Do classifiers predict differences in cognitive processing? A study of nominal classification in Mandarin Chinese

MAHESH SRINIVASAN*

Harvard University

Abstract

In English, numerals modify nouns directly (two tables), but in Mandarin Chinese, they modify numeral classifiers that are associated with nouns (two flat-thing table). Classifiers define a system of categories based on dimensions such as animacy, shape, and function (Adams and Conklin 1973; Dixon 1986), but do these categories predict differences in cognitive processing? The present study explored possible effects of classifier categories in a speeded task preventing significant deliberation and strategic responding. Participants counted objects in a visual display that were intermixed with distractor objects that had either the same Mandarin classifier or a different one. Classifier categories predicted Mandarin speakers' search performance, as Mandarin speakers showed greater interference from distractors with the same classifier than did Russian or English speakers. This result suggests that classifier categories may affect cognitive processing, and may have the potential to influence how speakers of classifier languages perform cognitive tasks in everyday situations. Two theoretical accounts of the results are discussed.

Keywords
nominal classification, cognitive processing, visual search

* Correspondence address: Mahesh Srinivasan, Department of Psychology, Harvard University, William James Hall, 1118, 33 Kirkland Street, Cambridge, MA 02138, USA. E-mail: mahesh@wjh.harvard.edu. Acknowledgements: I am grateful to Herb Clark and Susan Carey for their support. I thank also Peggy Li, Mary Erbaugh, and Chao Fen Sun for answering questions about Mandarin Chinese, Sasha Yakhkind for assistance in recruiting the Russian-speaking participants, and David Barner, Lera Boroditsky, Caitlin Fausey, Josh Hartshorne, Yi Ting Huang, Manizeh Khan, Teenie Matlock, Rebecca Nappa, Jesse Snedeker, Seenu Srinivasan, Nathan Winkler-Rhoades, Eva Wittenberg, and the editor and reviewers for helpful comments on earlier drafts.

Language and Cognition 2–2 (2010), 177–190 1866–9808/10/0002–0177
DOI 10.1515/LANGCOG.2010.007

AUTHOR'S COPY | AUTORENEXEMPLAR

1. Introduction

Languages vary in how they quantify nouns (Allan 1980; Comrie 1981; Croft 1990). In English, discrete physical objects are typically referred to with count nouns, which can be directly modified by numerals and marked plural (e.g. *two tables*). In contrast, substances are generally referred to with mass nouns, which are modified by introducing a unitizer that takes the plural marking instead of the noun (e.g. *two barrels of oil*). A different situation arises in other languages, such as Mandarin Chinese, where all nouns are modified by *classifiers*, and so no nouns are marked plural[1]. In these languages, every noun is treated much like uncountable substances are treated in English (e.g. *two flat-thing table*).

Classifiers are morphemes or words within the same noun phrase as the noun they qualify (Senft 2000). They have two major roles: they quantify objects and substances, and in doing so, also categorize these things (e.g. as flat things) (Allan 1977). Classifier languages are found throughout the world, and include Mandarin, Japanese, Southeast Asian languages, Austronesian languages, Mayan languages, and others (Aikhenvald 2000). The most common type of classifier is the kind found in Mandarin, in which classifiers form units with numerals or demonstratives. Some of these classifiers signify groups (e.g. a *flock*), containers (e.g. a *glass*), or standard measurements (e.g. a *kilogram*), and are quite flexible with regard to the nouns they can apply to. However, "individual" classifiers apply in a more restricted fashion, often with respect to semantic features—the classifier *tiao2*, for example, generally applies to long and slender objects, such as *wires* or *snakes* (Chao 1968; Gao and Malt 2009). Because each individual classifier can be used with a set of different nouns, these classifiers together define a system of categories, based on features such as animacy (with classes for human, non-human, etc.), shape (with classes for long, flat, etc.), and function (with classes for clothing, transportation, etc.) (Adams and Conklin 1973; Dixon 1986; Chao 1968; Norman 1988; Erbaugh 1986; see Table 1).

However, despite having semantic content, individual classifiers behave differently than other units of language that carry meaning (e.g. adjectives), in that the categories they define are relatively closed and can often seem heterogeneous. For example, although the Mandarin classifier *tiao2* generally classifies long and slender objects, it cannot apply to all objects that fit this description (e.g. *pencils*), while it can apply to objects that do not (e.g. *short pants*). The heterogeneity observed with *tiao2* also characterizes other Mandarin classifier categories: *zhang1*, which classifies flat and sheet-like ob-

1. Mandarin has an optional plural suffix (*-men*) but its use is restricted to nouns for humans and is thus infrequent.

Table 1. *A list of individual classifiers in Mandarin and the objects they typically apply to. The rightmost column is also a list of the objects used in Experiment 1.*

Classifier	Gloss in English	Objects Classified
ba3 (把)	Graspable with a hand	Scissors, Toothbrush, Knife, Fork, Pliers, Shovel, Key, Broom, Hammer, Comb, Spoon, Spatula, Umbrella
tiao2 (條)	Long, slender, bendable	Snake, Towel, Scarf, Wire, Skirt, Pants, Bread, Belt
zhang1 (張)	Flat, sheet-like	Paper, Ticket, Table, Bed, Napkin
gen1 (根)	Slender things	Stick, Cigarette, Match, Candle, Chalk
zhi1 (支)	Branchlike, long and narrow	Pen, Pencil, Chopstick (just one)
ke1 (顆)	Grain-like, small	Button, Pill, Peanut, Pearl
mian4 (面)	Horizontal level surface	Mirror, Drum, Flag
duo3 (朵)	Amorphous	Flower, Cloud

jects, can be applied to paper, tables, and beds, and *mian4*, which classifies objects with level surfaces, can be applied to mirrors, drums, and flags (Chao 1968).

Individual classifiers occupy a unique niche among linguistic devices—while they carry semantic content, they are also restricted in how freely they apply to nouns, and can often seem idiosyncratic and arbitrary. Due to their unique status, it is interesting to consider whether the categories defined by individual classifiers are meaningful for speakers of the language. Denny (1976), for example, has argued that while the categories labeled by nouns reflect the structure and properties of things in the world as they are, classifier categories are important in providing a classification of how humans interact with those things (see also Lakoff 1987). However, the apparent heterogeneity of many classifier categories could instead be taken to suggest that, synchronically, they are just etymological relics, and only serve grammatical functions (e.g. Greenberg 1972).

If the systems of categories defined by classifiers are meaningful for speakers of classifier languages, they might have a bearing on how these speakers conceptually represent and organize objects, and on how speakers perform cognitive tasks beyond when classifiers are explicitly invoked (e.g. as in sentence processing). But cognitive scientists have only recently begun to investigate these possibilities and the evidence has so far been mixed (Schmitt and Zhang 1998; Saalbach and Imai 2007; Gao and Malt 2009). For example, Saalbach and Imai (2007) tested Mandarin and German speakers on a range of tasks including forced-choice categorization and property induction. Objects in these studies could share a classifier, but could also be related taxonomically (e.g. as *robin* and *parrot* are related) or thematically (e.g. as *pencil* and *paper* are related)—and thus be related in ways that are known to be important for the conceptual organization of speakers of non-classifier languages (Markman 1989; Osherson et al. 1990; Lin and Murphy 2001). Of interest was whether Mandarin speakers would be more likely than German speakers to respond and make judgments patterning with classifier membership. However, on most of the tasks, Mandarin speakers were like the German speakers, with both groups responding in line with taxonomic and thematic relations.

Whatever role classifiers may play in conceptual organization, some studies have suggested that systems of nominal classification predict differences in cognitive processing. For example, Schmitt and Zhang (1998) presented Mandarin and English speakers with pairs of objects that either shared a classifier or did not, and found that Mandarin speakers rated pairs of objects sharing a classifier as being more similar than English speakers did (see also Saalbach and Imai 2007, who found a similar effect in their study of Mandarin and German speakers). When asked to describe commonalities between the pairs of objects in these studies, Mandarin speakers also more readily listed common features that were related to the classifier (e.g. "can be grasped" for *ba3*, and "long and slender" for *tiao2*). And finally, when participants were asked to recall a list of items from memory, the Mandarin speakers were more likely to recall items in clusters predicted by classifier categories.

While these results suggest that classifier categories predict differences in cognitive processing, the tasks that were used were not time-pressured, allowing for significant deliberation to take place. Because participants could deliberate about the objects they were shown, they could have responded while conscious of classifier membership, and so may not have limited their responses to inherent properties of the stimuli themselves. Classifier membership could also have been used to strategic advantage, in guiding responses for those tasks without clear answers (as in the similarity judgment task), or for tasks that encouraged a grouping of otherwise unrelated stimuli (as in the recall task) (see Boroditsky 2001; Gleitman and Papafragou 2005 for similar concerns). If possible effects of classifier category only emerge in these kinds of deliberative tasks, the scope of these categories' influence might be quite limited in everyday life. That is, if classifier categories have effects on cognitive processing, they should also predict performance in speeded tasks that prevent deliberation, and for which there is no strategic advantage of invoking classifier membership.

The present study investigates this possibility, using a speeded visual counting task, in which participants count target objects while having to ignore intermixed distractor objects. A counting task can be conceptualized as a visual search task with multiple target objects and ongoing enumeration. Because efficiency of visual search depends heavily on the visual similarity and conceptual relationship between the representations of the target and distractor objects (see Duncan and Humphreys 1989), this type of task provides an appropriate test of the question at hand. At the same time, because the task only involves counting, participants need not deliberate over the properties of the objects to decide on their responses, and as we shall later see, explicitly invoking classifier membership would actually place participants at a strategic disadvantage.

In visual search, the allocation of attention to items in the search array is based on a competition between these items (see Desimone and Duncan 1995). This competition can be influenced by both bottom-up and top-down factors. Top-down influence results from having a template of the target in mind that can prime the representation of a target object to the disadvantage of the other objects in the search array (Duncan and Humphreys 1989). But because attention will also spread to competitors that resemble this template, efficiency will depend on how competitors are related to the template. Competitors that share visual features with the template are more likely to be fixated (Dahan and Tanenhaus 2007), and have been long known to impede search performance (Neisser 1963; Corcoran and Jackson 1977; Treisman and Gormican 1988; Duncan and Humphreys 1989). But visual search is also affected by higher-level factors: competitor objects that are associated with the target (e.g. *bottle* with *glass*), or belong to the same semantic category (e.g. *comb* with *brush*) can attract attention in a comparable way (Moores et al. 2003; Meyer et al. 2007; see also Huettig and Altmann 2005).

Visual search tasks are therefore quite sensitive to the relations that exist between stimuli in the tasks. Thus, if classifier categories predict differences in cognitive processing, distractors that share a classifier category with a target could also be distracting for Mandarin speakers. If this is the case, Mandarin speakers could take longer to count target objects (e.g. *snakes*) pictured among distractor objects that take the same classifier (e.g. *ropes*), compared to distractors that take a different classifier (e.g. *tables*). This result would suggest an effect of classifier category, because it would emerge from a non-deliberative task in which invoking classifiers puts participants at a strategic disadvantage.

Because objects that share a classifier are often more similar than objects that take different classifiers (as in the case of *snakes* and *ropes*), it was necessary to use speakers of a language without classifiers as a comparison group. A group of native Russian speakers and a group of native English speakers each served this purpose. Whereas the Russian and English speakers should take longer to count same classifier pairs than different classifier pairs because of a baseline similarity, Mandarin speakers might experience *additional* interference on the same classifier pairs if classifier categories predict differences in cognitive processing.

2. Method

2.1. *Participants*

The participants were 104 members of a university community, mostly graduate students: 35 in the English group, 36 in the Russian group, and 33 in the Mandarin group. All were between 18 and 35 years old. The English participants were all native, monolingual English speakers, and both the Russian participants and Mandarin participants were native speakers of their language and had learned English only after age 11. All participants were right-handed. Translated instructions in Mandarin and Russian (supplementing the ones that appeared on-screen in English) were available for all participants. Three participants were excluded from the English group, four participants from the Russian group, and one from the Mandarin group, because of poor accuracy or time to complete the task (three standard deviations from the mean). That left 96 participants, 32 in each language group.

2.2. *Materials and procedure*

There were 80 displays of pictures used in the study. On each display, there were target objects (the objects to be counted) and distractor objects. Only one *kind* of target object (e.g. apples) and one kind of distractor object (e.g. oranges) appeared on each display. The target and distractor objects were scattered interchangeably and different images, obtained from the Internet, were used for each instance of an object. There were always between eight and eleven targets (and 8–11 distractors), such that there would be enough items to prevent subitizing (Mandler and Shebo 1982), but few enough to ensure that trials could be completed swiftly and accurately. All images were converted into grayscale to limit objects from standing out due to their color.

Objects were chosen that belonged to one of the following classifier categories: *tiao2*, *duo3*, *zhang1*, *ke1*, *gen1*, *zhi1*, *mian4*, and *ba3* (see Table 1). The objects for each classifier category were decided on after consulting Chao's *Grammar of Spoken Mandarin* (1968) and a classifier dictionary (Chen et al. 1988). The categories for *gen1* and *zhi1* (which classify similar, stick-like objects) were grouped together because it is possible to use either classifier for the objects in their categories, even though one classifier is preferred[2]. Thus, I

2. This was verified by showing pictures of the objects in the *gen1* and *zhi1* categories to four native Mandarin speakers, and asking them for acceptability judgments (1 = not acceptable at all; 10 = very acceptable) on the use of *gen1* and *zhi1* with those objects. While the judgments agreed with my classifications (see Figure 1), they also indicated that the opposite classifications were relatively acceptable (e.g. while "pencil" was more acceptable with *zhi1* (M = 8.0), *gen1* was not unacceptable (M = 4.4), and while "match" was more acceptable with *gen1*

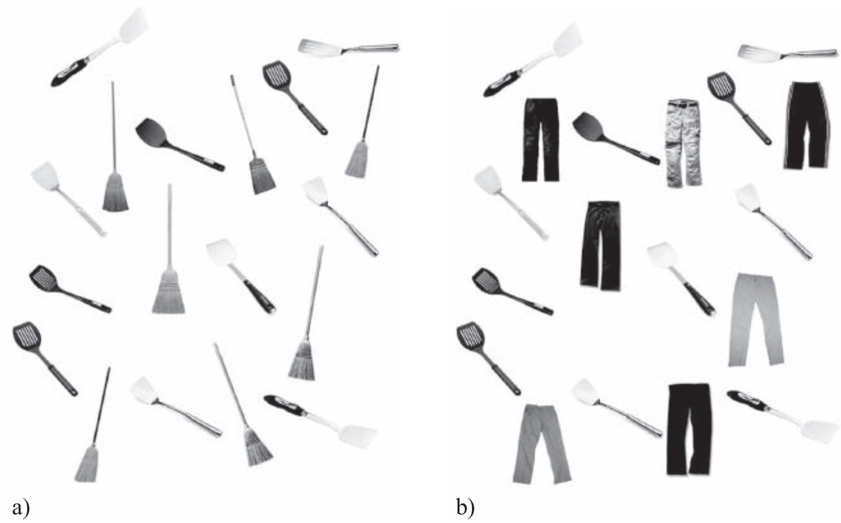

Figure 1. *Examples of items with a spatula as the search target with either a distractor sharing a) the same classifier (ba3, broom), or b) a different classifier (tiao2, pants)*

did not use objects that typically take *gen1* or *zhi1* as different-classifier or same-classifier distractor objects for target objects that typically take *zhi1* or *gen1*, respectively.

The 80 displays were arranged into 40 *sets* of displays. Displays in a set had the same target objects (e.g. spatulas), but different distractor objects. The distractor objects in one display (e.g. brooms, see Figure 1a) had the same classifier as the target objects (*same classifier pair*), whereas the distractor objects in the other (e.g. pants, see Figure 1b) had a different classifier (*different classifier pair*). In both displays of the set, the targets and distractors appeared in identical locations.

The participants completed 17 practice trials and then the 80 test trials. Each test display was preceded by a mid-screen image of the target object (which was not used in the test display) to tell the participants what kind of object to count. It stayed there for two seconds until the test display came up. In the lower right hand corner of the test display was a digit between 8 and 11. The participant had to press a key indicating whether it was the correct number of target objects. When the correct answer was "no", the displayed number was always incorrect by one, to ensure that participants would not be able to

(M = 7.9), *zhi1* was not unacceptable (M = 5.1). In contrast, *ba3* was deemed unacceptable for both of these objects.

immediately reject it as false. PsyScope (Cohen et al. 1993) measured the reaction times for how long a participant took from the presentation of the display to the initiation of the response, and recorded the response.

2.3. *Design and analyses*

Of the 80 displays, 40 were same classifier pairs, and 40 were different classifier pairs. The same classifier and different classifier pairs together formed 40 sets. For every participant, the answer to whether the number displayed was correct was "Yes" for 40 of the displays and "No" for the rest. There were four sub-groups of participants within each language group. In each of these sub-groups, one quarter of the sets were assigned Yes/No (where the correct answer for the same classifier pair was "Yes", and for the different classifier pair was "No"), one quarter Yes/Yes, one quarter No/Yes, and one quarter No/No. The items were presented in a $4 \times 4 \times 4$ Latin square design in which all four sets of displays received all four truth assignments, and each of the four sub-groups of participants received equal numbers of sets for each truth assignment. Participants were randomly assigned to sub-groups. The design was identical for all language groups.

Each participant's reaction time on the same and different classifier trials was taken by averaging over all of the trials in which they did not make errors. Measures were analyzed with F and t tests based on data across participants (F_1, t_1) and items (F_2, t_2) (see Clark 1973).

3. Results

Participants in the Russian group averaged 74 correct responses (36 for same classifier trials, 38 for different classifier trials) and participants in the Mandarin and English groups both averaged 75 correct responses (37 for same classifier trials, 38 for different classifier trials) for the 80 trial study. Figure 2 plots the mean reaction times (for all errorless trials) for the three groups, on trials with same and different classifier pairs. As expected, because objects that share a classifier are more visually similar than objects that take different classifiers, all groups were slower on same classifier trials (A *same-classifier disadvantage*). However, there was a greater difference in reaction time for Mandarin speakers between the same and different classifier trials ($M_{\text{same-different}}$ = 2219, $SE_{\text{same-different}}$ = 204) than for Russian speakers ($M_{\text{same-different}}$ = 1528, $SE_{\text{same-different}}$ = 142) or English speakers ($M_{\text{same-different}}$ = 1334, $SE_{\text{same-different}}$ = 144).

Statistical analyses support these observations. A 3×2 mixed-factor repeated-measures ANOVA examined the effects of the between-subjects factor of language group (Russian, Mandarin, or English), and the within-subjects

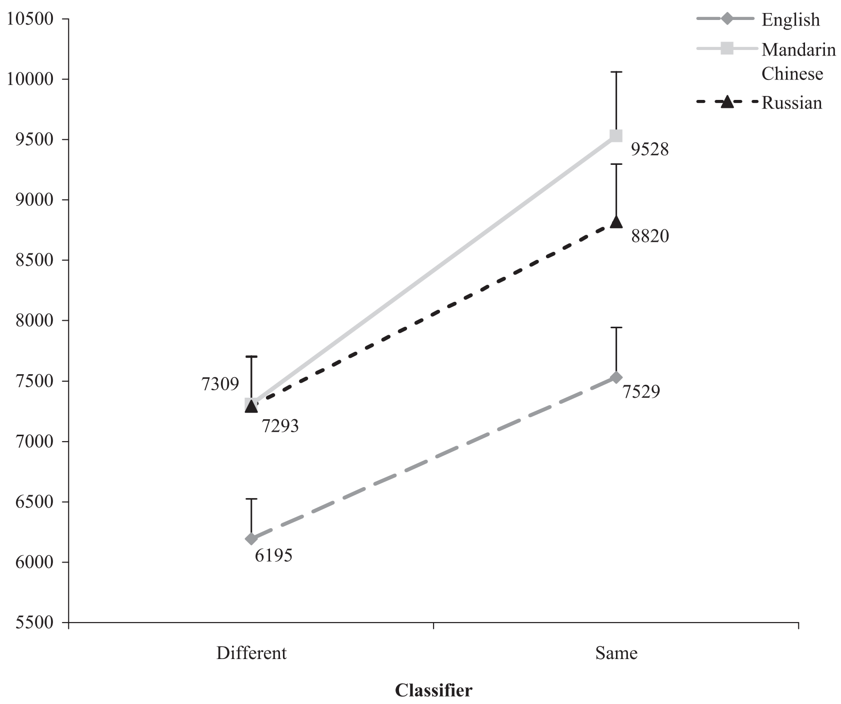

Figure 2. *The influence of classifier on reaction time for the English, Mandarin Chinese, and Russian groups (Error bars indicate +1 standard error).*

factor of classifier membership (same or different) on the participants' reaction times. The main effect of classifier membership (same/different) was strong, $F_1(1,93) = 308.0$, $p < .001$, $\eta^2 = .77$, and $F_2(1,39) = 18.19$, $p < .001$, $\eta^2 = .32$, confirming that speakers of all languages were slower on same classifier trials than on different classifier trials. The interaction of interest (classifier membership × language group) was also significant, $F_1(1,93) = 7.74$, $p < .001$, $\eta^2 = .14$, and $F_2(2,38) = 4.23$, $p < .05$, $\eta^2 = .18$, confirming that Mandarin speakers are subject to a greater same-classifier disadvantage than are Russian or English speakers.

While participants in the Mandarin group were slower than participants in the English group on the same classifier trials, they were also slower on different classifier trials, $t_1(62) = 2.18$, $p < .05$, $d = .54$ and $t_2(39) = 6.80$, $p < .01$, $d = .51$. This baseline difference may have been due to a difference between the groups in familiarity with taking experiments or in familiarity with English. Indeed, it is telling that the Russian group, which more closely matched the Mandarin group on these factors, did not differ from the Mandarin group on the different classifier trials, $t_1(62) = -.03$, $p = .98$ and $t_2(39) = .42$, $p = .67$, but were also slower on these trials than the English group, $t_1(62) = 2.08$, $p < .05$, $d = .52$, and $t_2(39) = 9.55$, $p < .001$, $d = .63$.

A 2 × 2 mixed-factor repeated-measures ANOVA examining only the data from the Russian and Mandarin groups found a significant main effect of classifier membership and interaction between classifier membership and language group (classifier membership: ($F_1(1,56) = 218.97$, $p < .001$, $\eta^2 = .80$, and $F_2(1,36) = 16.76$, $p < .001$, $\eta^2 = .32$); classifier membership × language group: $F_1(1,56) = 7.45$, $p < .001$, $\eta^2 = .12$, and $F_2(1,36) = 6.62$, $p < .05$, $\eta^2 = .16$). But while an ANOVA examining only the English and Russian data found a significant main effect of classifier membership, $F_1(1,56) = 184.34$, $p < .001$, $\eta^2 = .77$, and $F_2(1,36) = 17.78$, $p < .001$, $\eta^2 = .33$, it did not find a significant classifier membership × language group interaction, $F_1(1,56) = .843$, $p = .36$, and $F_2(1,36) = 1.65$, $p = .21$[3].

4. General discussion

Individual classifiers in Mandarin Chinese define a system of categories based on semantic dimensions such as animacy, shape, and function (Adams and Conklin 1973; Dixon 1986; Chao 1968; Norman 1988; Erbaugh 1986). But do these categories predict differences in cognitive processing, beyond when classifiers are explicitly invoked? In the experiment presented here, classifier membership predicted Mandarin speakers' search performance more than it did the search performance of Russian and English speakers, who do not speak a language with classifiers. This suggests that classifier categories may influence search efficiency, such that distractor objects that share a classifier with target objects create interference. These results validate previous evidence that Mandarin speakers' cognitive processing is predicted by classifier category

3. In another analysis, to test the robustness of the results, and because of the presence of some outliers, I trimmed the subject data by 5% on either side. This was done by replacing each of the subject means for the same and different classifier trials with the average of the middle 90% of their values. This reduced variability in the data in all language groups for both same-classifier means (Mandarin: standard error before trimming = 532, after trimming = 509; Russian: before trimming = 478, after trimming = 473, English: before trimming = 416, after trimming = 407) and different-classifier means (Mandarin: standard error before trimming = 391, after trimming = 362; Russian: before trimming = 413, after trimming = 378; English: before trimming = 331, after trimming = 317). The analyses with trimmed means including the data from all language groups closely resembled those with raw means (main effect of classifier: $F_1(1,93) = 322.1$, $p < .001$, $\eta^2 = .78$, classifier × language: $F_1(2,93) = 7.00$, $p < .001$, $\eta^2 = .13$). The results were also similar for the analyses with just the Mandarin and Russian data (classifier: $F_1(1,56) = 235.11$, $p < .001$, $\eta^2 = .81$, classifier × language: $F_1(1,56) = 5.88$, $p < .05$, $\eta^2 = .10$; baseline difference test: $t_1(62) = .029$, $p = .98$), as well as with just the English and Russian data (classifier: $F_1(1,56) = 179.81$, $p < .001$, $\eta^2 = .76$, classifier × language: $F_1(1,56) = 1.32$, $p = .26$; baseline difference test: $t_1(62) = 2.16$, $p < .05$, $d = .54$).

(e.g. Schmitt and Zhang 1998), but do so with a speeded task in which explicitly invoking classifier membership would lead to a strategic disadvantage. The observed effects are therefore unlikely to result from overt deliberation or strategic responding, and are consistent with the interpretation that classifier categories have automatic effects on cognitive processing.

A potential concern with this task is that Mandarin speakers might have used both number words *and* classifiers while counting (e.g. *1-zhang1, 2-zhang1, 3-zhang1*), in which case distractors with the same classifier might have created confusion because of an overt use of the classifier in the count list. However, this is not a concern because Mandarin speakers do not use the classifier form while counting: classifier forms are used only in completed statements of quantification (as in *three flat-thing table*). Indeed, participants who were debriefed confirmed that they did not use the classifier form while counting. I also asked five more native Mandarin speakers to verbally enumerate the target objects for the first 10 trials of the experiment, and observed no use of classifier forms.

Because the present experiment used a quasi-experimental design—participants were not randomly assigned to speak a language with or without classifier categories—it does not demonstrate a causal effect of classifier category on search performance. It remains possible that other factors, such as cultural differences, can explain both the linguistic and behavioral differences among the participant groups tested. For example, perhaps brooms and spatulas in Chinese (and Chinese-American) households look more alike than do brooms and spatulas in other cultures, resulting in both a common linguistic classification and a heightened conceptual similarity for Mandarin speakers[4]. While this interpretation is unlikely (e.g. the Russian-speaking participants also had cultural differences from the English-speaking participants, but the two groups performed similarly on the task with respect to classifier membership), future studies are required to demonstrate a causal effect of classifier category. In what follows, I outline the different mechanisms through which classifier categories could have had causal effects. I distinguish between two theoretical possibilities—one that attributes a possible linguistic locus to the effects, in modulating cognitive processing, and one a possible conceptual locus, in shaping conceptual representations.

On the first view, search efficiency may have been directly modulated by representations of classifier categories. For example, in seeing a picture of the target object, participants may have accessed the lexical representation corresponding to that object (see, e.g. Meyer et al. 2007, who report evidence for a rapid activation of lexical representations during visual search, but see

4. I thank the editor and reviewers for drawing my attention to this point.

also Zelinsky and Murphy 2000). If, for Mandarin speakers, lexical representations are organized into classes according to classifier category, activation could then spread from the lexical representation of the target to representations corresponding to other objects from the same classifier category, including those of the distractor objects in the same-classifier trials. This could ultimately activate representations of the distractor objects in the search array and make those objects stronger competitors for attention, impeding search efficiency. This would constitute some of the first evidence that grammatical categories can affect the allocation of attention in a task incorporating visual search.

On a second view, classifier categories may have only indirectly produced these results—instead, they may have had their effects well before the task in affecting conceptual representation itself. For example, in learning to use classifiers as children, and in using them in language throughout their lives, Mandarin speakers may have come to conceptually associate objects from the same category with one another, or to represent these objects as sharing features to a greater extent. Because visual similarity and associative relation are known to affect search efficiency (Moores et al. 2003; Meyer et al. 2007), Mandarin speakers would then be expected to experience more interference from same-classifier distractors.

These competing explanations raise a question of to what extent classifier categories might shape how Mandarin speakers think, when they are not using language. For example, the second account would predict that effects of classifier category would persist in a non-linguistic task. But while the visual counting task used here places minimal linguistic demands—participants use their count-list and might activate the name of the target object—the task is not entirely non-linguistic, and cannot decisively bear on whether classifier categories shape non-linguistic thought. What is clear is that the results cannot be attributed to an explicit use of the classifier, because participants did not use the classifier form while counting[5]. Thus, if linguistic representations of classifier categories mediated the results even under these conditions—which would put participants at a strategic disadvantage—this would indicate that these categories are quite robust, having the potential to influence how speakers of classifier languages perform a variety of cognitive tasks in everyday situations. Indeed, the scope of linguistically mediated effects on cognitive processing could be wide-ranging in light of evidence that people automatically name objects that they view (Meyer et al. 2007).

5. Even if classifiers are not explicitly used while counting, a context in which participants enumerate objects might be more likely to elicit effects of classifier category. This underscores the importance of using other types of speeded tasks in future studies that do not require enumeration (e.g. visual search). I thank the reviewers for making this point.

As a final note, it is interesting to consider how these results might generalize to other classifier categories, and to the classifier categories of other languages. Classifier categories differ in their structure and coherence, ranging from ones in which all objects within a category share features, to ones that are more arbitrary (see Gao and Malt 2009). The items included in this experiment tended to belong to classifier categories that are quite well-defined. Thus, future research should consider whether these results would generalize to more heterogeneous categories, as well as to the classifier categories of other languages.

References

Adams, K. & N. F. Conklin. 1973. Toward a theory of natural classification. *Papers from the 9th regional meeting of the Chicago Linguistics Society.* 1–10.

Aikhenvald, A. I. U. 2000. *Classifiers: A typology of noun categorization devices*. Oxford & New York: Oxford University Press.

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

Allan, K. 1980. Nouns and countability. *Language* 56(3). 541–567.

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

Chao, Y. 1968. *A grammar of spoken Mandarin*. Berkeley: University of California Press.

Chen, B., G. Che, H. Chen & Z. Zhang. 1988. *Hanyu liangci cidian* [A dictionary of Chinese numeral classifiers]. Fuzhou, Fujian Province: Fujian People's Publishing House.

Clark, H. H. 1973. The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior* 12. 335–359.

Cohen, J. D., B. MacWhinney, M. Flatt & J. Provost. 1993. PsyScope: A new graphic interactive environment for designing psychology experiments. *Behavioral Research Methods, Instruments, and Computers* 25(2). 257–271.

Comrie, B. 1981. *Language universals and linguistic typology: Syntax and morphology*. Chicago: University of Chicago Press.

Corcoran, D. W. J. & A. Jackson. 1977. Basic processes and strategies in visual search. In Dornic, S. (ed.), *Attention and performance* VI, 387–411. Hillsdale, NJ: Erlbaum.

Croft, W. A. 1990. *Typology and universals*. Cambridge: Cambridge University Press.

Dahan, D. & M. K. Tanenhaus. 2005. Looking at the rope when looking for the snake: Conceptually mediated eye movements during spoken-word recognition. *Psychonomic Bulletin and Review* 12. 453–459.

Denny, J. P. 1976. What are noun classifiers good for? *Papers from the twelfth regional meeting of the Chicago Linguistics Society*. 122–132.

Desimone, R. & J. Duncan. 1995. Neural mechanisms of selective visual attention. *Annual Review of Neuroscience* 18. 193–222.

Dixon, R. M. W. 1986. Noun classes and noun classification in typological perspective. In Craig, C. G. (ed.), *Noun classes and categorization*, 105–112. Amsterdam & Philadelphia: J. Benjamins.

Duncan, J. & G. W. Humphreys. 1989. Visual search and stimulus similarity. *Psychological Review* 96. 433–458.

Erbaugh, M. 1986. Taking stock: The development of Mandarin noun classifiers historically and in young children. In Craig, C. G. (ed.), *Noun classes and categorization*, 399–438. Amsterdam & Philadelphia: J. Benjamins.

Gao, M. Y. & B. C. Malt. 2009. Mental representation and cognitive consequences of Chinese individual classifiers. *Language and Cognitive Processes* 24(7). 1124–1179.

Gleitman, L. & A. Papafragou. 2005. Language and thought. In Holyoak, K. & R. Morrison (eds.), *Cambridge handbook of thinking and reasoning*, 633–661. Cambridge: Cambridge University Press.

Greenberg, J. H. 1972. Numeral classifiers and substantival number: Problems in the genesis of a linguistic type. *Stanford University Working Papers on Language Universals* 9. 1–39.

Huettig, F. & G. T. M. Altmann. 2005. Word meaning and the control of eye fixation: semantic competitor effects and the visual world paradigm. *Cognition* 96(1). B23–B32.

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

Lin, E. L. & G. L. Murphy. 2001. Thematic relations in adults' concepts. *Journal of Experimental Psychology: General* 130(1). 3–28.

Mandler, G. & B. J. Shebo. 1982. Subitizing: an analysis of its component processes. *Journal of Experimental Psychology*: *General* 111(1). 1–22.

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

Meyer, A. S., E. Belke, A. L. Telling & G. W. Humphreys. 2007. Early activation of object names in visual search. *Psychonomic Bulletin and Review* 14(4). 710–716.

Moores, E., L. Laiti & L. Chelazzi. 2003. Associative knowledge controls deployment of visual selective attention. *Nature Neuroscience* 6. 182–189.

Neisser, U. 1963. Decision-time without reaction-time. Experiments in visual scanning. *The American Journal of Psychology* 76(3). 376–385.

Norman, J. 1988. *Mandarin*. Cambridge: Cambridge University Press.

Osherson, D. N., E. E. Smith, O. Wilkie, A. López & E. Shafir. 1990. Category-based induction. *Psychological Review* 97(2). 185–200.

Saalbach, H. & M. Imai. 2007. The scope of linguistic influence: Does a classifier system alter object concepts? *Journal of Experimental Psychology: General* 136(3). 485–501.

Schmitt, B. H. & S. Zhang. 1998. Language structure and categorization: A study of classifiers in consumer cognition, judgment, and choice. *Journal of Consumer Research* 25(2). 108–122.

Senft, G. 2000. *Systems of nominal classification*. Cambridge & New York: Cambridge University Press.

Treisman, A. & S. Gormican. 1988. Feature analysis in early vision: Evidence from search asymmetries. *Psychological Review* 95. 15–48.

Zelinsky, G. J. & G. L. Murphy. 2000. Synchronizing visual and language processing: An effect of object name length on eye movements. *Psychological Science* 11. 125–131.