the Neurological Clinic of Hannover Medical School, Germany, while Experiment II was performed at the Department of Neurosciences, University of California, San Diego, CA. As the two experiments were quite similar in most respects, the methodological descriptions are combined with differences noted as necessary.

2.1. Participants

Participants (Ps) were paid volunteers recruited from university student populations at the universities of Hannover and San Diego. All were healthy, righthanded individuals with normal or corrected-to-normal vision. The Ps for experiment I included 13 native speakers of German (age range 20–31, seven women); the Ps for experiment II included 12 native speakers of English (age range 19–32, five women).

2.2. Stimulus material

The stimuli were selected from 23 categories of the Battig and Montague (1969) category norms plus one additional category (emotion), as no equivalent standardized category norms exist in German. For experiment I, these stimuli were translated into German. (Whenever translation was not appropriate, instances agreed to be equivalent by the two German experimenters were chosen as replacements.) Two typical (from among the first eight words) and two atypical (from among the words with a rank of 15 or more) were chosen from each category listing.

Two lists of 192 word pairs each were constructed using these materials, one list of each of these experimental runs. The first word of any given pair was always one of the 24 category names. The second word was always a member of some category, although not necessarily the category named by the first word; half of these ($n = 96$) were exemplars of the category named by the first word (true) whereas the other half were not (false). Of the true category members, half ($n = 48$) were typical members of the category, while the other half ($n = 48$) were atypical members.

The two word lists differed primarily in the nature of the false word pairs; as these constituted the context manipulation, each list contained only one type of false word pairs. In the false but related run (related false condition) the second word of each pair, while not an instance of that category, either shared some important features with the true exemplars (e.g. fish–whale, bird–bat) or was related in a more general way to the superordinate (e.g. fish–water). For the false but unrelated run (unrelated false condition) the list of false related items was scrambled to produce disjoint false word pairs (e.g. fish–car). Ratings obtained from eight independent participants confirmed that the category–exemplar relatedness was higher for the related false than for the unrelated false exemplars.

The true (typical and atypical) exemplars were different in the two lists in order to avoid repetition effects. To be sure that the typicality effects were equivalent in the related and unrelated false condition, half of the Ps were presented with list 1 with the true exemplars of list 2 and list 2 with the true exemplars of list 1.

The original lists of 192 word pairs for each condition were presented in eight blocks of 24 pairs each; these 24 pairs comprised a random mixture of six typical, six atypical and 12 false pairs.

In experiment I stimuli were flashed for a duration of 200 ms in green capital letters at the center of a CRT. At a viewing distance of 90 cm, the letters subtended a visual angle of 0.45° vertically. Word lengths varied from 4 to 14 letters, corresponding between 1.3–3.7° in width. Each stimulus was replaced by a pattern mask consisting of vertical bars which subtended a rectangle $3.7 \times 0.45°$ (width $\times$

height). The letters in experiment II were presented in light grey. No mask was used. Finally, a fixation cross remained in the center of the CRT for the duration of the experiment.

2.3. Procedure

All Ps read all 384 word-pairs, 192 in each experimental run. Half of the Ps experienced eight blocks of the related false condition, while the other half experienced the unrelated false condition. Experiment I was conducted across two separate sessions on two different days. For experiment II, both conditions were presented in one single session. Between the blocks, Ps were given ample time to relax. For experiment I, each single trial began with the presentation of the mask followed 1200 ms later by the category name. After 200 ms the category name was replaced by the mask. The onset asynchrony between the first and the second word of a pair was 800 ms. The second word stayed on the screen for 200 ms and was again replaced by the mask which stayed on for 2000 ms. The trial-to-trial interval was 5 s. The timing of the stimuli for experiment II was identical. Following the second word, Ps had to press one of two buttons located beneath each index finger as quickly and accurately as possible. The one button indicated that a true category exemplar occurred, while the other button was to indicate a false category member. The assignment of buttons was reversed for half of the Ps.

2.4. EEG-recording

In experiment I, ERPs were recorded from the scalp using silver–silver chloride electrodes placed at midfrontal, central, and parietal sites (Fz, Cz, Pz, International 10-20 system) as well as two lateral central sites placed halfway between Cz and the auditory canal (C3′, C4′). A reference electrode was placed on the right mastoid. Eye movements were recorded via electrodes placed on the left external canthus and the right inferior orbital ridge. Bioelectric signals were amplified with a VanGogh 50 000 system (time constant 10 s, low-pass 64 Hz) and digitized on-line with 8 ms resolution. Before averaging automatic artefact rejection algorithms were used to eliminate trials with blinks, eye movement, EMG and amplifier blocking.

In experiment II, ERPs were recorded from midline electrodes (Fz, Cz, Pz, International 10-20 system) and five pairs of lateral electrodes (F3/4, C3/4, T3/4, T5/6, O1/2, International 10-20 system) using tin electrodes mounted in an electrode cap. After digitization, data were algebraically rereferenced to the average of the two mastoids. Signals were amplified with a time constant of 8 s and a low-pass filter setting of 60 Hz with Grass amplifiers and digitized with 4 ms resolution.

2.5. Data analysis

Reaction times (RTs) and error rates were assessed. RTs exceeding 1700 ms after onset of the second word were excluded from further analysis. ERP waveforms were quantified exclusively by area measures. For the area measure of the ERPs to the second words the average voltage in the interval 100 ms prior to the onset of the second word served as baseline. Because the time course of the expected electrophysiological context effects was not known in advance, successive overlapping time windows were used (100–300 ms, 200–400 ms, ..., 600–800 ms). For the ERPs in the interval between the two words similar area measures were taken. ERP measures and behavioral data were analyzed using repeated measures analyses of variance (ANOVA) adjusting with the Greenhouse-Geisser epsilon coefficient when appropriate. ANOVAs were conducted with Typicality (typical vs. atypical), Condition (related false vs. unrelated false) and Electrode Site as factors and participants as repeated measures. Separate ANOVAs were calculated for the midline electrodes (with Fz, Cz, Pz as levels of the Electrode Site factor) and the lateral electrode pairs (left and right hemisphere as levels). The actual data from the false words were not included in the analyses since the different classes of false words were used to constitute the context conditions rather than influenced by them.

3. Results

3.1. Reaction times

RTs for the different stimulus classes are given in Table 1. The ANOVAs revealed significant main effects of Typicality (I: $P < 0.001$; II: $P < 0.001$) and Condition (I: $P < 0.007$; II: $P < 0.007$) in both experiments. The Typicality × Condition interaction was highly significant in I ($P < 0.001$) but not in II.

Table 1
RTs for the different stimulus classes

	Reaction times (ms)		Error rates (%)	
	Exp: 1		Exp: 2	
	RT	%	RT	%
Related false cond.				
Typical exemplar	796	2.1	761	5.1
Atypical exemplar	1039	5.3	902	16.8
False word	1079	4.4	982	18.74
Unrelated false cond.				
Typical exemplar	759	2.3	675	2.5
Atypical exemplar	908	4.9	814	12.2
False word	860	1.8	757	3.1

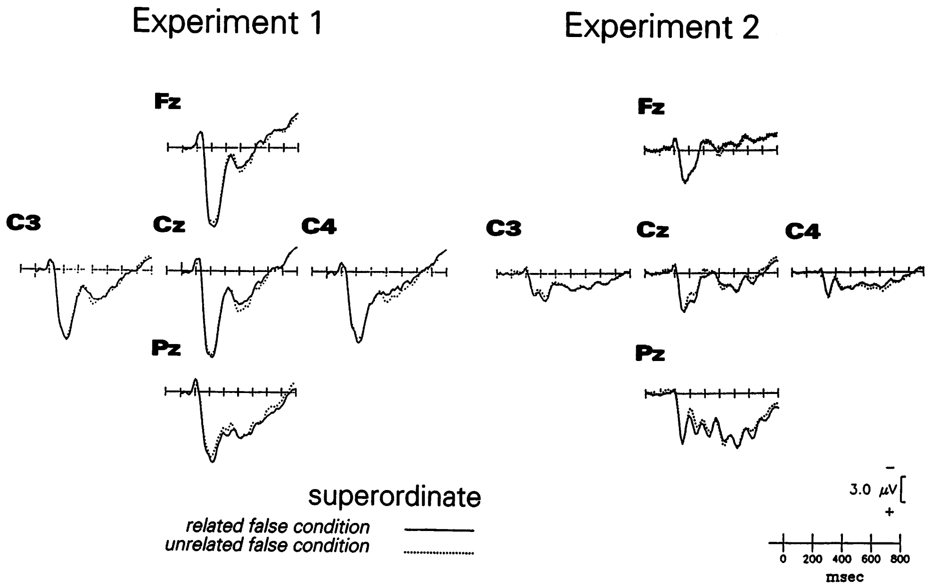

Fig. 1. Grand average ERPs to the first (superordinate) word of a pair in the related false and unrelated false condition in experiments I and II. Negative is plotted up in this and all subsequent figures.

3.2. Error rates

The percent error rate is provided in Table 1. Although statistical comparison between the two experiments was not carried out, it seemed that atypical exemplars were associated with a much higher error rate in II. There was a main effect of Typicality in both experiments (I: $P < 0.005$; II: $P < 0.001$), reflecting the higher error rates for atypical exemplars. A main effect for the Condition factor was obtained only in II ($P < 0.025$). In neither of the experiments was the Typicality × Condition interaction significant.

3.3. ERPs to the first words

The ERPs to the first words from both experiments are depicted in Fig. 1. Both studies revealed similar waveforms, characterized by an initial negative peak around 100 ms (N1) followed by a positivity peaking at about 180 ms (P2). Statistical analyses corroborated our observation that there was no significant difference in the first word's ERPs between the two context conditions in either experiments.

3.4. ERPs to the second words

3.4.1. Waveform morphology

The grand average ERPs in the unrelated false condition from both experiments are depicted in Fig. 2. All the waveforms showed a similar pattern with a sharp positivity peaking at about 180 ms followed by a broad negativity between 300 and 500 ms, the amplitude of which varied as a function of the stimulus type. False exemplars elicited the largest negativity, typical exemplars the smallest negativity. For both experiments this negativity was followed by a late positivity.

ERPs from the related false condition in both experiments are shown in Fig. 3. The principal characteristics of the ERPs are the same as those in the unrelated false condition. The main difference between the ERPs in the two conditions is in the response to the false items; the negativity to these stimuli in the 300–600 ms range is much smaller in the related than in the unrelated false condition.

To help visualize the differences between the two context conditions, we computed difference waves for each of the stimulus types. The ERPs to the typical, atypical and false exemplars in the related false condition were subtracted from the corresponding ERPs in the unrelated false condition (Fig. 4). As can be seen, the results were virtually identical in experiment I and II. While for the false items a negative wave with a peak latency of about 400 ms was present, the typical and atypical exemplars showed a positive difference in the 400–600 ms range.

3.4.2. Statistical results

3.4.2.1. Experiment I. A main effect of Typicality was obtained for the time window 200–400 ms (midline electrodes: $P < 0.001$; C3/4: $P < 0.04$) and the time window 300–500 ms (midline electrodes: $P < 0.01$; C3/4: $P < 0.007$). Neither a main effect of Condition nor a Typicality × Condition interaction was seen.

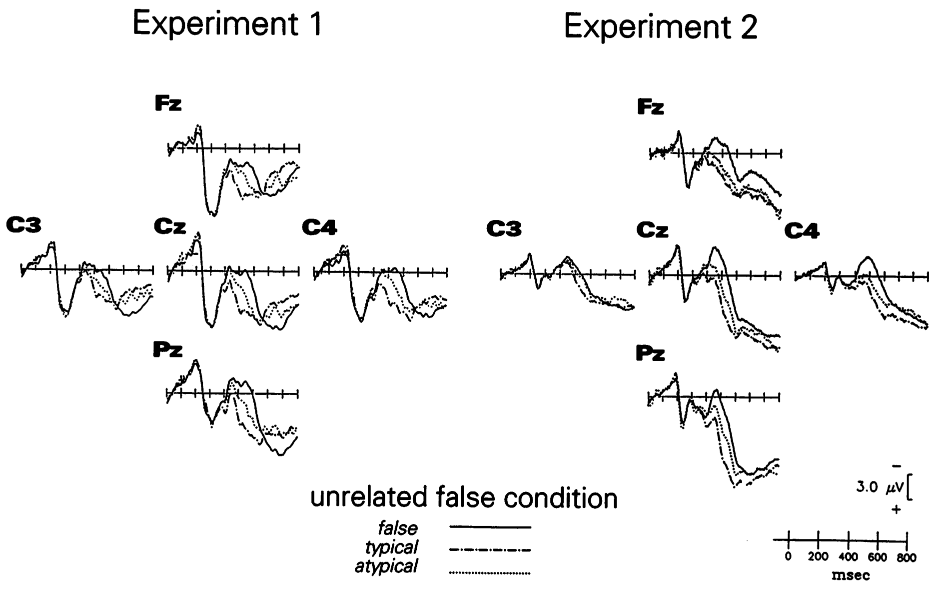

Fig. 2. Grand average ERPs to the second word of a pair in the unrelated false condition in experiments I and II. Note that in this and the subsequent figures the ERPs to the false pairs are presented to show the efficacy of the experimental manipulation but were not included in the statistical analyses.

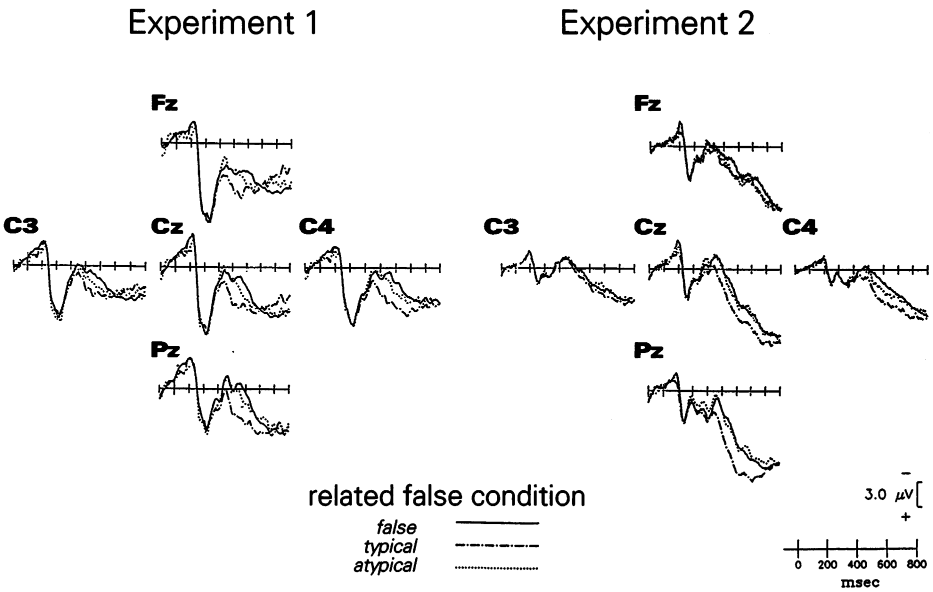

Fig. 3. Grand average ERPs to the second word of a pair in the related false condition in experiments I and II.

3.4.2.2. Experiment II. A main effect of Typicality was obtained for most scalp sites for the time window 200–400 ms (midline electrodes: $P < 0.003$; T3/4: $P < 0.02$; T5/6: $P < 0.03$; 01/2: $P < 0.01$). A similar result was seen for the time window 300–500 ms (midline electrodes: $P < 0.001$; C3/4: $P < 0.002$; T3/4: $P < 0.001$; T5/6: $P < 0.002$; 01/2: $P < 0.003$), for the time window 400–600 ms (F3/4: not significant; all others $P < 0.001$), for the time window 500–700 ms (F3/4: not significant; all others: $P < 0.05$), and for the time window 600–800 ms (midline electrodes and F3/4: not significant; all others: $P < 0.03$). A main effect of Condition was observed for the 400–600 ms interval (T3/4: $P < 0.02$; T5/6: $P < 0.005$), and for the 500–700 ms interval (T5/6: $P < 0.05$). No Typicality × Condition interaction was seen.

4. Discussion

The aim of the present experiments was to determine whether it is theoretically or empirically fruitful to view the structure internal to a category as a natural extension of relations (or structure) between categories. More specifically, the experiments used the list composition effect (context manipulation) to examine the extent to which decisions about category membership (between category structure) are dependent on or dissociable from the typicality structure (within category structure) of categories. To this end, behavioral and ERP measurements were obtained in a category verification task that incorporated a context manipulation that was previously shown to affect membership verification times (McCloskey and Glucksberg, 1979).

In its essential aspects, the behavioral data from both the German and American English versions of the present study replicated those of McCloskey and Glucksberg (1979). That is, there was a significant main effect of Typicality and Condition on verification times (experiment I and II) and error rates (experiment II). In addition, there was the expected Typicality × Condition interaction (experiment I) indicating that the context manipulation affected the verification times for atypical exemplars more than for typical ones. By contrast, the N400 component of the ERP to the true exemplars was sensitive only to typicality, not to the context (list) manipulation. Of course, the N400 to the false items was affected by relatedness: it was larger for unrelated than related false items. This was expected. The new finding was that the N400 to true items was unaffected by the nature of the false items which surrounded them in the list.

In both experiments, the ERPs elicited by the second word were characterized by a broad negativity in the 300–500 ms time range followed by a late positivity. The morphology and scalp distribution of this late negativity corresponded to that of the N400 component observed in response to both congruent and incongruent open class words presented in word lists, word pairs, and sentences (e.g. Kutas and Hillyard, 1980, 1983, 1984, 1989; van Petten and Kutas, 1990). In both experiments

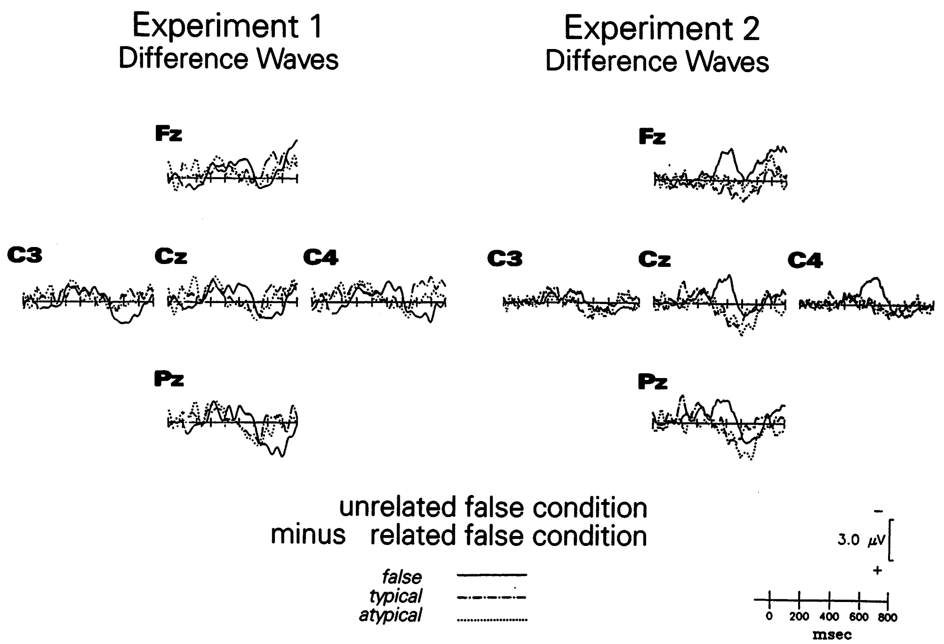

Fig. 4. Difference waves obtained by subtracting the ERPs to the typical, atypical and false exemplars in the related false condition from the ERPs to the corresponding stimuli in the unrelated false condition in experiments I and II.

of the present study, N400 amplitude was modulated by the semantic relation or association between the category name and the exemplar. The N400s elicited by typical instances were smaller than those by atypical instances; this N400 typicality effect was identical in the two context conditions showing an independence from context effects. Furthermore, in line with previous results, we found that N400 amplitudes varied primarily with the nature of the semantic relation or association between the category name and the eliciting word and not with the truth or falsity of the category verification judgement (see also Kounios and Holcomb, 1992). For example, Fischler et al. (1983) observed that in a sentence verification task when sentences contained negative verbs, an N400 was observed for both true (e.g. 'A robin is not a truck') and false (e.g. 'A robin is not a bird') statements, suggesting that sentence-level truth was not influencing the N400. In a subsequent study, however, Fischler et al. (1985) found that both lexical and sentence-level processes modulate N400 amplitude. In that experiment, subjects verified three basic types of sentences: (1) true statements with strong associations between context and final target word (TA; e.g. 'Tuna are fish'), (2) false statements with equally strong lexical associations (FA; e.g. 'Oysters are clams') and (3) false statements in which subjects and predicates were lexically unrelated (FU, e.g. 'Sardines are lips'). While N400 amplitudes to target words that were strongly associated with the context were reduced relative to unrelated targets whether they were true or false, the size of the reduction was greater for the true than false statements (Fischler, 1990). The ERPs to the associated targets in the false statements also differed from those in the true statements in the presence of a late positivity following the N400.

In the present experiments, we too found that the N400 to false items was smaller if an item was somehow related or associated with the category name than if it was not, although in neither case was the item a member of the designated category. The sensitivity of the N400 elicited by exemplars to their featural overlap, property overlap, similarity or associativeness with the category name regardless of the truth or falsity of the classification argues for a continuity between the typicality or relatedness effect and the false-relatedness effect. While some authors have implied a common underlying mechanism by referring to these two effects jointly as semantic relatedness or semantic distance effects, our data provide the first strong empirical support for this view.

While the N400 component of the ERP to the true exemplars was impervious to the list manipulation and thus at least partly independent of verification times, a late positive component (LPC) in the 400–600 ms range was not. Overall, our results underscore the previously shown dissociation between RT and N400 changes and suggest that the prolongation of membership verification time, known as the list composition effect, is a relatively late phenomenon, probably related to a change in criterion for the decision about category membership. These ERP and behavioral data provide strong evidence that category membership decision (as reflected in LPC and verification latencies) and semantic relatedness or typicality of an exemplar (as evidenced by N400 modulation) may vary independently from each other and, therefore, likely index the activity of different mechanisms of semantic memory.

One might argue that any interpretation of the results of these experiments is limited by the fact that the context manipulation may have induced very different processing modes in the two context conditions. According to this argument, the presence of the false related category–exemplar pairs might have forced subjects to process all the pairs in that condition more deeply (in terms of the levels-of-processing framework). If this were the case, then the differences in membership verification times as a function of the context manipulation could be attributed to differences in the amount of available information including membership relations. Clearly, the strength of this argument rests on whether or not the N400 amplitude varies with depth of semantic processing. And, indeed, as several experiments have demonstrated, the N400 is larger when attention is directed at the semantic relations between the stimuli (Kutas, 1985; Holcomb, 1988; Besson et al., 1992; Bentin et al., 1993). For example, Holcomb found that the N400 semantic priming effect obtained in a lexical decision task was significantly larger when 50% of the prime-target pairs were related than if only one-eighth of them were related. In the present experiments, the proportion of related category–exemplar pairs in the two context conditions was changed implicitly by interleaving either unrelated false or related false word pairs into the list of true pairs. Insofar as the differences in membership verification time in the present study are due to variations in the depth of semantic analysis of the exemplars and, consequently, in the amount of available information, these should be manifest in changes in N400 amplitude across the conditions. And, yet the N400s to true exemplars were equivalent despite the presence of the different types of false pairs. Consistent with this interpretation is the fact that the ERPs elicited by the category names in the two context conditions also did not differ suggesting that the mode of processing was not substantially different.

The results of this study also bear upon the question of how categories are organized in semantic memory. On the whole, the ERP data are more consistent with feature comparison than network models as outlined in the introduction. The decision about category membership appears to happen some time after a feature comparison stage that is performed similarly for both category members and non-members; here, features refer to sets of characteristics rather than defining or necessary features. Initially, the response to atypical category members is more similar to the response to non-members than it is to that of typical members. Moreover, the N400 response to category members is not altered by the presence of related false pairs, rather the latency of the LPC is delayed. So, category membership is not a simple function of the number of activated features but rather governed by a probabilistic decision function. Thus, as previously noted, the dissociation between category membership and typicality supports models of semantic memory that place emphasis on computing processes of category membership on the basis of a context dependent decision criterion. Although the data of the present study do not allow a more detailed analysis of these processes, what is of interest here is the implication that membership information is evidently not wholly and invariantly prestored in semantic memory.

Regarding the influence of context on category organization, Anderson and Shifrin (1980) take the extreme view that the meaning of a category is totally dependent upon context. Similarly, Roth and Shoben (1983) suggest that typicality in the absence of context is not a good predictor of category comprehension in constrained contexts. The present data partly support this view. On the other hand, the fact that we find a consistent N400 gradient to typical and atypical exemplars in both context conditions suggests that semantic relatedness or semantic distance is a relatively stable characteristic of categorial information.

Acknowledgements

This work was supported by a DFG Grant (INK15/A1(TP B1)) to H.J. Heinze. M. Kutas and some of the research herein was supported by an RSDA from NIMH (MH00322) and grants from NICHD (HD22614) and NIA (AG08313). The authors wish to thank C. van Petten for helpful comments on the manuscript and S. Luck and M. Scholz for technical assistance. The support of S.A. Hillyard's laboratory is gratefully acknowledged.

References

Anderson, R.C., Shifrin, Z., 1980. The meaning of words in context. In: Spiro, R.J., Bruce, B.C., Brewer, W.F. (Eds.), Theoretical Issues in Reading Comprehension. Erlbaum, Hillsdale, NJ, pp. 331–348.

Ashcraft, M.H., 1978. Property norms for typical and atypical items from 17 categories: a description and discussion. Mem. Cognit. 6, 227–232.

Battig, W.F., Montague, W.E., 1969. Category norms of verbal items in 56 categories: A replication and extension of the Connecticut category norms. J. Exp. Psychol. 80 (3,2), 1–46.

Bentin, S., 1987. Event-related potentials, semantic processes, and expectancy factors in word recognition. Brain Lang. 31, 308–327.

Bentin, S., McCarthy, G., Wood, C.C., 1985. Event-related potentials associated with semantic priming. Electroencephalogr. Clin. Neurophysiol. 60, 343–355.

Bentin, S., Kutas, M., Hileyard, S.A., 1993. Electrophysiological evidence for task effects on semantic priming in auditory word processing. Psychophysiology 30 (2), 161–169.

Besson, M., Fischler, I., Boaz, T., Raney, G., 1992. Effects of automatic associative activation on explicit and implicit memory tests. J. Exp. Psychol. Learn. Mem. Cognit. 18 (1), 89–105.

Berlin, B., Kay, P., 1969. Basic Color Terms: Their Universality and Evolution. Univeristy of California Press, Berkeley, CA.

Boddy, J., 1986. Event-related potentials in chronometric analysis of primed word recognition with different stimulus onset asynchronies. Psychophysiology 23, 232–245.

Cantor, N., Smith, E.E., French, R.D., Mezzich, J., 1980. Psychiatric diagnosis as prototype organization. J. Abnorm. Psychol. 89, 181–193.

Collins, A.J., Loftus, E.F., 1975. A spreading-activation theory of semantic processing. Psychol. Rev. 82, 407–428.

Fischler, I., 1990. Comprehending language with event-related potentials. In: Rohrbaugh, J.W., Parasuraman, R., Johnson, R. (Eds.), Event-Related Brain Potentials: Basic Issues and Applications. Oxford University Press, New York, pp. 165–177.

Fischler, I., Bloom, P.A., Childers, D.G., Roucos, S.E., Perry, N.W., 1983. Brain potentials related to stages of sentence verification. Psychophysiology 20, 400–409.

Fischler, I., Boaz, T., Childers, D.G., Perry, N.W., 1985. Lexical and propositional components of priming during sentence comprehension. Psychophysiology 22, 576 (Abstract).

Glass, A.L., Holyoak, K.J., 1975. Alternative conceptions of semantic memory. Cognition 3, 313–339.

Holcomb, P.J., 1988. Automatic and attentional processing: an event-related brain potential analysis of semantic priming. Brain Lang. 35, 66–85.

Holcomb, P.J., Neville, H.J., 1990. Auditory and visual semantic priming in lexical decisions: a comparison using event-related brain potentials. Lang. Cognit. Processes 5, 281–312.

Kempton, W., 1978. Category grading and taxonomic relations: a mug is a sort of a cup. Am. Ethnol. 5, 44–65.

Kounios, J., Holcomb, P.J., 1992. Structure and process in semantic memory: Evidence from event-related brain potentials and reaction times. J. Exp. Psychol. Gen. 121 (4), 459–479.

Kounios, J., Holcomb, P.J., 1994. Concreteness effects in semantic processing: ERP evidence supporting dual-coding theory. J. Exp. Psychol. Learn. Mem. Cognit. 20 (4), 804–823.

Kutas, M., 1985. ERP comparisons of the effects of single word and sentence contexts on word processing. Psychophysiology 22, 575–576, (Abstract).

Kutas, M., Hillyard, S.A., 1980. Reading senseless sentences: brain potentials reflect semantic incongruity. Science 207, 203–205.

Kutas, M., Hillyard, S.A., 1983. Event-related brain potentials to grammatical errors and semantic anomalies. Mem. Cognit. 11, 539–550.

Kutas, M., Hillyard, S.A., 1984. Brain potentials during reading reflect word expectancy and semantic association. Nature 307, 161–163.

Kutas, M., Hillyard, S.A., 1989. An electrophysiological probe of incidental semantic association. J. Cognit. Neurosci. 1, 38–49.

Kutas, M., Lindamood, T., Hillyard, S.A., 1984. Word expectancy and event-related brain potentials during sentence processing. In: Kornblum, S., Requin, J. (Eds.), Preparatory States and Processes. Erlbaum, Hillsdale, NJ, pp. 217–238.

Labov, W., 1973. The boundaries of words and their meanings. In: Bailey, C.J., Shuy, R. (Eds.), New Way of Analyzing Variation in English. Georgetown University Press, Washington, pp. 340–373.

Mccloskey, M.E., Glucksberg, S., 1978. Natural categories: well defined or fuzzy sets? Mem. Cognit. 6, 462–472.

Mccloskey, M.E., Glucksberg, S., 1979. Decision processes in verifying category membership statements: implications for models of semantic memory. Cognit. Psychol. 11, 1–37.

McFarland, C.E., Kellas, G., Klueger, K., Juola, J.F., 1974. Category similarity, instance dominance, and categorization time. J. Verbal Learn. Verbal Behav. 13, 698–708.

Medin, D.L., Shoben, E.J., 1988. Context and structure in conceptual combination. Cognit. Psychol. 20, 158–190.

Mervis, C.B., Rosch, E., 1981. Categorization of natural objects. Annu. Rev. Psychol. 32, 89–115.

Parasuraman, R., Beatty, J., 1980. Brain events underlying detection an recognition of weak sensory signals. Science 210, 80–83.

Paul, D.D., Sutton, S., 1972. Evoked potential correlates of response criteria in auditory signal detection. Science 177, 362–364.

Roth, E.M., Shoben, E.J., 1983. The effect of context on the structure of categories. Cognit. Psychol. 15, 346–378.

Smith, E.E., Osherson, D.N., 1984. Conceptual combination with prototype concepts. Cognit. Sci. 8, 337–361.

Smith, E.E., Osherson, D.N., 1987. Compositionality and typicality. In: Schifter, S., Steele, S. (Eds.), The 2nd Arizona Colloquium on Cognitive Science. University of Tucson Press, Tucson, AZ.

Smith, E.E., Rips, L.J., Shoben, E.J., 1974. Semantic memory and psychological semantics. In: Bower, G.H. (Ed.), The Psychology of Learning and Motivation 8. Academic Press, New York, pp. 1–45.

Squires, K.C., Squires, N.K., Hillyard, S.A., 1975. Decision-related cortical evoked potentials during an auditory signal detection task with cued observation intervals. J. Exp. Psychol. Hum. Percept. Perform. 1, 268–279.

Stuss, D.T., Sarazin, F.F., Leech, E.E., Picton, T.W., 1983. Event-related potentials during naming and mental rotation. Electroencephalogr. Clin. Neurophysiol. 56, 133–146.

van Petten, C., Kutas, M., 1990. Interactions between sentence context and word frequency in event-related brain potentials. Mem. Cognit. 18, 380–393.