Does language affect memory for object position?
A crosslinguistic comparison

Solveig Bosse & Anna Papafragou

To cite this article: Solveig Bosse & Anna Papafragou (2018): Does language affect memory for object position? A crosslinguistic comparison, Spatial Cognition & Computation

To link to this article: https://doi.org/10.1080/13875868.2018.1440396

Published online: 02 Mar 2018.
Does language affect memory for object position? A crosslinguistic comparison

Solveig Bosse and Anna Papafragou

Department of English, East Carolina University, Greenville, NC, USA; Department of Psychological and Brain Sciences, University of Delaware, Newark, DE, USA

ABSTRACT

We present an experimental investigation of the influence of language on spatial memory. German and English differ in how they encode the axial position (standing vs. lying) of an inanimate figure object which is supported by a ground object. In a series of four experiments, we show that German and English speakers’ ability to detect changes in axial position during recognition memory tasks is unaffected by this linguistic difference. Even when participants are required to use language to encode the spatial scenes, later recognition memory performance does not differ between the two language groups. We discuss implications of these findings for the relation between language and cognition.

KEYWORDS

Memory; spatial language; spatial cognition; orientation; positionals; language and thought

Introduction

It has been widely assumed that spatial meanings expressed by natural language must reflect core conceptual entities and relations which are important in spatial cognition (Clark, 1973; Jackendoff, 1983; Landau & Jackendoff, 1993; Talmy, 1985). Nevertheless, individual languages possess different means for encoding space (Bowerman, 1996; Levinson, 2003; Talmy, 1985). A question of central interest within cognitive science is whether one’s native language can itself shape the spatial representations recruited in perception, memory and other cognitive processes (see Gentner & Goldin-Meadow, 2003; Gleitman & Papafragou, 2005, 2012; Malt & Wolff, 2010; Wolff & Holmes, 2011; for reviews).

The literature has raised two influential hypotheses outlining a possible role for language in spatial cognition. According to a first hypothesis, habitual exposure to the spatial vocabulary of a language has long-term effects on spatial memory and spatial organization for speakers of that language (Brown & Levinson, 1993; Levinson, 2003; Lucy, 1996; Majid, Bowerman, Kita, Haun & Levinson, 2004; Pederson et al., 1998). On this position, language can lead to a kind of conceptual tuning, or loss of sensitivity to conceptual distinctions which are not captured by the semantics of one’s native language (a situation akin to phonological tuning, or loss of non-native sound distinctions; Werker & Tees,
For instance, languages such as Dutch and English typically make use of an egocentric, body-defined coordinate system to refer to locations of objects in small-scale arrays (“The fork is to the left of the plate”). However, Tseltal, a Mayan language spoken in the Tenejapa area of Chiapas (Mexico), uses geocentrically-defined terms which roughly correspond to the north-south axis (“The fork is to the south of the plate”; Brown & Levinson, 1993). Several studies have shown that Dutch- and Tseltal-speakers’ choice of coordinate systems in spatial reasoning reflects language-specific encoding preferences (Majid et al., 2004; Pederson et al., 1998). In another demonstration, prelinguistic infants appear to distinguish among various spatial relations (Casasola & Cohen, 2002; Hespos & Spelke, 2004; McDonough, Choi & Mandler, 2003). However, adults’ performance on nonlinguistic tasks seems to vary with the language they have acquired: adult speakers of Korean have been reported to attend to tightness-of-fit, unlike adult speakers of English whose language does not encode this distinction systematically (Hespos & Spelke, 2004; McDonough et al., 2003).

According to a second hypothesis, language offers an additional, flexibly recruited system of encoding, organizing and tracking spatial experience, even though it may not alter underlying spatial representations themselves (Gleitman & Papafragou, 2005; Landau, Dessalegn & Goldberg, 2010). In support of this possibility, there is evidence that linguistic terms can be implicitly brought online to bear on seemingly nonlinguistic visuo-spatial processes. For instance, eye-tracking studies have shown that, in challenging memory tasks, participants make implicit use of language to commit details of a motion event to memory; as a result, eye gaze patterns differ in speakers of different languages in accordance with the way their native language encodes motion (Papafragou, Hulbert & Trueswell, 2008). Such online effects of language on attention allocation disappear when people are given a secondary task which involves verbal interference and thus prevents them from using language covertly to code motion stimuli (Trueswell & Papafragou, 2010; see also Athanasopoulos et al., 2015; Athanasopoulos & Bylund, 2013). To take another example, people covertly access color boundaries in their native language to perform color discrimination; such effects, however, disappear under verbal interference (Winawer et al., 2007; cf. Mitterer, Horschig, Musseler & Majid, 2009). In both the domains of motion and color, online effects of language can be rapid and subtle (Flecken, Athanasopoulos, Kuipers & Thierry, 2015; Thierry, Athanasopoulos, Wiggett, Dering & Kuipers, 2009; Trueswell & Papafragou, 2010).

In further support of this second hypothesis, overtly present linguistic labels have been shown to affect spatial memory and cognition. In an early demonstration, Carmichael, Hogan and Walter (1932) presented participants with ambiguous images (e.g., two small circles with a line attaching them) in conjunction with one of two labels (e.g., “This figure resembles eyeglasses/dumbbells”). In a reproduction drawing task, participants generally did not produce an accurate representation of the images but instead, produced a representation resembling the
object labeled. In another classic study, Loftus (1979) presented participants first with visual stimuli and then with written texts. The texts either contradicted the visual stimuli or did not conflict with them. She found that participants were influenced by the written text when tested on the original, visual stimuli. More recent studies have shown that memory for motion events can be biased depending on whether path (exit) or manner (skip) verbs accompany the events, regardless of whether the verbs are provided by the experimenter (Billman & Krych, 1998) or generated by participants (Billman, Swilley & Krych, 2000). Relatedly, Feist and Gentner (2007) showed that, after participants viewed ambiguous spatial displays paired with spatial prepositions (e.g., on), they exhibited recognition memory distortions caused by the spatial terms. Lupyan (2008) reported similar effects of object names on recognition memory. Finally, Archambault, O’Donnell and Schyns (1999) showed that the linguistic level at which an object was categorized (general, e.g., “a mug”, or specific, e.g., “Steve’s mug”) influenced the time it later took people to detect a change in a picture containing the object: if objects had been introduced at a specific (individual) level, then the changes were detected faster than if the objects had been introduced at a general level.

At present, the nature and scope of the effects of language on spatial cognition are the subject of vigorous debate. The idea that language leads to “a deep-seated specialization of the mind” (Levinson, 2003, p.291) through mechanisms such as conceptual tuning has been challenged by evidence of similarities in underlying cognitive representations across children and adults from different language communities (Gennari, Sloman, Malt & Fitch, 2002; Li, Dunham & Carey, 2009; Munnich, Landau & Dosher, 2001; Papafragou, Hulbert & Trueswell, 2008; Papafragou, Massey & Gleitman, 2002; Ünal, Pinto, Bunger & Papafragou, 2016). Evidence in the specific domains of coordinate spatial systems (Li, Abarbanell, Gleitman & Papafragou, 2011) and tight/loose fit distinctions (Norbury, Waxman & Song, 2008) show that adults are sensitive to conceptual distinctions which go well beyond the linguistic resources of their native language.

Furthermore, the extent to which language exerts an online influence onto cognitive processes along with the mechanisms producing such effects is not clear. Some commentators have argued that linguistic intrusions into otherwise nonlinguistic tasks are pervasive, such that “linguistic representations normally meddle in even surprisingly simple objective perceptual decisions” (Winawer et al., 2007, p. 7784). However, other evidence suggests that linguistic intrusions may be more limited: for instance, participants do not consistently generate implicit verbal descriptions of spatial scenes to support visual recognition memory (Feist & Gentner, 2007; cf. Pecher, Dantzig, Von, Zwaan & Zeelenberg, 2009). Relatedly, it has been proposed that labeling to-be-remembered stimuli creates a representational shift which “cleans up” the stimuli and makes their representations “less noisy and more categorical” in ways which
might benefit later between-category recognition memory (Lupyan, 2008; p. 364; cf. also Feist & Gentner, 2007; Loftus, 1979). However, overt labeling does not always produce such effects: studies on motion events have found no effect of language on memory in circumstances in which people described each of the target events overtly (Gennari et al., 2002; Papafragou et al., 2002) or covertly (Trueswell & Papafragou, 2010) prior to the memory task (see also Bahrick & Boucher, 1968; Intraub, 1979).

In this article, we explore the extent of the influence of cross-linguistic differences on spatial memory with the aim to contribute to the theoretical discussion of the relation between spatial language and cognition. Our tasks specifically examine the relationship between spatial language and recognition memory. Even though recognition does not intrinsically require the use of language (cf. Gennari et al., 2003), prior theoretical accounts and results sketched above have argued for a strong role of language in recognition memory, both when linguistic labels are explicitly invoked (Lupyan, 2008) as well as when no overt labels accompany spatial encoding (Levinson, 2003). The spatial phenomenon we focus on is the orientation/axial position of objects, a novel empirical area which has not received a lot of attention in the literature. To preface our experiments, in the next section we present a brief overview of how object position is encoded cross-linguistically and outline predictions of the different hypotheses presented above about how linguistic positional encoding might interface with the nonlinguistic representation of object axial position.

**Object axial position in language and cognition**

Languages encode the position of objects in space in a variety of ways. Positional information is so basic as to have been considered by Aristotle a linguistic primitive, or a “noncomposite” category:

Expressions which are in no way composite signify substance, quantity, quality, relation, place, time, position, state, action, or affection. To sketch my meaning roughly, examples of substance are ‘man’ or ‘the horse’, of quantity, such terms as ‘two cubits long’ or ‘three cubits long’, of quality, such attributes as ‘white’, grammatical. ‘Double’, ‘half’, ‘greater’, fall under the category of relation; ‘in the market place’, ‘in the Lyceum’, under that of place; ‘yesterday’, ‘last year’, under that of time. ‘Lying’, ‘sitting’, are terms indicating position, ‘shod’, ‘armed’, state; ‘to lance’, ‘to cauterize’, action; ‘to be lanced’, ‘to be cauterized’, affection. (Aristotle, *Categories*, Part 4; cited by Ameka & Levinson, 2007, p.847, emphasis added)

Languages differ in the degrees of specificity with which they encode object position. According to an extensive recent cross-linguistic survey (Ameka & Levinson, 2007), languages can be classified into four basic types depending
on the kind of locative predication they use to answer the question *Where is the X?* (where X is the typical to-be-located object – i.e., a small moveable inanimate object): A first group of languages uses no verb in the basic locative construction (e.g., Austronesian, Papua New Guinea); a second group uses a single locative verb (e.g., copula *be* in English, or a locative verb in Japanese); a third group uses a small contrastive set of 3–7 locative verbs such as “standing”, “lying” and “sitting” to encode specific aspects of an object’s position and axial orientation. (e.g., Dutch, Arrernte); finally, a fourth group of languages uses a large set of positional verbs (e.g., German, Tzeltal). To illustrate these contrasts with a simple example, consider the pictures in Figure 1. Each picture depicts a roll of toilet paper (figure object) on a washing machine (ground object). The difference between the two pictures is the position (vertical vs. horizontal) of the roll of toilet paper on the washing machine. In German, Figure 1a is naturally described with sentence (1), whereas Figure 1b is naturally described with sentence (2). In English, both pictures in Figure 1 are typically described with sentence (3).

(1) Das Klopapier steht auf der Waschmaschine.
the toilet paper stands on the washing machine
(2) Das Klopapier liegt auf der Waschmaschine.
the toilet paper lies on the washing machine
(3) The toilet paper is on the washing machine.

As mentioned already, in German, the verbs *stehen* “stand” and *liegen* “lie” are part of a large set of frequently used positional verbs (see Kutscher & Schultze-
Berndt, 2007, for a detailed discussion of these terms). They both encode the position of free-standing (unattached) inanimate objects supported by a surface (they can also be used to refer to animate figures). Roughly speaking, *stehen* is used whenever the figure object is perceived to be upright, i.e., the vertical height of the object exceeds its width; *liegen* is used when the figure object is perceived to be in a horizontal position, i.e., the width of the object exceeds its height. In English, *stand* and *lie* are typically reserved for animate figure objects; the typical description of the position of inanimate objects which are supported by a surface involves the copula *be* (Ameka & Levinson, 2007; Newman, 2002).

This difference in the degree of linguistic specificity of object axial position extends to dynamic events. In German, the causative positional verbs *stellen* and *legen* are used to describe the act of putting an object in a vertical or horizontal position respectively. Thus, a dynamic event resulting in the configurations in Figure 1a and b would be encoded in German as in (4) and (5) below, respectively. In English, the corresponding ditransitive verbs *stand*/*lay* are very infrequent, and in most contexts *put*/*place* are used indiscriminately for both horizontal and vertical placements. Thus, (6) would be the natural English counterpart of both (4) and (5).

1. Jemand stellt das Klopapier auf die Waschmaschine.
   someone stands the toilet.paper onto the washing.machine
2. Jemand legt das Klopapier auf die Waschmaschine.
   someone laid the toilet.paper onto the washing.machine
3. Someone puts the toilet paper on the washing machine.

In sum, German and English represent two languages within a broader typology (Ameka & Levinson, 2007) which differ considerably in terms of how richly they encode object axial position in both static and dynamic scenes, with natural English descriptions being less detailed than their German counterparts. Could such cross-linguistic differences lead to differences in the nonlinguistic representation of object axial position? Ameka and Levinson (2007) point out that the semantic parameters which are relevant for positional verbs may show up elsewhere in the grammar of a language, as well as in other systems such as gesture, and conclude: “there is reason to think that the categorization that these [positional] verbs impose has cognitive consequences” (p. 848). However, such a proposal is hard to evaluate because little is known at present about how object position/orientation is cognitively represented or processed (see Gregory & McCloskey, 2010; Valtonen, Dilks & McCloskey, 2008).

1Interestingly, German lacks a corresponding distinction for “taking” verbs (even though such a distinction exists in other languages with contrastive positionals; Ameka & Levinson, 2007). The asymmetry between the putting and taking verbs in German most likely reflects a more general asymmetry in language, whereby goal-of-motion distinctions are more subtle and detailed compared to source-of-motion distinctions (see Lakusta & Landau, 2005; Papafragou, 2010; Regier & Zheng, 2007).
One suggestive piece of evidence shows that, during language comprehension, people seem to reconstruct the implied axial position of objects mentioned in event descriptions. Stanfield and Zwaan (2001) presented subjects with sentences such as *He hammered the nail into the wall* or *He hammered the nail into the floor*. In the first sentence, the nail’s orientation is horizontal; whereas in the second sentence, it is vertical. Each sentence was followed by a line drawing of an object and the participants’ task was to indicate whether the object was mentioned in the sentence. It was found that responses were faster if the line drawing of the object matched the orientation of the object implied in the preceding sentence than when there was a mismatch (see also Pecher, von Dantzig, Zwaan, & Zeelenberg, 2009; Zwaan, Stanfield & Yaxley, 2002). Thus, information about object position, even though not explicitly mentioned, seems to be accessed during language comprehension. Of interest for present purposes is whether, as Ameka and Levinson (2007) hypothesize, explicit encoding of object orientation in dedicated positional verbs within a language could sharpen sensitivity to object orientation in purely visual tasks, especially for objects which do not have a single, canonical orientation (e.g., the figure object in Figure 1a and b). In such cases, axial position cannot be unequivocally reconstructed from a sparse linguistic description of the scene (e.g., *The X is on the Y*).

More recent psycholinguistic evidence shows that L2 users show difficulty in acquiring the subtle semantic distinctions of positional verbs when such distinctions are not present in their native language (Alferink & Gullberg, 2014; Berthele, 2012; Gullberg, 2009, 2011). In one study, when listening to Dutch sentences containing positional placement verbs similar to the German ones in (4) and (5), both Dutch monolingual and German-Dutch bilingual speakers used information from positional verbs to launch anticipatory eye movements toward appropriate objects in a visual scene during incremental sentence processing; however, English-Dutch and French-Dutch speakers, whose L1 lacks a Dutch-like positional system, showed no such anticipatory patterns (Van Bergen & Flecken, 2017). Again, these studies leave it open whether such lexical semantic distinctions affect nonlinguistic scene apprehension and memory.

In the present article, we put this possibility to an empirical test. In a series of four experiments, we investigate the potential impact of the German-English difference in positional vocabulary on German and English speakers’ recognition memory for axial position of objects. Our visual stimuli involve both static positional scenes (Experiments 1, 2 and 3) and dynamic positional events (Experiment 4). We examine a variety of encoding conditions in which participants are either allowed to represent the visual stimuli freely (Experiments 1, 2 and 4), required to use language in addition to visual encoding (Experiments 3 and 4), or prevented from using language during encoding (Experiment 3).

Together, our experiments seek to distinguish among several major possibilities about the cognitive consequences of the linguistic typology of positional on visual memory. According to the conceptual tuning hypothesis,
the cross-linguistic differences should create a consistent cognitive bias, such that German speakers should perform better than English speakers on a recognition memory task involving changes of posture. This effect should persist regardless of whether language is overtly present as spatial scenes are encoded and committed to memory and should also persist when the online use of language is blocked through verbal interference.

Alternatively, if language serves as an online tool for encoding spatial position, participants may differ in recognition memory even when not asked to speak or understand language, with German speakers being more accurate than English speakers (the extent of these linguistic effects is open, see the following sections). However, these effects should be eliminated under a secondary interference task which blocks the online use of language but should persist in a similar interference task which does not block such a use. Furthermore, whatever the effects of language on performance, these effects should be even stronger when positional language is overtly present in the task.

Finally, it remains an open possibility that positional language does not influence memory in either of the ways described above. Recall that recognition does not require the use of language, and recognizing a scene can be accomplished by retrieval processes which give a sense of familiarity and can be accomplished in the absence of conscious awareness of familiarity with the scene (cf. Gennari et al., 2002). If so, native speakers of German and English should perform similarly on the recognition task.

**Experiment 1**

Experiment 1 had two goals. First, it sought to confirm the difference in the use of positional verbs between German and English. Second, it asked whether German and English speakers would differ in the recognition of positional information in a surprise recognition memory task.

**Participants**

We tested 26 native speakers of German and 28 native speakers of English. The former group was recruited at the Ruprecht-Karls-Universität Heidelberg (Germany) and the latter at University of Delaware (USA). An equal number of men and women were included. No participants were paid for their participation, but the American students received some course credit for their participation.

**Stimuli**

Stimuli consisted of 40 pairs of digital color pictures. Each picture depicted two household items in a figure-ground pairing. Each item appeared in only one pair of pictures. A full list of stimuli can be found in Appendix 1.
Sixteen of the 40 pairs were Test items. One member of each pair showed the figure on a ground object in a vertical/upright position (consistent with the German verb *stehen*) and the second showed the same figure on the same ground object but placed in a horizontal/lying position (consistent with the German verb *liegen*; see Figure 1 for an actual example). We avoided figures which resembled animate beings (e.g., dolls or stuffed animals) because both English and German allow the use of positional verbs (*stand, lie*) with animate beings. Figure items were medium-sized objects (e.g., a book, a lipstick, a roll of toilet paper) which did not generally have an expected axial orientation but looked natural in either a standing or a lying position. We confirmed this intuition in a norming study with 10 new participants in which standing and lying configurations of the test items were rated as equally likely to appear in the world as pictured (paired *t*-test, *t*(9) = −1.57, *p* = 0.14; see Appendix 2 for ratings by item). All ratings were obtained from English speakers (the group which does not routinely mark the relevant axial distinction linguistically).

Eight pairs of pictures were Changing Control items. These pictures depicted two objects in a spatial relation which was not consistent with the German positional verbs *stehen* and *liegen* but instead involved attachment (e.g., a clip on a pen), containment (e.g., a banana in a bowl), or piercing (e.g., a knife in an apple). The changes between the two pictures within each pair were either changes of state of the figure (e.g., a banana became a peeled banana) or non-axial positional changes of the figure (e.g., a clip attached to the cap of a pen was moved to the body of that pen).

Finally, 16 pairs of pictures were Non-Changing Control items. The two members of each pair were identical. Each pair of pictures depicted a figure object which was in a support, attachment or containment relation with respect to a ground object.

The 40 pairs of pictures were placed into two lists of 40 pictures each. One picture from each pair was included in List 1 and the other picture was included in List 2, such that each list contained 16 Test items, 8 Changing Control items and 16 Non-Changing Control items. We ensured that in each list, half of the Test items depicted a standing relation and half a lying relation. Within each list, the pictures were randomly ordered. For the Memory condition, these lists were arranged into two different presentation orders by counterbalancing which list was presented during the first (encoding) phase and which list was used during the test (memory) phase. Two further lists were created by reversing the orders of Lists 1 and 2, and two additional counterbalanced presentation orders were created for a total of four.

**Procedure**

Participants were assigned to either the Language task (*n* = 10 German speakers and *n* = 12 English speakers) or the Memory task (*n* = 16 speakers from each group):
Language Task: For the Language task, we only used a shortened version of Lists 1 and 2 which omitted the Non-Changing Control items. The participants viewed one of these Lists and were told that each picture depicted two everyday household objects and instructed to describe each picture with a single, complete sentence in their native language. The participants recorded their description on lined answer sheets. The pictures were presented individually and the participants could advance the display themselves.

Memory Task: The participants were instructed to carefully watch a set of pictures on a computer screen depicting two everyday household objects. They were then presented with the first list of one of the four presentation orders (each order was presented to four native speakers of each language). During this encoding phase, the display automatically advanced to the next picture after two seconds. After the pictures of the first list had been presented, the participants were told that they were to do a memory test: they would now see a second set of pictures in which some of the pictures would be exactly the same as the ones they had just seen and some pictures would be different from those they had just seen. The participants were instructed to verbally provide judgments of “same” or “different” on the picture as a whole. They were not tipped off to or trained on any particular type of change. The second list of pictures was then displayed on the screen for two seconds each. If the participants did not provide an answer within those two seconds, their response was discarded.

Results

Linguistic data: Beginning with the Language task, we found that 90% of the German descriptions for the test items included positional information (exclusively encoded by the verbs stehen/liegen); the remaining descriptions were existential constructions with the copula verb. In all instances, the positional verb was used appropriately for the axial position of the figure object. In contrast, only 32.3% of descriptions of the test items in English contained appropriate positional information about the figure object (primarily encoded by adverbials such as upright but occasionally by the verbs stand/lie). This difference between the two language groups is statistically significant (two-tailed t-test, p < 0.001). It shows that in the absence of clear instructions on which information about a scene to focus on, German speakers are significantly more likely to include information about the axial position of a figure object in their description than English speakers are.

Memory data: Figure 2 shows accuracy in the Memory task. Beginning with test items, a two-tailed t-test comparing the performance of the two language groups revealed no difference between German and English speakers (M_{German} = 67.6%, M_{English} = 71.9%, p = 0.398); both groups made errors
in remembering positional information about 30% of the time. Similar comparisons revealed no differences between language groups in either the Changing Control items ($M_{\text{German}} = 86.7\%$, $M_{\text{English}} = 90.3\%$, $p = 0.464$) or the Non-Changing Controls ($M_{\text{German}} = 91.0\%$, $M_{\text{English}} = 92.0\%$, $p = 0.808$).

**Discussion**

The results of Experiment 1 point to two significant conclusions. First, the descriptions provided in the Language task confirmed the presence of a linguistic difference between German and English, such that German speakers were vastly more likely to include information about the axial positioning of a figure object compared to English speakers. This finding aligns with prior typological observations by Ameka and Levinson (2007) and Kutscher and Schultze-Berndt (2007) about the availability and use of positional terms in the languages under discussion (and is reminiscent of previously reported, significant cross-linguistic differences in the spatial domain which have led to comparisons of cognitive performance across language communities; see Gennari et al., 2002; Papafragou et al., 2002). Second, the results of the Memory task show that German speakers did not perform differently from English speakers in a memory task targeting the axial position of a figure object despite the linguistic difference between the two languages. Both language groups made errors about 30% of the time; therefore, the task was sufficiently difficult for participants to allow differences between the groups to emerge. These results argue against a strong semantic tuning hypothesis,
according to which native language distinctions should constrain recognition memory for the spatial position of figure objects. Furthermore, they give no evidence of online use of positional language as a tool to aid spatial memory.

**Experiment 2**

In Experiment 2, we made two adjustments to the design of Experiment 1 to probe further into potential effects of the linguistic typology of positionals on recognition memory. First, we told participants beforehand that they needed to remember the pictures. We hypothesized that making the participants aware of their participation in a memory experiment would cause them to use all available resources to aid their memory, including perhaps language. Second, we inserted a temporal lag between the pictures during the encoding phase. We reasoned that participants could use this lag to construct a linguistic description and store this description alongside the visual representation of the pictures, even if they were not explicitly instructed to do so. To the extent that these manipulations might reveal differences in recognition memory between English and German speakers, they would provide evidence for flexible and selective use of language to support memory.

**Participants**

We recruited 16 native speakers of German at the Carl-von-Ossietzky Universität Oldenburg (Germany) and 16 native speakers of English at the University of Delaware (USA). None of these participants had participated in Experiment 1. None of the participants had native or near-native command of the other language. An approximately equal number of men and women was included. No participants were paid for their participation, but the American students received some course credit for their participation.

**Stimuli**

We used the same stimuli as in Experiment 1 (Memory task).

**Procedure**

The procedure was exactly the same as for the Memory task in Experiment 1; however, the participants were told beforehand that they were participating in a memory experiment and they needed to remember the first set of pictures for a later recognition test. Furthermore, we inserted 3 seconds of black screen between the pictures presented in the encoding phase.
Results

Results from Experiment 2 are presented in Figure 3. Two-tailed $t$-tests comparing accuracy on Test Items did not reveal any differences between the two language groups ($M_{\text{German}} = 69.1\%, M_{\text{English}} = 73.4\%, p = 0.486$). Similar comparisons showed no English-German differences in either Changing Controls ($M_{\text{German}} = 86.7\%, M_{\text{English}} = 93.8\%, p = 0.091$) or Non-Changing Controls ($M_{\text{German}} = 93.4\%, M_{\text{English}} = 92.9\%, p = 0.886$).

Discussion

As in Experiment 1, the findings of Experiment 2 show that participants did not rely on language for recognizing the spatial orientation of objects in a scene. Even when participants knew that they were participating in a memory experiment and provided with sufficient time to encode the scenes linguistically, there was no difference in memory for figure object orientation between German and English speakers.

Experiment 3

In Experiment 3, we sought to control the linguistic encoding of the axial position of a figure object more tightly than in the previous experiments so as to offer a more direct test of potential online effects of language on spatial memory. Specifically, we conducted three variations of Experiment 2 with distinct groups of participants to clearly manipulate whether access to linguistic labels during spatial encoding was required (Linguistic Completion),

![Figure 3. Memory accuracy in Experiment 2. (Error bars represent standard error of the mean.)](image-url)
blocked (Linguistic Shadowing) or simply possible (and potentially beneficial; Nonlinguistic Shadowing).

The group of participants in the Linguistic Completion condition was asked to linguistically (and explicitly) encode the scenes immediately after viewing them. We expected the content of the linguistic descriptions to reveal a difference between German and English speakers in terms of the frequency of positional information (as in Experiment 1). Of interest was whether this difference would create a corresponding difference in memory accuracy, because here the use of linguistic labels (and hence the emergence of cross-linguistic encoding differences) was guaranteed during encoding of the spatial scenes.

In the Linguistic Shadowing condition, a different group of participants was asked to linguistically shadow a rhythm during the encoding phase of the experiment. This task increased cognitive load compared to Experiment 2. Because this manipulation engaged the language faculty, participants could not use linguistic resources online to enhance/complement their visual memory for the scenes (see Trueswell & Papafragou, 2010).

Finally, in the Nonlinguistic Shadowing condition, participants were asked to shadow a rhythm by tapping on the table during the encoding phase. This task was chosen to be of comparable cognitive load to the previous condition while employing different cognitive resources (see Hermer-Vazquez, Spelke & Katsnelson, 1999). Unlike the previous condition, this task did not tax the language faculty. Thus, it remained possible that the increased cognitive load would cause participants to support their visual memory with strategies such as linguistic encoding of the pictures in preparation for the memory task (see Trueswell & Papafragou, 2010).

**Participants**

We recruited 36 native speakers of German at the Carl-von-Ossietzky Universität Oldenburg and the Gymnasium Nordenham (both in Germany) as well as 36 native speakers of English at the University of Delaware (USA). None of these participants had taken part in Experiments 1 or 2. We included an approximately equal number of men and women. No participants were paid for their participation but the American students received some course credit for their participation.

**Stimuli**

The same stimuli as in Experiment 2 were used (see Linguistic Completion Condition below for a modification).
**Procedure**

Participants were randomly assigned to one of three conditions (12 participants per language per condition):

**Linguistic Completion Condition:** The procedure was that of Experiment 2. However, we modified the stimuli used in Experiment 2 such that each blank screen between pictures during the encoding phase was replaced with a slide displaying a sentence fragment. The fragment described the previously seen picture with the verb and the ground object replaced by blanks. For instance, the sentence displayed after Figure 1a was “The toilet paper ____ on the ____”. for English speakers and “Das Klopapier ___ auf dem/der ____” for German speakers. The participants were instructed to read out loud and complete each sentence during the 3 seconds it was displayed on the screen. (We omitted the ground object in addition to the verb so that participants did not focus solely on the stand/lie relation.) The spoken sentences were tape-recorded and later transcribed and coded.

**Linguistic Shadowing Condition:** The procedure was that of Experiment 2. In addition, the participants wore headphones which continuously played an irregular rhythm. Participants had to repeat the rhythm verbally using the syllable “na” which they had to produce loud enough for the experimenter to hear. During the memory test phase, the headphones were taken off.

**Nonlinguistic Shadowing Condition:** The procedure was exactly as in the Linguistic Shadowing condition. However, instead of repeating the rhythm verbally, the participants had to repeat this rhythm by tapping on the table throughout the encoding phase.

**Results**

**Linguistic data:** We coded the linguistic descriptions in the Linguistic Completion Condition for the inclusion of positional information. As expected, German speakers overwhelmingly offered verbs (*stehen/liegen*) encoding the position of the figure object in Test items (M = 73.3%) but English speakers’ positional descriptions were vanishingly rare (M = 2.8%). This difference is significant (two-tailed *t*-test, *p* < 0.05).

**Memory data:** Results from all conditions are shown in Figure 4a–c. An omnibus ANOVA with the percentage of correct responses on Test Items as the dependent variable and Language Group (German and English) and

---

2These means are lower than those found in the Language Condition of Experiment 1, especially for English speakers. We believe that the reason for this difference lies in the fact that Experiment 1 elicited written descriptions without time pressure; this feature allowed participants to offer a more in-depth description of the pictures (including positional information) as well as emphasize differences between the pictures. In contrast, in Experiment 3, the descriptions were spontaneously produced within 3 seconds. We conclude that, even though the marking of the axial position of a figure object is available in English (as shown in Experiment 1), it is marginal in spontaneous, spoken language (as shown in the present experiment).
Condition (Linguistic Completion, Linguistic Shadowing and Nonlinguistic Shadowing) as factors revealed no main effects and no interactions.

A similar ANOVA with the percentage of correct responses on Changing Control Items as the dependent variable revealed only a main effect of Condition (F(1, 71) = 4.87, p = 0.01): performance on the Linguistic Completion condition was better (M = 92.1%) compared to either the Linguistic (M = 79.6%) or the Nonlinguistic (M = 80.2%) Shadowing conditions (p < 0.05) but the two Shadowing conditions did not differ from each other, p > 0.05). Similarly, an ANOVA for the Non-Changing Controls revealed only a main effect of Condition (F(1, 71) = 12.56, p < 0.0001): performance on the Linguistic Completion condition was better (M = 93.6%) compared to either the Linguistic (M = 78.6%) or the Nonlinguistic (M = 83.8%) Shadowing conditions (the two Shadowing conditions did not differ from each other, p > 0.05). Thus, performance was better for Control but not Test items when participants did not have to perform a secondary task.

Comparisons to Exp.2: Recall that Experiment 2 was identical to the present study but involved no shadowing and linguistic completions. To explore the effects of the present manipulations on the baseline performance in Experiment 2, we repeated the ANOVAs above with Language Group (German and English) and Manipulation (Exp.2, Exp.3/Linguistic Completion, Exp.3/Linguistic Shadowing and Exp.3/Nonlinguistic Shadowing) as factors. For Test Items, the
ANOVA revealed no main effects or interactions. For Changing Control Items, the ANOVA revealed only a main effect of Condition \((F(1, 55) = 5.18, p = 0.0023)\): performance in Experiment 2 did not differ from the Linguistic Completion condition of Experiment 3 \((p > 0.05)\) but was significantly better than both the Linguistic \((F(1, 55) = 10.31, p = 0.0022, M = 90.2\% \text{ vs. } 79.6\%)\), and the Nonlinguistic \((F(1, 55) = 5.21, p = 0.026, M = 90.2\% \text{ vs. } 80.2\%)\) Shadowing conditions of Experiment 3. Similarly, the ANOVA for Non-Changing Controls revealed only an effect of Condition \((F(1, 55) = 33.07, p < 0.0001)\): again performance in Experiment 2 was no different from the Linguistic Completion condition of Experiment 3 \((p > 0.05)\) but was higher than both the Linguistic Shadowing \((F(1, 55) = 33.14, p < 0.0001, M = 93.1\% \text{ vs. } 78.6\%)\) and the Nonlinguistic Shadowing conditions of Experiment 3 \((F(1, 55) = 12.78, p = 0.0007, M = 93.1\% \text{ vs. } 83.80\%)\). Thus, for the Control but not the Test Items, shadowing affected performance negatively compared to a baseline task which did not involve interference; linguistic encoding did not affect performance compared to the baseline for any type of item (for Test Items, performance remained error-prone and for Control Items, it was accurate over 90\% of the time). These effects held regardless of participants’ native language.

**Discussion**

The results of Experiment 3 confirm the conclusion that cross-linguistic differences in encoding figure orientation does not affect recognition memory in speakers of German and English. More specifically, our data show that participants did not rely on linguistic labels to improve their visual memory for positional information even when given a cognitively demanding secondary task which left the option of linguistic encoding of the scenes open to them (Nonlinguistic Shadowing condition). Most remarkably, the lack of differences between language groups persisted even when participants had to verbally encode the visual scenes after encountering them for the first time (Linguistic Completion condition). In fact, cases where participants used language to describe the scenes had identical effects on later recognition of positional details of figure objects compared to cases where language was blocked (Linguistic Shadowing condition) or not invoked during scene encoding (Experiment 2).

**Experiment 4**

In Experiments 1–3, we could not detect any influence of language on memory for object position despite cross-linguistic differences in the encoding of positional information in visual scenes. In Experiment 4, we turned to dynamic positional events, in which an object is placed on another object in a horizontal/vertical orientation. Recall that for these events, just as for their static counterparts, German has a more specific and readily available lexical
distinction (stellen/legen) compared to English which uses a single, undifferentiated term (put/place) regardless of figure orientation. Here, we tested whether this linguistic difference between German and English might have an observable effect on visual memory for dynamic placement events. We hypothesized that, unlike static positional arrangements, dynamic positional events highlight the intentionality of the motion and might, therefore, make positional information (including, potentially, its linguistic encoding) more central to the representation of the event.

Participants

We recruited 24 native speakers of German at the Gymnasium Nordenham, the Athenaeum Stade and the DFKI Projektbüro Berlin (all in Germany) and 24 native speakers of English at the University of Delaware (USA). None of these had participated in any of the previous experiments. Approximately equal numbers of men and women were included. No participants were paid for their participation but the American students received some course credit for their participation.

Stimuli

We created 32 pairs of video clips. In each clip, a hand in a white glove could be seen placing a figure object onto a ground object. Each object occurred in only one pair of clips. Each clip was between 4 and 5 seconds long. A full list of stimuli can be found in Appendix 3.

Of the 32 pairs, eight pairs of clips were Test items. In these pairs, the figure object was placed onto the ground object in a vertical position in one clip and in a horizontal position in the other clip (these actions were consistent with the German verbs stellen and legen, respectively). The test events were dynamic versions inspired by the static scenes used for Experiments 1–3 (e.g., a small tin can was placed on a cereal box in either a vertical or a horizontal position). The ground object stayed the same within a pair of clips. As before, we avoided figure objects which have a natural preference for a vertical or horizontal placement.

Another eight pairs of clips were Changing Control items. In these pairs, the figure object which was placed onto the ground object changed. For instance, in one clip, a toy horse was put on top of a box and in the second clip of that pair, a toy cow was placed on top of the same box. The figure objects in each pair were similar to each other in size and form.

The remaining 16 pairs of clips were Non-Changing Control items. In these pairs, the two clips were identical.

We arranged these 32 pairs of clips into two lists (Lists 1 and 2) with one clip of each pair included in each list. Within each list, members of Test, Changing Control and Non-Changing Control pairs were presented in the
same random order. Half of the Test items in List 1 were consistent with *stellen* and half with *legen* (and similarly for List 2). We then created a first presentation order, such that List 1 was used during the encoding phase and List 2 during the memory test phase. We also reversed the internal order of Lists 1 and 2 to create Lists 3 and 4, respectively. These new lists were placed into a second presentation order, such that List 3 was used during encoding and List 4 during the test phase.

**Procedure**

The participants were randomly assigned to one of the following two conditions (12 participants per language per condition):

*Memory Condition:* The participants were told that they were participating in a memory experiment. They were instructed to carefully watch the video clips because they would be tested on their recognition later on. The participants then watched the first list of clips. Between each clip, they saw 3 seconds of blank screen. After this exposure to the clips, they were then reminded that they needed to provide judgments on whether the clips had changed in any way. The participants were not tipped off on or trained to recognize any particular type of change in the clips. Responses were given by pressing keyboard keys for “same video clip” or “different video clip”. Once a response was made, the display advanced to the next clip.

*Linguistic Completion Condition:* The participants were told that they were participating in a memory experiment. They were instructed to carefully watch the video clips. After each of the clips, they saw a sentence fragment describing the clip in their native language. For instance, after seeing the event of placing a can onto a cereal box, English speakers saw “Someone _____ the can on the _____”. and German speakers saw “Jemand _____ die Dose auf die/den/das _____”. As in Experiment 3, the verb and the ground object were left out of the displayed sentences. The participants had 3 seconds to provide the complete description of the previously seen clip by reading out loud and filling out the fragment. Their sentences were recorded and then later transcribed and coded. During the memory test phase, the participants provided memory judgments on whether the video clips had changed by pressing keyboard keys for “same video clip” or “different video clip”. As in the memory condition, the participants were not instructed to look for any particular type of change but judge the clip holistically. Once a response was made, the display advanced to the next clip.

**Results**

*Linguistic data:* We coded the linguistic descriptions from the Linguistic Completion Condition for inclusion of dynamic positional information. As expected, German speakers encoded the axial position of the figure object 77.9%
of the time and English speakers did so only 1% of the time (one participant using transitive *lay* once). This difference is significant (two-tailed *t*-test, *p* < 0.0005).

**Memory data:** Memory accuracy for the Memory and Linguistic Completion conditions are displayed in Figure 5a and b, respectively. An ANOVA comparing performance on Test Items across Language Groups and Conditions revealed only a marginally significant effect of Condition (*F*(1, 47) = 3.96, *p* = 0.05), such that performance appeared to be worse in the Linguistic Completion (M = 68.2%) than the Memory condition (M = 80.2%).

Similar ANOVAs for the Changing Controls and Non-Changing Controls revealed no significant main effects or interactions: as shown in Figure 5a and b, accuracy was over 90% for these trials across both languages and conditions.

**Discussion**

In Experiment 4, we looked for an influence of language on memory for positional events in speakers of German who linguistically mark the distinction between horizontal and vertical positions of the figure object in dynamic events and speakers of English who do not systematically mark this distinction. As before, we hypothesized that the German speakers would benefit from marking the position linguistically but no such effect of language on memory was found. As with static positional scenes, this lack of influence was observed both when language was not explicitly involved (Memory condition) and when language was explicitly implicated in the task (Linguistic Completion condition).

Somewhat surprisingly, we found a marginally significant effect of condition showing that the participants in the linguistic completion condition performed worse on the memory task for the test items than participants in the memory condition. As this effect is not due to the native language, this might indicate that, in the memory for dynamic positional events, participants pay more attention to the figure and ground objects than the position.

![Figure 5](image-url)

**Figure 5.** Memory accuracy in the (a) memory and (b) linguistic completion condition of Experiment 4. (Error bars represent standard error of the mean.)
of the figure and the linguistic encoding of that information enforces a focus on these (over other information such as the position). However, this finding requires further investigation.

**General discussion**

In this article, we asked whether differences in the way languages encode the axial position of a figure object affect recognition memory. We chose two languages which are quite distinct in terms of the typology of encoding object position, English and German (Ameka & Levinson, 2007). Our data confirmed that, when describing spatial scenes or events, English speakers do not consistently specify the position (standing/lying) of an inanimate figure object but German speakers routinely do so. Nevertheless, these striking cross-linguistic differences in positional encoding have no influence on English and German speakers’ memory for spatial scenes (Experiments 1–3) or spatial events (Experiment 4). These results argue against the hypothesis that the positional typology in the world’s languages should have cognitive consequences (Ameka & Levinson, 2007).

These results bear on broader theoretical proposals about the role of language in cognition. First, our data argue against the hypothesis that habitual exposure to the spatial vocabulary of a language has long-term effects on spatial memory and spatial organization for speakers of that language – even in situations where no language is overtly present (Hespos & Spelke, 2004; Levinson et al., 2002; McDonough et al., 2003). Far from tuning the kinds of conceptual distinctions which speakers can and do entertain, the participants in our sample did not show evidence for either language altering the underlying conceptual representations of spatial position. Our findings are consistent with other experimental findings pointing to similarities in online perception, categorization and memory for space and motion across speakers of different languages (Gennari et al., 2002; Li et al., 2009; Malt et al., 2008; Munnich et al., 2001; Papafragou et al., 2008; Papafragou et al., 2002; Ünal et al., 2016; Ünal & Papafragou, 2016).

Second, our data, even though in principle compatible with the idea that language can intrude online into nonlinguistic tasks (e.g., Trueswell & Papafragou, 2010; Winawer et al., 2007), show that such intrusions are not as pervasive as sometimes assumed (Winawer et al., 2007). Across a variety of contexts which allowed access to linguistic descriptions of static positional scenes or dynamic positional events, participants’ later memory for these scenes or events was unaffected by native language distinctions. Strikingly, even when participants were required to provide linguistic descriptions of the spatial stimuli during encoding, native language distinctions failed to affect recognition memory (Linguistic Completion conditions of Experiments 3 and 4). These findings differ from previous studies reporting that implicit
or explicit labeling of visual stimuli affects performance in some nonlinguis-
tic tasks (see, e.g., Archambault et al., 1999; Billman et al., 2000; Billman &
Krych, 1998; Feist & Gentner, 2007; Lupsan, 2008; McCloskey & Zaragoza,
1985; for effects of language on visual memory). Nevertheless, our findings
are in line with reports which have pointed out boundaries for such linguistic
intrusions (Gennari et al., 2002; Feist & Gentner, 2007; Papafragou et al.,
2002, among others).

What is responsible for the lack of cross-linguistic effects on positional
memory, given divergent findings in the literature? A first possibility is that,
upon closer inspection, the underlying German-English differences in lin-
guistic positional encoding might turn out to be weaker than originally
thought. As several commentators have recently pointed out, even if a
semantic distinction is not formally marked in a language (e.g., through a
clear lexical distinction), speakers may nevertheless make subtle changes in
the way they use the terms they do have in their language in a way which
reveals sensitivity to the relevant semantic contrast (Coventry, Griffiths &
Hamilton, 2014; Johannes, Wilson & Landau, 2016; Landau, Johannes,
Skordos & Papafragou, 2017). If more fine-grained analysis revealed that
the German-English linguistic difference in positional encoding is smaller,
the lack of cognitive differences in positional memory between German and
English speakers would be unsurprising. However, our findings offer no
support for this hypothesis. The linguistic data collected in Experiments 1
and 2 were already coded very liberally for any device which could be used to
indicate positional information (including the use of positional verbs, but
also prepositions, positional expressions such as adverbs and so on). Even
with this liberal coding, there was a large gap in the frequency with which
axial information was encoded in the descriptions used by the English speak-
ers compared to the German speakers (and the descriptions did not seem to
differ in any other relevant respects).

A second possibility is that the particular spatial representations we inves-
tigated are fundamental enough so that they resist effects of language varia-
tion. If these structures constitute strong universals in spatial cognition (as
Aristotle seemed to suggest, see Section 2), then cross-linguistic variation on
this basic pattern might tend not to have any stable or online effects on
nonlinguistic spatial organization. Relatedly, object positional information is
closely tied to perception and effects of language on cognition might be
expected to arise in areas which are more abstract and further removed
from perceptual distinctions (Spelke & Tsivkin, 2001). One issue with these
hypotheses is that the notion of a strong universal is currently underspeci-
fied. Furthermore, the idea that visuo-spatial cognition might pre-empt
effects of cross-linguistic variation is not well-supported (even though the
nature of these effects remains contested; e.g., Hespos & Spelke, 2004;
McDonough et al., 2003).
A third possibility is that positional information about a figure object in a scene or event is a peripheral aspect of spatial representation and unlikely to be boosted by language. Notice that, across our studies, people from both language groups made considerable errors in detecting positional changes of the figure object (and appeared to be less successful with positional changes compared to other types of changes, such as those in the Changing Controls; see Figures 2–5b). In the vision literature, similar errors have been documented in a wider range of cases (see, e.g., Bornstein, 1982; Corballis & Beale, 1983; Sutherland, 1957) and have often been attributed to the fact that orientation differences are irrelevant for determining the identity of an object: as an object is always the same whether it is viewed upright, sideways or in some other orientation, perceptual processes often fail to encode or retain the relevant spatial distinctions. If this line of reasoning is correct, language may be insufficient to remedy failures in encoding, retaining or processing components of orientation representations (even when it makes potentially helpful lexical distinctions, as is the case with German). Nevertheless, two considerations cast doubt on the idea that positional information is inherently irrelevant for perception and memory systems. First, in addition to object identification, vision serves a variety of goals, many of which require orientation information (e.g., acting on objects, or interpreting scenes; Gregory & McCloskey, 2010). Second, the literature has shown that not all orientation differences are equally confusable and thus are not summarily ignored by perception (ibid.; even though those demonstrations do not include the configurations used in our current studies).

A fourth possibility is that language may be particularly useful for some tasks and domains but not others (Potter, 1979). As we have already pointed out, recognizing an event as one previously stored in memory can be carried out automatically by retrieval processes which give a sense of familiarity and can be accomplished in the absence of linguistic encoding of the stimulus. In that sense, recognition memory differs from other cognitive tasks such as similarity detection: the latter involves a more deliberate, open-ended process weighing several factors such as the set of objects under consideration, their features, and the higher-order goal of the task, and thus appears to be more susceptible to linguistic effects (Gennari et al., 2002). Notice that, in several cases where language has been shown to affect visual memory, the visual scenes to be remembered were ambiguous (Carmichael & Hogan, 1932; Feist & Gentner, 2007; cf. also Pinker, 1994) or could be categorized on several levels (Archambault et al., 1999), and thus allowed language to play a disambiguating role. More recently, Regier and Xu (2017) have explicitly argued that effects of language on cognition are more likely to be observed under conditions of uncertainty in the stimuli. Our stimuli included no ambiguity or uncertainty about spatial position (even though they targeted an aspect of...
spatial configuration which was difficult to remember and led to considerable errors, as already discussed).

Furthermore, given that the involvement of language in cognitive tasks is clearly malleable and context-bound, the specific domain which was the present target of investigation might have discouraged such involvement. As several commentators have noted, the representation of space and spatial relations across languages is rather coarse (Landau & Jackendoff, 1993; Talmy, 1985). The same is true for linguistic representations of the axial position of a figure object located within a scene or event. Despite the large number of lexical items representing different positionals in German, a linguistic description cannot be infinitely fine-grained: any sentence representing a standing/lying figure placed on a ground can describe several slightly different configurations or views of the figure. Because of the inherently abstract and underspecified nature of positional terms in language, such terms may have been considered unhelpful for the purpose of representing and storing specific exemplars of scenes and events in the context of our experiments. This might explain why German speakers in our study did not rely on positional verbs to remember and identify visual configurations (cf. also Gennari et al., 2002).

Regardless of the specific explanation which will turn out to be correct, our results provide evidence that the linguistic and visual encodings of spatial position belong to different levels of representation and are potentially independent of each other. Furthermore, even though language could be an additional route for encoding positional information online, spatial memory for object position is unaffected by language in many ordinary contexts. It is, of course, an open possibility that different tasks might uncover subtle, more transient effects which are unobservable by the current behavioral methods (see Thierry et al., 2009; Flecken et al., 2015). Be that as it may, the present study contributes boundary conditions on the potency of language effects on spatial memory.

Acknowledgments

Large parts of this research were completed while the first author was at the University of Delaware. We would like to thank the members of the Language and Cognition Lab at the University of Delaware and the audience at the 32nd Annual Meeting of the Cognitive Science Society for useful comments on earlier versions of this work. Furthermore, we would like to thank Ines Becker, Valia Kordoni and Kathrin Wagner for making possible the collection of the German data.

Funding

This research was partly supported by NIH grant 5R01HD55498-2 to A.P.
References


## Appendix 1. Stimuli for Experiment 1

<table>
<thead>
<tr>
<th>Item</th>
<th>Figure</th>
<th>Ground</th>
<th>Type of Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test</td>
<td>small vase</td>
<td>big pillow</td>
<td>Positional</td>
</tr>
<tr>
<td>Test</td>
<td>book</td>
<td>dining room chair</td>
<td>Positional</td>
</tr>
<tr>
<td>Test</td>
<td>backpack</td>
<td>staircase</td>
<td>Positional</td>
</tr>
<tr>
<td>Test</td>
<td>pineapple</td>
<td>windowsill</td>
<td>Positional</td>
</tr>
<tr>
<td>Test</td>
<td>roll of toilet paper</td>
<td>washing machine</td>
<td>Positional</td>
</tr>
<tr>
<td>Test</td>
<td>lipstick</td>
<td>dollar bill</td>
<td>Positional</td>
</tr>
<tr>
<td>Test</td>
<td>boot</td>
<td>frying pan</td>
<td>Positional</td>
</tr>
<tr>
<td>Test</td>
<td>CD case</td>
<td>notebook</td>
<td>Positional</td>
</tr>
<tr>
<td>Test</td>
<td>roll of paper towels</td>
<td>microwave</td>
<td>Positional</td>
</tr>
<tr>
<td>Test</td>
<td>cell phone</td>
<td>binder</td>
<td>Positional</td>
</tr>
<tr>
<td>Test</td>
<td>birthday card</td>
<td>lunch tray</td>
<td>Positional</td>
</tr>
<tr>
<td>Test</td>
<td>wicker basket</td>
<td>sidewalk</td>
<td>Positional</td>
</tr>
<tr>
<td>Test</td>
<td>cigarette lighter</td>
<td>mousepad</td>
<td>Positional</td>
</tr>
<tr>
<td>Test</td>
<td>wallet</td>
<td>newspaper</td>
<td>Positional</td>
</tr>
<tr>
<td>Test</td>
<td>walkman</td>
<td>glove</td>
<td>Positional</td>
</tr>
<tr>
<td>Test</td>
<td>ipod</td>
<td>lace doily</td>
<td>Positional</td>
</tr>
<tr>
<td>Changing Control</td>
<td>knife</td>
<td>apple</td>
<td>placement of knife</td>
</tr>
<tr>
<td>Changing Control</td>
<td>scissors</td>
<td>wine bottle</td>
<td>scissors open/close</td>
</tr>
<tr>
<td>Changing Control</td>
<td>key</td>
<td>keyring</td>
<td>key on/off keyring</td>
</tr>
<tr>
<td>Changing Control</td>
<td>banana</td>
<td>bowl</td>
<td>banana (not) peeled</td>
</tr>
<tr>
<td>Changing Control</td>
<td>guitar</td>
<td>guitar case</td>
<td>guitar in/on guitar case</td>
</tr>
<tr>
<td>Changing Control</td>
<td>chopsticks</td>
<td>glass</td>
<td>chopsticks whole/broken</td>
</tr>
<tr>
<td>Changing Control</td>
<td>clip</td>
<td>pen</td>
<td>placement of clip on pen</td>
</tr>
<tr>
<td>Changing Control</td>
<td>telephone receiver</td>
<td>phone set</td>
<td>receiver on/off phone</td>
</tr>
<tr>
<td>Non-Changing Control</td>
<td>post-it note</td>
<td>laptop screen</td>
<td></td>
</tr>
<tr>
<td>Non-Changing Control</td>
<td>glasses</td>
<td>table</td>
<td></td>
</tr>
<tr>
<td>Non-Changing Control</td>
<td>umbrella</td>
<td>door</td>
<td></td>
</tr>
<tr>
<td>Non-Changing Control</td>
<td>necklace</td>
<td>teddy bear</td>
<td></td>
</tr>
<tr>
<td>Non-Changing Control</td>
<td>teabag</td>
<td>wine glass</td>
<td></td>
</tr>
<tr>
<td>Non-Changing Control</td>
<td>soap</td>
<td>rubber mat</td>
<td></td>
</tr>
<tr>
<td>Non-Changing Control</td>
<td>hanger</td>
<td>doorknob</td>
<td></td>
</tr>
<tr>
<td>Non-Changing Control</td>
<td>pinecone</td>
<td>leaves</td>
<td></td>
</tr>
<tr>
<td>Non-Changing Control</td>
<td>teacup</td>
<td>bowl</td>
<td></td>
</tr>
<tr>
<td>Non-Changing Control</td>
<td>ribbon</td>
<td>sneaker</td>
<td></td>
</tr>
<tr>
<td>Non-Changing Control</td>
<td>hairclip</td>
<td>cabinet door</td>
<td></td>
</tr>
<tr>
<td>Non-Changing Control</td>
<td>button</td>
<td>plastic bag</td>
<td></td>
</tr>
<tr>
<td>Non-Changing Control</td>
<td>matches</td>
<td>deck of cards</td>
<td></td>
</tr>
<tr>
<td>Non-Changing Control</td>
<td>fortune cookie</td>
<td>tupperware box</td>
<td></td>
</tr>
<tr>
<td>Non-Changing Control</td>
<td>clothes pin</td>
<td>sheet of paper</td>
<td></td>
</tr>
<tr>
<td>Non-Changing Control</td>
<td>ring</td>
<td>napkin</td>
<td></td>
</tr>
</tbody>
</table>
Appendix 2. Typicality Ratings for Test Items (1 = very unlikely to be found together as shown, 7 = very likely to be found together as shown)

<table>
<thead>
<tr>
<th>Figure</th>
<th>Ground</th>
<th>Lying</th>
<th>Standing</th>
</tr>
</thead>
<tbody>
<tr>
<td>small vase</td>
<td>big pillow</td>
<td>1.4</td>
<td>1.8</td>
</tr>
<tr>
<td>book</td>
<td>dining room chair</td>
<td>3.2</td>
<td>4.8</td>
</tr>
<tr>
<td>backpack</td>
<td>staircase</td>
<td>5.3</td>
<td>4.8</td>
</tr>
<tr>
<td>pineapple</td>
<td>windowsill</td>
<td>2.8</td>
<td>2.0</td>
</tr>
<tr>
<td>roll of toilet paper</td>
<td>washing machine</td>
<td>2.8</td>
<td>2.8</td>
</tr>
<tr>
<td>lipstick</td>
<td>dollar bill</td>
<td>1.9</td>
<td>2.9</td>
</tr>
<tr>
<td>boot</td>
<td>frying pan</td>
<td>1.1</td>
<td>1.1</td>
</tr>
<tr>
<td>CD case</td>
<td>notebook</td>
<td>2.6</td>
<td>3.4</td>
</tr>
<tr>
<td>roll of paper towels</td>
<td>microwave</td>
<td>5.3</td>
<td>4.8</td>
</tr>
<tr>
<td>cell phone</td>
<td>binder</td>
<td>3.3</td>
<td>4.9</td>
</tr>
<tr>
<td>birthday card</td>
<td>lunch tray</td>
<td>2.3</td>
<td>2.2</td>
</tr>
<tr>
<td>wicker basket</td>
<td>sidewalk</td>
<td>3.4</td>
<td>3.2</td>
</tr>
<tr>
<td>cigarette lighter</td>
<td>mousepad</td>
<td>1.9</td>
<td>1.9</td>
</tr>
<tr>
<td>wallet</td>
<td>newspaper</td>
<td>3.2</td>
<td>3.8</td>
</tr>
<tr>
<td>walkman</td>
<td>glove</td>
<td>2.8</td>
<td>3.1</td>
</tr>
<tr>
<td>ipod</td>
<td>lace doily</td>
<td>2.8</td>
<td>4.2</td>
</tr>
<tr>
<td><strong>Mean</strong></td>
<td></td>
<td><strong>2.88</strong></td>
<td><strong>3.23</strong></td>
</tr>
</tbody>
</table>

Appendix 3. Stimuli for Experiment 4

<table>
<thead>
<tr>
<th>Item</th>
<th>Figure</th>
<th>Ground</th>
<th>Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test</td>
<td>alarm clock</td>
<td>newspaper</td>
<td>positional</td>
</tr>
<tr>
<td>Test</td>
<td>backpack</td>
<td>stairs</td>
<td>positional</td>
</tr>
<tr>
<td>Test</td>
<td>can</td>
<td>box of cornflakes</td>
<td>positional</td>
</tr>
<tr>
<td>Test</td>
<td>cassette</td>
<td>microsoft box</td>
<td>positional</td>
</tr>
<tr>
<td>Test</td>
<td>cigarette lighter</td>
<td>mouse pad</td>
<td>positional</td>
</tr>
<tr>
<td>Test</td>
<td>lipstick</td>
<td>book</td>
<td>positional</td>
</tr>
<tr>
<td>Test</td>
<td>scotch tape dispenser</td>
<td>laptop bag</td>
<td>positional</td>
</tr>
<tr>
<td>Test</td>
<td>wallet</td>
<td>folder</td>
<td>positional</td>
</tr>
<tr>
<td>Changing Control</td>
<td>brush</td>
<td>jewelery box</td>
<td>comb</td>
</tr>
<tr>
<td>Changing Control</td>
<td>CG bag</td>
<td>big camera bag</td>
<td>headphones</td>
</tr>
<tr>
<td>Changing Control</td>
<td>lipgloss</td>
<td>small plate</td>
<td>eye shadow</td>
</tr>
<tr>
<td>Changing Control</td>
<td>loud speaker</td>
<td>projector</td>
<td>wrapped gift</td>
</tr>
<tr>
<td>Changing Control</td>
<td>measuring tape</td>
<td>toy drawers</td>
<td>fake ivy</td>
</tr>
<tr>
<td>Changing Control</td>
<td>small fake apple</td>
<td>small flower pot</td>
<td>black cotton ball</td>
</tr>
<tr>
<td>Changing Control</td>
<td>toy cow</td>
<td>cube</td>
<td>toy horse</td>
</tr>
<tr>
<td>Changing Control</td>
<td>watch</td>
<td>playmobile box</td>
<td>hairclip</td>
</tr>
<tr>
<td>Non-Changing Control</td>
<td>bar of soap</td>
<td>carry-out container</td>
<td></td>
</tr>
<tr>
<td>Non-Changing Control</td>
<td>bottle</td>
<td>napkin</td>
<td></td>
</tr>
<tr>
<td>Non-Changing Control</td>
<td>button</td>
<td>wooden lid</td>
<td></td>
</tr>
<tr>
<td>Non-Changing Control</td>
<td>cup</td>
<td>shelf</td>
<td></td>
</tr>
<tr>
<td>Non-Changing Control</td>
<td>glasses</td>
<td>printer</td>
<td></td>
</tr>
<tr>
<td>Non-Changing Control</td>
<td>key</td>
<td>deck of cards</td>
<td></td>
</tr>
<tr>
<td>Non-Changing Control</td>
<td>ornament</td>
<td>lamp</td>
<td></td>
</tr>
<tr>
<td>Non-Changing Control</td>
<td>ring</td>
<td>linen</td>
<td></td>
</tr>
<tr>
<td>Non-Changing Control</td>
<td>Scissors</td>
<td>picture frame</td>
<td></td>
</tr>
<tr>
<td>Non-Changing Control</td>
<td>small box</td>
<td>wooden box</td>
<td></td>
</tr>
<tr>
<td>Non-Changing Control</td>
<td>Spoon</td>
<td>cookie box</td>
<td></td>
</tr>
<tr>
<td>Non-Changing Control</td>
<td>teabag</td>
<td>cup</td>
<td></td>
</tr>
<tr>
<td>Non-Changing Control</td>
<td>toy recorder (plastic)</td>
<td>toy table (plastic)</td>
<td></td>
</tr>
<tr>
<td>Non-Changing Control</td>
<td>USB stick</td>
<td>envelop</td>
<td></td>
</tr>
<tr>
<td>Non-Changing Control</td>
<td>wooden building block</td>
<td>frying pan</td>
<td></td>
</tr>
<tr>
<td>Non-Changing Control</td>
<td>wooden flower</td>
<td>plastic bag</td>
<td></td>
</tr>
</tbody>
</table>