

If John is taller than Jake, where is John? Spatial inference from magnitude comparison

William J. Skylark

Department of Psychology
University of Cambridge

Please address correspondence to:

William Skylark
Department of Psychology
University of Cambridge
Cambridge
CB2 3EB
United Kingdom

email: wjm22@cam.ac.uk
Telephone: +44 1223 333572

© 2017, American Psychological Association. This paper is not the copy of record and may not exactly replicate the final, authoritative version of the article. Please do not copy or cite without authors permission. The final article will be available, upon publication, via its DOI:
10.1037/xlm0000505

Abstract

We regularly compare magnitudes and describe these comparisons to other people. This paper reports 9 experiments which examine how messages about the relative magnitude of two items affect inferences about the items' spatial arrangement. Native English speakers were given sentences such as "One tree is taller than the other", and their beliefs about the left-right arrangement of the objects were probed. Across a wide range of dimensions and tasks, the choice of comparative shaped spatial inference: "smaller" comparatives (e.g., "shorter", "lighter", "less") led to the belief that the small item was on the left, whereas "larger" comparatives (e.g., "longer", "heavier", "more") led to the belief that the small item was on the right. These inferences match the tendency of message-senders to choose comparatives based on spatial layout, such that purely ordinal magnitude comparisons communicate information about the spatial arrangement of the compared objects. There was also evidence for a canonical "small-large" inference, consistent with the tendency of English speakers to associate "small" with "left" and "large" with "right"; however, this effect was task-dependent, indicating a flexible, language-based mapping rather than an immutable bias. Finally, there was evidence that the choice of comparative influenced the salience of particular response options. These results help to elucidate the deep inter-connections between language, space, and magnitude: linguistic tokens and structures reflect physical reality, and in turn shape mental representations of the physical world.

Keywords: relative judgment; magnitude comparison; language; spatial inference

If John is taller than Jake, where is John? Spatial inference from magnitude comparison

What do you infer when someone tells you that John is taller than Jake? The primary information conveyed by comparative adjectives such as “taller”, “bigger”, or “heavier”, concerns the relative placement of two entities on a single magnitude dimension – area, weight, height, and so forth. Because relative magnitude is often more important than absolute value (e.g., Boyce, Brown, & Moore, 2010; Matthews, 2013; Ono & Kawahara, 2007; Stewart, Chater, Stott, & Reimers, 2003), and relative judgment is better than absolute judgment (e.g., Laming, 1997; Mussweiler & Epstude, 2009), these kinds of comparative convey vital information about the world.

However, for a given dimension, English speakers may communicate relative magnitude by using a “smaller” adjective (e.g., smaller, shorter, lower, lighter, less) or a “larger” one (e.g., larger, longer, higher, heavier, more). These antonymic pairs convey the same ordinal information, but in English (and other languages), the “smaller” adjectives are marked, such that they are less common and harder to process (Allan, 1986; H.H. Clark, 1969; Hoorens & Bruckmüller, 2015; Matthews & Dylman, 2014). Moreover, they tend to be restricted to the lower end of the scale: people are more likely to use “smaller” adjectives when the compared items are both small than when they are both large (Matthews & Dylman, 2014), and tend to infer larger absolute magnitudes for items that are described with a “larger” comparative (e.g., Choplin, 2010; Holyoak & Mah, 1981; Rusiecki, 1985). Comparative adjectives therefore communicate more than just ordinal magnitude: the message-sender’s choice of comparative guides the message-receiver’s inferences about absolute stimulus magnitude.

The current studies test whether comparative adjectives also shape inferences about *spatial* relations. The hypothesis that descriptions of relative magnitude influence English speakers’ mental reconstruction of spatial layout was motivated by two considerations. The first concerns the spatiotemporal structure of English. For sentences that describe spatial or temporal relations, the word sequence shapes people’s mental representation of the structure of the world (e.g., H.H. Clark & E.V. Clark, 1968). For example, “the hammer is between the saw and the drill” is consistent with both saw-hammer-drill and drill-hammer-saw, but German-speakers were better at reasoning problems when the solution required the former layout, where the spatial order matches the sequence of the words in the premise (Jahn, Knauff, & Johnson-Laird, 2007). Similarly, when Spanish speakers were told that “the table is between the lamp and the TV” and asked to draw the scene, they typically produced lamp-table-TV rather than TV-table-lamp (Román, El Fathi, & Santiago, 2013; see also Román, Flumini, Lizano, Escobar, & Santiago, 2015). These demonstrations of *order matching* used sentences that explicitly describe spatial relations, but a similar principle might apply to inferences based on comparative adjectives because the choice of comparative dictates the order

in which objects are referenced: sentences such as “one thing is smaller than the other”, or “Jake is shorter than John”, reference the small item first and, in writing, to the left of the large object; swapping the comparative for “larger” reverses this sequence (“John is taller than Jake” has the larger man first/on the left). Thus, if people match spatial order to word order then the choice of comparative will correspondingly affect spatial inference.

The second motivation for the current work comes from recent studies of language production. Matthews and Dylman (2014) showed English-speakers pairs of objects and asked them to describe their relative magnitudes by completing sentences such as “One square is _____ than the other” (Figure 1a shows some example stimuli). Across a range of dimensions and several studies, participants were more likely to use larger comparatives than smaller ones, but this *higher use of larger comparatives* (HULC) effect was modulated by the spatial layout of the items: the use of larger comparatives was greater when the large object was on the left, consistent with participants matching the order in which the small- and large-objects are referenced in the sentence with their left-right arrangement in the world. The top panel of Figure 2 shows illustrative data. In preparation for the current experiments, I replicated this effect in a preliminary study that used pairs of objects that differed on more than just the focal dimension, which permits a richer type of comparative sentence (e.g., “The circle is larger than the square” or “The square is smaller than the circle”). The study is described in the Appendix and example stimuli are given in Figure 1b; the bottom panel of Figure 2 shows the results. Again, there is an overall HULC effect that is modulated by the left-right arrangement of the objects: people were more likely to say “Jake is taller than John” when Jake is on the left than when he is on the right, for example.

Taken together, these studies suggest that message-receivers may infer that the first-mentioned item is to the left of the item which is mentioned second, such that “smaller” comparatives imply a small-large layout whereas “larger” comparatives imply a large-small arrangement. Indeed, given that message-senders’ choice of comparative is contingent on layout, such order-matching by receivers would amount to the successful transmission of spatial information by a non-spatial magnitude comparison.

I report 9 experiments that test this possibility, using a variety of different tasks, stimuli, and attribute-dimensions. The basic approach is to present participants with comparative sentences that have previously been produced by people who were comparing two items, and to establish the spatial layout that the message-receivers infer from this comparison. In Studies 1 and 2, participants read sentences of the form “One X is [smaller/larger] than the other” and judged which of two spatial arrangements are being described; the studies find that the choice of comparative shapes spatial inference in both within- and between-subject designs. Studies 3 and 4 test whether this effect generalizes to spoken sentences and whether it requires a full comparative sentence, or

whether an isolated adjective is sufficient to affect spatial representations. Studies 5-7 probe spatial inference in a different way by asking people which of two items was in a particular location, or which location was occupied by a particular object. Finally, Studies 8 and 9 clarify the contribution of default or “canonical” layouts by asking participants which of two layouts a person saw, but without providing a comparative sentence.

Studies 1 and 2

In Studies 1 and 2, participants were presented with sentences such as “One tree is taller than the other” or “One tree is shorter than the other”, and indicated which of two images the speaker had seen – one where the shorter tree was on the left, or one where it was on the right. To ensure that the comparatives correspond to natural language descriptions of the presented objects, the visual stimuli were those used in the sentence production tasks of Matthews and Dylman (2014). Recall that these authors required participants to choose the word they would naturally use to complete sentences of the form “One X is _____ than the other”, where the choice of comparative is appropriate to the dimension of interest (e.g., “shorter” / “taller” when comparing the heights of two trees). In the current studies, participants were given these sentences – using either the modal (i.e., most common) “smaller” or modal “larger” comparative that the original participants had produced for each dimension – and asked which of two spatial layouts the message-sender had seen: small-large or large-small. To ensure generality, Studies 1 and 2 used different stimuli/dimensions and different designs: in Study 1 the type of comparative (smaller vs larger) was randomized for each dimension within-subjects; in Study 2, comparative condition was a between-subjects factor.

Method

Participants. All studies were run online using participants recruited via Amazon’s Mechanical Turk. Across all studies, only the first occurrence of a given ip address was included in the data set (e.g., Matthews, Gheorghiu, & Callan, 2016). Among these non-duplicate responders, responses were judged eligible if they completed the task, stated their age as 18 or over and their first language (“mother tongue”) as English, did not report problems viewing/hearing the stimuli, and did not self-report having started the task more than once or having taken part in an earlier study in the series. The demographics/debriefing questions at the end of each study included an attention-check that asked which one of several items had not been shown/mentioned/asked about during the task; participants who failed this were excluded prior to analysis. The studies received ethical approval from the University of Essex Faculty of Science and Health Ethics Committee.

The primary analyses used mixed-effects regression, for which power calculations can be difficult. Sample sizes were therefore chosen to give at least 80% power (based on “medium” effects, $d = 0.5$, or the size of the effects found in earlier studies in the series) in simple between-subject t -tests for contrasts between conditions. Because of the vagaries of on-line recruitment (e.g., the difficulty of knowing how many duplicate ips will appear) and because I was worried about type II errors, I consistently requested more participants than were needed and “overshot” the sample required for 80% power.

For Study 1, $N = 183$ (75 female, ages 19-69, $M = 32.9$, $SD = 10.2$; a further 16 failed the attention check); for Study 2, $N = 167$ (84 female, ages 19-65, $M = 34.6$, $SD = 10.1$; a further 28 failed the attention check).

Stimuli, Design, and Procedure. In Study 1, participants were told that on each page they would see a sentence that someone has used to describe two objects and that they should indicate which of two images the person saw, basing their answer on the sentence that the person had used to describe the relation between them. Participants completed 8 trials, one per dimension (here and throughout, dimension order was randomized for each participant). Each trial presented an appropriate comparative sentence (e.g., “The person said: One square is [smaller/larger] than the other”) followed by two images, one showing a small-large arrangement of the objects (e.g., a small square on the left, a large square on the right) and the other showing a large-small arrangement. The two images were placed one above the other, with top/bottom location randomized for each trial, and participants clicked a radio button next to the image that they thought the person had seen. Whether the sentence used “smaller” or “larger” was randomized for each trial. The stimuli and sentences were taken from Matthews and Dylman’s (2014) Study 2a; examples are presented in Figure 1a and full details are listed in Table 1.

Study 2 was identical except that (a) the visual stimuli were taken from the “large pair” condition of Matthews & Dylman (2014) Study 4c and the smaller/larger comparative sentences were based on the modal smaller/larger responses to these items in that study (see Table 1), and (b) comparative condition was manipulated wholly between-subjects, such that each participant was randomly assigned to read either all “smaller” or all “larger” sentences ($N_{\text{smaller}} = 91$; $N_{\text{larger}} = 76$).

Results

Figure 3 shows the proportion of participants in Studies 1 and 2 who selected the small-large arrangement for each dimension. For all dimensions, the majority of participants selected a small-large layout when the comparative was “smaller” and a large-small layout when the comparative was larger: when told (for example) that “One square is bigger than the other”, participants inferred

that the large square was on the left; when told “One square is smaller than the other”, they judged that the large square had been on the right.

I analysed these and all subsequent data with mixed-effects logistic regressions that included random intercepts for participants and dimensions and full random slopes (i.e. for within-subject predictors, random slopes within participant and dimension; for between-subject variables, random slopes within dimensions; random effects were uncorrelated). I report fixed effects coefficients with 95% Wald confidence intervals. For these studies, inferred layout (small-large = 1, large-small = 0) was the dependent variable and condition (“smaller” = -1, “larger” = +1) was the independent variable. These analyses confirmed the pattern evident in Figure 3: participants were far more likely to infer a small-large layout in the “smaller” condition (for Study 1, $B_{\text{cond}} = -2.677$, $CI = [-3.202, -2.151]$, $p < .001$; for Study 2, $B_{\text{cond}} = -2.234$, $CI = [-2.730, -1.738]$, $p < .001$). Notably, there was no overall bias towards either layout (Study 1: $B_{\text{int}} = -0.059$, $CI = [-0.309, 0.192]$, $p = .646$; Study 2: $B_{\text{int}} = 0.167$, $CI = [-0.278, 0.614]$, $p = .463$).

Discussion

Spatial inferences were strongly influenced by the choice of comparative: people who learned that “One X was smaller than the other” typically inferred a small-large layout whereas those told “One X is larger than the other” imagined a large-small layout. That is, people chose a spatial layout in which the left-right order of the objects matched their left-right sequence in the written sentence.

Study 3

An obvious question is whether the order-matching found in Studies 1 and 2 is explicitly visual: is it dependent on (or facilitated by) the ability to see the spatial correspondence between word locations and object locations? Study 3 tested this in two ways. First, it introduced a “text with delay” condition, in which the comparative sentence and the spatial choice task were presented on separate pages. **In Studies 1 and 2, the comparative sentence was presented on the same page as the images depicting the two candidate spatial layouts, which permitted direct visual mapping between the left-right layout of the objects and the left-right layout of the words (that is, participants could look back and forth between the words of the comparative sentence and the objects in each image). Splitting the sentences and the choice task across separate pages prevents such visual mapping, and might therefore be expected to reduce the effect of comparative adjective (and hence word order) on spatial inference.** Second, and more important, the study introduced an auditory condition, in which the comparative sentence was presented aurally.

Methods

Participants. A sample of 416 (239 female, ages 19-79, $M = 38.7$, $SD = 12.8$) completed the study (a further 4 failed the attention check).

Stimuli, Design, and Procedure. Participants were randomly assigned to one of three conditions. The “Text-No Delay” condition ($N = 139$) was identical to Study 1. In the “Text-Delay” condition ($N = 129$), the comparative sentence was on one page and the participant only saw the two candidate layouts when he/she advanced to the next page. In the “Auditory” condition ($N = 148$), the first page presented an embedded audio file; participants clicked to hear the comparative sentence, which was spoken by a neutral computer-generated female voice. Participants could listen as many times as they wished (analogous to being able to read the written sentence at their leisure) before progressing to the next page, which gave the two layout options one above the other as before.

Results

The top panel of Figure 4 plots the proportion of “small-large” and “large-small” inferences for each condition. Two regressions were run to make two planned comparisons. First, the effects of a delay were tested by comparing the Text-Delay (coded 1) and Text-No-Delay (coded -1) conditions. There was no overall response bias, $B_{\text{int}} = -0.008$, $CI = [-0.298, 0.281]$, $p = .954$, but a strong effect of comparative adjective: participants were far more likely to infer a small-large layout when the adjective was “smaller” than when it was “larger”, $B_{\text{cond}} = -2.041$, $CI = [-2.396, -1.686]$, $p < .001$. There was no overall effect of a delay between comparative sentence and visual display of the two possible layouts, $B_{\text{delay}} = 0.020$, $CI = [-0.171, 0.211]$, $p = .839$, and the delay did not modulate the effect of the comparative adjective, $B_{\text{interaction}} = -0.230$, $CI = [-0.534, 0.074]$, $p = .138$.

The second analysis compared the Text-Delay condition (coded -1) with the Auditory condition (coded +1). Again, there was no overall bias, $B_{\text{int}} = 0.021$, $CI = [-0.275, 0.317]$, $p = .890$ but people were more likely to infer a small-large layout for “smaller” comparatives, $B_{\text{cond}} = -2.553$, $CI = [-2.946, -2.160]$, $p < .001$. There was no overall effect of presentation modality, $B_{\text{modality}} = 0.047$, $CI = [-0.185, 0.280]$, $p = .688$, but modality did modulate the effects of comparative, with the difference between “smaller” and “larger” being more pronounced in the auditory condition than in the text condition, $B_{\text{interaction}} = -0.743$, $CI = [-1.056, -0.428]$, $p < .001$.

Discussion

As before, this study found no evidence of an overall tendency to infer a small-large layout but found strong evidence for an effect of comparative adjective on the inferred left-right layout of the original objects. This effect was robust against a brief delay between the comparative sentence

and the candidate object layouts, and generalized to spoken sentences – indeed, the effect was slightly stronger in this modality.

Order-matching is therefore not contingent on the message-receiver being able to simultaneously inspect the layout of the comparative sentence and the described-objects, or on the comparative sentence having an explicitly spatial structure. Rather, matching seems to happen at a more abstract level of representation. In particular, English speakers are thought to represent temporal order in terms of a left-right mapping, perhaps because of the left-right reading direction (e.g., Fuhrman & Boroditsky, 2010; Pérez, García, Valdés-Sosa, & Jaśkowski, 2011; Ulrich & Maienborn, 2010; Santiago, Lupiáñez, Pérez, & Funes, 2007). Correspondingly, when it comes to inferring the left-right layout of the original objects, the temporal directionality of “One square is smaller than the other” – that is, the temporal sequence in which the referenced objects occur – may serve the same function as the spatial directionality that comes with the written form of the sentence. This does not explain why the speech effect seems to be somewhat *stronger* than the written form, unless one makes post hoc assumptions such as greater engagement with the spoken materials and/or a deeper processing of the order information with that modality. **One interesting possibility, suggested by a Reviewer, is that the rapidity of reading serves to suppress the effect of spatial layout: with a spoken sentence, one has to hold the first object in memory while the sentence unfolds, perhaps thereby binding the representation to a particular spatiotemporal position for longer.** In any case, Study 3 shows that the effects of comparative adjective on spatial inferences are robust and likely to rest on abstract mental representations rather than the direct spatial correspondence.

Study 4

Studies 1-3 presented the comparative adjective in a complete sentence, where the small and large objects are explicitly referenced and occupy explicit positions in a spatial or temporal sequence of words. Study 4 asked whether such explicit labelling is required for the use of the comparative as the basis for spatial inference, or whether an isolated adjective (e.g., “smaller”, or “larger”) is sufficient to shape people’s beliefs about the layout of the objects whose magnitudes are being described.

Method

Participants. The sample comprised 168 participants (80 female, ages 20-74, $M = 36.1$, $SD = 11.8$; a further 50 failed the attention check).

Stimuli , Design, and Procedure. The experiment used a 2x2 design in which comparative (smaller vs larger) and format (single adjective vs sentence) were both manipulated between-

subjects. The sentence condition was identical to Study 2; the single adjective condition was the same except the comparative was presented as an isolated word (e.g., “The person said: “shorter”. Which of the following did they see?”).

Results

The bottom panel of Figure 4 plots the response proportions for each condition; as before, “smaller” comparatives elicited the belief that the small object was on the left and “larger” comparatives elicited the belief that it was on the right. The data were analysed with a mixed-effects logistic regression with condition (“smaller” = -1, “larger” = +1), format (word = -1, sentence = +1) and their interaction as predictors. There was no overall tendency to infer a small-large or large-small layout, $B_{\text{int}} = 0.407$, $CI = [-0.416, 1.230]$, $p = .333$, but there was a strong effect of comparative adjective, $B_{\text{cond}} = -3.499$, $CI = [-4.680, -2.317]$, $p < .001$. The format did not influence people’s inferences about the layout, $B_{\text{format}} = 0.465$, $CI = [-0.307, 1.237]$, $p = .238$, and, more importantly, did not modulate the effects of the comparative adjective, $B_{\text{interaction}} = -0.313$, $CI = [-1.073, 0.448]$, $p = .420$.

Discussion

Study 4 again found strong support for order-matching -- even though the comparative adjectives were presented in isolation, people responded as if they were mapping the left-right order of objects onto the left-right order in which the objects would be mentioned *if* they were described in a written comparative sentence, or to the order in which they would be uttered in a spoken comparative sentence. One possibility is that participants implicitly constructed a sentence around the comparative; alternatively, they may simply have learned through experience to associate “smaller” adjectives with a small-large layout, and vice-versa for “larger” comparatives.

Studies 5-7

The next experiments tested the generality of the foregoing results by probing people’s spatial inferences in different ways. Studies 1-4 tested inferences by presenting two possible layouts and asking which one the message-sender saw. Studies 5-7 instead asked participants which of two objects, one small and one large, was on the left or right of the original display, or by asked them to identify the location of a specified object.

Study 5 used the same kind of sentence as Studies 1-4 (e.g., “One square is smaller than the other”) and presented the two response options as visual images (e.g., “Which of these squares was on the left?”). Studies 6 and 7 used the more elaborate stimuli and comparative sentences introduced in the preliminary experiment -- e.g., “The circle is smaller than the square” (Figures 1

and 2). Because these sentences identify the objects by explicit labels, we can use these labels rather than the visual objects themselves when probing inference (e.g., “Which shape was on the left?” response options: “The circle” vs “The square” [Study 6] or “Where was the square?” response options: “On the left” vs “On the right” [Study 7]).

Methods

Participants. In Study 5, $N = 203$ (91 female, ages 19-73, $M = 36.0$, $SD = 11.7$; a further 21 failed the attention check). In Study 6, $N = 226$ (114 female, ages 19-78, $M = 36.1$, $SD = 11.9$; for this study, the debriefing questions probing attention and self-reported problems with image display were ambiguous so responses to these questions were ignored). Study 7 was run in two versions: the first had a smaller sample size than intended and produced a slightly ambiguous result for one contrast, so the second run sought to resolve these problems. [I also noticed a large number of attention-check failures in this first version (29 out of 141) and therefore made a minor change to reduce potential ambiguity in the second version, although in practice this made little difference (74 of 215 people failed); the relatively poor performance may reflect the fact that, by this point in the experimental sequence, quite a lot of participants were ineligible to sign up for the study because of past participation, and thus the participant pool may have comprised a somewhat less experienced/motivated population.] The combined sample comprised $N = 253$ (140 female, ages 18-66, $M = 36.7$, $SD = 12.1$).

Stimuli, Design, and Procedure. In Study 5, participants were told that they would read sentences that a person had used to describe the relation between two objects that had been presented side-by-side. They were asked to picture what the person was looking at and then to select which of two objects had occupied a particular location. For example: “The person saw the two squares shown below. The squares were shown side-by-side. The person said: “One square is [smaller/bigger] than the other”. Try to picture what the person saw. Which of the squares was on the [RIGHT/LEFT] of the screen?”, followed by the small square and large square, one above the other (location randomized). Participants completed 8 trials, one per dimension. The visual stimuli were taken from the “large jump” condition of Matthews & Dylman’s (2014) Study 4a and the smaller/larger comparative sentences were the modal smaller/larger responses to these items in that study (see Table 1 for details). Both the comparative (smaller vs larger) and side (left vs right) were randomized for each trial giving a 2x2 within-subject design.

Study 6 was the same, except that (a) the objects were the stimuli from the preliminary experiment of the current paper (see Appendix), (b) the sentences had the same form as the responses to the stimuli in that experiment (e.g., “The circle is larger than the square” or “The square is smaller than the circle”, with the smaller/larger adjectives being based on the modal

responses of each type), and (c) the two response options were identified by verbal labels rather than physical depictions of the two objects (e.g., people selected “The circle” or “The square”). Descriptions of the stimuli and comparatives for each dimension are listed in Table 2.

Study 7 was like Study 6, except that participants judged the location of the smaller or larger object, rather than indicating the object at a particular location. For example, they were asked “Where was the [circle/square]?” followed by the response options “On the left”/“On the right” arranged one above the other (with random assignment to location).

Results

Responses were coded according to whether they indicated a small-large or large-small layout. For example, in Studies 5 and 6, if participants chose the smaller item when asked “what was on the left?” or chose the large item when asked “what was on the right?”, their response implies a small-large inference and was coded 0; if the response indicated a large-small layout it was coded 1.

The left-hand panels of Figure 5 plot the proportion of responses that indicate a small-large layout for each experiment, collapsed across dimensions for simplicity (the pattern looks very similar when the dimensions are examined separately). For Studies 5 and 6, I analysed these data with mixed-effects logistic regressions with comparative adjective (“smaller” = -1, “larger” = +1), probed-side (“which was on the left” = -1, “which was on the right?” = +1) and their interaction as predictors; for Study 7, probed-side was replaced with probed-object (small object = -1, large object = +1); for this study the analysis included “version” and all its interactions as fixed-effects predictors (but not as random effects) to check for differences between the two waves of data collection; these terms were all trivially small, indicating consistency across the two runs of the study.

The right panels of Figure 5 plot the regression coefficients, labelled according to the theoretical effects to which they correspond (and signed such that a positive coefficient indicates positive evidence for the effect).

All three studies provide strong evidence of an order-matching effect (corresponding to the main effect of comparative condition): people were more likely to infer that the small item was on the left when they were told (for example) that “X is smaller than Y” than when told that “Y is larger than X”: for Study 5, $B = 0.561$, $CI = [0.300, 0.822]$, $p < .001$; for Study 6, $B = 1.012$, $CI = [0.805, 1.233]$, $p < .001$; for Study 7, $B = 0.536$, $CI = [0.400, 0.671]$, $p < .001$ (Version 1: $B = 0.547$, $CI = [0.303, 0.792]$, $p < .001$; Version 2: $B = 0.527$, $CI = [0.360, 0.695]$, $p < .001$).

Unlike Studies 1-4, these experiments also found that participants had an overall preference for inferring a small-large layout, as assessed by the intercept term: Study 5: $B = 0.358$, $CI = [0.164-0.553]$ $p < .001$; Study 6: $B = 1.159$, $CI = [0.905, 1.414]$, $p < .001$; Study 7: $B = 0.176$, $CI = [0.068,$

0.283], $p = .001$ (Version 1: $B = 0.235$, $CI = [0.054, 0.416]$, $p = .011$; Version 2: $B = 0.151$, $CI = [0.020, 0.282]$, $p = .024$). That is, participants inferred a default or canonical representation, in which the objects increase in magnitude from left to right.

In Studies 5 and 6, the main effect of “side” indicates whether participants were biased towards choosing the large item, irrespective of whether they were asked about the object on the left or the object on the right. Such a “Large bias”, if present at all, is not strong (Study 5: $B = 0.147$, $CI = [-0.006, 0.300]$, $p = .060$; Study 6: $B = 0.028$, $CI = [-0.091, 0.147]$, $p = .642$). Likewise, in Study 7 there was little indication of a bias towards responding “On the left”, $B = -0.060$, $CI = [-0.161, 0.040]$, $p = .238$ (Version 1: $B = -0.063$, $CI = [-0.237, 0.112]$, $p = .483$; Version 2: $B = -0.058$, $CI = [-0.188, 0.071]$, $p = .378$).

In Studies 5 and 6 the interaction coefficient indicates whether participants who were told that “One X is smaller than the other” were more likely to select the small object than those who were informed that “One X is larger than the other”; I label this a “magnitude relevance” effect, because it is as if the comparative is increasing the relevance or salience of the congruent response (e.g., “smaller” makes people more likely to choose the small object, irrespective of whether they are asked “what was on the left?” or “what was on the right?”; Sperber & Wilson, 1986). Such an effect was found in Study 5, $B = 0.841$, $CI = [0.652, 1.029]$, $p < .001$. However, this effect was trivially small in Study 6, $B = 0.063$, $CI = [-0.076, 0.203]$, $p = .375$. Pooling the data and including Study (coded -1 for Study 5, +1 for Study 6) and its interaction with all other variables as fixed-effects predictors confirmed that the magnitude-relevance effect was much larger in Study 5 than in Study 6, $B = -0.346$, $CI = [-0.452, -0.240]$, $p < .001$. In Study 7, the interaction tests for a “location relevance” effect, such that a “smaller” comparative makes people more likely to respond “on the left” whereas a “larger” comparative makes people more likely to respond “on the right”. Study 7 found no indication of such an effect, $B = -0.054$, $CI = [-0.159, 0.050]$, $p = .301$ (Version 1: $B = -0.098$, $CI = [-0.266, 0.070]$, $p = .252$; Version 2: $B = -0.038$, $CI = [-0.171, 0.096]$, $p = .581$).

Discussion

These studies required the message-receiver to engage in a greater degree of mental reconstruction when answering the spatial-inference questions: rather than simply choosing between two layouts, they had to select the object that occupied a named location, or the location occupied by a named object. Nonetheless, the order-matching effect remains strikingly apparent, indicating a robust and deep-seated tendency.

These studies also produced two other findings. First, Study 5 found evidence for a magnitude relevance effect: people were more likely to select the small object when the comparative was “smaller”, whereas Study 6 found no such effect. The difference between the

results likely arises from the different format of the response options: Study 5 presented the response options as visual objects, whereas Study 6 used verbal labels (e.g., “The circle” vs “The square”). With visual presentation of the two objects, the “largeness” of the larger item is very salient, and the congruency between the comparative and that object is readily-accessible. In contrast, the largeness of “The circle” exists only as a mental representation, inferred from the comparative sentence itself.

More importantly, Studies 5-7 demonstrated a canonical layout effect: irrespective of the comparative adjective, participants tended to infer a small-large spatial arrangement for the two objects. However, this robust and intriguing effect was not found in Studies 1-4. To probe this difference further, Studies 8 and 9 examined inferences when participants were not given any description of comparative magnitude.

Studies 8 and 9

Studies 8 and 9 respectively replicated Studies 1 and 5, but without giving participants a comparative sentence.

Method

Participants. I ran two versions of Study 8 because the initial sample was smaller than anticipated. (Again, I thought this may have been because of a slight ambiguity in the attention check and therefore made a minor change to this for Version 2, but it made little difference; 27 out of 115 and 45 out of 166 failed for Version 1 and 2, respectively). The final sample comprised $N=209$ (98 female, ages 19-87, $M = 35.9$, $SD = 13.0$).

For Study 9, $N = 108$ (55 female, ages 18-68, $M = 39.6$, $SD = 12.5$; a further 14 failed the attention check).

Stimuli, design and procedure. Study 8 was like Study 1 except that the comparative sentences were not presented. Thus, participants were told that “On each page you will see two objects that we have previously shown to other participants. You simply have to indicate which of two images the person saw.” Each trial then presented the small-large and large-small arrangements and asked “Which of the following did they see?”

Similarly, Study 9 replicated Study 5 but again withheld the comparative sentences so that participants were told, for example: “The person saw the two squares shown below. The squares were shown side-by-side. Try to picture what the person saw. Which of the squares was on the [LEFT/RIGHT] of the screen?”, whereupon they selected one of the two objects.

Results

Figure 6 shows the proportion of participants who selected the small-large arrangement for each dimension in Study 8. There is no indication of a preference for inferring a small-large layout; if anything, there is a slight trend in the opposite direction. A mixed-effects logistic regression that included Version (coded -1 for Version 1, +1 for Version 2) and random intercepts for participant and stimulus correspondingly indicated no substantial tendency to choose one layout rather than the other, $B_{\text{int}} = -0.286$, $CI = [-0.647, 0.079]$, $p = .125$, and no difference between the two runs of the study, $B_{\text{version}} = 0.059$, $CI = [-0.176, 0.294]$, $p = 0.623$. (Version 1: $B = -0.366$, $CI = [-0.790, 0.059]$, $p = .091$; Version 2: $B = -0.267$, $CI = [-0.656, 0.203]$, $p = .301$).

In contrast, Study 9 found an overall tendency to infer a small-large layout, $B_{\text{int}} = 0.697$, $CI = [0.424, 0.971]$, $p < .001$. This tendency was independent of whether people were asked about the item on the left (coded -1) or right (coded +1), $B_{\text{side}} = 0.115$, $CI = [-0.042, 0.272]$, $p = 0.151$.

To test the difference between the results of these two studies, I combined the data (collapsing over Version for Study 8 and probed-side for Study 9) and included Study (coded -1 for Study 8 and +1 for Study 9) as a fixed-effects predictor. This analysis revealed a weak overall tendency to infer a small-large layout, $B_{\text{int}} = 0.237$, $CI = [-0.004, 0.477]$, $p = 0.054$, but more importantly a substantial difference between the studies, $B_{\text{study}} = 0.0509$, $CI = [0.268, 0.750]$, $p < .001$.

Discussion

The absence of a default small-large inference in Studies 1-4, and its presence in Studies 5-7, were replicated in tasks where the participant judged the layout without being given any verbal description of the objects. Although there are a number of minor differences between the studies that produce a canonical small-large inference and those that do not, the most obvious difference is that the studies showing the effect are those in which the message-receiver was explicitly asked to consider locations – “what was on the left?” or “where was X? On the left or on the right?” -- rather than choosing between two visually-presented layouts. I consider the implications of this in the General Discussion.

General Discussion

We use other people’s linguistic productions to construct mental representations of the world. The present studies demonstrate a new aspect of this process by showing that descriptions of relative magnitude influence the inferences that English speakers make about spatial structure. The following discussion relates these findings to three aspects of the relationship between language, space, and magnitude: (1) the communication of spatiotemporal information via the spatiotemporal

structure of language; (2) the existence of “default” or “canonical” spatial representations of magnitude, and the role of language in shaping these space-magnitude associations; and (3) the role of language in shaping the perceived relevance of particular response options, and the dependence of such effects on the format of those options. In each case, I link the current results to existing data and theory, and outline key directions for future work.

Magnitude comparisons convey spatial information

Previous work has shown that the message-sender’s choice of magnitude comparative is influenced by the absolute size of the items (Matthews & Dylman, 2014) and that this choice in turn affects inferences about the sizes of unseen objects (Choplin, 2010; Rusiecki, 1985). The present results extend this principle to spatial inference: English speakers modulate their choice of comparative based on the left-right arrangement of magnitudes (Matthews & Dylman, 2014), and message-receivers use the sender’s choice of comparative as the basis for inference about object layout (Studies 1-7).

The effects of comparative adjectives on spatial inference held irrespective of whether the objects differed on one or multiple dimensions, whether the comparative sentence was of the form “One X is [smaller/larger] than the other” or “X is [smaller/larger] than Y”, and of whether participants chose between two layouts, chose which object had been in a particular location, or chose the location of a particular object. In addition, the effect was consistent across a wide range of stimuli and dimensions, including basic perceptual attributes (e.g., length), multiply-determined attributes (e.g., age), and values expressed with Arabic numerals (e.g., probability).

I have labelled this effect “order matching” because it is as if the message-receiver takes the order of reference in the comparative sentence to indicate the left-right layout of the objects in the world (cf Jahn et al., 2010; Román et al., 2013, 2015). However, the effect also arose with spoken sentences and isolated adjectives, indicating that English speakers have internalized the associations between particular comparatives and particular spatial layouts such that these comparatives guide the construction of spatial mental models even without the context of a written sentence.

These results support a general framework in which the structure of the world shapes the structure of language, so that these linguistic structures convey information about physical reality (e.g., A. Clark, 2006; Goodhew, McGaw, & Kidd, 2014; Louwerse, 2011). Previous work has shown that the message-sender’s choice of magnitude comparative is influenced by the absolute size of the items (Matthews & Dylman, 2014) and that this choice in turn affects inferences about the sizes of unseen objects (Choplin, 2010; Rusiecki, 1985). The present work extends this principle to spatial inference: English speakers modulate their choice of comparative based on the left-right arrangement of magnitudes (Preliminary Experiment; Matthews & Dylman, 2014), and message-

receivers use the sender's choice of comparative as the basis for inference about object layout (Studies 1-7).

However, the information-transmission is imperfect. For example, participants in the preliminary experiment were equally likely to see small-large and large-small layouts for each dimension, and overall they produced messages of the form "X is larger than Y" 38% of the time when the layout was small-large, and 94% of the time when it was large-small. The adjective is therefore informative regarding layout. In Study 6, participants received the modal "smaller" and "larger" versions of these sentences for each dimension with equal frequency, and inferred a small-large layout 71% of the time for the "smaller" sentences and 34% of the time for the "larger" sentences. Based on these values, the expected probability that the message-receiver will infer the true original layout is 60%, which represents reasonable performance given that the sender did not explicitly mention the left-right layout of the objects. However, the optimal strategy would be to infer small-large whenever the sender uses "smaller" and large-small whenever they use "larger", which would give 78% correct.

Transmission is limited because message-senders do not entirely base their choice of comparative on the spatial layout: there is a strong bias towards "larger" comparatives (the HULC effect), and choice of comparative is also modulated by other variables such as absolute magnitude. Correspondingly, the expected probability of correct inference for Study 1 is only 54%, but the main limitation here is that the message-senders (i.e., the participants in Matthews & Dylman, 2014, Study 2a) were only moderately affected by the spatial layout (Figure 2), which caps optimal performance at 57.5%. Likewise, receivers do not base their inferences entirely on the choice of comparative; this was especially true in Study 5, where there was essentially no transmission (only 51% expected proportion correct, again with a 57.5% maximum). This is not surprising given that, in this study, as well as the order-matching process participants based their responses a default small-large layout and on the congruence between the comparative and the response options (see below). Thus, ordinal comparisons communicate spatial information, but both the production and interpretation of comparative sentences are multiply determined; the choice of adjective partly reflects the spatial structure of the world – and in turn partly shapes the mental reconstruction of that structure – but this is only one facet of the "work" that language is doing in this context.

It will be important to test the effects of comparative on spatial inference in languages with different reading directions, and among bilingual speakers who have been "primed" to process in one direction or the other (e.g., Fischer, Shaki, & Cruise, 2009; Román et al., 2015; Zebian, 2005). I did not have access to appropriate participants, but this line of enquiry will be vital to establishing the basis for the order-matching effect found in the current studies. **Similarly, it would be interesting to manipulate the spatiotemporal presentation of text for English speakers – for example, by**

stacking the words vertically so that the sentence is read from top to bottom, or by presenting words one after the other at a central location (or perhaps at varying locations). More generally, it will be productive to test whether the effect is limited to left-right inference, or whether people infer other structures in the same way (for example, top-bottom layout and temporal order), and to see how the effects reported here generalize to scenes comprising more than two objects. An additional question concerns whether the spatial inferences examined here arise spontaneously, or only when people are asked about the layout of the objects. This is a difficult issue, because we can never know what mental representations have been constructed without in some way asking people to report them, but it might be possible to do some in less direct ways than were used in the present studies. In particular, future work might ask people to draw the scene that the message-sender is describing (e.g., Shaki et al., 2013), or give them physical objects to arrange and position in a way that recreates the original scene. A related conceptual issue concerns whether people's comprehension of comparative adjectives depends upon the construction of mental models in which spatial information is explicit (but perhaps based on arbitrary assumptions). That is, does the mental reconstruction of magnitudes from comparative descriptions require a mental reconstruction of their spatiotemporal arrangement?

A default “small-large” inference depends on linguistic labelling

In addition to the main finding, the current studies also found that English speakers sometimes have an overall bias to infer a small-large layout for objects that have been described purely in terms of relative magnitude. That is, just as people have default, canonical representations of individual objects (e.g., Blanz, Tarr, & Bülhoff, 1999; Matthews & Adams, 2008) and for the layout for complex scenes (Ehringer & Oliva, 2011), the current data suggest a canonical small-large layout for two objects that differ on a single magnitude dimension. However, the effect was task-dependent, emerging in Studies 5-7 and 9 but not in Studies 1-4 and 8.

The existence of a canonical small-large layout accords with other evidence that English speakers preferentially associate smaller items with “left” and larger ones with “right”. Much of this work originates with the “spatial-numerical association of response codes” (SNARC) effect (Dehaene, Bossini, & Giraux, 1993), wherein participants making parity (odd-even) judgments of numerical digits are faster when small digits require a left-hand button press and when large digits require a right-hand response than vice-versa. The tendency to associate low magnitude items with the left side of space has also been generalized to other dimensions (e.g., Fumarola, Prpic, Da Pos, Murgia, Umiltà, & Agostini, 2014; Ren, Nicholls, Ma, & Chen, 2011; Vallesi, Binns, & Shallice, 2008) and tasks (e.g., Calabria & Rossetti, 2005; Fischer, Castle, Dodd, & Pratt, 2003; Holmes, Ayzenberg, & Lourenco, 2016; see Fischer & Shaki, 2014, for a review).

These effects show considerable flexibility and task-sensitivity, leading to multiple theoretical accounts (e.g., Cheung & Lourenco, 2016; Dehaene et al., 1993; de Hevia, Izard, Coubart, Spelke, & Streri, 2014; Eikmeier, Alex-Ruf, Maienborn, & Ulrich, 2015; Fischer, Shaki, & Cruise, 2009; Fitousi, Shaki, & Algom, 2009; Gevers et al., 2010; Holmes & Lourenco, 2011; Hutchinson & Louwse, 2014; Lourenco & Longo, 2010; Patro & Shaki, 2016; Santens & Gevers, 2008; Shaki & Fischer, 2008; van Dijck & Fias, 2011; Zebian, 2005). The current data cannot arbitrate between the myriad accounts of the SNARC effect and its cousins, but those theories do help illuminate the current findings and the variation in the canonical-layout effect across studies. In particular, Gevers et al. (2010) distinguish between theories that emphasize a visuo-spatial mapping (for example, a left-right “mental number line”; e.g., Cheung & Lourenco, 2016) and those that emphasize verbal-spatial associations (e.g., Proctor & Cho, 2006). In the latter class of model, spatial-mapping effects are not due to internal continuous magnitude representations with a left-right orientation; rather, they are due to “an association between verbal concepts such as “small” and “left” and “large” and “right”” (Gevers et al., 2010, p. 181). For example, Gevers et al. (2006) developed a neural network model in which magnitude labels like “small” and “large” become associated with spatial labels (“left” and “right”) such that “small” automatically activates “left” and “large” activates “right”.

In the present work, the studies which found an overall small-large inference bias were those that used verbal labels to refer to spatial locations, either in the probe question (e.g., “Which shape was on the left?”; Studies 5, 6, 9) or in the response options (e.g., “Where was the square? On the left/On the right”; Study 7). In contrast, the studies which found no overall bias were those in which people chose between two visually-presented layouts, small-large and large-small, with no verbal labelling of locations (Studies 1-4 and 8). We must be cautious about accepting this dissociation without more controlled experimentation involving random-assignment. Nonetheless the consistency of the findings suggests that verbal-spatial mapping plays a role in spatial inference, such that a default “small-large” inference emerges when there is scope for the labels “left” and “right” to shape response selection – or perhaps more generally for any kind of “polarity-correspondence” between stimuli and responses (Proctor & Cho, 2006). This idea could be tested by future work that factorially manipulates the spatial labelling of locations within probe questions and across response options.

Comparatives influence response selection

A final observation is that, when participants had to choose which of two visually-presented objects had occupied a particular location, they showed a magnitude-relevance effect: “smaller” comparatives increased selection of the small item and “larger” comparatives increased selection of the large item, irrespective of the target location (Study 5). This effect could be driven by an

automatic process of activation, where the semantic overlap between the comparative (e.g., “larger”) and one of the response options (e.g., the “largeness” of the large square) activates or facilitates selection of that response (cf H.H. Clark, 1969). Alternatively, message-receivers might engage in active inference, treating the sender’s choice of comparative as a “hint” about the correct or desired response, in keeping with conversational norms (e.g., Grice, 1989). These alternatives could be disambiguated by manipulating the perceived informativeness of the sender’s message (e.g., Zhang & Schwarz, 2012).

In contrast to the effects with visually-presented stimuli, no relevance effects were found when the target objects were identified by verbal labels (e.g., “The circle” and “The square”; Studies 6 and 7). This difference echoes results from other studies of the interplay between word- and image-processing (e.g., Louwerse & Jeuniaux, 2010), and a straightforward explanation is that, with visual presentation, the perceptual properties of the two objects are directly activated: it is easy to identify the “larger” or “smaller” of two visible shapes, but with verbal labels such as “The circle” and “The square” the largeness (or smallness) of each item is not immediately apparent and must be accessed indirectly from a representation constructed during processing of the comparative sentence. Such a differential-accessibility account accords with dual-coding models of meaning (e.g., Paivio, 1986) and generates a straightforward prediction: if the response options were identified by verbal labels that explicitly referenced their size (e.g., “The large circle/The small square”) we would expect to see a resurgence of the relevance effect.

Conclusion

The relative sizes and locations of objects form fundamental information about the physical world, and describing these relations is a basic function of language. The current studies support the idea that the mental representations of magnitude and space are connected, in part, by linguistic codes (e.g., Clark, 2006; Dehaene et al., 1993; Gevers et al., 2006; Matthews & Meck, 2016; Román et al., 2013; Ulrich & Maienborn, 2010), and demonstrate the flexibility and adaptive value of these space-magnitude-language associations. These results are consistent with a general framework in which the spatiotemporal structure of language can be tuned to match the structure of the physical world. The result is a communication system which exploits redundancy, allowing a single message to convey multiple types of information: a simple statement of ordinal magnitude carries implications about absolute size (e.g., Choplin, 2010) and, as found in the current studies, about the spatial arrangement of the compared items – and possibly about other things, too.

Acknowledgments

I am grateful to George Farmer for help with coding free responses to the preliminary study and to Mair Thomas for helpful discussion.

References

- Allan, K. (1986). Interpreting English comparatives. *Journal of Semantics*, 5, 1-50.
- Blanz, V., Tarr, M.J., & Bülthoof, H.H. (1999). What object attributes determine canonical views? *Perception*, 28(5), 575-599. doi: 10.1068/p2897
- Boyce, C. J., Brown, G. D. A., & Moore, S. C. (2010). Money and happiness: Rank of income, not income, affects life satisfaction. *Psychological Science*, 21(4), 471-475. doi: 10.1177/0956797610362671
- Calabria, M., & Rossetti, Y. (2005). Interference between number processing and line bisection: a methodology. *Neuropsychologia*, 43, 779-783. doi: 10.1016/j.neuropsychologia.2004.06.027
- Cheung, C-N., & Lourenco, S.F. (2016). The associations between space and order in numerical and non-numerical sequences. *Consciousness and Cognition*, 45, 124-134. doi: 10.1016/j.concog.2016.08.013
- Choplin, J.M. (2010). I am "fatter" than she is: Language-expressible body-size comparisons bias judgments of body size. *Journal of Language and Social Psychology*, 29, 55-74. doi: 10.1177/0261927X09351679
- Clark, A. (2006). Language, embodiment, and the cognitive niche. *Trends in Cognitive Sciences*, 10(8), 370-374. doi: 10.1016/j.tics.2006.06.012
- Clark, H.H. (1969). Linguistic processes in deductive reasoning. *Psychological Review*, 76, 387-404.
- Clark, H.H., & Clark, E.V. (1968). Semantic distinctions and memory for complex sentences sentences. *Quarterly Journal of Experimental Psychology*, 20(2), 129-138. doi: 10.1080/14640746808400141
- De Hevia, M.D., Izard, V., Coubart, A., Spelke, E.S., & Streri, A. (2014). Representations of space, time, and number in neonates. *Proceedings of the National Academy of Sciences*, 111(13), 4809-4813. doi: 10.1073/pnas.1323628111
- Dehaene, S., Bossini, S., & Giraux, P. (1993). The mental representation of parity and number magnitude. *Journal of Experimental Psychology: General*, 122(3), 371-396. doi: 10.1037/0096-3445.122.3.371
- Ehringer, K.A., & Oliva, A. (2011). Canonical views of scenes depend on the shape of the space. Paper presented at the 33rd Annual Cognitive Science Conference, Boston, MA.
- Eikmeier, V., Alex-Ruf, S., Maienborn, C., & Ulrich, R. (2015). How strongly linked are mental time and space along the left-right axis? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 41(6), 1878-1883. doi: 10.1037/xlm0000129
- Fischer, M.H., Castel, A.D., Dodd, M.D., & Pratt, J. (2003). Perceiving numbers causes spatial shifts of attention. *Nature Neuroscience*, 6(6), 555-556. doi:10.1038/nn1066
- Fischer, M.H., & Shaki, S. (2014). Spatial associations in numerical cognition – From single digits to arithmetic. *Quarterly Journal of Experimental Psychology*, 67(8), 1461-1483. doi: 10.1080/17470218.2014.927515
- Fischer, M.H., Shaki, S., & Cruise, A. (2009). It takes just one word to quash a SNARC. *Experimental Psychology*, 56(5), 361-366. doi: 10.1027/1618-3169.56.5.361
- Fuhrman, O., & Boroditsky, L. (2010). Cross-cultural differences in mental representations of time: Evidence from an implicit nonlinguistic task. *Cognitive Science*, 34, 1430-1451. doi: 10.1111/j.1551-6709.2010.01105.x

- Fumarola, A., Prpic, V., Da Pos, O., Murgia, M., Umiltà, C., & Agostini, T. (2014). Automatic spatial association for luminance. *Attention, Perception, & Psychophysics*, *76*, 759-765. doi: 10.3758/s13414-013-0614-y
- Gevers, W., Santens, S., Dhooge, E., Chen, Q., Van den Bossche, L., Fias, W., & Verguts, T. (2010). Verbal-spatial and visuospatial coding of number-space interactions. *Journal of Experimental Psychology: General*, *139*(1), 180-190. doi: 10.1037/a0017688
- Gevers, W., Verguts, T., Reynvoet, B., Caessens, B., & Fias, W. (2006). Numbers and space: A computational model of the SNARC effect. *Journal of Experimental Psychology: Human Perception and Performance*, *37*(1), 32-44. doi: 10.1037/0096-1523.32.1.32
- Goodhew, S.C., McGaw, B., & Kidd, E. (2014). Why *is the sunny side* always up? Explaining the spatial mapping of concepts by language use. *Psychonomic Bulletin & Review*, *21*, 1287-1293. doi: 10.3758/s13423-014-0593-6
- Grice, H. (1989). *Studies in the way of words*. Cambridge, MA: Harvard University Press.
- Holmes, K.J., Ayzenberg, V. & Lourenco, S.F. (2016). Gamble on gaze: Eye movements reflect the numerical value of blackjack hands. *Psychonomic Bulletin & Review*. doi: 10.3758/s13423-016-1055-0
- Holmes, K.J., & Lourenco, S.F. (2011). Common spatial organization of number and emotional expression: A mental magnitude line. *Brain and Cognition*, *77*, 315-323. doi: 10.1016/j.bandc.2011.07.002
- Holyoak, K.J. & Mah, W.A. (1981). Semantic congruity in symbolic comparisons: Evidence against an expectancy hypothesis. *Memory & Cognition*, *9*(2), 197-204. doi:10.3758/BF03202335
- Hoorens, V., & Bruckmüller, S. (2015). Less is more? Think again! A cognitive-fluency-based more-less asymmetry in comparative communication. *Journal of Personality and Social Psychology*, *109*(5), 753-766. doi: 10.1037/pspa0000032
- Hutchinson, S., & Louwerse, M.M. (2014). Language statistics explain the spatial-numerical association of response codes. *Psychonomic Bulletin & Review*, *21*, 470-478. doi: 10.3758/s13423-013-0492-2
- Jahn, G., Knauff, M., & Johnson-Laird, P.N. (2007). Preferred mental models in reasoning about spatial relations. *Memory & Cognition*, *35*(8), 2075-2087.
- Laming, D.R.J. (1997). *The measurement of sensation*. Oxford: Oxford University Press.
- Lourenco, S.F., & Longo, M.R. (2010). General magnitude representation in human infants. *Psychological Science*, *2*(6), 873-881. doi: 10.1177/0956797610370158
- Louwerse, M.M. (2011). Symbol interdependency in symbolic and embodied cognition. *Topics in Cognitive Science*, *3*, 273-302. doi: 10.1111/j.1756-8765.2010.01106.x
- Louwerse, M.M., & Jeuniaux, P. (2010). The linguistic and embodied nature of conceptual processing. *Cognition*, *114*(1), 96-104. doi: 10.1016/j.cognition.2009.09.002
- Matthews, W.J. (2013). Relatively random: Context effects on perceived randomness and predicted outcomes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(5), 1642-1648. doi: 10.1037/a0031081
- Matthews, W. J., & Adams, A. (2008). Another reason why adults find it hard to draw accurately. *Perception*, *37*, 628-630. doi: 10.1068/p5895
- Matthews, W.J., & Dylman, A.S. (2014). The language of magnitude comparison. *Journal of Experimental Psychology: General*, *143*, 510-520. doi: 10.1037/a0034143
- Matthews, W.J., Gheorghiu, A.I., & Callan, M.J. (2016). Why do we overestimate others' willingness to pay? *Judgment and Decision Making*, *11*(1), 21-39.

- Matthews, W.J., & Meck, W.H. (2016). Temporal cognition: Connecting subjective time to perception, attention, and memory. *Psychological Bulletin*, *142*(8), 865-907. doi: 10.1037/bul0000045
- Mussweiler, T. & Epstude, K. (2009). Relatively fast! Efficiency advantages of comparative thinking. *Journal of Experimental Psychology: General*, *138*, 1-21. doi: 10.1037/a0014374.
- Ono, F., & Kawahara, J.-I. (2007). The subjective size of visual stimuli affects the perceived duration of their presentation. *Perception & Psychophysics*, *69*(6), 952-952. doi:10.3758/BF03193932
- Paivio, A. (1986). *Mental representations: A dual coding approach*. Oxford: Oxford University Press
- Patro, K., & Shaki, S. (2016). SNARC for numerosities is modulated by comparative instruction (and resembles some non-numerical effects). *Cognitive Processes*, *17*, 127-137, doi: 10.1007/s10339-015-0745-2
- Pérez, A., García, L., Valdés-Sosa, M., & Jaśkowski, P. (2011). Influence of the learnt direction of reading on temporal order judgments. *Psychology*, *2*(2), 103-108. doi: 10.4236/psych.2011.22017
- Proctor, R.W., & Cho, Y.S. (2006). Polarity correspondence: A general principle for performance of speeded binary classification tasks. *Psychological Bulletin*, *132*(3), 416-442. doi: 10.1037/0033-2909.132.3.416
- Ren, P., Nicholls, M.E.R., Ma, Y-y, & Chen, L. (2011). Size matters: Non-numerical magnitude affects the spatial coding of response. *PLoS ONE*, *6*(8), e23553, 1-6. doi: 10.1371/journal.pone.0023553
- Román, A., El Fathi, A., & Santiago, J. (2013). Spatial biases in understanding descriptions of static scenes: the of reading and writing direction. *Memory & Cognition*, *41*, 588-599. doi: 10.3758/s13421-012-0285-2
- Román, A., Flumini, A., Lizano, P., & Santiago, J. (2015). Reading direction causes spatial biases in mental model construction in language understanding. *Scientific Reports*, *5*, 18248, 1-8. doi: 10.1038/srep18248
- Rusiecki, J. (1985). *Adjectives and comparison in English: A semantic study*. London: Longman.
- Santens, S., & Gevers, W. (2008). The SNARC effect does not imply a mental number line. *Cognition*, *108*, 263-270. doi: 10.1016/j.cognition.2008.01.002
- Santiago, J., Lupiáñez, J., Pérez, E., & Funes, M.J. (2007). Time (also) flies from left to right. *Psychonomic Bulletin & Review*, *14*(3), 512-516. doi:10.3758/BF03194099
- Shaki, S., & Fischer, M.H. (2008). Reading space into numbers – a cross-linguistic comparison of the SNARC effect. *Cognition*, *108*, 590-599. doi:10.1016/j.cognition.2008.04.001
- Sperber, D., & Wilson, D. (1986). *Relevance: Communication and cognition*. Oxford: Blackwell.
- Stewart, N., Chater, N., Stott, H.P., & Reimers, S. (2003). Prospect relativity: How choice options influence decisions under risk. *Journal of Experimental Psychology: General*, *132*(1), 23-46. doi: 10.1037/0096-3445.132.1.23
- Ulrich, R., & Maienborn, C. (2010). Left-right coding of past and future in language: The mental timeline during sentence processing. *Cognition*, *117*, 126-138. doi: 10.1016/j.cognition.2010.08.001
- Vallesi, A., Binns, M., & Shallice, T. (2008). An effect of spatial-temporal association of response codes: Understanding the cognitive representations of time. *Cognition*, *107*(2), 501-527. doi: 10.1016/j.cognition.2007.10.011
- van Dijck, J-P., & Fias, W. (2011). A working memory account for spatial-numerical associations. *Cognition*, *119*, 114-119. doi: 10.1016/j.cognition.2010.12.013

- Zebian, S. (2005). Linkages between number concepts, spatial thinking, and directionality of writing: The SNARC effect and the REVERSE SNARC effect in English and Arabic monoliterates, biliterates, and illiterate Arabic speakers. *Journal of Cognition and Culture*, 5(1-2), 165-190. doi: 10.1163/1568537054068660
- Zhang, Y.C., & Schwarz, N. (2012). How and why 1 year differs from 365 days: A conversational logic analysis of inferences from the granularity of quantitative expressions. *Journal of Consumer Research*, 39(2), 248-259. doi: 10.1086/662612

Table 1. Details of Stimuli used in Studies 1-5.

Study	Dimension	Small	Large	Comparatives
1,3	Area (squares)	79 px	304 px	smaller/larger
	Delay	A 2 week delay	A 6 month delay	shorter/longer
	Height (trees)	178 px	378 px	shorter/taller
	Length (lines)	44 px	353 px	shorter/longer
	Number (dots)	6	19	fewer/more
	Probability	A 2% chance	A 98% chance	less/greater
	Reward	Win \$10	Win \$1000	smaller/bigger
	Weight	1 kg	10 kg	lighter/heavier
2,4	Area (circles)	179 px	233 px	smaller/bigger
	Height (trees)	318 px	350 px	shorter/taller
	Length (lines)	161 px	225 px	shorter/longer
	Money	\$8000	\$10000	smaller/larger
	Number (squares)	14	21	less/more
	Probability	70% chance	98% chance	less/greater
	Time	300 years	400 years	shorter/longer
	Weight	300 tons	500 tons	lighter/heavier
5	Area (squares)	19 px	141 px	smaller/bigger
	Height	42 px	252 px	lower/higher
	Length	14 px	182 px	shorter/longer
	Money	\$10	\$1400	less/greater
	Number (stars)	3	20	less/more
	Probability	A 4% chance	A 90% chance	lower/greater
	Time	5 seconds	6 years	shorter/longer
	Weight	4 grams	7 tons	less/heavier



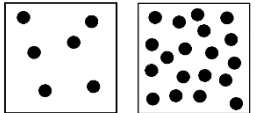
Note: px = pixels, and refers to diameter/height of visual items. Studies 1 and 3 used stimuli and comparative-sentences from Matthews & Dylman (2014, Study 2a); Studies 2 and 4 used stimuli and comparatives from Matthews & Dylman (2014, Study 4c); Study 5 used stimuli and comparatives from Matthews and Dylman (2014, Study 4a).

Table 2: Description of stimuli from Studies 6 and 7

Dimension	Small item	Large item	Comparatives
Age	Young woman	Old man	younger/older
Area	Small circle	Large square	smaller/larger
Height	Short man called John	Tall man called Jake	shorter/taller
Length	101 pixel blue line	301 pixel yellow line	shorter/longer
Number	Pink square with 4 spots	Brown square with 15 spots	less/more
Price	\$25 shirt	\$50 skirt	less/more
Probability	5% chance of hail	20% chance of sleet	less/greater
Weight	2600 lb Ford car	3300 lb Toyota car	lighter/heavier

Note: The visual stimuli were those used in the preliminary study and the comparatives were the model “smaller” and “larger” adjectives for each dimension produced by participants in that study. Age, Area, Length, and Number were all presented as photos/drawings without any text. Height, Price, and Weight combined an image (e.g., a picture of a car) with text (e.g., “Ford: 2600 lbs”). Probability was purely textual.

A. Example stimuli from Matthews & Dylman (2014) Study 2a

Area		“One square is _____ than the other”
Length		“One line is _____ than the other”
Number		“One box contains _____ than the other”
Probability	A 2% chance A 98% chance	“One probability is _____ than the other”

B. Example stimuli from Preliminary Experiment




Area		“The _____ is _____ than the _____”		
Length		“The _____ line is _____ than the _____ line”		
Number		“The _____ square has _____ spots than the _____ square”		
Probability	<table border="1" style="display: inline-table; border-collapse: collapse;"> <tr> <td style="padding: 5px; text-align: center;">5% chance of hail</td> <td style="padding: 5px; text-align: center;">20% chance of sleet</td> </tr> </table>	5% chance of hail	20% chance of sleet	“The probability of _____ is _____ than the probability of _____”
5% chance of hail	20% chance of sleet			

Figure 1. Examples of stimuli used by Matthews and Dylman (2014, Study 2a) and for the preliminary study conducted in preparation for the current experiments. In both cases, only a subset of the stimuli/dimensions are shown, and the stimuli are always shown in a small-large layout; in practice, the left-right layout of the objects was varied and participants completed sentences to describe the

relative magnitudes of the items. In the Matthews and Dylman (2014) studies, stimuli differed on a single attribute (e.g., area) and so could not be readily identified by a verbal label; in the preliminary experiment, the items differed in other properties, too, allowing more complex comparative sentences. Note that in the preliminary experiment, the 2 lines are of different colours.

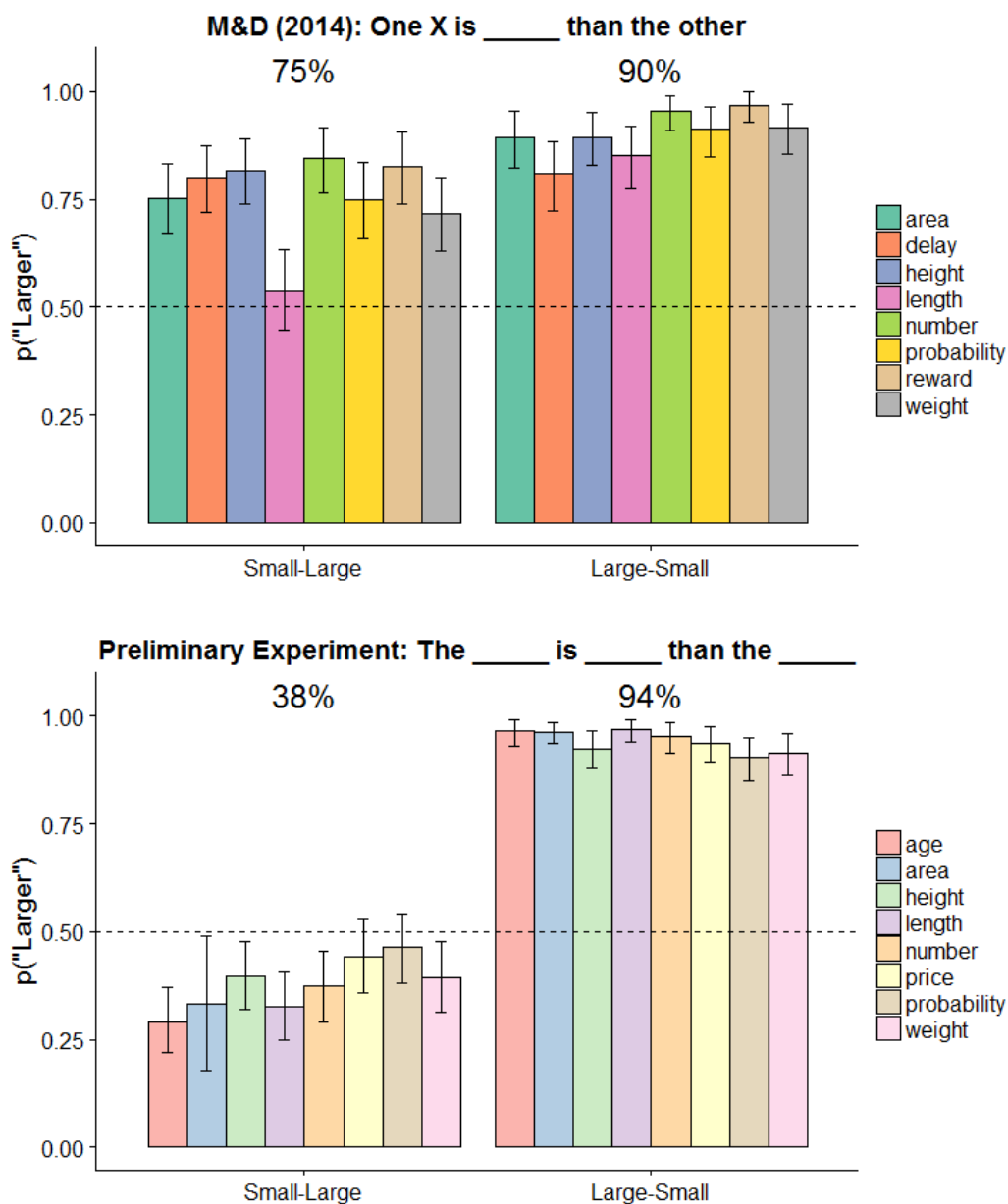


Figure 2. Spatial layout shapes the choice of comparative adjective. The bars show the proportion of participants who chose “larger” (rather than “smaller”) comparatives for each of multiple dimensions, grouped according to whether the two items were arranged with the small item on the left (Small-Large) or on the right (Large-Small); the numbers above the bars show the overall response proportions. The top panel plots the results from Matthews and Dylman (2014, Study 2a): English speakers typically selected “larger” comparatives over “smaller” ones (the HULC effect), but this tendency was more pronounced when the small object was on the left – consistent with order-matching. The lower panel shows an even stronger version of the same modulation in a preliminary experiment to the current series, where the objects differed on more than the focal dimension and could be identified by verbal labels (see Appendix). (Note that the same attribute dimensions recur in multiple experiments but are often tested with different stimuli; the bar colours signal the specific stimuli used to test a given dimension, and studies that used the same stimuli are plotted with the

same colours.) Here and throughout, the error bars show percentile-based 95% confidence intervals obtained by bootstrap resampling.

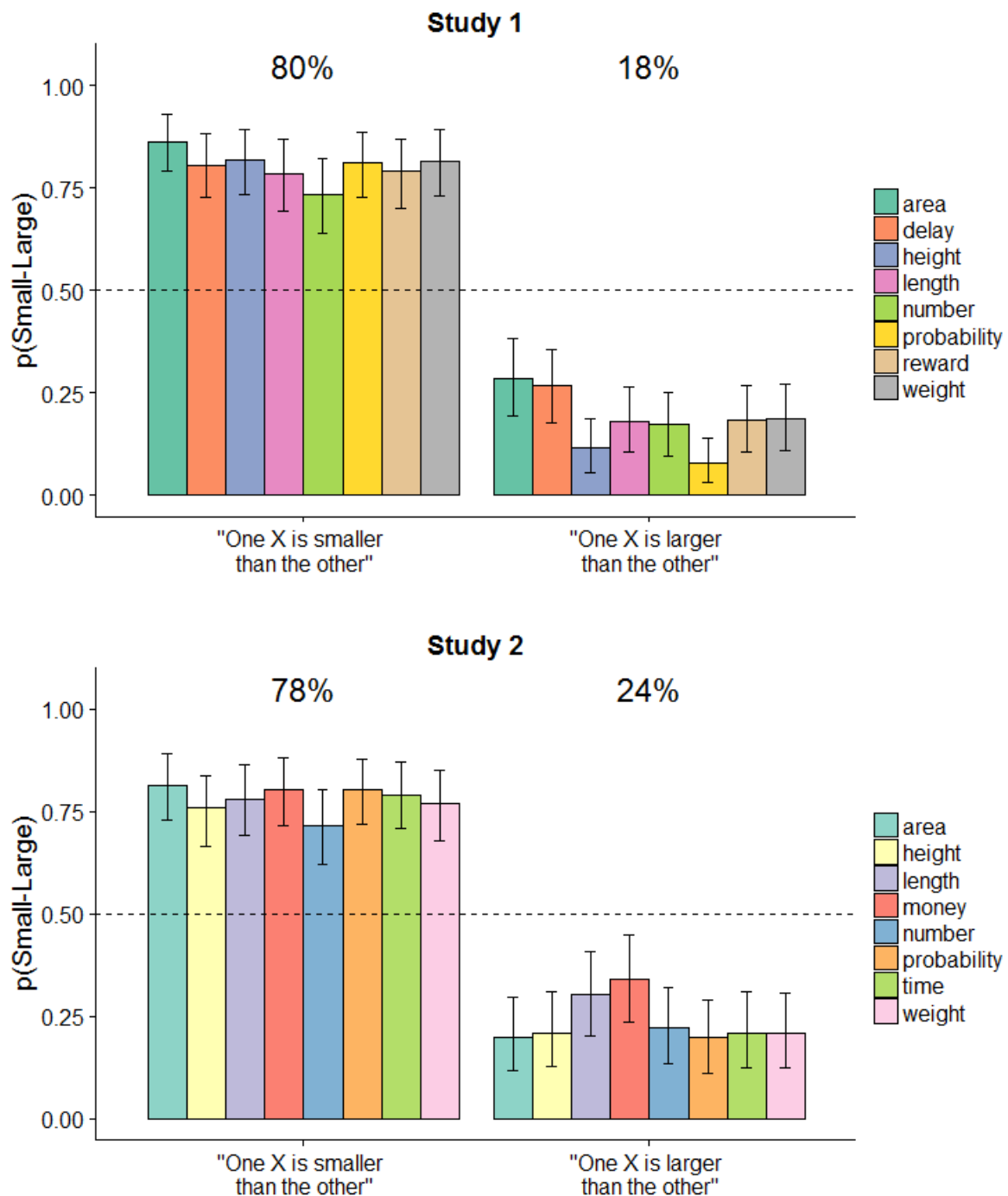


Figure 3. Results from Studies 1 and 2. The plots show the proportion of participants who inferred a “small-large” layout for each dimension, grouped by comparative adjective condition. The numbers above the bars for each condition indicate the overall proportion of small-large responses. “Smaller” comparatives (e.g., “smaller”, “shorter”, “less”) led to the belief that the small item was on the left;

“larger” comparatives (e.g., “larger”, “longer”, “more”) led to the inference that the small object was on the right.

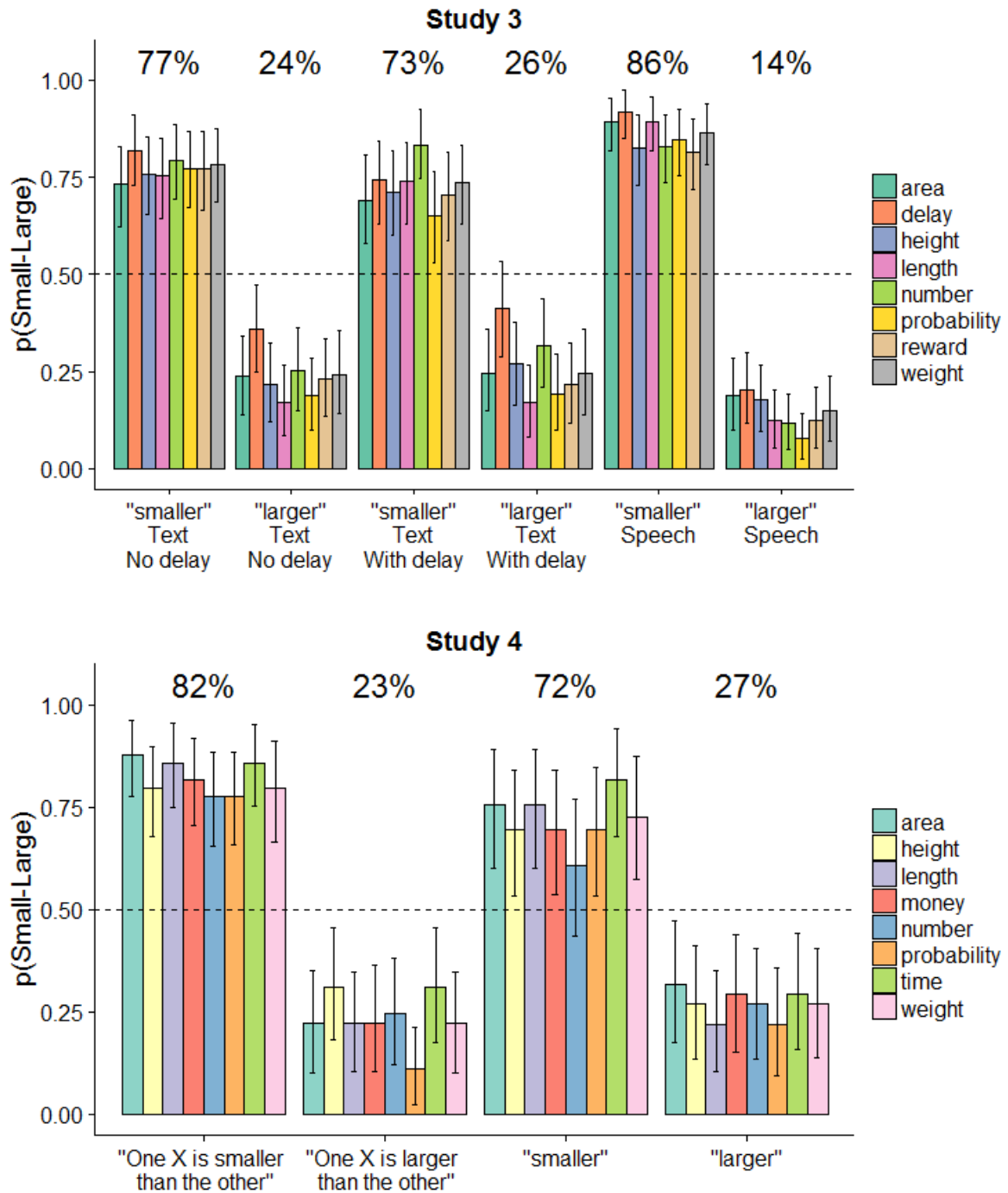


Figure 4. Results of Studies 3 and 4. "smaller" comparatives lead to the inference that the objects had a small-large layout, "larger" comparatives lead to the opposite belief. In Study 3, this effect is stronger with auditory than with written presentation of the comparative sentences, but for written sentences the effect is not modulated by whether the two possible layouts are presented at the same time as the sentence or after a short delay. In Study 4, the effect of comparative is not substantially weakened by replacing the comparative sentences with isolated comparative adjectives.

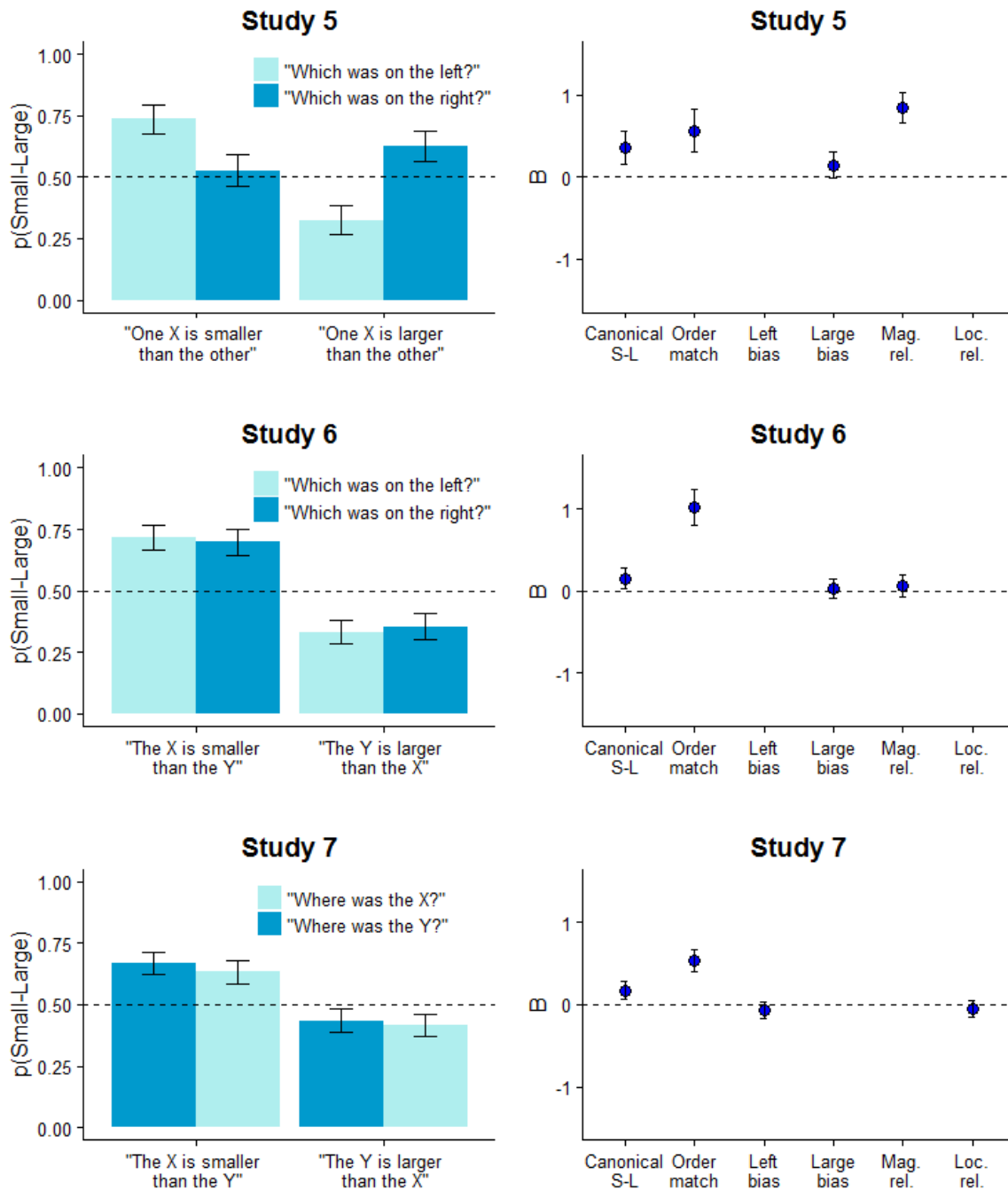


Figure 5. Results of Studies 5-7. The left panels show the proportion of small-large inferences for each condition. The right panel shows the regression coefficients, coded and signed such that larger positive values indicate greater evidence for the theoretical effect of interest: The Canonical S-L coefficient tests whether participants showed an overall tendency to infer a small-large layout; the Order match coefficient tests whether they were more likely to infer a small-large layout when the adjective was “smaller” than when it was “larger”; Left bias tests for a tendency/bias to indicate a left location; Large bias tests for a tendency to select the large object; Mag. rel. is short for “magnitude relevance”, and tests whether participants were more likely to select a small object when the comparative was “smaller”; Loc. rel. is short for “location relevance” and tests whether participants were more likely to select a left-hand location when the comparative was “smaller”.

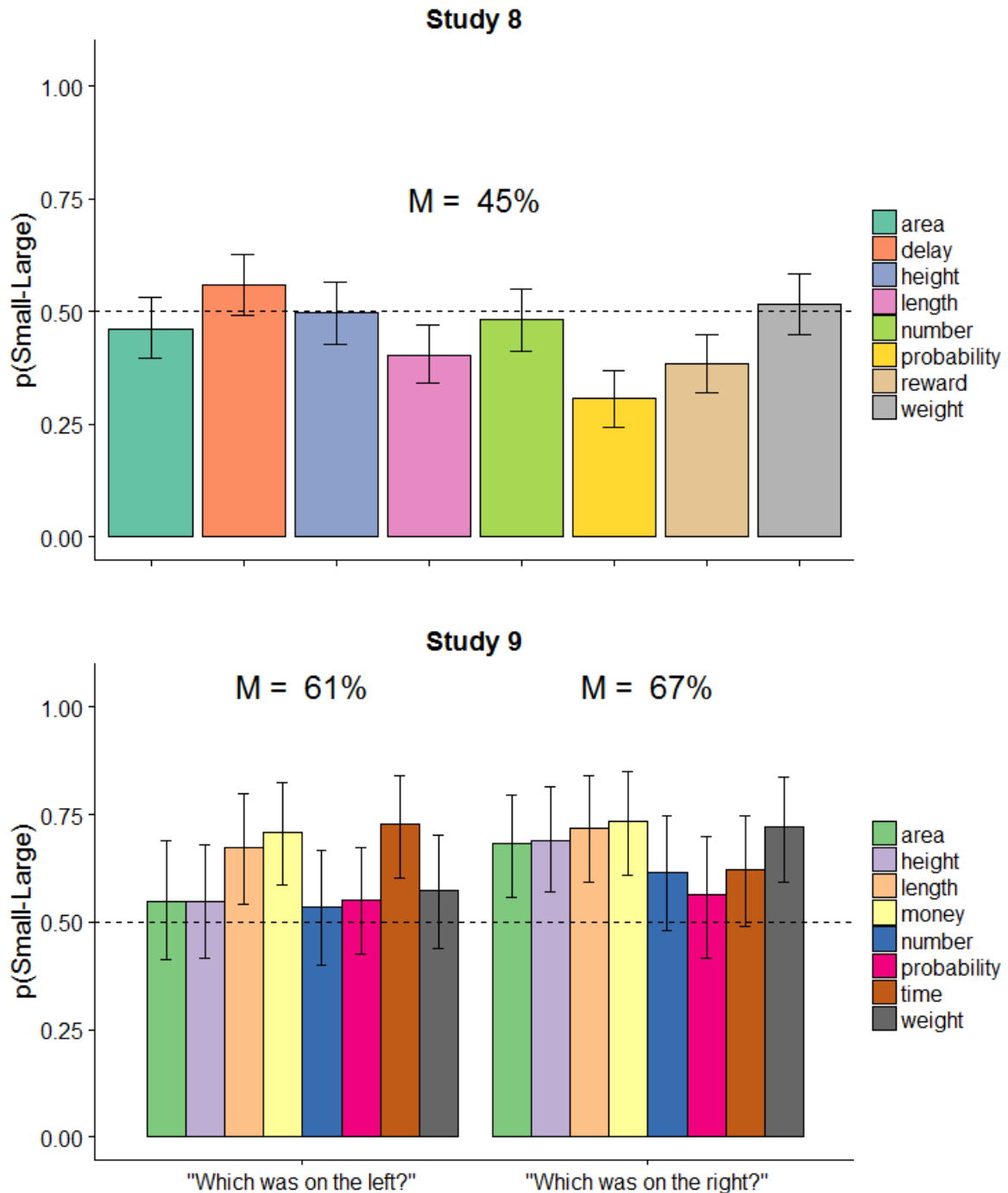


Figure 6. Results of Studies 8 and 9. The plots show the proportion of participants who judged that the smaller member of the presented objects was to the left of the larger item (i.e., a small-large layout). Study 8 required participants to select between two visually-presented layouts (small-large vs large-small) and finds no evidence for a default small-large inference. In Study 9 participants chose which of two visually-presented items had occupied a particular location (left or right); there is a modest tendency to assume a small-large layout, and this tendency is not modulated by whether participants were asked to identify the item on the left or on the right of the display. (Note that even though many of the attribute dimensions tested in these two studies are the same, they are labelled with different colours because the specific stimuli were different.)

Appendix: Preliminary Study

This study was conducted in preparation for the main experimental series. It generalized the approach of Matthews and Dylman (2014) to more complex objects and comparative sentences by presenting English-speakers with pairs of objects that differed on multiple dimensions so that each item could be identified by a non-magnitude linguistic label (e.g., “The man” and “The woman” or “The Toyota” and “The Ford”), allowing comparative sentences of the form “The X is [comparative] than the Y”.

Method

Participants. The approach to sampling was as for the main experiments. The sample comprised 288 participants (147 female, ages 19-71, $M = 37.0$, $SD = 11.8$; a further 45 were ineligible and 11 failed the attention check).

Design and procedure. On each page participants saw two objects side by side. The objects were shown pictorially and differed in magnitude but were also identifiable by some other property – for example, a small circle and a large square. Participants were asked how they would describe the relative magnitudes of the two items (e.g., “How would you describe the relative sizes of these two shapes?”). There were 8 pairs of stimuli; examples are given in Figure 1; full details are listed in Table 2.

Participants were asked to “Please type a sentence of the form: ‘The _____ is _____ than the _____.’ You can use whatever words you would naturally use to complete the sentence, but make sure it has that structure”. The wording of this instruction was tailored to the dimension (e.g., for price: “The _____ costs _____ than the _____”; for number: “The _____ square has _____ spots than the _____ square”). The left-right arrangement of the items was randomized for each dimension, as was the order of the dimensions. (A coding error meant that, for the first 219 participants, all participants received the Large-Small layout for the Area dimension.)

Results

Participants’ free-text responses were independently classified by two coders as “smaller” (e.g., “smaller”, “younger”, “shorter”, “less”), “larger” (e.g., “larger”, “bigger”, “greater”, “more”), or “unclassifiable”. Responses were judged unclassifiable when:

- They did not have the structure: object – comparative adjective – object
- They explicitly mentioned the spatial location of one or both objects

- The comparative was irrelevant or ambiguous with respect to the target dimension (e.g., “better”)
- The sentence did not seem reasonable to a native English speaker
- There was more than one magnitude comparative in the sentence, or the comparative was not describing the relation between the two objects
- One or both objects were ambiguously identified/described (i.e., the sentence references an object that wasn’t actually one of the presented pair)

The use of additional adjectives/adverbs/embellishments (e.g., “much longer...”) were acceptable, as were responses that treated the task as one of sentence completion (i.e., where the participant wrote a partial sentence but the overall structure was clear from the fill-in-the-blanks format of the question).

Inter-rater agreement was very high (99.2% identical classifications); the handful of disagreements were resolved by discussion. Unclassifiable response constituted 5.7% of the total and were removed from the analysis when computing response proportions. The proportion of “larger” responses in each condition is plotted in Figure 2.

I conducted a mixed-effects logistic regression with comparative (“smaller” = 0, “larger” = 1) as the dependent variable and layout (coded -1 for small-large, +1 for large-small) as the fixed-effects predictor, with random intercepts for participant and dimension and random slopes for layout within participants and dimensions (random effects uncorrelated). There was an overall tendency to use “larger” comparatives, $B_{\text{int}} = 2.000$, $CI = [1.660, 2.339]$, $p < .001$, and an effect of layout on adjective-choice, $B_{\text{cond}} = 3.59$, $CI = [2.526, 3.593]$, $p < .001$. Analysing responses from the small-large and large-small layout conditions separately confirmed the pattern in Figure 2: there was an overall tendency to use “smaller” comparatives when the small item was on the left, $B_{\text{int}} = -0.939$, $CI = [-1.373, -0.506]$, $p < .001$, and a tendency to use “larger” comparatives when the small item was on the right, $B_{\text{int}} = 8.286$, $CI = [6.618, 9.953]$, $p < .001$.