Cognition 194 (2020) 104023

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/cognit

Original Articles

Whereof one cannot speak: How language and capture of visual attention interact

Florian Goller[a,*], Soonja Choi[b,c], Upyong Hong[d], Ulrich Ansorge[a]

[a] Faculty of Psychology, University of Vienna, Austria
[b] Department of Linguistics and Asian/Middle-Eastern Languages, San Diego State University, United States
[c] Faculty of Philological and Cultural Studies, University of Vienna, Austria
[d] Department of Media and Communication, Konkuk University, Seoul, South Korea

ARTICLE INFO

Keywords:
Attention capture
Psycholinguistics
Cross-linguistic differences
Language and cognition
Spatial semantics

ABSTRACT

Our research addresses the important question whether language influences cognition by studying crosslinguistic differences in *nonlinguistic* visual search tasks. We investigated whether capture of visual attention is mediated by characteristics corresponding to concepts that are differently expressed across different languages. Korean grammatically distinguishes between tight- (*kkita*) and loose-fit (*nehta*) containment whereas German collapses them into a single semantic category (*in*). Although linguistic processing was neither instructed nor necessary to perform the visual search task, we found that Korean speakers showed attention capture by non-instructed but target-coincident (Experiment 1) or distractor-coincident (Experiments 4 and 5) spatial fitness of the stimuli, whereas German speakers were not sensitive to it. As the tight- versus loose-fit distinction is grammaticalized only in the Korean but not the German language, our results demonstrate that language influences which visual features capture attention even in non-linguistic tasks that do *not* require paying attention to these features. In separate control experiments (Experiments 2 and 3), we ruled out cultural or general cognitive group differences between Korean and German speaking participants as alternative explanations. We outline the mechanisms underlying these crosslinguistic differences in nonlinguistic visual search behaviors. This is the first study showing that linguistic spatial relational concepts held in long-term memory can affect attention capture in visual search tasks.

1. Introduction

Spatial cognition and spatial language have a deep influence on human life. Humans do not only use spatial language to determine the relation between objects (e.g., the bicycle is in front of the house), they also use spatial language to express temporal relations (e.g., Friday is *before* Saturday, see Boroditsky, 2000, 2001) or affective concepts (e.g., *down* as synonymous to a feeling of negative affect, see Lakoff & Johnson, 2008). Since languages differ in the way they encode and express spatial relations (Bowerman & Choi, 2001; Levinson, Meira, & the Language and Cognition Group, 2003), human's native language could shape spatial cognition and perception in a fundamental way (Athanasopoulos et al., 2015; Goller, Lee, Ansorge, & Choi, 2017; Whorf, 1964). Because of the broad implications of spatial language such as its impact on cognition (for a review see Majid, Bowerman, Kita, Haun, & Levinson, 2004), this question has interested researchers for decades.

In the present study, we address this question by studying German and Korean speakers in nonlinguistic visual search tasks. In a series of experiments, we demonstrate that between the two language groups, the types of visual features that capture attention differ in ways that correspond to the spatial semantics of their native language. In arguing for language influencing visual cognition, we also outline the mechanisms by which the influence takes place. Specifically, we show that language-specific semantics bias speakers' visual attention through linguistic long-term memory representation. But first, we provide brief backgrounds on the theories about the relationship between language and spatial cognition and about possible mechanisms. We also explain the basic theoretical concept of attention capture as a tool that can illuminate the language and cognition relation.

1.1. Relationship between language and cognition

Although an influence of language on cognition is plausible at least

* Corresponding author at: Faculty of Psychology, University of Vienna, Liebiggasse 5, 1010 Vienna, Austria.
E-mail address: florian.goller@univie.ac.at (F. Goller).

https://doi.org/10.1016/j.cognition.2019.104023
Received 7 August 2018; Received in revised form 2 July 2019; Accepted 4 July 2019

where active verbal processing is required for performance (e.g., to support short-term memory), some studies have argued that language does *not* principally influence spatial cognition and perception in nonlinguistic tasks (Li & Gleitman, 2002; Munnich, Landau, & Dosher, 2001). For example, Munnich et al. (2001) showed that while English and Korean differ in the way they linguistically categorize contact/support with respect to the reference object (*on* vs. *over* in English; *wui* for both *on* and *over* in Korean), speakers of the two languages do not differ in nonverbal memory tasks. These results, along other similar findings that show no significant language effects on nonverbal behaviors (Li & Gleitman, 2002; Malt, Sloman, Gennari, Shi, & Wang, 1999; Papafragou, Massey, & Gleitman, 2002), have led to the claim that cognition is independent of language and is not altered, in a fundamental way, by language-specific grammar (Gleitman & Papafragou, 2013; Landau, Dessalegn, & Goldberg, 2010).

However, other studies have demonstrated that language-specific semantics do influence spatial cognition and perception (e.g., Athanasopoulos & Bylund, 2013; Goller et al., 2017; Özyürek & Kita, 1999; Soroli, 2012). For example, Soroli (2012; Soroli & Hickmann, 2010) found that, while watching videos of motion events (e.g., someone running into a house), speakers of satellite-framed or manner-framed languages (in which manner is expressed in the verb, e.g., English: *He ran into the house*) allocate more attention to the visual manner of motion and also judge the similarity of videos based on the manner. On the other hand, speakers of verb-framed or path languages (in which path is expressed in the verb, e.g., French: *Il est entré dans la maison en courant* 'He entered the house running') allocate more attention to visual path (Flecken, von Stutterheim, & Carroll, 2014). Goller et al. (2017) have also reported language-specificity in a nonverbal task concerning spatial categorization of containment/support and tight-/loose-fit in German and Korean speakers (cf. Section 1.4).

1.2. If language influences cognition, how does it happen? Long-term memory versus working-memory influences

Even if the latter studies summarized in Section 1.1 do demonstrate that language differences influence cognition, it is unclear how the influence comes about. Do linguistic long-term memory representations bias how visual attention is allocated in these situations (cf. Lupyan, 2012; Whorf, 1964)? Or are these influences due to currently activated linguistic representations in working memory that the participants use to solve their task at hand (cf. Huettig, Olivers, & Hartsuiker, 2011; Knoeferle & Crocker, 2006; Soto & Humphreys, 2007)?

According to Whorf's concept of linguistic relativity, language-specific semantic representations in long-term memory could affect performance in nonlinguistic tasks, for example, by biasing perceptual discrimination along the boundaries of linguistically defined categories in an enduring way (Gilbert, Regier, Kay, & Ivry, 2006; Winawer et al., 2007). In Lupyan's model, such biases, or *warped perception* as he calls it, would be due to reciprocally connected linguistic and non-linguistic memory representations. Through training, humans first build up long-term memory associations between category names or labels and perceptual category exemplars. As a consequence, perceptual processing activity can (but does not have to) spread along the connections to also activate associated linguistic representations, and linguistic representations could activate related perceptual processes. Due to this spread of activation in long-term memory, linguistic representations could affect nonlinguistic task performance.

In contrast, researchers such as Huettig et al. (2011) believe that only those linguistic representations that are currently relevant and, therefore, actively held in working memory account almost entirely for influences of language on cognition/perception. In this view, even if a long-term memory association connects perceptual and linguistic representations, as long as a task does not require or encourage the use of language representations, these representations would not be activated in working memory and, hence, would be without effect in a visual task. This conclusion is based on the observation that language and cognition/perception typically interact in tasks that require (or at least encourage) the use of linguistic representations. However, tasks that do *not* require the active use of linguistic representations do not show the same interactions. For example, if participants have to keep a color label (e.g., *red*) in working memory for later recall, presenting a visual distractor (e.g., a red disk) with a color similar to this working memory representation in the retention interval captures attention and interferes with visual search for a target (Soto & Humphreys, 2007). Yet, the same interference is not observed if only the color label (e.g., the word 'red') is presented to the participants and the participants do not have to keep the color label in working memory (Soto & Humphreys, 2007).

1.3. Visual attention as a tool to study the relationship between language and cognition

In the present study, we studied long-term memory influences of linguistic representations on cognition/perception by looking at visual attention. Here, visual attention denotes the selection of information from the visual surroundings. In the following, we will first review some of the findings from this research area.

To start with, for a long time, studies of human vision have been a testing ground for interactions between language and cognition/perception. For example, verbal information can influence what stimuli/objects a participant looks at when observing a visual scene (Allopenna, Magnuson, & Tanenhaus, 1998; Cooper, 1974; Huettig & Altmann, 2005; Spivey, Tanenhaus, Eberhard, & Sedivy, 2002; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). Take the instruction, *put the apple on the towel in the box*, as an example. In this situation, participants would first look at the towel and then at the box. Likewise, in visual search tasks, where in each trial a to-be-searched-for target is presented together with several distractors, the interplay between language and visual attention is well established (e.g., Lupyan & Spivey, 2008, 2010). For example, Spivey, Tyler, Eberhard, and Tanenhaus (2001) found more efficient search if auditory verbal information about the target was presented simultaneously with the search display as compared to before the search display. This indicates that language that humans hear immediately affects the way humans attend to objects in a visual scene, providing further evidence for rapid interactions between linguistic processing and visual attention. Along similar lines, Walenchok, Hout, and Goldinger (2016) found that visual distractors phonologically similar to the verbally instructed targets (e.g., a picture of a beaver when participants look for a beaker) interfere with search performance (see also Meyer, Belke, Telling, & Humphreys, 2007).

Critically for our discussion of the underlying mechanisms, in the research so far, most of such interactions between language and visual attention were demonstrated in tasks that either explicitly required or at least invited the usage of the linguistic concepts in question[1] (Baier & Ansorge, 2019; Gleitman & Papafragou, 2013; Li & Gleitman, 2002; Soto & Humphreys, 2007). In these studies, the language effects on cognition, such as the capture of attention by a stimulus resembling a word's meaning, could be based on the currently activated linguistic representations in working memory (cf. Huettig et al., 2011; Landau et al., 2010). Therefore, more compelling evidence for influences of language on nonlinguistic cognition would come from more indirect tasks: Long-term linguistic representations that are having an effect on the performance in *nonlinguistic* cognitive tasks, in particular in those tasks that do not *require* the activation of the linguistic representations in question to complete the tasks.

Only a few studies have done this (e.g., Gilbert et al., 2006; Winawer et al., 2007), although each can be criticized for one or

[1] The linguistic representation in question is that representation which accounts for the interaction between language and performance in a cognitive task.

another reason. As an illustration, take the elegant study by Thierry, Athanasopoulos, Wiggett, Dering, and Kuipers (2009), who wanted to demonstrate linguistic influence in a nonlinguistic visual task. These authors presented a sequence of visual stimuli of the same green or blue color. This sequence was only occasionally interrupted by an oddball stimulus in a different (deviant) shade of green or blue color. Thierry et al. measured the visual mismatch negativity (vMMN), an event-related potential (ERP) triggered by the oddball shade, even when the participant was not asked to pay attention to the shade feature. Specifically, participants had to press a button to an occasional square-shaped stimulus, but oddball disks and their colors were task-irrelevant. Native Greek and English speakers participated in this study because only Greek but not English has distinct mono-morphemic words for light and dark blue. The results showed that the vMMN discriminated more between different shades of blue than different shades of green in Greek speakers but not in English-speakers.

Although the study by Thierry et al. (2009) can be seen as evidence that linguistic representations could influence (pre-attentive) visual processing in a nonlinguistic task, their results should be taken with a grain of salt. For example, there were also other ERP differences both before and following the vMMN that did not fit the picture. Maybe most critically, at the time of the P300 (an ERP component indicating stimulus evaluation), the Greek speakers showed a stronger discrimination between greens than between blues, whereas the English speakers now showed a stronger discrimination between blues than between greens (see Fig. 2B of Thierry et al., 2009). But why should Greek speakers show a smaller P300 difference between light and dark blue than between light and dark green, even though these participants would discriminate the different blue colors but not the different green colors by different mono-morphemic words? Also, why should English speakers show a stronger P300 difference for different blues than greens at all, if the English language does neither discriminate between different blues nor between different greens by different mono-morphemic words? Given these open questions, sceptics may still doubt the existence of linguistic influences on performance in non-linguistic tasks.

1.4. The present study

In the present study, we address the question of influence of linguistic representations on performance in a nonlinguistic visual-attention task with a more clear-cut methodology. We compared the results of Korean speakers with those of German speakers, as German lacks the same linguistic characteristics on which our study focuses.

To be precise, we investigated how variants of spatial 3D object fits captured attention of Korean and German speakers in a nonlinguistic visual-search task (see below), the motivation being that these languages classify spatial 3D-fits in a fundamentally different way. As is now well known (Bowerman & Choi, 2003; Yun & Choi, 2018), Korean grammatically specifies the degree of 3D object fits (tight vs. loose) with spatial verbs: When such objects fit tightly (e.g., a key in a lock), Korean speakers use the verb *kkita* ('fit tightly/interlock'), but when the objects fit loosely (e.g., a teabag in a cup), they use other verbs (e.g., *nehta,* meaning *put.in.loosely*, *nohta,* meaning *put.on.loosely*; Bowerman & Choi, 2003; Yun & Choi, 2018). The distinction of spatial fitness is grammaticalized and pervasive in Korean, as in Korean a spatial verb is necessary for expressing a spatial relation[2]. In contrast, German (similar to English) encodes spatial relations grammatically and obligatorily with prepositions and does so based on whether an object is contained in (*in*) or supported by (*auf* meaning *on*) another object, *regardless* of degree of fit. More specifically, while Korean speakers routinely distinguish between tight-fit (*kkita*) and loose-fit (*nehta*) containment, German speakers collapse them non-differentially into a single semantic category of containment (*in*). Of course, German (and English) speakers can specify tight- or loose-fit using adverbs (*fest/tight*; *lose/loose*). However, such specification occurs only occasionally (Yun & Choi, 2018) and does not contribute to the systematic categorization of spatial relations. Thus, only in Korean, but not in German, tight and loose 3D-fits are systematically contrasted. Given the fundamental nature of spatial cognition and language and early influence of language on spatial categorization (Casasola, 2005; Casasola, Cohen, & Chiarello, 2003; Choi, 2006; McDonough, Choi, & Mandler, 2003), we hypothesized that effects of these linguistic *long-term memory* differences between German and Korean adults affect not only spatial tasks that directly involve those terms but also extend to nonlinguistic tasks that do not explicitly require or induce the linguistic expressions to solve them: Here, a visual search for a specific color.

In our visual-search task, these linguistic concepts were completely task irrelevant, meaning that we studied a long-term memory influence of language on attention (see Lupyan, 2012). For our test, we used visual attention capture. The visual capture of attention denotes the quick allocation of attention to visual stimuli. Stimuli resembling a memory representation can capture attention (Folk, Remington, & Johnston, 1992; Theeuwes, 2013), such as when one quickly looks at a banana because one was thinking of a banana. However, salient stimuli that stand out by visual features can also capture attention. For example, one might quickly look at a banana in a bowl of oranges, as the banana stands out as a color- and shape-singleton (singleton meaning that the banana is unique in color or shape among the oranges) (cf. Theeuwes, 1992, 2013).

We assessed the differential capture of visual attention by tight versus loose 3D object fits in Korean and German speakers in three color-search experiments (Experiments 1, 4, and 5). In all three experiments, participants had to search for target colors. Object fits were task-irrelevant and were not mentioned. All three experiments showed capture by 3D object fits among Korean speaking participants but not among German speakers. Furthermore, in two control experiments, we demonstrate the following: First, Korean speakers do not show capture by task-irrelevant 2D object fits, which the Korean language does not grammatically discriminate (Experiment 2, see also Experiment 5). Second, German speakers show capture by task-*relevant* 3D object fits, once German speakers are explicitly instructed to search for these fits as target-defining characteristics (Experiment 3), ruling out language-independent cognitive or cultural group differences between Korean and German speaking participants as better explanations. We, thus, conclude that it is the language-specific semantics that influences nonlinguistic visual search performance.

2. Experiment 1

In Experiment 1, participants searched for a color-defined target (e.g., red) at one out of four possible positions. Prior to each target display, a non-predictive positional cue (i.e., a cue that across trials does not predict the target's most likely position) was shown, either at target position or at a different position. If the cue captures attention, target search is facilitated under same-position conditions, as attention would already be at the target location by the time of target onset (Posner, 1980). In contrast, different-position cues require participants to shift their attention from the cue position to the target position when the target display comes on. The resulting cueing effect (longer reaction times [RTs] in the different- than in the same-position condition when cue-target interval does not exceed about 300 ms) indicates attention capture by the cue.

With non-predictive positional cues, such cueing effects often depend on cue-target similarity: They are stronger or selectively present

[2] Korean also has locative nouns such as *an* (inside) or *wui* (on/above) to express spatial relations. However, these are optional element. Studies by Choi and her colleagues (Yun & Choi, 2018) report 12%–24% of locative nouns in spatial descriptions by Korean speakers. Moreover, the use of the locative nouns is concentrated in the description of loose-fit relations (e.g., loose containment, loose support).

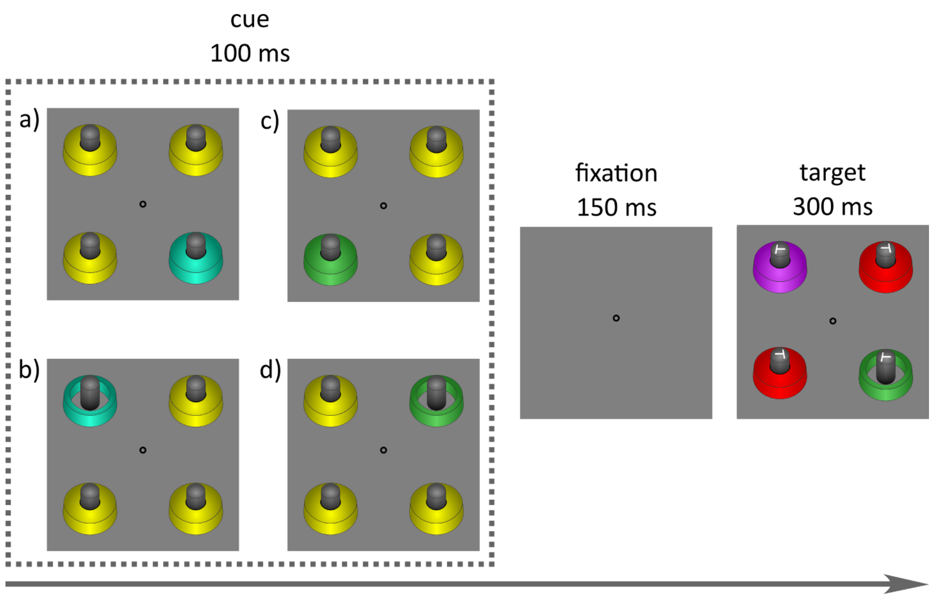

Fig. 1. Procedure of Experiment 1. Depicted are example trials with a green target (in the rightmost square). Four different cue conditions of Experiment 1 are illustrated on the left side (see panel labelled with *cue*): On the left side (a and b), cues with a target-dissimilar color are shown, on the right side (c and d) cues with a target-similar color. The upper row of the cue displays (a and c) shows target-dissimilar fit cues, the lower row (b and d) target-similar fit cues. The arrow depicts the flow of time. The stimuli are not drawn to scale.

with target-similar cues (e.g., cues with the same color as the target, red in this example) and weaker or absent with target-dissimilar cues (e.g., cues with a different color than the target, see Folk et al., 1992).

In Experiment 1, we used this differential cueing effect to test the sensitivity of Korean and German speakers for attention-capture by graphical depictions of task-irrelevant tight-fit and loose-fit 3D stimuli: a 2D illustration of a 3D piston and a 3D cylinder that surrounds it; see target display of Fig. 1. Participants were instructed to search for a color-defined target (e.g., green, see Fig. 1, right panel labelled *target*). Per each trial, one of two kinds of salient color-singleton cues preceded the target displays. We expected a (larger) cueing effect by singleton cues with a target-similar color (cue displays c and d of Fig. 1) but no (or weaker) cueing effects by singleton cues with a target-dissimilar color (cue displays a and b of Fig. 1).

Crucially, we also integrated different tight- and loose-fits (Hespos & Spelke, 2004) into the design, but this was not mentioned at all during the entire experiment: For half the participants, the color-defined target was consistently a single 3D loose-fit singleton presented together with three (= non-singleton) tight-fit distractors at other positions (cf. target display in Fig. 1), and for half the participants, this was the reverse–that is, a singleton tight-fit 3D target with three loose-fit 3D distractors. We hypothesized that due to the grammaticalized distinction, Korean but not German speakers would pick up on this uninstructed but target-coincident characteristic. To measure this, in the cue-display, the singleton cues could have either the same spatial fit as the target (cue displays b and d of Fig. 1) or the alternative one (cue displays a and c of Fig. 1). Therefore, the cueing effect as a reflection of attention capture could be measured for singleton cues similar to the target in two different features: cue color and spatial tight versus loose fit of the cue. If the language-specific semantic distinction between tight- and loose-fit in Korean (Bowerman & Choi, 2003; Yun & Choi, 2018) has a long-term memory impact on the performance in a nonlinguistic visual attention task, Korean speakers' attention should be captured by the tight/loose feature of the cues, depending on which of the two is coincident with the target. This should be the case even where such characteristics are surreptitiously embedded and hidden in a color search task as was done here. More specifically, among Korean speakers, cues with the target-coincident loose or tight-fit (cue displays b and d of Fig. 1) should capture attention more efficiently than cues with a target-dissimilar fit (cue displays a and c of Fig. 1): When the targets are loose-fit, loose-fit cues should lead to a stronger cueing effect than tight-fit cues. Conversely, for targets that are tight-fit, tight-fit cues should lead to a stronger cueing effect than loose-fit cues. In contrast, German speakers, who do not distinguish between tight- and loose-fit in a grammaticalized way but rather collapse them together as tokens of containment (*in*), should be less sensitive to the variants of the spatial loose and tight-fits of the cues and, thus, not show a differential cueing effect based on cue-target loose- versus tight-fit similarity.

As task-specific linguistic representations might impact visual processing in the left but not in the right cortical hemisphere (cf. Brown, Gore, & Pearson, 1998; Franklin et al., 2008; Gilbert et al., 2006; Weekes & Zaidel, 1996; Zhou et al., 2010; but see Witzel & Gegenfurtner, 2011), we also conducted complementary analyses splitting up our results for the additional variable *side of cue presentation* (to the left or to the right visual field)[3]. This was done to check if the expected language-dependent effect on attention capture was maybe restricted to cues presented to the right visual field, and, thus, initially projected to the left hemisphere (i.e., activating the relevant linguistic representations for the task).

2.1. Methods

2.1.1. Participants

Twenty-four participants (14 female, 10 male, M_{age} = 22.84 years, SD_{age} = 4.35) from the University of Vienna (Austria) and 24 participants (10 female, 14 male, M_{age} = 22.25 years, SD_{age} = 2.01) from the Konkuk University (Seoul, Korea) were tested. The sample size was based on an a-priori power calculation: We assumed a moderate effect size and a statistical power of 80%. Here and in the following experiments, all participants were students, received about 10€ monetary compensation, and provided informed consent. All participants were native speakers of their respective language (Korean in Korea and German in Austria). They were raised as monolinguals, and did not learn any second language before the age of 8 years. A language questionnaire after the experiment ensured these criteria. One participant with an error rate of more than 25% was excluded from the Korean speaking sample.

2.1.2. Apparatus

We tested the Korean speaking participants in Seoul and the German speaking participants in Vienna. We carefully ensured that testing conditions were as similar as possible. In both laboratories, stimuli were

[3] We are grateful to one anonymous reviewer for suggesting this additional analysis of the data.

presented on a screen with a resolution of 1,024 × 768 pixels, at a vertical refresh rate of 60 Hz. All colors used for the stimuli and for the background (see Fig. 1, upper left panel) were equiluminant at 54 cd/m^2. The viewing distance was kept at 57 cm by a chin and forehead rest, and the room was dark. The experiment program was controlled using EPrime 2.0.10.353 (Psychological Software Tools Inc., Sharpsburg, PA, USA) in both laboratories.

2.1.3. Stimuli

All stimuli were presented against a grey background (CIE Lab coordinates: −3.6/−11.9) with a black fixation cross (0.5°) at screen center. Cues and targets occupied the four corners of a centered imaginary square, at an eccentricity of 5.0°. All singleton cues, targets, and distractor stimuli were red (67.9/44.7), green (−63.8/38.4), pink (86.2/−70.1), yellow (−17.5/47.7), or turquoise (−32.9/−23.2). Cues, distractors, and targets consisted of a central piston (3.3° × 1.2°), surrounded by either a tightly or loosely fitting cylinder (3.3° × 3°). Stimuli colors and sizes were identical in both laboratories.

2.1.4. Procedure

The experiment followed a contingent-capture protocol (Folk et al., 1992) that is also illustrated in Fig. 1. After a fixation display (1,500 ms), for 100 ms, a singleton cue of a unique color was shown at one position, together with three non-singleton distractors of a second color at the other positions. In 50% of the time, the singleton cue also had a unique fit (tight or loose) different from the non-singleton distractors in the cueing display of the alternative fit (i.e., tight-fit if the singleton cue was loose-fit; loose-fit if the singleton cue was tight-fit). The cue was similar or dissimilar to the target in either its color or its loose/tight-fit, or in both features. After another fixation display of 150 ms, a target was shown at one out of the four positions for 300 ms. Cue and target positions were equally likely and uncorrelated to one another. Hence, across trials, cues were unpredictive of target positions. (This is important because if all cues would indicate the likely target position, all participants, Korean and German speakers alike, would start to search for all cues as if they were targets.) The target consisted of an instructed color and an uninstructed but consistently target-coincident loose- or tight-fit that were both different from the three distractors at the other positions in the target display. During a blank screen, after target offset, participants reported via key press the tilt orientation (to the left vs. to the right) of a letter *T* superimposed on the target. We measured RTs to the nearest millisecond as the time intervals between target onset and key press. Release of the response key started the next trial. Participants were instructed to answer as fast and accurately as possible. If no key was pressed within 1,500 ms, participants received an on-screen feedback ("react faster!") and the trial counted as an error. From trial to trial, cue positions, cues' loose versus tight-fits, target positions, and target-letter orientations varied pseudo-randomly, but across trials, both cues and targets were counterbalanced across all four positions. Cue and target positions, cue and target colors, and cue and target loose- versus tight-fits were all uncorrelated across trials. Again, this was important so as to prevent that participants start searching for any of the cue features in their own rights (i.e., in the extreme case of 100% predictive cues as if the cues were the targets).

Thirty practice trials (with feedback), not analyzed, preceded the 30-min experiment consisting of 384 trials. After the experiment, participants filled out a language questionnaire and were debriefed.

2.2. Results

The dataset underlying the results of all experiments presented in this study can be found at Goller, Choi, Hong, and Ansorge (2019).

2.2.1. Reaction times

In Experiment 1, error trials and trials with an RT below or above 2.5 *SD*s of the mean (per participant and condition) were excluded (4.23%). Based on the remaining correct RTs, we calculated cueing effects (CEs) as indices of attention capture in the different cue-target similarity and dissimilarity conditions regarding color and/or loose- versus tight-fit by subtracting RTs in same-position trials from RTs in different-position trials. Positive CEs indicate attention capture by the cue (i.e., a same-position cue facilitated target identification relative to a different-position cue). CEs near zero indicate that cues did not capture attention. In the following, unless otherwise noted, CEs are significantly different from zero, tested with a *t*-test against zero. CEs were subjected to a mixed-model analysis of variance (ANOVA), with the within-subject variables cue color (target-similar; target-dissimilar) and cue loose- versus tight-fit (target-similar; target-dissimilar), and the between-subjects variable language (Korean speaker; German speaker). Fig. 2 shows the results. Most importantly, we found a significant interaction between all variables, $F(1, 45) = 5.25$, $p = 0.027$, $n_p^2 = 0.10$. This was further analyzed by two separate follow-up ANOVAs for Korean and German speakers, summarized below. In addition, we obtained significant main effects of language, $F(1, 45) = 13.79$, $p < .001$, $n_p^2 = 0.24$, cue fit, $F(1, 45) = 11.84$, $p = 0.001$, $n_p^2 = 0.21$, and cue color, $F(1, 45) = 152.60$, $p < 0.001$, $n_p^2 = 0.77$, and a significant interaction between cue fit and cue color, $F(1, 45) = 10.16$, $p = 0.003$, $n_p^2 = 0.18$. To note, as these lower-order interaction and main effects were all qualified by the higher-order interaction between all three variables, these main effects and lower-order interactions were not analyzed in their own right. Instead, we investigated these main effects and interactions in the context of the follow-up tests conducted for the three-way interaction.

In the follow-up ANOVA for Korean speakers, the CE for the target-similar color cue was larger if the cue also had a target-similar fit (107 ms) than if it had a target-dissimilar fit (69 ms), $t(22) = 3.25$, $p = 0.004$, $d = 0.68$ (see Fig. 2b right panel), indicated by an interaction between cue fit and cue color, $F(1, 22) = 6.48$, $p = 0.018$, $n_p^2 = 0.23$. No such fit-dependence was found with target-dissimilar color cues (target-similar fit: 12 ms; target-dissimilar fit: 3 ms), both CEs not significantly different from zero, $ps > 0.080$), $t(22) = 1.35$, $p = 0.190$, $d = 0.28$. These results indicate that when the cue color was target-similar, Korean speakers showed more attention capture when the cue *fit* was target-similar than when it was not. For German speakers, only a main effect of cue color was found, $F(1, 23) = 165.49$, $p < 0.001$, $n_p^2 = 0.88$, indicating a larger CE for target-similar color cues (143 ms) than for target-dissimilar color cues (7 ms, not significantly different from zero, $p = 0.094$) (cf. Fig. 2b). Neither a main effect of cue fit nor an interaction between cue fit and cue color were found, all $F < 2.32$, all $p > 0.142$.

Two results of this analysis merit a closer inspection. First, Korean speakers reacted overall slower than German speakers, indicated by the main effect of language. To check for possible floor effects in the German sample, we compared the 50% slowest answers of the German sample (median RT: 666 ms) with the 50% fastest reactions of the Korean sample (mean RT: 578 ms) with a median split comparison. Our results were essentially replicated, meaning that the critical three-way interaction was significant, $F(1, 45) = 5.95$, $p = 0.019$, $n_p^2 = 0.12$, and the post-hoc tests yielded results similar to the omnibus ANOVA. Different to the first analysis, we also obtained an interaction between language and cue color $F(1, 45) = 10.25$, $p = 0.003$, $n_p^2 = 0.19$, that was not significant in the previous ANOVA. As this interaction is qualified by the three-way interaction, we will not discuss it in detail here. Second, our results could be due to a faster learning of the target-coincident loose- or tight-fit in the Korean sample. Therefore, we compared the performance of the Korean speakers in the first half of the experiment with the second half by inserting a respective variable into the analysis. There was no main effect nor any interaction in which the variable experiment half was involved, all $Fs < 2.40$, all $ps > 0.135$.

As mentioned earlier, we also repeated the analysis of the CEs with the additional factor cue hemifield (left visual field; right visual field, see Footnote 3), to examine whether participants were thinking in language. We found a main effect of cue hemifield, $F(1, 45) = 7.26$, $p = 0.010$, $\eta_p^2 = 0.14$, that results from a higher CE when the cue was shown in the left (68 ms) as compared to the right (56 ms) hemifield. More importantly though, cue hemifield did not interact with any other factors of the ANOVA, all Fs < 2.15, all ps > 0.150. The same analysis was conducted with target hemifield instead of cue hemifield. Neither a main effect of target hemifield, nor any interaction with target hemifield was found, all Fs < 3.31, all ps > 0.075.

2.2.2. Error rates

We calculated CEs as the difference between the arcsine transformed error rates (ERs) in different-position minus same-position trials. A mixed ANOVA, with the same variables as above, yielded a main effect of cue color, $F(1, 45) = 29.95$, $p < 0.001$, $\eta_p^2 = 0.40$, indicating a lower CE after target-similar color cues (2.80%) than target-dissimilar color cues (4.31%). No other significant effects were found, all Fs < 2.87, all ps > 0.097. As before, we repeated this analysis with the additional factor cue hemifield but neither a significant main effect nor an interaction with this factor was found, all Fs < 1.26, all ps > 0.249. The same was true for an analysis with the factor target hemisphere instead of cue hemisphere, all Fs < 2.45, all ps > 0.125.

2.3. Discussion

In Experiment 1, we found evidence that Korean speakers are more sensitive to tight- and loose-fits than German speakers who do not encode these features in a grammaticalized way. This was reflected in a capture effect in a nonlinguistic visual search task. We also did not find any evidence that the corresponding interaction between language group and cue-target fit similarity was restricted to processing in the left hemisphere, minimizing the possibility that participants were actively engaging linguistic representations (i.e., thinking for speaking) during the task.

However, in addition to these hypothesis-related results, we observed a few unexpected findings. First of all, the Korean speakers were slower than the German speakers. Although this was not predicted, this result fits nicely with the rest of the data, in that stronger CEs by target-similar cue fits among the Korean speakers indicated that the Korean speakers either noted the target-coincident fit feature and incorporated this into their search sets for the targets (cf. Folk et al., 1992) or that the participants used the fits of the target-dissimilar distractors and cues to actively filter out these otherwise not helpful stimuli more efficiently (cf. Theeuwes, Atchley, & Kramer, 2000). In the first case, a target-similar cue would have created its CE by top-down contingent capture: the cue's match to the participant's search setting set up for the targets (cf. Folk et al., 1992). In the second case, both cues would have captured attention initially, but the fact that the target was never a specific fit (e.g., never a loose fit because all targets were tight-fits) would have allowed Korean speakers to quickly suppress their capture of attention by the target-dissimilar fit cues and to return to a neutral position in the cue-target interval before target onset (Theeuwes et al., 2000). In either case, the differential cueing effect regarding cue-target fit similarity indicated that the Korean but not the German speakers used two features (color and fit) rather than only one feature for their visual search task performance. As with each additional feature used in a visual search task, processing demands increase (cf. Büsel, Pomper, & Ansorge, 2018), it is no wonder that on average the RTs (see Fig. 2a) increased among the Korean relative to the German speakers. In line with this argument, for a related reason, the CE in color-similar conditions was on average somewhat smaller among the Korean than among the German speakers (see Fig. 2b). For example, if it takes longer to search for two features (color and fit; by the Korean speakers) than for one feature (color only; by the German speakers) because with two features participants would occasionally have to switch between them for the search (Büsel et al., 2018; Moore & Weissman, 2010), then the CE by each feature under two-feature search conditions could be smaller than under one-feature search conditions, simply because it is less likely that the cue feature that the participants currently see would always correspond to the target feature that the participants currently search for[4].

A second unanticipated result concerns the lacking CE of the target-similar fit cues in color-dissimilar conditions (cf. Fig. 2b). Here, the color difference between target-dissimilar color cue and the color of the target at the cued position could have delayed target perception (Carmel & Lamy, 2014, 2015). If this happened, the updating cost created by the change from the cue color to the target color in cue-target dissimilar conditions could have masked an attention capture advantage toward the cue's position and, thereby, nullified the capture effect measured in a net CE.

Experiment 1 provided tentative evidence for a linguistic long-term memory influence on attention capture in a nonlinguistic task. However, other group differences between Korean and German speakers might also account for the differences in CEs based on fit-similarity. Experiments 2 and 3 address two possibilities, one by one: possible group differences in holistic versus analytical visual processing of scenes (Experiment 2) and lack of perceptual ability to categorically distinguish between tight and loose fit among German speakers (Experiment 3).

3. Experiment 2

Korean and German speakers might have differed in more than their languages' semantics, and some of these group differences could also account for the found differences in terms of CEs by tight versus loose fit objects. One possibility is that Asians, process visual scenes more holistically than Westerners, meaning that Asians pay more attention to the relations between entities in a visual scene than Westerners: "…, we believe there is considerable evidence that shows that Asians are inclined to attend to, perceive and remember contexts and relationships whereas Westerners are more likely to attend to, perceive and remember the attributes of salient objects and their category memberships" (p. 469, Nisbett & Miyamoto, 2005). Translated into the situation of Experiment 1, as representatives of the Asian culture, Korean speakers might have simply spent more time on processing and using the fit relation between piston and cylinder than German speakers, so as to note the target-coincident fit between the two, regardless of their linguistic relevance. Another possibility is that the Korean speakers were better than the German speakers in terms of their cognitive control abilities. For instance, Korean speakers might have simply spent more time on processing all target features – instructed target colors as well as target-coincident fits - than German speakers. One could argue that the overall slower responses and the lower CEs of the Korean speakers are in line with this possibility, providing further evidence of a more thought- and careful approach to target processing.

Experiment 2 was our first control experiment to rule out such general cultural or cognitive differences as better explanations for the capture effect differences we found between the language groups in Experiment 1. To that end, we used flat 2D "fits" that consisted of 2D disk and surrounding ring that were either contiguous with each other or that were separate, as the Korean language does not grammatically distinguish as *tight* (contiguous) versus *loose* (non-contiguous) relations

[4] A weaker cueing effect with two searched-for features than one searched-for feature is not always found (e.g., Irons, Folk, & Remington, 2012). However, studies that failed to find the effect typically used two colors rather than color and fit, as were used here. As a change of dimensions (e.g., between color and shape) seems to incur a cost (Müller, Heller, & Ziegler, 1995) that is different from that of a change in features of the same dimension (e.g., between two colors), this procedural difference might be responsible for why the interaction between number of searched-for target features and cueing effects was found here but not everywhere else (e.g., Kerzel & Witzel, 2019).

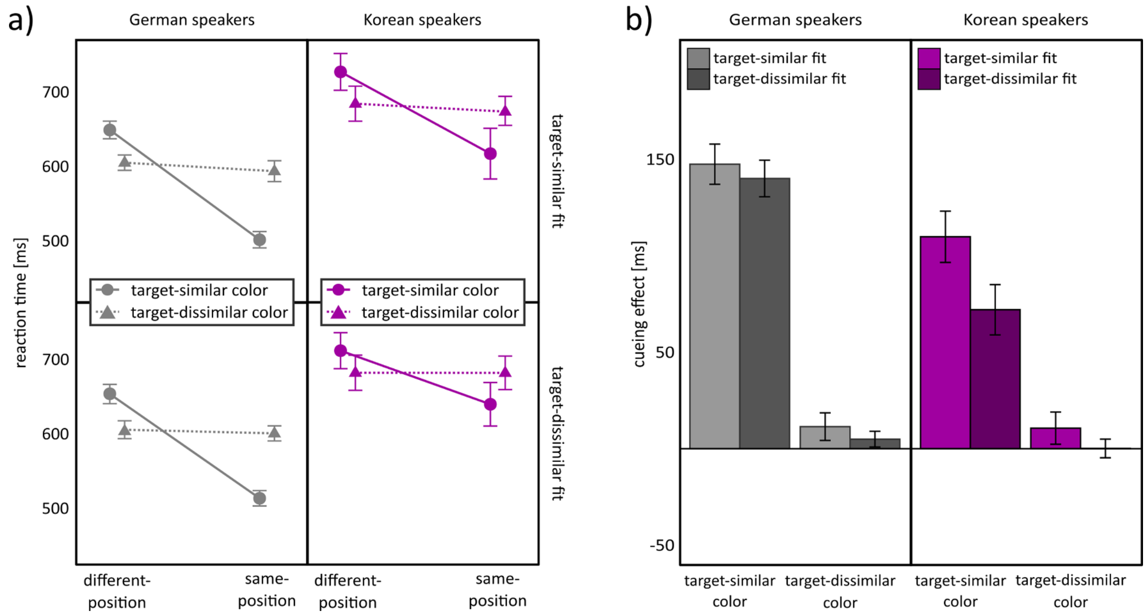

Fig. 2. Illustration of the results of Experiment 1. (a) Reaction times: The left panel (grey lines) shows the data from the German speakers, the right panel (purple lines) shows the data from the Korean speakers. Solid lines indicate performance with target-similar color cues, dotted lines indicate target-dissimilar color cues as a function of cueing (different-position vs. same-position cue conditions) on the *x* axis. Finally, the upper panel shows target-similar fit cues, the lower panel shows target-dissimilar fit cues. (b) Cueing effects (CEs): The left panel (grey) shows the data from the German speakers, the right panel (purple) shows the data from the Korean speakers. The shades of the colors indicate target-similar and target dissimilar fit cues, both depicted for target-similar versus target dissimilar color cue conditions on the *x* axis. All error bars represent *SEM*.

in 2D objects (Bowerman & Choi, 2003).[5] Because neither the Korean nor the German language grammatically discriminates between the two different types of 2D structure, Korean and German speaking participants should show a similar insensitivity to the (cues resembling the) target-coincident 2D fits. However, if general cultural or cognitive differences set the Asian and the Western sample apart, in Experiment 2, the same CE differences as found in Experiment 1 can be expected.

3.1. Methods

3.1.1. Participants

Twenty-four participants (16 female, 8 male, M_{age} = 21.44 years, SD_{age} = 3.03) from University of Vienna (Austria) and 24 participants (13 female, 11 male, M_{age} = 21.97 years, SD_{age} = 2.44) from Konkuk University (Seoul, Korea) were tested. One participant with an error rate of more than 25% was excluded from the German speaking sample.

3.1.2. Apparatus, stimuli and procedure

These were the same as in Experiment 1 (see Fig. 3), with the notable exception that cues, targets and distractors each consisted of a (flat) 2D central disk (1.2°), surrounded by a (flat) 2D ring (3.3°) that either were contiguous all around the inner disk (i.e., no gap between the two) or left a small gap between disk and ring (0.5°). See Fig. 3 for an illustration.

3.2. Results

3.2.1. Reaction times

In Experiment 2, a total of 4.16% trials (error trials and RT outliers) were excluded. The analysis was analogous to Experiment 1. Fig. 4 shows the results. The CE was larger following target-similar (109 ms) than target-dissimilar color cues (10 ms), indicated by a main effect of cue color, $F(1, 45) = 144.57$, $p < 0.001$, $\eta_p^2 = 0.76$. In line with a linguistic and, therefore, 3D-specific explanation of the findings of Experiment 1, no other effects were significant, especially no interaction with language, all *F*s < 1.32, all *p*s > 0.249. This is at odds with an explanation according to which the Korean-German differences obtained in Experiment 1 with 3D-fit stimuli were due to a higher context sensitivity or more cognitive control among Asian participants than among Westerners.

We further assessed a potential influence of the cue and the target hemifield [Footnote 3]. First, we repeated the above analysis, with the additional factor cue hemifield–that is, left hemisphere (= right visual field) versus right hemisphere (= left visual field). We found the main effect of cue color as before, but no other main effects or interactions were significant, all *F*s < 1.93, all *p*s > 0.171. Second, we explored a possible influence of the target hemifield. The ANOVA with the additional factor target hemifield (left versus right) yielded a main effect of this factor, $F(1, 45) = 5.51$, $p = 0.024$, $\eta_p^2 = 0.11$. The CE was slightly larger if the target was presented in the left (65 ms) as compared to the right hemifield (56 ms). Besides an also significant main effect of cue color, no other main effects or interactions were significant, all *F*s < 1.93, all *p*s > 0.171.

3.2.2. Error rates

CEs based on the arcsine transformed ERs were calculated analogous to Experiment 1. An analysis of the CEs yielded a main effect of cue color, $F(1, 45) = 5.39$, $p = 0.025$, $\eta_p^2 = 0.11$. The CE was larger for target-similar (3.95%) than for target-dissimilar color cues (3.83%). No other effects were found, all *F*s < 1.04, all *p*s > 0.249. We repeated the analysis with the additional factor cue hemifield (left visual field; right visual field) and found an interaction between cue color and cue hemifield. The CE was higher if a different-color cue was shown in the right (4.12%) as compared to the left hemifield (3.55%), $t(46) = 2.36$, $p = .023$, $d = 0.34$. No such difference was found for same-color cues (4.18% vs. 3.72%), $t(46) = 1.45$, $p = 0.152$, $d = 0.21$. No other significant results were found, all *F*s < 3.22, all *p*s > 0.080. We conducted a similar analysis with target hemifield instead of cue hemifield, but only found the already known main effect of cue color and a main effect of target hemisphere, $F(1, 45) = 4.67$, $p = 0.036$, $\eta_p^2 = 0.09$. The

[5] For both types of fit in 2D, in Korean, one would use the phrase *an-ey issta*, meaning *inside-at located* referring to *X being located inside (of) Y*.

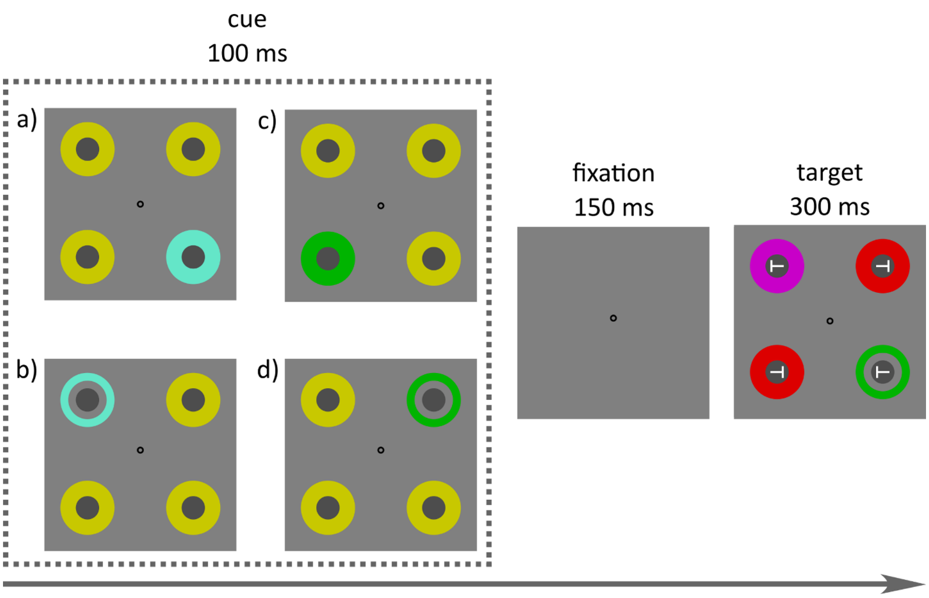

Fig. 3. Procedure of Experiments 2. Depicted are trials with a green target (see rightmost panel). Four different cue conditions for Experiment 1 are illustrated on the left side (see panel labelled with *cue*): On the left side (a and b), cues with a target-dissimilar color are shown, on the right side (c and d), cues with a target-similar color. The upper row of the cue displays (a and c) shows target-dissimilar fit cues, the lower row (b and d) target-similar fit cues. The arrow depicts the flow of time. The stimuli are not drawn to scale.

CE was slightly higher if the target was presented in the right (3.97%) versus the left (3.81%) hemifield. No other effects were found, all non-significant all $Fs < 2.16$, all $ps > 0.148$.

3.3. Discussion

Experiment 2 confirmed that when the two languages do not semantically differentiate between two types of spatial relation (2D [tight] fit versus 2D [loose] non-fit), there was no attention capture by cue-target fit similarity in neither Korean nor German speakers. This was found despite the use of otherwise very similar tasks and stimuli, and despite recruiting participants from the same environments as in Experiment 1. As there is no reason why general cultural or cognitive differences between Asian and Western participants should have vanished from Experiment 1 to Experiment 2 (with two independent and randomly assigned groups of participants), it is likely that neither cultural nor language-independent cognitive differences contributed to the capture-effect differences between Korean and German speakers that we found in Experiment 1.

4. Experiment 3

As Experiment 1 did not show any significant capture by target-similar fit cues among German speakers, we wanted to make sure that German speakers are able to perceive and use these fits if explicitly instructed to do so. For this aim, we conducted Experiment 3 where German speakers were instructed to search for target-defining 3D-fits (either tight-fit or loose-fit, between participants). If some persistent

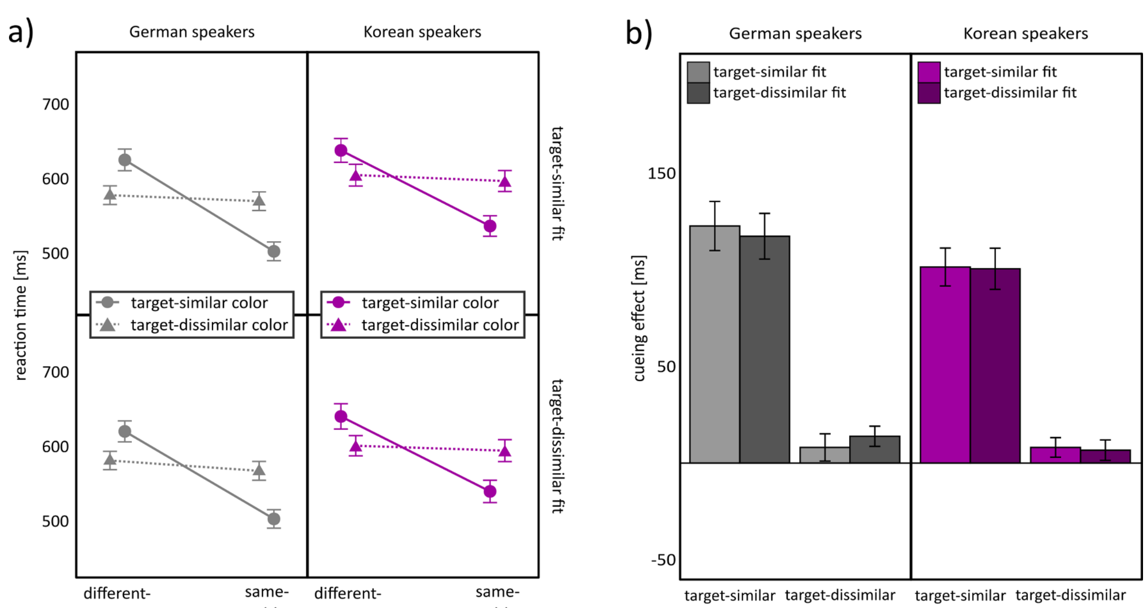

Fig. 4. Illustration of the results of Experiment 2. (a) Reaction times: The left panel (grey lines) shows the data from the German speakers, the right panel (purple lines) shows the data from the Korean speakers. Solid lines indicate target-similar color cues, dotted lines indicate target-dissimilar color cues as a function of cueing (different-position vs. same-position cue conditions) on the *x* axis. Finally, the upper panel shows target-similar fit cues, the lower panel shows target-dissimilar fit cues. (b) Cueing effects (CEs): The left panel (grey) shows the data from the German speakers, the right panel (purple) shows the data from the Korean speakers. The shades of the colors indicate target-similar and target dissimilar fit cues. They are depicted for target-similar versus target-dissimilar color cues on the *x* axis. All error bars represent *SEM*.

cognitive or cultural difference between the two groups prevented the use of target-coincident fits among German speakers in Experiment 1, we expected to see no capture by target-similar fit cues in Experiment 3 either. However, if German speakers failed to register and use target-defining fits in Experiment 1 simply because their habitual language use did not increase their sensitivity for these features, an explicit instruction to search for these features should yield a capture effect based on target-similar cue fits among German participants in Experiment 3.

4.1. Methods

4.1.1. Participants

Twenty-four German speaking participants (18 female, 6 male, $M_{age} = 24.55$ years, $SD_{age} = 5.47$) from the University of Vienna (Austria) were tested. Two participants with an error rate exceeding 25% were excluded.

4.1.2. Apparatus, stimuli and procedure

These were the same as in Experiment 1, with the only difference being that participants were now instructed to search for a specific 3D-fit (tight-fit or loose-fit, counterbalanced across participants) and color now served as the additional target-coincident feature. In other words: In comparison with Experiment 1, color and fit switched their roles, and although the target had both one specific color and one specific fit throughout the whole experiment, only one of the features – here: the fit – was ever mentioned to the participants and instructed as the to-be-searched-for target feature.

4.2. Results

4.2.1. Reaction times

All error trials and RT outliers (12.96%) were excluded by the same criteria as before. The analysis was the same as in Experiments 1 and 2, with the notable exception that there was no between-subjects variable language, as only German speakers were tested. Fig. 5 shows the results. When explicitly instructed to search for the target-defining 3D-fit, German speakers showed attention-capture by target-similar fit cues: We found a main effect of cue fit, $F(1, 21) = 10.59$, $p = 0.004$, $n_p^2 = 0.34$, indicating a CE for target-similar fit cues (36 ms) and no CE for target-dissimilar fit cues (−16 ms, not significantly different from zero, $p = 0.123$). Additionally, we found a significant main effect for cue color, $F(1, 21) = 5.87$, $p = 0.025$, $n_p^2 = 0.22$, indication of a larger CE following target-similar (25 ms) than target-dissimilar (−6 ms, not significantly different form zero, $p > 0.249$) color cues. There was no interaction between cue color and cue fit, $p > 0.249$.

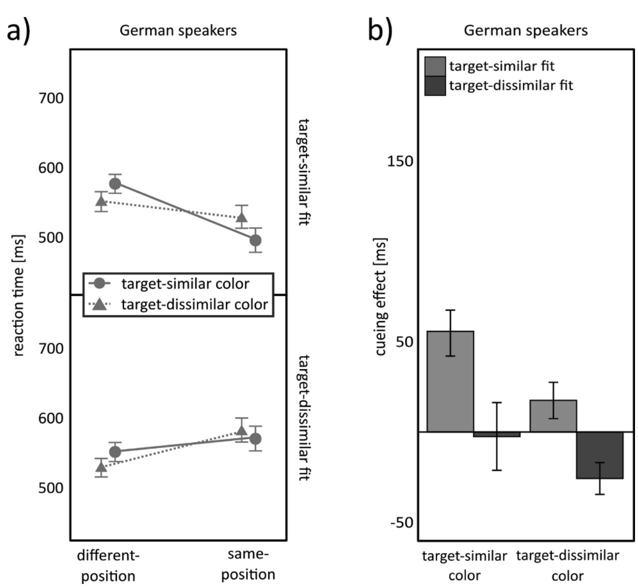

Fig. 5. Illustration of the results of Experiment 3. (a) Reaction times: Solid lines indicate target-similar color cues, dotted lines indicate target-dissimilar color cues, illustrated as a function of cueing (different-position vs. same-position cue conditions) on the *x* axis. Finally, the upper panel shows target-similar fit cues, the lower panel shows target-dissimilar fit cues. (b) Cueing effects (CEs): The shades of grey indicate target-similar and target dissimilar fit cues. They are depicted for target-similar versus target-dissimilar color cues on the *x* axis. All error bars represent *SEM*.

4.2.2. Error rates

CEs based on the arcsine transformed ERs were calculated analogous to Experiment 1. The analysis revealed a borderline significant main effect of cue color (target-similar color: 13.27%; target-dissimilar color: 13.38%), $F(1, 21) = 4.13$, $p = 0.057$, $n_p^2 = 0.19$. No other effects were found, both Fs < 1.38, both ps > 0.249.

4.3. Discussion

We tested if explicit instructions to search for loose or tight-fits could replace the lacking long-term influence of grammaticalized language characteristics among German speakers. In line with this, the instructions allowed German speakers to also use target fits to search for targets or to use distractor fits to suppress distractors, in Experiment 3: German speakers showed a stronger CE by target-similar fit cues than by target-dissimilar fit cues. This finding rules out that some general cultural specificity or cognitive insensitivity or perceptual inability prevented attention capture based on stimulus fit among German speaking participants.

In addition to this theoretically most important effect, there was an interesting side observation. The participants showed more capture of attention by target-similar color cues than by target-dissimilar color cues. As this capture effect was based on uninstructed colors and as Germans do not discriminate colors in a grammaticalized way, the color-based CEs suggest that grammaticalized language characteristics are not the only factor driving attention capture. Instead, visual characteristics per se probably also played a role. For example, if German-speaking participants registered target-coincident colors somewhat earlier than instructed target fits, participants would have noticed target colors more easily in Experiment 3 than they would have noticed target fits in Experiment 1. Such a salience difference between the different target characteristics could have increased chances that the participants incorporated target-coincident colors into their top-down search settings. This was in comparison to chances to incorporate target-coincident fits into their search settings in Experiment 1, unless some linguistic long-term memory representation facilitated processing of target-coincident fits, which was only the case for the Korean speakers of Experiment 1. We will get back to this point following the next experiment.

5. Experiment 4

As explained in the Discussion of Experiment 1, different accounts are possible to explain more attention capture by target-similar cues. This is also true of the CEs based on fit-similarity among the Korean speakers that we found in Experiment 1. According to one explanation, stronger CEs for target-similar fit (and color) cues could be due to top-down contingent capture: more capture by cues matching a set of top-down searched-for target features than by cues not matching such a set, right from cue onset onward (see Folk et al., 1992). Yet, according to another explanation, both target-similar and -dissimilar fit (and color) singleton cues might initially capture attention in a bottom-up way based on their visual/perceptual salience. However, ultimately the participants want to find the target and therefore have to dismiss all cues. (Remember that all cues were not predictive of the targets.) Dismissing a cue as irrelevant following initial capture by this cue is, of course, easier for target-dissimilar cues than for target-similar cues, as stimulus discrimination is a function of stimulus differences. Hence, participants could deallocate attention away from a target-dissimilar cue at an earlier point in time than away from a target-similar cue (Theeuwes et al., 2000). In Experiment 1, this was possibly not noted, as an interval between cue and target could have allowed for active deallocation away from the target-dissimilar fit cues to be concluded even before the target was presented. As a consequence, nonsignificant CEs could have been due to capture by the target-dissimilar cue followed by quick deallocation of attention from this cue.

If this deallocation took place in Experiment 1, it should be possible to also demonstrate bottom-up capture even by target-dissimilar singleton fit distractors among Korean speakers, where no interval would allow to disengage attention from the distractor and before the target is presented. Therefore, in Experiment 4, we used an additional-singleton protocol (cf. Theeuwes, 1992), in which the irrelevant fit singleton as a distractor was always different from the target's fit and in which the singleton distractor was presented *at the same time* as the target but away from the target. Such a singleton should only capture attention in a bottom-up way – that is, different from Experiment 1, participants cannot incorporate this feature into a top-down search set for the target as the targets are different. As there was no interval between singleton distractor and target, there would also be no time to deallocate attention before the target.

These characteristics of the protocol have two additional desirable implications. First, by focusing on this type of bottom-up attention capture, we decreased the likelihood for an active use of language to represent the feature in question even further because it would not make any sense at all for the participants to name or otherwise linguistically represent such a distracting feature[6]. Secondly, through bottom-up capture we were able to put the language-dependent sensitivity difference for object fits to another, independent test. If the grammaticalized discrimination between tight and loose fits among the Korean speakers but not among the German speakers sensitizes Koreans speakers to visual fits between objects, we expected more interference by a target-dissimilar singleton fit distractor presented away from the targets among Korean than among German speakers. This interference in singleton-fit distractor conditions was measured relative to a baseline condition without singleton distractor.

In addition, to rule out once more that differences in terms of cognitive control abilities between the different language groups were responsible for the expected attention capture differences, we included a control condition with color singleton distractors. In this control condition, we expected to see capture by the color distractor in both Korean and German speakers (e.g., Theeuwes, 1992; Weichselbaum & Ansorge, 2018), as the linguistic differences concerned the fits and not the colors. In contrast, if a group difference in cognitive control in general existed, for example, because German speakers are better able to suppress bottom-up capture by any task-irrelevant singletons, be that now task-irrelevant fits or task-irrelevant colors, we expected to see more evidence of successful suppression (i.e., less evidence of capture) in irrelevant fit-singleton *and* irrelevant color-singleton conditions among German than among Korean speaking participants.

An upshot in this experiment is that it conceptually replicates Experiment 1 with a different protocol. We expected to replicate more capture by 3D-fit singleton distractors among Korean speakers that linguistically discriminate this singleton characteristic in a grammaticalized way. As explained, under the linguistic perspective, color-singleton distractors were expected to capture attention among both Korean and German speakers (Theeuwes, 1992).

5.1. Methods

5.1.1. Participants

Thirty-two participants (18 female, 14 male, M_{age} = 23.21 years, SD_{age} = 2.87) from the University of Vienna and 32 participants (17 female, 15 male, M_{age} = 22.50 years, SD_{age} = 2.53) from the Konkuk University were tested.

5.1.2. Apparatus

These were similar to the previous experiments, but programming was in PsychoPy 1.84 (Peirce, 2007, 2009).

5.1.3. Stimuli

All stimuli were presented against a black background (0.5/0.8) with a grey fixation cross (0.5°; −3.6/−11.9) in the center of the screen and with an eccentricity of 5.0°. The stimuli were as in Experiment 1, except that we added an additional blue (49.7/–98.5) stimulus that served as a singleton-color distractor only. In addition, letters *T* of different orientations were replaced by horizontal and vertical bars.

5.1.4. Procedure

After a fixation display (1 s), a search display, consisting of eight equidistant stimuli arranged on an imaginary circle around the fixation cross, was presented until the participant pressed a key (see Fig. 6). The participants' task was to report the orientation (horizontal vs. vertical) of a bar superimposed on a red or green color-singleton target. Participants were told that the color-singleton target was either a single green stimulus among red non-targets or a single red stimulus among green non-targets. Across trials, the color of the singleton target was chosen randomly but counterbalanced.

The experiment consisted of three blocks (see Fig. 6): In the no-singleton distractor block, the color-singleton target was shown together with non-singleton non-targets at all other positions. In the 3D-fit-singleton distractor block, a fit-singleton distractor replaced one of the non-singleton non-targets, and all other non-singletons had the same color. In the color-singleton distractor block, we presented a blue color-singleton distractor, and the target and all non-singleton non-targets had the same fit. All participants were first tested in the no-singleton-distractor block, whereas the order of the fit-singleton and color-singleton distractor blocks was counterbalanced across participants.

From trial to trial, target positions, target colors (red or green), target-line orientations, and, where it applied, the positions of the additional (distractor) singletons varied pseudo-randomly. The experiment consisted of 640 trials and lasted about 45 min.

5.2. Results

5.2.1. Reaction times

We excluded 8.15% of all trials by the same criteria as in the previous experiments. We subjected the mean correct RTs to a mixed ANOVA, with the within-subject variable block (no-singleton distractor, fit-singleton distractor, color-singleton distractor) and the between-subjects variable language (Korean speaker; German speaker). Fig. 7 illustrates the following results. We found a significant main effect of block, $F(2, 124) = 15.08$, $p < 0.001$, $n_p^2 = 0.20$, that also interacted with language, $F(2, 124) = 3.50$, $p = 0.048$, $n_p^2 = 0.05$. There was no significant main effect of language, $F(1, 62) = 3.05$, $p = 0.086$, $n_p^2 = 0.05$. Post-hoc tests showed that RTs of Korean and German speakers did not differ significantly in the no-singleton distractor (772 ms vs. 783 ms), $t(62) = -0.42$, $p > 0.249$, $d = -0.10$, and the color-singleton distractor (947 ms vs. 899 ms), $t(62) = 0.94$, $p > 0.249$, $d = 0.24$, blocks. However, in line with a language-based

[6] One reviewer noted that participants could have also set up linguistic representations for the suppression of distracting features. However, as this would have only been necessary if a feature would have captured attention in the first place, such a task-activated linguistic representation would have been a consequence, not a cause of the processing of the distracting visual feature. This, in turn, makes it likely that – other factors ruled out – long-term memory representations of grammaticalized linguistic characteristics biased visual processing among the Korean speakers even in a nonlinguistic task.

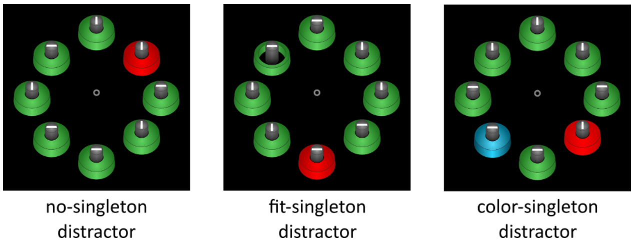

Fig. 6. Illustration of the different conditions in Experiment 4. Each display is an example of a different block. In the no-singleton distractor block (left screen), the target (here: red) was unique in color whereas all the non-targets (here: green) were homogenous in color and spatial fit. In the fit-singleton distractor block (center screen), the target (red) had again a unique color and all the non-targets had shared the same color that was different from the target color (here: green). However, one of the non-targets was a fit-singleton (upper left position in this example) that had a different spatial fit (loose-fit) than the other non-targets (tight-fit). In the color-singleton block (right screen) the target (red) was surrounded by green non-targets and one color-singleton distractor (blue). All stimuli had the same spatial fit. Stimuli are not drawn to scale. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

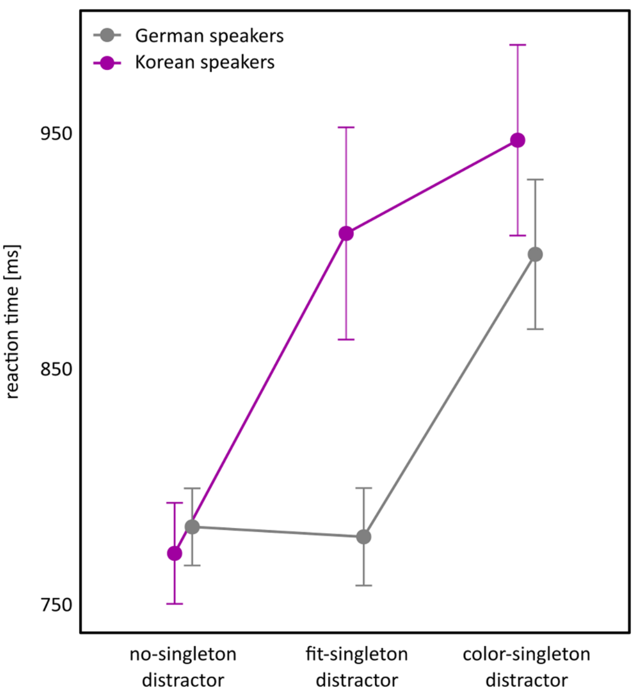

Fig. 7. Illustration of the reaction times (**y** axis) of Experiment 4. The **x** axis indicated the different blocks of the experiment. The separate lines show the results of Korean speakers (purple) and German (grey) speakers. All error bars represent *SEM*. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

difference in capture by 3D-fit singletons, Korean speakers took longer than German speakers to respond in fit-singleton distractor blocks (907 ms vs. 779 ms), $t(62) = 2.60$, $p = 0.012$, $d = 0.65$.[7]

5.2.2. Error rates

We also subjected the arcsine transformed ERs to a mixed ANOVA with the same variables. We only found a main effect of language, $F(1, 62) = 36.47$, $p < 0.001$, $n_p^2 = 0.37$, indicating a significantly lower ER for Korean (1.16%) than for German (4.45%) speakers. No other effects were found, both non-significant *F*s < 2.37, both *ps* > 0.098.

5.3. Discussion

Experiment 4 showed more attention capture by fit-singleton distractors among Korean than among German speakers. This result suggests that more capture by target-similar than target-dissimilar cues in Experiment 1 could have been due to more deallocation following *bottom-up* attention capture by target-dissimilar than target-similar fit cues (cf. Theeuwes et al., 2000). Otherwise – that is, if CE differences between target-similar and target-dissimilar fits among Korean speakers in Experiment 1 had been due to selective capture only by the top-down matching target-similar cues, thus, lacking capture by target-dissimilar cues (cf. Folk et al., 1992) – we would have expected to see no capture by the target-dissimilar fit cues in Experiment 4. Moreover, control conditions with task-irrelevant (blue) color-singleton cues showed equal attention capture among Korean speakers as among German speakers. This result rules out the possibility that German speakers were different from Korean speakers in regards to some other, nonlinguistic cognitive function such as lower vulnerability to attention capture by just any singleton distractor in general. This result demonstrates that the differences between target-similar and target-dissimilar color cues in Experiment 1 were probably also due to successful suppression following initial capture by target-dissimilar color cues.

Finally, following Experiment 3, we speculated that German speakers might have featured a temporal advantage in processing color relative to fit, and that this temporal advantage could have allowed the German speakers to register and use the target-coincident colors for their visual search performance in Experiment 3. Experiment 4 now supported this conclusion. Evidently, German speakers processed a singleton in the color dimension faster than a singleton in the fit dimension: For the German speakers, the irrelevant color-singleton distractor captured attention before the target was found and, thus, the color singleton interfered with the target responses. In contrast, no such interference by the irrelevant fit-singleton distractor was found. This suggested that in Experiment 1, German speakers registered a fit-singleton too late – if they registered it at all — to cause any interference with search for a color-defined target.

6. Experiment 5

In Experiment 5, we tested the 2D versus 3D difference in attention capture that we hitherto only investigated as a between-participants difference (between Korean speakers of Experiment 1 and those of Experiment 2) as a within-participant difference. We used the same additional-singleton procedure as in Experiment 4. Only Korean speakers were tested. In one block, the Korean speakers were presented with depictions of 3D object fits (as in Experiments 1 and 4) and in the other condition, they were presented with 2D disks and rings (as in Experiment 2). If we can replicate more capture by 3D-fit singleton distractors than by 2D fit singleton distractors within participants in Experiment 5, we can rule out that a mere chance difference between different samples accounted better for the corresponding CE difference between Korean speakers of Experiments 1 and 2.

Participants. Thirty-two participants (15 female, 17 male, $M_{age} = 21.45$ years, $SD_{age} = 4.66$) from Konkuk University were tested.

Apparatus, Stimuli and Procedure. These were similar to Experiment 4, with the exception that the color-singleton distractor block was replaced by a 2D-fit singleton distractor block. The rings and disks in this block were the same as in Experiment 2. All participants were first tested in the no-singleton-distractor block. The order of 2D-fit-singleton

[7] For the sake of completeness, we checked whether, within the two language groups, the classical findings from additional singleton experiments were replicated in Experiment 4: A higher RT in the color-singleton distractor block than in the no-singleton distractor block. Bonferroni-corrected pairwise comparisons between the three different blocks, confirmed these expectations for Korean (no-singleton distractor: 772 ms vs. color-singleton distractor: 947 ms, $p > 0.001$) and German speakers (no-singleton distractor: 783 ms vs. color-singleton distractor: 899 ms, $p > 0.001$).

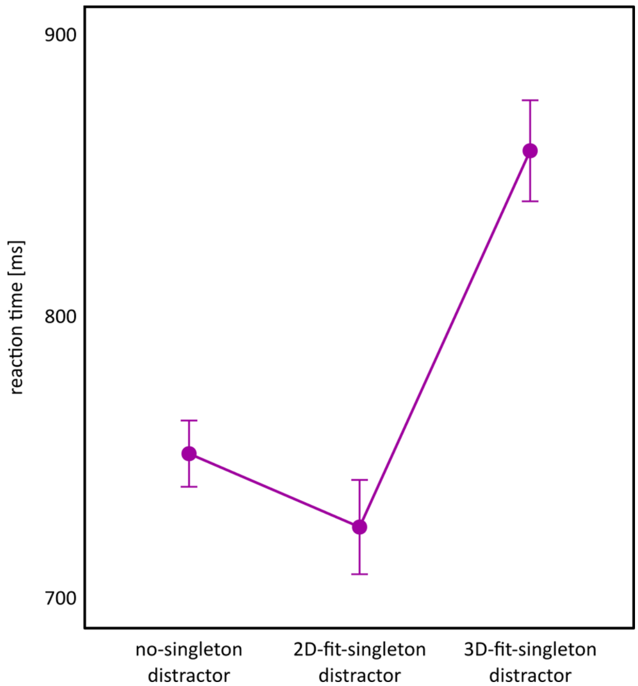

Fig. 8. Illustration of the reaction time (*y* axis) of Experiment 5. The *x* axis indicated the different blocks of the experiment. The solid line shows the results of Korean speakers. All error bars represent *SEM*.

and 3D-fit singleton distractor blocks was counterbalanced across participants.

6.1. Results

6.1.1. Reaction times

We excluded 7.51% of all trials by the same criteria as in the previous experiments (error trials and RT outliers). We subjected the mean correct RTs to a mixed ANOVA, with the within-subject variable block (no-singleton distractor; 3D-fit-singleton distractor; 2D-fit-singleton distractor) and the between-subjects variable block order (2D-3D; 3D-2D). For the results, see Fig. 8. We found a significant main effect of block, $F(2, 60) = 45.43$, $p < 0.001$, $n_p^2 = 0.60$. No other effects were found, both non-significant $Fs < 0.44$, both $ps > 0.249$. Pairwise Holm corrected comparisons showed that participants reacted significantly slower in the 3D-fit singleton distractor block (859 ms) than in the 2D-fit singleton distractor block (725 ms, $p < 0.001$) and in the no-singleton distractor block (751 ms, $p < 0.001$). There was no significant difference between the 2D-fit singleton distractor block and the no-singleton distractor block ($p > 0.249$).

6.1.2. Error rates

The arcsine transformed ERs were subjected to the same analysis as the RTs. The analysis revealed a significant main effect block, $F(2, 60) = 4.50$, $p = 0.021$, $n_p^2 = 0.13$. Pairwise Holm corrected comparisons between all blocks, yielded no significant differences (all $p > 0.300$). A significant main effect of block order, $F(1, 30) = 17.89$, $p < 0.001$, $n_p^2 = 0.37$, indicates that participants made overall significantly more errors if they started with the 2D-composite-singleton distractor block (7.35%) as compared to the 3D-composite-singleton distractor block (3.71%). No interaction was found, $F = 2.73$, $p = 0.083$.

6.2. Discussion

Using a sample of Korean speakers, Experiment 5 showed clearly that attention capture by the irrelevant fit singletons was restricted to the 3D shapes that are distinguished in a grammaticalized way in the Korean language, and that 2D shape fits that were not linguistically distinguished in a grammaticalized way in the Korean language do not capture attention. This finding shows that the difference in attention capture between 2D and 3D-fits can be replicated within participants, thereby, ruling out participant differences as an account for the findings.

7. General discussion

To conclude, Experiments 1, 4, and 5 showed a higher sensitivity for attention capture by 3D objects' tight- versus loose-fits among Korean than German speakers. We also showed that these differences were not due to a generally larger sensitivity to contexts among Korean than German speakers (Experiments 2 and 5) or to a general inability of German speakers to direct their attention to such fits (Experiment 3). Together, these results are in line with a higher sensitivity of the Korean speakers to 3D tight- and loose-fits in a nonlinguistic task that was predicted on the basis of the Korean speakers' language-specific grammaticalized distinction between these objects. In contrast, German speakers, who do not systematically distinguish between these two types of fits were correspondingly not sensitive to capture by 3D tight- and loose-fits. This holds true even though the successful task performance (here: visual search) did not depend on an instruction mentioning the fits and though it was not necessary to attend to these fits to solve the task, as the targets in Experiments 1, 4 and 5 were defined by color. Additionally, in Experiments 4 and 5, the target never carried the fit-singleton distractor's fit feature: If the target was a non-singleton loose-fit stimulus, the distractor fit singleton was a tight-fit and vice versa. If anything, the explicit instruction to search for color may have even hidden the relevance of the linguistic fit concepts. Thus, our results are well in line with a long-term memory effect of linguistic representations on visual attention in a nonlinguistic task. Even if a visual search task, such as ours, would sometimes invite the use of linguistic representations in working memory to solve the task (cf. Meyer et al., 2007; Walenchok et al., 2016; but see, e.g., Baier & Ansorge, 2019; Najemnik & Geisler, 2005; Zelinsky, 2008), once an object is currently not relevant for the task, this object is not represented in active working memory responsible for attention capture (Soto & Humphreys, 2007).

One might argue that in Experiment 1, Korean participants could have used a verbal representation to actively search for target-coincident fits, once participants registered the coincidence. Note, however, that this leaves open the question of why the target fit aroused the interest of the Korean but not the German speakers in the first place, as the target-coincident fit was never mentioned to the participants. Therefore, such a post-hoc explanation would still require explaining the higher sensitivity of the Korean speakers for registering the visual fits. In Experiments 4 and 5, there were even less reasons for the inclusion of a linguistic representation of the fits, as these fits were only used for distractor singletons and could never be used to find the target [Footnote 6]. The participants' activation of linguistic working-memory representations concerning the spatial fit was, thus, unlikely in the context of our visual search tasks. This means, in turn, that some more persistent linguistic representations in long-term memory must have elicited the language-dependent attention-capture effects in our study.

Our findings are therefore in line with some form of linguistic relativity, but how exactly can this language-specific long-term memory representations influence the capture of visual attention? In the Introduction, we have referred to Lupyan's feedback-label hypothesis (2012), according to which reciprocal connections between linguistic long-term memory representations and visual feature detectors could be responsible for language-cognition interactions (see also Miller, Schmidt, Blankenburg, & Pulvermüller, 2018). However, for a full explanation of the present findings, a crucial ingredient is missing: The reciprocal activation of representations would not explain why fit-singletons captured attention in Experiments 4 and 5. Say that attention capture is merely a function of the activation of some linguistic representation by a visual feature that then feeds back on the corresponding visual feature detector. In this case, the non-singletons rather than the singletons should have attracted attention the most. For example, together with a tight-fit singleton distractor, participants would have seen several loose-fit non-singletons. Thus, attention capture on the basis of visual detector activation plus feedback from co-activated linguistic representations corresponding to the detected visual features would have fostered capture by the non-singletons rather than by the singletons – that is, the “opposite” of what we observed.

Thus, the grammaticalized distinction between tight- and loose-fits must have worked differently. One possibility is that the linguistic discrimination increased the sensitivity for the corresponding visual discrimination between these objects. Instead of the activation of either the one or the other of the concept representations involved in a grammaticalized discrimination, it would have been the *operation* of linguistically discriminating between concept representations that somehow fed down to the operation of visually discriminating between the corresponding visual features. For example, when a language persistently requires to discriminate between visual events for the selection of the correct grammaticalized linguistic representation, then the corresponding perceptual discrimination is practiced and learned: Every time a speaker makes a correct perceptual discrimination, s/he can use the appropriate linguistic label and chances for her/him being understood and, thus, to be rewarded increase. We think that this type of language-use supervised learning could be part and parcel of the origin of visual categorizations. For example, during repeated search for objects of a category (e.g., animals), search templates can be derived from the prototypical or the necessary features that indicate the membership of a stimulus to a specific category (e.g., Hout, Robbins, Godwin, Fitzsimmons, & Scarince, 2017; Levin, Takarae, Miner, & Keil, 2001; Maxfield, Stalder, & Zelinsky, 2014; Robbins & Hout, 2015; Yu, Maxfield, & Zelinsky, 2016). However, such learning of a category requires an external *supervision criterion* by which the weights of visual feature detectors that facilitate successful categorization (e.g., search) are increased and by which the weights of feature detectors that failed to do so are decreased (cf. McClelland & Rumelhart, 1985). We propose that grammaticalized distinctions provide exactly one such supervision criterion: For a language, to serve the propositional act (of appropriately informing about the world, here: the visual surroundings; Searle, 1969), the visual features discriminated by pervasive grammatical distinctions have to be reliably discriminated. We propose that it is this necessity that is reflected in the higher sensitivity of the Korean speakers for visual features discriminated by grammaticalized distinctions. This is also in line with previous work: Choi and Hattrup (2012) showed heightened sensitivity in Korean speakers (compared to English speakers) to the tight/loose distinction in a nonlinguistic similarity judgment task involving real objects. More recently, Yun and Choi (2018) found a significant correspondence between language-specific spatial semantics and nonlinguistic sensitivity levels for relevant spatial features (e.g., tight-/loose-fit in Korean, containment in English).

In so far as the repeated classification of objects along the lines of a required grammaticalized distinction is initially probably a deliberately executed task, for which linguistic representations are held in working memory, we can also understand how these long-term effects of linguistic relativity are just a consequence of what researchers have emphasized when pointing out the importance of task relevance and flexibility when it comes to interactions between language and cognition (Huettig et al., 2011; Landau et al., 2010; Munnich et al., 2001). The long-term memory effects that we have studied here are the results of the repeated necessity to frequently and reliably discriminate between stimuli which correspond to grammaticalized distinctions.

Another but less likely possibility is that language merely influences inter-trial priming of capture of the target-similar 3D-fit cues in trial *n* (by language's boosting of the memory of target 3D-fits in the preceding trial *n-1*) (Huang, Holcombe, & Pashler, 2004). Though this might account for the findings of Experiment 1, this could not explain results of Experiment 4, where Korean speakers showed bottom-up capture of attention by fit singletons that were never similar to the preceding targets. Hence, it seems that Korean speakers' sensitivity towards tight- and loose- 3D-fits is generally heightened (cf. Choi & Hattrup, 2012), demonstrating significant influence of language-specific grammar on visual processing (cf. Meteyard, Bahrami, & Vigliocco, 2007).

Although future research is necessary to understand the exact (neuronal) mechanisms underlying linguistic long-term memory effects on attention capture, our main conclusion is clear: The current data provides evidence for spatial language permeating into nonlinguistic cognitive domains – here, visual attention – even without explicit instructions.

Acknowledgments

This research is supported by the Vienna Science and Technology Funds (WWTF) Grant CS15-001 to Ulrich Ansorge and Soonja Choi.

We would like to thank Sungwha Choi, Alexandra Kroiss, Daniel Mitic, Effrosyni Ntemou, and Hongoak Yun for their help with the data collection.

References

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language, 38*(4), 419–439.

Athanasopoulos, P., & Bylund, E. (2013). Does grammatical aspect affect motion event cognition? A cross-linguistic comparison of English and Swedish speakers. *Cognitive Science, 37*(2), 286–309.

Athanasopoulos, P., Bylund, E., Montero-Melis, G., Damjanovic, L., Schartner, A., Kibbe, A., ... Thierry, G. (2015). Two languages, two minds: Flexible cognitive processing driven by language of operation. *Psychological Science, 26*(4), 518–526.

Baier, D., & Ansorge, U. (2019). Investigating the role of verbal templates in contingent capture by color. *Attention, Perception, & Psychophysics* Advance online publication.

Boroditsky, L. (2000). Metaphoric structuring: Understanding time through spatial metaphors. *Cognition, 75*(1), 1–28.

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

Bowerman, M., & Choi, S. (2001). Shaping meanings for language: Universal and language specific in the acquisition of spatial semantic categories. In M. Bowerman, & S. Levinson (Eds.). *Language acquisition and conceptual development* (pp. 475–511). Cambridge: Cambridge University Press.

Bowerman, M., & Choi, S. (2003). Space under construction: Language-specific spatial categorization in first language acquisition. In D. Gentner, & S. Goldin-Meadow (Eds.). *Language in mind: Advances in the study of language and cognition* (pp. 387–428). Cambridge, MA: MIT Press.

Brown, T. L., Gore, C. L., & Pearson, T. (1998). Visual half-field Stroop effects with spatial separation of words and color targets. *Brain and Language, 63*(1), 122–142.

Büsel, C., Pomper, U., & Ansorge, U. (2018). Capture of attention by target-similar cues during dual-color search reflects reactive control among top-down selected attentional control settings. *Psychonomic Bulletin & Review* Advance online publication.

Carmel, T., & Lamy, D. (2014). The same-location cost is unrelated to attentional settings: An object-updating account. *Journal of Experimental Psychology: Human Perception and Performance, 40*, 1465–1479.

Carmel, T., & Lamy, D. (2015). Towards a resolution of the attentional-capture debate. *Journal of Experimental Psychology: Human Perception and Performance, 41*, 1772–1783.

Casasola, M. (2005). Can language do the driving? The effect of linguistic input on infants' categorization of support spatial relations. *Developmental Psychology, 41*(1), 183–192.

Casasola, M., Cohen, L. B., & Chiarello, E. (2003). Six-month-old infants' categorization of containment spatial relations. *Child Development, 74*(3), 679–693.

Choi, S. (2006). Influence of language-specific input on spatial cognition: Categories of containment. *First Language, 26*, 207–232.

Choi, S., & Hattrup, K. (2012). Relative contribution of cognition/perception and language on spatial categorization. *Cognitive Science, 36*, 102–129.

Cooper, R. M. (1974). The control of eye fixation by the meaning of spoken language: A new methodology for the real-time investigation of speech perception, memory, and language processing. *Cognitive Psychology, 6*(1), 84–107.

Flecken, M., von Stutterheim, C., & Carroll, M. (2014). Grammatical aspect influences motion event perception: Findings from a cross-linguistic non-verbal recognition task. *Language and Cognition, 6*, 45–78.

Folk, C. L., Remington, R. W., & Johnston, J. C. (1992). Involuntary covert orienting is contingent on attentional control settings. *Journal of Experimental Psychology: Human Perception and Performance, 18*, 1030.

Franklin, A., Drivonikou, G. V., Clifford, A., Kay, P., Regier, T., & Davies, I. R. L. (2008). Lateralization of categorical perception of color changes with color term acquisition. *Proceedings of the National Academy of Sciences of the United States of America, 105*,

18221–18225.
Gilbert, A. L., Regier, T., Kay, P., & Ivry, R. B. (2006). Whorf hypothesis is supported in the right visual field but not the left. *Proceedings of the National Academy of Sciences of the United States of America, 103*, 489–494.
Gleitman, L., & Papafragou, A. (2013). Relations between language and thought. In D. Reisberg (Ed.). *Oxford handbook of cognitive psychology* (pp. 504–523). New York, NY: Oxford University Press.
Goller, F., Choi, S., Hong, U., & Ansorge, U. (2019). Whereof one cannot speak: How language and capture of visual attention interact: dataset. [figshare dataset] https://doi.org/10.6084/m9.figshare.832026. [dataset].
Goller, F., Lee, D., Ansorge, U., & Choi, S. (2017). Effects of language background on gaze behavior: A crosslinguistic comparison between Korean and German speakers. *Advances in Cognitive Psychology, 13*(4), 267.
Hespos, S. J., & Spelke, E. S. (2004). Conceptual precursors to language. *Nature, 430*(6998), 453–456.
Hout, M. C., Robbins, A., Godwin, H. J., Fitzsimmons, G., & Scarince, C. (2017). Categorical templates are more useful when features are consistent: Evidence from eye movements during search for societally important vehicles. *Attention, Perception, & Psychophysics, 79*, 1578–1592.
Huang, L., Holcombe, A. O., & Pashler, H. (2004). Repetition priming in visual search: Episodic retrieval, not feature priming. *Memory & Cognition, 32*(1), 12–20.
Huettig, F., & Altmann, G. T. M. (2005). Word meaning and the control of eye fixation: Semantic competitor effects and the visual world paradigm. *Cognition, 96*(1), B23–B32.
Huettig, F., Olivers, C. N., & Hartsuiker, R. J. (2011). Looking, language, and memory: Bridging research from the visual world and visual search paradigms. *Acta Psychologica, 137*(2), 138–150.
Irons, J. L., Folk, C. L., & Remington, R. W. (2012). All set! Evidence of simultaneous attentional control settings for multiple target colors. *Journal of Experimental Psychology: Human Perception and Performance, 38*(3), 758–775.
Kerzel, D., & Witzel, C. (2019). The allocation of resources in visual working memory and multiple attentional templates. *Journal of Experimental Psychology: Human Perception and Performance, 45*(5), 645–658.
Knoeferle, P., & Crocker, M. W. (2006). The coordinated interplay of scene, utterance, and world knowledge: Evidence from eye tracking. *Cognitive Science, 30*(3), 481–529.
Lakoff, G., & Johnson, M. (2008). *Metaphors we live by.* Chicago, IL: University of Chicago press.
Landau, B., Dessalegn, B., & Goldberg, A. M. (2010). Language and space: Momentary interactions. In V. Evans, & P. Chilton (Eds.). *Language, cognition, and space: The state of the art and new directions* (pp. 51–78). London: Equinox Publishing.
Levin, D. T., Takarae, Y., Miner, A. G., & Keil, F. (2001). Efficient visual search by category: Specifying the features that mark the difference between artifacts and animals in preattentive vision. *Perception & Psychophysics, 63*(4), 676–697.
Levinson, S. C., Meira, S., & the Language and Cognition Group (2003). 'Natural concepts' in the spatial topological domain – adpositional meanings in crosslinguistic perspective: an exercise in semantic typology. *Language, 79*(3), 485–516.
Li, P., & Gleitman, L. (2002). Turning the tables: Language and spatial reasoning. *Cognition, 83*, 265–294.
Lupyan, G. (2012). Linguistically modulated perception and cognition: The label-feedback hypothesis. *Frontiers in Psychology, 3*, 54.
Lupyan, G., & Spivey, M. J. (2008). Perceptual processing is facilitated by ascribing meaning to novel stimuli. *Current Biology, 18*(10), R410–R412.
Lupyan, G., & Spivey, M. J. (2010). Redundant spoken labels facilitate perception of multiple items. *Attention, Perception, & Psychophysics, 72*, 2236–2253.
Majid, A., Bowerman, M., Kita, S., Haun, D. B., & Levinson, S. C. (2004). Can language restructure cognition? The case for space. *Trends in Cognitive Sciences, 8*(3), 108–114.
Malt, B., Sloman, S., Gennari, S., Shi, M., & Wang, Y. (1999). Knowing versus naming: Similarity and the linguistic categorization of artifacts. *Journal of Memory and Language, 40*, 230–262.
Maxfield, J. T., Stalder, W. D., & Zelinsky, G. J. (2014). Effects of target typicality on categorical search. *Journal of Vision, 14*(12) 1-1.
McClelland, J. L., & Rumelhart, D. E. (1985). Distributed memory and the representation of general and specific information. *Journal of Experimental Psychology: General, 114*(2), 159–188.
McDonough, L., Choi, S., & Mandler, J. (2003). Understanding spatial relations: Flexible infants, lexical adults. *Cognitive Psychology, 46*, 229–259.
Meteyard, L., Bahrami, B., & Vigliocco, G. (2007). Motion detection and motion verbs: Language affects low-level visual perception. *Psychological Science, 18*(11), 1007–1013.
Meyer, A. S., Belke, E., Telling, A. L., & Humphreys, G. W. (2007). Early activation of object names in visual search. *Psychonomic Bulletin & Review, 14*(4), 710–716.
Miller, T. M., Schmidt, T. T., Blankenburg, F., & Pulvermüller, F. (2018). Verbal labels facilitate tactile perception. *Cognition, 171*, 172–179.
Moore, K. S., & Weissman, D. H. (2010). Involuntary transfer of a top-down attentional set into the focus of attention: Evidence from a contingent attentional capture paradigm. *Attention, Perception, & Psychophysics, 72*, 1495–1509.
Müller, H. J., Heller, D., & Ziegler, J. (1995). Visual search for singleton feature targets within and across feature dimensions. *Perception & Psychophysics, 57*(1), 1–17.
Munnich, E., Landau, B., & Dosher, B. A. (2001). Spatial language and spatial representation: A cross-linguistic comparison. *Cognition, 81*(3), 171–208.
Najemnik, J., & Geisler, W. S. (2005). Optimal eye movement strategies in visual search. *Nature, 434*(7031), 387–391.
Nisbett, R. E., & Miyamoto, Y. (2005). The influence of culture: Holistic versus analytic perception. *Trends in Cognitive Sciences, 9*(10), 467–473.
Özyürek, A., & Kita, S. (1999). Expressing manner and path in English and Turkish: Differences in speech, gesture, and conceptualization. In M. Hahn, & S. C. Stones (Eds.). *Twenty-first annual conference of the cognition science society* (pp. 507–512). Mahwah, NJ: Lawrence Erlbaum Associates.
Papafragou, A., Massey, C., & Gleitman, L. (2002). Shake, rattle, 'n' roll: The representation of motion in language and cognition. *Cognition, 84*, 189–219.
Peirce, J. W. (2007). PsychoPy—psychophysics software in Python. *Journal of Neuroscience Methods, 162*(1–2), 8–13.
Peirce, J. W. (2009). Generating stimuli for neuroscience using PsychoPy. *Frontiers in Neuroinformatics, 2*, 10.
Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology, 32A*(1), 3–25.
Robbins, A., & Hout, M. C. (2015). Categorical target templates: Typical category members are found and identified quickly during word-cued search. *Visual Cognition, 23*(7), 817–821.
Searle, J. R. (1969). *Speech acts. An essay in the philosophy of language.* Cambridge, England: Cambridge University Press.
Soroli, E. (2012). Variation in spatial language and cognition: Exploring visuospatial thinking and speaking cross-linguistically. *Cognitive Processing, 13*(S1), 333–337.
Soroli, E., & Hickmann, M. (2010). Spatial language and cognition in French and English: some evident from eye-movements. In G. Marotta, A. Lenci, L. Meini, & F. Rovai (Eds.). *Space in language: Proceedings of the Pisa international conference* (pp. 581–597). Pisa, Italy: University of Pisa.
Soto, D., & Humphreys, G. W. (2007). Automatic guidance of visual attention from verbal working memory. *Journal of Experimental Psychology: Human Perception and Performance, 33*, 730–757.
Spivey, M. J., Tanenhaus, M. K., Eberhard, K. M., & Sedivy, J. C. (2002). Eye movements and spoken language comprehension: Effects of visual context on syntactic ambiguity resolution. *Cognitive Psychology, 45*(4), 447–481.
Spivey, M. J., Tyler, M. J., Eberhard, K. M., & Tanenhaus, M. K. (2001). Linguistically mediated visual search. *Psychological Science, 12*(4), 282–286.
Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science, 268*(5217), 1632–1634.
Theeuwes, J. (1992). Perceptual selectivity for color and form. *Perception & Psychophysics, 51*, 599–606.
Theeuwes, J. (2013). Feature-based attention: It is all bottom-up priming. *Philosophical Transactions of the Royal Society B: Biological Sciences, 368*(1628), 20130055.
Theeuwes, J., Atchley, P., & Kramer, A. F. (2000). On the time course of top-down and bottom-up control of visual attention. In S. Monsell, & J. Driver (Eds.). *Control of cognitive processes: Attention and performance XVIII* (pp. 105–125). Cambridge, MA: MIT Press.
Thierry, G., Athanasopoulos, P., Wiggett, A., Dering, B., & Kuipers, J. R. (2009). Unconscious effects of language-specific terminology on preattentive color perception. *Proceedings of the National Academy of Sciences of the United States of America, 106*, 4567–4570.
Walenchok, S. C., Hout, M. C., & Goldinger, S. D. (2016). Implicit object naming in visual search: Evidence from phonological competition. *Attention, Perception, & Psychophysics, 78*, 2633–2654.
Weekes, N. Y., & Zaidel, E. (1996). The effects of procedural variations on lateralized Stroop effects. *Brain and Cognition, 31*(3), 308–330.
Weichselbaum, H., & Ansorge, U. (2018). Bottom-up attention capture with distractor and target singletons defined in the same (color) dimension is not a matter of feature uncertainty. *Attention, Perception, & Psychophysics, 80*, 1350–1361.
Whorf, B. L. (1964). *Language, thought, and reality.* Boston, MA: MIT Press.
Winawer, J., Witthoft, N., Frank, M. C., Wu, L., Wade, A. R., & Boroditsky, L. (2007). Russian blues reveal effects of language on color discrimination. *Proceedings of the National Academy of Sciences of the United States of America, 104*, 7780–7785.
Witzel, C., & Gegenfurtner, K. R. (2011). Is there a lateralized category effect for color? *Journal of Vision, 11*(12), 16.
Yu, C. P., Maxfield, J. T., & Zelinsky, G. J. (2016). Searching for category-consistent features: A computational approach to understanding visual category representation. *Psychological Science, 27*, 870–884.
Yun, H., & Choi, S. (2018). Spatial semantics, cognition, and their interaction: A comparative study of spatial categorization in English and Korean. *Cognitive Science.* https://doi.org/10.1111/cogs.12622.
Zelinsky, G. J. (2008). A theory of eye movements during target acquisition. *Psychological Review, 115*, 787–835.
Zhou, K., Mob, L., Kay, P., Kwok, V. P. Y., Ip, T. N. M., & Tan, L. H. (2010). Newly trained lexical categories produce lateralized categorical perception of color. *Proceedings of the National Academy of Sciences of the United States of America, 107*, 9974–9978. https://doi.org/10.1073/pnas.1005669107.