

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

Chao Liu,<sup>1,2,3</sup> Twila Tardif,<sup>1,2\*</sup> Xiaoqin Mai,<sup>2,3</sup> William J. Gehring,<sup>1,2</sup>  
Nina Simms,<sup>4</sup> and Yue-Jia Luo<sup>5\*</sup>

<sup>1</sup>Department of Psychology, University of Michigan, Ann Arbor, Michigan

<sup>2</sup>Center for Human Growth and Development, University of Michigan, Ann Arbor, Michigan

<sup>3</sup>Key Laboratory of Mental Health, Institute of Psychology, Chinese Academy of Sciences, Beijing, China

<sup>4</sup>Department of Psychology, Northwestern University, Evanston, Illinois

<sup>5</sup>State 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* 车 in the noun “train” *huo3che1* 火车) 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

---

Contract grant sponsor: NSF; Contract grant number: BCS-0350272; Contract grant sponsor: NIH; Contract grant number: TW000035-07S1; Contract grant sponsor: Ministry of Education (PCSIRT); Contract grant number: IRT0710; Contract grant sponsor: NSF China; Contract grant number: 30930031; Contract grant sponsor: Center for Human Growth and Development.

\*Correspondence to: Twila Tardif, Department of Psychology, University of Michigan, Ann Arbor, MI 48109-1109, USA. E-mail: Twila@umich.edu; or Yue-Jia Luo, State Key Laboratory of Cogni-

tive Neuroscience and Learning, Beijing Normal University, Beijing 100875, China. E-mail: Luoyj@bnu.edu.cn

Received for publication 5 April 2009; Revised 21 August 2009; Accepted 22 November 2009

DOI: 10.1002/hbm.20974

Published online 26 March 2010 in Wiley Online Library (wileyonlinelibrary.com).

## 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; Chumbley, 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 *mei2gui4hua* 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., 1998; 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* 苍蝇), butterfly (*hu2die2* 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 1

### 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 categories for English and Mandarin speakers and that the category-level labels were equally appropriate “substitutions” for both languages, as these stimuli were to be used in subsequent studies. Finally, we wanted to also assess equivalences in labeling and substitutions at the item level for the two language communities. The bilingual and bicultural research team selected a total of 16 categories, each with 3–11 items and one or more possible category level labels. By design, both the Mandarin and English stimuli contained category items that were morphologically transparent and morphologically nontransparent. The Mandarin stimuli additionally contained items that were orthographically transparent. Twenty native speakers in

each location were then asked about the acceptability of replacing each term with another term in that same cluster. The questionnaire was organized by category, but neither the category- nor the item-level labels within a given category were explicitly identified as such and the category-level labels appeared in every possible position (first, second, etc.) across the 16 lists. For example, given the nouns fly *cang1ying1* 苍蝇, worm *qiu1ying3* 蚯蚓, bug *chong2zi* 虫子, and mosquito *wen2zi* 蚊子 (from the category BUG), a participant could say that “fly” can be used in place of “bug,” or that “fly” can be used in place of “mosquito,” and so on.

On the basis of the acceptability of replacement judgments from Pilot Study 1, we then eliminated those items or category labels that the majority of participants in one or the other group did not rate as acceptable “replacements” and found corresponding gray-scale photographs for the remaining two to eight items per category.

In Pilot Study 2, 29 English- and 24 Mandarin-speaking participants were asked to rate the typicality of each picture, given either the category- (e.g., vehicle/车) or item- (e.g., car/轿车) level label, on a six-point scale, with 1 representing not at all typical (完全不典型) and 6 representing extremely typical (极端典型). These ratings were then used to identify typical and atypical items for each 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

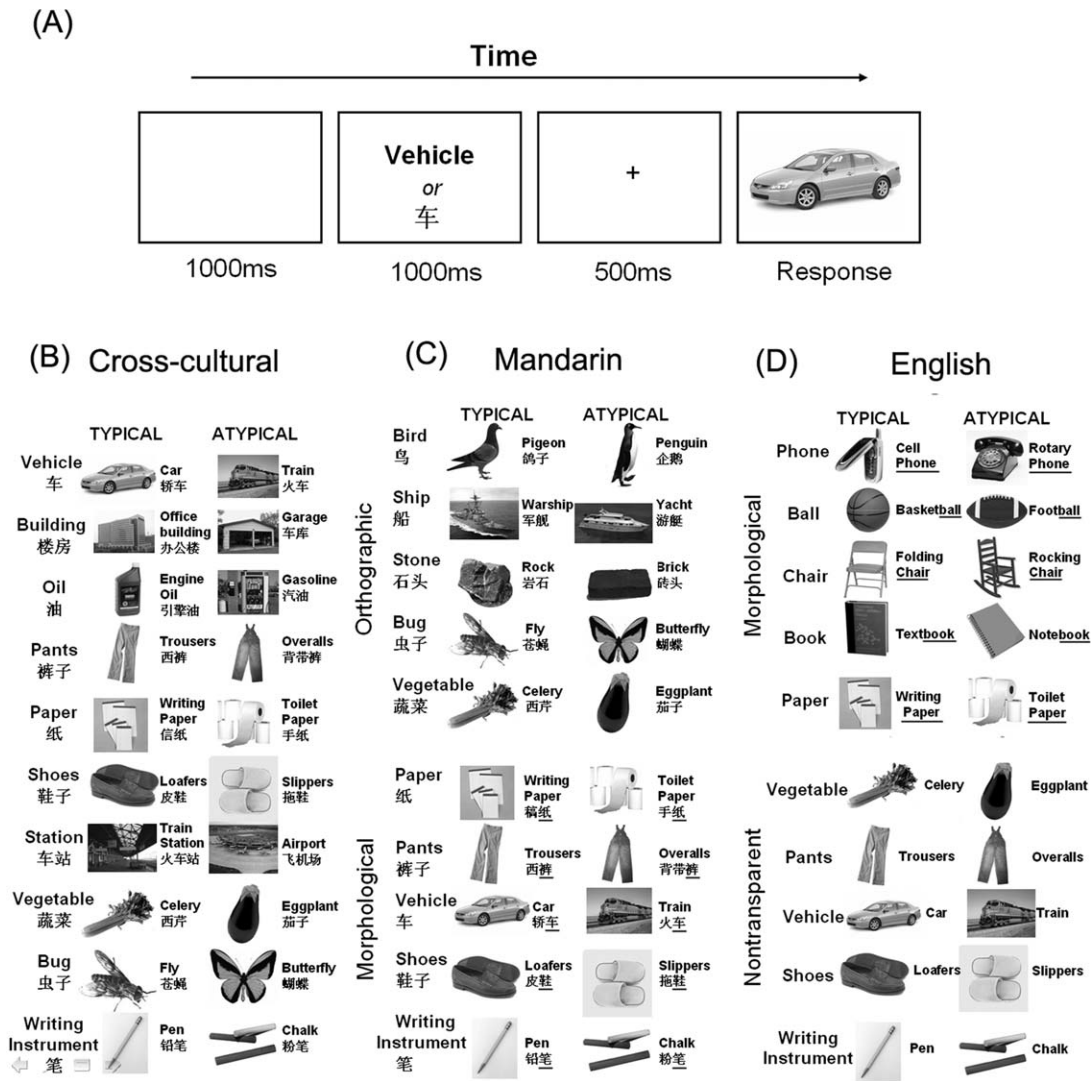


Figure 1.

(A) Experimental procedure and (B–D) Materials with labels and grayscale photographs in the three studies.

Mandarin) or an item-level label<sup>1</sup> (e.g., “CAR” in English or *jiao4che1* “轿车” 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” (“判断图片是否为其前面的词所代表的概念的一个例子”).

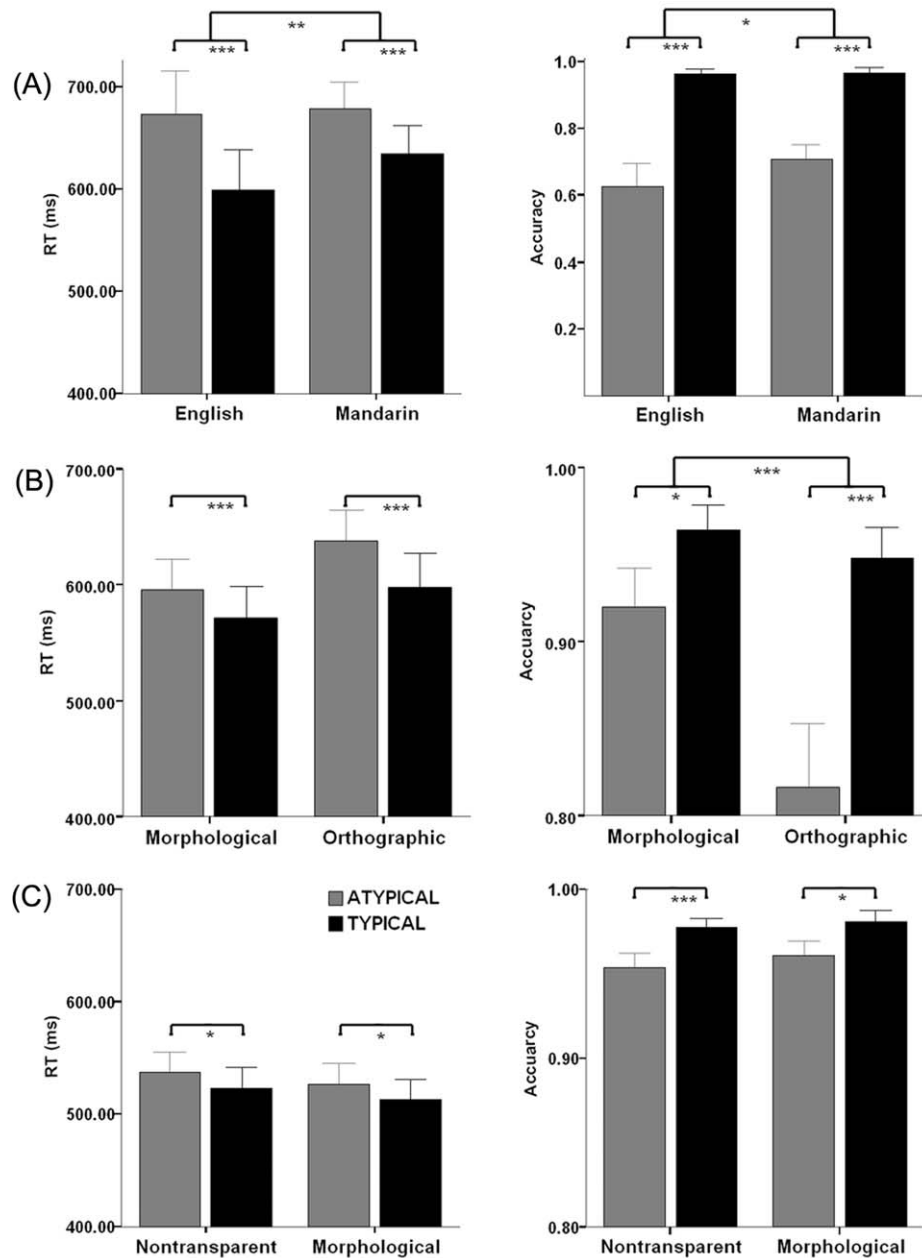
A total of 1,212 trials were presented, randomly ordered with the constraint that the same words or pictures were not repeated for three consecutive trials. The first 12 trials were practice trials. For 500 trials, pictures were preceded by a category-level label and for 700 trials pictures were preceded by an item-level label. Half of all trials required

a *Yes* response (e.g., category-level label VEHICLE *che1* 车 followed by a picture of car; or item-level label CAR *jiao4che1* 轿车 followed by a picture of car) and half required a *No* response (e.g., category label VEHICLE *che1* 车 or item level CAR *jiao4che1* 轿车 followed by a picture of eggplant). Half of the 250 “Yes” category-level trials (e.g., VEHICLE *che1* 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

<sup>1</sup>The data from item-level labels were not included in the current report, but it is important to note that no interactions between typicality and language were found.



**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 pic-

tures 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.

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 k $\Omega$ . The

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 in both 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  $F_s(1,51) = 31.70$  and  $27.14$ ,  $P_s < 0.001$  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. In fact, the typicality effect was almost completely absent from the N300 and N400 components in Mandarin speakers. This typicality by language interaction was also present for the F7–F8 pair in the N400 component,  $F(1,51) = 8.07$ ,  $P = 0.006$ , but did not reach significance for the N300.

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 *gong1ji1* 公鸡, duck *ya1z3* 鸭子, penguin

*qi3e2* 企鹅, pigeon *ge1zi* 鸽子) 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. On the basis of 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 *niao3* 鸟, STONE *shi2-tou2* 石头, SHIP *chuan2* 船 (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* 车, followed by a picture of car) and half were wrong that required a *No* response (e.g., label VEHICLE *che1* 车 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 six electrodes placed bilaterally in the external canthi (128 and 125), supraorbital (26 and 8), and infraorbital (127 and

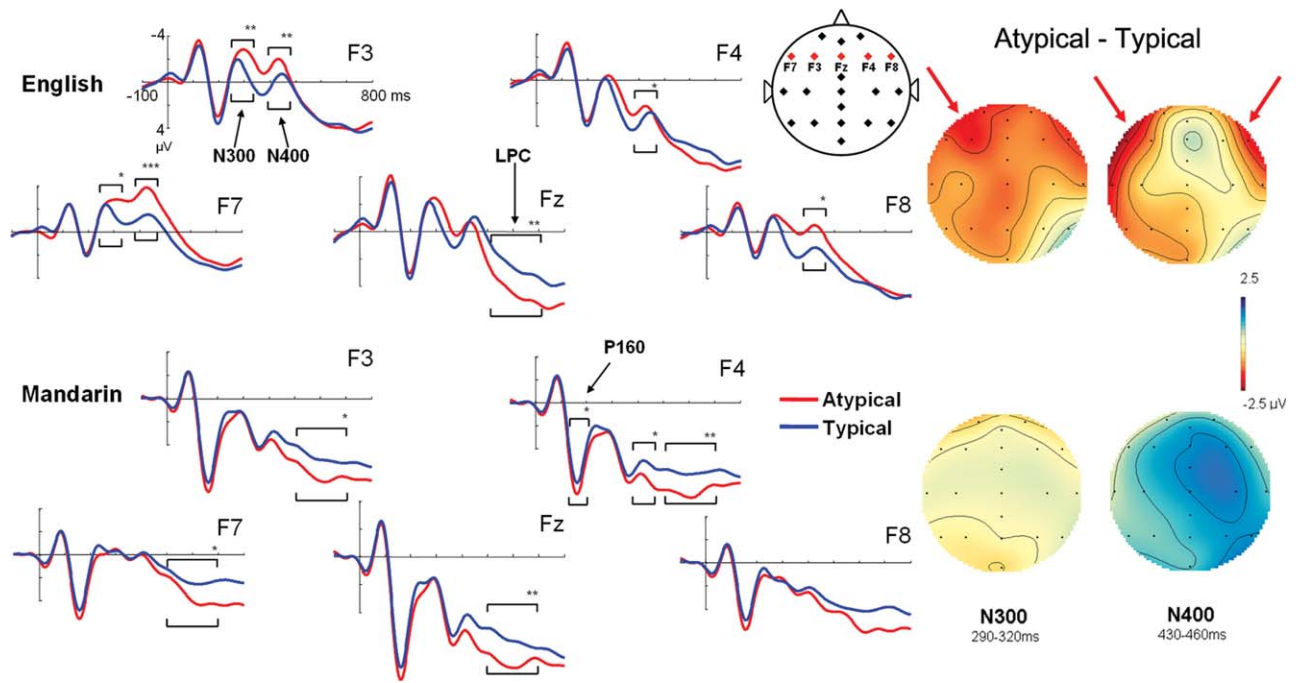


Figure 3 (legend on page 1794).

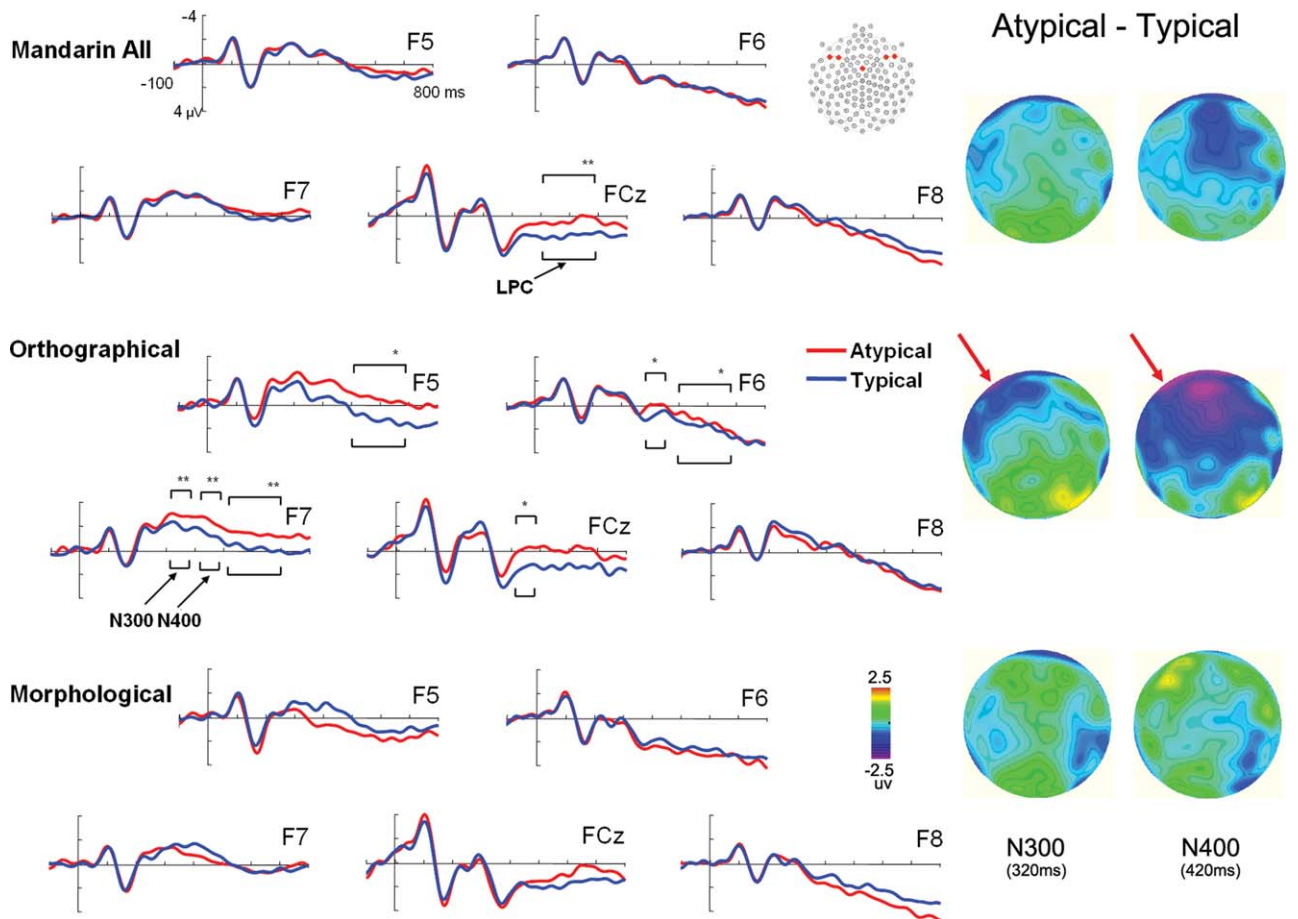


Figure 4 (legend on page 1794).



126) regions. Impedances for each electrode were measured prior to recording and kept below 50 k $\Omega$  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  $\mu$ V 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  $\mu$ V or difference average amplitude of 100  $\mu$ V 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 re-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 (F7, F5, 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

**Figure 3.**

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

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 4.**

ERP waves and scalp topographies show divergent difference waves (atypical–typical, correct Yes responses only) for pictures of (top) both orthographically and morphologically transparent, (middle) orthographically transparent only and (bottom) morphologically transparent only typical vs. atypical items after viewing category-level labels. Mandarin-speaking participants showed

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.

0.371,  $P_s > 0.549$ ), which repeated our finding in Study 1. However, the F5–F6 pair showed a marginally significant typicality by label type interaction for the N400 component,  $F(1,21) = 2.93$ ,  $P = 0.10$ , such that orthographically transparent items indeed showed left-lateralized typicality effects for both the N300 and N400 components (Fig. 4), whereas morphologically transparent items showed no significant differences between typical and atypical items.<sup>2</sup> A further typicality (typical vs. atypical) by side (left [F5 or F7] vs. right [F6 or F8]) ANOVA for orthographically transparent items only revealed that atypical items elicited a larger N400 (F5–F6 pair  $F(1,21) = 4.92$ ,  $P = 0.038$  and F7–F8 pair  $F(1,21) = 5.24$ ,  $P = 0.033$ ) than the typical items.

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

Interestingly, and consistent with the findings from Study 1, native Mandarin Chinese speakers did not show a typicality effect in the N300 and N400 components for morphologically transparent items. In contrast, and parallel to the cross-linguistic findings when we compared English vs. 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 significant typicality effect for both the N300 and N400 components 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 category information in Mandarin, it appears to be less accessible than the information provided by morphologically transparent nouns. Since the orthographic information embedded 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 studies on Chinese characters processing have found that orthographical 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;

<sup>2</sup>An exception was an almost inverse typicality effect in F5 (although it was not significant, N300:  $P = 0.547$ , N400:  $P = 0.119$ ). When we conducted a second set of analyses with an average mastoid referencing procedure for Morphologically transparent items, the main effect for typicality was still not significant,  $F(1,21) = 0.139$  and  $0.138$ , for N300 and N400, respectively,  $P_s > 0.713$ , but an interaction between label type and typicality was significant for the N400 effect at the F5–F6 pair,  $F(1,21) = 4.68$ ,  $P = 0.042$ .

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 differences 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 morphologically 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 categorization 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 Mandarin 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 category 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 processing 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, participated for course credit. Eight participants were

excluded from further analysis, four for poor behavioral performance (accuracy <80% in either *Yes* or *No* responses) and four for too many eye-blinks or artifacts in the electroencephalogram data. The final sample consisted of 25 participants (eight females,  $M$  age = 19.12,  $SD$  = 1.04).

### 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.

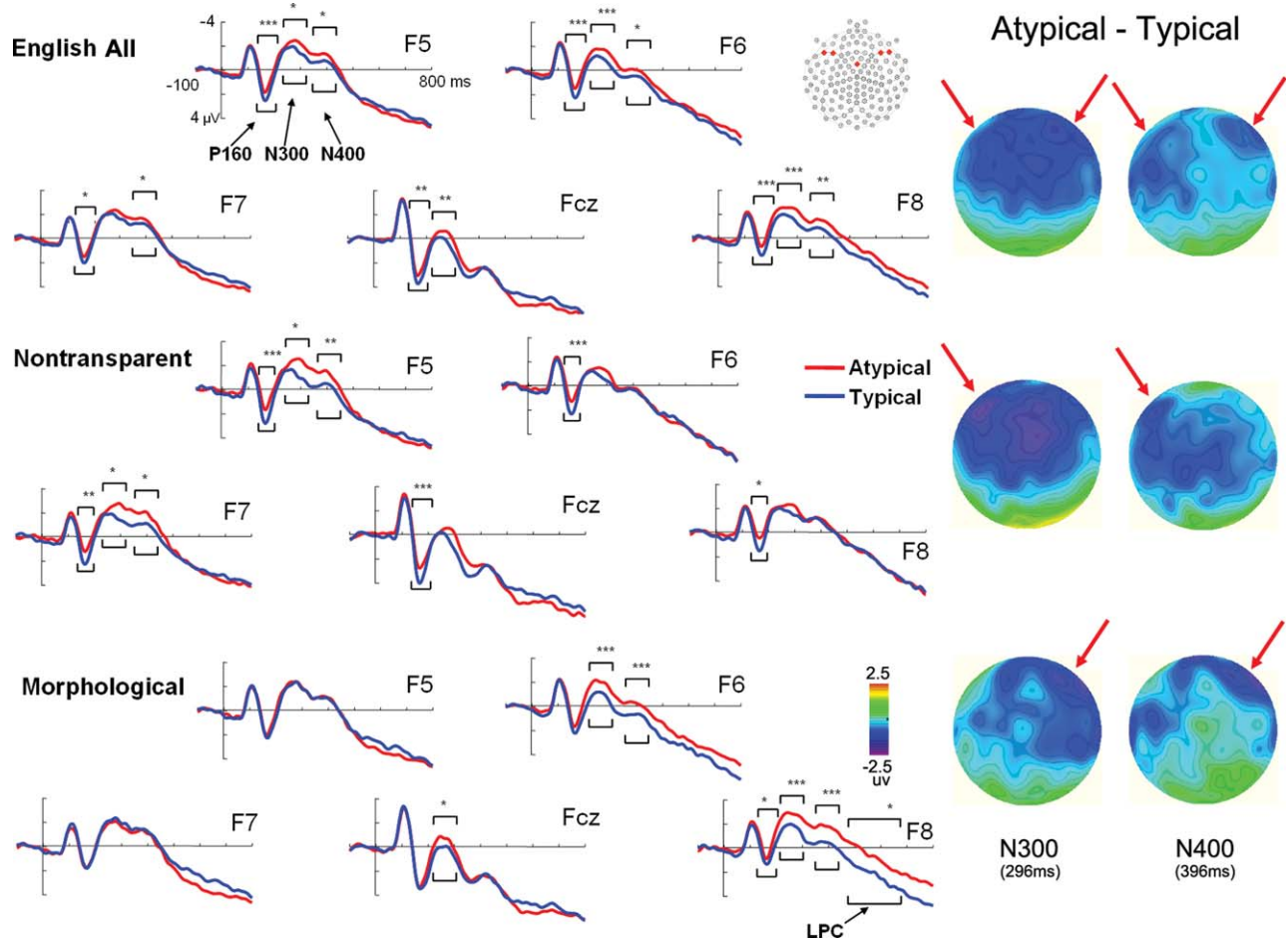
As in Studies 1 and 2, we focused our analysis on the horizontal line encompassing the bilateral frontal electrodes (F7, F5, Fcz, F6, and F8, corresponding to 34, 28, 6, 123, 122 in the Geodesic Sensor Net, respectively). Voltage data for peak amplitude of the P160, N300, and N400 components from two 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 ( $F(1,24) = 9.48$  and  $8.82$ ,  $P_s < 0.01$ ) and N400 ( $F(1,24) = 5.45$  and  $5.44$ ,  $P_s < 0.05$ ) 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



**Figure 5.**

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

electrodes (F5, F7) when viewing nontransparent items, but in the right frontal electrodes (F6, F8) activity when viewing morphologically transparent items for both N300 and N400 components. Significant typicality effects in bilateral frontal electrodes were found when combined together: \*  $P < 0.05$  \*\*  $P < 0.01$  \*\*\*  $P < 0.001$  for Bonferroni corrected post-hoc comparisons.

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,

2002; Han et al., 2004] might be involved for English speakers' analysis of the morphologically transparent category information. We interpret the relatively greater activation for the atypical morphologically transparent nouns at right side electrodes to be a result of this additional processing load for English speakers. These data alone, however, do not allow us to clarify whether the nature of the typicality effect is identical for morphologically transparent vs. nontransparent items in English.

## 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*) 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 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 *che1* 车) in the common morphological (e.g., car *jiao4che1* 轿车) 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.

To return to Whorf, and perhaps to Shakespeare—"What's in a name? That which we call a rose by any other name [may] smell as sweet." However, our data show that when a rose, or a canola, is called by a name that includes category information (e.g., *you2cai4hua1* 油菜花, or "canola flower"), it changes the way we think and the ways our brains access semantic information. These data also demonstrate when these differences occur in the brain. They suggest that typicality is a useful heuristic for deciding whether a rose is a flower when one's language does not regularly embed category-level terms in the labels for members of the category. Both the patterns of similarity and the patterns of divergence in the N300 and N400 ERP components between English and Mandarin speakers and between the two different types of nouns for English and Mandarin speakers in this categorization task suggest that category level judgments can undergo differences in processing at early- to mid-stages of stimulus processing (~ 300–400 ms after stimulus presentation), and yet show similarities at later stages of processing (e.g., LPC after 500 ms) and in behavioral responses. In other words, different brain processes can produce similar behavioral outcomes. At the least, these data demonstrate that whether one finds support for the linguistic relativity hypothesis (and for the typicality effect) may depend on the strength and pervasiveness of the linguistic information provided.

An interesting implication of these findings pertains to patients who present with symptoms of semantic dementia (SD) [Basso et al., 1988]. In several studies of English-speaking individuals, patients diagnosed with SD tended to not only have general word-finding and other semantic and conceptual difficulties, but a specific regularity in their behavior involving an over-reliance on information that is "typical" of the category or knowledge base being tested. This is true not only for words, but for pictures of real and imaginary animals as well as for real and nonsense words with typical and atypical spelling patterns [Hauk et al., 2006; Rogers et al., 2004; Woollams et al., 2008]. An interesting question given the present data is whether Chinese patients with SD would also show similar typicality effects in their symptoms and behaviors, or whether the morphological information provided in Chinese nouns could be used to help ameliorate these symptoms.

## 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

- Bailenson JN, Shum MS, Atran S, Medin DL, Coley JD (2002): A bird's eye view: Biological categorization and reasoning within and across cultures. *Cognition* 84:1–53.

- Barrett SE, Rugg MD (1990): Event-related potentials and the semantic matching of pictures. *Brain Cogn* 14:201–212.
- Basso A, Capitani E, Laiacona M (1988): Progressive language impairment without dementia—A case with isolated category specific semantic defect. *J Neurol Neurosurg Psychiatry* 51:1201–1207.
- Bjorklund DF, Thompson BE, Ornstein PA (1983): Developmental-trends in childrens typicality judgments. *Behav Res Methods Instrum* 15:350–356.
- Boster JS (1988): Natural sources of internal category structure—Typicality, familiarity, and similarity of birds. *Mem Cogn* 16:258–270.
- Chumbley JI (1986): The roles of typicality, instance dominance, and category dominance in verifying category membership. *J Exp Psychol Learn Mem Cogn* 12:257–267.
- Corbetta M, Shulman GL (2002): Control of goal-directed and stimulus-driven attention in the brain. *Nat Rev Neurosci* 3:201–215.
- Federmeier KD, Kutas M (2001): Meaning and modality: Influences of context, semantic memory organization, and perceptual predictability on picture processing. *J Exp Psychol Learn Mem Cogn* 27:202–224.
- Fujihara N, Nageishi Y, Koyama S, Nakajima Y (1998): Electrophysiological evidence for the typicality effect of human cognitive categorization. *Int J Psychophysiol* 29:65–75.
- Ganis G, Kutas M, Sereno MI (1996): The search for “common sense”: An electrophysiological study of the comprehension of words and pictures in reading. *J Cogn Neurosci* 8:89–106.
- Gelman SA, Wilcox SA, Clark EV (1989): Conceptual and lexical hierarchies in young-children. *Cogn Dev* 4:309–326.
- Gratton G, Coles MGH, Donchin E (1983): A new method for off-line removal of ocular artifact. *Electroencephalogr Clin Neurophysiol* 55:468–484.
- Han SH, Jiang Y, Gu H, Rao HY, Mao LH, Cui Y, Zhai RY (2004): The role of human parietal cortex in attention networks. *Brain* 127:650–659.
- Hauk O, Patterson K, Woollams A, Watling L, Pulvermuller F, Rogers TT (2006): [Q:] When would you prefer a SOSSAGE to a SAUSAGE? [A:] At about 100 msec. ERP correlates of orthographic typicality and lexicality in written word recognition. *J Cogn Neurosci* 18:818–832.
- Hauk O, Patterson K, Woollams A, Cooper-Pye E, Pulvermuller F, Rogers TT (2007): How the camel lost its hump: The impact of object typicality on event-related potential signals in object decision. *J Cogn Neurosci* 19:1338–1353.
- Heinze HJ, Munte TF, Kutas M (1998): Context effects in a category verification task as assessed by event-related brain potential (ERP) measures. *Biol Psychol* 47:121–135.
- Junghofer M, Elbert T, Tucker DM, Braun C (1999): The polar average reference effect: A bias in estimating the head surface integral in EEG recording. *Clin Neurophysiol* 110:1149–1155.
- Kiefer M (2001): Perceptual and semantic sources of category-specific effects: Event-related potentials during picture and word categorization. *Mem Cogn* 29:100–116.
- Kiefer M (2005): Repetition-priming modulates category-related effects on event-related potentials: Further evidence for multiple cortical semantic systems. *J Cogn Neurosci* 17:199–211.
- Komatsu LK (1992): Recent views of conceptual structure. *Psychol Bull* 112:500–526.
- Kutas M, Federmeier KD (2000): Electrophysiology reveals semantic memory use in language comprehension. *Trends Cogn Sci* 4:463–470.
- Kutas M, Hillyard SA (1980a): Event-related brain potentials to semantically inappropriate and surprisingly large words. *Biol Psychol* 11:99–116.
- Kutas M, Hillyard SA (1980b): Reading senseless sentences—Brain potentials reflect semantic incongruity. *Science* 207:203–205.
- Liu C, Zhang WT, Tang YY, Mai XQ, Chen HC, Tardif T, Luo YJ (2008): The visual word form area: Evidence from an fMRI study of implicit processing of Chinese characters. *Neuroimage* 40:1350–1361.
- McPherson WB, Holcomb PJ (1999): An electrophysiological investigation of semantic priming with pictures of real objects. *Psychophysiology* 36:53–65.
- Mervis CB, Rosch E (1981): Categorization of natural objects. *Annu Rev Psychol* 32:89–115.
- Mervis CB, Catlin J, Rosch E (1976): Relationships among goodness-of-example, category norms, and word-frequency. *Bull Psychonomic Soc* 7:283–284.
- Poldrack RA, Wagner AD, Prull MW, Desmond JE, Glover GH, Gabrieli JDE (1999): Functional specialization for semantic and phonological processing in the left inferior prefrontal cortex. *Neuroimage* 10:15–35.
- Rogers TT, Ralph MAL, Garrard P, Bozeat S, McClelland JL, Hodges JR, Patterson K (2004): Structure and deterioration of semantic memory: A neuropsychological and computational investigation. *Psychol Rev* 111:205–235.
- Schwanenflugel PJ, Rey M (1986): The relationship between category typicality and concept familiarity—Evidence from Spanish-speaking and English-speaking monolinguals. *Mem Cogn* 14:150–163.
- Siok WT, Jin Z, Fletcher P, Tan LH (2003): Distinct brain regions associated with syllable and phoneme. *Hum Brain Mapp* 18:201–207.
- Siok WT, Perfetti CA, Jin Z, Tan LH (2004): Biological abnormality of impaired reading is constrained by culture. *Nature* 431:71–76.
- Sitnikova T, Kuperberg G, Holcomb PJ (2003): Semantic integration in videos of real-world events: An electrophysiological investigation. *Psychophysiology* 40:160–164.
- Sitnikova T, Holcomb PJ, Kiyonaga KA, Kuperberg GR (2008): Two neurocognitive mechanisms of semantic integration during the comprehension of visual real-world events. *J Cogn Neurosci* 20:2037–2057.
- Stuss DT, Picton TW, Cerri AM (1988): Electrophysiological manifestations of typicality judgment. *Brain Lang* 33:260–272.
- Tan LH, Laird AR, Li K, Fox PT (2005): Neuroanatomical correlates of phonological processing of Chinese characters and alphabetic words: A meta-analysis. *Hum Brain Mapp* 25:83–91.
- Tardif T (2006): But are they really verbs?: Chinese words for action. In: Hirsh-Pasek K, Golinkoff RM, editors. *Action Meets Word: How Children Learn Verbs*. NY: Oxford University Press. pp477–498.
- West WC, Holcomb PJ (2002): Event-related potentials during discourse-level semantic integration of complex pictures. *Cogn Brain Res* 13:363–375.
- Whorf BL (1956): *Language, Thought and Reality: Selected Writings of Benjamin Lee Whorf*. Cambridge, MA: MIT Press. pp212–214.
- Woollams AM, Cooper-Pye E, Hodges JR, Patterson K (2008): Anomia: A doubly typical signature of semantic dementia. *Neuropsychologia* 46:2503–2514.
- Zhou XL, Marslen-Wilson W, Taft M, Shu H (1999): Morphology, orthography, and phonology in reading Chinese compound words. *Lang Cogn Processes* 14:525–565.
- Zhou YG (1978): To what degree are the “phonetics” of present-day Chinese characters still phonetic? *Zhongguo Yuwen* 146:172–177.

**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**

	Mandarin			
	Typical		Atypical	
SHOES	loafers	5.33	slippers	4.30
PANTS	trousers	5.38	overalls	4.83
VEHICLE	car	5.52	train	4.21
WRITING INSTRUMENT	pencil	5.00	chalk	3.17
STATION	train station	4.74	airport	1.96
BUG	fly	4.74	butterfly	4.08
BUILDING	office building	5.75	garage	1.50
OIL	engine oil	4.39	gasoline	3.41
VEGETABLE	celery	5.30	eggplant	4.74
PAPER	writing paper	5.13	toilet paper	4.87
English				
	Typical		Atypical	
SHOES	loafers	5.30	slippers	3.15
PANTS	trousers	5.85	overalls	4.19
VEHICLE	car	5.96	train	3.59
WRITING INSTRUMENT	pencil	5.81	chalk	4.33
STATION	train station	5.77	airport	3.59
BUG	fly	5.85	butterfly	4.74
BUILDING	office building	5.85	garage	2.84
OIL	engine oil	5.54	gasoline	3.74
VEGETABLE	celery	5.59	eggplant	4.96
PAPER	writing paper	5.89	toilet paper	3.85

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**

	Morphological			
	Typical		Atypical	
SHOES <i>xie2zi</i>	loafers <i>pi2xie2</i>	5.33	slippers <i>tuo1xie2</i>	4.30
PANTS <i>ku4zi</i>	trousers <i>xi1ku4</i>	5.38	overalls <i>bei1dai4ku4</i>	4.83
VEHICLE <i>che1</i>	car <i>jiao4che1</i>	5.52	train <i>huo3che1</i>	4.21
WRITING INSTRUMENT <i>bi3</i>	pencil <i>qian1bi3</i>	5.38	chalk <i>fen3bi3</i>	3.17
PAPER <i>zhi3</i>	writing paper <i>xin4zhi3</i>	5.13	toilet paper <i>shou3zhi3</i>	4.87
Orthographic				
	Typical		Atypical	
VEGETABLE <i>cai4</i>	celery <i>xi1qin2</i>	5.30	eggplant <i>qie2zi3</i>	4.74
BUG <i>chong2zi</i>	fly <i>cang1ying</i>	4.74	butterfly <i>hu2die2</i>	4.08
BIRD <i>niao3</i>	pigeon <i>ge1zi3</i>	5.43	penguin <i>qi3e2</i>	3.00
SHIP <i>chuan2</i>	warship <i>jun1jian4</i>	5.22	yacht <i>you2ting3</i>	4.39
STONE <i>shi2tou2</i>	rock <i>yan2shi</i>	4.74	brick <i>zhuān1tou2</i>	3.17

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**

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

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.