Category and Semantic Distance Modulate the Effects of Expectation on Memory

Authors: Dai Jiaojian, Sun Mingze, Wang Dongfang, Xinrui Mao, Guo Chunyan, Mao Xinrui, Guo Chunyan

Date: 2025-07-11T16:26:13+00:00

Abstract

Humans often need to make predictions based on current situations to guide subsequent behavior. However, researchers currently debate vigorously regarding whether predictable or unpredictable items yield better memory performance. In this study (Experiment 1: item recognition; Experiment 2: associative recognition), we manipulated expectancy through category and semantic distance to examine memory performance under different levels of expectancy. Both experiments consistently showed that during the learning phase, participants achieved higher accuracy and shorter reaction times when processing items that conformed to category rules. Experiment 2 found that closer semantic distance produced more positive P600 amplitudes (an ERP index of semantic integration). In both experiments, multivariate pattern analysis indicated significant differences in neural representations across conditions. Both experiments found that as category rules were violated and semantic distance increased, the amplitude of N400 (an ERP index of predictability) during the learning phase became more negative, and memory performance during the test phase became worse. Additionally, N400 during the learning phase showed a significant positive correlation with memory performance during the test phase. These results suggest that category and semantic distance modulate the effect of expectancy on memory through different mechanisms: category influences memory by modulating encoding load, while semantic distance influences memory by modulating semantic integration. These findings provide a new perspective for resolving the debate on memory performance for predictable versus unpredictable items.

Full Text

Category and Semantic Distance Modulate the Impact of Prediction on Memory

DAI Jiaojian[1], **SUN Mingze**[3], **WANG Dongfang**[3], **MAO Xinrui**[2], **GUO Chunyan**[1]

[1] College of Psychology, Beijing Key Lab of Learning and Cognition, Capital Normal University, Beijing 100048, China
[2] College of Elementary Education, Capital Normal University, Beijing 100048, China
[3] School of Psychology and Center for Studies of Psychological Application, South China Normal University, Guangzhou 510631, China

Abstract

Humans constantly generate predictions based on current contexts to guide subsequent behavior. However, researchers remain divided over whether predictable or unpredictable items yield superior memory performance. In this study (Experiment 1: item recognition; Experiment 2: associative recognition), we manipulated prediction through category and semantic distance to examine memory performance across different levels of predictability. Both experiments consistently demonstrated that during the study phase, participants exhibited higher accuracy and shorter reaction times when processing items that conformed to category rules. Experiment 2 revealed that shorter semantic distances elicited more positive P600 amplitudes (an ERP index of semantic integration). Multivariate pattern analysis in both experiments showed significant differences in neural representations across conditions. Both experiments found that as category rule violations increased and semantic distance grew, N400 amplitudes (an ERP index of predictability) became more negative during encoding, while memory performance during the test phase declined. Moreover, study-phase N400 amplitudes showed significant positive correlations with test-phase memory performance. These results indicate that category and semantic distance modulate prediction's impact on memory through distinct mechanisms: category influences memory by modulating encoding burden, whereas semantic distance influences memory through semantic integration. These findings provide a novel perspective for resolving the debate surrounding memory performance for predictable versus unpredictable items.

Keywords: category rules, semantic distance, prediction, recognition, EEG
Classification Number: B842

Introduction

Humans frequently need to make predictions based on current situations to guide subsequent behavior (Bar, 2007, 2009). However, humans are not born prophets; we must continuously learn from past predictive experiences to improve our predictive abilities. Understanding memory performance under different predictive conditions is therefore crucial. Currently, researchers debate whether prediction enhances or impairs memory performance. In this study, we used EEG to investigate memory across different levels of predictability.

Some studies have found that when learning materials can be predicted based on semantic context, memory performance improves (Craik & Tulving, 1975; Kutas, 1993; Schulman, 1974; Silcox et al., 2023). For example, Schulman (1974) manipulated prediction through categories to examine memory for predictable versus unpredictable items. Participants saw expected events (e.g., "is a corkscrew an opener?") and unexpected events (e.g., "is a corkscrew a scholar?"), with better memory observed for expected items. Subsequently, Silcox et al. (2023) manipulated both category and semantic distance to examine the relationship between the ERP index of prediction (N400; DeLong et al., 2005; Federmeier et al., 2007; Kutas & Hillyard, 1984) and memory performance. During the study phase, they applied transcranial magnetic stimulation (TMS) to the left inferior frontal gyrus (LIFC), a hub of the speech production network (Bonhage et al., 2015; Giglio et al., 2022; Menenti et al., 2011; Silbert et al., 2014), with right inferior frontal gyrus (RIFC) stimulation as a control condition. No stimulation was applied during memory testing. Results showed that regardless of stimulation site, N400 amplitudes became progressively more negative across three conditions (high typicality: flower–rose, low typicality: flower–poppy, and incongruent: flower–opera), indicating decreasing target word predictability. With RIFC stimulation, memory performance declined across the three conditions. With LIFC stimulation, high-typicality memory performance dropped to the low-typicality level, yet both remained significantly better than the incongruent condition. These findings demonstrate that both category and semantic distance influence memory, though the mechanisms remain unclear.

Conversely, other studies have found better memory for unexpected learning materials (Frank et al., 2020; Kafkas & Montaldi, 2015; Rajaram, 1998; Nyberg, 2005). For instance, Rajaram (1998) found that "bank" showed better memory performance in an unexpected context ("river") than in an expected one ("money"). Kuperberg et al. (2003) found that anomalous items (e.g., "eggs" in "For breakfast the eggs would only eat toast and jam") elicited a more positive P600, reflecting the brain's effort to repair or integrate anomalous items. Similarly, Schotter et al. (2023) found that anomalous items elicited a more positive LPC than expected items, indicating greater effort to integrate anomalous target words into semantic context. These studies suggest that anomalous (or unpredictable) items elicit more positive ERP activity following the N400 time window, reflecting integration processes. However, Kuperberg et al. (2003)

and Schotter et al. (2023) did not examine the relationship between P600 and memory performance.

The present study aimed to compare memory performance across different predictability levels while measuring the neural processes of memory encoding. We manipulated prediction through category and semantic distance, establishing three conditions: C+S+ (within-category and near semantic distance; e.g., furniture: sofa, as sofa belongs to furniture), where participants could effectively predict the target word; C-S+ (out-of-category but near semantic distance; e.g., furniture: decoration), where the target word did not belong to the category but was semantically close; and C-S- (out-of-category and far semantic distance; e.g., furniture: phase), where the target word neither belonged to the category nor was semantically close. Our design differs from previous studies (Federmeier et al., 2010; Ryskin et al., 2020; Silcox et al., 2023). Specifically, Silcox et al. (2023) used "flower" as context, with "rose" in the high-expectation condition, "poppy" in the low-expectation condition, and "opera" in the incongruent condition. Their high-expectation condition resembles our C+S+, their incongruent condition resembles our C-S-, but their low-expectation condition (poppy as a low-typicality flower member) differs from our C-S+ (decoration is not furniture but semantically close). Our design allows us to investigate prediction's impact on memory while strictly controlling category and semantic distance.

Previous studies comparing high-typicality, low-typicality, and incongruent conditions found graded N400 effects. However, these effects could reflect decreasing category typicality, increasing semantic distance, or a combination of both, making it impossible to isolate the unique mechanisms of category and semantic distance. By comparing C+S+ and C-S+, we can examine category's influence while controlling semantic distance. By comparing C-S+ and C-S-, we can examine semantic distance's influence while controlling category. This clarifies how different aspects of prediction affect memory.

Based on previous research, we hypothesized: (1) N400 amplitudes would become progressively more negative across C+S+, C-S+, and C-S- conditions, reflecting decreasing target word predictability; (2) P600 amplitudes would be most positive in the C-S- condition, as these word pairs violate category rules and have far semantic distance, requiring the greatest integration effort. To provide convergent evidence, we complemented univariate analyses with machine learning, an effective method for distinguishing neural representations across conditions (Haxby et al., 2001) that trains classifiers on partial data and tests on remaining data (Treder, 2020). We also conducted temporal generalization analysis—training classifiers at one time point and testing across all time points—to examine the stability of neural representations across conditions (King & Dehaene, 2014).

Experiment 1

Participants   A power analysis based on effect sizes from Elmer et al. (2022; $p^2 = 0.114$) indicated that 22 participants were needed for 95% statistical power ($\alpha = 0.05$) to detect prediction effects on memory. Experiment 1 recruited 28 undergraduate and graduate students aged 18–25 years (mean age = 22 years; 18 females, 10 males; native Chinese speakers). The study was approved by the Psychological Research Ethics Committee. Participants provided informed consent and received compensation. All participants were right-handed with normal or corrected-to-normal vision and no history of psychiatric disorders.

Materials   Study-phase materials consisted of word pairs (e.g., "fruit: banana::furniture: sofa"; Zhao et al., 2012). In each pair, the second word belonged to the first word's category (e.g., "banana" belongs to "fruit," "sofa" belongs to "furniture"). The first word pair ("fruit: banana") provided semantic context, indicating that participants should evaluate the relationship in the subsequent pair. Word pairs with consistent relationships served as C+S+ trials.

We also included C-S+ and C-S- conditions. In C-S+, we used pairs like "fruit: banana::furniture: decoration," where "decoration" does not belong to "furniture" but is semantically close. In C-S-, we used pairs like "fruit: banana::furniture: phase," where "phase" neither belongs to "furniture" nor is semantically close.

Two experimenters initially selected materials from a noun database. Twelve participants who did not participate in the formal experiment rated the materials on acceptability, semantic distance, and familiarity. Acceptability assessed the degree to which the second word belonged to the first (1 = completely unacceptable, 9 = completely acceptable). To ensure high acceptability, C+S+ materials received ratings above 7.0, while C-S+ and C-S- materials received ratings below 3.0. Semantic distance ratings were M(C+S+) = 8.10, SE = 0.13; M(C-S+) = 7.87, SE = 0.16; M(C-S-) = 1.23, SE = 0.03. C+S+ and C-S+ did not differ significantly ($t(11) = 2.31$, p = 0.123, Cohen's d = 0.67) but both were significantly higher than C-S- (C+S+ vs. C-S-: $t(11) = 51.80$, p < 0.001, Cohen's d = 14.95; C-S+ vs. C-S-: $t(11) = 42.36$, p < 0.001, Cohen's d = 12.23). Familiarity ratings did not differ across conditions ($F(2, 22) = 0.93$, p = 0.411, $p^2 = 0.08$), preventing familiarity differences from confounding memory performance. The final material set included 336 word pairs (112 per condition, excluding filler pairs used during study).

Procedure   Participants completed the experiment in a sound-attenuated, well-lit room. They were instructed to maintain a comfortable posture and remain still. The computer screen was positioned 100 cm from participants. All stimuli appeared in white on a black background. The experiment consisted of study and test phases.

During the study phase, stimulus presentation followed Zhao et al. (2012). Par-

ticipants first completed a practice block identical to the formal experiment. Each trial began with a central fixation point (1400–1600 ms), followed by Word Pair 1 (e.g., "fruit: banana"; 6.8° × 1.3°) for 1000 ms. Next, the prime word (the first word of Pair 2, e.g., "furniture:"; 3.1° × 1.3°) appeared for 1500 ms. Finally, the target word (the second word of Pair 2, e.g., "sofa"; 2.6° × 1.3°) appeared for 2000 ms. A jitter of 300–500 ms separated adjacent stimuli. When the target word appeared, participants responded by pressing keys: "1" if the target belonged to the prime, "2" if it did not belong but was semantically close, and "3" if it neither belonged nor was semantically close. Response mappings were counterbalanced across participants, and trials from different conditions were randomly intermixed. To minimize primacy and recency effects, each study block began and ended with two filler trials excluded from analysis. Participants completed six 3-minute study blocks, each containing 8 C+S+, 8 C-S+, and 8 C-S- trials. After all study blocks, participants completed a surprise memory test (see Figure 1 [Figure 1: see original paper]A).

In the test phase, participants first completed a practice block. Each trial began with a central fixation point (900–1300 ms), followed by a target word from the study phase (old) or a new word not previously seen. Participants rated each word on a 6-point scale (1 = definitely not studied, 2 = probably not studied, 3 = guess not studied, 4 = guess studied, 5 = probably studied, 6 = definitely studied). "Definitely," "probably," and "guess" indicated different confidence levels. Old and new words appeared in a 1:1 ratio with counterbalanced response mappings. Participants completed six 3-minute blocks, each containing 24 old and 24 new words (see Figure 1A). Words in Test Block 1 came from Study Block 1, and so on.

EEG Recording and Analysis   We recorded EEG using a Neuroscan ESI-64 system with an Ag/AgCl electrode cap based on the international 10-20 extended electrode system. The left mastoid served as online reference, the right mastoid as recording electrode, with offline re-referencing to the average of both mastoids. Vertical EOG was recorded from electrodes above and below the left eye, horizontal EOG from electrodes at the outer canthi of both eyes, and the ground electrode was positioned between Fpz and Fz. The sampling rate was 500 Hz, with electrode impedance maintained below 5 kΩ.

We preprocessed EEG data using EEGLAB (Delorme & Makeig, 2004) in MAT-LAB. Data were bandpass filtered from 0.1 Hz to 40 Hz and re-referenced to the average mastoids. Independent component analysis (ICA) identified blink and saccade components, which were manually removed. We segmented data from 200 ms pre-target (baseline) to 1000 ms post-target, rejecting trials with amplitudes exceeding ± 75 V at any time point or electrode. Mean valid trial counts (range) were: C+S+ = 43.5 (25–48), C-S+ = 43.0 (25–48), C-S- = 43.2 (21–48).

ERP Analysis: Based on previous research and grand-average waveforms, we selected three time windows for statistical analysis: (1) 300–500 ms for N400;

(2) 500–700 ms for early P600; (3) 700–1000 ms for late P600. We calculated regional averages for frontal (F1, Fz, F2), central (C1, Cz, C2), and parietal (P1, Pz, P2) electrodes. For example, parietal N400 was the average amplitude across P1, Pz, and P2 during 300–500 ms. Greenhouse-Geisser correction was applied when sphericity was violated, with post-hoc comparisons conducted via paired t-tests and Bonferroni correction for multiple comparisons.

EEG Decoding: We conducted multivariate pattern analysis using the MVPA-light toolbox (Treder, 2020) in MATLAB. Linear discriminant analysis (LDA) distinguished EEG data across conditions. We used preprocessed EEG data excluding EOG (VEOG, HEOG) and reference (M1, M2) electrodes. To reduce computational load, we downsampled data to 250 Hz. Trial counts were balanced across conditions via random sampling. Data were converted to z-scores to improve model performance and stability. To enhance signal-to-noise ratio, we randomly averaged 5 trials into 1 until fewer than 5 trials remained. During decoding, data were randomly split into three folds, with two folds used for training and one for testing. This process was repeated three times so each fold served as test data, with the entire cross-validation repeated 10 times.

After decoding, we performed cluster-based permutation testing on decoding accuracy to identify time points significantly above chance level (chance = 33.33%, $\alpha = 0.05$, permutations = 1000).

Regression Analysis: To directly examine relationships between study-phase ERP components and subsequent memory performance, we conducted regression analyses. For each participant, we calculated correlations between ERP components (N400, early P600, late P600) and memory performance (Pr). Each participant thus had 3 ERP values (mean amplitudes for C+S+, C-S+, C-S-) and 3 behavioral values (mean Pr for each condition), yielding one regression coefficient per participant. We then tested whether standardized beta coefficients differed significantly from zero (Lorch & Myers, 1990).

Experiment 1 Results

Behavioral Results   Test Phase: Pr
To eliminate response bias in old/new judgments, we calculated Pr as hit rate minus false alarm rate for each condition. One-way repeated measures ANOVA revealed significant differences across conditions ($F(2, 54) = 30.59$, $p < 0.001$, $p^2 = 0.53$). Post-hoc tests showed Pr was significantly higher for C+S+ than C-S+ ($t(27) = 3.62$, $p = 0.004$, Cohen's d = 0.68) and C-S- ($t(27) = 6.42$, $p < 0.001$, Cohen's d = 1.21), and significantly higher for C-S+ than C-S- ($t(27) = 5.11$, $p < 0.001$, Cohen's d = 0.97) (see Table 1 , Figure 2 [Figure 2: see original paper]A). These results indicate that memory performance declined as predictability decreased.

Study Phase: Accuracy

One-way repeated measures ANOVA revealed significant differences across conditions ($F(2, 54) = 16.70$, $p < 0.001$, $p^2 = 0.38$). Post-hoc tests showed accuracy was significantly higher for C-S- than C-S+ ($t(27) = 6.09$, $p < 0.001$, Cohen's d = 1.15) and C+S+ ($t(27) = 3.31$, $p = 0.008$, Cohen's d = 0.63), and significantly higher for C+S+ than C-S+ ($t(27) = 2.67$, $p = 0.038$, Cohen's d = 0.50) (see Table 1, Figure 2A).

Reaction Time (RT)
One-way repeated measures ANOVA revealed significant differences across conditions ($F(2, 54) = 19.68$, $p < 0.001$, $p^2 = 0.42$). Post-hoc tests showed RT was significantly longer for C-S+ than C+S+ ($t(27) = 5.15$, $p < 0.001$, Cohen's d = 0.97) and C-S- ($t(27) = 4.85$, $p < 0.001$, Cohen's d = 0.92), with no significant difference between C+S+ and C-S- ($t(27) = -0.64$, $p = 1$, Cohen's d = -0.12) (see Table 1, Figure 2A).

Study-phase behavioral results showed that C+S+ and C-S- conditions yielded higher accuracy and shorter RTs than C-S+, indicating processing advantages (higher accuracy, faster responses) in C+S+ and C-S-.

EEG Results   Leveraging EEG's high temporal resolution, we examined neural encoding processes across three time windows (300–500 ms: N400; 500–700 ms: early P600; 700–1000 ms: late P600) under different predictability levels.

N400
A 3 (Region: frontal, central, parietal) × 3 (Condition: C+S+, C-S+, C-S-) repeated measures ANOVA revealed a significant main effect of condition ($F(2, 54) = 22.89$, $p < 0.001$, $p^2 = 0.50$) and no significant region × condition interaction ($F(4, 108) = 0.85$, $p = 0.496$, $p^2 = 0.03$). Post-hoc tests showed N400 was more positive for C+S+ (M = 4.28 V, SE = 0.94) than C-S+ (M = 2.54 V, SE = 0.74, $t(27) = 4.18$, $p < 0.001$, Cohen's d = 0.79) and C-S- (M = 1.41 V, SE = 0.79, $t(27) = 5.66$, $p < 0.001$, Cohen's d = 1.07), and more positive for C-S+ than C-S- ($t(27) = 3.30$, $p = 0.008$, Cohen's d = 0.62). Given N400's parietal distribution, we conducted a separate one-way ANOVA for the parietal region, finding significant differences ($F(2, 54) = 19.67$, $p < 0.001$, $p^2 = 0.42$). Parietal N400 was more positive for C+S+ (M = 5.00 V, SE = 0.85) than C-S+ (M = 3.46 V, SE = 0.68, $t(27) = 4.12$, $p < 0.001$, Cohen's d = 0.78) and C-S- (M = 2.32 V, SE = 0.64, $t(27) = 5.76$, $p < 0.001$, Cohen's d = 1.09), and more positive for C-S+ than C-S- ($t(27) = 2.59$, $p = 0.046$, Cohen's d = 0.49) (see Figure 2B). This indicates that target word predictability decreased progressively across C+S+, C-S+, and C-S- conditions.

Early P600
A 3 × 3 repeated measures ANOVA revealed no significant main effect of condition ($F(2, 54) = 1.49$, $p = 0.235$, $p^2 = 0.05$) but a significant region × condition interaction ($F(4, 108) = 4.43$, $p = 0.002$, $p^2 = 0.14$). Simple effects analysis revealed no significant differences among the three conditions (see Figure 2B).

Late P600

A 3 × 3 repeated measures ANOVA revealed no significant main effect of condition ($F(2, 54) = 0.67$, $p = 0.52$, $p^2 = 0.02$) and no significant region × condition interaction ($F(4, 108) = 1.24$, $p = 0.297$, $p^2 = 0.04$) (see Figure 2B).

EEG Decoding

We conducted MVPA (decoding) on study-phase EEG data using LDA. Cluster-based permutation tests revealed that LDA successfully distinguished conditions during 340–440 ms and 584–912 ms post-target ($p < 0.05$, see Figure 2C). Temporal generalization analysis showed no cross-temporal decoding between N400 and P600 time windows (see Figure 2D).

These results indicate significant differences in neural representations across conditions during both N400 and P600 time windows, with temporal generalization analysis showing that N400 and P600 reflect distinct cognitive processes (He et al., 2024; Schotter et al., 2023).

Regression Analysis

We conducted paired t-tests on standardized regression coefficients between ERP components and Pr to test whether they differed from zero. N400-Pr regression coefficients were significantly greater than zero in frontal (M = 0.40, SE = 0.14, $t(27) = 2.88$, $p = 0.008$, Cohen's d = 0.55), central (M = 0.47, SE = 0.12, $t(27) = 3.93$, $p < 0.001$, Cohen's d = 0.74), and parietal regions (M = 0.42, SE = 0.12, $t(27) = 3.51$, $p = 0.002$, Cohen's d = 0.66). Early P600-Pr coefficients did not differ from zero in frontal (M = 0.18, SE = 0.13, $t(27) = 1.44$, $p = 0.162$, Cohen's d = 0.27), central (M = 0.21, SE = 0.14, $t(27) = 1.45$, $p = 0.159$, Cohen's d = 0.27), or parietal regions (M = 0.02, SE = 0.15, $t(27) = 0.11$, $p = 0.916$, Cohen's d = 0.02). Late P600-Pr coefficients also did not differ from zero in frontal (M = 0.22, SE = 0.12, $t(27) = 1.78$, $p = 0.087$, Cohen's d = 0.34), central (M = 0.21, SE = 0.14, $t(27) = 1.48$, $p = 0.150$, Cohen's d = 0.28), or parietal regions (M = -0.11, SE = 0.14, $t(27) = -0.83$, $p = 0.417$, Cohen's d = -0.16) (see Figure 2E).

These results demonstrate that study-phase N400 effectively predicted subsequent memory performance, whereas P600 (early and late) did not.

Experiment 1 Discussion

Experiment 1 examined item recognition memory across different predictability levels. During the study phase, C+S+ showed higher accuracy and shorter RTs than C-S+, indicating a processing advantage. N400 amplitudes became progressively more negative across C+S+, C-S+, and C-S- conditions. During the test phase, memory performance declined across these conditions, and study-phase N400 significantly predicted test-phase memory performance. These results show that predictability level effectively predicts subsequent memory performance, supporting the view that predictable items yield better memory (Craik & Tulving, 1975; Kutas, 1993; Schulman, 1974; Silcox et al., 2023).

P600 reflects semantic integration of word meaning with context (Brouwer et al., 2017; Federmeier, 2022; Schotter et al., 2023). Previous studies found that anomalous words elicited more positive P600, reflecting greater cognitive effort required for integration (Kuperberg et al., 2003; Schotter et al., 2023). However, Experiment 1 found no P600 differences across conditions. We speculate that the study-phase task (word-pair judgment) did not require semantic integration—participants could complete judgments without integrating word meanings. To investigate the relationship between P600 and memory performance, we conducted Experiment 2 (associative recognition).

In Experiment 2, we examined associative recognition memory across predictability levels. Compared to Experiment 1, we made three changes: First, participants were explicitly informed about both study-phase (word-pair judgment) and test-phase (associative recognition) tasks before the experiment began. Second, participants received immediate associative recognition tests after each study block to encourage integration of prime and target words.

Experiment 2

Participants Based on the effect size from Experiment 1 ($p^2 = 0.53$), 12 participants were needed for 95% power ($\alpha = 0.05$). Experiment 2 recruited 26 participants aged 18–25 years (mean age = 22 years; 16 females, 10 males). All procedures, including ethics approval and informed consent, were identical to Experiment 1. The entire experiment lasted approximately 100 minutes.

Materials Identical to Experiment 1.

Procedure Experiment 2 differed from Experiment 1 in three ways: (1) The test phase assessed associative recognition of prime-target combinations (e.g., "furniture sofa") rather than target words alone. Participants judged pairs as old, recombined, or new; (2) Recombined pairs served as filler trials to prevent judgments based on single words. These were created by randomly selecting one C+S+ pair (e.g., "Chinese medicine: ginseng"), one C-S+ pair (e.g., "furniture: decoration"), and one C-S- pair (e.g., "color: bullet"), then recomposing them (e.g., "Chinese medicine: decoration," "furniture: bullet," "color: ginseng"). Recombined pairs were excluded from analysis; (3) Participants completed immediate associative recognition tests after each study block. The experiment comprised 7 study-test blocks. Each study block contained 8 C+S+, 8 C-S+, and 8 C-S- trials. Each test block included 7 old and 7 new pairs per condition, plus 3 recombined filler pairs (see Figure 1B).

EEG Recording and Analysis After preprocessing, mean valid trial counts (range) were: C+S+ = 50.7 (34–56), C-S+ = 50.6 (41–56), C-S- = 51.0 (38–56). All other parameters were identical to Experiment 1.

Experiment 2 Results

Behavioral Results   Test Phase: Pr
One-way repeated measures ANOVA revealed significant differences across conditions ($F(2, 50) = 54.44$, $p < 0.001$, $p^2 = 0.69$). Post-hoc tests showed Pr was significantly higher for C+S+ than C-S+ ($t(25) = 4.61$, $p < 0.001$, Cohen's d = 0.90) and C-S- ($t(25) = 9.43$, $p < 0.001$, Cohen's d = 1.85), and significantly higher for C-S+ than C-S- ($t(25) = 6.52$, $p < 0.001$, Cohen's d = 1.28) (see Table 1, Figure 3 [Figure 3: see original paper]A). Memory performance declined as predictability decreased.

Study Phase: Accuracy
One-way repeated measures ANOVA revealed significant differences across conditions ($F(2, 50) = 14.85$, $p < 0.001$, $p^2 = 0.37$). Post-hoc tests showed accuracy was significantly lower for C-S+ than C+S+ ($t(25) = -3.52$, $p = 0.005$, Cohen's d = -0.69) and C-S- ($t(25) = -5.85$, $p < 0.001$, Cohen's d = -1.15), with no difference between C+S+ and C-S- ($t(25) = -1.03$, $p = 0.934$, Cohen's d = -0.20) (see Table 1, Figure 3A).

Reaction Time (RT)
One-way repeated measures ANOVA revealed significant differences across conditions ($F(2, 50) = 29.51$, $p < 0.001$, $p^2 = 0.54$). Post-hoc tests showed RT was significantly longer for C-S+ than C+S+ ($t(25) = 7.48$, $p < 0.001$, Cohen's d = 1.47) and C-S- ($t(25) = 6.10$, $p < 0.001$, Cohen's d = 1.20), with no difference between C+S+ and C-S- ($t(25) = -1.88$, $p = 0.214$, Cohen's d = -0.37) (see Table 1, Figure 3A).

EEG Results   N400
A 3×3 repeated measures ANOVA revealed a significant main effect of condition ($F(2, 50) = 20.76$, $p < 0.001$, $p^2 = 0.45$) and no significant region × condition interaction ($F(4, 100) = 1.21$, $p = 0.312$, $p^2 = 0.05$). Post-hoc tests showed N400 was more positive for C+S+ (M = 3.55 V, SE = 0.70) than C-S+ (M = 2.60 V, SE = 0.63, $t(25) = 2.67$, $p = 0.040$, Cohen's d = 0.52) and C-S- (M = 0.86 V, SE = 0.66, $t(25) = 5.37$, $p < 0.001$, Cohen's d = 1.05), and more positive for C-S+ than C-S- ($t(25) = 4.35$, $p < 0.001$, Cohen's d = 0.85). A separate one-way ANOVA for the parietal region showed significant differences ($F(2, 50) = 27.92$, $p < 0.001$, $p^2 = 0.53$). Parietal N400 was more positive for C+S+ (M = 4.03 V, SE = 0.81) than C-S+ (M = 3.13 V, SE = 0.65, $t(25) = 2.41$, $p = 0.071$, marginally significant, Cohen's d = 0.47) and C-S- (M = 1.07 V, SE = 0.63, $t(25) = 6.00$, $p < 0.001$, Cohen's d = 1.18), and more positive for C-S+ than C-S- ($t(25) = 6.17$, $p < 0.001$, Cohen's d = 1.21) (see Figure 3B). This indicates decreasing predictability across C+S+, C-S+, and C-S-.

Early P600
A 3×3 repeated measures ANOVA revealed no significant main effect of con-

dition (F(2, 50) = 1.80, p = 0.175, p^2 = 0.07) and no significant region × condition interaction (F(4, 100) = 1.24, p = 0.298, p^2 = 0.05) (see Figure 3B).

Late P600

A 3 × 3 repeated measures ANOVA revealed a significant main effect of condition (F(2, 50) = 2.85, p = 0.046, p^2 = 0.12) and no significant region × condition interaction (F(4, 100) = 0.62, p = 0.653, p^2 = 0.02). Post-hoc tests showed amplitude was more positive for C-S+ (M = 3.32 V, SE = 0.74) than C-S- (M = 2.26 V, SE = 0.63, t(25) = 2.90, p = 0.023, Cohen's d = 0.57). C+S+ (M = 3.16 V, SE = 0.79) did not differ from C-S+ (t(25) = -0.33, p = 1, Cohen's d = -0.07) or C-S- (t(25) = 1.81, p = 0.249, Cohen's d = 0.35). Given P600's parietal distribution, a separate one-way ANOVA for the parietal region showed significant differences (F(2, 50) = 6.18, p = 0.004, p^2 = 0.20). Parietal amplitudes were more positive for C+S+ (M = 3.90 V, SE = 0.75, t(25) = 2.63, p = 0.044, Cohen's d = 0.52) and C-S+ (M = 4.28 V, SE = 0.67, t(25) = 3.80, p = 0.003, Cohen's d = 0.74) than C-S- (M = 2.91 V, SE = 0.64), with no difference between C+S+ and C-S+ (t(25) = -0.83, p = 1, Cohen's d = -0.16) (see Figure 3B). This indicates more positive parietal amplitudes for C+S+ and C-S+ versus C-S-.

EEG Decoding

LDA decoding of study-phase EEG data revealed significant condition discrimination during 336–500 ms and 668–1000 ms post-target (p < 0.05, see Figure 3C). Temporal generalization analysis showed no cross-temporal decoding between N400 and P600 windows (see Figure 3D), consistent with Experiment 1.

Regression Analysis

Paired t-tests on standardized regression coefficients showed N400-Pr coefficients were significantly greater than zero in frontal (M = 0.45, SE = 0.12, t(25) = 3.78, p < 0.001, Cohen's d = 0.74), central (M = 0.56, SE = 0.11, t(25) = 4.94, p < 0.001, Cohen's d = 0.97), and parietal regions (M = 0.60, SE = 0.11, t(25) = 5.37, p < 0.001, Cohen's d = 1.05). Early P600-Pr coefficients did not differ from zero in frontal (M = 0.08, SE = 0.12, t(25) = 0.67, p = 0.511, Cohen's d = 0.13), central (M = 0.09, SE = 0.14, t(25) = 0.68, p = 0.501, Cohen's d = 0.13), or parietal regions (M = 0.25, SE = 0.15, t(25) = 1.75, p = 0.092, Cohen's d = 0.34). Late P600-Pr coefficients also did not differ from zero in frontal (M = 0.17, SE = 0.16, t(25) = 1.07, p = 0.297, Cohen's d = 0.21), central (M = 0.17, SE = 0.15, t(25) = 1.20, p = 0.243, Cohen's d = 0.23), or parietal regions (M = 0.15, SE = 0.15, t(25) = 1.00, p = 0.327, Cohen's d = 0.20) (see Figure 3E).

Experiment 2 Discussion

Experiment 2 examined associative recognition memory across predictability levels. During the study phase, C+S+ showed higher accuracy and shorter RTs than C-S+, with N400 amplitudes becoming progressively more negative

across C+S+, C-S+, and C-S-. During the test phase, associative recognition memory declined across these conditions, and study-phase N400 significantly predicted associative memory performance. These results replicate Experiment 1. Additionally, C+S+ and C-S+ showed more positive P600 amplitudes and better associative memory than C-S-, indicating that semantic distance facilitates prime-target integration, which enhances subsequent associative memory.

General Discussion

This study (Experiment 1: item recognition; Experiment 2: associative recognition) manipulated prediction through category and semantic distance to examine encoding differences and memory performance across predictability levels. Both experiments consistently found that C+S+ and C-S- yielded higher accuracy and shorter RTs than C-S+ during the study phase. N400 amplitudes became progressively more negative across C+S+, C-S+, and C-S-. Only Experiment 2 found more positive P600 amplitudes for C+S+ and C-S+ versus C-S-. In both experiments, LDA successfully distinguished neural representations across conditions during N400 and P600 time windows. During the test phase, item recognition (Experiment 1) and associative recognition (Experiment 2) both showed graded memory decline across C+S+, C-S+, and C-S-. Moreover, both experiments consistently demonstrated that study-phase N400 significantly predicted test-phase memory performance. These results indicate that prediction's impact on memory is modulated by category and semantic distance.

Both experiments found that C+S+ and C-S- yielded higher accuracy and shorter RTs than C-S+. While both conditions showed behavioral advantages, we propose they reflect different cognitive mechanisms. Specifically, C+S+'s superior performance stemmed from facilitative effects of semantic context (Silcox & Payne, 2021), where appropriate semantic context enhances semantic processing of matching words. C-S-'s advantage may reflect a self-termination procedure: LeFevre and Bisanz (1986) found that when processing anomalous sequences, identification of an anomalous element triggers processing termination, saving subsequent processing time. This may explain why C-S- showed shorter RTs than C-S+. Notably, Rommers et al. (2013) found longer RTs for anomalous conditions using sentence frames, suggesting that anomalous item processing varies across paradigms—a point we will revisit in relation to EEG results.

Category and semantic distance significantly affected predictability. Both experiments found progressively more negative N400 amplitudes across C+S+, C-S+, and C-S-, consistent with previous research showing highest predictability for high-typicality category members (Federmeier et al., 2010; Silcox et al., 2023). The more positive N400 for C+S+ versus C-S+ demonstrates category's influence on predictability, while the more positive N400 for C-S+ versus C-S-

demonstrates semantic distance's influence. These results indicate that prediction is a complex cognitive process integrating multiple information sources (e.g., category, semantic distance) to forecast upcoming events.

Semantic distance modulated integration encoding of prime-target pairs. Both experiments showed LDA distinguished conditions during the P600 window, and Experiment 2 found more positive P600 amplitudes for C+S+ and C-S+ versus C-S-. This indicates that semantic distance facilitates prime-target integration encoding. While Experiment 1 found no P600 differences, we attribute this to task demands. First, Experiment 2's explicit associative recognition test encouraged participants to invest more effort in integration encoding. P600 indexes semantic integration (Brouwer et al., 2017; Federmeier, 2022; Schotter et al., 2023), and the closer semantic distance in C+S+ and C-S+ promoted integration, yielding more positive P600 amplitudes. C-S-'s self-termination procedure (LeFevre & Bisanz, 1986) may have reduced integration effort, producing smaller P600. Sentence-frame studies showing longer RTs and more positive P600 for anomalous words (Rommers et al., 2013; Schotter et al., 2023) likely reflect higher semantic integration demands compared to word-pair judgments, explaining paradigm differences. Second, MVPA is more sensitive than univariate analysis, detecting differences ERP cannot (Davis et al., 2014; Petit et al., 2024). Experiment 1 may not have motivated integration effort, so condition differences appeared only in the more sensitive MVPA rather than P600 amplitude.

Category and semantic distance influenced memory performance. Both experiments found higher Pr for C+S+ than C-S+, indicating that conforming to category rules enhances memory. Higher Pr for C-S+ than C-S- indicates that near semantic distance improves memory. The significant positive correlation between study-phase N400 (an index of prediction) and test-phase memory performance supports the view that predictable materials yield better memory (Craik & Tulving, 1975; Kutas, 1993; Schulman, 1974; Silcox et al., 2023). Combined with previous research (Silcox et al., 2023; Silcox & Payne, 2021), we propose that predictable contexts reduce encoding burden and facilitate integration of target words with semantic context, thereby improving memory performance.

We propose that category and semantic distance influence memory through distinct mechanisms. First, both factors affect memory. Previous studies manipulating typicality (high/low) found better memory for high-typicality (e.g., flower: rose) than low-typicality (flower: poppy) items (Silcox et al., 2023). Here, both targets were category members but differed in semantic distance, demonstrating semantic distance's influence. Our finding that C+S+ (furniture: sofa) yielded better memory than C-S+ (furniture: decoration) shows category's influence when semantic distance is controlled. Thus, memory performance is modulated by both category and semantic distance.

Second, we speculate these factors operate through different mechanisms. Some studies suggest that rule-congruent items enhance memory by reducing encoding

burden (Frank & Kafkas, 2021; Silcox & Payne, 2021). Our study-phase results showing shorter RTs, higher accuracy, and better memory for C+S+ than C-S+ support this encoding advantage for rule-congruent items. Conversely, our finding that C+S+ and C-S+ showed more positive P600 and better memory than C-S- suggests that near semantic distance enhances memory by facilitating prime-target integration. Additionally, a recent retrieval study (Dai et al., 2025) found significant LPC old/new effects for semantic predictability (C+S+) but FN400 old/new effects for semantic relatedness (C-S+), indicating distinct retrieval mechanisms. Silcox et al. (2023) used TMS to show that LIFC is associated with semantic distance's effect on memory, not category's effect, providing neural evidence for distinct mechanisms. These findings suggest category and semantic distance influence memory through different mechanisms, offering a potential resolution to the predictable-unpredictable memory debate.

We propose that the predictable-unpredictable memory discrepancy may be determined by the relative contributions of encoding difficulty and encoding effort—two separable factors. Although researchers debate the precise functional significance of N400 and P600, there is consensus that they are distinct ERP components reflecting different cognitive processes (He et al., 2024; Kutas & Federmeier, 2011; Troyer et al., 2024). Our temporal generalization results showing no cross-decoding between N400 and P600 support this distinction. We therefore propose that at least two independent cognitive processes operate during prediction's influence on memory. N400 reflects semantic processing difficulty (Hagoort et al., 2009; Kutas & Federmeier, 2011), while P600 reflects cognitive effort invested in processing (Rosburg et al., 2015). We suggest that N400 indexes encoding difficulty and P600 indexes encoding effort. When cognitive effort is equivalent, lower encoding difficulty yields better memory. When encoding difficulty is equivalent, greater effort yields better memory. Critically, N400 and P600 are not necessarily linked—easy encoding does not imply less effort, and greater effort does not imply difficult encoding. Predictable conditions typically involve easier encoding because they align with prior knowledge (Frank & Kafkas, 2021). Encoding effort varies across paradigms. When participants invest more effort in predictable conditions (as in our study), predictable items show better memory. When they invest more effort in unpredictable conditions, memory performance depends on the relative balance between encoding difficulty and encoding effort.

Our results support the view that predictable materials yield better memory performance. However, given variations across studies in paradigms, materials, and prediction manipulations (see Frank & Kafkas, 2021), we must carefully delimit our conclusions. First, we used a word-pair judgment paradigm manipulating prediction through category and semantic distance. The preceding C+S+ pair may have primed participants to generate predictions along category and semantic dimensions, potentially limiting generalizability to real-world prediction formation, which involves more dimensions. Second, while we propose two sub-processes (encoding difficulty and encoding effort) in prediction's effect on memory, ongoing debate about N400 and P600 functional significance means

our interpretation remains somewhat speculative and requires future validation. Third, P600 results for unpredictable conditions vary across paradigms, suggesting paradigm influences on P600. Future research should examine P600-memory relationships to resolve the predictable-unpredictable memory debate.

In summary, this study demonstrates that category and semantic distance modulate prediction's impact on memory performance. Predictable semantic contexts enhance subsequent memory by reducing encoding burden and/or facilitating semantic integration. These findings support the view that predictable items yield better memory performance and illuminate the relationship between encoding processes and subsequent memory, advancing our understanding of the prediction-memory relationship.

References

Bar, M. (2007). The proactive brain: Using analogies and associations to generate predictions. *Trends in Cognitive Sciences, 11*(7), 280−289.

Bar, M. (2009). Predictions: A universal principle in the operation of the human brain. *Philosophical Transactions of the Royal Society B-Biological Sciences, 364*(1521), 1181−1182.

Bonhage, C. E., Mueller, J. L., Friederici, A. D., & Fiebach, C. J. (2015). Combined eye tracking and fMRI reveals neural basis of linguistic predictions during sentence comprehension. *Cortex, 68*, 33−47.

Brouwer, H., Crocker, M. W., Venhuizen, N. J., & Hoeks, J. C. J. (2017). A neurocomputational model of the N400 and the P600 in language processing. *Cognitive Science, 41*, 1318−1352.

Craik, F., & Tulving, E. (1975). Depth of processing and retention of words in episodic memory. *Journal of Experimental Psychology-General, 104*(3), 268−294.

Dai, J., Liang, P., Li, X., Zhang, J., Tian, L., Mao, X., & Guo, C. (2025). Semantic predictability and semantic relevance through different neural mechanisms to improve memory performance. *Brain and Cognition, 186*, 105998.

Davis, T., LaRocque, K. F., Mumford, J. A., Norman, K. A., Wagner, A. D., & Poldrack, R. A. (2014). What do differences between multi-voxel and univariate analysis mean? How subject-, voxel-, and trial-level variance impact fMRI analysis. *Neuroimage, 97*, 271−283.

DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience, 8*(8), 1117−1121.

Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis.

Journal of Neuroscience Methods, 134(1), 9−21.

Elmer, S., Besson, M., & Rodríguez-Fornells, A. (2022). The electrophysiological correlates of word pre-activation during associative word learning. *International Journal of Psychophysiology, 182*, 12−22.

Federmeier, K. D. (2022). Connecting and considering: Electrophysiology provides insights into comprehension. *Psychophysiology, 59*(1), e13940.

Federmeier, K. D., Kutas, M., & Schul, R. (2010). Age-related and individual differences in the use of prediction during language comprehension. *Brain and Language, 115*(3), 149−161.

Federmeier, K. D., Wlotko, E. W., De Ochoa-Dewald, E., & Kutas, M. (2007). Multiple effects of sentential constraint on word processing. *Brain Research, 1146*, 75−84.

Frank, D., & Kafkas, A. (2021). Expectation-driven novelty effects in episodic memory. *Neurobiology of Learning and Memory, 183*, 107466.

Frank, D., Montemurro, M. A., & Montaldi, D. (2020). Pattern separation underpins expectation-modulated memory. *Journal of Neuroscience, 40*(17), 3455−3464.

Giglio, L., Ostarek, M., Weber, K., & Hagoort, P. (2022). Commonalities and asymmetries in the neurobiological infrastructure for language production and comprehension. *Cerebral Cortex, 32*(7), 1405−1418.

Hagoort, P., Baggio, G., & Willems, R. M. (2009). Semantic unification. In M. S. Gazzaniga (Ed.), *The Cognitive neurosciences IV*. Cambridge, MA: MIT press.

Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science, 293*(5539), 2425−2430.

He, Y., Sommer, J., Hansen-Schirra, S., & Nagels, A. (2024). Multivariate pattern analysis of EEG reveals nuanced impact of negation on sentence processing in the N400 and later time windows. *Psychophysiology, 61*(4), e14491.

Kafkas, A., & Montaldi, D. (2015). Striatal and midbrain connectivity with the hippocampus selectively boosts memory for contextual novelty. *Hippocampus, 25*(11), 1262−1273.

King, J.-R., & Dehaene, S. (2014). Characterizing the dynamics of mental representations: The temporal generalization method. *Trends in Cognitive Sciences, 18*(4), 203−210.

Kuperberg, G. R., Sitnikova, T., Caplan, D., & Holcomb, P. J. (2003). Electrophysiological distinctions in processing conceptual relationships within simple sentences. *Cognitive Brain Research, 17*(1), 117−129.

Kutas, M. (1993). In the company of other words—Electrophysiological evidence for single-word and sentence context effects. *Language and Cognitive Processes, 8*(4), 533−572.

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology, 62*(1), 621–647.

Kutas, M., & Hillyard, S. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature, 307*(5947), 161−163.

LeFevre, J. A., & Bisanz, J. (1986). A cognitive analysis of number-series problems: Sources of individual differences in performance. *Memory & Cognition, 14*(4), 287−298.

Lorch, R., & Myers, J. (1990). Regression-analyses of repeated measures data in cognitive research. *Journal of Experimental Psychology-Learning Memory and Cognition, 16*(1), 149−157.

Menenti, L., Gierhan, S. M. E., Segaert, K., & Hagoort, P. (2011). Shared language: Overlap and segregation of the neuronal infrastructure for speaking and listening revealed by functional MRI. *Psychological Science, 22*(9), 1173−1182.

Nyberg, L. (2005). Any novelty in hippocampal formation and memory? *Current Opinion in Neurology, 18*(4), 424−428.

Petit, S., Brown, A., Jessen, E. T., & Woolgar, A. (2024). How robustly do multivariate EEG patterns track individual-subject lexico-semantic processing of visual stimuli? *Language Cognition and Neuroscience, 39*(9), 1134−1148.

Rajaram, S. (1998). The effects of conceptual salience and perceptual distinctiveness on conscious recollection. *Psychonomic Bulletin & Review, 5*(1), 71−78.

Rommers, J., Dijkstra, T., & Bastiaansen, M. (2013). Context-dependent semantic processing in the human brain: Evidence from idiom comprehension. *Journal of Cognitive Neuroscience, 25*(5), 762−776.

Rosburg, T., Johansson, M., Weigl, M., & Mecklinger, A. (2015). How does testing affect retrieval-related processes? An event-related potential (ERP) study on the short-term effects of repeated retrieval. *Cognitive Affective & Behavioral Neuroscience, 15*(1), 195−210.

Ryskin, R., Ng, S., Mimnaugh, K., Brown-Schmidt, S., & Federmeier, K. D. (2020). Talker-specific predictions during language processing. *Language, Cognition and Neuroscience, 35*(6), 797−812.

Schotter, E. R., Milligan, S., & Estevez, V. M. (2023). Event-related potentials show that parafoveal vision is insufficient for semantic integration. *Psychophysiology, 60*(7), e14246.

Schulman, A. I. (1974). Memory for words recently classified. *Memory & Cognition, 2*(1), 47−52.

Silbert, L. J., Honey, C. J., Simony, E., Poeppel, D., & Hasson, U. (2014). Coupled neural systems underlie the production and comprehension of naturalistic narrative speech. *Proceedings of The National Academy of Sciences of The United States of America, 111*(43), E4687−E4696.

Silcox, J. W., Mickey, B., & Payne, B. R. (2023). Disruption to left inferior frontal cortex modulates semantic prediction effects in reading and subsequent memory: Evidence from simultaneous TMS-EEG. *Psychophysiology, 60*(9), e14312.

Silcox, J. W., & Payne, B. R. (2021). The costs (and benefits) of effortful listening on context processing: A simultaneous electrophysiology, pupillometry, and behavioral study. *Cortex, 142*, 296−316.

Treder, M. S. (2020). MVPA-Light: A classification and regression toolbox for multi-dimensional data. *Frontiers in Neuroscience, 14*, 289.

Troyer, M., Kutas, M., Batterink, L., & McRae, K. (2024). Nuances of knowing: Brain potentials reveal implicit effects of domain knowledge on word processing in the absence of sentence-level knowledge. *Psychophysiology, 61*(1), e14422.

Zhao, M., Xu, Z. Y., Liu, T., Du, F. L., Li, Y. X., & Chen, F. Y. (2012). The neuromechanism underlying language analogical reasoning: Evidence from an ERP study. *Acta Psychologica Sinica, 44*(6), 711−719.

Note: Figure translations are in progress. See original paper for figures.

Source: ChinaXiv — Machine translation. Verify with original.