Journal of Experimental Psychology: General
2007, Vol. 136, No. 3, 485–501

0096-3445/07/$12.00 DOI: 10.1037/0096-3445.136.3.485

Scope of Linguistic Influence: Does a Classifier System Alter Object Concepts?

Henrik Saalbach
Max Planck Institute for Human Development

Mutsumi Imai
Keio University at Shonan-Fujisawa

Whether and to what extent conceptual structure is universal is of great importance for understanding the nature of human concepts. Two major factors that might affect concepts are language and culture. The authors investigated whether these 2 factors affect concepts of everyday objects in any significant ways. Specifically, they tested (a) whether the system of grammatical categorization by classifiers influenced the conceptual structure of speakers of classifier languages, and (b) whether Westerners organized object concepts around taxonomic relations whereas Easterners organized them around thematic relations, as proposed by R. E. Nisbett (2003). The relative importance of 3 types of relations—taxonomic, thematic, and classifier—for Chinese and German speakers was tested using a range of tasks, including categorization, similarity judgment, property induction, and fast-speed word–picture matching. Some support for linguistic relativity as well as for the cultural-specific cognition proposal was found in some tasks, but these effects were miniscule compared with the importance of taxonomic and thematic relations for both language–culture groups. The authors conclude that the global structure of everyday object concepts is strikingly similar across different cultures and languages.

Keywords: concepts, linguistic relativity, culture-specific cognition, classifiers, thematic relations

Language labels and refers to categories. Object categories are referred to by nouns, but in many languages, nouns are further categorized into grammatical categories, and languages have developed a broad variety of nominal classification systems such as count/mass grammar, gender marking grammar, and classifier grammar systems (e.g., Aikhenvald, 2000; Senft, 2000). The count/mass grammar system divides nouns into 2 categories: nouns denoting individuated things (typically objects, such as animals and machines) and nouns denoting nonindividuated things (typically substances, such as water and sand). Most gender grammar systems categorize nouns into 2 (masculine and feminine) or 3 (masculine, feminine, and neuter) categories. Unlike the count/mass or the gender grammar systems, the classifier grammar system categorizes nouns into more than 100 semantic categories (Craig, 1986; Grinevald, 2000).

How classifiers categorize the world is very different from how nouns categorize the world. Whereas the noun lexicon is structured hierarchically around taxonomic relations, classifier systems are usually organized around semantic features such as animacy, shape, function, size, rigidity, and social importance, and do not have hierarchical structures (e.g., Adams & Conklin, 1973; Allan, 1977; Croft, 1994; Denny, 1986; Downing, 1996). For example, animal classifiers in Chinese are determined by semantic features of size, shape, and social importance: *Tou* includes large animals such as cows, elephants, and rhinos; *zhi* typically includes small animals such as birds, insects, or smaller mammals. Horses have their own special classifier, *pi*, presumably because of their special historical importance for the society. Fish, snakes, and other long, thin, flexible animals are associated with *tiao*. *Tiao* also includes members across the ontological boundary of animacy, including long, thin, flexible inanimate things such as ropes, rivers, ties, and roads. One classifier for artifact objects, *ba*, is associated with objects that have a handle or that can be grasped by the hand, such as umbrellas, screwdrivers, brooms, keys, combs, chairs, and fans; but it can also be used for fire, effort or power, efficient worker, and so on, presumably because these things are metaphorically or metonymically associated with hands.[1]

Henrik Saalbach, Center for Educational Research, Max Planck Institute for Human Development, Berlin, Germany; Mutsumi Imai, Department of Environmental Information, Graduate School of Media and Governance, Keio University at Shonan-Fujisawa, Kanagawa, Japan.

This research was supported by fellowships from the Japan Society of the Promotion of Science and Friedrich-Ebert-Stiftung awarded to Henrik Saalbach, and Ministry of Education Grant-in-Aid for Scientific Research 15300088 and research grants from Keio University (Keio Gijuku Academic Development Funds and Keio Gijuku Mori Memorial Research Fund) awarded to Mutsumi Imai.

We are deeply indebted to Zhou Xiaolin, Elsbeth Stern, Sachiko Saitoh, Lennart Schalk, Shu Hua, Zhu Liqi, and Huangfu Gang for help with data collection and discussion. We also thank Ed Wisniewski, Barbara Malt, and Victoria Muehleisen for their thoughtful comments on earlier versions of this article. We both equally contributed to the article, and the order of authorship is arbitrary.

Correspondence concerning this article should be addressed to Mutsumi Imai, Keio University at Shonan-Fujisawa, Endo, Fujisawa, 252-8520 Japan; or to Henrik Saalbach, who is now at ETH Zurich, Institute for Behavioral Sciences, Research on Learning and Instruction, Universitaetsstrasse 6, Zurich CH-8092, Switzerland. E-mail: imai@sfc.keio.ac.jp or henrik.saalbach@ifv.gess.ethz.ch

[1] In standard Chinese, classifiers must occur with a number and/or a demonstrative (i.e., a word with meaning similar to *this*, *that*, *which*), or certain quantifiers (e.g., words equivalent to *whole*, *many*, *several*, *a few*, *a certain*, *every*) before a noun (Li & Thompson, 1981) as shown here: (a) *Ta* (she) *kan* (see) *le* (grammatical particle indicating completed action) *yi* (one) *tiao* [classifier] *she* (snake) = She saw a snake. (b) *Ni* (you) *xiang* (think of/want) *mai* (buy) *ji* (how many) *ba* [classifier] *yusan* (umbrella)? = How many umbrellas do you want to buy?

An extremely interesting and important question here is whether and to what extent classifier categories produce any significant cognitive consequences in the minds of the speakers of a classifier language beyond the act of linguistic reference. Some researchers have assumed that semantic structures of the classifier system are a reflection of the conceptual structure of the speaker. For example, Lakoff (1987) saw linguistic categories as categories within the cognitive system per se:

> whether they are used in non-linguistic tasks or not, linguistic categories are categories—and they are part of our overall cognitive apparatus. Whether one wants to dignify them with the term "conceptual" or not, linguistic categories are categories within our cognitive system and a study of all categories within our cognitive system will have to include them. (p. 110)

Lakoff further argued that classifier categories are among the best examples of language expressing conceptual categories. He wrote: "Classifier languages–languages where nouns are marked as being members of certain categories–are among the richest sources of data that we have concerning the structure of *conceptual* [italics added] categories as they are revealed through language" (p. 91).

Further specific questions arise concerning the question of whether classifier categories have any significant cognitive consequences. Given that a classifier system carves up the world in a way that cross-cuts taxonomic categories, what cognitive consequences, if any, should we expect with respect to representation and cognitive processes? One possibility is that classifier systems provide an alternative organization of object concepts that results in categories that are nonexistent for speakers of nonclassifier languages. If this is the case, one would expect that speakers of a classifier language and those of a nonclassifier language would behave very differently in almost all cognitive activities, including category formation, similarity judgment, and, most importantly, inductive reasoning. People categorize things in the world in order to communicate, learn, and remember more efficiently, and it is the basis for inductive inference about unseen properties of novel objects (e.g., Murphy, 2002). Thus, if the classifier system alters the organization of object concepts, one would expect that speakers of a classifier language would draw inductions very differently from speakers of a nonclassifier language. Furthermore, if the classifier system provides a way of organizing object concepts, one may expect that classifier relations would be accessed automatically in online processing.

A second possibility is that speakers of a classifier language do not organize their concepts in the sense that they do not use classifier-category membership as a basis for categorization, nor do they draw inductive inferences from it. However, the experience of linguistically categorizing objects by the use of classifiers may heighten attention to semantic features underlying classifier categories, and, as a consequence, the similarity among objects that are members of the same classifier category may be magnified. If this is the case, the difference between speakers of a classifier language and those of a nonclassifier language may be observed in similarity judgment but not in categorization or inductive reasoning.

A third possibility is that classifiers are "frozen" linguistic conventions and do not have any cognitive impact on speakers of classifier languages. In this case, one would not see any difference between speakers of a classifier language and those of a nonclassifier language, although the two groups may differ due to factors other than the classifier categorization system (e.g., culture).

To our knowledge, there is only a handful of studies in the literature that have directly addressed the question of whether classifiers affect concepts and categorization of objects beyond the contexts of linguistic references. One of these is the study by Zhang and Schmitt (1998), in which the authors asked whether classifiers influence perception of similarity between two objects. They had speakers of Chinese and English rate the similarity of pairs of everyday objects. Half of the pairs consisted of objects that shared the same classifier in Chinese, and half of the pairs consisted of objects from different classifier categories. Zhang and Schmitt found that the Chinese speakers rated the similarity of the same classifier pairs more highly than the native English speakers did, whereas ratings of the different classifier pairs did not differ cross-culturally. On the basis of these results, the authors concluded that classifier categories strongly affect speakers' conceptual organization:

> The results obtained were a strong indication that objects sharing the classifier are grouped into schematic organizations in Chinese speakers' mental representations. That is, although English speakers may group these objects on the basis of their conceptual similarity, Chinese speakers seem to add a linguistic categorization to the classification of objects. (p. 381)

However, their results do not provide a clear answer as to whether the classifier system truly influences Chinese speakers' conceptual structure. Specifically, it is not clear whether their results suggest that Chinese speakers' organization of object concepts is significantly different from that of English speakers due to classifier categories, because their experimental design does not allow researchers to determine how to interpret the effect they found in light of the criteria we suggested earlier. It is possible that the classifier effect found by Zhang and Schmitt (1998) was limited to similarity ratings, and that the classifier relations were not utilized in inductive reasoning or activated even in automatic processing. If so, to claim that the classifier system adds a new way of organizing concepts might be an overstatement, though it may still be taken as a weak form of linguistic relativity. If Chinese speakers draw inductive inferences on the basis of classifier category membership and the classifier category membership is automatically evoked even when a classifier is not expressed, we would be more comfortable in agreeing that the classifier system does indeed provide Chinese speakers with a way of organizing objects that English speakers do not possess.

To distinguish the two possibilities, the influence of classifier categories must be attested in multiple cognitive tasks such as categorization, similarity judgment, inductive reasoning, and a task in which automatic semantic access is required. Furthermore, to evaluate the claim that the classifier system adds a new way of organizing concepts, the impact of the classifier system should be evaluated with respect to the other major conceptual relations that are said to organize speakers' concepts. If we find that classifier relations have an impact of greater magnitude than other major conceptual relations or have a comparable impact, we can comfortably conclude that classifiers function as one of the major organizers of concepts.

What are some other conceptual relations that are considered to organize concepts? One is, of course, taxonomic relations. Taxonomic categories are denoted by nouns and include items of the same kind. They are differentiated into levels of varying specificity (e.g., animal, dog, collie) related by class inclusion (e.g., a collie is a dog, a dog is an animal, a collie is an animal). Numerous studies have shown that taxonomic relations organize concepts and provide a basis for categorization, similarity judgment, and inductive reasoning (e.g., Gelman & E. M. Markman, 1986; E. M. Markman, 1989; Osherson, Smith, Wilkie, López, & Shafir, 1990; Waxman & Gelman, 1986) in both children and adults. However, recently researchers have noted that taxonomic relations do not capture the full spectrum and richness of human concepts and categories, and they have pointed out that thematic relations are also an integral and important part of conceptual structure (e.g., Lin & Murphy, 2001; Wisniewski & Bassok, 1999; see also Bassok & Medin, 1997). Lin and Murphy (2001; see also E. M. Markman, 1989) suggested that many human concepts include knowledge about nontaxonomic relations, with thematic relations being the most important sort among them. These authors defined thematic relations as external relations that arise through objects co-occurring or interacting together in space or time, or objects being linked by functional or causal relationships (e.g., table/chair, morning/newspaper, scissors/paper). Through a series of experiments with varying paradigms, including speeded and nonspeeded forced choice categorization and property induction, Lin and Murphy demonstrated that thematic relations play a prominent role in the conceptual structure of well-educated young American adults.

Our Research

In this research, we examined whether classifier relations had any impact on speakers of a classifier language by testing Chinese and German speakers on a range of tasks including categorization, similarity judgment, and property induction. We also included a speeded word–picture matching task to see whether membership of the same classifier category was automatically invoked in the semantic network. Furthermore, we designed the stimuli and experiments in such a way that, if classifier relations indeed exerted any influence on the speakers' concepts, the degree of the impact could be compared with that of taxonomic and thematic relations. We designed a stimulus set of everyday objects that allowed us to examine four types of relations around the same target (e.g., flower). One object type was taxonomically related (e.g., tree) and a second was thematically related (e.g., vase) to the target. A third object type shared the same classifier as the target in Chinese (e.g., cloud), and the fourth type served as a control, having no relation with the target (e.g., shoe). The four relations were orthogonally crossed, so that the object serving as the same-classifier item was not related to the target taxonomically or thematically. Likewise, the taxonomic item did not belong to the same classifier category as the target, nor did it have any thematic relation to the target.

If classifier categories function as an organizer of object concepts, speakers of a classifier language should utilize them not only in similarity judgment but also in category formation and inductive reasoning. However, it is possible that classifier category membership heightens construal of similarity through heightening attention to semantic features underlying the classifier system, but these features (e.g., shape,[2] flexibility, size, hand manipulability [e.g., *ba* example described earlier on page 1], social importance) may not be readily used as basis of inductive reasoning, especially when other conceptual relations or background knowledge is available, because these features do not usually support commonality of internal constituents or spatiotemporal proximity as taxonomic or thematic categories do. If this is the case, we may expect to see higher similarity ratings for the same-classifier pairs in Chinese speakers than in German speakers, but we may not necessarily see such a cross-cultural difference in property induction, especially when participants are able to access background knowledge. Furthermore, in this case, with Chinese speakers, we should expect that the rated similarity for the same-classifier pairs will not surpass that for the taxonomically or thematically related pairs.[3]

Culture and Language

Our design also provided an excellent opportunity to evaluate a proposal that has attracted much attention in the recent literature on cross-cultural cognition. Nisbett and his colleagues have put forward a hypothesis that the philosophy, values, and customs that have been nursed in a culture throughout its history lead to a "culturally specific" style of cognition (e.g., Choi, Nisbett, & Smith, 1997; Nisbett, 2003; Nisbett, Peng, Choi, & Norenzayan, 2001). In their empirical work, Nisbett and colleagues compared East Asians and Westerners. Characterizing the former as "holistic" and the latter as "analytic," they argued that whereas East Asians tend to view the environment as a unified whole and pay a

[2] Shape here does not mean concrete shape of objects. Rather, shape relevant for classifier categories is dimensional shape, such as one dimensionally extended (long and thin), two dimensionally extended (flat), and three dimensionally extended.

[3] Note that, just as nouns in Chinese are categorized by the classifier system, nouns in German are categorized into three gender categories of masculine, feminine, and neuter. In two recent studies, Vigliocco and colleagues (Vigliocco, Vinson, Indefrey, Levelt, & Hellwig, 2004; Vigliocco, Vinson, Paganelli, & Dworzynski, 2005) examined the cognitive impact of grammatical gender in German. Vigliocco et al. (2004) found that, when semantic substitution errors were induced, German speakers tended to preserve grammatical gender when they produced phrases with gender-marked determiners but not when they produced bare nouns. Vigliocco et al. (2005) further demonstrated that in an odd-one-out categorization task, German speakers were not influenced by grammatical gender (whereas an influence was observed in Italian speakers in the same task). Based on these results, Vigliocco and colleagues suggested that grammatical gender is not part of German speakers' lexicosemantic or conceptual representation. Given Vigliocco and colleagues' results, it is not likely that the performance of German participants would be affected by the German grammatical gender system. However, to avoid any potential confounding from the grammatical gender effect, it would have been ideal if we could have designed the materials in such a way that German grammatical gender assignment would have been orthogonally crossed with other relations (the taxonomic, thematic, classifier) in the stimuli. However, in practice this manipulation would have made the stimulus construction virtually impossible. We thus did not incorporate this manipulation in preparing the stimuli. Instead, in each experiment we carried out an analysis to test if there was any language-specific effect due to grammatical gender on the part of the German participants, which would have been manifested by a significant interaction between culture–language and gender (whether or not pairs of nouns were of the same grammatical gender).

great deal of attention to relations that tie elements into the environment, Westerners tend to focus on individual elements of the environment separately. Based on this schema, Nisbett and colleagues made a specific prediction regarding the conceptual structure of East Asians and Westerners: East Asians, with their predisposition to see a scene or event as a whole, would be expected to categorize the world around thematic relations; Westerners, with their focus on properties of individual objects, would be expected to categorize the world by taxonomic relations. If this argument holds, in our study we would expect a difference between Germans (Westerners) and Chinese (Easterners) in their reliance on the taxonomic and thematic relations in the four tasks, with Germans relying more heavily on the taxonomic relations and Chinese relying more heavily on the thematic relations.

Ji, Zhang, and Nisbett (2004) further attempted to specify whether this cross-cultural difference could be attributed to differences in language rather than culture. For this purpose, they tested four language-culture groups on an odd-one-out categorization task. The four groups were (a) American college students, (b) Chinese college students in mainland China, (c) Chinese students living in the United States who were from mainland China or Taiwan, and (d) Chinese students living in the United States who were from Singapore or Hong Kong. The participants were asked to choose two items out of three that were most closely related to each other; items were grouped on the basis of thematic relations, taxonomical relations, or neither. The American participants were tested only in English. The Chinese groups were tested in Chinese as well as in English.

Ji et al. (2004) found that not only the Americans but also the Chinese bilinguals from Hong Kong and Singapore made groupings based on taxonomic relations more often than did the mainland and Taiwan Chinese. The authors attributed this difference among the Chinese groups to culture rather than language, stating that

> learning English at an early age is also an indicator of the environment where the children grow up. Hong Kong and Singapore Chinese are more westernized than Mainland China and Taiwan. A mixture of English and Chinese languages and a mixture of Chinese and Western ways become the reality that Hong Kong and Singapore Chinese live. (p. 64)

This quote itself succinctly expresses a problem of using bilinguals for the purpose of separating the influence of culture and language in cognition: It is a chicken or the egg problem. In other words, it is virtually impossible to determine which of the two is the cause and which is the effect. In this case, one could easily argue that the Hong Kong and Singapore Chinese responded more like Americans because English had been established as a medium of their thought from early on. Furthermore, as reviewed earlier, previous research has suggested that thematic relations are an integral part of the conceptual structure even for educated American adults (Bassok & Medin, 1997; Lin & Murphy, 2001; Wisniewski & Bassok, 1999). Thus, it is worthwhile to revisit Ji et al.'s conclusion that Westerners organize their concepts around taxonomic relations.

Our research can provide an unique opportunity to test Nisbett and colleagues' culture-specific cognition proposal on one hand (Ji et al., 2004) and the proposal that thematic relations are integral part of the conceptual structures even for well-educated people in Western culture on the other hand (Lin & Murphy, 2001; Wisniewski & Bassok, 1999), in addition to examining whether and to what degree linguistic categories affect conceptual structure. For this purpose, instead of testing bilinguals as Ji et al. did, we compared Chinese and German participants who were living in their own cultural settings and used their native language in administering the tasks. We are aware of a concern that instructions given in different languages may not be exactly the same (Boroditsky, 2001). However, this worry may not be much greater than the worry that arises from difficulties in controlling the participants' age of acquisition and proficiency in the nondominant language, as well as the need for the cultural background to be homogeneous within the same language–cultural group.

Another important difference between our approach and Ji et al.'s (2004) lies in the notion of the "effect of language" per se. Rather than defining the effect of language broadly, and hence vaguely, we focused on a particular aspect of language that is directly relevant to object categorization (i.e., the classifier system). We can thus make clear a priori predictions about the effect of language in our studies. It is important to note that in our research, we did not need to evaluate the effect of culture and language in a mutually exclusive, black-and-white fashion. For example, it was possible for us to find that Chinese speakers showed stronger sensitivity both to classifier relations and to thematic relations than did German speakers. Our paradigm also allowed us to evaluate the relative importance of taxonomic relations, thematic relations, and classifier relations within the culture. Thus, it was possible that we could find that the people from the two culture–language groups showed reliance on the three types of relations in the same order but nonetheless find that the two groups relied on the three relations to different degrees. In this way, we could place the effects of culture and language, if we found any, in a global picture of our conceptual structures of everyday objects.

Experiment 1: Forced Choice Categorization Task

In our first experiment, we compared Chinese and German participants in a categorization task. Following previous research (e.g., Ji et al., 2004; Lin & Murphy, 2001), we employed a forced choice match-to-sample task: The participants were asked to determine which of the two test items best matched the target item. As stated earlier, our stimuli included a taxonomic item, a thematic item, a classifier item, and a control item around the same target object. We thus constructed six types of contrasts around the same target item by making pairwise combinations of the four relations: (a) classifier versus taxonomic, (b) classifier versus thematic, (c) classifier versus control, (d) taxonomic versus thematic, (e) taxonomic versus control, and (f) thematic versus control. In this way, we were able to test whether classifier categories influenced people's classifications of everyday objects (through the classifier–control contrasts) and to assess the strength of this influence relative to taxonomic relations and the thematic relations. In addition, the contrast between the taxonomic and the thematic items allows us to test the proposal by Nisbett and colleagues that Easterners (Chinese) organize their concepts around thematic relations, whereas Westerners (Germans) organize their concepts around taxonomic relations (Ji et al., 2004).

Method

Participants

Twenty-three Chinese undergraduates living in Beijing and 24 German undergraduates living in Berlin participated in this study. The Chinese participants were all native speakers of Mandarin Chinese, and the German participants were all native speakers of German. In both cultural groups, most of the participants were undergraduates majoring in psychology, but those from other majors, such as engineering, physics, and social sciences, were also included. The participants in both groups were paid for participation. The demographic backgrounds of the participants were the same for all of the studies reported in this article.

Materials

Preparation of stimulus sets common to all studies. We constructed 14 quintuplets of objects, each consisting of one target and four objects representing each of the four types of relations to the target. The first type was from the same classifier class as the target item but was not related to it either taxonomically or thematically (e.g., flower–cloud: same classifier). The second type was taxonomically related to the target item (e.g., flower–tree: taxonomic), and the third type was thematically related to the target item (e.g., flower–vase: thematic). Neither the taxonomic nor thematic items belonged to the same classifier class as the target item. The fourth type served as a control, unrelated taxonomically or thematically, and with objects from different classifier classes (control). The 14 quintuplets were used for all of the studies reported in this research. We first describe how these 14 quintuplets were constructed before describing how we prepared the stimulus materials for Experiment 1.

To select items for constructing the quintuplets, we conducted a pre-study on native speakers of Mandarin Chinese who did not participate in any of the main studies reported in this article. A group of Chinese graduate students from Peking University selected a set of 21 prominent classifiers and the nouns that were prominently associated with each of the classifiers from a classifier dictionary (Guo, 2002). Next, 100 object names were selected and were randomly arranged in a questionnaire booklet. We then instructed 10 native Mandarin Chinese speakers from the Beijing region to write the corresponding classifier for each of the nouns listed. We selected the nouns for which at least 8 out of 10 wrote the classifier we had originally expected on the basis of dictionary definitions. One fifth of the nouns did not meet this requirement and were thus dropped from the list. A new questionnaire of 100 nouns, including the remaining nouns from the first round as well as newly chosen ones, was distributed to four graduate students from the Chinese Academy of Sciences in Beijing for a final check. A noun was excluded from the final list if its most appropriate classifier was not agreed on by at least three of the four judges.

From the final list of the nouns, we selected 14 pairs of nouns that belonged to the same classifier category yet were not either taxonomically or thematically related to each other. One of the pair was assigned to serve as a target, and the other to serve as the same-classifier item for that target. Also from the list we selected the taxonomic, thematic, and control items for each target based on common sense, with the constraint that none of the three items belonged to the same classifier category as the target item. Finally, we checked whether the taxonomic and thematic items we had selected indeed carried taxonomic and thematic relations. In this task, we presented all item sets (each including target, taxonomic item, thematic item, same-classifier item, and control) in a questionnaire and asked 10 Chinese and 18 German undergraduates who did not participate in any of the studies in their native language to select the items that shared the best taxonomic and thematic relation to the target. To make sure that participants understood the terminology, definitions and examples of the two relations were provided.[4] Overall, the participants of the relation verification task had 97.5% agreement on the taxonomic choices and 93% on the thematic choices. There were no significant cross-cultural differences in the agreement rates either for the taxonomic choices (Chinese: 98.6%; German: 96.4%) or for the thematic choices (Chinese: 92.9%; German: 93.1%). The final list of stimulus sets is presented in Table 1.

Material used in Experiment 1. A questionnaire booklet containing 84 triads of objects was prepared. Each triad consisted of the target and two choice items and represented one of the six contrasts: classifier item versus taxonomic item, classifier item versus thematic item, classifier item versus control, taxonomic item versus thematic item, taxonomic item versus control, and thematic item versus control. For each triad, the name of the target object was presented on the top with the names of the two choice objects beneath it, accompanied by an instruction asking which of the two objects below went best together with the object above (German: "[A] *passt am besten zu* [B] *oder* [C]?"; Chinese:"[A] *he shenme zai yiqi zui heshi:* [B] *huozhe* [C]?" Approximate English translation: "Which of [B] or [C] best goes together with [A]?").

Three versions of the questionnaire booklet were prepared, each with a different random order of the stimuli. For each contrast type (e.g., classifier vs. taxonomic, taxonomic vs. thematic, etc.), the relative left/right position of the two alternatives with respect to the target was counterbalanced. Each object was presented in the word form (see the General Discussion for why we chose to present the stimuli verbally but not pictorially).

Procedure

The participants were tested in groups and randomly received one of the three versions of the booklet. They were instructed to go through the questionnaire carefully at their own pace and to rely on their intuition.

---

[4] The definitions and examples for taxonomic relations and thematic relations were given as follows. *Taxonomic relations:* Two things are considered to have taxonomic relations when they belong to the same conceptual category and share common behavioral, functional, or internal properties. For example, cats and dogs are taxonomically related because they are both animals and have many behavioral, physical, and internal properties in common. Motorcycles and buses can also be considered as taxonomically related, because they are both vehicles and have the common function of serving as a tool for transportation. *Thematic relations:* Two things are considered to have thematic relations when they are related externally (e.g., seen together in the same scene or the same event, or have causal relations) rather than internally (i.e., have common internal properties). For example, dogs and dog houses are thematically related because they are often seen together, but the two objects do not share common properties that reside in them. Cows and milk are also thematically related because cows produce milk, but cows and milk do not share internal properties.

Table 1
Stimulus Items Used for Experiments 1 Through 4

Target	Classifier	Same classifier	Taxonomic item	Thematic item	Control
Comb	*Ba*	Key	Hair dryer	Hair	Ticket
Pistol	*Ba*	Umbrella	Canon	Bullet	Stamp
Scissors	*Ba*	Fan	Cutter	Paper	TV
Chain	*Tiao*	Carp	Rope	Lock	Poster
Necklace	*Tiao*	Blanket	Ring	Dress	Book
Towel	*Tiao*	Eel	Handkerchief	Shower	Potato
Mountain	*Zuo*	Tower	Hill	Snow	Necklace
Bell	*Zuo*	Building	Buzzer	Temple/church	Bike
Piano	*Jia*	Ladder	Violin	Music book	Scarf
Plane	*Jia*	Swing	Boat	Airport	Chain
Flower	*Duo*	Cloud	Tree	Vase	Cup
Newspaper	*Zhang*	Bed	Book	Morning	Tube
Drum	*Mian*	Wall	Trumpet	Sticks	Scissors
Tent	*Ding*	Hat	Sleeping bag	Campfire	Table

Results

Table 2 shows the mean percentages of the choices for classifier, taxonomic, and thematic items as well as control items across the different contrasts. We first report the results for the contrasts involving the classifiers to see whether Chinese speakers classified objects differently from Germans in a way that was in accord with the way classifiers categorize them. We then report the results concerning the question of whether Chinese and German participants showed culture-specific preferences with regard to the taxonomic-based and thematic-based groupings. We also checked whether grammatical gender in German affected German speakers' performance in this forced choice categorization task.[5]

Are Classifiers Used as the Basis of Categorization by Chinese Speakers?

We calculated the proportion of the same-classifier choices for each participant and carried out single-sample t tests to see whether the same-classifier choice exceeded chance (50%). When pitted against the control item, both groups' choices of the same-classifier item reliably exceeded chance: Chinese, 76.1%, $t(22) = 12.49$, $d = 2.61$; Germans, 71.1%, $t(23) = 8.98$, $d = 1.82$, same-classifier choice, both ps $< .01$, Bonferroni corrected. When the same-classifier item was pitted against the taxonomic item, both the Chinese and German participants selected the taxonomic item (Chinese: 82.9%; German: 89%) over the same-classifier item almost exclusively. Similarly, when the thematic relation was pitted against the same classifier relation, both the Chinese and Germans predominantly chose the thematic item (Chinese: 82.9%; German: 84.8%) over the same-classifier item.

Table 2
Mean Percentages of Classifier, Taxonomic, Thematic, and Control Item Choices Across Conditions of Experiment 1

Relation	Chinese ($n = 23$)	German ($n = 24$)
	Classifier vs. taxonomic	
Classifier	17.1	11.0
Taxonomic	82.9	89.0
	Classifier vs. thematic	
Classifier	17.1	15.2
Thematic	82.9	84.8
	Classifier vs. control	
Classifier	76.1	71.1
Control	23.9	28.9
	Taxonomic vs. thematic	
Taxonomic	36.0	34.5
Thematic	64.0	65.5

To examine the effect of culture on the proportion of classifier choice, we conducted a 3 (contrast type: same classifier vs. taxonomic, same classifier vs. thematic, same classifier vs. control) × 2 (culture) analysis of variance (ANOVA). Naturally, the effect of contrast type was significant, $F(2, 90) = 36.00$, $p < .001$, but the two groups did not differ from each other on the overall rate of classifier choice, $F(1, 45) = 2.45$, $p = .127$. The Culture × Contrast Type interaction was not significant, $F(2, 90) = 0.35$. Thus, there was no classifier effect here. Objects in the same classifier category in Chinese were chosen over the control items not only by the Chinese but also by the German participants.

Do Chinese Speakers Categorize Objects on the Basis of Thematic Relations?

We next examined whether Chinese speakers predominantly categorized objects on the basis of thematic relations and German speakers categorized objects on the basis of taxonomic relations, as predicted by Nisbett and his colleagues (Ji et al., 2004). The critical contrast to test this hypothesis was the pairs contrasting the

[5] Because we did not manipulate this as a variable orthogonally crossed with other variables, of the total 84 pairs there were only 36 pairs in which the grammatical gender of the two choice items differed from each other. For these 36 pairs, the object whose gender was the same as the target was selected 61% of the time by the Chinese group and 65% of the time by the German group. This difference was not statistically significant, $t(35) = 1.53$, $p > .1$. Thus, it is unlikely that the results of Experiment 1 were confounded by grammatical gender for the German participants.

taxonomic and thematic items directly. It is interesting to note that, in this case, both the Chinese and Germans favored the thematic item—Chinese: 64%, $t(22) = 3.25$, $p < .01$, $d = 0.63$; Germans: 65.5%, $t(23) = 5.22$, $p < .01$, $d = 1.06$—over the taxonomic item. There was no cross-cultural difference in the rate of the thematic choices here, $t(45) = .25$, $p > .10$ (observed power = .08).

For the pairs in which the taxonomically related or thematically related object was pitted against the control object, speakers of both languages naturally selected the taxonomic (Chinese: 94.9%; German: 94.0%) or the thematic object (Chinese: 96.9%; German: 96.2%) almost exclusively as a match to the target, and no cross-cultural difference was detected in the rate of the taxonomic or the thematic choice over the control.

Discussion

Overall, classifier categories were not used as the basis for categorization by Chinese speakers. When the same-classifier item was pitted against the taxonomic or the thematic item, the Chinese as well as the German speakers made categories exclusively on the basis of taxonomic or thematic relations. When the same-classifier item was contrasted to the control, not only the Chinese participants but also the German participants judged the same-classifier item to be the better match to the target. This finding suggests that there is an inherent similarity among objects belonging to the same classifier category, even when they do not share any taxonomic or thematic relations. This inherent similarity is detectable by speakers of a nonclassifier language, and people use it when there is no other kind of similarity to resort to in forming categories. However, it is not the kind of similarity even speakers of a classifier language use spontaneously for categorization, especially when taxonomic or thematic relations are present.

In addition, we did not find evidence for the proposal put forward by Ji et al. (2004) that Westerners organize object concepts around taxonomic relations whereas Easterners organize them around thematic relations. It was particularly interesting that not only the Chinese but also the Germans preferred the thematic match over the taxonomic match. This finding is consistent with Lin and Murphy's (2001) results and further suggests that thematic relations are a universally important and integral part of human concepts. At this point, we do not know why our results are different from those of Ji et al., who found that the Chinese group showed a preference for thematic grouping and the American group showed preference for taxonomic grouping. One possibility is that the difference arose from the stimuli. Another reason could be the difference in the instructions, as previous research has shown that different instructions can result in different categorization behavior within the same age/cultural community (Waxman & Namy, 1997). We asked the participants to select the object that "best goes together" with the target object ("Which of [B] or [C] best goes together with [A]?"). In contrast, Ji et al. asked their participants to indicate which two of the three were "most closely related." The phrase *best goes together* may highlight thematic relations to a greater degree than the phrase *most closely related*, which may have more neutral nuance between taxonomic- and thematic-based groupings.

In our view, the question of what instruction is most appropriate to elicit people's most representative categorization behavior is not so important, however, given that people's categorization behavior is very context sensitive and susceptible to change according to different instructions. It is very possible that within-culture variation due to instructions is greater than across-culture differences in the representative categorization behavior, if there is any such thing (see Saalbach & Imai, 2006, for relevant data from Chinese and German children and adults). In any case, our results indicate that Ji et al.'s proposal may not hold as a general claim, as the preference between thematic-based grouping and taxonomic-based grouping they found was not replicated when different stimuli and different instructions were used.

In summary, in Experiment 1 we did not find evidence that classifiers affect Chinese speakers' concepts of everyday objects, nor did we find evidence that Chinese people rely on thematic relations more strongly than do Germans in grouping objects. Given the results of Experiment 1, it is unlikely that the Chinese and Germans organize their concepts in drastically different ways. However, this does not exclude the possibility that the classifier system does exert an influence in a more subtle fashion (e.g., heightening attention to semantic features underlying classifier categories), which may be detected only by more sensitive, more finely grained tasks using rating scales. It was important to conduct the forced choice categorization task, as this task has been most widely used in the closely relevant literature (e.g., Ji et al., 2004; Lin & Murphy, 2001; E. M. Markman & Hutchinson, 1984; Smiley & Brown, 1979; Vigliocco, Vinson, Paganelli, & Dworzynski, 2005). However, there is a limitation to this methodology, as it forces the participants to choose only one match to the target, even in cases in which the participants think that neither of the two choice items would go together well with the target or that both items would go together well. In contrast, similarity judgment on a rating scale allows participants to judge similarity between the target and each item independently of other items in the set, and may hence be more sensitive to any subtle differences between the two cultural groups, if any.

However, as discussed earlier, even if we had found some influence of classifier categories in people's similarity judgment, it would be premature to conclude only on the basis of such a finding that classifier categories provide an alternative way of organizing object concepts for Chinese speakers. Therefore, in addition to similarity judgment, we also tested whether Chinese speakers generalized properties on the basis of classifier category membership. We followed Lin and Murphy and used the property "have the same bacteria," because this property can be applied to both animate and inanimate artifact stimuli and is assumed to cover various kinds of relations, including taxonomic and thematic ones (Lin & Murphy, 2001). As in the similarity judgment study, we used a rating scale to have participants judge the likelihood of the two items sharing the same property.

Experiment 2: Similarity Judgments and Inductive Reasoning

In our second experiment, we examined (a) whether classifier relations had any impact on people's similarity judgments and (b) whether classifier relations could promote inductive inference by having participants respond on a rating scale. We presented Chinese and German participants with pairs of objects representing four relations around the same target (belonging to the same classifier category, belonging to the same taxonomic category, thematically related, and unrelated) and asked them (a) to rate the similarity of the two objects presented in each pair, or (b) to rate the likelihood that the two objects shared the same property.

In addition, in parallel to Experiment 1, we also examined whether the Chinese and Germans utilized taxonomic and thematic

relations differently when judging similarity and making inductive inferences in the way predicted by Nisbett (Ji et al., 2004). Although Ji et al. concluded that object concepts are organized differently by East Asians and Westerners, they made this strong claim only on the basis of the results of an odd-one-out categorization task. If we found that the Chinese relied on thematic relations more strongly than taxonomic relations and Germans relied on taxonomic relations more strongly than thematic relations in judging similarity and making inductive inferences, it would be much more convincing evidence for the culture-specific cognition proposal. Even if this strong pattern were not borne out, if we found that, in the two tasks, Chinese participants displayed higher ratings for thematic relations than did Germans and German participants displayed higher ratings for taxonomic relations than did Chinese participants, these results would be some support for the proposal, if not as strong as in the first case.

Method

Participants

Thirty-seven Chinese undergraduates from Beijing and 38 German undergraduates from Berlin participated in this study. The Chinese students were all native speakers of Mandarin Chinese, and the German students were all native speakers of German. The participants in this study had not participated in Experiment 1.

Materials and Procedure

Questionnaire booklets were prepared using the same 14 quintuplets of objects used in Experiment 1. Each booklet consisted of two sections: similarity judgment questions and property induction questions. The 14 quintuplets were divided into two groups (Item Group A and Item Group B) and were used either for the similarity judgment task or for the property induction task to construct two versions of the booklet. Booklet Type 1 consisted of the similarity judgment questions constructed from Item Group A and the property induction questions constructed from Item Group B. Booklet Type 2 was constructed with the reverse combination of Item Groups A and B: Item Group B was used for the similarity judgment questions and Item Group A was used for the property induction questions. We constructed the booklets this way to avoid monotonous responses from participants making the same type of judgments all the way through the task.

Each booklet included 28 object pairs for similarity judgments and 28 object pairs for property induction, representing four pairs of relations around the same target object from the 14 sets of quintuplets. Each of the item pairs represented one of the four types of relations around the same target (same classifier, taxonomically related, thematically related, unrelated). Twelve pairs of unrelated objects that were drawn from outside the 14 quintuplets were added to each part of the questionnaire as filler items in order to prevent pairs from the same target item following too close to one another. Thus, there were 40 similarity judgment questions and 40 property induction questions in the booklet. In both versions of the booklet, the similarity judgment questions were presented before the property induction questions, as we believed that similarity judgments were more likely to be influenced by a preceding task (e.g., A. B. Markman & Gentner, 1993). Within each section, the 40 similarity judgment questions (the 28 pairs plus 12 fillers) and the 40 property induction questions (the 28 pairs plus fillers) were presented in different orders across participants.

For the similarity judgment task, each pair was presented with a scale of 1 (*very dissimilar*) to 7 (*very similar*). For the property induction task, we followed Lin and Murphy (2001), using the property "same bacteria," as most of our items were artifacts. Participants were asked "How likely is it that [Object 1] and [Object 2] carry the same bacteria?" and judged the likelihood on a rating scale of 1 (*not likely at all*) to 7 (*very likely*).

The participants in both culture–language groups randomly received either Booklet Type 1 or Booklet Type 2. They were instructed to go through the questionnaire carefully at their own pace and to rely on their intuition.

Results

We report the results of similarity judgments followed by property induction. For both similarity judgments and property induction, we first report the pattern of results for the four types of relations separately for the Chinese and German groups so that readers can see the relative order of the four relations within each culture. We also examined the possible influence of German gender grammar on similarity judgments[6] and property induction.[7]

[6] To test whether German participants' similarity judgments were affected by grammatical gender categories, we classified all item pairs as either same-gender pairs or different-gender pairs. The German speakers' mean similarity rating for the same-gender pairs was 2.77 (averaged across the four relation types), and that for the different-gender pairs was 2.42. The Chinese speakers' ratings for the same-gender pairs and for the different-gender pairs were 3.30 and 3.00, respectively. Similarity ratings across same-gender and different-gender pairs differed both for Germans, $t(37) = 4.56$, $p < .01$, $d = 0.741$, and for Chinese, $t(36) = 2.44$, $p < .05$, $d = 0.401$. These results suggested that same-gender pairs were rated more highly than the different-gender pairs in both cultures. To test whether this effect was stronger in Germans than in Chinese, we conducted a 2 (culture) × 2 (gender congruency) ANOVA. A main effect for gender congruency was detected, $F(1, 72) = 112.313$, $p < .001$, $\eta^2 = .609$, but the critical effect for our object of concern (a Culture × Gender Congruency interaction) was not significant, $F(1, 72) = 0.023$. It has been pointed out that there are semantic regularities, though subtle, for gender class assignment in German (Zubin & Köpcke, 1986). For example, beasts of prey, birds, types of minerals, and engine-powered boats tend to be assigned to the masculine gender, whereas lower animals (such as reptiles, insects, and spiders) and wind-powered boats tend to receive the feminine gender. Thus, it is not particularly surprising that the same-gender pairs received higher similarity ratings than the different-gender pairs. However, unlike the case with the same-classifier relation, this gender congruency effect was not any stronger for the Germans than for the Chinese. Thus, it is unlikely that the similarity judgments in this study were confounded by grammatical gender on the part of the German participants.

[7] We also examined a potential influence from German grammatical gender on property induction (Experiment 2). In contrast to similarity judgments, a 2 (culture) × 2 (gender congruency) ANOVA revealed an interaction effect, $F(1, 68) = 8.63$, $p < .05$, $\eta^2 = .113$, but not in the expected way: It was the Chinese participants who rated same-gender pairs more highly. This result may be explained by an interplay of two facts: The Chinese participants in general tended to give higher ratings than the Germans, and there were more same-gender pairs than different-gender pairs in the questionnaire. In any case, grammatical gender did not seem to have affected the performance of German speakers on this task either.

Similarity Judgments

Within-culture analysis. Figure 1 shows the mean similarity ratings for the four relation types for the Chinese and German groups. The overall response pattern across the four relation types was very similar across the two cultural groups. Participants in both groups rated similarity in the following order: taxonomic pairs, thematic pairs, same-classifier pairs, and control (unrelated) pairs. A one-way repeated measures ANOVA on the mean similarity score across the four relations was carried out separately for each cultural group. For both cultures, the effect for the relation type was highly significant, $F(3, 108) = 126.70$, $p < .01$, $\eta^2 = .779$, for the Chinese, and $F(3, 111) = 113.17$, $p < .01$, $\eta^2 = .754$, for the Germans. A set of preplanned contrasts was conducted to examine (a) whether and how ratings for same-classifier pairs as well as taxonomic and thematic pairs differed from control pairs, (b) whether the classifier effect was comparable to the taxonomic or thematic relations, and (c) whether there was a significant difference between ratings for taxonomic and thematic pairs. For both cultures, we found that similarity among items sharing the same classifier in Chinese was rated significantly higher than similarity among unrelated items, $F(1, 36) = 74.83$, $p < .01$, $\eta^2 = .675$, for the Chinese; and $F(1, 37) = 23.69$, $p < .01$, $\eta^2 = .390$, for the Germans. Likewise, the taxonomically related items and the thematically related items were rated higher than the control items: taxonomic items, $F(1, 36) = 307.134$, $p < .01$, $\eta^2 = .895$, for the Chinese, and $F(1, 37) = 413.611$, $p < .01$, $\eta^2 = .918$, for the Germans; thematic items, $F(1, 36) = 148.06$, $p < .01$, $\eta^2 = .804$, for the Chinese, and $F(1, 37) = 46.95$, $p < .01$, $\eta^2 = .559$, for the Germans. In both cultures, however, the classifier effect was significantly smaller than both the taxonomic effect, $F(1, 36) = 125.51$, $p < .01$, $\eta^2 = .777$, for the Chinese, and $F(1, 37) = 261.90$, $p < .01$, $\eta^2 = .876$, for the Germans; and the thematic effect, $F(1, 36) = 33.01$, $p < .01$, $\eta^2 = .478$, for the Chinese, and $F(1, 37) = 18.48$, $p < .01$, $\eta^2 = .333$, for the Germans. A final contrast revealed that taxonomic pairs were rated more highly than thematic pairs both by Chinese and German participants, $F(1, 36) = 35.94$, $p < .01$, $\eta^2 = .500$, and $F(1, 37) = 79.73$, $p < .01$, $\eta^2 = .683$, respectively.

Comparison across the two culture–language groups. In testing whether the magnitude of the effect for each target relation differed across the two cultures, we were interested only in the Culture × Relation Type interaction and not in the main effect of culture. However, as seen in Figure 1, Chinese participants tended to give higher ratings overall, including for the control pairs. To adjust this baseline difference across the two groups, we transformed each participant's rating scores into standardized z scores (within each participant) and conducted a 2 (culture) × 4 (relation type) repeated measures ANOVA on them (see Table 3). We only report the Relation Type × Culture interaction, because the main effect of relation type is not of our interest here.

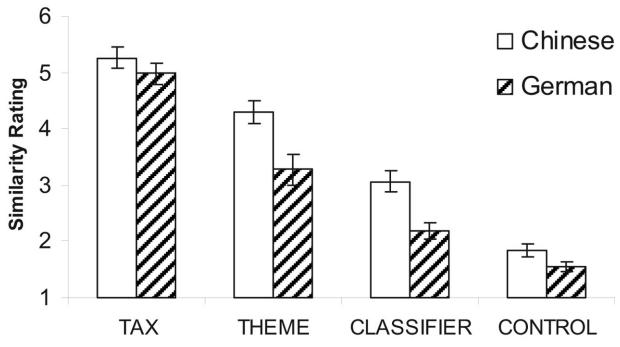

Figure 1. Mean similarity ratings for each relation type in each culture in Experiment 2 (similarity judgment). Tax = taxonomic item; Theme = thematic item.

The Relation Type × Culture interaction effect was significant, $F(3, 216) = 6.53$, $p < .05$, $\eta^2 = .083$. We then conducted a set of preplanned contrasts to decompose the overall interaction effect. Of primary interest was (a) whether the classifier effect was stronger for Chinese than for Germans and (b) whether Chinese and German participants relied on taxonomic and thematic relations in different degrees. We thus examined the effect of culture on the classifier versus control contrast and on the taxonomic versus thematic contrast. The effect of culture was significant in the classifier versus control contrast, $F(1, 72) = 10.54$, $p < .01$, $\eta^2 = .128$, suggesting that the classifier effect was larger for the Chinese (−0.283 vs. −0.934) than for the Germans (−0.414 vs. −0.740). As we noted in reporting the within-culture analysis, both the Chinese and the Germans rated the taxonomic pairs higher than the thematic pairs; however, relative reliance of taxonomic and thematic relations was slightly, but statistically significantly, different across the two groups, as indicated by a significant effect of culture on the taxonomic versus thematic contrast (0.863 vs. 0.354, taxonomic vs. thematic, respectively, for Chinese; and 1.046 vs. 0.108 for Germans), $F(1, 72) = 10.44$, $p < .01$, $\eta^2 = .127$.

Property Induction

Within-culture analysis. Figure 2 shows the mean rating scores for the likelihood of the pairs sharing the same property (carrying the same bacteria) for each relation type. For similarity judgments, we first tested the within-culture pattern across the taxonomic, thematic, same-classifier, and unrelated (control) items within each cultural group. We found a highly significant effect for relation type for both the Chinese, $F(3, 102) = 58.71$, $p < .01$, $\eta^2 = .633$, and the Germans, $F(3, 102) = 51.47$, $p < .01$, $\eta^2 = .602$. A series of preplanned contrasts revealed that, unlike in the case of similarity judgments, classifier membership did not influence participants' ratings relative to the control condition either for the Chinese, $F(1, 34) = 0.01$, $p > .10$ or for the Germans, $F(1, 34) = 1.36$, $p > .1$. Both Chinese and German participants judged the taxonomically related pairs and thematically related pairs as more likely to carry the same bacteria than the unrelated pairs: Chinese, $F(1, 34) = 106.29$, $p < .01$, $\eta^2 = .758$, for taxonomic, and $F(1, 34) = 37.78$, $p < .01$, $\eta^2 = .526$, for thematic; German, $F(1, 34) = 78.10$, $p < .01$, $\eta^2 = .697$, for taxonomic, and $F(1, 34) = 70.96$, $p < .01$, $\eta^2 = .676$, for thematic. The taxonomic pairs received higher likelihood ratings than the thematic pairs from the Chinese participants, $F(1, 34) = 14.64$, $p < .01$, $\eta^2 = .301$. In contrast, the likelihood ratings for the taxonomic and thematic pairs did not differ among the German participants, $F(1, 34) = 0.002$, $p > .10$.

Comparison between the two culture–language groups. Just as for the similarity judgments, we conducted a 2 (culture) × 4 (relation type) ANOVA on standardized scores in order to test

Table 3
Z Scores for Each Relation Type in Each Culture in Experiment 2

		Similarity Judgments				Property Induction			
Culture	*n*	Tax	Theme	Classifier	Control	Tax	Theme	Classifier	Control
Chinese	37	0.86	0.35	−0.28	−0.93	0.63	0.26	−0.45	−0.45
German	38	1.05	0.11	−0.41	−0.74	0.50	0.45	−0.44	−0.50

Note. Tax = taxonomic item; Theme = thematic item.

whether the pattern of response differed across the two cultural groups (see Table 3). The overall interaction effect was not significant, $F(3, 204) = 2.20$, $p > .05$ (observed power = .57). The two groups did not differ on the classifier versus control contrast, $F(1, 68) = 0.46$, $p > .1$. There was a significant culture effect on the contrast between the taxonomic pairs and thematic pairs, $F(1, 68) = 4.44$, $p < .05$, $\eta^2 = .061$, but in an unexpected direction: The taxonomic pairs were rated more highly by the Chinese than the Germans (0.63 vs. 0.50), whereas the thematic pairs were rated more highly by the Germans than by the Chinese (0.45 vs. 0.26).

Comparison Between Similarity Judgment and Property Induction

The results of the analyses so far have indicated that the effect of culture was observed in the similarity judgment task but not in the property induction task. Confirming this, a 2 (task: similarity judgment vs. property induction) × 2 (culture) × 4 (relation type) ANOVA revealed a highly significant three-way (Task × Culture × Relation Type) interaction, $F(3, 420) = 7.66$, $p < .01$, $\eta^2 = .052$.

Discussion

The pattern of results in Experiment 2 suggests that Chinese and Germans utilize both taxonomic and thematic relations as bases for similarity ratings and property induction. In both groups, similarity ratings for the four types of pairs had the order of taxonomic, thematic, same classifier, and control. We first discuss the influence of the classifier system in light of the results and then discuss whether the pattern of the results bears out for Nisbett and colleagues' culture-specific cognition proposal.

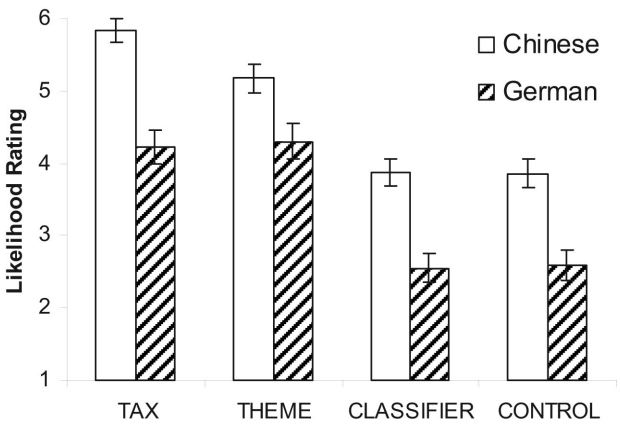

Figure 2. Mean likelihood ratings for property induction (on the "carrying the same bacteria" question) for each target type in each culture in Experiment 2. Tax = taxonomic item; Theme = thematic item.

Consistent with the results of Experiment 1, both the Chinese and German participants rated the same-classifier pairs as more similar than the control pairs. This result suggests that even speakers of a nonclassifier language can detect an inherent similarity between objects belonging to the same classifier category. However, this inherent similarity may be magnified for speakers of the classifier language, as Chinese people's similarity judgments for pairs drawn from the same classifier classes were higher than those of the Germans.

Although both Chinese and German participants noted the similarity underlying classifier categories in the similarity judgment task, neither utilized this similarity in inductive reasoning, as participants in neither group rated the same-classifier items as having a higher probability than the control items of carrying the same bacteria as the target object. The absence of any classifier effect in property induction may suggest that speakers of a classifier language do not use classifier categories as basis for inductive reasoning. This pattern is perfectly consistent with the second possibility we discussed earlier regarding the form of the influence of the classifier system: The classifier system does not function as an organizer of object concepts for speakers of classifier languages, but the classifier categorization system heightens sensitivity to semantic features underlying the classifier categories.

The fact that Chinese speakers gave higher similarity ratings for thematically related object pairs than the Germans did may indeed be taken as support for the culture-specific cognition hypothesis advanced by Nisbett and colleagues (Ji et al., 2004; Nisbett et al., 2001). However, these effects need to be qualified in two respects: (a) In both the similarity judgment and property induction tasks, Chinese speakers gave higher ratings for the taxonomic pairs than for the thematic pairs; and (b) Nisbett and colleagues' hypothesis did not bear out for the pattern of the results in the property induction task.

Why were the Chinese participants sensitive to classifier relations in judging similarity but did not utilize them as a basis for inductive reasoning? It has been well established that inductive inferences depend on the kind of properties to be projected (e.g., Heit & Rubinstein, 1994; Nisbett, Krantz, Jepson, & Kunda, 1983; see also Gelman, Collman, & Maccoby, 1986). As discussed earlier, the participants may have judged that the taxonomic and thematic items were likely to carry the same bacteria by recruiting specific knowledge that things of the same kind may provide similar living conditions for a kind of bacteria, and that things that

co-occur in space and time may carry the same bacteria though physical contact (Lin & Murphy, 2001). In this situation, there is little basis for the induction of this property (carrying the same bacteria) across objects in the same classifier category, because classifier categories are not kinds of categories in which members are held together by being in contact with one another or by having the same internal constitution.

To examine whether classifier categories carried any inductive potential, and whether this effect (if any) was culture specific, we replicated the property induction task of Experiment 2 with a blank property.

Experiment 3: Inductive Reasoning With a Blank Property

Method

Participants

Twenty-three Chinese undergraduates from Beijing and 22 German undergraduates from Berlin participated in this study. The Chinese participants were all native speakers of Mandarin Chinese, and the German participants were all native speakers of German. None of the participants had participated in any of the previous studies.

Materials

The stimulus set was the same as that used in Experiments 1 and 2. Fifty-six item pairs of objects constructed from the 14 quintuplets and 24 unrelated filler pairs were presented in a questionnaire booklet. Three different versions of the booklet were prepared, in which the item pairs were presented in different orders. Participants were randomly assigned to a version of the booklet. For every item pair, participants saw the following question: "Suppose that property X is an important property for [Object 1]. If [Object 1] has property X, how likely is it that [Object 2] has also property X?" They were asked to judge the likelihood on a rating scale of 1 (*not likely at all*) to 7 (*very likely*). Object 1 was the target object of each quintuplet, and Object 2 was one of the four items representing each target relation.

Procedure

Each participant received a copy of the booklet. Participants were instructed to go through the questionnaire carefully at their own pace and to rely on their intuition.

Results

Within-Culture Analyses

Figure 3 shows the mean likelihood ratings for the four relation types for each of the two cultural groups. As in Experiment 2, we first examined the pattern of likelihood ratings for the four target relations separately for each culture. Again, the overall response pattern across the four target relations was strikingly similar across the two culture–language groups. The effect of relation type was highly significant for both groups, $F(3, 66) = 142.20$, $p < .01$, $\eta^2 = .866$, for the Chinese, and $F(3, 63) = 114.55$, $p < .01$, $\eta^2 = .845$, for the Germans. Preplanned contrasts revealed that the likelihood ratings for items belonging to the same classifier category in Chinese were significantly higher than the ratings for unrelated items in both the Chinese, $F(1, 22) = 53.90$, $p < .01$, $\eta^2 = .710$, and German, $F(1, 21) = 31.81$, $p < .01$, $\eta^2 = .602$, groups. The taxonomically related items and thematically related items were also rated higher than the unrelated items in both cultural groups: taxonomic, $F(1, 22) = 325.78$, $p < .01$, $\eta^2 = .937$, for the Chinese, and $F(1, 21) = 549.71$, $p < .01$, $\eta^2 = .963$, for the Germans; and thematic, $F(1, 22) = 112.39$, $p < .01$, $\eta^2 = .836$, for the Chinese, and $F(1, 21) = 28.12$, $p < .01$, $\eta^2 = .573$, for the Germans. As in similarity judgments, the classifier effect turned out to be significantly smaller in both cultures than both the taxonomic effect, $F(1, 22) = 269.97$, $p < .01$, $\eta^2 = .925$, for the Chinese, and $F(1, 21) = 424.81$, $p < .01$, $\eta^2 = .953$, for the Germans; and the thematic effect, $F(1, 22) = 52.28$, $p < .01$, $\eta^2 = .704$, for the Chinese, and $F(1, 21) = 10.25$, $p < .01$, $\eta^2 = .328$, for the Germans. Finally, the taxonomic effect was found to be significantly stronger than the thematic effect among both the Chinese and Germans, $F(1, 22) = 37.63$, $p < .01$, $\eta^2 = .925$, and $F(1, 21) = 67.38$, $p < .01$, $\eta^2 = .762$, respectively.

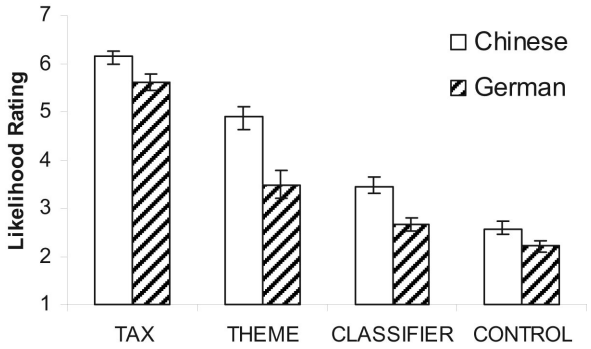

Figure 3. Mean likelihood ratings for property induction (on the blank property) for each target type in each culture in Experiment 3. Tax = taxonomic item; Theme = thematic item.

Comparison of the Chinese and German Groups

To test the effect of the culture and/or language, we again adjusted the baseline difference between the two groups by transforming the likelihood rating scores into standardized *z* scores (see Table 4). A significant Relation Type × Culture interaction effect was found, $F(3, 129) = 5.34$, $p = .01$, $\eta^2 = .110$, which was then followed up by a set of preplanned contrasts. As in similarity judgments in Experiment 2, the classifier effect was larger for the Chinese (same classifier: −0.427, control: −0.880) than for the Germans (−0.423 vs. −0.679), which was manifested by a significant effect of culture on the classifier versus control contrast, $F(1, 43) = 6.86$, $p < .05$, $\eta^2 = .138$. Likewise, relative reliance on thematic relations was stronger for the Chinese (thematic: 0.329, control: −0.880) than for the Germans (thematic: 0.048; control: −0.679), as suggested by a significant effect of culture on the taxonomic versus thematic contrast, $F(1, 43) = 4.77$, $p < .05$, $\eta^2 = .10$, although the taxonomic pairs received higher ratings than the thematic pairs in both cultural groups.

Table 4
Z Scores for Each Relation Type in Each Culture in Experiment 3 (Induction of a Blank Property)

Culture	*n*	Tax	Theme	Classifier	Control
Chinese	23	0.98	0.33	−0.43	−0.88
German	22	1.05	0.05	−0.42	−0.68

Note. Tax = taxonomic item; Theme = thematic item.

As in the previous experiments, we also tested the possibility that grammatical gender in German might have influenced likelihood ratings.[8]

Comparison to the Results of Experiment 2

The results from the blank property induction in this experiment, especially that concerning the effect of culture, were very similar to the results of the similarity judgment task in Experiment 2. In contrast, the effect of culture seemed very different across the two contexts of property induction—the context in which participants were able to access background knowledge for their reasoning (the bacteria question in Experiment 2) and the context in which participants needed to draw induction without any background knowledge (this study). To confirm this, we conducted two ANOVAs, one comparing the similarity judgments and the blank property induction (this study) and the other comparing the two property induction tasks (bacteria vs. blank). The critical effect of interest here was the three-way interaction among Task, Culture, and Relation Type. Confirming the above observations, the three-way interaction was highly significant for the analysis comparing the two property induction tasks, $F(3, 333) = 6.01$, $p < .01$, $\eta^2 = .051$, but not for the analysis comparing the similarity judgment and the blank property induction tasks, $F(3, 345) = 0.641$, $p > .10$.

Discussion

The pattern of the results of this study was strikingly similar to the pattern observed for similarity judgments in Experiment 2. Participants in both cultures rated the likelihood in the order of taxonomic, thematic, classifier, and unrelated items. The results of this study were somewhat different from those of the property induction task in Experiment 2, in that culture-specific effects of the classifier relations and thematic relations were found here.

Unlike the bacteria property case, inference of a blank property did not allow the participants to recruit any specific knowledge. In this situation, they had nothing to resort to except for similarity; when there were no salient relations such as taxonomic or thematic relations, the participants might have sought for any common properties, including features underlying classifier categories (e.g., both objects are thin and long, or both objects have handles) as bases for induction. In this process, magnified sensitivity to features underlying classifier categories may have led Chinese participants to give higher likelihood ratings, as was the case with similarity judgments. In any case, it seems to be reasonable to conclude that same-classifier relations have some inductive potential in the sense that they can be a basis for similarity. However, their inductive potential is limited in situations in which background knowledge is not readily available for the induction.

Experiment 4: Priming

The final experiment examined whether the language- and culture-specific differences that we had observed in the similarity judgments (Experiment 2) and inductive inference of a blank property (Experiment 3) tasks were also observed in a task that accessed fast and automatic processes. For this purpose we used a version of a semantic priming paradigm. It is widely known that recognition of a word involves activation of its corresponding node in a semantic network, and a priming effect is observed when two objects that are presented sequentially are conceptually related (Anderson, 1983; Joordens & Becker, 1997; Tulving & Schacter, 1990). Yokosawa and Imai (1997) demonstrated that the conceptual priming effect is observed in picture recognition as well. In their study, participants (Japanese adults) saw a cue, which was presented either orthographically or pictorially, followed by a target picture. The participants were to judge whether the target matched the cue. Yokosawa and Imai found that regardless of whether the cue was presented as a word or a picture, when the cue was taxonomically (e.g., dog) or thematically (e.g., carrot) related to the target (e.g., rabbit), the participants took more time to judge that the cue and the target were different objects (at the basic level) than when the cue was unrelated to the target (e.g., hammer).

We borrowed this paradigm to test whether (a) the priming effect for taxonomic and thematic relations in the automatic picture recognition process found in Japanese adults could be replicated in Chinese and Germans, (b) whether the influence of the classifier relations would be observed in Chinese but not in German speakers, and (c) whether the thematic effect would be larger for the Chinese than for Germans and the taxonomic effect larger for Germans than for the Chinese. We presented the cues verbally but presented the target pictorially for the following reasons. If we had presented both the cue and the target in words, culture-specific influence due to writing systems might have influenced the response latencies (e.g., morphosyllabic Chinese characters may provide faster access to word meaning; see Saalbach & Stern, 2004). However, if we had presented both the cue and the target in pictures, the visual similarity of the two pictures (rather than conceptual relations between the two objects) might have affected the response latencies (see the General Discussion for further discussion of this issue).

Method

Participants

Twenty-one Chinese undergraduates from Beijing and 23 German undergraduates from Berlin participated in this study. The Chinese students were all native speakers of Mandarin Chinese,

[8] The German speakers' average likelihood rating for the same-gender pairs was 3.27, and that for the different-gender pairs was 2.92. The Chinese speakers' ratings for the same-gender pairs and for the different-gender pairs were 3.95 and 3.68, respectively. A 2 (Culture) × 2 (Gender Congruency) ANOVA on the standardized scores revealed a significant main effect for Gender Congruency, $F(1, 45) = 149.58$, $p < .01$, $\eta^2 = .769$, but not a Gender Congruency × Culture interaction, $F(1, 45) = 0.10$, $p < .01$. Thus, again, it is unlikely that the cultural differences obtained in this experiment were confounded with any influence of grammatical gender in German.

and the German students were all native speakers of German. None of the participants had been involved in the previous experiments.

Materials and Procedure

The same 14 quintuplets used for Experiments 1 through 3 were used for this study. The target object was presented pictorially, and the cue representing one of the four relations (taxonomic, thematic, same classifier, unrelated) was presented as a written word. The participants were instructed to verify whether the picture matched the word they had just seen. The pictures (simple black-and-white line drawings) were derived from Snodgrass and Vanderwart (1980) when available. For the items that were not available from the Snodgrass and Vandewart stimuli, we had an artist draw pictures in the same style. Each word (in the respective language) and each picture was presented in black on a white background in the middle of the screen. Each participant went through a total of 112 test trials, half of which required a positive response (prime and target matched) and the remaining half of which required a negative response (prime and target did not match). Target pictures appeared only once in the positive trials. The rest of the positive trials consisted of filler items to avoid presenting the target pictures too many times. The order of the prime–target pairs was random with the constraint that the same target could not appear within a three-trial window.

The Chinese and German participants were all tested individually in a university laboratory. They were told that they would see a word followed by a short presentation of a picture and were asked to judge as fast as possible whether the word and the picture matched. The participants first saw a fixation cross for 1,500 ms followed by the presentation of the prime word for 1,000 ms. Another fixation cross appeared and stayed for 500 ms. The target picture was then presented for 200 ms, followed by a pattern mask that remained until the participants gave their response. After the response, a new trial started with the presentation of the fixation cross for 1,500 ms. All fixations, words, and pictures presented during the trials appeared in the center of the screen. Before starting the test trials, the participants received 15 practice trials to guarantee their correct understanding of the task and the keys assigned to yes and no.

Results

Analysis of Error Responses

In general, error rates for the negative trials were low across all relation types for both groups (see Table 5), except for two taxonomically related pairs (mountain/hill: 73%, bell/buzzer: 54%). These pairs were excluded from further analyses on error rates as well as on response latencies. In general, error rates differed across four different relation types, with highest error rates on the taxonomic items (9.4%, averaged across the two groups), $F(3, 126) = 13.22$, $p < .01$, $\eta^2 = .239$. Furthermore, the Chinese participants overall made more errors (5.26%) than the Germans (3.35%), $F(1, 42) = 3.99$, $p = .05$, $\eta^2 = .087$. However, there was no interaction between the two factors, $F(3, 126) = 1.64$, $p > .1$. In particular, the error rates for the same classifier were very low for both groups (Chinese: 1.98%, German: 1.45%) and did not differ from the error rates for the control item.

Table 5
Error Rates for Each Relation Type in Each Culture in Experiment 4 (Priming)

Culture	*n*	Tax	Theme	Classifier	Control
Chinese	21	11.91	4.76	1.98	2.36
German	23	6.88	2.17	1.45	2.90

Note. Tax = taxonomic item; Theme = thematic item.

Analysis of Response Latencies

Trials including false responses were excluded in the analysis of response latencies. The response latencies for the positive trials (in which the word prime and picture target matched) did not differ across the two cultures. We thus only focus on the response latencies for the negative trials. Figure 4 shows the mean response latencies for each relation type across the two groups. As in Experiments 2 and 3, the overall response pattern across the four relation types was very similar across the two culture–language groups. The effect for relation type was highly significant for both the Chinese, $F(3, 60) = 10.91$, $p < .01$, $\eta^2 = .343$, and the Germans, $F(3, 66) = 23.81$, $p < .01$, $\eta^2 = .520$. Somewhat unexpectedly, the preplanned contrasts revealed that the latencies for the same-classifier primes were slower than those for the unrelated primes in the German group, $F(1, 22) = 5.83$, $p < .05$, $\eta^2 = .209$, but not in the Chinese group, $F(1, 20) = 0.41$, $p > .10$. Further contrasts revealed that in both groups, response latencies for the trials with taxonomically and thematically related primes were significantly longer than those for the trials with unrelated primes: taxonomic, $F(1, 20) = 20.31$, $p < .01$, $\eta^2 = .504$, for Chinese, and $F(1, 22) = 33.23$, $p < .01$, $\eta^2 = .602$, for German; thematic, $F(1, 20) = 8.09$, $p < .01$, $\eta^2 = .288$, for Chinese, and $F(1, 22) = 49.28$, $p < .01$, $\eta^2 = .209$, for German. The latencies for the trials with taxonomic and thematic primes were not different from one another in either culture, $F(1, 20) = 1.08$, $p > .1$, for Chinese, and $F(1, 22) = 0.33$, $p > .5$, for German. Thus, the results revealed a strong influence of taxonomic relations and thematic relations on the word–picture matching task for both culture groups. However, classifier relations did not affect the latencies for the Chinese participants.

Comparison of the Two Groups

To test whether there were any culture- or language-specific effects, we transformed the reaction latencies into standardized *z* scores (see Table 6). The critical interaction effect between relation type and culture was not significant, $F(3, 126) = 2.15$, $p = .10$, (power = .54) Thus, unlike the similarity judgment task and the inductive reasoning with a blank property task, a culture- or language-specific effect of classifiers was not observed here.

As in the earlier studies, we also tested whether German grammatical gender had any particular impact on the response latencies.[9]

Discussion

The pattern of Experiment 4 was similar to the results of the earlier studies in that taxonomic and thematic relations strongly affected the participants' performance in both the Chinese and German groups. In contrast, the classifier relation did not influence picture recognition of the target object in the Chinese group. Taken together, it appears that the language-specific impact of classifier classification system obtained for the unspeeded similarity judgments did not hold for a task that required fast, automatic cognitive processes. The lack of the classifier effect among the Chinese speakers suggests that objects belonging to the same classifier category are not included in a semantic network that is automatically activated when the target object is accessed.

It should also be noted that even though thematic items were perceptually very dissimilar to the target objects, there was a pattern of interference in the responses of both culture groups. This result again supports the proposal that thematic relations are a universally important and integral part of concept structure (Lin & Murphy, 2001; Wisniewski & Bassok, 1999), and, unlike classifier relations, thematically related objects are included in the semantic network that is automatically activated with the target object. However, this thematic effect was no larger for the Chinese group than for the German group, inconsistent with the culture-specific cognition proposal (Ji et al., 2004).

General Discussion

Does the Classifier System Influence the Speakers' Conceptual Structure of Everyday Objects?

The main aim of this research was to examine whether a classifier categorization system affects speakers' conceptual structures in significant ways. Specifically, we tested three possible patterns regarding this question: (a) Classifier categories function as alternative organizers of concepts and categories (cf. Lakoff, 1987; Zhang & Schmitt, 1998), (b) speakers of classifier languages attend to features underlying classifier categories more strongly than speakers of nonclassifier languages in certain cognitive contexts (cf. Hunt & Agnoli, 1991), and (c) classifiers are frozen linguistic conventions without any cognitive impact. We found some support for the second but not the first possibility. The Chinese participants gave higher ratings to the same-classifier pairs than the German participants both in similarity judgments and inductive inference of a blank property, which suggests that classifier categories have some impact on Chinese speakers' conceptual structure of everyday objects. However, it would be an overstatement, given this effect, to say that classifier categories serve as an additional or alternative basis for organizing concepts, because the magnitude of the classifier effect was limited compared to that of the other major relations that organize people's concepts. Throughout the four experiments, which included a variety of cognitive activities, taxonomic and thematic relations revealed themselves to be important organizers of conceptual structures. In contrast, the impact of classifiers was much smaller in magnitude when it was found at all, and the culture- or language-specific influence of classifier categories was limited to just two tasks (i.e., similarity judgments and inductive reasoning of blank properties). Furthermore, it is important to note that German speakers were also sensitive to the similarity underlying classifier categories, even though the kind of similarity conveyed by classifier categories is very different from that conveyed by taxonomic relations or thematic relations. All in all, a plausible conclusion seems to be that the classifier categorization system does not organize speakers' concepts of objects. The classifier effect found among Chinese speakers is perhaps best characterized as a magnified sensitivity to semantic features underlying classifier categories developed through the habitual use of classifiers in association with the names of objects.

How are classifier categories different from taxonomic or thematic categories, which apparently play a much more important role in organizing concepts? As discussed earlier, classifier categories are held together only by a single or a combination of at most a few semantic features. This characteristic naturally leads to the consequence that category members do not have much in common, which probably makes classifier categories at best only weak conceptual categories. Seen this way, it does not seem unreasonable that the classifier effect was observed in similarity judgments and blank property induction but not in other tasks. People are very versatile in perceiving similarity, and even the commonality of a single feature can significantly affect the construal of similarity among objects. However, it is reasonable that people (both Chinese and Germans) did

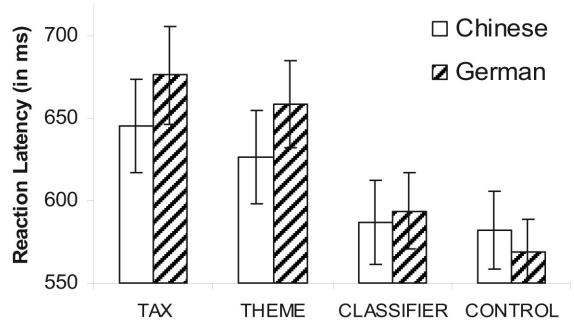

Figure 4. Mean response latencies for each relation type in each culture in Experiment 4. Tax = taxonomic item; Theme = thematic item.

Table 6
Z Scores for Each Relation Type in Each Culture in Experiment 4 (Priming)

Culture	*n*	Tax	Theme	Classifier	Control
Chinese	21	0.27	0.17	−0.20	−0.18
German	23	0.36	0.24	−0.21	−0.31

Note. Tax = taxonomic item; Theme = thematic item.

[9] Consistent with the results of the earlier studies, the latencies for the same-gender pairs were slower than the latencies for different-gender pairs, $F(1, 42) = 30.32$, $p < .01$, $\eta^2 = .419$. However, the critical Gender Congruency × Culture interaction was not detected, $F(1, 42) = 0.80$, $p > .10$. Thus, it is not likely that the results of Experiment 4 were influenced by German grammatical gender.

not make inductive generalizations of a property based on the kind of similarity that underlies classifier categories when they could access background knowledge.

Does the classifier system influence thought in any significant way other than magnifying attention to the semantic features underlying classifier categories? Perhaps cognitive influence of classifiers may be seen not so much as a long-term influence on the representation of objects per se but as a temporary shift in the construal of the referent by highlighting a feature that is not spontaneously activated by the noun label itself. In future research it will be interesting to see whether and to what extent classifiers change the construal of a given object in the context in which the classifier is used. This would be similar to the way speakers of English have very different images when they hear a noun used with mass syntax versus count syntax, as in "Jim had some chocolate" and "Jim had three chocolates" (e.g., Middleton, Wisniewski, Trindel, & Imai, 2004; Wisniewski, Lamb, & Middleton, 2003).

Do Classifier Categories Influence Cognition Beyond the Realm of Language?

Even though limited in magnitude, our research has demonstrated the influence of classifier categories at some level of cognition in Chinese speakers in that Chinese speakers utilized classifier relations to a larger degree than did German speakers in similarity judgments and inductive reasoning of an unknown property. This influence could hence be taken as support for a weak version of linguistic relativity but not for a strong version. One may question, however, whether the classifier effect in our research is truly an effect in a "nonlinguistic" cognitive realm or if the effect is still in the realm of language, as the stimuli were presented verbally (i.e., in words) rather than pictorially in our experiments (cf. Slobin, 1996).

We believe that the answer depends on how one defines *linguistic* and *nonlinguistic cognition*, which, of course, is not at all a simple question. One way to draw a line between linguistic and nonlinguistic effects is according to whether the effect is obtained with or without the explicit invocation of the target linguistic categories. For example, as reviewed earlier, Vigliocco, Vinson, Indefrey, Levelt, and Hellwig (2004) found that, in their semantic substitution paradigm, German speakers tended to produce substitution errors within the same gender category when speakers produced phrases with determiners marked for gender, but that this gender preservation effect disappeared when they produced bare nouns or phrases with determiners not marked for gender (a phrase with an indefinite determiner plus noun). In our case, the culture- or language-specific classifier effect was obtained when objects were presented without classifiers. Thus, it could be argued that the impact of classifier categories are stronger than the impact of gender categories in German, and the classifier effect goes beyond the realm of linguistic reference. Another way of distinguishing between linguistic and nonlinguistic effects is suggested in a study by Vigliocco et al. (2005). They used an odd-one-out categorization task to test whether grammatical gender categories would make a direct impact on semantic representation of Italian (whose syntax–semantic mapping concerning grammatical gender and sex is clearer than that of German) beyond the context of speech production. These authors found that Italian speakers chose two objects of the same gender category (in Italian) at a rate higher than did English speakers. Furthermore, the gender effect was not observed when the objects were presented in pictures instead of words. On the basis of these results, Vigliocco et al. (2005) concluded that gender categories in Italian exert their influence at the lexicosemantic level but not at the conceptual level.

If one follows this line of reasoning, one might conclude that our results demonstrate the influence of the classifier system within the realm of language only but not beyond. In our view, however, linguistic and nonlinguistic cognitive processes cannot be so simply distinguished, as pictorial presentations have their own limitations for assessing people's nonlinguistic concepts. First, even though the objects were presented in pictures, the participants may have unconsciously named the objects. Second, a picture may invoke a strong visual image of the particular instance of the object depicted in it. Thus, visual similarity among pictures may affect judgment of similarity more strongly than when objects are presented in words without specific visual images. In fact, one possible account for the disappearance of the gender effect with the pictorial stimuli in Vigliocco et al.'s (2005) study is that the concrete visual images of objects wiped out the subtle conceptual similarity between objects arising from gender category membership which had shown up in the similarity judgment task with words.

We, of course, acknowledge that it is extremely difficult to determine whether the kind of effects we found in this research stay in the realm of language or go beyond it. As stated earlier, we believe that the interpretation hinges on how one defines *linguistic* and *nonlinguistic cognition*, and we realize that almost every researcher in the field may have a different idea about this (e.g., Gennari, Sloman, Malt, & Fitch, 2002; Imai & Gentner, 1997; Imai & Mazuka, 2003, in press; Lucy, 1992; Malt, Sloman, Gennari, Shi, & Wang, 1999; Slobin, 1996; Vigliocco et al., 2005). In any case, we believe that it is more important to specify exactly when and how classifier categories affect cognitive processes and representations involving objects than to characterize the classifier effect along the dimension of being linguistic or nonlinguistic.

Universal Significance of Thematic Relations for Conceptual Structures and Evaluation of the Culture-Specific Cognition Proposal

The results of this research are also important in that they extend the recent recognition of the importance of thematic relations in human object concepts (e.g., Bassok & Medin, 1997; Lin & Murphy, 2001; Wisniewski & Bassok, 1999) not only to a sample from a different Western cultural background (European rather than American) but also to a sample from an East Asian cultural background. In fact, in the forced choice categorization task (Experiment 1) and the word–picture matching study (Experiment 4), the thematic effect was stronger than, or equally as strong as, the taxonomic effect in both the Chinese and German groups. Furthermore, thematic relations proved to be important not only in categorization and similarity judgments, but also in inductive reasoning both when some prior knowledge was accessible (carrying-the-same-bacteria property) and when it was not (a blank property). These results endorse Lin and Murphy's proposal that thematic relations have actual conceptual content, and they hence support the view that thematic relations are an integral part of people's concepts, along with taxonomic relations. As suggested

by Lin and Murphy, concepts may be intertwined with background knowledge of events, scenes, and causal relations, and people may use this knowledge spontaneously and automatically in all major cognitive activities, even when this is task inappropriate (Wisniewski & Bassok, 1999), as in similarity judgments and in word–picture matching.

The universal importance of thematic relations is deeply relevant to Nisbett and colleagues' proposal concerning the cross-cultural difference in cognition across Easterners and Westerners (Ji et al., 2004; Nisbett, 2003). Despite the universal importance of thematic relations, our results in part support their hypothesis, in that we did find that the Chinese participants gave thematic relations higher similarity ratings (Experiment 2) as well as higher likelihood judgments in inductive inferences of an unknown property (Experiment 3) than did the Germans. However, as was the case for the culture-specific classifier effect, the culture-specific preference effect was miniscule rather than global and pervasive, as the culture-specific thematic preference in the Chinese people was not observed in the categorization (Experiment 1), inductive inference of a known property (Experiment 2), or speeded word–picture matching (Experiment 4) tasks. In contrast, what was more noted across all of the studies was the strong similarity across the two groups from very different cultural backgrounds in their reliance on taxonomic relations and thematic relations. In this sense, the conclusion advanced by Ji et al. that Westerners organize their concepts around taxonomic category relations and Easterners organize their concepts around thematic relations should clearly be tempered. However, of course, this cross-cultural similarity does not preclude the possibility that there are significant cross-cultural differences in the style of cognition in other conceptual domains, such as reasoning and/or attention to social relations (e.g., Choi, Nisbett, & Smith, 1997).

Implications for Research on the Relation Between Language and Thought

This research provides important implications for the field of language and thought both theoretically and methodologically. First, this research highlights the importance of examining the effect of given language-specific categories in a range of cognitive tasks, as the effect may be observed in one type of cognitive activity but not in others. The fact that we obtained the classifier influence in the inductive reasoning task with a blank property but not in the same task with a concrete property (carrying the same bacteria) should be particularly noted in this respect, as it suggests that the influence of linguistic categories deeply interacts with availability of background knowledge. Second, this research also highlights the importance of examining the influence of linguistic categories not in light of whether there is one, but in light of how large the influence is in a whole spectrum of concepts, and how it is related to other major conceptual relations underlying conceptual structure. Relevant to this point, the fact that German participants judged objects belonging to the same classifier category as more similar than unrelated objects supports the notion that grammatical categories are motivated (but of course not determined) by universally shared cognitive and perceptual experience (e.g., Zubin & Köpcke, 1986). In this sense, our results are not incongruent with the view that grammatical categories are a reflection of cognitive categories (e.g., Lakoff, 1987). At the same time, our results suggest that grammatical categories such as classifier categories do not function as a major organizer of concepts to the same extent that taxonomic and thematic categories do. More importantly, these two aspects of our results—that similarity underlying classifier categories can be detected by German speakers on one hand and that similarity due to classifier relations is magnified by Chinese speakers on the other hand—cogently suggest that the relation between language and thought is not unidirectional: Linguistic categories reflect universally perceived commonalities in the world, but at the same time they modify universally perceived (unbiased) similarities (see Imai and Mazuka, in press, for a relevant discussion).

Another aspect that should be considered in examining the effect of language, which was not dealt with in this article, is whether the effect of a target grammatical categorization holds across all languages having that grammatical function. For example, Vigliocco et al. (2005) found the effect of grammatical gender in Italian speakers' lexicosemantic representation, but this effect was not found in German speakers. In our case, it is important to examine whether the classifier effect we found among Chinese is also found in speakers of other classifier languages, languages in which the semantic function as well as the grammatical function of the classifier categories are not identical to those of Chinese. An examination of this question is currently underway with Japanese.

To conclude, we suggest that the simple Whorfian/non-Whorfian dichotomy does not deepen our understanding of the nature of these concepts and categories very much, given the complexity of the interactions among many of the factors that affect the structure concepts and cognitive processes. What is important, then, is to clarify how, rather than whether, language-specific categories, be they grammatical or lexical, affect concepts, categories, and cognitive processes. It is particularly important to specify how the effect interacts with universal cognitive biases and constraints, the structure of the world, and constraints placed by the task or cognitive activity at hand (e.g., what type of information or knowledge is most relevant for the inference).

References

Adams, K., & Conklin, N. F. (1973). Toward a theory of natural classification. *Papers from the 9th meeting of the Chicago Linguistic Society.*

Aikhenvald, A. (2000). *Classifiers: A typology of noun categorization devices.* New York: Oxford University Press.

Allan, K. (1977). Classifiers. *Language, 53,* 285–311.

Anderson, J. R. (1983). *The architecture of cognition.* Cambridge, MA: Harvard University Press.

Bassok, M., & Medin, D. L. (1997). Birds of a feather flock together: Similarity judgments with semantically rich stimuli. *Journal of Memory and Language, 36,* 311–336.

Boroditsky, L. (2001). Does language shape thought? Mandarin and English speakers' conceptions of time. *Cognitive Psychology, 43,* 1–22.

Choi, I., Nisbett, R. E., & Smith, E. E. (1997). Culture, category salience, and inductive reasoning. *Cognition, 65,* 15–32.

Craig, C. G. (1986). Jacaltec noun classifiers: A study in language and culture. In C. Craig (Ed.), *Noun classes and categorization* (pp. 263–293). Philadelphia: John Benjamins.

Croft, W. (1994). Semantic universals in classifier systems. *Word, 45,* 145–171.

Denny, J. P. (1976). Semantic analysis of selected Japanese numeral classifiers for units. *Linguistics, 17,* 317–335.

Denny, J. P. (1986). The semantic role of noun classifiers. In C. Craig

(Ed.), *Noun classes and categorization* (pp. 297–308). Philadelphia: John Benjamins.

Downing, P. (1996). *Numeral classifier systems: The case of Japanese.* Philadelphia: John Benjamins.

Gelman, S. A., Collman, P., & Maccoby, E. E. (1986). Inferring properties from categories versus inferring categories from properties: The case of gender. *Child Development, 57,* 396–404.

Gelman, S. A., & Markman, E. M. (1986). Categories and induction in young children. *Cognition, 23,* 183–209.

Gennari, S. P., Sloman, S. A., Malt, B. C., & Fitch, T. (2002). Motion events in language and cognition. *Cognition, 83,* 49–79.

Grinevald, C. (2000). A morphosyntactic typology of classifiers. In G. Senft (Ed.), *Systems of nominal classification* (pp. 50–93). Cambridge, England: Cambridge University Press.

Guo, X. Z. (2002). *Xiandai hanyu liangci yongfa cidian* [Modern Chinese classifier usage dictionary]. Beijing, China: Yuwen chubanshe.

Heit, E., & Rubinstein, J. (1994). Similarity and property effects in inductive reasoning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 411–422.

Hunt, E., & Agnoli, F. (1991). The Whorfian hypothesis: A cognitive psychology perspective. *Psychological Review, 98,* 377–389.

Imai, M., & Gentner, D. (1997). A cross-linguistic study of early word meaning: Universal ontology and linguistic influence. *Cognition, 62,* 169–200.

Imai, M., & Mazuka, R. (2003). Reevaluating linguistic relativity: Language-specific categories and the role of universal ontological knowledge in the construal of individuation. In D. Gentner & S. Goldin-Meadow (Eds.), *Language in mind* (pp. 429–464). Cambridge, MA: MIT Press.

Imai, M., & Mazuka, R. (in press). Revisiting language universals and linguistic relativity: Language-relative construal of individuation constrained by universal ontology. *Cognitive Science.*

Ji, L.-J., Zhang, Z., & Nisbett, R. E. (2004). Is it culture or is it language? Examination of language effects in cross-cultural research on categorization. *Journal of Personality and Social Psychology, 87,* 57–65.

Joordens, S., & Becker, S. (1997). The long and short of semantic priming effects in lexical decision. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 1083–1105.

Lakoff, G. (1987). *Women, fire, and dangerous things: What categories reveal about the mind.* Chicago: University of Chicago Press.

Li, C. N., & Thompson, S. A. (1981). *Mandarin Chinese: A functional reference grammar.* Berkeley: University of California Press.

Lin, E. L., & Murphy, G. L. (2001). Thematic relations in adults' concepts. *Journal of Experimental Psychology: General, 130,* 3–28.

Lucy, J. A. (1992). *Grammatical categories and cognition: A case study of the linguistic relativity hypothesis.* New York: Cambridge University Press.

Malt, B. C., Sloman, S. A., Gennari, S., Shi, M., & Wang, Y. (1999). Knowing versus naming: Similarity and the linguistic categorization of artifacts. *Journal of Memory and Language, 40,* 230–262.

Markman, A. B., & Gentner, D. (1993). Structural alignment during similarity comparisons. *Cognitive Psychology, 25,* 431–467.

Markman, E. M. (1989). *Categorization and naming in children: Problems of induction.* Cambridge, MA: MIT Press.

Markman, E. M., & Hutchinson, J. E. (1984). Children's sensitivity to constraints on word meaning: Taxonomic versus thematic relations. *Cognitive Psychology, 16,* 1–27.

Middleton, E. L., Wisniewski, E. J., Trindel, K. A., & Imai, M. (2004). Separating the chaff from the oats: Evidence for a conceptual distinction between count noun and mass noun aggregates. *Journal of Memory and Language, 50,* 371–394.

Murphy, G. L. (2002). *The big book of concepts.* Cambridge, MA: MIT Press.

Nisbett, R. E. (2003). *The geography of thought: How Asians and Westerners think differently and why.* New York: Free Press.

Nisbett, R. E., Krantz, D. H., Jepson, C., & Kunda, Z. (1983). The use of statistical heuristics in everyday inductive reasoning. *Psychological Review, 90,* 339–363.

Nisbett, R. E., Peng, K., Choi, I., & Norenzayan, A. (2001). Culture and systems of thought: Holistic versus analytic cognition. *Psychological Review, 108,* 291–310.

Osherson, D. N., Smith, E. E., Wilkie, O., López, A., & Shafir, E. (1990). Category-based induction. *Psychological Review, 97,* 185–200.

Saalbach, H., & Imai, M. (2006). Categorization, label extension, and inductive reasoning in Chinese and German preschoolers: Influence of a classifier system and universal cognitive constraints. *Proceedings of the 28th Annual Conference of the Cognitive Science Society* (pp. 703–708). Mahwah, NJ: Erlbaum.

Saalbach, H., & Stern, E. (2004). Differences between Chinese morphosyllabic and German alphabetic readers in the Stroop interference effect. *Psychonomic Bulletin & Review, 11,* 709–715.

Senft, G. (2000). What do we really know about nominal classification systems? In G. Senft (Ed.), *Systems of nominal classification* (pp. 11–50). Cambridge, England: Cambridge University Press.

Slobin, D. I. (1996). From "thought and language" to "thinking for speaking." In J. J. Gumperz & S. C. Levinson (Eds.), *Rethinking linguistic relativity* (pp. 70–96). Cambridge, England: Cambridge University Press.

Smiley, S. S., & Brown, A. L. (1979). Conceptual preference for thematic or taxonomic relations: A nonmonotonic trend from preschool to old age. *Journal of Experimental Child Psychology, 28,* 437–458.

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory, 6,* 174–215.

Tulving, E., & Schacter, D. L. (1990, January 19). Priming and human memory systems. *Science, 247,* 301–306.

Vigliocco, G., Vinson, D., Indefrey, P., Levelt, W., & Hellwig, F. (2004). The interplay between meaning and syntax in language production. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30,* 483–497.

Vigliocco, G., Vinson, D., Paganelli, F., & Dworzynski, K. (2005). Grammatical gender effects on cognition: Implications for language learning and language use. *Journal of Experimental Psychology: General, 134,* 501–520.

Waxman, S., & Gelman, R. (1986). Preschoolers' use of superordinate relations in classification and language. *Cognitive Development, 1,* 139–156.

Waxman, S. R., & Namy, L. L. (1997). Challenging the notion of a thematic preference in young children. *Developmental Psychology, 33,* 555–567.

Wisniewski, E. J., & Bassok, M. (1999). What makes a man similar to a tie? Stimulus compatibility with comparison and integration. *Cognitive Psychology, 39,* 208–238.

Wisniewski, E. J., Lamb, C. A., & Middleton, E. L. (2003). On the conceptual basis for the count and mass noun distinction. *Language and Cognitive Processes, 18,* 583–624.

Yokosawa, K., & Imai, M. (1997, April 28). *Conceptual interference on object verification: Effect of superordinate vs. thematic associative relations.* Paper presented at the meeting of the Association for Research in Vision and Ophthalmology, Fort Lauderdale, FL.

Zhang, S., & Schmitt, B. (1998). Language-dependent classification: The mental representation of classifiers in cognition, memory, and ad evaluations. *Journal of Experimental Psychology: Applied, 4,* 375–385.

Zubin, D. A., & Köpcke, K. M. (1986). Gender and folk taxonomy: The indexical relation between grammatical and lexical categorization. In C. Craig (Ed.), *Noun classification and categorization* (pp. 139–180). Philadelphia: John Benjamins.

Received April 13, 2006
Revision received December 15, 2006
Accepted December 20, 2006 ■