The Effect of Language-specific Characteristics on English and Japanese Speakers' Ability to Recall Number Information

Abstract

The current paper presents two experiments investigating the effect of presence versus absence of compulsory number marking in a native language on a speaker's ability to recall number information from photos. In Experiment 1, monolingual English and Japanese adults were shown a sequence of 110 photos after which they were asked questions about the photos. We found that the English participants showed a significantly higher accuracy rate for questions testing recall for number information when the correct answer was '2' (instead of '1') than Japanese participants. In experiment 2, English and Japanese adults engaged in the same task as in Experiment 1 with an addition that explored reasons for the results found in Experiment 1. The results of Experiment 2 were in line with the results of Experiment 1, but also suggested that the results could not be attributed to differences in guessing patterns between the two groups or the type of linguistic constructions used in the test situations. The current study suggests that native language affects speakers' ability to recall number information from scenes and thus provides evidence for the *Whorfian hypothesis.*

Keywords: Whorfian hypothesis; Linguistic Relativity; number marking; visual memory; plurality; recall; Japanese; English

The Effect of Language-specific Characteristics on English and Japanese Speakers' Ability to Recall Number Information

The question as to whether native language does (e.g., Lucy, 2016; Slobin, 1996, 2003; Wolff & Holmes, 2011; Whorf, 1956; Zlatev & Blomberg, 2015) or does not (e.g., Berlin & Kay, 1969; Bloom & Kiel, 2001; Pinker, 1994; 2007; Fodor, 1975) structure speakers' cognition has been debated in the cognitive sciences for decades. The *Whorfian hypothesis* is commonly divided into the largely abandoned strong form, *linguistic determinism*, which assumes that knowledge of a linguistic element allows conceptualizations that would not otherwise be possible. The weak form, *linguistic relativity,* (e.g., Gumperz & Levinson, 1996; Lucy, 1992, 2016) holds that language does not determine, but influences, speakers' thought. For example, due to differences in grammatical conventions, a Japanese speaker might conceptualize or categorize the world differently and might engage in different cognitive processes when interpreting, perceiving and recalling the world than, say, an English speaker.

Research conducted in the past 25 years has demonstrated that language can have an effect on non-linguistic cognitive processes. For example, the availability (i.e., whether or not a language has a particular linguistic construction) and frequency of grammatical constructions, lexical items and metaphors in a speaker's language(s) have an effect on their color cognition (e.g., Roberson, Davies & Davidoff, 2000; Paramei, Griber & Mylonas, 2018), the conceptualizations of time (e.g., Boroditsky, 2001; Bylund & Athanasopoulos, 2017; Gu, Zheng & Swerts, 2019), assignment of gender (e.g., Sato & Athanasopoulos, 2018; Sato, Casaponsa & Athanasopoulos, in press; Sedlmeier, Tipandjan & Jänchen, 2016; Speed & Majid, 2019; Vigliocco, Vinson, Paganelli & Dworzynski, 2005), space (e.g., Levinson, 1996; Majid, Bowerman, Kita, Haun & Levinson, 2004) and number cognition (e.g., Athanasiadou & Athanasopoulos, 2017; Athanasopoulos, 2006; Everett & Madora, 2012; Frank, Everett, Fedorenko & Gibson, 2008; Gordon, 2004; Lucy and Gaskins, 2003; Pica, Lemer, Izard & Dehaene, 2004). The current paper contributes to this body of research by focusing on a previously unexplored link between language and cognition and presents two experiments testing whether cross-linguistic differences in the presence (English) or absence (Japanese) of compulsory number marking result in differences in adult speakers' recall of number information from scenes seen – an area of thought that has been previously shown to be at least to some extent resistant to Whorfian effects, as we explain below.

Whorfian effects on number cognition

Number cognition research can be divided into two broad areas. First, several studies have investigated number cognition in speakers of languages that have limited ways to refer to number (e.g., Everett & Madora, 2012; Frank et al., 2008; Gelman & Gallistel, 2004; Gordon, 2004; Pica et al., 2004; Spaepen, Coppola, Spelke, Carey, & Goldin-Meadow, 2011) as well as in animals and pre-linguistic infants (e.g., Brannon & Roitman, 2003; Feigenson & Carey, 2003, 2005; Gallistel & Gelman, 2000; Gelman & Cordes, 2001; Wang & Feigenson, 2019) to see if number cognition would exhibit universal or language-specific influences. Gordon (2004), Frank et al. (2008) and Everett and Madora (2012) conducted several experiments on adult Pirahã-speakers, whose native language lacks morphological number marking and systematic number terms beyond 'one' and 'two'. In these experiments Pirahã-speakers were asked (a) to choose the same number of items (e.g., *nuts* or *AA batteries*) that matched the number of items that had been presented to them; in one condition the items presented were left visible and in another they were hidden after presentation (Everett & Madora, 2012; Frank et al., 2008; Gordon, 2004), and (b) to recall if there were still items (e.g., *nuts*) left in an opaque container when the experimenter had shown the participant the nuts placed in the container and then took the nuts out one at a time (Frank et al., 2008; Gordon, 2004). In all three studies Pirahã-speakers were at ceiling in the tasks involving small (1-3) numbers of items. However, large (4 -10) numbers of items provided different results in the matching tasks where the items were visible during the task – Gordon (2004) and Everett and Madora (2012) found that matching >3 objects was difficult even when visible to the participant, while Frank et al. (2008) found no significant difference in the performance between small and large number of items, a discrepancy which might be related to the Frank et al.'s participants' having had exposure to languages with number terms (Everett & Madora, 2012). In the hidden tasks in which the participant had to memorise a larger number of items than just 1-3, the response accuracy significantly dropped in all three studies. Studies with pre-linguistic infants echo these findings and suggest that while matching and recalling a small number of items is possible without knowledge of number terms, tasks with greater than three entities is difficult without linguistic labels (e.g., Feigenson & Carey, 2003, 2005; Wang & Feigenson, 2019). These studies suggest that even though visual attention and conceptualization of number information without linguistic expression is possible to some extent, memory is an area of cognition that can be affected by cross-linguistic differences in number marking, and when participants are tested with a small number of objects (<4), there seems to be little evidence of Whorfian effects.

The second major strand of *linguistic relativity* research on number cognition has investigated the effects of cross-linguistic differences in noun quantification to reveal whether or not speakers from different linguistic backgrounds differ in the way that they conceptualize entities as substances (e.g., *water*, *cream*, *mud*) or discrete objects (e.g., *a dog*, *a comb*, *a lemon squeezer*) (e.g., Athanasopoulos & Kasai, 2011; Cook, Bassetti, Kasai, Sasaki & Takahashi, 2006; Imai & Mazuka, 2003; Iwasaki, Vinson & Vigliocco, 2010; Lucy & Gaskins, 2003; Masuda, Ishii, Miwa, Rashid, Lee & Mahdi, 2017; Mazuka & Friedman, 2000). Non-classifier languages, such as English, refer to the number of count nouns with numerals (e.g., *two apples*) and to mass nouns with quantifiers (e.g., *some sugar*, *a lot of butter*) and have inflectional number marking on count nouns (e.g., *an apple* vs. *two apple-**s***).[1] Classifier languages, such as Yucatec Maya, Japanese and Chinese do not have systematic inflectional number marking on nouns, nor can nouns be modified by numerals in the same way as count nouns in English. Instead, the number information is expressed by the use of a classifier determined, for example, by animacy, shape, size, function and social importance of the entity being quantified (Croft, 1994) (e.g., in Japanese: *san biki no inu* 'three small animal of dog', English: 'three dogs').

Research shows that the way languages refer to number information (classifier vs. non-classifier) can have an effect on the speakers' conceptualizations. First, in similarity judgment tasks, speakers of classifier languages perceive nouns from the same classifier category as more similar than speakers of non-classifier languages, suggesting that classifier categories have an effect on speakers' concepts (e.g., Gao & Malt, 2009; Huettig, Chen, Bowerman & Majid, 2010; Lucy, 1992; Lucy & Gaskins, 2003; Saalbach & Imai, 2007, 2012). Second, tasks in which participants have to match an object with one of two other objects, one of which is a shape match with the target object and the other is a substance match, show that while English speakers are more likely to choose the item whose shape matches the target object than Japanese and Yucatec speakers, there is not such a distinct difference in mass nouns (e.g., Athanasopoulos & Kasai, 2011; Cook et al., 2006; Imai & Mazuka, 2003; Lucy & Gaskins, 2003). However, these types of cross-linguistic effects between classifier and non-classifier languages have not been found in all tasks. Tasks tapping into automatic cognitive processes such as fast-speech picture matching in which the participants have to

[1] Note that some non-classifier languages (e.g., English) do not distinguish between different types of plural information (i.e., nouns referring to 2 and >2 entities are marked with the same plural morpheme), while others (e.g., some dialects of Slovenian, Marušic, Žaucer, Plesničar, Razboršek, Sullivan, & Barner, 2016) contrast nouns for singular, dual (2) vs. plural (>2) entities.

match a word to a picture (out of two pictures) do not find differences between classifier and non-classifier languages (Saalbach & Imai, 2007, 2012). Furthermore, eye-tracking (Chen, Bowerman, Huettig, & Majid, 2010; Huettig, et al., 2010), recall (Gao & Malt, 2009) and picture/word matching (Speed, Chen, Huettig & Majid, in press) studies suggest that the classifier categories affect speakers' responses only if the classifier is overtly present in the stimuli. Thus, the way languages refer to number information has an effect on how count and mass nouns are categorized, but only in some contexts and only if the classifier is present.

As the use of classifiers in languages like Japanese can be seen as being similar to English speakers referring to quantities of substances (e.g., *two cups of coffee*), some have suggested that classifier language speakers treat (inanimate) count nouns as mass nouns (Athanasopoulos, 2006; Lucy, 1992; Whorf, 1956). However, several studies have failed to find evidence that classifier language speakers would perceive inanimate countable objects as mass items (e.g., Cheng & Sybesma, 1998; Cook et al. 2006; Imai & Gentner, 1997; Imai & Mazuka, 2003; Iwasaki, et al., 2010; Mazuka & Friedman, 2000). First, in similarity judgment tasks, both classifier (e.g., Japanese) and non-classifier (e.g., English) language speakers are more likely to match objects according to their shape than their substance. This tendency is stronger in non-classifier language than in classifier language speakers, in particular with simple objects such as a pyramid (Cook et al., 2006; Imai & Mazuka, 2003). Thus, the use of classifiers to express number information does not seem to result in Japanese speakers conceptualizing discrete objects as mass, but they are able to distinguish objects from substances. Second, Imai & Gentner's (1997) similarity judgment study with simple and complex objects and substances found that both American and Japanese adults and children (2-year olds; 2 ½ to nearly 3-year olds; 4-year olds) made a distinction between (complex) objects and substances, suggesting that the use of classifiers in Japanese does not result in Japanese speakers perceiving discrete objects as mass items. Third, it has been suggested that the use of different classifiers for objects and substances in classifier languages results in different syntactic processing for objects and mass nouns (Cheng & Sybesma, 1998; Imai & Gentner, 1997; Iwasaki, et al., 2010; Mazuka & Friedman, 2000). For example, Iwasaki et al.'s (2010) experiment 1 compared slips of the tongue errors in picture naming of food items (e.g., saying 'carrot' when presented with a picture of a cucumber) between English and Japanese adults and found the Japanese participants errors reflected English mass/count distinctions. They suggest that both speaker groups have mass and count categories regardless of the use of classifiers in Japanese and morphological markings and quantifiers in English.

To recap, previous research suggests that (a) lack of number labels (e.g., Pirahã) might result in speakers' failing to recall number information beyond small numbers (≥ 3), and (b) classifier language (e.g., Japanese) speakers, who do have grammatical number markers but mark number information similarly to mass nouns in non-classifier languages (e.g., English) might have different categorizations of objects than non-classifier language speakers, but nevertheless do perceive discrete entities as count rather than mass items (in particular if the classifier is not presented with the noun).

The current study extends the previous research conducted on number cognition by focusing on the presence versus absence of compulsory number marking on discrete entities (i.e., count nouns) and tests if cross-linguistic differences in the necessity to express number information have an effect on speakers' recall of singular versus plural entities.

Number marking in English and Japanese

Japanese has an elaborate counting system that precedes nouns. It has between 150 (Downing, 1996, p. 17) and 360 numerical classifiers (Iida cited in Tojyo, 2014), for example, *san biki no inu* ('three small animal of dog' Eng: 'three dogs'). However, more importantly, nouns themselves rarely take grammatical number markers and can be used without indication of number, even when referred to plural entities.[2] Thus, for example, the sentence *Inu ga aruite-imasu* is ambiguous as to whether it refers to *A dog is walking* or *Dogs are walking*. This feature of Japanese allows Japanese speakers to refer to scenes without overtly expressing whether there is one or more than one entity in the scene. On the other hand, whenever English speakers are talking about regular count nouns, they need to express whether they are referring to one or more than one entity by the use of grammatical marking (presence/absence of the bound morpheme -(e/s) and additionally by using articles (a/an), quantifiers (e.g., *some*, *several*) or numerals (1, 2, 3…). Thus, for example, *A dog is walking* versus *Dogs are walking* have different meaning, and using the former sentence to describe the latter scene would be considered ungrammatical or false. This linguistic feature might force attention to and retention of number information in English speakers, similarly for

[2] Note that in Japanese one can optionally express plural marking (e.g., *komodo-tachi* 'child-PLU') but these markings are usually only used with human referents (although, *-tachi* can be used when referring to pets that are perceived as part of one's family). In the current study no human referents were used – only animals and inanimate entities. That is, the materials were such that the items have no grammatical plural marking – and importantly for the current study, even if an item could be used with plural marking, the use of it would be optional.

instance to the way that absolute orientation forces speakers to monitor deixis (e.g., Levinson, 1996). More specifically, in line with the Whorfian hypothesis, the linguistic difference between English and Japanese, whereby one language necessitates reference to number (English) and the other one does not (Japanese), raises a possibility that speakers whose language routinely forces them to encode information about number linguistically might show a higher sensitivity to number information, e.g., by better visual memory for the number of entities seen, than speakers whose language does not have this linguistic feature. Furthermore, given that in Japanese the meaning of unmarked noun forms is more closely associated with singularity (Asami, 2006), the strongest difference in sensitivity is likely to surface with plural items (rather than singular items).

The use of number words

Even though number marking is optional in Japanese, it does not appear to be the case that reference to number information by using number words would be absent from Japanese. Sarnecka, Kamenskaya, Yamana, Ogura, and Yodovina's (2007) CHILDES corpus analysis reports that cardinal uses (excludes e.g., counting) of 'three' was equally frequent in the English and Japanese corpora (English: 426; Japanese: 435 per 1 million words) and while the use of 'two' was more frequent in English than in Japanese, it was not absent from the Japanese data either (English 1925; Japanese 796 per 1 million words). If we assume that the Child (Direct) speech analyzed by Sarnecka et al. (2007) can be applied to adult language use, both English and Japanese speakers are likely to have similar number concepts overall and no clear cultural difference in referral to number by using numerals.

Thus, number information is not culturally unimportant or systematically unexpressed by Japanese speakers and Japanese would not be considered a language that has an impoverished number system or unsystematic use of number words (cf. Pirahã). Instead, both Japanese and English speakers have the ability to talk about number, and frequently do so but while the English speakers must convey number information every time they produce a regular count noun by the provision/omission of a bound morpheme, and by use of the article and/or the optional numeral/quantifier, Japanese speakers do not have similar grammatical ways to express number information, and more importantly, have the option of omitting any reference to number even when referring to >1 entities.

Previous research on the effect of language on recall

Previous research suggests that structural/labelling differences between languages can result in differences in recall for example on source evidence (Tosun et al., 2013, but see Ünal, Pinto, Bunger, & Papafragou, 2016), agency (e.g., Fausey & Boroditsky, 2011), resultative events (Sakarias & Flecken, 2019), figure-ground information (e.g., Nisbett & Masuda, 2001; Tajima & Duffiel, 2012) and colour (e.g., Davidoff, Davies, & Roberson, 1999; Roberson, Davidoff, Davies, & Shapiro, 2005; but see Cibelli, Xu, Austerweil, Griffiths, & Regier 2016; Regier and Xu, 2017; Rosch Heider & Olivier, 1972). Tajima and Duffield (2012) investigated whether the linguistic difference in the conventional way to express figure-ground information by English and Chinese versus Japanese speakers results in difference in recall of figure or ground information. They found that the Japanese speakers whose language puts more syntactic focus on the ground information told stories and described pictures with more reports of the ground information than the English and Chinese speakers, whose language syntactically focuses on the figure. Furthermore, when shown small clips of the ground information from the pictures the participants had previously seen and asked if the clips were from the previously seen photos, the Japanese speakers showed a superior ability to recall this information than the English and Chinese speakers. Tosun, Vaid and Geraci (2013) investigated the effect of obligatory syntactic source marking (Turkish) versus the lack of it (English) on source memory. They found that the presence of obligatory syntactic marking indicating whether information was hearsay or first-hand knowledge in Turkish and the lack of such marking in English resulted in differences in the recall of source information in the two groups predicted by the linguistic differences. Fausey and Boroditsky (2011) report similar effects for causality, whereby the conventional way in a speaker's language to refer to deliberate versus accidental actions had an effect on visual recall. Namely, speakers of English – a language that conventionally uses agentive sentence structures (e.g., *He broke the vase*) in deliberate and accidental contexts were more likely to remember the agent of accidental scenes than Spanish-speakers whose language conventionally uses agentive structures in deliberate actions but in accidental contexts a structure that de-emphasizes agency (e.g., *Se rompió el florero* 'The vase broke itself'). Sakarias and Flecken's (2019) study investigated Estonian and Dutch speakers and found that the presence of a morphological distinction (partitive vs. accusative) in Estonian between resultative (e.g., *peeling a potato*) and non-resultative (e.g., *stirring a pan*) events, respectively, and the absence of such distinction in Dutch affected the recall of the endings of causative events. Namely, the Estonian speakers were more accurate in recalling if the action they had seen in a video clip had concluded, compared to Dutch speakers in the condition where the participants described the videos with one sentence immediately after having seen them. Finally, Davidoff et al. (1999) and Roberson et al. (2005) investigated Dani, Berinmo, Himba and English speakers – languages in which the colour spectrum is divided differently linguistically. They found that the different speaker groups' ability to accurately select the same colour they had seen 30 seconds earlier reflected the colour naming systems of their native language.

The aforementioned studies suggest that cross-linguistic differences can influence speakers' visual memory recall, but as far as we are aware, the question as to whether the presence or absence of obligatory number marking on count nouns affects recall in speakers that have systematic number labels is yet to be investigated. Given that both of our speaker groups have number labels and use them, the current study informs us as to whether systematic expression of plural marking boosts recall of plurality information in scenes seen.

The present study

To further our understanding as to whether cross-linguistic differences in compulsory grammatical markings can impact the recall of visual information, we conducted two experiments investigating Japanese and English-speaking adults' ability to retain number information from photos seen. Even though a large number of studies have investigated the effect of native language on number cognition, as far as we are aware, this is the first study investigating if the presence versus absence of compulsory singularity/plurality marking affects visual memory for number information in a language that has number terms (cf. Pirahã).

Given that Japanese lacks obligatory number marking thus allowing Japanese speakers to talk about nouns without explicitly giving number information, while English necessitates its speakers to give singularity and plurality information (1 vs. >1, e.g., *an apple vs. apple***s**) whenever they refer to regular count nouns, based on the Whorfian hypothesis, we hypothesized that English and Japanese speakers would differ in their cognitive processes relating to numeric information. More specifically, English speakers' habitual and systematic marking of number information should result in them recalling having seen one or more than one entity equally well, while Japanese should show a lower recall accuracy rate when having seen more than one entity.

In the present study, we compare recall of 1 (singular) versus 2 (plural) items for the following reasons. First, comparing 1 versus 2 presents the smallest difference between singularity and plurality in English and Japanese as neither English nor Japanese have dual marking (i.e., they do not make a grammatical distinction between 1 vs. 2 vs. > 2) yet this minimal difference should be salient to English speakers due to morphological distinction between 1 versus ≥ 2 items. Second, small numbers ≤ 3 have been found to be resistant to Whorfian effects (Gordon; 2004; Feigenson & Carey, 2003, 2005; Frank et al., 2008; Wang & Feigenson, 2019). Thus, comparing small numbers (1 vs. 2) can inform us about the presence/absence of universals in number cognition.

Experiment 1

Method

An experiment testing visual memory recall was conducted. This consisted of two stages: (1) photo presentation and (2) memory test of the items depicted in the photo presentation.

Participants

Thirty-three monolingual Japanese-speakers (19 male, 14 female) and 30 monolingual English-speakers (13 male, 17 female) aged between 18 - 34 years (English: M_{age}= 21.8 years, SD_{age} = 4.6; Japanese M_{age}= 20.4, SD_{age}= 1.7) took part in the study. One additional participant (an English male) was tested, but excluded for not focusing on the task. The sample sizes were chosen to broadly reflect sample sizes in a number of previous studies testing the effect of cross-linguistic grammatical features on the recall of visual information (e.g., Sakarias & Flecken, 2019; Tajima & Duffield, 2012; Tosun et al., 2013). The participants were recruited from their universities, workplace or through friends and were individually tested either in a quiet room at their university, workplace or private home. All participants reported being monolingual and having only very limited knowledge of a foreign language. However, due to English being a foreign language that all Japanese students study at schools from junior high school to university, we only included Japanese participants who reported that their standardized English language scores were low (TOEIC <600, TOEFL <500/35, EIKEN <2), or had not taken these tests (indicating low ability), that they had not been outside of Japan for longer than one month, and that they had no foreign friends or colleagues. None of the participants had known cognitive disabilities. If the participant had corrected to normal vision, they were instructed to wear their glasses during the test.

Materials

Items. Forty items in testing recall of number information were created. To ensure that the participants were familiar with the animals or objects depicted in the photos, apart from two items (*crocodile* and *cushion*), the items were chosen from the Japanese (Ogura & Watamaki, 2004) and British English versions (Lincoln University Babylab, 2001) of the MacArthur-Bates Communicative Development Inventory (CDI) (Fenson, Dale, Reznick, Thal, Bates, Hartung, Pethick & Reilly, 1993), toddler version.[3] Given that Japanese and English children are likely to be familiar with these words and therefore also with these entities, we assumed that our native monolingual adult speakers would also be familiar with them. The test items were largely inanimate entities (*n=31*) such as vehicles, household objects and food items. In addition, nine animate entities (animals) were included. Any item that could have been perceived as the same item (e.g., *chair* and *stool*) were excluded. All items were count nouns in English and perceived as discrete entities by the second author of this paper, a native Japanese speaker.

We selected items that were natural in singular and plural forms and included only items that were not in any obvious way strongly associated with singularity (e.g., *the sun*) or duality (e.g., *socks*). For some items, the singular and dual forms may not have been the most typical association (e.g., *bee, biscuits*) as these might be associated with larger numbers of entities (e.g., *a swarm of bees, a plateful of biscuits*). However, these possible stronger > 2 associations (a) would not directly prime either of our response options (1 vs. 2), and (b) can be expected to be similar for English and Japanese speakers, thus unlikely to affect our results.

Thirty control items testing memory of other than number detail (6 x color, 6 x shape, 6 x material, 6 x location, 6 x action) were chosen. Eleven of these were animate entities (although one item, '*turtle*', was depicted as an inanimate ornament), 19 were inanimate. [4]

40 filler items were included to distract the participants from noticing that many of the photos depicted one or two animals/objects and to avoid primacy or recency effects. The fillers depicted (a) human characters (e.g., *five people lying on the floor, three people on a beach*) (*n*=20), (b) landscapes (e.g., *amusement park, a garden*) (*n*=3), mass entities (*a close up of grass, a large pile of money*) (*n*=2), (c) discrete entity/entities, none of which were

[3] *Crocodile* and *cushion* were chosen even thought they did not occur in the two CDIs. This was because it was impossible to find 40 suitable count nouns that would have been listed in both English and Japanese CDIs. We assumed that adult English and Japanese speakers would be familiar with the concepts and words 'crocodile' and 'cushion'.
[4] Full list of test and control items are available at https://osf.io/bh5dj/, along with iStock photo IDs where possible.

semantically similar to the test or control items (e.g., *a snowman, 5 squirrels, 7 potatoes, 6 pegs hanging on a washing line*) (*n*=15).

Photo presentations. To minimize the effect of a given test item to be associated more strongly with singularity or plurality, and to enable analyses on the potential item effects, two PowerPoint presentations (PP1 and PP2) each consisting of 110 photos were created. The two presentations consisted of the same set of test items, but each item was manipulated so that in one PowerPoint presentation, the photo depicted one animal/object (e.g., *one elephant*) and in the other two animals/objects (e.g., *two elephants*). The two presentations consisted of the same control items, but these were manipulated so that the picture in one presentation depicted one of the answer options (e.g., *blue butterfly*), while the other presentation depicted the other answer option (e.g., *orange butterfly*). Participants were randomly allocated to either seeing PP1 or PP2.

Out of the 110 photos, 40 photos depicted test items, 30 depicted control items, and 40 were filler photos. Items were counterbalanced between the two PowerPoint presentations so that 20 of the number questions in each presentation depicted one entity, and 20 depicted two entities. The two photos (e.g., *one elephant* vs. *two elephants*) were as closely matched as was possible. Due to the fact that Japanese speakers pay more attention to background information in scenes than English speakers (e.g., Masuda & Nisbett, 2001; Tajima & Duffield, 2012), the photos depicting test items focused on the test item (i.e., the figure) and no other salient entities were present in the background (see example photos in Appendix A). The same procedure was used with the control photos.

Filler photos were identical in the two presentations. To avoid primacy or recency effects, 20 of the filler photos were placed as the first 10 and the last 10 photos. The rest were randomly mixed with the test and control photos.

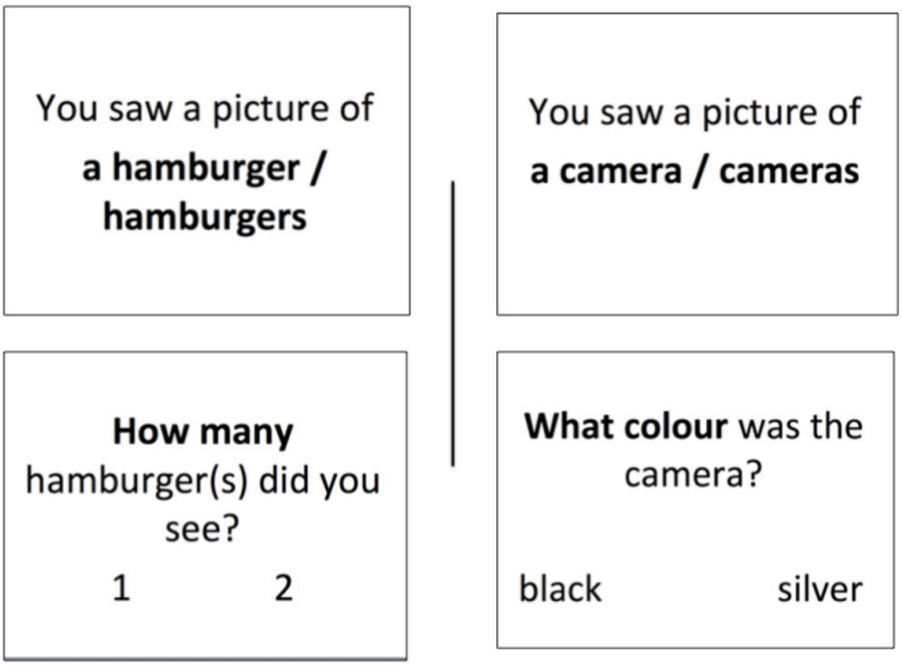
To minimize any order effects, other than the first 10 and last 10 fillers, which were always in the same order, the order of the photos presented was randomized separately for every participant with PowerPoint's "randomize slides" function. The participant saw each slide for two seconds after which the slide automatically changed to present the next photo.

PP presentations testing recall. Seventy questions, one question for each of the test (40) and control (30) photos, were created. No questions were asked about the filler photos. The questions were inserted in random order into four PP presentations and each participant randomly allocated to be tested with one of the four presentations.

Each question consisted of three slides. Slide 1 instructed the participant to recall a photo they had seen during the photo presentation (e.g., test (number) question: *You saw a* *picture of a hamburger/hamburgers* or control question: *You saw a picture of a camera/cameras*). Slide 2 presented a forced choice question (e.g., test (number) question: *How many hamburger(s) did you see? 1 or 2?;* control question: *What color was the camera? Black or Silver*) (see Figure 1.). Slide 3 was blank indicating the test question was complete. The questions were translated into Japanese by a bilingual English-Japanese speaker. The Japanese question slides were worded without classifier information, that is, without number marking as (a) unmarked forms are grammatical and frequently used in Japanese to refer to singular and plural entities and (b) the wording of question slides including the unmarked as well as the marked form would have been unnatural in the context of Japanese (see example slides in Appendix B).

The PP presentation was set to automatically change the slides so that the participant saw Slide 1 for four seconds, Slide 2 for three seconds and Slide 3 for three seconds, after which the next triad would automatically start.

Figure 1. Examples of Number and Control PowerPoint Slides for Testing Recall



Procedure

Instructions were given to the Japanese participants in Japanese and English participants in English.

The participants were told that they were taking part in a memory test in which they would first see 110 photos on the computer screen and that their memory for the photos would be tested straight after the photo presentation. To avoid the participants predicting what kinds of questions would be asked (namely, that they would have to choose from two options) they were told that they would be asked questions about the photos and would have to tell the experimenter what they had seen. As per their random allocation, the participant was then shown PP1 or PP2. The presentation of the 110 photos took approx. 4 minutes.

After having seen the photos, the experimenter told the participant that she would test their memory. The participant was told that this was done by 70 questions for which they would have to give an answer orally. They were shown three practice questions that referred to three of the filler pictures in the same manner as the test questions but were asking the participant to recall details other than the numbers or the control questions (see Appendix C for the practice questions). The participant was told that they should give an answer to every question and if they could not remember the answer they should guess. The experimenter took note of the participant's answers during the test session, and the test situations were also audio-recorded. If the experimenter failed to take note of an answer given during the test, the answer was played back and coded later.[5]

Results

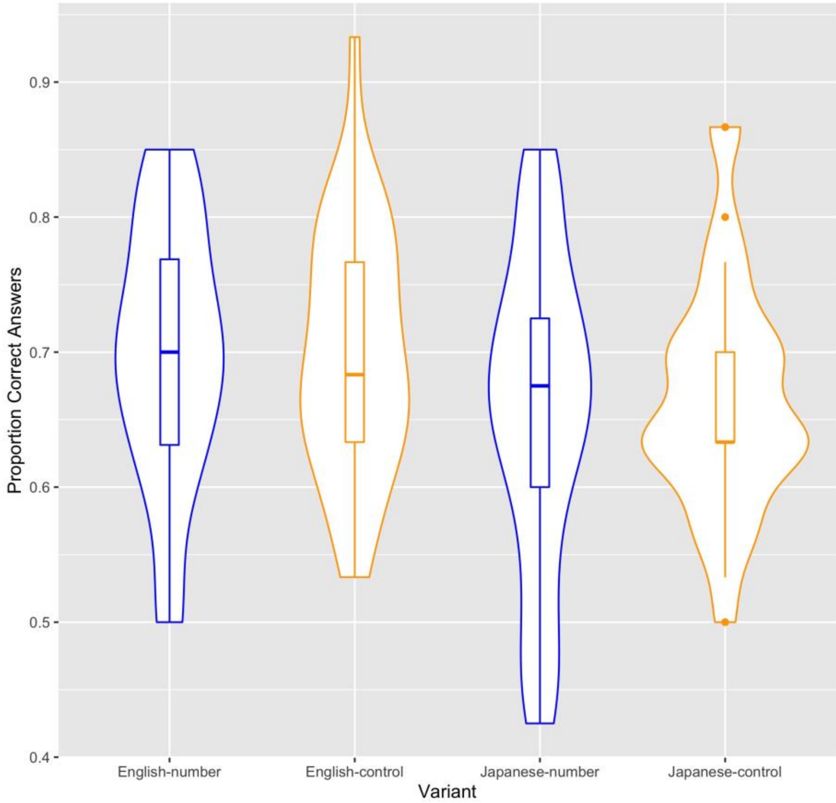
If the obligatory inflectional number marking affects speakers' retention of number information from photos seen, we should find that the English participants outperform the Japanese participants in the items that test memory for number information, in particular, for questions for which the correct answer was 2 (instead of 1). Control items should show no such difference.

We found that although the Japanese group's recall accuracy was slightly weaker than the English group's, there was no difference between the number and control questions in either English or Japanese groups (English group: number: *M* = 70% correct, *SD* = 9.7; range = 50-85, control: *M* = 70%, *SD* = 9.1; range = 53-93; Japanese group: number: *M* = 66%, *SD* = 11, range = 43-85, control: *M* = 67%, *SD* = 9, range = 50-87). See Fig. 2.

Figure 2.

The Distribution of Correct Answers for Number and Control Questions in Experiment 1.

[5] All data and R-codes for analyses are available at https://osf.io/bh5dj/.



We used *R* (R version 4.0.2) and blme package (Chung, Rabe-Hesketh, Dorie, Gelman & Liu, 2013) to perform logistic mixed effects analyses of the effect of language, question type, the correct answer to the question (1 vs. 2) and the on screen location of answers on accuracy. Prior to analysis, we excluded one data point, which lacked a repsonse from a participant. This left the initial data set comprising 4409 data points. All analyses included participants and items as crossed random effects, which meant to counteract any potential effects of individual variation in memory capacity and item artefacts.

To address the question of whether the accuracy of responses to certain question types was determined by a speaker's native language, we fitted a simple main effects model with primary-interest predictors native language (lang) and question type (i.e., number vs. control questions) (qtype), and with secondary-interest variables gender, age and photo presentation (PP1 vs. PP2). We also tested the interaction between language and question type. Table 1 depicts the fixed effects in our final model[6].

[6] Note that whilst all initial regression models in this study included the secondary-interest variables, age, gender and photo presentation, none of these variables reached significance as predictors or improved the model fit. For this reason, we have decided to exclude them from

Table 1.

Final Model for Number and Control Question Data in Experiment 1. Intercept level: Japanese, number question type.

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	0.67	0.09	7.34	0.00
Lang(English)	0.20	0.12	1.76	0.08
Qtype(control)	0.05	0.11	0.45	0.66
Lang:Qtype	-0.04	0.13	-0.29	0.77

The results show that the Japanese speakers performed marginally worse in the test than the English speakers (p = 0.08). However, the lack of a main effect of question type, or interaction between language and question type suggests that both groups performed similarly with the number and control questions.

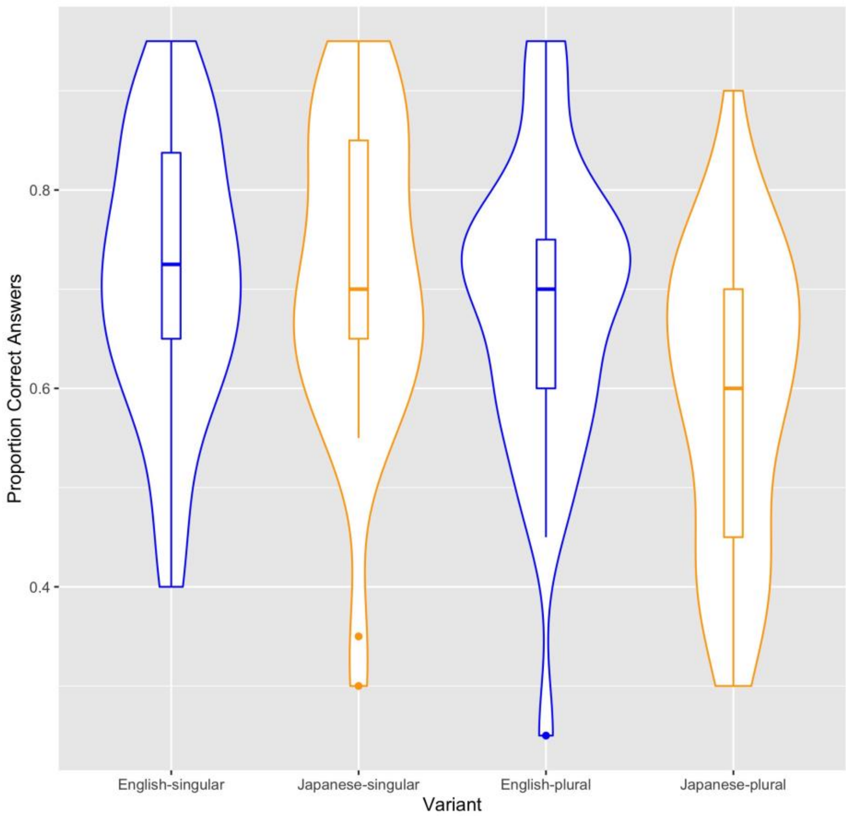
We then looked at the number question data in more detail. Based on our prediction, we investigated if the Japanese and English participants differed in their ability to recall number information when they were responding to questions for which the correct answer was 1 versus 2.

Japanese speakers' mean response accuracy was lower for questions for which the correct answer was 2 (M = 60% correct, SD = 16, range: 30-90) than those questions for which the correct answer was 1 (M = 71% correct, SD = 15, range: 30-95), and Japanese speakers' response accuracy for questions for which the correct answer was 2 was lower than that of English speakers (M = 68% correct). English speakers, on the other hand, had similar mean accuracy in both question types (one: M = 71%, SD = 14, range: 40-95; two: M = 68%, SD = 15, range: 25-95). See Fig. 3.

Figure 3

reporting here and elsewhere, which is in line with existing literature (e.g. Baayen, 2008; Balling, 2008; Jaeger, 2008; Miwa et al., 2014).

The Distribution of Correct Answers for Number Questions for which the Correct Answer was 1 vs. 2 in Experiment 1.



A separate regression analysis of responses was carried out. The same predictors as in the previous analysis were included in the statistical model, with a particular focus on the interaction between language (lang) and the correct answer (1 vs. 2) (answer). The final model is shown in Table 2.

Table 2.

Final Model for Correct Answer (1 vs. 2) Data in Experiment 1. Intercept level: Japanese, answer 1.

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	0.96	0.12	7.86	0.00
Lang(English)	-0.00	0.16	-0.03	0.98
Answer(2)	-0.55	0.12	-4.56	0.00
Lang:Answer	0.40	0.18	2.27	0.02

Table 2 shows that the questions, whose answer was 2, had significantly lower response accuracy than questions for which the correct answer was 1 (β = -.55, z = -4.56, p < .01). The significant language (eng) x answer (2) interaction (β = .40, z = 2.27, p = .02) suggests that this was driven by the Japanese speakers producing significantly fewer correct answers than English speakers, when responding to questions for which the correct answer was 2. The presence of an interaction effect and lack of main effect for language prompt further investigation of the phenomenon within the groups (Japanese and English). Follow-up regression analyses found that English speakers' response accuracy was similar when responding to the number questions for which the correct answer was 1 and those for which the answer was 2 (β = -.15, z = -1.17, p = .24) while Japanese speakers' performance was statistically significantly different between these two question types (β = -.55, z = -4.54, p < .01), with a poorer performance on questions whose answer was 2.

The above results suggest that Japanese speakers are more likely to answer "1" more frequently compared to "2" when responding to number questions. However, as it was impossible to counterbalance the location of the correct answers (i.e., numbers 1 and 2) on the screen in the PP presentations testing recall – as ordering the numbers as 2 or 1 in half of the question slides would have been highly unconventional – it is possible that the difference between English and Japanese speakers in giving correct answer to questions for which the answer was 2 could be due to Japanese speakers having a preference to select the answer option on the left side of the screen. Thus, we ran a separate regression analysis investigating the on screen location of answers given. The same predictors as in the previous analyses were included in the statistical model, with a particular focus on the interaction between language spoken (lang) and the answer location (loc). The statistical model showed that the interaction between language and location was non-significant (p = .26). This suggests that Japanese speakers did not differ from English speakers in their preference for selecting an answer on the screen. Thus, it was the answer and not the location of the answer that brought about the results in the previous analyses.

Discussion

Experiment 1 tested Japanese and English speakers' ability to recall number and control information from photos. We hypothesized that if the compulsory versus optional number information impacts speakers' visual memory, we would find a difference between English and Japanese speakers' ability to recall number information from photos seen, in particular when it came to recalling having seen two entities. In line with our hypothesis, we found that while there was no significant difference in English and Japanese speakers' response accuracy for questions for which the correct answer was 1 or for control questions, the English group outperformed the Japanese group in the questions for which plurality information was needed (i.e., questions for which the correct answer was 2), which was not brought about by different answer location preferences between English and Japanese speakers.

Experiment 1 provides evidence for the Whorfian hypothesis and suggests that the presence/absence of compulsory number marking in speakers' native language affects their ability to retain visual number information. It could be that the presence of compulsory number information with count nouns in English has an effect on the visual memory trace of discrete items, different from the memory trace that Japanese speakers have. More specifically, the plurality information retained from scenes seen seems to be stronger or more salient in English speakers than in Japanese speakers. However, there are at least two possible alternative explanations for our results. First, given that Experiment 1 used a forced choice method, it could be that the presence of systematic plurality marking in English and the absence of such marking in Japanese resulted in language-specific differences in guessing patterns. Namely, it could be that the Japanese participants were more likely to guess 1 when they could not remember the answer or when the memory trace of the number information was fragile, while the English participants' guessing patterns might have been different (e.g., no preference to guess 1 or 2 when they could not recall the answer), bringing about the differences in response accuracy found. Second, it could be that the wording of the English test questions primed English speakers to give more "2" responses than the Japanese speakers, as, due to the characteristics of English, the English questions included singular and plural forms (e.g., *You saw a picture of an elephant/elephants*) and due to the characteristics of Japanese (which allows unmarked forms for both singular and plural entities), Japanese did not (e.g., *象の写真を見ました. Zoo no shashin wo mimashita.* '*You saw elephant*'). The presence of a plural form in the English questions could have primed the English participants to give more plural responses in particular when the memory trace was fragile, while the lack of plural marking in the Japanese questions might have primed the Japanese participants to give more singular responses. Having said that, the fact that at the memory test (i.e., when the questions were presented) the participants had already seen the target photos (with no second chance to see them) and thus they either did or did not remember the answers to the questions presented. This means it is unlikely the wording of the questions helped the memory process. However, to investigate the possibility that cross-linguistic guessing/priming patterns could explain our results, we conducted Experiment 2.

Experiment 2

Experiment 2 used the same method and largely the same materials as Experiment 1, but included a guessing component to shed light on the potential differences in the guessing patterns between Japanese and English participants.

Method

Participants

Twenty-four monolingual Japanese (9 male, 15 female) and 21 monolingual English (10 male, 11 female) speakers aged between 18 - 29 years (English: M_{age} = 22.7 years, SD_{age} = 2.5; Japanese M_{age} = 21.5 years, SD_{age} = 1.1) took part in the study. In addition, one Japanese male and one English female were tested but were excluded due to disruptions during testing (background noise and participant's mobile phone ringing, respectively). The participants were recruited from their universities or through friends, and were tested either in a quiet room at each university or in a private home. None of them had taken part in our Experiment 1. None had any known cognitive disabilities. If the participant had corrected to normal vision, they were instructed to wear their glasses during the test.

Materials

Photo presentations. The same PP picture presentations (PP1 and PP2) were used as in Experiment 1 and participants were randomly allocated to see one of the orders.

PP presentations testing recall. The same question triads (*N*=70) as in Experiment 1 were used. Along with testing recall as in Experiment 1, we further tested the participants' guessing patterns, with an additional 20 test number question triads added in random order in the PP presentations (4 different versions). These 20 new items were to elicit responses for entities that were not depicted in the picture PowerPoint presentations (see https://osf.io/bh5dj/ for these additional items). The fact that the participants had not seen pictures of the entities referred to in these questions allowed us to compare guessing patterns on responses that we knew were guesses. None of the participants expressed during or after the test that they had noticed the fact that they were asked questions about photos that they had not seen.

Procedure

The procedure was the same as in Experiment 1.

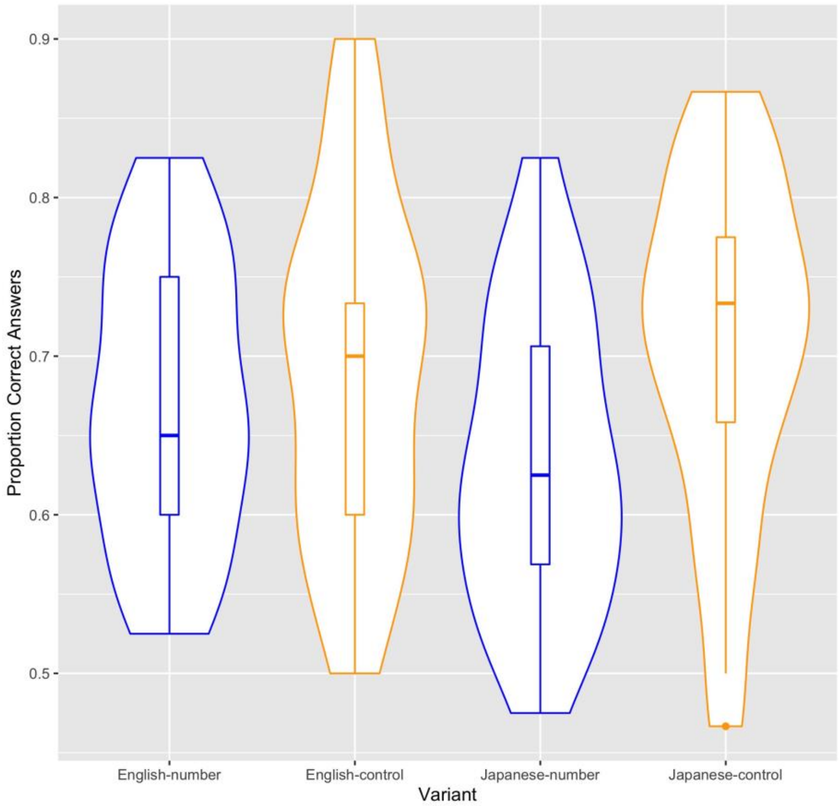
Results

Experiment 2 investigated guessing patterns between English and Japanese speakers when giving responses to questions for which the possible answers were 1 or 2. However, because we used the same materials as in Experiment 1, we will also report the results of the recall data for this second set of participants.

Experiment 2 found very similar results as Experiment 1. First, the Japanese participants' memory recall accuracy for number information (*M* = 63% correct, *SD* = 9, range = 48-83) was lower than their recall of control information (*M* = 71%, *SD* = 11, range = 47-87) while the English participants' recall of number (*M* = 67%, *SD* = 9, range = 53-83) and control (*M* = 69%, *SD* = 11, range = 50-90) information were similar. See Fig. 4.

Figure 4

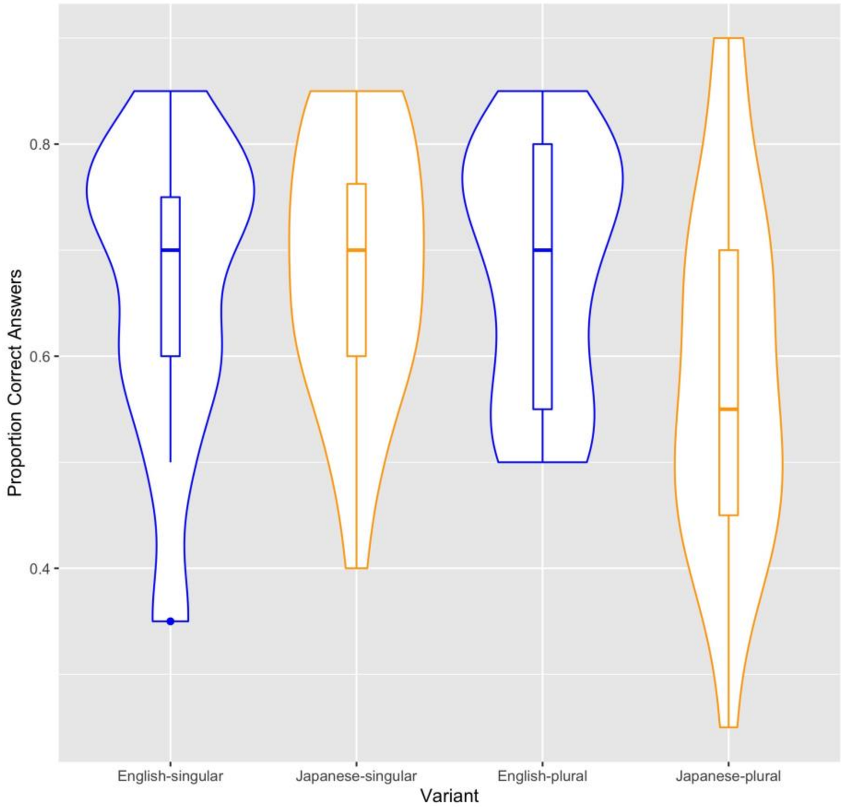
The Distribution of Correct Answers for Number and Control Questions in Experiment 2.



Second, as in Experiment 1, Japanese speakers' response accuracy was lower for questions for which the correct answer was 2 (*M* = 58%, *SD* = 17, range = 25-90) than those questions for which the correct answer was 1 (*M* = 68%, *SD* = 12 , range = 40-85), and that Japanese speakers' response accuracy for questions for which the correct answer was 2 was lower than that of English speakers (*M* = 68%). English speakers, on the other hand, had similar accuracy rates for both question types (1: *M*= 66%, *SD* = 14, range: 35-85; 2: *M* = 68%, *SD* = 12, range: 50-85). See Fig. 5.

Figure 5

The Distribution of Correct Answers for Questions for which the Correct Answer was 1 vs. 2 in Experiment 2.



The same statistical procedures were used as in experiment 1, with the data set comprising 3150 data points (excluding the guessing data). The statistical model fitted to the data included as primary-interest predictors native language (lang) and the question type (i.e., number vs. control questions) (qtype), as secondary-interest variables gender, age and photo presentation (PP1 vs. PP2), and participants and items as crossed random effects. Table 3 summarises the fixed effects in our final model.

Table 3

Final Model for Number and Control Questions in Experiment 2 . Intercept level: Japanese, number question type.

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	0.57	0.10	5.50	0.00
Lang(English)	0.17	0.14	1.26	0.21

Qtype(control)	0.33	0.13	2.56	0.01
Lang:Qtype	-0.24	0.16	-1.50	0.13

There was a significant main effect of question type (qtype), with the control questions showing an overall higher accuracy rate than the number questions ($\beta = 0.33$, $z = 2.56$, $p = .01$). The lack of a significant interaction between language and question type, however, suggests that both the Japanese and the English speakers performed similarly with the number and control questions, and did not warrant a further analysis of the main effect for each group.

Next, we investigated the responses given to the number questions. The same fixed and random effects were used as in Experiment 1 and a separate analysis of responses was carried out, with a particular focus on the interaction between language (lang) and the correct answer (1 vs. 2) (answer). The final model results are shown in Table 4.

Table 4

Final Model for Correct Answer (1 vs. 2) Data in Study 2. Intercept level: Japanese, answer 1.

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	0.80	0.13	6.27	0.00
Lang(English)	-0.11	0.17	-0.69	0.49
Answer(2)	-0.45	0.14	-3.28	0.00
Lang:Answer	0.56	0.20	2.73	0.00

Table 4 shows a significant main effect of correct answer ($\beta = -.45$, $z = -3.28$, $p < .01$), questions for which the correct answer was 2 had lower response accuracy than questions for which the correct answer was 1. It also shows an interaction between the two variables ($\beta = .56$, $z = 2.73$, $p < .01$), suggesting that there is a significant difference in performance between language groups depending on the correct answer. In line with the findings reported in Experiment 1, follow-up regression analyses found that English speakers' response accuracy was similar when responding to the number questions for which the correct answer was 1 and those for which the answer was 2 (β = .10, z = 0.67, p = .50). In contrast, Japanese speakers' performance was statistically significantly different between these two question types (β = -.46, z = -3.28, p < .01), with a poorer performance on questions whose answer was 2.[7]

Lastly, we analyzed the novel aspect of Experiment 2, the participants' guessing behavior to see if the difference in the recall found could be explained by a difference in English and Japanese speakers' guessing patterns in relation to numeric information or if the differences could be a result of differential priming due to the differences in the wording of the experimental materials between English and Japanese. Fig. 6 shows the mean proportion of answers (1 and 2) to the guesses for Japanese and English speakers.

Figure 6

Mean Percentage of Guesses

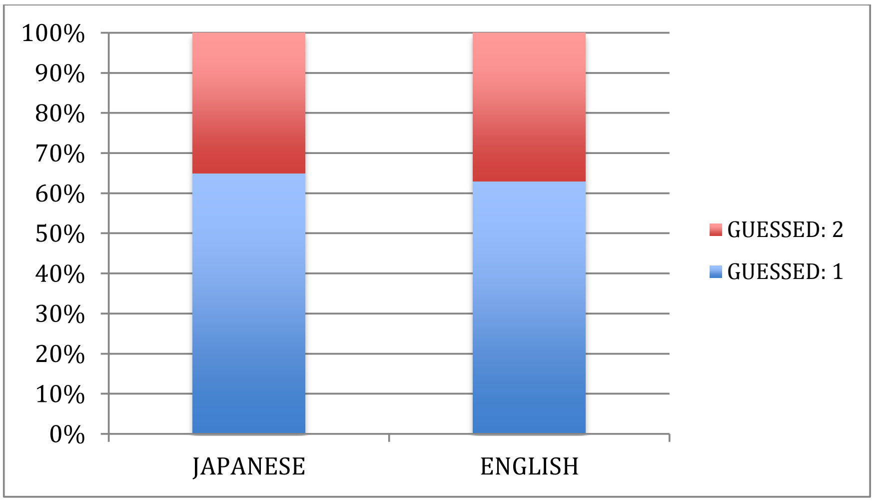


Figure 6 shows that when the participants had not seen a picture corresponding to the question and hence must have been guessing their answer, both Japanese speakers and English speakers were more likely to respond by guessing having seen 1 entity and the proportions of 1 and 2 responses were very similar between Japanese and English speakers

[7] Note that as in experiment 1, the poorer performance with questions for which the correct answer was 2 was not created by Japanese speakers' preference to choose answers on the left side of the screen (main effect of location, p = .48, interaction between language and location p = .54).

(Japanese: *M* = 65% *SD* = 13, range 35-95; English: *M* = 63%, *SD* = 16, range = 35-85) with no evidence of statistical difference (β = .12, *z* = 0.01, *p* = .99).

Discussion

Experiment 2 investigated guessing patterns between Japanese and English speakers, but following Experiment 1, it also analysed the second groups' visual memory of number information. We found that the participants' ability to recall information was in line with the results of Experiment 1. The English group outperformed the Japanese group in the response accuracy in the questions for which the correct answer was 2. The only difference between Experiment 1 and 2 was the significant main effect of question type in Experiment 2, whereby answers to the control questions were overall more accurate than answers to the number questions. The reason we found a significant main effect here but not in Experiment 1 could be because in Experiment 2 the participants answered 20 additional number questions (i.e., the guesses) which is likely to have put extra burden on recall, in particular of information that was difficult to retain, such as number information for participants whose language does not have systematic plurality marking.

The guessing patterns between the two groups in Experiment 2 suggest that guessing behaviour dependent on the differences in the culture or language does not in any obvious way explain the results of Experiments 1 and 2. Furthermore, had the question format (plural vs. no plural wording) between English and Japanese affected the guessing patterns, we would have expected the English group to produce a higher proportion of "two" responses than the Japanese group, but this was not observed. Thus, our analysis of Japanese and English speakers' guessing patterns suggested that English speakers' better memory recall for plurality information does not derive from differences in cultural, linguistic, or experimental priming-related guessing patterns.

Even though the current study does not provide an answer to the question as to why both groups preferred guessing 1 over 2, it might be useful to consider this. One explanation could be related to assumed perceptual salience. When responding to questions referring to photos that the participant cannot recall having seen, their reasoning might follow the assumption that they would be more likely to recall the item or the numeric information if they had seen two entities due to two entities contributing to saliency relative to just one entity. Consequently, if the participant cannot recall a photo, they assume that they only saw one animal or object. Alternatively, even though the items were selected so that they were not in any obvious way related to singularity (e.g., *the sun*) or plurality (e.g., *socks*), it could be that some of the items were more strongly associated with singularity, that is, with no memory trace of the photo, the frequency-based best guess might have been ‘1’. Furthermore, it is possible that the pattern of guessing 1 more often than 2 might have been brought about by different reasons between the two groups. Future research on guessing patterns of number information could shed light on this question.

General discussion

The current study investigated whether cross-linguistic differences in number marking between English and Japanese have an effect on English and Japanese speakers’ recall of number information (1 vs. 2) from photos seen. Experiments 1 and 2 found that English speakers whose language necessitates expression of number information with regular count nouns had a superior ability to recall number information from photos than Japanese speakers whose language allows the omission of this information. This was observed in the condition where the participant had to recall having seen two entities, while there was no difference in recall of having seen one entity. Our study provides evidence for the Whorfian hypothesis by linking structural characteristics of a language to speakers’ visual memory and suggests that language-specific characteristics can boost recall of number information.

Previous studies have investigated the effect of availability/frequency of grammatical constructions on speakers’ ability to recall information (e.g., Fausey & Boroditsky, 2011; Sakarias & Flecken, 2019; Tajima & Duffield, 2012; Tosun et al., 2013) but the current study is the first linking the lack of compulsory number information with a decrease in the ability to recall plurality information in a language that has number concepts and linguistic means to talk about number.

Why the difference in recall accuracy between English and Japanese speakers?

The fact that Japanese speakers showed a lower recall ability for questions for which the correct answer was “2”, could not be explained by preferences in choosing an answer on a particular location on the screen (Experiment 1 and 2) or on different guessing patterns between Japanese and English speakers (Experiment 2). Thus, we argue that our results suggest that native language can influence speakers’ memory trace of objects, possibly because the use of plural inflectional morphemes in English entails procedural (i.e., automatic and unconscious) processing in the English speakers, which is different from the Japanese group. However, it might be useful to consider if anything else could have brought about our results. First, it could be that Japanese speakers did not remember the photos to which the numeric questions referred and thus were guessing and subsequently more frequently responding by the answer “1.” This is unlikely given that the Japanese participants were more or less as good as the English participants in remembering non-numeric details in the photos (control questions). Thus, the effects found cannot be attributed to different abilities in recalling the photos.

Second, given that Japanese speakers are more sensitive to background information in scenes than English speakers, while English speakers focus more on objects in scenes they see (e.g., Masuda & Nisbett, 2001; Senzaki, Masuda & Ishii, 2014; Tajima & Duffield, 2012), it could be that the Japanese participants failed to perceive the entity (i.e., figure) and its characteristics in the photos to the same extent as the English speakers did. However, the pictures were selected so that they depicted very little background information, forcing attention to the target item that was the only salient entity in the photos. Furthermore, the fact that the Japanese participants recalled answers equally well to English speakers to control questions and to number questions when correct answer was “1” suggests that the Japanese participants did sufficiently observe the target entities.

Third, even though we selected the test items so that they were not in any obvious way associated with singularity or plurality, it could be that some of the items were culturally more strongly associated with singularity or plurality in Japanese speakers than in English speakers. As explained in the method section, we created two PowerPoint presentations. Each test item had both a singular entity (e.g., *one elephant*) or a plural entity (e.g., *two elephants*). Each PowerPoint presentation had one of those test items as either a singular or a plural variant which were randomly allocated. This should have in itself minimized the effect of any potential preference for singularity or plurality on an item level, but in addition, based on the response accuracy for each item, we created a rank order for the test items separately for the Japanese and English data so that we deducted the mean response accuracy score for each item when the participants had seen 2 entities (e.g., *two elephants*) from the score when they had seen 1 entity (e.g., *one elephant*). The rank orders of the items based on these scores for English and Japanese highly correlated ($r = 0.507$, p (2-tailed) < 0.01), suggesting that there were no major overall differences in the association of singularity or plurality to the test items between the two groups.

Fourth, some Whorfian effects found might actually not reflect linguistic differences affecting non-linguistic cognition, but that culture specific cognitive biases might be one explanation for the different cognitive processes (e.g., Imai, Kanero & Masuda, 2015; Masuda & Nisbett, 2001), which might be also the case with the results of the current study. However, the fact that Japanese allows number marking by the use of numerals and classifiers means that both Japanese and English speakers can express number information should they want to do so. Thus, there is no obvious cultural reason for the difference in number recall found. That being said, it might very well be that there are distributional frequency differences in reference to number in discourse, thus cultural traditions in that sense may play a part here. Future work could tease the effect of language and culture apart in relation to singular-plural distinction by including bilingual groups of Japanese and English speakers or by including speakers of other languages exhibiting similar grammatical number systems as Japanese and English.

We argue that the fact that Japanese language allows speakers to talk about discrete entities without reference to number, while English does not, results in different cognitive processes in the two language groups. The cognitive processes affected by the presence versus absence of compulsory number marking could be related to perception or memory.

A number of previous studies have investigated Whorfian effects on visual memory, some providing evidence for these effects (e.g., Fausey & Boroditsky, 2011; Frank et al. 2008; Gordon, 2004; Sakarias & Flecken, 2019; Tajima & Duffield, 2012; Tosun et al., 2013), but some failing to do so (e.g., Rosch Heider & Olivier, 1972 and Regier & Xu, 2017 on colour; Ünal et al., 2016 for source memory). Even though some previous studies have failed to find Whorfian effects for recall, previous studies investigating number cognition suggest that limited number vocabulary does have an effect on speakers' ability to recall number information (e.g., Everett & Madora, 2012; Frank et al., 2008; Gordon, 2004). This is while some previous studies have found that lack of number vocabulary does not have an effect on perception of even large numbers of items (Frank et al., 2008). Furthermore, given that Japanese has number words and Japanese speakers can, and do, give number information (Sarnecka et al., 2007), it is unlikely that our Japanese speakers failed to perceive the number of entities seen in the photos. Thus, the difference between English and Japanese-speakers' response accuracy is likely to result from differences in visual memory, but a future eye tracking study might better shed light onto this question.

Alternatively, it could be that due to the classifier system used to refer to number information in Japanese, Japanese speakers perceive both substances and discrete entities as 'mass' and thus fail to retain number information to the same level as English speakers. However, this is unlikely as both Japanese and English speakers have been found to distinguish between substances and objects (Cook et al., 2006; Imai & Gentner, 1997; Iwasaki et al., 2010). Furthermore, the difference in object categorization between classifier and non-classifier language speakers seems to only surface when the classifier is overtly presented in the task (e.g., Gao & Malt, 2009; Huetting, et al., 2010). As none of our test items occurred with the classifier, both groups are likely to have perceived the test items as discrete entities.

To recap, we assume that both our Japanese and English participants observed the number of entities in the photos, but our Japanese participants' memory trace of the numerical information was weaker than that of our English participants'.

Whorfian effects with a small number of entities/objects recalled

Previous research with pre-linguistic infants (e.g., Feigenson & Carey, 2003, 2005; Wang & Feigenson, 2019) and speakers of languages that do not have systematic number labels and morphological number marking (e.g., Everett & Madora, 2012; Frank et al., 2008; Gordon, 2004) suggest that conceptualization and recall of a small number of entities (≤ 3) can be resistant to Whorfian effects. The current study found that structural and distributional characteristics of native language had a significant effect on recall of 2 entities, suggesting Whorfian effects even with a small number of entities. Two factors could explain this discrepancy. First, the methods used in the current and previous studies were different. In the current study the participants were shown a large number of scenes and had to retain the information from the scenes for several minutes, thus involving storage of the visual information, while in most tasks in previous studies the participants relied on their visual short-term memory. Thus, the current and previous studies might tap into different memory processes, and suggest that linguistic features (such as number marking) can affect different types of memory processes differently.

Second, the current study tested recall of two speaker groups whose languages have number terms (cf. Pirahã), but while the native language of one group (English) necessitates speakers to routinely express number information, the other group's (Japanese) language does not. Based on the Whorfian hypothesis, the systematic linguistic difference when referring to 1 versus > 1 entities in English and the absence of such clear distinction in Japanese is likely to have resulted in the different recall patterns observed.

Taken together, previous studies suggest that number information is cognitively salient and thus not ignored even by pre-linguistic infants or adult speakers whose language does not have systematic number words/grammatical marking (e.g., Feigenson & Carey, 2003, 2005;

Frank et al., 2008; Gordon, 2004; Pica et al., 2004; Spaepen, Coppola, Spelke, Carey, & Goldin-Meadow, 2011; Wang & Feigenson, 2019). However, the current study indicates that speakers of a language that has systematic number marking put more cognitive resources into retaining number information from scenes seen than speakers of a language that has no systematic marking.

Singular (1) versus plural (2)

In the current study we tested recall of 1 and 2 entities as this pair represented the smallest distance between singularity and plurality, given that neither English nor Japanese has a duality marker like some other languages have, for example, Slovenian (e.g., Marušic et al., 2016). However, it is possible that the recall of 2 versus > 2 entities is not identical. For example, higher numbers and thus the higher distance from 1 might make plurality information more salient and thus create a stronger memory trace even in Japanese speakers. Alternatively, *two elephants* versus *a herd of elephants* might result in a different memory trace altogether representing the item occurring in plural versus a group. The effect of singular versus dual versus plural marking on visual memory would benefit from further investigation.

Linguistic or non-linguistic effects?

One important issue when investigating potential Whorfian effects is the extent to which research conducted demonstrates a link between language and other areas of cognition. Some research might provide evidence that the characteristics of native language affect the speakers' language-related cognitive processes. For example, plural marking would have to be accessed online when, for example, describing events, thus bringing about *Thinking for Speaking* effects (Slobin, 1996, 2003). Other studies can be taken to demonstrate that language has an effect on how speakers encode and/or process experiences outside the realm of language. For example, studies of how we perceive scenes without linguistic output during a task, contributes to the wider theoretical debate about the organization of language and other areas of cognition.

The most convincing way to demonstrate that language shapes non-linguistic cognition is by the use of non-verbal tasks where there is no inner speech (Bloom & Keil, 2001:358; Gleitman & Papafragou, 2005; Pinker, 1994, 2007) as these types of tasks have the potential of showing that language affects non-linguistic rather than just linguistic cognition. However, this is a major problem only if the task is such that inner speech or language use during an experiment would have an obvious effect on the completion of the task (e.g., a categorization task, online description of scenes). In relation to the current study, in the memory phase, the participants were simply presented with 110 photos and instructed to try and remember what they saw. Although this phase involved no linguistic material it is possible that the participants were silently verbalizing what they had seen, hence adding a linguistic aspect to the memory trace. However, it seems that the effect of these possible verbalizations might be relatively small. For example, Masuda, et al. (2017) study found that labelling and not overtly labelling items seen in photos (speaking out loud) resulted in very similar responses in similarity judgement tasks. Thus, it tentatively suggests that our participants potentially silently verbalizing *'(one) elephant'* or *'(two) elephant(s)'* while seeing a picture of elephant(s) might not have had a major effect on the results. Second, although the recall phase was verbal (PowerPoint slides with questions and an oral answer was required), the participants had no access to the photos seen at the memory phase stage, and so either did or did not remember the answer to the questions. This being the case, the linguistic aspects of our method are unlikely to have had a huge impact on the response accuracy rates during recall. Thus, we argue that the fact that English systematically encodes grammatical number on (regular) count nouns, while Japanese does not, leads to English speakers developing a habit of putting more cognitive resources on number information than Japanese speakers, whether the particular process lies more towards the linguistic or non-linguistic end of the cognition continuum. Future research could provide stronger evidence for compulsory number information affecting non-linguistic cognition, for example, by conducting a study in which at the memory recall phase the participants were presented with pictorial questions to avoid use of overt language at this stage or a linguistic distractor task during the memory phase to minimize opportunity to silently verbalize the items seen.

Conclusion

The current study found that English speakers whose language necessitates expression of singular vs. plural information whenever regular count nouns are produced showed a superior recall accuracy of number information of plural items than Japanese speakers, whose language does not necessitate expression of plurality information. This suggests that the structure of native language affects speakers' visual memory, providing support for the Whorfian hypothesis.

References

Asami, S. (2006). Nichiei-go kan no fukusu hyogen: Meishi ni donoyouni kakawaruka. [Plural expressions in Japanese and English: How it relates with nouns]. *Hokkaido University Papers, 22,* 15-32.

Athanasiadou, I. & Athanasopoulos, P. (2017). Plural mass nouns and the construal of individuation: Crosslinguistic evidence from verbal and nonverbal behaviour in labelling and non-labelling contexts. *Cognitive Semantics, 3*, 62-94.

Athanasopoulos, P. (2006). Effects of the grammatical representation of number on cognition in bilinguals. *Bilingualism: Language and Cognition*, *9*, 89-96.

Athanasapoulos, P., & Kasai, C. (2011). Effects of second language on cognition in English users of L2 Japanese. *Vigo International Journal of Applied Linguistics*, *8,* 13-27.

Baayen, R. H. (2008). *Analyzing linguistic data. A practical introduction to statistics using R.* Cambridge: Cambridge University Press.

Balling, L. W. (2008). A brief introduction to regression designs and mixed-effects modelling by a recent convert. *Copenhagen Studies in Language, 36*, 175–192.

Berlin, P., & Kay, B. (1969). *Basic color terms: Their universality and evolution.* Berkeley: University of California Press.

Bloom, P., & Keil, F. C. (2001). Thinking through language. *Mind & Language, 16*, 351-367.

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

Brannon, E. M., & Roitman, J. (2003). Nonverbal representations of time and number in non-human animals and human infants. In W. Meck (Ed.), *Functional and neural mechanism of interval timing* (pp. 143-182). New York: CRC Press.

Bylund, E., & Athanasopoulos, P. (2017). The Whorfian time warp: Representing duration through the language hourglass. *Journal of Experimental Psychology: General*, *146*, 911-916.

Chen, J., Bowerman, M., Huettig, F., & Majid, A. (2010). Do language-specific categories shape conceptual processing? Mandarin classifier distinctions influence eye gaze behaviour, but only during linguistic processing. *Journal of Cognition and Culture, 10,* 39-58.

Cheng, L., & Sybesma, R. (1998). Yi-wan tang, yi-ge Tang: Classifiers and massifiers. *Tsing-Hua Journal of Chinese Studies, New Series, 28,* 385-412.

Chung Y., Rabe-Hesketh S., Dorie V., Gelman A., & Liu., J. (2013). A nondegenerate penalized likelihood estimator for variance parameters in multilevel models. *Psychometrika,78*, 685–709.

Cibelli, E., Xu, Y., Austerweil, J. L., Griffiths, T. L., & Regier, T. (2016). The Sapir-Whorf Hypothesis and probabilistic interference: Evidence from the domain of color. *PloS ONE 11*(7): e0158725.

Cook, V., Bassetti, B., Kasai, C., Sasaki, M., & Takahshi, J. (2006). Do bilinguals have different concepts? The case of shape and material in Japanese L2 users of English. *International Journal of Bilingualism, 10*, 137-152.

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

Davidoff, J., Davies, I., & Roberson, D. (1999) Colour categories in a stone-age tribe. Nature 1999, 398: 203-204.

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

Everett, C. & Madora, K. (2012). Quantity recognition among speakers of an anumeric language. *Cognitive Science, 36*, 130-141.

Fausey, C. M., & Boroditsky, L. (2011). Who dunnit? Cross-linguistic differences in eye-witness memory. *Psychonomic Bulletin & Review*, *18,* 150-157.
Feigenson, L., & Carey, S. (2003). Tracking individuals via object-files: Evidence from infants' manual search. *Developmental Science, 6,* 568-548.
Feigenson, L., & Carey, S. (2005). On the limits of infants' quantification of small object arrays. *Cognition, 97,* 295-313.
Fenson, L., Dale, P. S., Reznick, J. S., Thal, D., Bates, E., Hartung, J. P., Pethick, S., & Reilly, J. S. (1993). *The MacArthur communicative development inventories: User's guide and technical manual.* Baltimore: Paul H. Brookes Publishing Co.
Fodor, J. A. (1975). *The language of thought.* New York: Crowell.
Frank, M. C., Everett, D. L., Fedorenko, E., & Gibson, E. (2008). Number as a cognitive technology: Evidence from Pirahã language and cognition. *Cognition*, *108*, 819-824.
Gallistel, C. R., & Gelman, R. (2000). Non-verbal numerical cognition: From reals to integers. *Trends in Cognitive Sciences, 4*, 59-65.
Gao, M. Y., & Malt, B. C. (2009). Mental representation and cognitive consequences of Chinese individual classifiers. *Language and Cognitive Processes, 24*, 1124-1179.
Gelman, R., & Gallistel, C. R. (2004). Language and the origin of numerical concepts. *Science*, *306*, 441-443.
Gelman, R., & Cordes, S. (2001). Counting in animals and humans. In E. Dupoux (Ed.), *Language, brain and cognitive development: Essays in honor of Jacques Mehler* (pp. 279-303). Cambridge, MA: The MIT Press.
Gleitman, L., & Papafragou, A. (2005). Language and thought. In K. J. Holyoak & B. Morrison (Eds.), *Cambridge handbook of thinking and reasoning* (pp. 633-661). Cambridge: Cambridge University Press.
Gordon, P. (2004). Numerical cognition without words: Evidence from Amazonia. *Science, 306*, 496-499.
Gu, Y., Zheng, Y., & Swerts, M. (2019). Which is in front of Chinge people, past or future? The effect of language and culture on temporal gestures and spatial conceptions of time. *Cognitive Science, 43*, e12804.
Gumperz, J., & Levinson, S. (Eds.). (1996). *Rethinking linguistic relativity.* Cambridge, UK: Cambridge University Press.
Huettig, F., Chen, J., Bowerman, M., & Majid, A. (2010). Do language-specific categories shape conceptual processing? Mandarin classifier distinctions influence eye gaze behavior, but only during linguistic processing. *Journal of Cognition and Culture, 10*, 39-58.
Imai, M., & Gentner, D. (1997). A crosslinguistic study on constraints on early word meaning: Linguistic influence vs. universal ontology. *Cognition, 62,* 169-200.
Imai, M., Kanero, J., & Masuda, T. (2015). The relation between language, culture, and thought. *Current Opinion in Psychology*, *8*, 79-77.
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. Golding-Meadow (Eds.), *Language in mind; Advances in the study of language and thought* (pp. 429-464). A Bradford Book, Cambridge, MA: The MIT Press.
Iwasaki, N., Vinson, D. P., & Vigliocco, G. (2010). Does the grammatical count/mass distinction affect semantic representations? Evidence from experiments in English and Japanese. *Language and Cognitive Processes,* 25, 189-223.
Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language, 59(4),* 434–446.
Levinson, S. C. (1996). Language and Space. *Annual Review of Anthropology*, *25*, 353-382.

Lucy, J. A. (1992). *Grammatical categories and cognition: A case study of the linguistic relativity hypothesis.* Cambridge: Cambridge University Press.
Lucy, J. A. (2016). Recent advances in the study of linguistic relativity in historical context: A critical assessment. *Language Learning, 66*, 487-515.
Lucy, J. A., & Gaskins, S. (2003). Interaction of language type and referent type in the development of nonverbal classification preferences. In D. Gentner & S. Goldin-Meadow (Eds.), *Language in mind: Advances in the study of language and thought* (pp. 465-492). Cambridge, MA: The MIT Press.
Ogura, T., & Watamaki, T. (2004). Nihongo makkaasaa nyuuyooji gengo hattatsu shitsumon yooshi [The Japanese MacArthur communicative development inventory]. Kyoto: Kyoto Kokusai Shakai Fukushi Center.
Majid, A., Bowerman, M., Kita, S., Haun, D. B. M., & Levinson, S. C. (2004). Can language structure cognition? The case for space. *Trends in Cognitive Sciences*, *8*, 108-114.
Marušic, F., Žaucer, R., Plesnićar, V., Razboršek, T., Sullivan, J., & Barner, D. (2016). Does grammatical structure accelerate number word learning? Evidence from learners of dual and non-dual dialects of slovenian. *PLoS One, 11*(8):e0159208.
Masuda, T., Ishii, K., Miwa, K., Rashid, M., Lee, H., & Mahdi, R. (2017). One label of two? Linguistic influences on the similarity judgment of objects between English and Japanese speakers. *Frontiers in Psychology*, *8*, 1637.
Masuda, T., & Nisbett, R. E. (2001). Attending holistically versus analytically: Comparing the context sensitivity of Japanese and Americans. *Journal of Personality and Social Psychology, 81*. 922-934.
Mazuka, R., & Friedman, R. S. (2000). Linguistic relativity in Japanese and English: Is language the primary determinant in object classification? *Journal of East Asian Linguistics, 9,* 353-377.
Miwa, K., Dijkstra, T., Bolger, P., & Baayen, R. H. (2014). Reading English with Japanese in mind: Effects of frequency, phonology, and meaning in different-script bilinguals. *Bilingualism: Language and Cognition, 17(3)*, 445 – 463.
Paramei, G. V., Griber, Y. A., & Mylonas, D. (2018). An online colour naming experiment in Russian using Munsell colour samples. *Color Research and Application, 43*, 358-374.
Pica, P., Lemer, C., Izard, V., & Dehaene, S. (2004). Exact and approximate arithmetic in an Amazonian indigene group. *Science*, *306*, 499-503.
Pinker, S. (1994). *The language instinct.* New York: Harper Collins.
Pinker, S. (2007). *The stuff of thought: Language as a window into human nature.* New York, NY: Viking.
Regier, T. & Xu, Y. (2017). The Sapir-Whorf hypothesis and infernece under uncertainty. *WIREs Cognitive Science, 8*:e1440.
Roberson, D., Davies, I., & Davidoff, J. (2000). Color categories are not universal. Replications and new evidence from a stone-age culture. *Journal of Experimental Psychology: General, 129*, 369-398.
Rosch Heider, E. & Olivier, D. C. (1972). The structure of the colour space in naming and memory for two languages. *Cognitive Psychology, 3,* 337-354.
Saalbach, H., & Imai, M. (2007). The scope of linguistic influence: Does a classifier system alter object concepts? *Journal of Experimental Psychology: General, 136,* 485-501.
Saalbach, H., & Imai, M. (2012). The relation between linguistic categories and cognition: The case of numeral classifiers. *Language and Cognitive Processes, 27,* 381-428.
Sakarias, M., & Flecken, M. (2019). Keeping the results on sight and mind: General cognitive principles and language-specific influences in the perception and memory of resultative events. *Cognitive Science, 43*, e12708.

Sarnecka, B. W., Kamenskaya, V. G., Yamana, Y., Ogura, T., & Yudovina, Y. B. (2007). From grammatical number to exact numbers: Early meanings of 'one', 'two', and 'three' in English, Russian, and Japanese. *Cognitive Psychology*, *55*, 136-168.

Sato, S. & Athanasopoulos, P. (2018). Grammatical gender affects gender perception: Evidence for the stuctural-feedback hypothesis. *Cognition, 176,* 220-231.

Sato, S., Casaponsa, A., & Athanasopoulos, P. (in press). Flexing gender perception: Brain potentials reveal the cognitive permeability of grammatical informataion. *Cognitive Science.*

Sedlmeier, P. Tipandjan, A., & Jänchen, A. (2016). How persistent are grammatical gender effects? The case of German and Tamil. *Journal of Psycholinguistic Research, 45*, 317-336.

Senzaki, S., Masuda, T., & Ishii, K. (2014). When is perception top-down and when is it not? Culture, narrative, and attention. *Cognitive Science*, *28*, 1493-1506.

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

Slobin, D. I. (2003). Language and thought online: Cognitive consequences of linguistic relativity. In D. Gentner & S. Goldin-Meadow (Eds.), *Language in mind: Advances in the study of language and thought (*pp. 157-192). London: The MIT Press.

Spaepen, E., Coppola, M., Spelke, E. S., Carey, S. E., & Goldin-Meadow, S. (2011). Number without a language model. *Proceedings of the National Academy of Sciences*, *108*, 3163-3168.

Speed, L., Chen, J., Huettig, F., & Majid, A. (in press). Classifier categories reflect, but do not affect conceptual organization. *Journal of Experimental Psychology: Learning, Memory and Cognition.*

Speed, L. & Majid, A. Linguistic features of fragrances: The role of grammatical gender and gender association. *Attention, Perception & Psychophysics, 81*(6), 2063-2077.

Tojyo, K. (2014). Two types of nominal classifiers in Japanese: Quasi-classifiers and pseudo-classifiers. *The Society of Japanese Linguistics, 10*(4),16-32.

Tosun, S., Vaid, J., & Geraci, L. (2013). Does obligatory linguistic marking of source of evidence affect source memory? A Turkish/English investigation. *Journal of Memory and Language, 69,* 121-134.

Tajima, Y., & Duffield, N. (2012). Linguistic versus cultural relativity: On Japanese-Chinese differences in picture description and recall. *Cognitive Linguistics*, *23*, 675-709.

Ünal, E., Pinto, A., Bunger, A., & Papafragou, A. (2016). Monitoring sources of event memories: A cross-linguistic investigation. *Journal of Memory and Language, 87*, 157-176.

Vigliocco, G., Vinson, D. P., 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.

Wang, J., & Feigenson, L. (2019). Infants recognize counting as numerically relevant. *Developmental Science, 22*:e12805.

Whorf, B. (1956). *Language, thought, and reality*. Cambridge: MA: The MIT Press.

Wolff, P., & Holmes, K. (2011). Linguistic relativity. Wiley Interdisciplinary Reviews: *Cognitive Science*, *2*, 253-265.

Zlatev, J., & Blomberg, J. (2015). Language may indeed influence thought. *Frontiers in Psychology,* 6:1631.

Appendices

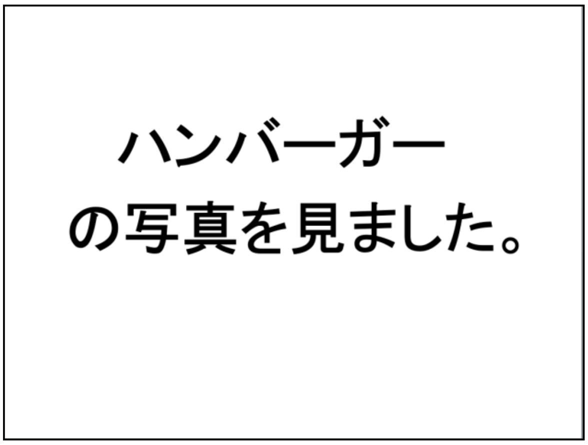
Appendix A. Examples of photos used

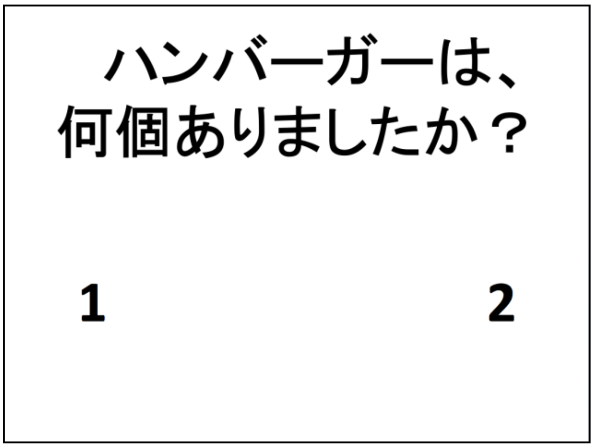
Lion, singular

Lion, plural

Appendix B. Examples of the wording in the Japanese slides.

a) Hamburger





b) Camera

カメラ
の写真を見ました。

カメラは、
何色でしたか？

黒 銀

Appendix C. Practice questions.

Slide1: You saw a picture of a chef / chefs making pizza.
Slide2: Were the chefs boys girls
Slide 3: (blank)
Slide 4: You saw a picture of a cleaner / cleaners.
Slide 5: What did the cleaner clean with? hoover mop
Slide 6: (blank)
Slide 7: You saw a picture of a fisherman / fishermen.
Slide 8: What did the fisherman use for fishing? net rod
Slide 9: (blank)