What’s in a Name? Brain Activity Reveals Categorization Processes Differ Across Languages

Chao Liu,1,2,3 Twila Tardif,1,2* Xiaoqin Mai,2,3 William J. Gehring,1,2 Nina Simms,4 and Yue-Jia Luo5*

1Department of Psychology, University of Michigan, Ann Arbor, Michigan
2Center for Human Growth and Development, University of Michigan, Ann Arbor, Michigan
3Key Laboratory of Mental Health, Institute of Psychology, Chinese Academy of Sciences, Beijing, China
4Department of Psychology, Northwestern University, Evanston, Illinois
5State Key Laboratory of Cognitive Neuroscience and Learning, Beijing Normal University, Beijing, China

Abstract: The linguistic relativity hypothesis proposes that speakers of different languages perceive and conceptualize the world differently, but do their brains reflect these differences? In English, most nouns do not provide linguistic clues to their categories, whereas most Mandarin Chinese nouns provide explicit category information, either morphologically (e.g., the morpheme “vehicle” che1 車 in the noun “train” hua3che1 車 in the noun “train” hua3che1 車 or orthographically (e.g., the radical “bug” chong2 虫 in the character for the noun “butterfly” hu2die2 蝴蝶). When asked to judge the membership of atypical (e.g., train) vs. typical (e.g., car) pictorial exemplars of a category (e.g., vehicle), English speakers (N = 26) showed larger N300 and N400 event-related potential (ERP) component differences, whereas Mandarin speakers (N = 27) showed no such differences. Further investigation with Mandarin speakers only (N = 22) found that it was the morphologically transparent items that did not show a typicality effect, whereas orthographically transparent items elicited moderate N300 and N400 effects. In a follow-up study with English speakers only (N = 25), morphologically transparent items also showed different patterns of N300 and N400 activation than nontransparent items even for English speakers. Together, these results demonstrate that even for pictorial stimuli, how and whether category information is embedded in object names affects the extent to which typicality is used in category judgments, as shown in N300 and N400 responses. Hum Brain Mapp 31:1786–1801, 2010. © 2010 Wiley-Liss, Inc.

Key words: event-related potentials; categorization; typicality effect; N300; N400; morphological and orthographical processing; mandarin Chinese; linguistic relativity hypothesis; crosscultural
INTRODUCTION

Is a canola a flower? Is a rose a flower? When asked questions such as this, North American English speakers generally respond to “rose” more quickly and more accurately than “canola” (Brassica napus, or the flower used in canola oil). Many researchers [Bjorklund et al., 1983; Boster, 1988; Chumley, 1986; Komatsu, 1992; Mervis and Rosch, 1981; Mervis et al., 1976; Schwanenflugel and Rey, 1986] believe that this is because most people think a rose is a more typical example of the category “flower” than canola. Nonetheless, the linguistic relativity hypothesis proposed by Whorf [1956] argues that the language one speaks influences the way one thinks. Would Chinese speakers then be as fast and accurate to classify canola you2cai4hua1 油菜花 as rose mei2gui4hua1 玫瑰花 because both words have the category term flower hua1 花 embedded in their names?

The typicality effect is one of the most consistent indexes of categorization processes in behavioral studies [Mervis and Rosch, 1981; Mervis et al., 1976], and has also been investigated with neuroimaging techniques such as the event-related potential (ERP). ERP studies with English speakers have found that typicality effects in linguistic stimuli are marked by a negative component, peaking approximately at 400 ms after the stimulus onset (N400) [Kutas and Federmeier, 2000; Kutas and Hillyard, 1980a,b] such that atypical items of a category elicit a larger N400 than typical items [Fujihara et al., 1998; Heinze et al., 1988; Stuss et al., 1988]. Thus, the N400 is thought to be related not just to semantic access at its most general level, but specifically to how integral the meaning of a word or category member is to the category as a whole [Kiefer, 2005]. The N400 can also be elicited by pictorial stimuli, and is usually distributed more frontally and occurs together with an additional component (N300) when pictures instead of words are used [Federmeier and Kutas, 2001; Ganis et al., 1996; McPherson and Holcomb, 1999; West and Holcomb, 2002].

Additional components involving typical vs. atypical stimuli include very early components at 100–160 ms (P160) in occipital areas and 280–300 ms (N300) in frontal, temporal, and parietal areas [Barrett and Rugg, 1990; Hauk et al., 2007; Kiefer, 2001; McPherson and Holcomb, 1999]. In addition, other researchers have suggested there may also be a late occurring positive component (LPC) involved in the typicality effect, which might indicate different levels of decision making and evaluative processes, word frequency differences, [Heinze et al., 1998; Stuss et al., 1988] or violations of rules or goal-related requirements [Sitnikova et al., 2003, 2008]. Although these additional components are clearly of interest in determining the nature of the typicality effect, it is not clear that they all represent semantic access, which is of primary interest for the present question of whether or not what one calls a rose (or a canola) influences categorization and, specifically, the effect of typicality on categorization.

Specifically, it is not clear when typicality is a useful cue for categorization and when or whether other types of information might supersede it. In studies involving both experts and nonexperts, for instance, only nonexperts (US undergraduates vs. Itza’ Mayans or US bird experts) relied on typicality to make judgments about bird classifications, which suggests that typicality is used as a “crutch” for categorization when other information is lacking [Bailenson et al., 2002]. In the present studies, we explore the role of explicit linguistic labels as an alternative to a reliance on typicality during a categorization task. We focus our examination on the N300 and N400 effects to pictorially presented stimuli so as to avoid any potential confounds with differences in the physical stimuli provided to speakers of different languages.

Importantly, both Mandarin Chinese and English can provide explicit linguistic labels to category membership (e.g., you2cai4hua1 油菜花 [canola] or mei2gui4hua1 玫瑰花 [rose]; pufferfish, catfish). In studies with English-speaking children, moreover, this type of cue allows children to more easily learn nouns with the category name embedded in the item label (e.g., oak tree) than those that do not contain explicit category information (e.g., oak) [Gelman et al., 1989]. However, the prevalence of words containing such labels differs substantially across these two languages—they are relatively rare in English, but are highly prevalent in Mandarin [Tardif, 2006; Zhou, 1978; Zhou et al., 1999]. For example, in Mandarin, all wheeled vehicles share a common root morpheme (vehicle che1 车—e.g., bicycle zi4xing2che1 自行车, truck ka3che1 卡车, car jiao4che1 轿车, bus gong1gong4qi4che1 公共汽车, train huo3che1 火车). In addition to this level of morphological cueing, over 80% of Chinese characters provide orthographic labels to the category by including a “radical” which labels semantic information [Zhou, 1978; Zhou et al., 1999]. For example, although the noun bug chong2 虫 is a simple character in its own right, it can also be found as a radical component in many nouns for insects such as fly (cang1ying 蚊蝇), butterfly (hu2di2 蝴蝶), mosquito (wen2zi3 蚊子), and ant (ma2yi3 蚂蚁). This kind of orthographic cue can even be found in the oracle bone characters used 3,500 years ago (e.g., the radical of water shui3 水 in the characters for river he2 河 and wine jiu3 酒).

Here, we report that the explicit linguistic labels to category membership in Mandarin nouns facilitate categorization for Mandarin speakers and thus remove a relative reliance on typicality, as reflected both in behavioral measures and neural activity. Specifically, it is the morphologically transparent items in Mandarin that influences the typicality effect in Mandarin speakers. In addition, even in English, items containing morphologically transparent labels to category membership show different types of reliance on typicality, compared with items that do not provide such information. Nonetheless the attenuation of this effect in English is not as strong as for morphologically transparent items in Mandarin and behavioral responses in English show minimal differences for morphologically transparent vs. nontransparent items.
STUDY I

Method

Participants

Thirty native English speakers in Ann Arbor, MI and 30 native Mandarin speakers in Beijing, China, all right-handed undergraduates with normal vision, participated in this study for payment. Four participants in the English group and three participants in the Mandarin group were excluded from analysis due to poor behavioral performance (accuracy <80% in either Yes or No responses, three English-speaking participants) or too many artifacts in the electroencephalogram data (three Mandarin- and one English-speaking participant). The final samples consisted of 27 Mandarin speakers (14 females, mean age = 21.81, SD = 1.86 years) and 26 English speakers (13 females, mean age = 20.30, SD = 1.99) in the behavioral and ERP data analysis.

Development and design of stimuli

Because our main interest was in comparing the nature of categorization processes across languages, several steps were taken to ensure cross-linguistic comparability and the validity of our results. First, we avoided the use of linguistic labels as our time-locked stimulus to avoid eliciting ERP components specific to a particular orthography in a word that would differ across languages. Thus, we chose to provide linguistic labels for the categories and pictorial stimuli for the actual items. Moreover, in Study 1 we chose to present identical pictures to both groups. This necessitated additional pilot testing (see below) to ensure that they were valid and equally typical/atypical across languages. In addition, for Study 1 it also meant that we had unequal numbers of different types of transparency in the items due to natural variation across languages. This was rectified within each language in Study 2 and 3.

In Pilot Study 1, participants were given a questionnaire about the acceptability of replacing certain terms for each other (e.g., “Can car be used to replace the word vehicle?”). This task was used for several reasons. First, it was necessary to verify that the items belonged to the same category and to create the final set of 10 categories (e.g., vehicle and 20 objects, half typical (e.g., car) and half atypical (e.g., train), used in Study 1 (see Fig. 1B for labels and pictures and Appendix 1 for the corresponding typicality rating results). The presence of explicit linguistic labels in the item-level labels was reflective of the natural presence/absence of such labels in these two languages and was not possible to control while also ensuring that identical pictures were presented in all conditions. Thus, in English, nine of the ten categories were nontransparent (e.g., car) and only one was morphologically transparent (e.g., writing paper). In Mandarin, eight of the ten categories contained morphologically transparent labels, and two were nontransparent. Importantly, there were no differences in the mean ratings of either typical (English $M = 5.74$, SD = 0.19, Mandarin $M = 5.13$, SD = 0.39) or atypical items across languages (English $M = 3.90$, SD = 0.63, Mandarin $M = 3.70$, SD = 1.13), with only a main effect of typicality for these items, $F(1, 18) = 53.68, P < 0.001$.

Procedure and task

English- and Mandarin-speaking participants were tested in their native language in either the US or China, respectively. Presentation of the stimuli was controlled with the E-prime program, with participants sitting ~20–28 in. away from the screen (resulting ~ 3’ of visual angle for words, 1° for crosshair, and 10° for pictures) in each location. The presentation procedure can be found in Figure 1A. Participants first saw either a category-level label (e.g., “VEHICLE” in English or che1 “车” in
Mandarin) or an item-level label (e.g., “CAR” in English or jiao4che1 车 in Mandarin), followed by a picture of either a typical, atypical or out-of-category object (e.g., a car, a train or a pen, respectively). The participant’s instructions were to “judge whether or not the picture is an example of the concept represented by the preceding word” (判断图片是否为前面的词所代表的概念的一个例子).

A total of 1,212 trials were presented, randomly ordered with the constraint that the same words or pictures were not repeated for three consecutive trials. The first 12 trials were practice trials. For 500 trials, pictures were preceded by a category-level label and for 700 trials pictures were preceded by an item-level label. Half of all trials required a Yes response (e.g., category-level label VEHICLE 车 followed by a picture of car; or item-level label CAR 车 followed by a picture of car) and half required a No response (e.g., category label VEHICLE 车 or item level CAR 车 followed by a picture of eggplant). Half of the 250 “Yes” category-level trials (e.g., VEHICLE 车) were pictures of atypical items (e.g., a train), and half were pictures of typical items (e.g., a car). The experimental session lasted ~90 min.

**EEG recording**

The recording equipment and procedures were nearly identical across the two laboratories, except for the display monitor (12” refresh rate 75 Hz in US and 14” refresh rate 85 Hz in China) and recording software (Neuroscan 4.0 in
US and Neuroscan 4.3 in China). The EEG for both sites was recorded using Ag/AgCl electrodes embedded in a nylon mesh cap (21 scalp sites Easy-Cap, Falk Minow Systems, Herrsching-Breitbrunn, Bavaria, Germany) with a left mastoid reference and a forehead ground. An average mastoid reference was derived off-line. The vertical electrooculogram (VEOG) was recorded with electrodes placed above and below the left eye and the horizontal electrooculogram (HEOG) on the outer canthi of both eyes. All interelectrode impedance was maintained below 5 kΩ. The

**Figure 2.**

(A) English- and Mandarin-speaking participants’ mean accuracy and RT data (error bars show 2 SE) in Study 1 for typical and atypical pictures in response to category-level labels. Reaction time data is presented only for correct responses. (B) Mandarin participants’ mean accuracy and RT in Study 2 for typical and atypical pictures with orthographically and morphologically transparent items. (C) English participants’ mean accuracy and RT in Study 3 for typical and atypical pictures with nontransparent and morphologically transparent items. *P < 0.05 **P < 0.01 ***P < 0.001 for Bonferroni corrected post-hoc comparisons of conditions.
EEG and EOG were amplified using a 0.1–100-Hz band-pass filter and continuously sampled at 500 Hz/electrode for off-line analysis with a SynAmps data acquisition system (Neuroscan Labs, Sterling, VA). EEG data were corrected for ocular movement artifacts using the Gratton algorithm [Gratton et al., 1983]. Prior to analysis the data were filtered with a 9-point Chebyshev type II low-pass zero-phase shift digital filter (Matlab 7.0, Mathworks, Natick, MA), with a half-amplitude cutoff at 12 Hz.

Results

Behavioral results

Trials with a response time >1,200 ms or <200 ms were excluded as outliers (English: 2,487 trials, 7.97% of all responses; Mandarin: 2,203 trials, 6.79% of all responses). A typicality (typical vs. atypical) by language (English vs. Mandarin) repeated measures ANOVA with Bonferroni corrections for post-hoc analyses was conducted for both accuracy and RT data to explore the effect of typicality and language in the Yes responses.

As found in numerous previous studies, participants made more errors and responded more slowly when shown pictures of atypical than typical members of a category (Accuracy: $M = 0.67$ and $0.97$, $F(1, 51) = 270.37$, $P < 0.001$; RT: $M = 675.37$ and $616.44$, $F(1, 51) = 142.80$, $P < 0.001$, respectively). However, a significant typicality by language interaction was also present for the accuracy ($P = 0.042$) and reaction time data ($P = 0.004$) indicated that the typicality effect was attenuated for Mandarin speakers (Fig. 2A).

ERP results

The P160, N300, and N400 ERP components were quantified as the positive peak amplitude in the 140–240 (P160) range, and the negative peak amplitude in the 240–340 ms (N300) and the 370–470 ms (N400) range, respectively. In addition, the LPC component was calculated as the mean positive amplitude in the 500–700-ms interval. All epochs were measured from the onset of the target picture to 800-ms later, relative to a 100 ms prestimulus baseline.

On the basis of the previous studies of N400 responses and the typicality effect for pictorial stimuli [Federmeier and Kutas, 2001; Fujihara et al., 1998; Ganis et al., 1996; Heinze et al., 1998; Stuss et al., 1988; West and Holcomb, 2002] as well as the scalp topography of the difference waves (atypical–typical) for the category-level labels (Fig. 3), we focused our analysis on the horizontal line encompassing the bilateral frontal electrodes (F7, F3, Fz, F4, and F8). Voltage data for peak amplitude of the P160, N300, and N400 components from two left-right pairs, F3–F4 and F7–F8, and mean amplitude for the LPC component for the Yes responses were used in a typicality (typical vs. atypical) by side (left [F3 or F7] vs. right [F4 or F8]) by language (English vs. Mandarin) repeated measures ANOVA with Bonferroni corrections for post-hoc analyses (Fig. 3).

We focus here on the results for the N300 and N400 components.

For both components, main effects of side and language were significant for both the F3–F4 and F7–F8 pairs. The left side elicited a larger negative peak than the right side (F3–F4 pair $F(1,51) = 27.82$, $P < 0.001$ and F7–F8 pair $F(1,51) = 22.01$, $P < 0.001$ for N300 and $F(1,51) = 31.70$ and $27.14$, $P < 0.001$ for N400). In addition, English-speaking participants elicited larger N300s and N400s than Mandarin-speaking participants (F3–F4 pair $F(1,51) = 5.39$, $P = 0.024$ and F7–F8 pair $F(1,51) = 5.10$, $P = 0.028$, for N300 and $F(1,51) = 9.89$, $P = 0.003$ and $F(1,51) = 5.39$, $P = 0.024$ for N400). Moreover, the F3–F4 pair also showed a significant typicality by language interaction for both the N300 and N400 components, $F(1,51) = 7.33$, $P = 0.009$ and $F(1,51) = 12.33$, $P = 0.001$, respectively, such that the difference between atypical and typical items was larger in English- than Mandarin-speaking participants. Interestingly, both English and Mandarin speakers showed similar and significant typicality effects in the LPC component such that atypical items showed larger LPC amplitude at the Fz electrodes than typical items (see Fig. 3). No systematic P160 differences were found between typical and atypical items in either language.

Discussion

Despite overall similar behavioral results (albeit attenuated for Mandarin speakers) (Fig. 2A), the ERP results in Study 1 showed dramatic differences in the English and Mandarin speakers’ processing of typical vs. atypical items. For English speakers, atypical items elicited larger N300 and N400 components than typical items (see Fig. 3), consistent with several previous studies in English, German, and Japanese [Fujihara et al., 1998; Heinze et al., 1998; Stuss et al., 1988]. In contrast, Mandarin speakers did not show a typicality effect for either the N300 or N400 components, although they still showed a similar medial frontal LPC (see Fig. 3), which might contribute to decision making and evaluative processes or to some sort of post-semantic process such as those found by West and Holcomb [2002] when the goal-related expectations of a pictorially presented story were violated. In our study, the pictorial stimuli were relatively simple, but participants had to decide whether the picture was a “member of the category” with either item- or category-level labels. The fact that the atypical items took longer and also elicited greater positivity across cultures suggests that there were indeed post-lexical processes that may have been related to a final decision process that was more taxing for atypical than for typical items for both groups of speakers. Importantly, when we...
simplified the decision processes by only including category-level judgment in Studies 2 and 3 (see below), this effect either diminished or disappeared altogether.

Nonetheless, the focus in this study is on the role of typicality in category-level decisions for speakers of a language that contains category information even in atypical item labels (Mandarin) vs. speakers of a language that does not contain category information in the item level (English). The absence of a typicality effect in the N300 and N400 ERP components is not something that has been reported in previous studies and is thus a unique and intriguing aspect of the present study. The Chinese speakers, who have category-level information embedded in the names of the items, appear not to rely on typicality during semantic access reflected by the N300 and N400 components. Nonetheless, there were important differences in the numbers of morphologically transparent vs. nontransparent items that were used across the two languages (as a result of their natural frequencies in each language since identical pictures were used) and it is possible that the differences in the language-specific labels for the pictures, rather than the use of morphological transparency as an organizing feature of categories in Mandarin vs. English per se were responsible for the cross-linguistic differences that we found. Study 2 and 3 control these variations in each language separately by adopting additional stimuli. Study 2 controls for the explicitness of linguistic information in Mandarin by using pictures with orthographically vs. morphologically transparent labels, and thus asks whether the typicality effect in Mandarin speakers is equally absent for orthographically (vs. morphologically) transparent items. Study 3 specifically controls for the explicitness of linguistic information in English by using pictures with morphologically transparent vs. nontransparent labels, and thus asks whether the typicality effect would be similarly reduced in English speakers when provided with pictures that have explicit category information embedded morphologically in their verbal labels.

**STUDY 2**

**Method**

**Participants**

Twenty-two (13 females, Mean age = 22.10, SD = 1.97) native Mandarin-speaking undergraduates in Beijing, all right-handed undergraduates with normal vision, participated in this study and received payment.

**Stimuli**

To generate new categories that included more Mandarin items with orthographically transparent items, we first selected three to four typical or atypical items for three candidate categories, BIRD, STONE, and SHIP (e.g., for BIRD: rooster gongji1 公鴨, duck yai1z3 鴨子, pigeon qi3e2 鶴, pigeon gelizi 鵝子) and created two to three corresponding grayscale pictures for each item. To select more typical and atypical items and their corresponding pictures for each category, we conducted two more pilot studies. In Pilot Study 3, we used a naming task in which participants were asked to choose a name to describe each of these new pictures. Twenty native Mandarin-speaking undergraduates in Beijing participated and received souvenir pens for participating. Using these results, we then removed pictures that received <60% naming accuracy to the intended item name. This is a conservative criterion since we were looking for exact hits and many nonhits were very close in meaning with similar orthographic or morphological structures as the intended names. In Pilot Study 4, we asked participants to rate typicality for all remaining item pictures using the same scale (1–6) as in Pilot Study 2. Twenty-three native Mandarin-speaking undergraduates in Beijing participated and received souvenir pens. Using these results, we then discarded items that had intermediate or below-threshold typicality ratings and kept the two items for each category with the lowest and highest typicality ratings (Appendix 2). Finally, we selected six pictures for these three new orthographical categories: BIRD niu03 牛 STONE shi2- tou2 石头, SHIP chuan2 guo3 (Fig. 1C, Appendix 2). In our final stimulus set, we had five morphologically transparent and five orthographically transparent categories for Study 2, each with one typical and one atypical picture (Fig. 1C).

**Procedure and task**

The procedure, apparatus, and task of Study 2 were the same as for Study 1 except that only category-level trials were included so that we could reduce the testing time and task difficulty. For practical reasons, we also changed the ERP recording system from Neuroscan with 21 scalp sites (Easy-Cap) to EGI with a 128-channel Geodesic sensor Net and EGI NetStation 4.1.

A total of 412 trials were presented in pseudo-random order to each participant. The first 12 trials were practice trials. As with Study 1, half of all trials were correct that required a Yes response (e.g., category label VEHICLE che1 車, followed by a picture of a car) and half were wrong that required a No response (e.g., label VEHICLE che1 車 followed by a picture of eggplant). Among the 200 Yes trials, half were typical and half were atypical, and this was crossed with Label type (morphological vs. orthographical), yielding 50 trials for each condition (e.g., typical morphological items). The experimental session lasted ~25–30 min.

**EEG recording**

The electroencephalogram (EEG) was recorded using a 128-electrode Geodesic Sensor Net. The EEG signal was amplified using a 0.01–100-Hz bandpass and digitized at 500 Hz. The electro-oculogram (EOG) was monitored with 6 electrodes placed bilaterally in the external canthi (128 and 125), supraorbital (26 and 8), and infraorbital (127 and 126).
Figure 3 (legend on page 1794).

Figure 4 (legend on page 1794).
126) regions. Impedances for each electrode were measured prior to recording and kept below 50 kΩ during testing. Recording in every electrode was vertex-referenced. Data were recorded and processed using Net Station 4.1 (EGI software).

After acquisition, the data were lowpass filtered below 20 Hz. The continuous EEG was segmented into an epoch starting at 100 ms before the onset of the stimulus and lasting until 800 ms after stimulus onset. Segmented files were scanned for artifacts with the Artifact Detection toolbox in NetStation 4.1 using a threshold of 70 μV for excessive muscular activity, eye blinks, and eye movements. Segments containing eye blinks or movements as well as segments with more than 20 bad electrodes were rejected. Within each segment, electrodes with either an average amplitude of greater than 200 μV or difference average amplitude of 100 μV were also discarded from further processing. Finally, particular electrodes were rejected if they contained artifacts of any kind in more than 50% of the segments. Artifact-free segments for correct responses were averaged separately for Yes and No trials over the 800-ms epoch across subjects and referenced against the average of all electrodes. To compare data across studies, we also conducted a second set of analyses with an average mastoid referencing procedure. Results from both methods were nearly identical, but the average of all electrodes method will be presented as our primary findings since this is the method most appropriate for the EGI system and is a better representation of a true zero for the Geodesic Sensor Net [Junghofer et al., 1999]. Deviations across referencing methods will be noted where relevant. The 100 ms preceding the target served as baseline.

Results

Behavioral results

Trials with a response time >1,200 ms or <200 ms were cut-off as outliers (304 trials, 3.6% of all responses). Figure 2B shows the accuracy and RT data to pictures with morphologically vs. orthographically transparent labels and atypical vs. typical items in Yes responses. A typicality (typical vs. atypical) by label type (morphological vs. orthographic) ANOVA found that the main effects of typicality and label type were significant specifically for both RT and accuracy data. Participants made more errors, \( F(1,21) = 16.12, P < 0.001 \) and responded more slowly, \( F(1,21) = 37.23, P < 0.001 \) for atypical than typical items. They also made more errors, \( F(1,21) = 15.86, P < 0.001 \) and responded more slowly, \( F(1,21) = 17.09, P < 0.001 \) for orthographically transparent items than morphologically transparent items. A significant typicality by label type interaction was also found in the accuracy data only, \( F(1,21) = 13.39, P = 0.001 \), such that the typicality effect was larger for the orthographically transparent items than morphologically transparent items.

ERP results

The P160, N300, and N400 ERP components were quantified as the negative or positive peak amplitude in the 130–190 ms (P160), 240–340 ms (N300), and the 370–470 ms (N400) range, respectively. In addition, the LPC component was calculated as the mean amplitude in the 500–700-ms interval. All epochs were measured following the onset of the target picture, relative to a 100 ms prestimulus baseline.

As in Study 1, we focused our analysis on the horizontal line encompassing the bilateral frontal electrodes (F5, F7, FCz, F6, and F8, corresponding to electrodes 34, 28, 6, 123, 122 in the Geodesic sensor Net, respectively) (EGI software). Voltage data for peak amplitude of the P160, N300, and N400 components from two left-right pairs, F5–F6 and F7–F8 for the Yes responses were used in a typicality (typical vs. atypical) by side (left [F5 or F7] vs. right [F6 or F8]) by label type (morphological vs. orthographic) repeated measures ANOVA with Bonferroni corrections for post-hoc analyses (Fig. 4).

As can be seen in Figure 4, no significant typicality effects were found when combining both types of stimuli for both N300 and N400 components (\( F(1,21) = 0.183 \) and N400 differences were not found for Mandarin-speaking participants. Both English and Mandarin speaking-participants also show a strong typicality effect in the middle frontal electrode (Fz) for the LPC component, for which Mandarin speaking participants showed a more widespread distribution in the bilateral frontal electrodes (F3, F4, F7). * \( P < 0.05 \) ** \( P < 0.01 \) *** \( P < 0.001 \) for Bonferroni corrected post-hoc comparisons.

Figure 3.

Figure 4.

ERP waves and scalp topographies show divergent difference waves (Atypical-Typical, Correct Yes responses only) for pictures of items as judged by (top) English-speaking participants and (bottom) Mandarin-speaking participants. English-speaking participants show a strong typicality effect in the left frontal electrodes (F3, F7) for the N300 component and bilateral frontal electrodes (F3, F7, F4, F8) for the N400 component. In contrast, the expected N300 and strong typicality effect in the left frontal electrodes (F5, F7) when viewing orthographically transparent items, but no differences when viewing morphologically transparent items for both N300 and N400 components. No significant typicality effect was found when combine them together. * \( P < 0.05 \) ** \( P < 0.01 \) *** \( P < 0.001 \) for Bonferroni corrected post-hoc comparisons.
significant differences between typical and atypical items. A
whereas morphologically transparent items showed no sig-
.transparent items indeed showed left-lateralized typicality
further typicality (typical vs. atypical) by side (left [F5 or
[F7] vs. right [F6 or F8]) ANOVA for orthographically
order to be accessed [Liu et al., 2008; Siok et al., 2003, 2004;
ies on Chinese characters processing have found that ortho-
transparent labels may thus be more
spoken Chinese radicals is not pronounced and does
transparent nouns. Since the orthographic information em-

cie judgments for
pictures are used, different label types influence Mandarin
speakers, orthographically transparent items revealed larger
typicality effects in both behavioral and ERP results than morphologically transparent items. For Mandarin
speakers, orthographically transparent items showed a sig-
ificant typicality effect for both the N300 and N400 compo-
nents in the left frontal electrodes (see Fig. 4).

These results suggest even when category judgments for
pictures are used, different label types influence Mandarin
speakers’ use of typicality to make these judgments. Although orthographically transparent nouns provide cate-
gory information in Mandarin, it appears to be less accessi-
ble than the information provided by morphologically
transparent nouns. Since the orthographic information em-
bedded in Mandarin radicals is not pronounced and does
not also provide phonological information when the label of
an item is accessed, the category information provided by
orthographically transparent labels may thus be more
implicit, and also does not become available for use until
Mandarin speakers become fluent readers (i.e., not from the
beginning of productive language). In addition, fMRI stud-
ies on Chinese characters processing have found that ortho-
 graphical information in Chinese characters may also
require additional orthographic-to-semantic mappings in
order to be accessed [Liu et al., 2008; Siok et al., 2003, 2004;
Tan et al., 2005] and thus may need more semantic processing
to process than morphologically transparent items. However, it is not clear whether orthographic information
might still confer an additional advantage, albeit very small,
relative to a completely nontransparent item. This was not
tested and would perhaps be undetectable in the current
design, but one that is worthy of future study, particularly
given that the differences between the morphologically and
orthographically transparent items in the present study
were significant, but relatively small compared to the differ-
ences between English and Mandarin in Study 1.

Nonetheless, this label type difference in Mandarin is
similar to the cross-cultural difference found in Study 1. Just
as nontransparent English nouns that provide no category
information need more semantic processing in order to
make category judgments, orthographically transparent
Mandarin items that provide less salient linguistic category
information need more semantic processing than morpho-
logically transparent items. Relying on typicality, for less
transparent items, is thus a useful way to reduce the amount
of semantic processing required. Because morphological
transparency, in Mandarin, is both a regular feature of the
language and it provides explicit and solid linguistic cues to
category membership, typicality is not needed for initial
category access and thus morphologically transparent items in
Mandarin do not show a typicality effect.

Our next question then is whether or not morphologically
transparent items in English (e.g., catfish) facilitate the cate-
gorization processes in English speakers. Although this type
of transparency does occur in English, it is not as productive
and regular as it is in Mandarin [Tardif, 2006; Zhou et al.,
1999]. Thus we are able to use English to distinguish
between one of two possibilities. First, our results in Man-
darin may simply be an immediate effect of the relation
between a picture’s label and its category; if so, we might
expect that English speakers can also rely on the explicit cat-
gory information provided by morphologically transparent
items and thus show a reduction in the typicality effect just
like Mandarin speakers when pictures of morphologically
transparent items are used in a category judgment task.
However, if it is not only an immediate effect of label type,
but also the conventions of a language that play a role, then
English speakers might not be able to extract such linguistic
information as efficiently as Mandarin speakers, who have
presumably used such linguistic cues in their implicit proc-
essing of language since they first began to understand and
produce words [Tardif, 2006].

STUDY 3
Method

Participants
Thirty-three native English speakers in Ann Arbor, MI,
all right-handed undergraduates with normal vision, par-
ticipated for course credit. Eight participants were
Behavioral results

ANOVA revealed that participants made more errors, $F(1,24) = 6.07, P = 0.021$ and responded more slowly, $F(1,24) = 13.94, P = 0.001$, for atypical items than typical items. In addition, English speakers were slightly faster to categorize exemplars that contained morphologically transparent items than those that did not, $F(1,24) = 6.20, P = 0.020$, thus suggesting that even for English, morphological transparency can convey a slight advantage during categorization. However, unlike the cross-linguistic comparisons in Study 1, no interactions were observed between typicality and the morphological transparency conditions for either the RT or accuracy data (Fig. 2C).

Stimuli

To generate new categories that included more English items with morphologically transparent items, we followed the procedures of Pilot Studies 3 and 4. First we selected three to four typical or atypical items for eight candidate categories, PHONE, POOL, BAG, BOOK, BALL, CHAIR, STATION, and PAPER (e.g., BALL: basketball, football, soccer ball, and baseball) and produced corresponding grayscale pictures for each item. In Pilot Study 5, 29 native English-speaking undergraduates provided labels for and rated the typicality of all pictures on a scale from 1 to 6. On the basis of these results, we then discarded items that had poor label agreements and/or intermediate typicality ratings and kept the two items for each category with the lowest and highest typicality ratings (Appendix 3).

For Study 3, we selected four of these morphologically transparent categories (BALL, BOOK, CHAIR, PHONE), each with a typical and an atypical item, and combined them with the one morphologically transparent category (PAPER) and five nontransparent categories used in Study 1, to produce a total of five morphologically transparent and five nontransparent categories (Fig. 1D, Appendix 3).

Procedure and task

The procedure, apparatus, and task of Study 3 were the same as for Study 2. A total of 812 trials were presented pseudo randomly to each participant as in Study 2. The first 12 trials were practice trials. Half of all trials required a Yes response and half required a No response. Among the 400 “Yes” trials, half were typical and half were atypical, and this was crossed with label type (morphological vs. nontransparent), yielding 100 trials for each condition (e.g., typical morphological items). The experimental session lasted ~45–50 min.

EEG recording

The electroencephalogram (EEG) recording and analyses procedures and equipment for Study 3 were identical to Study 2. The study was conducted in the US with identical EGI equipment and software as the laboratory in Beijing.

Results

Behavioral results

Trials with a response time >1,200 ms or <200 ms were excluded as outliers (774 trials, 3.9% of all responses). As with Study 2, a Label type (morphological vs. nontransparent) by typicality (typical vs. atypical) repeated measures ANOVA revealed that participants made more errors, $F(1,24) = 6.07, P = 0.021$ and responded more slowly, $F(1,24) = 13.94, P = 0.001$, for atypical items than typical items. In addition, English speakers were slightly faster to categorize exemplars that contained morphologically transparent items than those that did not, $F(1,24) = 6.20, P = 0.020$, thus suggesting that even for English, morphological transparency can convey a slight advantage during categorization. However, unlike the cross-linguistic comparisons in Study 1, no interactions were observed between typicality and the morphological transparency conditions for either the RT or accuracy data (Fig. 2C).

ERP results

The P160, N300, N400, and LPC ERP components were quantified the same way as in Study 2 and analyzed using the same software.

As in Studies 1 and 2, we focused our analysis on the horizontal line encompassing the bilateral frontal electrodes (F7, F5, Fcz, F6, and F8, corresponding to 34, 28, 6, 123, 122 in the Geodesic Sensor Net, respectively). Voltage data for peak amplitude of the P160, N300, and N400 components from two left–right pairs, F5–F6 and F7–F8 for the “Yes” responses were used in a typicality (typical vs. atypical) by side (left [F5 or F7] vs. right [F6 or F8]) by label type (morphological vs. nontransparent) repeated measures ANOVA with Bonferroni corrections for post-hoc analyses (see Fig. 5). As with Study 2, we focus here on the N300 and N400 components. As can be seen in Figure 5, significant typicality effects were found for both types of stimuli, but laterality interacted with the type of information provided in the stimulus label—the typicality effect for the nontransparent items appeared in left-side scalp electrodes whereas the typicality effect for the morphologically transparent items appeared in right-side scalp electrodes and this was true for both the N300 and N400 components.

Atypical items elicited a larger N300 (F5–F6 pair $F(1,24) = 15.21$ and F7–F8 pair $F(1,24) = 15.22, P < 0.001$) and N400 (F5–F6 pair $F(1,24) = 8.46, P < 0.01$ and F7–F8 pair $F(1,24) = 14.51, P < 0.001$) than the typical items. In addition, both the F5–F6 and the F7–F8 pairs showed significant typicality by label type by side interactions for both the N300 ($F(1,24) = 19.28$ and 30.97, $Ps < 0.001$) and N400 ($F(1,24) = 5.45$ and 5.44, $Ps < 0.05$) components, such that the nontransparent items showed a significant typicality effect only on the left side electrodes, whereas morphologically transparent items showed a significant typicality effect only on the right side electrodes (Fig. 5).

The LPC component showed no significant main effects of typicality or typicality by label type interaction, which was true also for the average mastoid referencing results.

Discussion

Although the behavioral results in Study 3 were almost identical to those for the English speakers in Study 1, the ERP results revealed some interesting differences in the typicality effect for nontransparent vs. morphologically
ERP waves and scalp topographies show divergent difference waves (atypical–typical, correct Yes responses only) for pictures of (top) both nontransparent and morphologically transparent, (middle) nontransparent only, and (bottom) morphologically transparent only typical vs. atypical items after viewing category-level labels. English-speaking participants showed strong typicality effect in the left frontal transparent items for English speakers. Nontransparent atypical items elicited larger N300 and N400 components than typical items only in the left frontal electrodes, whereas for the morphologically transparent nouns, this difference was apparent in the right frontal electrodes, as shown in Figure 5. The laterality of the typicality effect for English nouns is intriguing, but it is not clear from these data alone whether it truly reflects differences in right vs. left hemisphere processing of morphologically transparent vs. nontransparent nouns, or whether there are some other differences contributing to the appearance of this effect at different electrode sites. Nonetheless, as has been found in previous studies and Study 1 and 3, the increased activation for atypical nouns indicates increased semantic processing [Fujihara et al., 1998; Heinze et al., 1998; Poldrack et al., 1999; Stuss et al., 1988]. In English, this is particularly prominent for pictures of nouns that do not contain any linguistic cues to category information, which is the predominant pattern in English. In contrast, when pictures of items that have morphologically transparent cues to the category are provided, even English speakers appear to make use of these cues to facilitate semantic access and category judgments, as evidenced by the decrease in RTs for morphologically transparent items. However, since most English nouns are nontransparent rather than linguistically transparent like Mandarin nouns, providing pictures of morphologically transparent items is not enough to allow English speakers to circumvent the use of typicality as an aid to categorization. As a result, additional executive attention load [Corbetta and Shulman,
GENERAL DISCUSSION

Our ERP data suggest that language has a direct impact on categorization processes. Speakers of different languages show different patterns of reliance on typicality during category judgment tasks. Moreover, within a language such as Mandarin, objects that have category labels embedded in their names also show different reliance on typicality during these same types of category judgment tasks. These results are impressive not only because they show the impact of language on categorization both between and within languages, but also because we provided pictures of the objects and not their labels. Thus, the effects of language on categorization hold even with pictorial stimuli. In addition, the absence of a typicality effect in the N300 and N400 ERP components has not been reported in previous studies and is thus a unique and intriguing aspect of the present study.

In the present study, the category judgment task required that the participant first read and keep in mind a category label (e.g., vehicle) and then judge whether a picture (e.g., a sedan), shown 1,500-ms later, was an example of the label. Because participants ultimately had to make a link between the visual characteristics of the picture and the linguistic stimulus shown before it, we assume that all participants, both in China and in the US, engaged in some sort of semantic access before the final decision was made. Of great interest, however, is whether participants in the two cultures engaged in the same type or level of semantic access and/or whether they engaged in additional semantic processing of both the pictorial stimulus and the word label (e.g., car jiao4che1 辆车) given to the stimulus. Our assumption, based on the present data, is that speakers of English not only accessed a verbal label for the pictures, but that they engaged in additional semantic processing, evidenced by the presence of a typicality effect at both the N300 and N400 ERP components for both morphologically transparent (attenuated) and nontransparent items, to facilitate judgments in this task. In contrast, speakers of Mandarin Chinese were able to bypass this additional semantic processing because of the presence of the category name (e.g., vehicle chêl 车) in the common morphological (e.g., car jiao4che1 辆车) label that speakers accessed even when shown a relevant picture. However, they were not able to fully bypass this when given pictures with orthographically transparent information which, as discussed above, is available for use much later in development and does not share the additional phonological cues provided by morphologically transparent nouns.

ACKNOWLEDGMENTS

The authors thank Yanni Liu, Liming Zhou, Sheila Wang, Johnny Parker, Jiayin Wu, and Dr. Yaxu Zhang (Peking University) for help in conducting these studies. They also thank Drs. Julie Boland, Henry Wellman, and Wutian Zhang for comments. The recruitment of participants was approved by IRBs at the University of Michigan (B04-00001580-M1), the Beijing Normal University and the Institute of Psychology, Chinese Academy of Sciences.

REFERENCES


APPENDIX 1. Category-level typicality ratings (1-6) in Pilot Study 2 show similarities across languages for Typical vs. Atypical items for the ten categories used in Study 1

<table>
<thead>
<tr>
<th>Category</th>
<th>Typical</th>
<th>Atypical</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mandarin</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SHOES</td>
<td>loafers</td>
<td>5.33</td>
</tr>
<tr>
<td>PANTS</td>
<td>trousers</td>
<td>5.38</td>
</tr>
<tr>
<td>VEHICLE</td>
<td>car</td>
<td>5.52</td>
</tr>
<tr>
<td>WRITING INSTRUMENT</td>
<td>pencil</td>
<td>5.00</td>
</tr>
<tr>
<td>STATION</td>
<td>train station</td>
<td>4.74</td>
</tr>
<tr>
<td>BUG</td>
<td>fly</td>
<td>4.74</td>
</tr>
<tr>
<td>BUILDING</td>
<td>office building</td>
<td>5.75</td>
</tr>
<tr>
<td>OIL</td>
<td>engine oil</td>
<td>4.39</td>
</tr>
<tr>
<td>VEGETABLE</td>
<td>celery</td>
<td>5.30</td>
</tr>
<tr>
<td>PAPER</td>
<td>writing paper</td>
<td>5.13</td>
</tr>
<tr>
<td>English</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SHOES</td>
<td>loafers</td>
<td>5.30</td>
</tr>
<tr>
<td>PANTS</td>
<td>trousers</td>
<td>5.85</td>
</tr>
<tr>
<td>VEHICLE</td>
<td>car</td>
<td>5.96</td>
</tr>
<tr>
<td>WRITING INSTRUMENT</td>
<td>pencil</td>
<td>5.81</td>
</tr>
<tr>
<td>STATION</td>
<td>train station</td>
<td>5.77</td>
</tr>
<tr>
<td>BUG</td>
<td>fly</td>
<td>5.85</td>
</tr>
<tr>
<td>BUILDING</td>
<td>office building</td>
<td>5.85</td>
</tr>
<tr>
<td>OIL</td>
<td>engine oil</td>
<td>5.54</td>
</tr>
<tr>
<td>VEGETABLE</td>
<td>celery</td>
<td>5.59</td>
</tr>
<tr>
<td>PAPER</td>
<td>writing paper</td>
<td>5.89</td>
</tr>
</tbody>
</table>

Typical items (Chinese, $M = 5.13$, $SD = 0.39$; English, $M = 5.74$, $SD = 0.19$), Atypical items (Chinese, $M = 3.70$, $SD = 1.13$; English, $M = 3.90$, $SD = 0.63$). A Typicality (Typical vs. Atypical) by Language (English vs. Mandarin) ANOVA revealed only a main effect of Typicality, $F(1, 18) = 53.68$, $P < 0.001$. No significant effects of Language or Typicality by Language interactions were found.

APPENDIX 2. Category-level typicality ratings (1-6) from Pilot Study 3 show similarities across Label types for Typical vs. Atypical exemplar pictures for the ten categories used in study 2

<table>
<thead>
<tr>
<th>Category</th>
<th>Typical</th>
<th>Atypical</th>
</tr>
</thead>
<tbody>
<tr>
<td>Morphological</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SHOES $xie2zi$</td>
<td>loafers $pi2xie2$</td>
<td>5.33</td>
</tr>
<tr>
<td>PANTS $ku4zi$</td>
<td>trousers $xi1ku4$</td>
<td>5.38</td>
</tr>
<tr>
<td>VEHICLE $che1$</td>
<td>car $jiao4che1$</td>
<td>5.52</td>
</tr>
<tr>
<td>WRITING INSTRUMENT $bi3$</td>
<td>pencil $qian1bi3$</td>
<td>5.38</td>
</tr>
<tr>
<td>PAPER $zhi3$</td>
<td>writing paper $xiu4zhi3$</td>
<td>5.13</td>
</tr>
<tr>
<td>Orthographic</td>
<td></td>
<td></td>
</tr>
<tr>
<td>VEGETABLE $cai4$</td>
<td>celery $xi1qin2$</td>
<td>5.30</td>
</tr>
<tr>
<td>BUG $chong2zi$</td>
<td>fly $cang1ying$</td>
<td>4.74</td>
</tr>
<tr>
<td>BIRD $niao3$</td>
<td>pigeon $ge1zi3$</td>
<td>5.43</td>
</tr>
<tr>
<td>SHIP $chuan2$</td>
<td>warship $jun1jian4$</td>
<td>5.22</td>
</tr>
<tr>
<td>STONE $shi2tou2$</td>
<td>rock $yin2shi$</td>
<td>4.74</td>
</tr>
</tbody>
</table>

Typical items (Morphological, $M = 5.34$, $SD = 0.14$; Orthographic, $M = 5.09$, $SD = 0.32$), Atypical items (Morphological, $M = 4.28$, $SD = 0.69$; Orthographic, $M = 3.88$, $SD = 0.76$). A Typicality (Typical vs. Atypical) by Label type (Morphological vs. Orthographic) ANOVA revealed only a main effect of Typicality, $F(1, 4) = 22.97$, $P = 0.009$. No main effect of Label type or interactions between Typicality and Label type were found.
APPENDIX 3. Category-level typicality ratings (1-6) from Pilot Study 5 show similarities across Label types for Typical vs. Atypical exemplar pictures for the ten categories used in study 3

<table>
<thead>
<tr>
<th></th>
<th>Morphological</th>
<th></th>
<th>Nontransparent</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Typical</td>
<td>Atypical</td>
<td>Typical</td>
<td>Atypical</td>
</tr>
<tr>
<td>PAPER</td>
<td>writing paper</td>
<td>5.13</td>
<td>toilet paper</td>
<td>4.87</td>
</tr>
<tr>
<td>PHONE</td>
<td>cell phone</td>
<td>5.76</td>
<td>rotary phone</td>
<td>3.68</td>
</tr>
<tr>
<td>BALL</td>
<td>basketball</td>
<td>5.55</td>
<td>football</td>
<td>4.48</td>
</tr>
<tr>
<td>BOOK</td>
<td>textbook</td>
<td>5.13</td>
<td>notebook</td>
<td>3.48</td>
</tr>
<tr>
<td>CHAIR</td>
<td>folding chair</td>
<td>5.48</td>
<td>rocking chair</td>
<td>3.24</td>
</tr>
<tr>
<td>SHOES</td>
<td>loafers</td>
<td>5.30</td>
<td>slippers</td>
<td>3.15</td>
</tr>
<tr>
<td>PANTS</td>
<td>trousers</td>
<td>5.85</td>
<td>overalls</td>
<td>4.19</td>
</tr>
<tr>
<td>VEHICLE</td>
<td>car</td>
<td>5.96</td>
<td>train</td>
<td>3.59</td>
</tr>
<tr>
<td>WRITING INSTRUMENT</td>
<td>pencil</td>
<td>5.81</td>
<td>chalk</td>
<td>4.33</td>
</tr>
<tr>
<td>VEGETABLE</td>
<td>celery</td>
<td>5.59</td>
<td>eggplant</td>
<td>4.96</td>
</tr>
</tbody>
</table>

Typical items (Morphological, M = 5.45, SD = 0.07; Nontransparent, M = 5.67, SD = 0.09), Atypical items (Morphological, M = 3.67, SD = 0.15; Nontransparent, M = 4.03, SD = 0.18). A Typicality (Typical vs. Atypical) by Label type (Morphological vs. Nontransparent) ANOVA revealed only a main effect of Typicality, F (1, 4) = 35.60, P = 0.004. No significant main effect of Label type or interactions between Typicality and Label type were found.