Journal of Memory and Language 127 (2022) 104355

Contents lists available at ScienceDirect

Journal of Memory and Language

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

Language concatenates perceptual features into representations during comprehension

Bruno R. Bocanegra [a,b,*], Fenna H. Poletiek [b,c], Rolf A. Zwaan [a]

[a] Department of Psychology, Education and Child Studies, Erasmus University Rotterdam, the Netherlands
[b] Institute of Psychology, Leiden University, the Netherlands
[c] Max Planck Institute for Psycholinguistics, Nijmegen, the Netherlands

ARTICLE INFO

Keywords:
Language
Comprehension
Perception
Conjunction
Simulation

ABSTRACT

Although many studies have investigated the activation of perceptual representations during language comprehension, to our knowledge only one previous study has directly tested how perceptual features are combined into representations during comprehension. In their classic study, Potter and Faulconer [(1979). Understanding noun phrases. Journal of Verbal Learning and Verbal Behavior, 18, 509–521.] investigated the perceptual representation of adjective-noun combinations. However, their non-orthogonal design did not allow the differentiation between conjunctive vs. disjunctive representations. Using randomized orthogonal designs, we observe evidence for disjunctive perceptual representations when participants represent feature combinations simultaneously (in several experiments; $N = 469$), and we observe evidence for conjunctive perceptual representations when participants represent feature combinations sequentially (In several experiments; $N = 628$). Our findings show that the generation of conjunctive representations during comprehension depends on the concatenation of linguistic cues, and thus suggest the construction of elaborate perceptual representations may critically depend on language.

Introduction

A currently influential view is that conceptual processing depends on the reactivation of perceptual representations (e.g. Damasio, 1994; Edelman, 1992; Feldman, 2010; Gallese & Lakoff, 2005; Barsalou, 1999; Goldstone & Barsalou, 1998; Prinz, 2002; Pulvermüller, 2005). For example, Prinz claims that "all (human) concepts are copies or combinations of copies of perceptual representations" (Prinz, 2002, p. 108). The idea is that our perceptual system analyzes objects in terms of different features, each of which can be attended to individually through the allocation of selective attention. By allocating attention to different objects, people can, over time, store individual perceptual features in memory. Once established, these features can be reactivated episodically in different combinations (see Goldstone & Barsalou, 1998). In this manner, we can represent known objects in their absence, and simulate novel objects that we have never seen before. When thinking of a house, for example, people can simulate a representation of the overall shape of the house, the windows, color and texture[1] (Barsalou, 1999).

In order to ensure that the simulation of perceptual representations is productive (i.e., that a bounded set of constituent features can generate a potentially unbounded set of composite representations), current models assume that people are able to simulate specific conjunctions of perceptual features during comprehension (see Barsalou, 1999; Goldstone & Barsalou, 1998). For instance, a perceptual simulation of a *BLUE HOUSE* should represent a house (excluding other objects), that is blue (excluding other colors).

We define a conjunctive representation as the selective representation of an intersection of two sets of features, and a disjunctive representation as a general representation of the union of two sets of features. In operational terms, we claim that participants are using a conjunctive representation, if, during the process of perceptual decision-making (e.g., the process of target identification by matching the shape and color dimensions of a target to those encoded in the representation), they have access to information pertaining to feature presence vs. absence for both

* Corresponding author at: Department of Psychology, Education and Child Studies, Erasmus University Rotterdam, Burg. Oudlaan 50, 3062 PA Rotterdam, the Netherlands.
E-mail address: bocanegra@essb.eur.nl (B.R. Bocanegra).

[1] Please note that throughout this paper we use the term visualization to describe the activation of visual features through language, not the direct activation of visual features through perception.

https://doi.org/10.1016/j.jml.2022.104355
Received 17 May 2021; Received in revised form 7 July 2022; Accepted 15 July 2022
Available online 25 July 2022

features simultaneously. If, on the other hand, features can only be interrogated independently from each other during perceptual decision-making, this would constitute a disjunctive representation[2].

According to standard arguments, being able to generate a conjunctive representation is necessary for meanings of concepts to have combinatorial productivity, which is one of the presumed hallmarks of human cognition (e.g., Rips, 1995). The assumption here is that the conjunction operates as an independent unit which inherits its meaning from its constituent features, and that it provides the internal structure which is necessary for (novel) combinations to be composed. This composition function serves to specify the constituency relations that mental representations can enter into (e.g. Barsalou, 1999; Fodor & Lepore, 1996). Importantly, this wide-held claim is made independently of separate claims regarding the (perceptual) nature of the features involved in conceptual representation.

Although previous studies have convincingly established that people activate perceptual features during language comprehension (e.g., Lupyan & Ward, 2013; Ostarek & Huettig, 2017; Zwaan, Stanfield, & Yaxley, 2002), the critical (and often implicit) assumption that people can generate conjunctive representations has not often been investigated.

To date, surprisingly few studies have investigated how people combine features when they generate perceptual representations (i.e., Potter & Faulconer, 1979; Wu & Barsalou, 2009). To our knowledge, Potter and Faulconer (1979) were the first to investigate how combinations of words in a sentence can cue the generation of a perceptual representation during comprehension. In their classic study, participants listened to sentences containing either an adjective-noun or simple noun phrase, for example, "it was already late when the man saw the {burning house vs. house} ahead of him", and were asked to discriminate a picture of a burning house from an unrelated picture presented immediately after the noun phrase (see Fig. 1). They argued that if participants are cued by the adjective-noun phrase "burning house", then they should be faster to respond to a picture of a burning house than when they are cued by the simple noun phrase "house". The rationale for this benefit in performance is that in the adjective-noun condition all features in the picture can potentially be matched to the generated representation, whereas in the noun-only condition only the features relating to the building can be matched to the representation but the features relating to the flames cannot (see Fig. 2a).

Potter and Faulconer (1979) indeed confirmed that participants responded faster in the adjective-noun condition (e.g. "burning house"), compared to the noun-only condition (e.g. "house"). Although this finding shows that participants were generating a representation using both the adjective and the noun, their experimental design did not allow one to differentiate between a conjunctive representation of the intersection of two sets of features and a disjunctive representation of the union of two sets of features (see the two left-most panels in Fig. 2b). The reason for this is that, within their design, each target was either a positive or a negative instance of a unique combination of features: in other words, participants were presented a *BURNING HOUSE*, or an *UPSIDE-DOWN AIRPLANE* (see the middle panel in Fig. 2b), but participants were never presented an instance of *BURNING AIRPLANE* or a *UPSIDE-DOWN HOUSE*. Because of this, the adjective and noun features were perfectly correlated and could not be evaluated independently of each other. If, within an experimental context, every house is a burning one and every burning thing is a house, there is no way of telling whether being fast at identifying a picture of a burning house is due to the expectation that it will be a house that is burning, or due to the expectation that it will be an object that is burning or a house or both.

In the present study, we investigated the combinatorial properties of perceptual representations generated during language comprehension, similar to the way this has previously been investigated in the visual perception literature (e.g. Kahneman, Treisman, & Gibbs, 1992; Gordon & Irwin, 1996; Saiki, 2003; Hommel, 2005; Bocanegra & Hommel, 2014). We used an orthogonal set of 4 stimuli consisting of two colors (i.e., red or green) and two shapes (i.e., a diamond or square)[3]. In our tasks, participants were presented a verbal cue and instructed to visualize the object described by the cue in terms of its color and shape. Subsequently, they were presented with a visual target, and they had to indicate as fast as possible whether the target matched or mismatched their visualization. For example, participants would be cued with the phrase "red square" and then would have to identify a *RED SQUARE* as a match or a *GREEN DIAMOND* as a mismatch, within an experimental context where they also had to identify a *RED DIAMOND* and a *GREEN SQUARE* (see right-most panel in Fig. 2b).

Because color and shape features were uncorrelated within our design this allowed us to test whether people were generating conjunctive or disjunctive representations. Our critical comparison is between *single-feature* trials where participants were asked to visualize a one visual feature (i.e., "red", "square", "green" or "diamond") and *dual-feature* trials where participants were asked to visualize two visual features simultaneously (i.e., "red square", "green diamond", "red diamond" or "green square"). If participants take less time to identify the target when they visualized two features vs. one feature we take this as evidence for the representation of a feature conjunction (e.g., both *RED* and *SQUARE*). On the other hand, if perceptual identification times are similar in dual-feature and single-feature trials we interpret this as evidence for the representation of a feature disjunction (e.g., either *RED* or *SQUARE*).

To clarify these predictions, let's first assume that the participant activated a conjunctive representation of both *RED* and *SQUARE* during a dual-feature trial. This conjunctive representation is illustrated schematically by the single box located in the R-S (top-left) quadrant in the space of possible stimuli (see the left column in Fig. 3). A feature conjunction is therefore a selective representation of only *one* of the four possible target stimuli in the experiment (i.e., the red square), excluding all other possible targets (i.e., the green square, the red diamond, and the green diamond). At the end of the trial, the color or the shape of the target may be checked against this representation (both color and shape were cued features). Given that both the target color and shape will match the representation, and given our assumption of a conjunctive representation, only a single comparison is needed in order to uniquely match the target to the represented quadrant (i.e., R-S).

In contrast, the situation is different if we assume that the participant activated a disjunctive representation of either *RED* or *SQUARE* during a dual-feature trial. This disjunctive representation is illustrated schematically by the three boxes located in the R-S (top-left), G-S (top-right), and R-D (bottom-left) quadrants in the space of possible stimuli (see the middle column in Fig. 3). A feature disjunction is therefore a more

[2] This implies that our general notion of conjunctive and disjunctive representation could be implemented in various process models using different decision rules / thresholds, as well different types of information representation. It is also worth pointing out that there are various ways one might represent multiple objects as distinct entities (e.g., a white cloud above a blue house) within frameworks that use both conjunctive and disjunctive representations. For instance, this could be achieved using conjunctive normal form (CNF) where a set of objects is represented as an overall disjunction of conjunctive literals. For example, one could use a representation like ((*A* and *B*) or (¬*A* and ¬*B*)). Alternatively, one could use an equivalent disjunctive normal form (DNF) where a set of objects is represented as an overall conjunction of disjunctive literals. Here, the previous example would be represented as ((¬*A* or *B*) and (*A* or ¬*B*)) (see; Mooney, 1995).

[3] One may argue that a diamond and a square are identical shapes which differ only in their orientation. Importantly however, participants in our experiments experienced no difficulty interpreting the rotated square as a 'diamond'. It is also worth noting that in the visual search literature color and orientation are standard features for investigating conjunctions (e.g., Spivey, Tyler, Eberhard, & Tanenhaus, 2001).

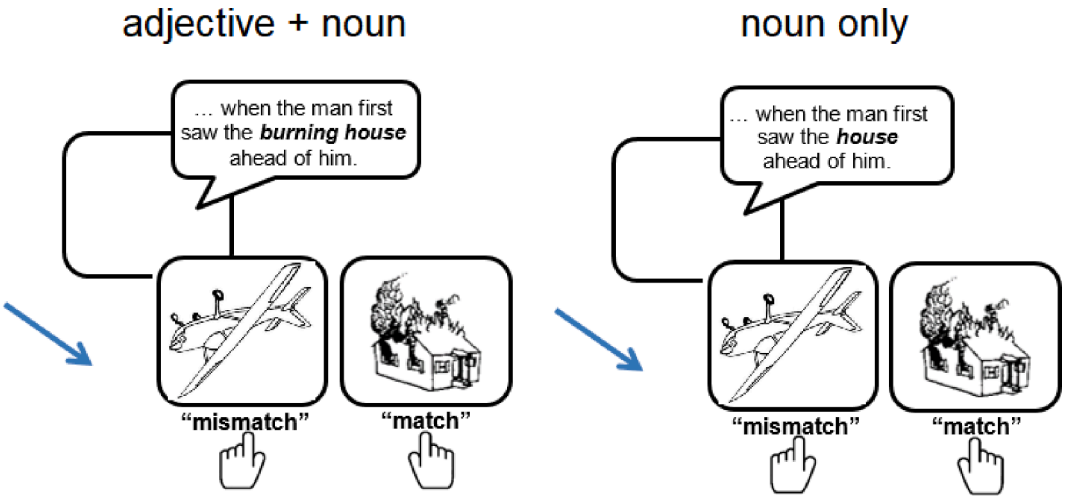

Fig. 1. Schematic examples of trials in the experiment reported in Potter and Faulconer (1979). In this experiment participants identified a visual picture after hearing a sentence describing the object presented in the picture. They had to indicate whether the object matched or mismatched the sentence. The left panel displays a trial where the object was described using an adjective and noun, the right panel displays a trial where the object was described using a noun only. On mismatching trials participants responded to an unrelated picture.

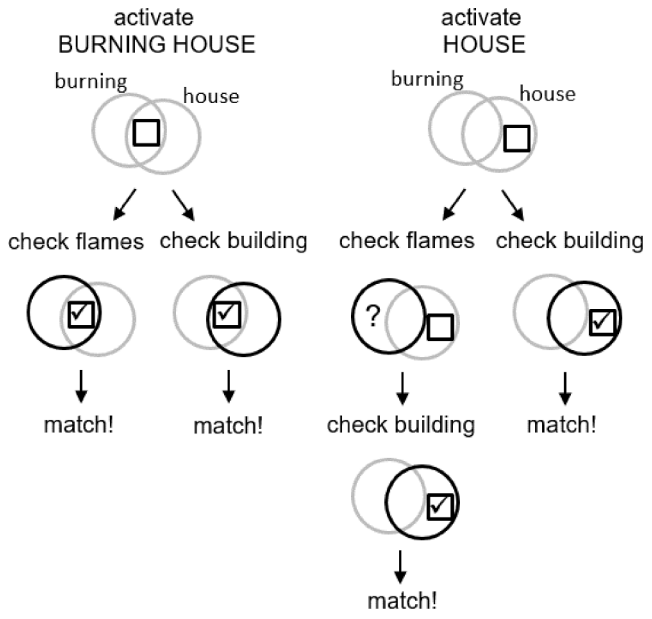

Fig. 2a. A theoretical interpretation of the activation and identification process underlying the speed-up for adjective + noun compared to noun-only trials observed in Potter and Faulconer (1979). The left column displays the interpretation that the representation of *HOUSE* is selectively constrained by the adjective *BURNING*. The right column displays the activation of the representation of *HOUSE* in the absence of the adjective *BURNING*.

general representation of three of the four possible target stimuli in the experiment (i.e., the red square, the green square, and the red diamond), excluding only one possible target (i.e., the green diamond). At the end of the trial, the color or the shape of the target may be checked against this representation (both color and shape were cued features). Given that both the target color and shape will match the representation, and given our assumption of a disjunctive representation, in this case *two* comparisons are needed in order to uniquely match the target to one of the represented quadrants. For instance, if color is checked first, two of the represented quadrants will coincide with the target (R-S and R-D), and shape will additionally need to be checked in order to determine that the target uniquely matches the R-S quadrant.

Critically, we can compare the conjunctive and disjunctive representations to a situation when a participant activated a simple representation of only *SQUARE* during a single-feature trial. This representation is illustrated schematically by the two boxes located in the R-S (top-left) and G-S (top-right) quadrants in the space of possible stimuli (see the right column in Fig. 3). At the end of the trial, the color of the target may be checked against this representation (shape was the only cued feature). In this case, two comparisons are also needed in order to uniquely match the target to one of the represented quadrants. For instance, after initially checking shape, two of the represented quadrants will coincide with the target (R-S and G-S), and color will additionally need to be checked in order to determine that the target uniquely matches the R-S quadrant.

Following this logic, we interpret a performance benefit for dual-feature vs. single-feature trials as suggesting that participants activated a conjunctive representation, whereas a null-effect is interpreted as suggesting that participants activated disjunctive representation. Please note that this logic applies symmetrically to match trials where both target features match the representation (see top panel of Fig. 3), and to mismatch trials where both target features mismatch the representation (see bottom panel of Fig. 3). In the latter case, the same number of comparisons are needed for each type of representation in order to determine that the target uniquely *mismatches* one of the represented quadrants.

It is important to note that we posit that the perceptual identification process is agnostic to meta-cognitive knowledge the participant might have of task constraints. For instance, within our experimental context both target features always either match or mismatch the verbal cue. Therefore, one might argue that there is no need for participants to evaluate both features: they could employ a task strategy where they only focus on a single feature (e.g., color) in order to determine the correct response. However, given that the task-relevant features in the cues varied randomly at the trial-level, participants would not be able to consistently apply this strategy throughout the course of the experiment (see Bocanegra & Hommel, 2014, for examples of similar tasks that either enable or prevent participants from applying this type of strategy). Given the orthogonal design and the variability in task-relevant features, we posit that participants will focus their attention on both color and shape features and engage in an exhaustive perceptual identification of the target. From this it follows that in match trials a target will have been positively identified once it is determined that it uniquely matches one of the potential targets in the representation, whereas in

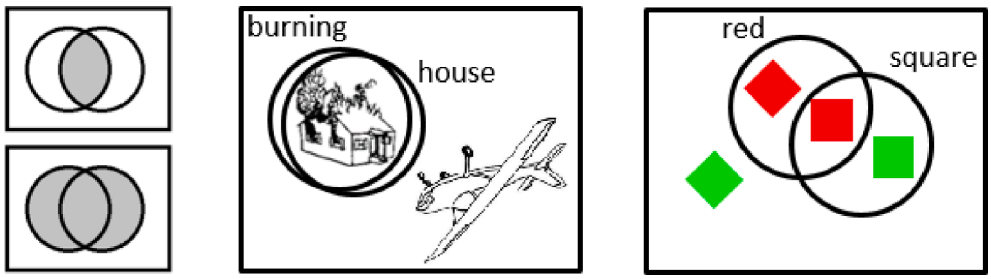

Fig. 2b. The two left-most panels illustrate the difference between a conjunctive representation of the intersection between two sets of features (top) and a disjunctive representation of the union of two sets of features (bottom). The middle panel illustrates the non-orthogonal manipulation of feature combinations used in Potter and Faulconer (1979). The right panel illustrates the orthogonal manipulation of feature combinations used in the present study.

mismatch trials a target will have been negatively identified if it uniquely mismatches one of the potential targets in the representation (see Fig. 3).

Experiments 1a-d

In Experiments 1a and 1b we instructed participants to visualize features using an auditory verbal cue, whereas in Experiments 1c and 1d we used a visual verbal cue. According to our predictions (see Fig. 3), if participants activate a conjunctive representation of color and shape, then we should observe faster reaction-times for dual-feature vs. single-feature trials; on the other hand, if participants activate a disjunctive representation of color and shape then we should observe similar reaction-times for dual-feature and single-feature trials.

Alongside the experimental tasks (i.e., the visualization tasks in Experiments 1a-d; see Fig. 4), participants also performed various types of control tasks (i.e., a purely verbal task in Experiments 1a-b, a purely visual task in Experiment 1c, and a verbalization task in Experiment 1d) investigating several interpretations of potential effects observed in the visualizations tasks (see Fig. 5). These controls had task structures that were identical to the experimental visualization tasks, but systematically varied characteristics of the cues and targets, where we instructed participants either to verbalize (Experiment 1a, 1b, and 1d) or to visualize (Experiment 1c).

By systematically comparing our original visualization task (where participants were required to actively generate a visual representation based on a verbal description) to identical tasks that either (a) could be performed using only visual maintenance (where participants were presented an actual picture as a cue; Experiment 1c), (b) could be performed using only verbal maintenance (where participants were presented verbal stimuli as targets; Experiments 1a and 1b), and (c) could only be performed using verbalization (where participants were required to actively generate a *verbal* representation based on a visual stimulus; Experiment 1d), we aimed to show that any effects observed in the experimental conditions could only be attributed to visualization, rather than other task parameters. Their outcomes are discussed in the results and discussion section below.

Method

Participants

Forty-eight participants with normal or corrected-to-normal vision participated in Experiments 1a-d (twelve in each experiment). All participants were undergraduate students at Leiden University, the Netherlands, participating for course credit or a small monetary reward. All experiments were conducted in accordance with relevant regulations and institutional guidelines and was approved by the local ethics committees from the Faculty of Social and Behavioural Sciences.

Materials

We used 8 different auditory cue stimuli. In the dual-feature trials, we used 4 auditory word combinations: “red square”, “red diamond”, “green square” and “green diamond”. In the single-feature trials, we used 4 auditory words: “red”, “green”, “square” and “diamond”. The cues were spoken by a digitally generated female speaker with a neutral prosody. We used 8 different visual target stimuli. In the visualization task, we used 4 colored geometric shapes: a red square, red diamond, green square and a green diamond. In the verbal task, we used 4 visual word combinations: “red square”, “red diamond”, “green square” and “green diamond”.

Procedure

Experiment 1a had a 2 (visualization task vs. verbal control task) × 2 (dual-feature trial vs. single-feature trial) design. In the first half of the experiment participants performed the visualization task and in the second half they performed the verbal task, or vice versa. Within each task, single-feature and dual-feature trials were blocked and alternated (4 blocks per task, with 32 trials per block). The order of the tasks and blocks were counterbalanced over participants. All other variables varied randomly from trial to trial within each block. The experiment consisted of 128 trials.

Each trial started with a fixation cross (500 ms), followed by an auditory verbal cue stimulus (between 700 ms and 1200 ms), a blank screen (2000 ms), and finally a visual perceptual target stimulus (until response). After responding, performance feedback was given (500 ms).

In conditions with visual perceptual targets, participants were instructed to visualize the cued object (in terms of its color and / or shape), whereas in conditions with visual word targets participants were instructed to verbalize the cued words (in terms of its color and / or shape). After a 2 s interval, they were presented with a target stimulus, and had to indicate as fast as possible whether the target matched or mismatched their visualization or verbalization (see Figs. 4 and 5). They were instructed to press the right response if the target matched the cue they were presented at the start of the trial (the “m” key on the keyboard), and to press the left response if the target mismatched the cue (the “z” key on the keyboard). Participants were instructed that there would never be ambiguity as to which response had to be given: for dual-feature trials, the target would always either match or mismatch on both color and shape dimensions (see Figs. 4 and 5).

Experiment 1b was identical to Experiment 1a except for the visual targets used in the visualization task: We spatially separated out the two perceptual features in order to match the target presentation in the verbal control task.

Experiment 1c was identical to Experiment 1a, except for in the following ways. Instead of using a purely verbal task as a control task, we used a purely visual task alongside the visualization task. The cue and target stimuli were therefore always presented in the visual modality. In the visual control task, cue stimuli consisted of colored patches and geometric shapes. In the visualization task, cue stimuli consisted of visually presented words. Cue stimuli were presented for 1000 ms.

Experiment 1d was identical to Experiment 1c, except for in the following ways. Here, we used a verbalization task as the control task.

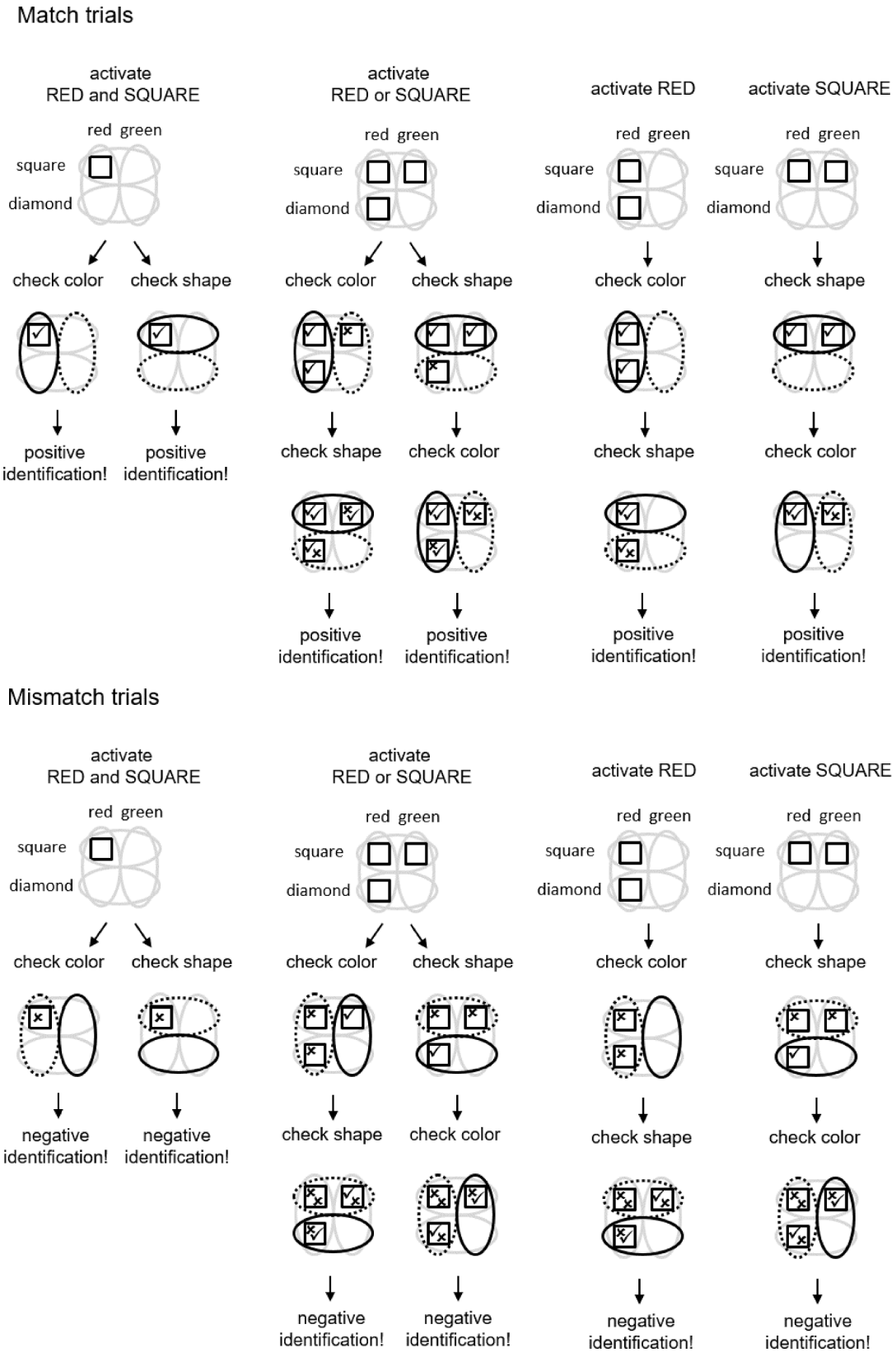

Fig. 3. Schematic representation of the process of perceptual representation and identification underlying the theoretical predictions for the visualization tasks in Experiments 1a-d. We represent the situation where participants activate a feature conjunction (e.g., *RED* and *SQUARE*; see left column), a feature disjunction (e.g. *RED* or *SQUARE*; see middle column), or a single feature (e.g., *SQUARE*; see right column). The top panel displays match trials that required "yes" responses, whereas bottom panel displays mismatch trials that required "no" responses.

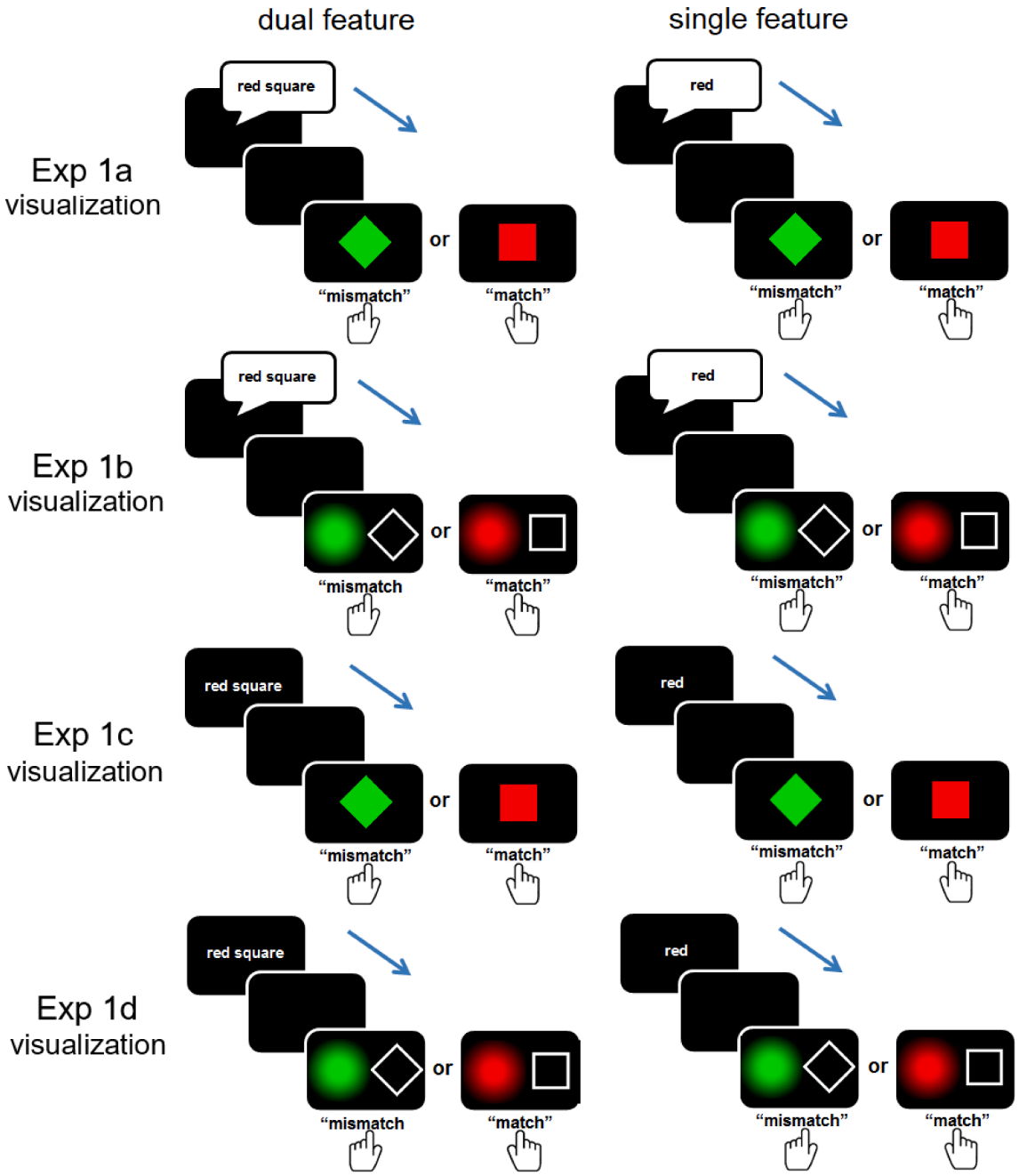

Fig. 4. Examples of trials in the visualization tasks of Experiments 1a-d. In these visualization tasks participants identified a visual perceptual target using a verbal linguistic cue (i.e., in order to identify the perceptual target they had to visualize the linguistic cue). The left panels display dual-feature trials, the right panels display single-feature trials. In Experiment 1a and 1b, the sequence consisted of an auditory cue (700–1200 ms), followed by a blank screen (2000 ms), and finally a visual target (until response). In Experiments 1c and 1d, the sequence consisted of a visual cue (1000 ms), followed by a blank screen (2000 ms), and finally a visual target (until response).

The features in both the cue and target stimuli were always presented in a spatially separated manner. In the verbalization control task, cue stimuli consisted of colored patches and geometric shapes and target stimuli consisted of visually presented words. In the visualization task, this was the reverse: cue stimuli consisted of visually presented words and target stimuli consisted of colored patches and geometric shapes.

Data analysis

Reaction times faster than 100 ms and slower than 2000 ms were discarded (<4% in Experiment 1a, < 2% in Experiments 1b-d). The mean RTs for correct trials, as well as the proportion of accurate responses were included in the statistical analyses. All raw data files for the experiments reported in this study are available from https://osf.io/gbuqm/.

Results

RTs and accuracies for each experiment were analyzed using a 2 × 2 (Task [visualization vs. control] × Trial-type [dual-feature vs. single-feature]) analysis of variance (ANOVA). In Experiment 1a, we found significant main effects in RT for task, $F(1,11) = 19.23$, $p < .01$, $\eta_p^2 = 0.64$, trial-type, $F(1,11) = 17.27$, $p < .01$, $\eta_p^2 = 0.61$, and a significant interaction between task and trial-type, $F(1,11) = 10.24$, $p < .01$, $\eta_p^2 = 0.48$. Dual-feature trials ($M = 637$, $SD = 256$) did not differ from single-

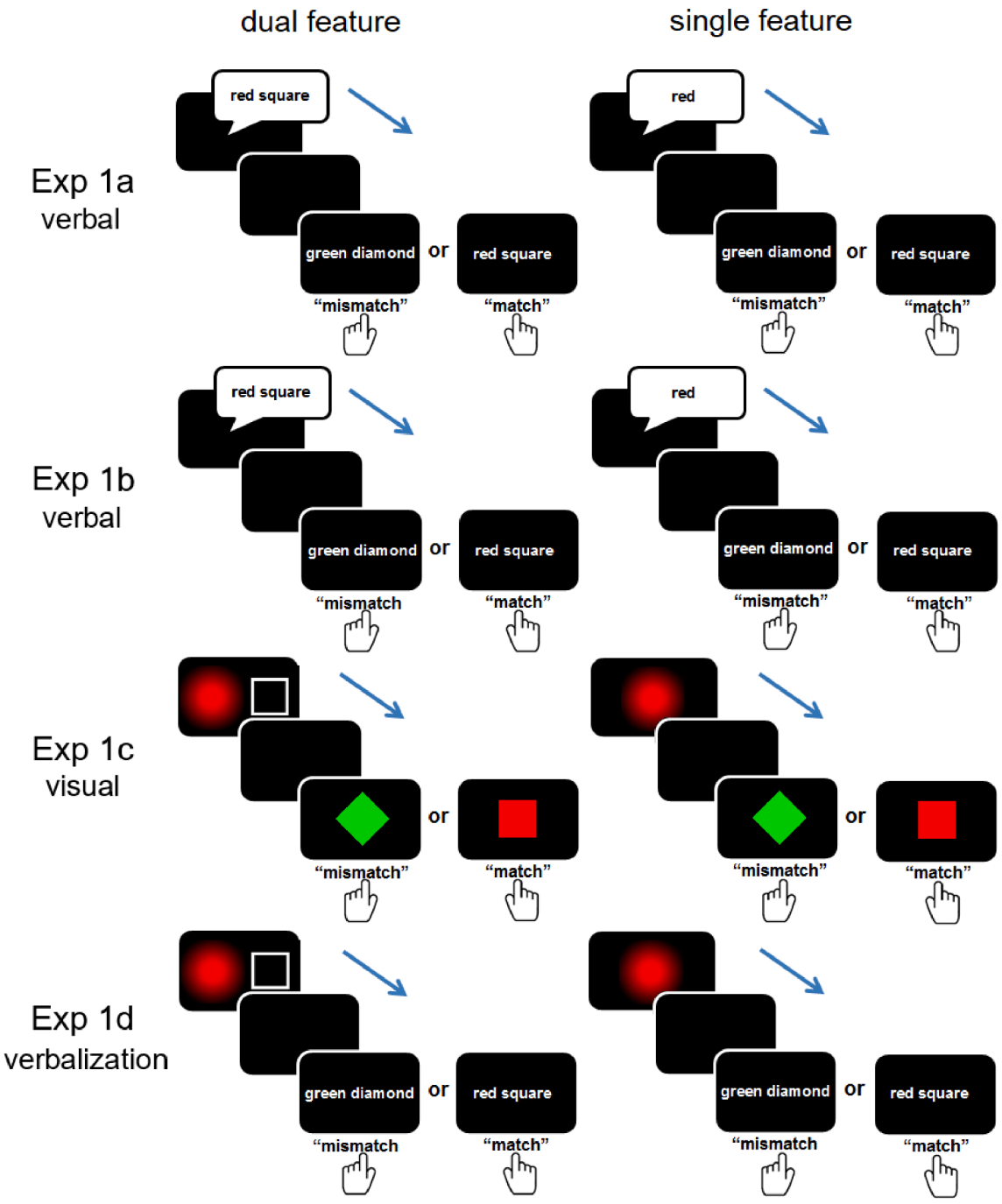

Fig. 5. Examples of trials in the control tasks of Experiments 1a-d. In Experiments 1a and 1b, the control task was a purely verbal task where both cues and target were verbal linguistic stimuli. In Experiment 1c the control task was a purely visual task where both cues and targets were visual perceptual stimuli. In Experiment 1d the control task was a verbalization task where participants identified a verbal linguistic target using a visual perceptual cue (i.e., in order to identify the verbal target they had to verbalize the perceptual cue). The left panels display dual-feature trials, the right panels display single-feature trials. In Experiment 1a and 1b, the sequence consisted of an auditory cue (700–1200 ms), followed by a blank screen (2000 ms), and finally a visual target (until response). In Experiments 1c and 1d, the sequence consisted of a visual cue (1000 ms), followed by a blank screen (2000 ms), and finally a visual target (until response).

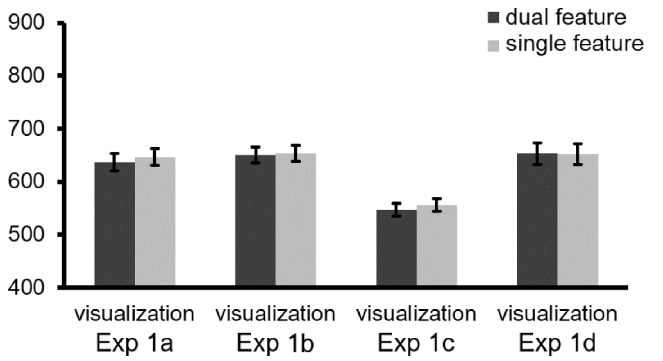

Fig. 6. RTs for the dual and single feature conditions in the visualization tasks of Experiments 1a-d. In these visualization tasks participants identified a perceptual target using a linguistic cue. Error bars represent within-subjects standard errors (Loftus & Masson, 1994).

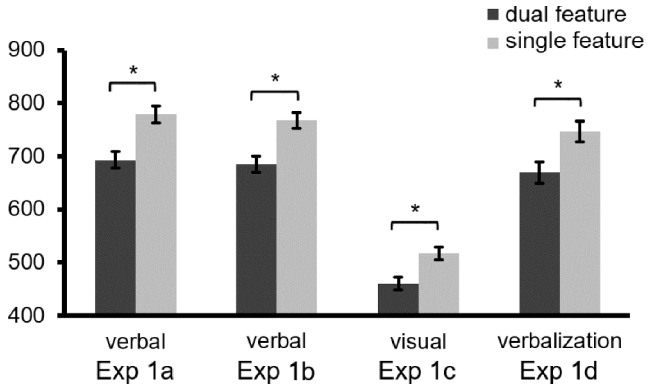

Fig. 7. RTs for the dual and single feature conditions in the control tasks in Experiments 1a-d. In Experiments 1a and 1b, the control task was a purely verbal task (both cues and target were linguistic stimuli). In Experiment 1c the control task was a purely visual task (both cues and targets were perceptual stimuli). In Experiment 1d the control task was a verbalization task (participants identified a linguistic target using a perceptual cue). Error bars represent within-subjects standard errors (Loftus & Masson, 1994).

feature trials ($M = 646$, $SD = 264$) in the visualization task, $|t| < 1$, $p > .50$ (see Fig. 6). However, participants were faster for dual-feature trials ($M = 693$, $SD = 275$) compared to single-feature trials ($M = 779$, $SD = 297$) in the verbal control task, $t(11) = 4.90$, $p < .001$ (see Fig. 7). In the accuracies, we did not observe significant main effects or interaction, $F < 3.7$, $p > .08$. Participants were equally accurate for dual-feature trials (verbal control task: $M = .98$, $SD = .03$; visualization task: $M = .96$, $SD = .03$) and single-feature trials (verbal control task: $M = .96$, $SD = .02$; visualization task: $M = .96$, $SD = .03$)[4].

In Experiment 1b, we found significant main effects in RT for task, $F(1,11) = 15.82$, $p < .01$, $\eta_p^2 = 0.59$, trial-type, $F(1,11) = 11.21$, $p < .01$, $\eta_p^2 = 0.50$, and a significant interaction between task and trial-type, $F(1,11) = 14.04$, $p < .01$, $\eta_p^2 = 0.56$. Dual-feature trials ($M = 650$, $SD = 129$) did not differ from single-feature trials ($M = 653$, $SD = 169$) in the visualization task, $|t| < 1$, $p > .80$ (see Fig. 6). However, participants were faster for dual-feature trials ($M = 684$, $SD = 144$) compared to single-feature trials ($M = 766$, $SD = 151$) in the verbal control task, $t(11) = 7.18$, $p < .001$ (see Fig. 7). In the accuracies, we only observed a significant main effect for trial-type $F(1,11) = 7.09$, $p < .05$, $\eta_p^2 = 0.39$. Participants were more accurate for dual-feature trials (verbal control task: $M = .96$, $SD = .03$; visualization task: $M = .96$, $SD = .03$) compared to single-feature trials (verbal control task: $M = .94$, $SD = .05$; visualization task: $M = .93$, $SD = .05$).

In Experiment 1c, we found significant main effects in RT for task, $F(1,11) = 9.84$, $p < .01$, $\eta_p^2 = 0.47$, trial-type, $F(1,11) = 18.69$, $p < .01$, $\eta_p^2 = 0.63$, and and a significant interaction between task and trial-type, $F(1,11) = 5.47$, $p < .05$, $\eta_p^2 = 0.33$. Dual-feature trials ($M = 548$, $SD = 111$) did not differ from single-feature trials ($M = 556$, $SD = 108$) in the visualization task, $|t| < 1$, $p > .50$ (see Fig. 6). However, participants were faster for dual-feature trials ($M = 460$, $SD = 62$) compared to single-feature trials ($M = 517$, $SD = 77$) in the visual control task, $t(11) = 4.90$, $p < .001$ (see Fig. 7). In the accuracies, we only observed a significant main effect for trial-type $F(1,11) = 6.42$, $p < .05$, $\eta_p^2 = 0.37$. Participants were more accurate for dual-feature trials (visual control task: $M = .96$, $SD = .03$; visualization task: $M = .95$, $SD = .05$) compared to single feature trials (visual control task: $M = .95$, $SD = .05$; visualization task: $M = .92$, $SD = .06$).

In Experiment 1d, we found significant main effects in RT for task, $F(1,11) = 5.55$, $p < .05$, $\eta_p^2 = 0.34$, trial-type, $F(1,11) = 4.96$, $p < .05$, $\eta_p^2 = 0.31$, and a significant interaction between task and trial-type, $F(1,11) = 10.40$, $p < .01$, $\eta_p^2 = 0.49$. Dual-feature trials ($M = 652$, $SD = 139$) did not differ from single-feature trials ($M = 651$, $SD = 134$) in the visualization task, $|t| < 1$, $p > .90$ (see Fig. 6). However, participants were faster for dual-feature trials ($M = 668$, $SD = 105$) compared to single-feature trials ($M = 746$, $SD = 127$) in the verbalization control task, $t(11) = 3.40$, $p < .01$ (see Fig. 7). In the accuracies, we only observed a significant main effect for trial-type $F(1,11) = 5.12$, $p < .05$, $\eta_p^2 = 0.31$. Participants were more accurate for dual-feature trials (verbalization control task: $M = .96$, $SD = .03$; visualization task: $M = .96$, $SD = .03$) compared to single-feature trials (verbalization control task: $M = .94$, $SD = .03$; visualization task: $M = .95$, $SD = .03$).

Discussion

In Experiment 1a, we did not observe a dual-feature benefit in the visualization task, which is consistent with the idea that participants were representing a disjunctive combination of shape and color. However, there are various potential alternative explanations for this null-effect. We will go through each of these in turn.

Although we instructed participants to visualize two features during dual-feature trials, there were no task requirements that forced participants to do so. In other words, participants could have potentially disregarded the instructions and attended to only one of the two features presented in the auditory cue (for example, the first word). The selective representation of only a single feature in during dual-feature trials would make them informationally equivalent to the single-feature trials and therefore explain the observed null-effect. Given that participants could potentially perform the task by processing only one of the words presented in the auditory cue, we therefore included a verbal control task in order to assess whether participants were processing both cue words. If participants were only processing one of the two cue words, then we should also observe similar reaction-times for dual-feature trials and single-feature trials in the verbal control task that we tested alongside the visualization task in Experiment 1a. This was not the case: we observed a clear dual-feature benefit in the verbal control task, suggesting that the null-effect observed in the visualization task cannot be explained by a failure of participants to process both cue words.

A salient procedural difference between the visualization task and verbal control task in Experiment 1a is that target features in the verbal control task were spatially segregated (see Figs. 4 and 5), whereas target features in the visualization task were spatially integrated. This may have created a difference in the way that participants performed the two

[4] In order to get an indication of whether the type of feature had an overall effect on RT in our experiments, we analyzed potential differences between color vs shape in the single feature trials for Experiments 1a-d. We failed to find significant differences in the visualization tasks ($ps > .12$), and in the control tasks ($ps > .62$), indicating that judging the color and shape features during visualization and verbalization was approximately comparable in terms of difficulty.

tasks, which could potentially offer an explanation for the differential effect of cueing two features vs. one feature in the visualization vs. verbal control task. In order to assess the impact of this difference, we replicated these two tasks in Experiment 1b, while segregating the target features in the visualization task (see Figs. 4 and 5). As in Experiment 1a, we did not observe a dual-feature benefit in the visualization task, whereas we replicated the dual-feature benefit that we previously observed in the verbal control task. This suggests that the spatial positioning of the target features cannot account for the null-effect observed in the visualization task.

Our failure to observe a dual-feature benefit in Experiments 1a and 1b may also be explained by the nature of the identification task performed on the target. Perhaps identifying the target in a visualization task is very easy and does not allow us to observe any additional benefit of visualizing two features vs. one feature. In Experiment 1c, we therefore compared the visualization task to a purely visual control task (see Figs. 4 and 5). If some aspect of the target identification task is preventing us from observing a dual-feature benefit in the visualization task, then we should also fail to observe this benefit in a purely visual control task. As in Experiments 1a and 1b, we again observed a null-effect in the visualization task. However, we did observe a clear dual-feature benefit in the purely visual control task. Given that the targets were identical in the visualization and purely visual control tasks, this suggests that the ease of target identification cannot account for our failure to observe a dual-feature benefit in the visualization task (i.e. due to floor effects in RT masking a difference). This interpretation is bolstered by the fact that participants showed the dual-feature benefit in the visual control task despite being overall *faster* (see Figs. 6 and 7).

A critical feature of the visualization tasks in Experiments 1a-c is that they always required participants to translate information from the verbal to the visual domain. In contrast, while performing the control tasks in Experiments 1a-c participants always processed information *within* a domain (i.e., within the verbal domain in Experiment 1a and 1b, and the visual domain in Experiment 1c). Due to this difference, one might ask whether cross-domain translation may be masking a potential dual-feature benefit in the visualization tasks. To test this, we compared the visualization task to a verbalization control task in Experiment 1d (see Fig. 5), where participants translated information from the visual to the verbal domain. If cross-domain translation is responsible for masking a dual-feature benefit in the visualization task, then we should also observe a null-effect in the verbalization control task. Our results indicate that this was not the case. In addition to again observing the null-effect in the visualization task, we observed a clear dual-feature benefit in the verbalization task, suggesting that cross-domain translation cannot account for our failure to observe a dual-feature benefit in the visualization tasks.

In sum, in all four experiments we observed consistent null-effects in the visualization tasks where reaction-times for dual-feature trials and single-feature trials were similar, and we observed consistent dual-feature benefits in the various control tasks. Please note that the dual-features benefits occurred while we systematically varied various properties of the cues and targets in the control tasks, which allows us to exclude many alternative explanations of the null-effects observed in the visualizations tasks. The consistent pattern across Experiments 1a-d is that null-effects are observed whenever participants are visualizing a verbal cue. Therefore, in light of our predictions, this pattern of results suggests that participants were activating a general disjunctive representation instead of a specific conjunctive representation of the shape and color features in the visualization tasks.

Experiments 2a-b

Given that we did not observe dual-feature benefits when participants were visualizing combinations of color and shape features in the visualization tasks in Experiments 1a-d, this leaves us with the following question: under what conditions do participants represent specific feature conjunctions during language comprehension?

Some researchers have proposed that language may play an important role in the generation of complex perceptual representations (e.g. Borghi & Binkofski, 2014; Carruthers, 1996; Gomila, Travieso, & Lobo, 2012; Lupyan & Bergen, 2016; Paivio, 1986; 2007; Spelke, 2003; Zwaan & Madden, 2005). For example, Zwaan and Madden (2005) view linguistic sequences as a series of cues that can activate and combine previously stored perceptual features. If this is the case, then the temporal concatenation of a sequence of cues may play an important role in the activation of feature combinations. The rationale behind this idea is that an ordered sequence of words may, over time, incrementally activate (sub-) sets of perceptual features that provide the referential domain for a phrase. Through this process, sets of perceptual features may become increasingly specified in terms of its perceptual content.

This prediction is illustrated schematically in Fig. 8. Based on our results in Experiments 1a-d, we hypothesized a sequential process where a set of features is initially activated, leading to a disjunctive representation of features, and subsequently a subset of features is selected, leading to a conjunctive representation of features (see left column in Fig. 8). For example, initially, a participant may activate a representation of *RED*, illustrated schematically by the two boxes located in the R-S (top-left) and R-D (bottom-left) quadrants. Subsequently, the participant may select the subset *SQUARE* within the previously activated representation, illustrated by the single box located in the R-S (top-left) quadrant. Finally, at the end of the trial, either the color or the shape of the target may be checked against this representation in order to identify the target (as explained in our introduction, see Fig. 8).

In Experiments 2a-b, we tested the prediction that participants generate a conjunctive representation of shape and color after being presented a sequence of verbal cues (see Fig. 9). Therefore, we predicted a dual-feature benefit in a sequential version of the visualization task (Experiment 2b), whereas we predicted similar reaction times for dual-feature and single-feature trials in a simultaneous version of the visualization task (Experiment 2a).

Method

Participants

Twenty-four participants with normal or corrected-to-normal vision participated, twelve in each experiment. All participants were undergraduate students at Leiden University, the Netherlands, participating for course credit or a small monetary reward.

Materials

Stimuli in Experiments 2a and 2b were identical to the visualization task introduced in Experiment 1c, except in the following. In both experiments, the cue words were separated vertically (see Fig. 9). Single-feature trials presented the same (color or shape) feature twice, whereas dual-feature trials presented a color and a shape feature.

Procedure

Procedures in Experiments 2a and 2b were identical to the visualization task introduced in Experiment 1c, except for in the following ways. In Experiment 2a, both cue features were presented simultaneously (for 1000 ms, followed by a 2000 ms blank screen), so that they would be visualized simultaneously (i.e., a *simultaneous* visualization task). In Experiment 2b, cue features were presented consecutively (each for 1000 ms, followed by a 2000 ms blank screen), so that they would be visualized one-at-a-time (i.e., a *sequential* visualization task; see Fig. 9). Task was therefore manipulated as a between-subjects factor, and trial-type as a within-subjects factor. Within each experiment, two feature orderings (color-shape vs shape-color) were blocked (2 blocks per experiment) and counterbalanced over participants.

Data analysis

Reaction times faster than 100 ms and slower than 2000 ms were

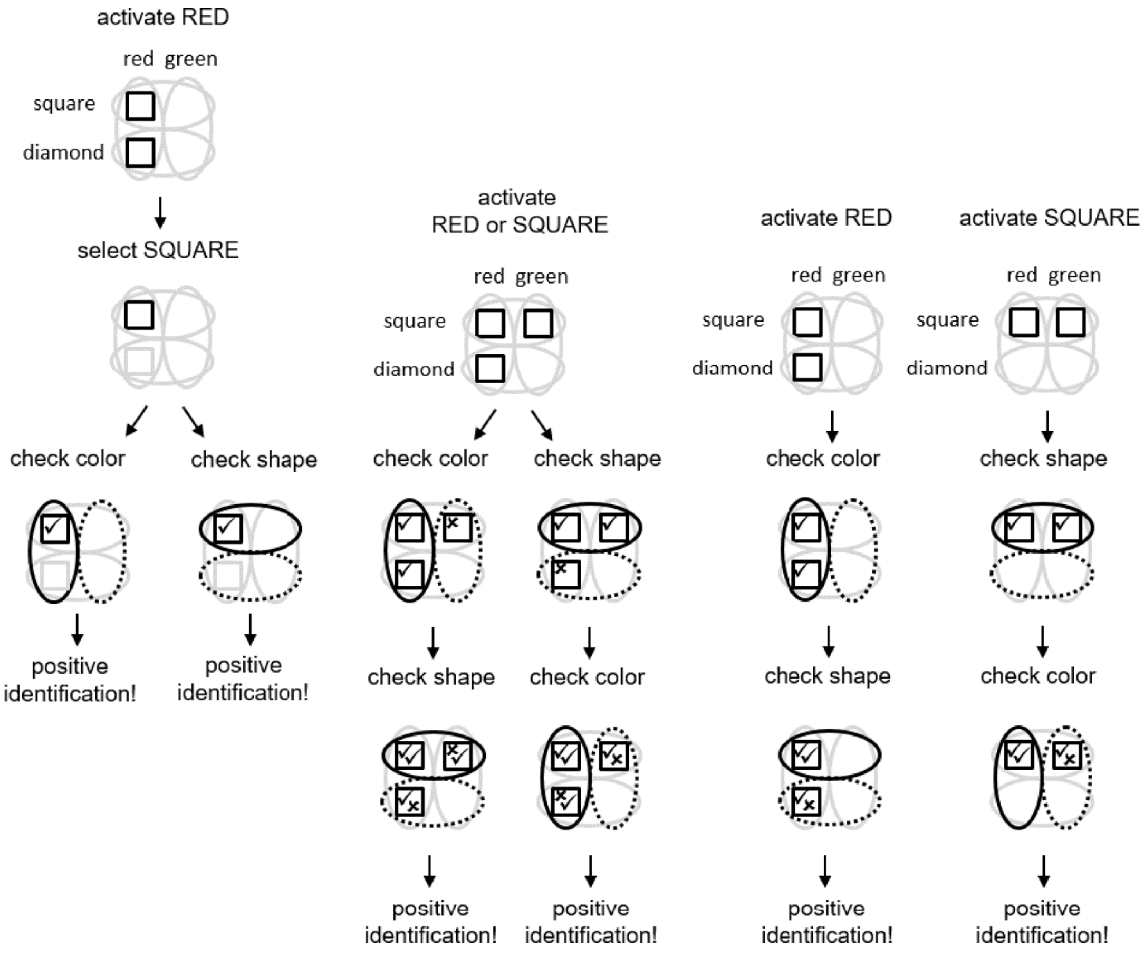

Fig. 8. Schematic representation of the perceptual simulation and identification process underlying the theoretical predictions for the visualization tasks in Experiments 2a-b. We represent the situation where participants activate (a) a feature conjunction (e.g., *RED* and *SQUARE*) by first activating a feature and then selecting a subset representation within the feature (see left column), (b) a feature disjunction (e.g. *RED* or *SQUARE*; see middle column), or a single feature (e.g., *SQUARE*; see right column).

discarded (<3%). The mean RTs for correct trials, as well as the proportion of accurate responses were included in the statistical analyses. Given that we did not observe any significant main-effects or interaction effects due to feature ordering, we collapsed the data over this factor.

Results and discussion

RTs and accuracies for each experiment were analyzed using a 2 × 2 (Task [simultaneous vs. sequential] × Trial-type [dual-feature vs. single-feature]) analysis of variance (ANOVA). In the RTs, we found a significant main effect for trial-type, $F(1,22) = 12.77$, $p < .01$, $\eta_p^2 = 0.37$, and a significant interaction between task and trial-type, $F(1,22) = 4.49$, $p < .05$, $\eta_p^2 = 0.17$. Dual-feature trials ($M = 581$, $SD = 207$) did not differ from single-feature trials ($M = 605$, $SD = 160$) in the simultaneous visualization task, $|t| < 1$, $p > .30$ (see Fig. 10). However, participants were faster for dual-feature trials ($M = 587$, $SD = 160$) compared to single-feature trials ($M = 682$, $SD = 210$) in the sequential visualization task, $t(11) = 3.54$, $p < .01$. In the accuracies, we only observed a significant main effect for trial-type $F(1,22) = 5.95$, $p < .05$, $\eta_p^2 = 0.21$. Participants were more accurate for dual-feature trials (sequential visualization task: $M = .93$, $SD = .05$; simultaneous visualization task: $M = .93$, $SD = .13$) compared to single-feature trials (sequential visualization task: $M = .92$, $SD = .07$; simultaneous visualization task: $M = .91$, $SD = .14$).

Consistent with Experiments 1a-d, we again observed a null-effect in a visualization task where both cue words were presented simultaneously. Importantly, however, we observed a clear dual-feature benefit in a sequential version of the visualization task. This finding is consistent with idea that participants initially activate a set of features, and subsequently select a subset of features, leading to a conjunctive representation.

Experiments 3a-b

In Experiments 1a-d and Experiment 2a, we failed to observe dual-feature benefits when participants were instructed to visualize two features simultaneously. However, one might wonder whether these null-effects may be explained by the fact that, in each experiment, single-feature and dual-feature conditions were presented in a blocked fashion where each block contained a number of trial repetitions. In addition, although well-controlled, it is conceivable that our previous experiments were sampling from a homogenous population of university students which may limit the generality of our results.

In light of these issues, we tested whether we could observe the pattern of results of Experiments 2a-b, testing a larger, more heterogeneous sample of participants and using a simpler experimental paradigm within a design where single-feature and dual-feature conditions were not blocked (i.e. the cueing conditions were randomized at the trial-level and each trial was presented once). In Experiment 3a, the cue was presented for a short temporal interval of 1000 ms, and was followed by a blank screen for 2000 ms until the onset of the target. In Experiment 3b, the cue was presented continuously for a long temporal interval of

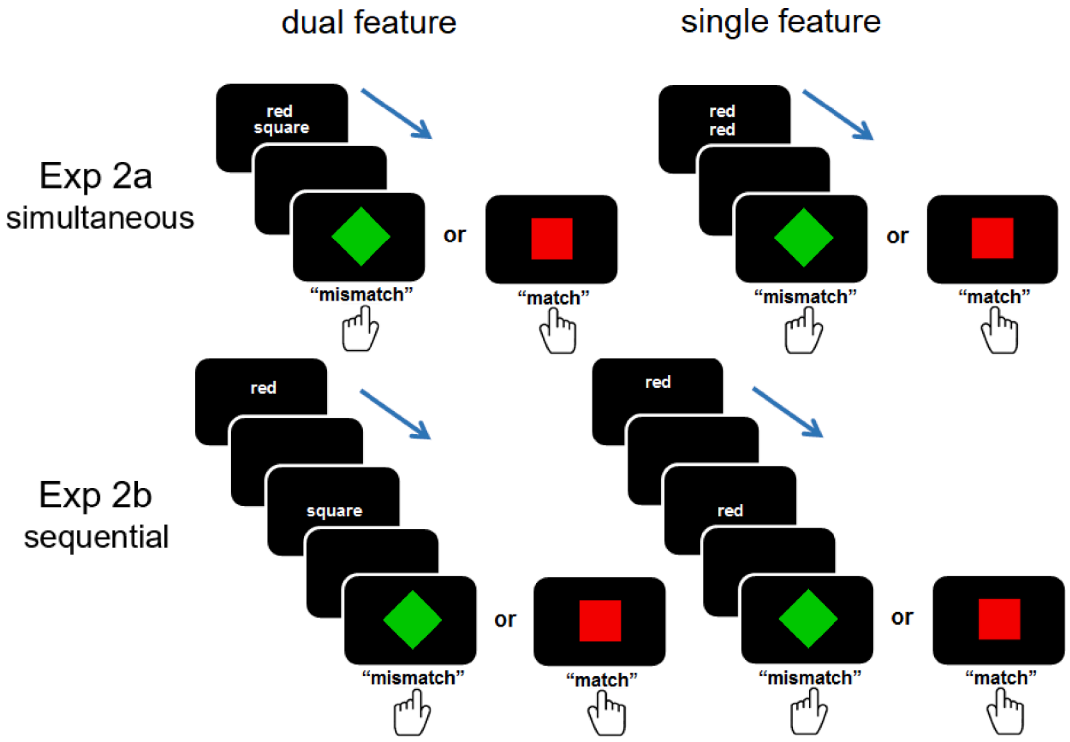

Fig. 9. Illustrations of the trials in Experiment 2a-b. The top panels display the simultaneous visualization task, the bottom panels display the sequential visualization task. The left panels display dual-feature trials, the right panels display single-feature trials. Each cue (1000 ms), was followed by a blank screen (2000 ms), and the sequence ended with a target (until response).

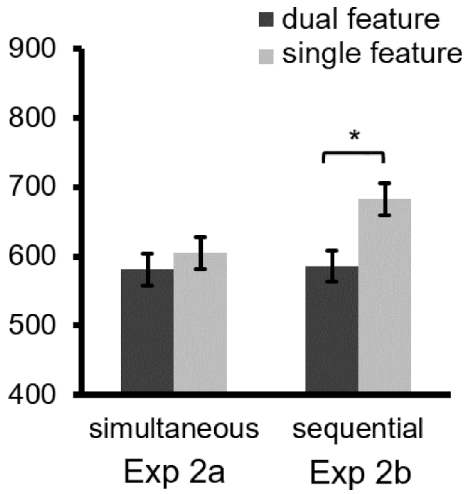

Fig. 10. RTs for each of the conditions in Experiments 2a-b. Error bars represent within-subjects standard errors (Loftus & Masson, 1994).

3000 ms until the onset of the target (see Fig. 11).

Please note that, based on our results in Experiments 1a-b and 2a, we expect that the temporal restriction in cue presentation will force participants to visualize the two features simultaneously in Experiment 3a, whereas we expect that relieving the temporal restriction in cue presentation will make it possible for participants to visualize the two feature sequentially in Experiment 3b. Therefore, we predicted similar reaction times for dual-feature and single-feature trials in Experiment 3a and we predicted a dual-feature benefit in Experiment 3b.

Method

Participants

The participants were recruited using the Amazon Mechanical Turk (https://www.mturk.com)[5]. Four-hundred and ten participants participated[6], 203 in Experiment 3a and 207 in Experiment 3b. All participants completed an informed consent form prior to the start of the experiment, were from the United States and were paid $1.00 for approximately 5–10 min of their time (see Buhrmester, Kwang, & Gosling, 2011).

Materials

Stimuli in Experiments 3a and 3b were identical to Experiments 2a and 2b, except for the vertical vs. horizontal positioning of the two words in the verbal cues.

Procedure

Both Experiments 3a and 3b were similar to Experiments 2a and 2b, except for the following aspects. In Experiment 3a, the cue features were presented for 1000 ms, followed by a 2000 ms blank screen (short cue task). In Experiment 3b, cue features were presented continuously for 3000 ms until target onset (long cue task; see Fig. 11). Task was therefore manipulated as a between-subjects factor, and trial-type as a within-subjects factor constituting a 2 (task: short cue vs. long cue) × 2 (trial:

[5] Internet-based experimentation was chosen due to the simple nature of the experiments and the large number of participants that we desired. Previous studies had shown that Internet-based behavioral experiments generate reliable data comparable to those based on more traditional data acquisition in the lab (e.g., Zwaan & Pecher, 2012).

[6] The planned sample sizes were 200 participants for each condition. The actual sample sizes varied because participants were included who successfully completed the experiment but did not sign off on the task on the Amazon Mechanical Turk website. We did not have any good a priori reasons to exclude the latter group of participants, so we included them in our analyses.

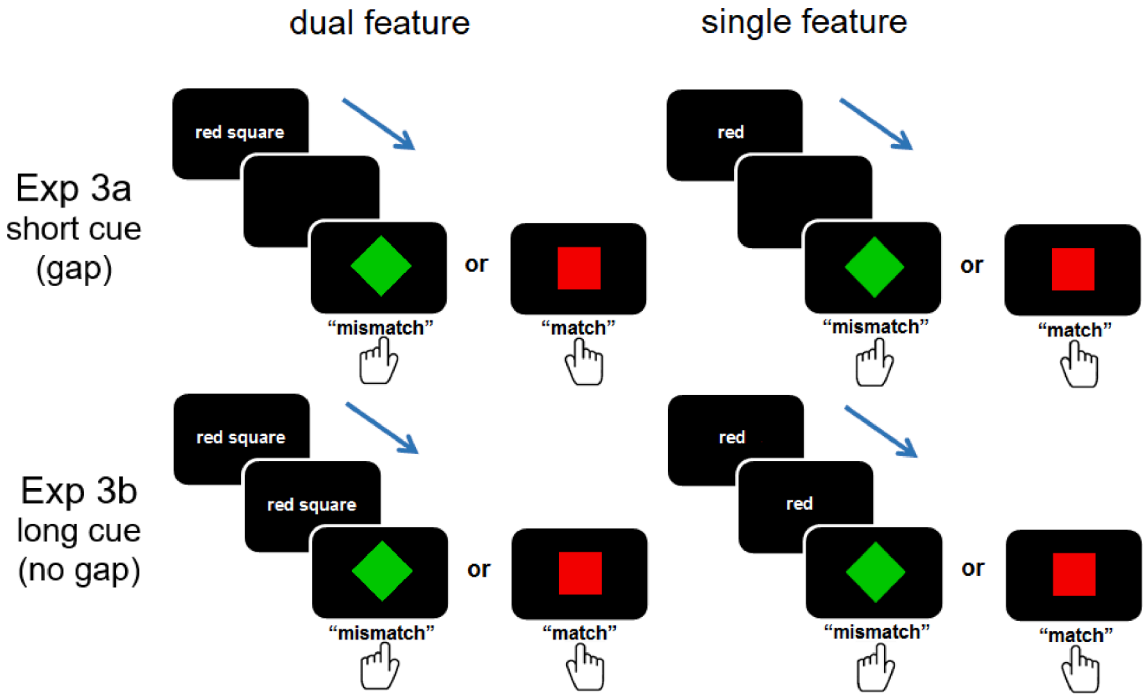

Fig. 11. Illustrations of the trials in Experiment 3a-b. The top panels display the short cue visualization task, the bottom panels display the long cue visualization task. The left panels display dual-feature trials, the right panels display single-feature trials. In the short cue task, each cue was presented for 1000 ms, followed by a 2000 ms blank screen, and the target (until response). In the long cue task, each cue was presented continuously for 3000 ms and was followed by the target (until response).

single-feature vs. dual-feature) design. Within each task, unique single-feature and dual-feature trials were presented once in a random order. The experiment consisted of 32 trials.

Data analysis

Reaction times faster than 100 ms and slower than 3000 ms were discarded (<4% in both experiments). The mean RTs for correct trials, as well as the proportion of accurate responses were included in the statistical analyses.

Results and discussion

RTs and accuracies for each experiment were analyzed using a 2 × 2 (Task [short cue vs. long cue] × Trial-type [dual-feature vs. single-feature]) analysis of variance (ANOVA). In the RTs, we found significant main effects for task, $F(1,408) = 8.90$, $p = .003$, $\eta_p^2 = 0.02$, trial-type, $F(1,408) = 16.65$, $p < .001$, $\eta_p^2 = 0.04$, and a significant interaction between task and trial-type, $F(1,408) = 23.914$, $p < .001$, $\eta_p^2 = 0.06$. Dual-feature trials ($M = 938$, $SD = 348$) did not differ from single-feature trials ($M = 930$, $SD = 337$) in the short cue task, $|t| < 1$, $p > .50$ (see Fig. 12). However, participants were faster for dual-feature trials ($M = 991$, $SD = 374$) compared to single-feature trials ($M = 1083$, $SD = 396$) in the long cue task, $t(206) = 6.41$, $p < .001$. In the accuracies, we found significant main effects for task, $F(1,408) = 35.02$, $p < .001$, $\eta_p^2 = 0.08$, trial-type, $F(1,408) = 14.38$, $p < .001$, $\eta_p^2 = 0.03$, and an interaction between task and trial-type that approached significance, $F(1,408) = 2.81$, $p = .09$, $\eta_p^2 = 0.01$. Participants were more accurate for dual-feature trials ($M = .97$, $SD = .05$) compared to single-feature trials ($M = .94$, $SD = .08$) in the long cue task, $t(206) = 4.27$, $p < .001$. However, dual-feature trials ($M = .93$, $SD = .08$) did not differ from single-feature trials ($M = .92$, $SD = .09$) in the short cue task, $t(202) = 1.37$, $p = .17$.

Consistent with Experiments 2a-b, we did not observe a dual-feature benefit when participants were presented a brief verbal cue, but we did observe a clear dual-feature benefit when participants were presented a longer verbal cue (see Fig. 11). This is consistent with the idea that the

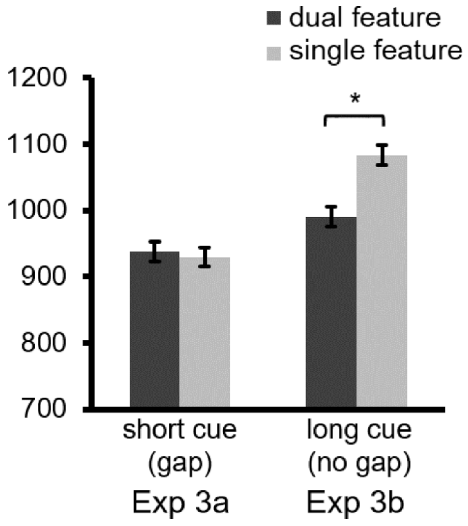

Fig. 12. RTs for each of the conditions in Experiments 3a-b. Error bars represent within-subjects standard errors (Loftus & Masson, 1994).

generation of a conjunctive representation of features during language comprehension is an incremental process that needs time to unfold. Interestingly, although participants in Experiments 3a-b overall took longer to respond (which is a general difference often observed between university students and more general populations, see; Zwaan & Pecher, 2012), the magnitude of the dual-feature benefit in Experiment 3b was very comparable to the magnitude of the dual-feature benefit in Experiment 2b (92 ms vs. 95 ms).

In addition to the differential effects of cueing (i.e., the contrast between dual- vs single-feature trials), we also observed a between-subjects main-effect where the short-cue group (Experiment 3a) responded overall faster than the long-cue group (Experiment 3b). This

effect had the same directionality as the non-significant difference in response-times observed between the simultaneous and sequential cue groups in Experiment 2a-b. Importantly, the between-subjects effect observed in the current experiment cannot be explained by our proposed mechanism and could represent a general facilitation in response preparation triggered by the offset of the cue (e.g., Forbes & Klein, 1996). We will address this point in Experiment 5 where we performed a within-subject manipulation of sequential vs. simultaneous conditions in a visualization task.

Experiments 4a-b

Experiments 2a-b and Experiments 3a-b suggest that generating representations of feature conjunctions depends on the visualization of features over time. Although our original rationale was based on the idea that language may play a key role in the incremental construction of perceptual representations (cf. Zwaan & Madden, 2005), Experiments 2a-b and 3a-b did not explicitly manipulate linguistic structure in the verbal cues (i.e., they manipulated word pairs that could be interpreted as having linguistic structure, but could also be interpreted as merely disconnected words). In Experiments 4a-b we tested whether the linguistic structure of verbal cues influences the way feature combinations are represented in a visualization task.

In Experiment 4a, we presented participants with two types of dual-feature sentences as cues. In simultaneous dual-feature trials, the two features were presented simultaneously within a sentence fragment (e.g., "the object is" => "a RED SQUARE"), whereas in sequential dual-feature trials, the two features were described consecutively over sentence fragments (e.g., "the RED object is" => "a SQUARE" or "the SQUARE object is" => "RED"). In addition, we also presented simple sentences as single-feature trials (e.g., "the object is" => "RED" or "the object is" => "a SQUARE"). Our initial predictions were similar to the predictions we made in Experiments 2a-b: sequential dual-feature trials (e.g., "the RED object is" => "a SQUARE") should be faster than single-feature trials (e.g., "the object is" => "a SQUARE"), whereas simultaneous dual-feature trials (e.g., "the object is a RED SQUARE") and single-feature trials should show similar reaction-times.

We ran an additional experiment (Experiment 4b) where we replicated two of the conditions in Experiment 4a: the sequential dual-feature condition (e.g., "the RED object is" => "a SQUARE") and the single-feature condition (e.g., "the object is" => "a SQUARE"). In addition, we tested a new simultaneous dual-feature condition where the two features were presented at the start of the sentence (e.g., "the RED SQUARE is" => "an object"). Please note that this additional dual-feature condition is the last remaining possibility to describe two features simultaneously in the specific sentence format we used: We were interesting in examining whether we could replicate part of our results observed in Experiment 4a, and whether performance in the new dual-feature condition would confirm our predictions.

Method

Participants

Participants were recruited using the Amazon Mechanical Turk. Four-hundred and ten participants participated, 203 in Experiment 4a and 207 in Experiment 4b. All participants completed an informed consent form prior to the start of the experiment, were from the United States and were paid $1.00 for approximately 5–10 min of their time.

Materials

Both Experiments 4a and 4b were similar to Experiments 3a and 3b, except for the following aspects. In Experiment 4a, we displayed (a) dual-feature trials where the two features were presented simultaneously within the second fragment of the cue sentence (e.g., "the OBJECT is" => "a RED SQUARE"), (b) dual-feature trials where the two features were described consecutively over fragments of the cue sentence (e.g., "the RED OBJECT is" => "a SQUARE" / "the SQUARE OBJECT is" => "RED"), and (c) single-feature trials where one feature was presented in the last fragment of the cue sentence (e.g., "the OBJECT is" => "RED" / "the OBJECT is" => "SQUARE"). In Experiment 4b, (a) dual-feature trials where the two features were described consecutively over fragments of the cue sentence (e.g., "the RED OBJECT is" => "a SQUARE" / "the SQUARE OBJECT is" => "RED"), (b) single-feature trials where one feature was presented in the last fragment of the cue sentence (e.g., "the OBJECT is" => "RED" / "the OBJECT is" => "SQUARE"), and (c) dual-feature trials where the two features were presented simultaneously in the first fragment of the cue sentence (e.g., "the RED SQUARE is" => "an object").

Procedure

Both Experiments 4a and 4b were similar to Experiments 3a and 3b, except for the following aspects. In both Experiments 4a and 4b, the first part of the cue sentence was presented continuously until the onset of the target (for a duration of 3000 ms), whereas the second part of the cue sentence was presented 1500 ms after onset of the first part of the sentences until the onset of the target (for a duration of 1500 ms). Finally, this was followed by the target (until response; see Fig. 13). Both Experiments 4a and 4b each consisted of 3 within-subjects conditions, one for each type of cue sentence. Within each within-subjects condition, unique trials were presented once in a random order. Each experiment consisted of 48 trials.

Data analysis

Reaction times faster than 100 ms and slower than 3000 ms were discarded (<5% in both experiments). The mean RTs for correct trials, as well as the proportion of accurate responses were included in the statistical analyses.

Results

RTs and accuracies for each experiment were analyzed using one-way analysis of variance (ANOVA) (Trial-type [simultaneous dual-feature vs. sequential dual-feature vs, single-feature]). In Experiment 4a, we found a significant effect for trial-type in the RTs, $F(2,404) = 11.14$, $p < .001$, $\eta_p^2 = 0.05$. Participants were faster for dual-feature trials where the two features were presented simultaneously in the second fragment of the cue sentence ($M = 950$, $SD = 371$, e.g., "the OBJECT is a RED SQUARE") than sequential dual-feature trials ($M = 1006$, $SD = 369$, e.g., "the RED OBJECT is a SQUARE"), $t(202) = 3.74$, $p < .001$. The simultaneous dual-feature trials were also faster than single-feature trials ($M = 1022$, $SD = 377$, e.g., "the OBJECT is a SQUARE"), $t(202) = 4.41$, $p < .001$. Sequential dual-feature trials did not differ from single-feature trials, $t(202) = 0.98$, $p = .33$ (see Fig. 14). In the accuracies, the effect for trial-type approached significance, $F(2,404) = 2.71$, $p = .07$, $\eta_p^2 = 0.01$. Participants were equally accurate for dual-feature trials where the two features were presented simultaneously in the second fragment of the cue sentence ($M = .95$, $SD = .07$) and sequential dual-feature trials ($M = .96$, $SD = .07$), $t(202) = 0.36$, $p = .72$. Simultaneous dual-feature trials did not differ from single-feature trials ($M = .94$, $SD = .08$), $t(202) = 1.70$, $p = .09$. Sequential dual-feature trials were marginally more accurate than single-feature trials, $t(202) = 2.07$, $p = .04$.

In Experiment 4b, we found a significant effect for trial-type in the RTs, $F(2,412) = 11.63$, $p < .001$, $\eta_p^2 = 0.05$. Sequential dual-feature trials ($M = 1023$, $SD = 285$, e.g., "the RED OBJECT is a SQUARE") did not differ from single-feature trials ($M = 1042$, $SD = 280$, e.g., "the OBJECT is a SQUARE"), $t(206) = 1.62$, $p = .11$. Sequential dual-feature trials were faster than dual-feature trials where the two features were presented simultaneously in the first fragment of the cue sentence ($M = 1085$, $SD = 324$, e.g., "the RED SQUARE is an OBJECT"), $t(206) = 4.41$, $p < .001$. Single-feature trials were also faster than dual-feature trials where the two features were presented simultaneously in the first fragment of the cue sentence, $t(206) = 3.17$, $p = .002$ (see Fig. 14). In the accuracies, we

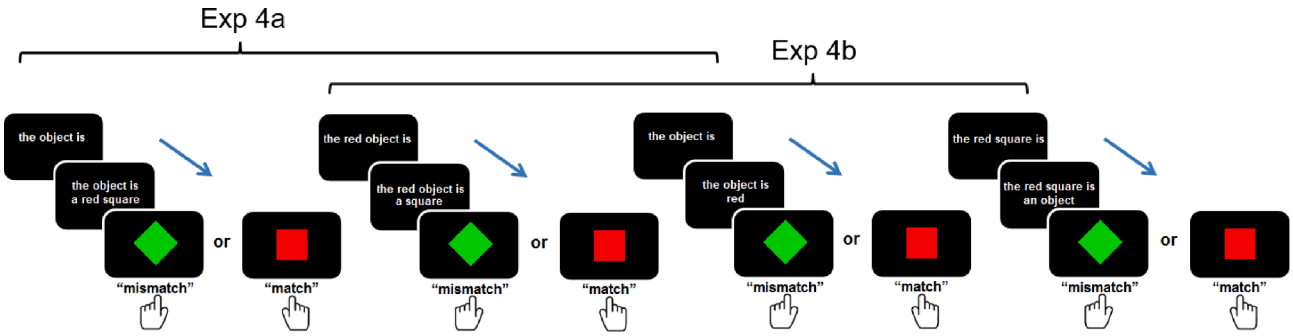

Fig. 13. Illustrations of the trials in Experiment 4a-b. In Experiment 4a, we displayed two types of dual-feature trials (e.g., "the OBJECT is a RED SQUARE", and "the RED OBJECT is a SQUARE" / "the SQUARE OBJECT is RED"), and one type of single-feature trials (e.g., "the OBJECT is RED" / "the OBJECT is SQUARE"). In Experiment 4b, we displayed two types of dual-feature trials (e.g., "the RED OBJECT is a SQUARE" / "the SQUARE OBJECT is RED", and "the RED SQUARE is an OBJECT"), and one type of single-feature trials (e.g., "the OBJECT is RED" / "the OBJECT is SQUARE"). In each task, the first part of the sentence was presented continuously until the onset of the target (for a duration of 3000 ms), whereas the second part of the sentence was presented 1500 ms after onset of the first part of the sentences until the onset of the target (for a duration of 1500 ms). Finally, this was followed by the target (until response).

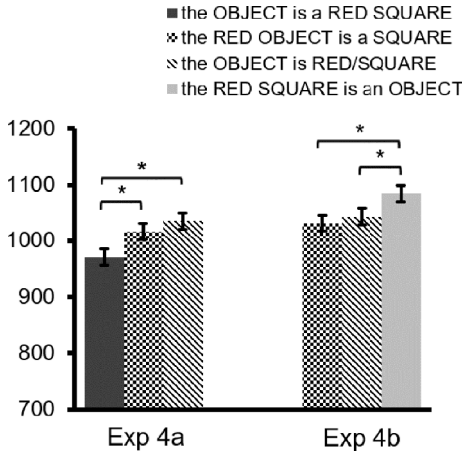

Fig. 14. RTs for each of the conditions in Experiments 4a-b. Error bars represent within-subjects standard errors (Loftus & Masson, 1994).

found a significant effect for trial-type, $F(2,412) = 4.23$, $p = .02$, $\eta_p^2 = 0.02$. Participants were equally accurate for sequential dual-feature trials ($M = .96$, $SD = .06$) and single-feature trials ($M = .96$, $SD = .06$), $t(206) = 0.54$, $p = .59$. Sequential dual-feature trials were more accurate than dual-feature trials where the two features were presented simultaneously in the first fragment of the cue sentence ($M = .95$, $SD = .08$), $t(206) = 2.13$, $p = .03$. Single-feature trials were also more accurate than the simultaneous dual-feature trials, $t(206) = 2.56$, $p = .01$.

Discussion

In Experiment 4a, we observed a pattern of differences between the cueing conditions that did not agree with our initial predictions. Given the large sample-size and therefore high statistical power in this experiment, we decided to interpret these differences between conditions and reconsider the way we had originally translated our theoretical assumptions (see introduction to Experiment 2a-b) into specific predictions for the cues used in Experiment 4a.

Critically, our initial prediction assumed that the word "object" would function as a simple place-holder in the verbal cue and would therefore not play any role in the generation of the perceptual representation. If, instead, one assumes that the word "object" acts as a meaningful cue in its own right, and refers to any of the four potential target stimuli in the experiment (in other words, that it acts just like the cue words "red" or "square", only more general in scope), the results in Experiment 4a become interpretable and consistent with our assumptions of an initial disjunctive activation and subsequent conjunctive selection of features (see introduction to Experiments 2a-b). Under this assumption, we generated a new set of postdictions (see Fig. 15).

First, let's look at the cue "the OBJECT is a RED SQUARE". When confronted with the initial fragment "the OBJECT is" the participant first activates a general representation of all possible targets which is illustrated by four boxes in the space of stimuli (see leftmost column in Fig. 15). Then, when confronted with the subsequent fragment "a RED SQUARE" the participant selects a restricted representation of only *one* of the four possible target stimuli in the experiment (i.e., the red square). This representation is illustrated schematically by the single box located in the R-S (top-left) quadrant. At the end of the trial, the color or the shape of the target may be checked against this representation. Whichever may be the case, only a single comparison is needed in order to uniquely match the target to the represented quadrant (i.e., R-S). On average, therefore, the visualization activated using cue "the OBJECT is a RED SQUARE" only requires a single comparison to identity.

The situation is different with the cue "the RED OBJECT is a SQUARE". When confronted with the initial fragment "the RED OBJECT is" the participant first activates a disjunctive representation of all targets and all red targets which is illustrated by four boxes in the space of stimuli (see second column from the left in Fig. 15). Then, when confronted with the subsequent fragment "a SQUARE" the participant then selects a more restricted representation of the two square target stimuli in the experiment (i.e., the red square and the green square). This representation is illustrated schematically by the two boxes located in the R-S (top-left) and G-S (top-right) quadrants. At the end of the trial, the color or the shape of the target may be checked against this representation. If the color feature is checked first, then only a single comparison is needed in order to uniquely match the target to the represented quadrant (i.e., R-S). If, on the other hand, the shape feature is checked first, then two of the represented quadrants will coincide with the target (R-S and G-S), and color will additionally need to be checked in order to determine that the target uniquely matches the R-S quadrant. On average, therefore, the visualization activated using cue "the RED OBJECT is a SQUARE" requires 1.5 comparisons to identity.

The cue "the OBJECT is RED" results in a similar situation as the previous case. When confronted with the initial fragment "the OBJECT is" the participant first activates a general representation of all possible targets which is illustrated by four boxes in the space of stimuli (see second column from the right in Fig. 15). Then, when confronted with the subsequent fragment "RED" the participant selects a more restricted representation of the two red target stimuli in the experiment (i.e., the red square and the red diamond). This representation is illustrated schematically by the two boxes located in the R-S (top-left) and R-D

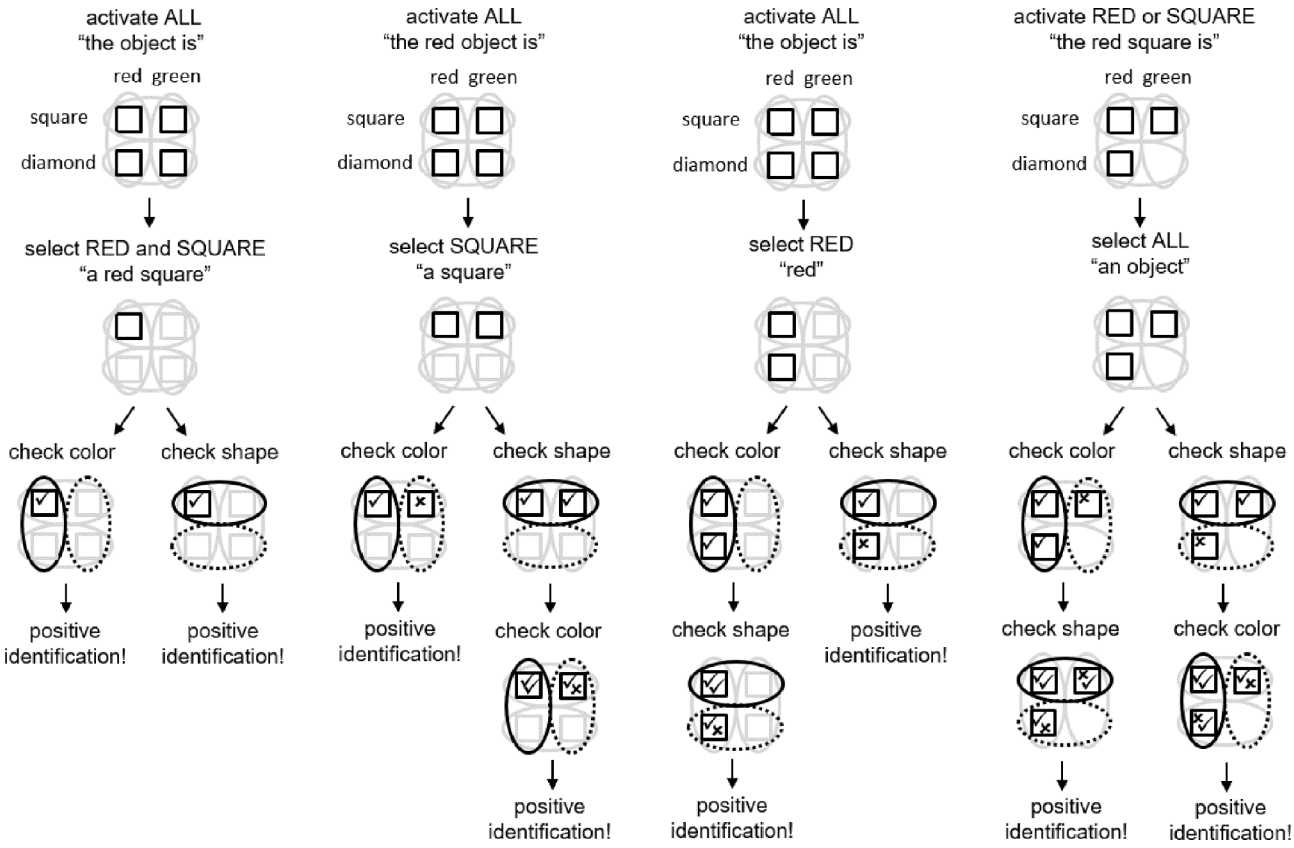

Fig. 15. Schematic representation of the perceptual simulation and identification process underlying the theoretical postdictions for the visualization tasks in Experiments 4a-b.

(bottom-left) quadrants. At the end of the trial, the color or the shape of the target may be checked against this representation. If the color feature is checked first, then two of the represented quadrants will coincide with the target (R-S and R-D), and shape will additionally need to be checked in order to determine that the target uniquely matches the R-S quadrant. If, on the other hand, the shape feature is checked first, then only a single comparison is needed in order to uniquely match the target to the represented quadrant (i.e., R-S). On average, therefore, the visualization activated using cue "the OBJECT is RED" also requires 1.5 comparisons to identity.

Finally, the situation is again different with the cue "the RED SQUARE is an OBJECT". When confronted with the initial fragment "the RED SQUARE is" the participant first activates a disjunctive representation of targets which are red or squared, illustrated by three boxes in the space of stimuli (see rightmost column in Fig. 15). Then, when confronted with the subsequent fragment "an OBJECT" the participant selects the whole representation of the three activated target stimuli (i.e., the red square, the red diamond and the green square). This representation is illustrated schematically by the three boxes located in the R-S (top-left), R-D (bottom-left) and G-S (top-right) quadrants. At the end of the trial, the color or the shape of the target may be checked first against this representation. Whichever may be the case, two comparisons are needed in order to uniquely match the target to the represented quadrant (i.e., R-S). If the color feature is checked first, then two of the represented quadrants will coincide with the target (R-S and R-D), and shape will additionally need to be checked in order to determine that the target uniquely matches the R-S quadrant. If the shape feature is checked first, then two of the represented quadrants will coincide with the target (R-S and G-S), and color will additionally need to be checked in order to determine that the target uniquely matches the R-S quadrant. On average, therefore, the visualization activated using cue "the RED SQUARE is an OBJECT" requires 2 comparisons to identity.

Using this amended translation of assumptions into predictions, one can interpret the results in Experiments 4a-b in terms of an initial disjunctive activation and subsequent conjunctive selection of features. This account can explain why "the OBJECT is a RED SQUARE" is the fastest condition, and why "the RED OBJECT is a SQUARE" and "the OBJECT is RED" conditions are slower and equal to each other in Experiment 4a (see Fig. 14). In addition, it explains the finding that the last condition tested in Experiment 4b ("the RED SQUARE is an OBJECT") is the slowest condition overall. Also, it is noteworthy that the numerical difference in reaction-time between "the OBJECT is a RED SQUARE" and "the RED SQUARE is an OBJECT" conditions is +- 100 ms, which is similar to the dual-feature benefits observed in Experiments 2b and 3b. The dual-feature benefit was predicted based on the expectation that the fastest condition would requires 1 comparison whereas the slowest condition requires 2 comparisons for identification (see also Fig. 8), which coincides with the difference in number of comparisons postulated for identification in the fastest and slowest condition in Experiments 4a-b (see Fig. 15). Overall, these findings demonstrate that the linguistic structure of verbal cues influences the way people represent features in a visualization task.

Experiment 5

In Experiment 2b we observed a dual-feature benefit when participants visualized perceptual features sequentially, which is consistent with the idea that the generation of a conjunctive representation depends on the temporal concatenation of linguistic cues. Although we hypothesized that language plays a key role in the incremental construction of perceptual representations (cf. Zwaan & Madden, 2005), Experiment 2b did not present verbal cues containing explicit linguistic structure. In Experiment 5 we therefore used linguistically structured cues where perceptual features were presented either sequentially or

simultaneously. We presented participants with two types of dual-feature cues and single-feature cues.

In sequential dual-feature trials, the two features were presented consecutively over the cue displays (e.g., "the RED" => "SQUARE", or "the SQUARE is" => "RED"). In simultaneous dual-feature trials, both features were presented in the second cue display (e.g., "—" => "the RED SQUARE", or "—" => "the SQUARE is RED"). In addition, in single-feature trials one feature was presented in the second cue display (e.g., "—" => "SQUARE", or "—" => "RED"). In this manner, any potential differences between sequential dual-feature trials and single-feature trials can be attributed to the extra feature presented in the first cue display (i.e., the second cue displays were identical), and any potential differences between simultaneous dual-feature trials and single-feature trials can be attributed to the extra feature presented in the second cue display (i.e., the first cue displays were identical).

In Experiment 2a-b and Experiment 3a-b we unexpectedly observed that the participants assigned to a simultaneous dual-feature condition responded overall faster than the participants assigned to a sequential dual-feature condition. Importantly, this effect was not hypothesized and cannot be explained by our proposed mechanism. Because these conditions were manipulated between-subjects this effect could represent a general response preparation effect triggered by differences in temporal onsets/offsets between conditions.

In Experiment 5 we again investigated the contrast between simultaneous and sequential cueing conditions in a design aimed to minimize any strategic response preparation effects for each cueing condition. To do this, we manipulated the three cueing condition within-subjects at the level of individual trials in order to be able to meaningfully interpret potential differences between the simultaneous and sequential dual-feature conditions.

Our predictions were similar to the predictions we made in Experiments 2a-b: sequential dual-feature trials should be faster than single-feature trials, whereas simultaneous dual-feature trials and single-feature trials should show similar reaction-times. Also, given our within-subjects design, we additionally predicted that sequential dual-feature trials should be faster than simultaneous dual-feature trials.

Method

Participants

Participants were recruited using the Amazon Mechanical Turk. Two-hundred and six participants participated in Experiment 5. All participants completed an informed consent form prior to the start of the experiment, were from the United States and were paid $1.00 for approximately 5–10 min of their time.

Materials

Experiment 5 was similar to Experiment 4a, except for in the following ways. We displayed (a) dual-feature trials where the two features were presented consecutively over the cue displays (e.g., "the RED" => "SQUARE", "the SQUARE is" => "RED"), (b) dual-feature trials where the two features were presented simultaneously in the second cue display (e.g., "—" => "the RED SQUARE", "—" => "the SQUARE is RED"), and (c) single-feature trials where one feature was presented in the second cue display (e.g., "—" => "SQUARE", "—" => "RED").

Procedure

Experiment 5 was similar to Experiment 4a, except for in the following ways. The first cue display was presented for a duration of 1500 ms, whereas the second cue display was presented for a duration of 1500 ms (for a total cue duration of 3000 ms). Finally, this was followed by the target (until response; see Fig. 16). The experiment consisted of 3 within-subjects conditions, one for each type of cue. Within each within-subjects condition, unique trials were presented once in a random order. Each experiment consisted of 48 trials.

Data analysis

Reaction times faster than 100 ms and slower than 3000 ms were discarded (<6%). The mean RTs for correct trials, as well as the proportion of accurate responses were included in the statistical analyses.

Results and discussion

RTs and accuracies for each experiment were analyzed using one-way analysis of variance (ANOVA) (Trial-type [simultaneous dual-feature vs. sequential dual-feature vs, single-feature]). We observed a

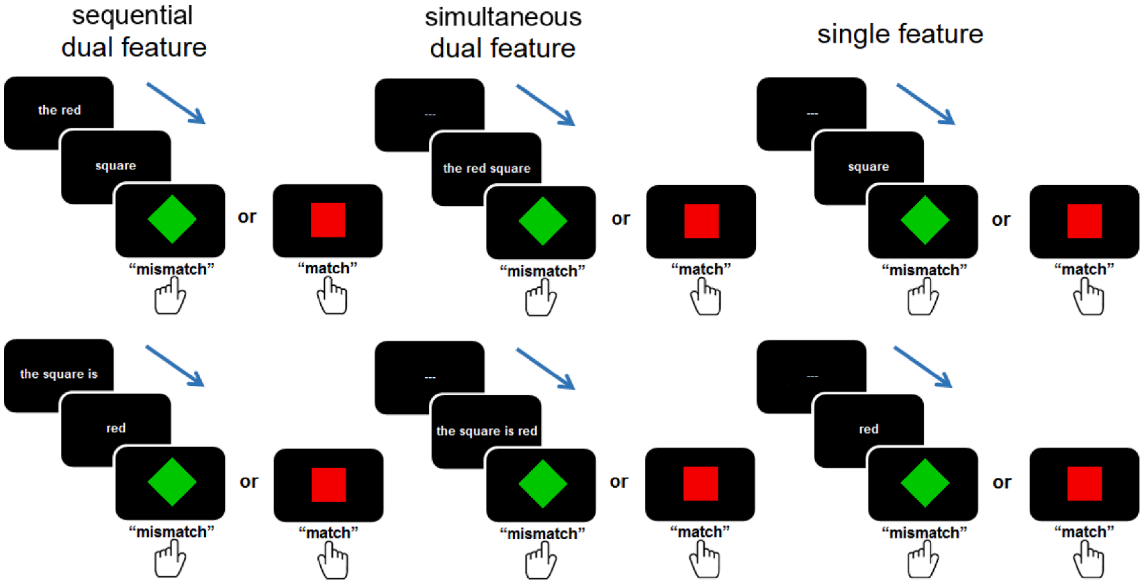

Fig. 16. Illustrations of the trials in Experiment 5. In the experiment we displayed sequential dual-feature trials, simultaneous dual-feature trials and single-feature trials (the rows display different instantiations of the trials). In all tasks, the two cue displays were presented for 1500 ms each (for a total duration of 3000 ms). The cue displays were followed by the target (until response).

significant effect of trial-type in the RTs, $F(2,410) = 19.75$, $p < .001$, $\eta_p^2 = 0.09$. Sequential dual-feature trials ($M = 864$, $SD = 342$, e.g., "the RED" => "SQUARE", "the SQUARE is" => "RED") showed faster responses than simultaneous dual-feature trials ($M = 933$, $SD = 341$, e.g., "—" => "the RED SQUARE", "—" => "the SQUARE is RED"), $t(205) = 4.17$, $p < .001$, and showed faster responses than single feature trials ($M = 958$, $SD = 357$, e.g., "—" => "SQUARE", "—" => "RED"), $t(205) = 6.00$, $p < .001$. Simultaneous dual-feature trials did not differ from single-feature trials, $t(205) = 1.80$, $p = .07$ (see Fig. 17). In the accuracies, the effect for trial-type was significant, $F(2,410) = 5.40$, $p < .01$, $\eta_p^2 = 0.03$. Participants were equally accurate for sequential dual-feature trials ($M = .94$, $SD = .09$) and simultaneous dual-feature trials ($M = .93$, $SD = .11$), $t(205) = 1.04$, $p = .30$. Sequential dual-feature trials were more accurate than single-feature trials ($M = .92$, $SD = .11$), $t(205) = 3.25$, $p = .001$. Simultaneous dual-feature trials were more accurate than single-feature trials, $t(205) = 2.10$, $p = .04$.

Consistent with Experiments 1a-d, Experiment 2a, and Experiment 3a we did not observe a dual-feature benefit when participants were visualizing the perceptual features simultaneously, and consistent with Experiment 2b and Experiment 3b we did observe a dual-feature benefit when participants were visualizing the features sequentially (see Fig. 17). Please note that the magnitude of this sequential dual-feature benefit (94 ms) was very comparable to the benefits observed previously in Experiment 2b (95 ms) and Experiment 3b (92 ms).

Combined, these findings are again consistent with our proposal that the representation of specific feature combinations is driven by a temporal concatenation of verbal cues: initially, participants activate a disjunctive representation and subsequently they select a conjunctive representation of features. This interpretation is bolstered by the fact that the sequential dual-feature trials were not only faster than single-feature trials, but were also faster than simultaneous dual-feature trials in a direct within-subject comparison (cf. Experiment 2a-b and Experiment 3a-b).

General discussion

In Experiments 1a, 1b, 1c, 1d, 2a, 3a, and 5 we observed that participants were equally fast at identifying a perceptual target when they were verbally cued to simultaneously visualize both of its two features (i.e., color and shape), vs. only one of its two features (i.e., color or shape). Importantly however, in Experiments 2b, 3b, 4a, and 5 we observed clear dual-feature benefits (i.e., two features were faster than one feature) when they were forced (or allowed to) perform a sequence of visualizations. In Experiments 4a, 4b and 5 we observed that when target features were cued incrementally in a sentence this increased the speed of target identification relative to when target features were cued immediately in a sentence. We interpret these findings as showing that within a cued visualization participants tend to activate a general disjunctive set of features, and only across a sequence of cued visualizations are they able to select a specific conjunctive subset of features. This suggests that the generation of conjunctive perceptual representations during comprehension depends of the concatenation of linguistic cues.

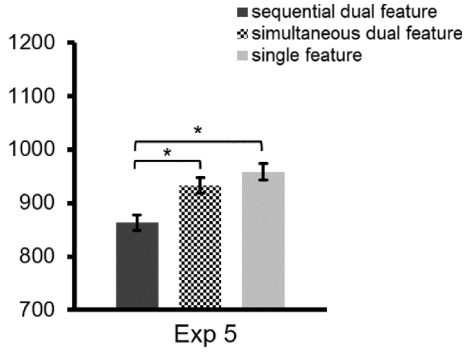

Fig. 17. RTs for each of the conditions in Experiments 5a-b. Error bars represent within-subjects standard errors (Loftus & Masson, 1994).

Alongside the visualization tasks (i.e., where participants were verbally cued to generate a visual representation), we included various control tasks in Experiments 1a-d in order to exclude alternative explanations for the null-effects observed during visualization. In all these control tasks we found that participants were faster (+- 90 ms) when cued with two features (both color and shape) compared to only one feature (color or shape). This allowed us to conclude that the null-effects observed in the visualization tasks cannot be spuriously explained by (a) a selective processing of the cue words (i.e., participants attending to only one of the two words), b) the spatial integration of the target's visual features (i.e., color and shape being presented at the same location), c) the task demands of visual identification (i.e., the relative ease of identifying a two-feature visual object), and d) a mismatch between the modality of the verbal cue and the visual target (i.e., the fact that words and pictures are different types of stimuli). In addition, Experiments 3a and 3b showed that both the null-effects as well as the dual-feature benefits during visualization can be observed in more general, heterogeneous non-student samples and Experiment 5 also demonstrated this generalization when the simultaneous and sequential visualization tasks were manipulated within-subjects.

We argue that the null-effects we observed in Experiments 1a, 1b, 1c, 1d, 2a, 3a, and 5 are due to the simultaneous visualization manipulation. However, one may be tempted to argue that even in the simultaneous condition participants would have had to perform a sequence of two fixations in order to read the two words. However, we know from previous research this is not the case: participants can process a few short words within a single fixation (McConkie & Rayner, 1975).

Could the null-effects observed during simultaneous visualization be explained by participants not having sufficient time to process the two words in the verbal cue? This explanation appears unlikely considering that participants had 1.5 s to process each word, whereas on average, people only need about.3 s per word (see Trauzettel-Klosinski & Dietz, 2012). In addition, clear dual-feature benefits were observed in control tasks–presented alongside the visualization tasks–that used identical verbal cues (Experiments 1a-b), which demonstrates that participants were processing both cue words in these experiments.

Can the dual-feature benefits observed in the verbal control tasks (Experiments 1a and 1b), be explained by a general tendency of participants to read target words from left to right? The idea behind this account is that left-to-right reading will selectively slow down responses for single-feature trials in the verbal control task (i.e., on half of the single-feature trials participants will need to read both words in order to respond). However, there is a problem with this explanation, in that it implicitly presupposes that the representation generated during dual-feature trials was conjunctive to begin with. Only if the two features are represented conjunctively (in other words, the participant expects the target to be both *RED* and *SQUARE*) can a participant decide what response to perform after reading only the first word. If, instead, the features are not represented conjunctively, a participant will need to process both target words on all the dual-feature trials to determine the response (see also Introduction). Therefore, this account doesn't rule out conjunctive representation as an explanation of the dual-feature benefit, but rather presupposes it.

It is important to emphasize that our interpretation of the null-effects observed during simultaneous visualization does not hinge on a comparison to a verbal control task. In Experiment 1c we observed a dual-feature benefit in a *visual* control task where targets that were identical to those in the visualization task (see Fig. 5). This suggests that the null-effects in the visualization tasks can be attributed to differences in the way features were represented due to the modality of the cue (i.e., it being verbal), instead of potential differences in identification strategy induced by the modality of the target. Lastly, although Experiment 1c

demonstrates that participants do generate a conjunctive representation of two features in a visual control task (for comparable findings see Kahneman et al., 1992; Gordon & Irwin, 1996; Saiki, 2003; Hommel, 2004), it is important to point out that this finding is not in conflict with the null-effects observed in the visualization tasks. Instead, we would interpret this as strong evidence that the representation of perceptual features during comprehension is distinct from the representation of perceptual features during perception proper.

Our results extend previous investigations on the combinatorial properties of perceptual representations elicited during language comprehension (i.e., Potter & Faulconer, 1979; Wu & Barsalou, 2009). Our findings are consistent with Potter and Faulconer (1979) in showing that participants are fast at generating a perceptual representation using a verbal cue. Both our study and theirs suggest that immediately after reading a sentence, participants are able to represent information pertaining to both an adjective and a noun within a noun phrase. Our study extends these findings to an orthogonal design to show when participants activate a general disjunctive representation vs. a specific conjunctive representation of adjective-noun combinations. One interesting point is that, in their experiments, Potter and Faulconer observed dual-feature benefits about half the size of those observed in the present study. One explanation for this may be that, unlike our study, they did not manipulate the presentation of the adjectives and nouns orthogonally across the dual-feature and single-feature conditions. Due to this, participants did not know which visual features of the target would be task-relevant and which would be task-irrelevant. This means that on half of the single-feature trials participants could initially be distracted by the adjective-related target features (e.g., the unexpected flames in the picture), whereas on the other half of trials they could directly identify the target using the noun-related target features (e.g., the expected house in the picture, see Fig. 2a), which is consistent with a dual-feature benefit about half the size of those observed in our study.

It is important to point out that we do not know whether our findings generalize to more complex (naturalistic) pictures and / or sentences. In the present studies, we aimed to test the distinction between conjunctive vs. disjunctive representations in a tightly controlled experimental design using basic and strongly contrasting visual features presented randomly in all possible combinations. Our observed effects may be difficult to test in more complex designs, while at the same time excluding alternative interpretations based on strategic effects, feature imbalance and / or counterbalancing issues[7].

In a seminal paper, Wu and Barsalou (2009) presented participants with either single nouns (e.g., "lawn") or noun phrases containing an adjective and noun (e.g., "rolled-up lawn") and asked them to verbally report their properties. Using this property generation task, they observed that participants were more likely to report occluded features for the noun phrases than for the nouns (i.e., features that are not perceptually accessible such as *dirt* or *roots*), and conversely, participants were more likely to report non-occluded features for the nouns than for the noun phrases (i.e., features that are perceptually accessible such as *blades* or *green*). Given that occluded features by definition fall within the intersection of the sets of adjective and noun-related features this finding is consistent with the notion that participants in their experiments generated conjunctive representations. Interestingly, subsequent experiments have shown that participants report both verbal associates as well as perceptual properties during the property generation task, and that the verbal associates tend to be reported prior to the perceptual properties (Santos, Chaigneau, Simmons, & Barsalou, 2011). Although different from the authors' own interpretation, this finding is consistent with our proposal that (internally generated) verbal representations may be cueing the generation and elaboration of perceptual representations.

It is noteworthy that the generation of complex perceptual representations always seems to occur in a sequential manner (e.g. Craver-Lemley, Arterberry, & Reeves, 1999; Finke, Pinker, & Farah, 1989; Kosslyn, Cave, Provost, & von Gierke, 1988). Why would this be the case? Why does the generation of elaborate perceptual representations appear to be temporally constrained in the same way that language is? Although not directly tested in the present study, the fact that linguistic structure is inherently sequential, points to the possibility that sequences of internally generated linguistic cues may be driving the incremental construction of perceptual representations.

Barsalou, Santos, Simmons, & Wilson (2008) proposed the language and situated simulation (LASS) theory as a framework for integrating verbal and perceptual approaches to conceptual processing. In this account, they assume that language provides a powerful system for indexing perceptual representations, and for manipulating them during comprehension. Although most accounts of perceptual simulation posit that conceptual processing can involve interactions between verbal and perceptual representations, they commonly claim that language plays only an indirect role in the construction of perceptual representations during comprehension. For instance, Barsalou (1999) argues that the mechanisms underlying perceptual simulation are the ones that implement basic symbolic processes such as predication, conceptual combination, and recursion. Although linguistic mechanisms are not capable of implementing symbolic operations on their own within this framework, they are claimed to play a central role in controlling perceptual simulation during interpersonal communication (i.e., it is through language that individuals are able to share perceptual representations; Barsalou, Santos, Simmons, & Wilson, 2008). However, the episodic recombination of features into perceptual representations is generally posited to occur independently of language (Barsalou, 2012; Barsalou & Prinz, 1997).

Although we agree with the position that conceptual processing depends on interactions between verbal and perceptual representations, we interpret our findings as suggesting that linguistic structure may play a central role in the construction of perceptual representations during comprehension. Within this perspective, an ordered sequence of linguistic cues may be able to incrementally select (sub-)sets of perceptual features that provide a referential domain. Through this incremental process, the represented set may become increasingly specified in terms of its perceptual content.

Our proposal is that perceptual representations are constructed incrementally, guided step-by-step by linguistic sequences, and that, depending on the task at hand, the perceptual features are incrementally specified to the level of representation that would be necessary to carry out the task at hand. For example, when hearing the sentence *The ranger saw the eagle in the sky*, location constrains the shape of the object. Initially, all perceptual features are activated that are associated with the verbal cue *eagle*. Subsequently, a more specified subset is selected that is additionally associated with the verbal cue *sky* (for example, spread out wings; see Zwaan et al., 2002). In order to accommodate our findings, current accounts would need to assume that language-mediated incremental construction of perceptual representations also applies generally (not just to the case when individuals communicate information through the use of language, see Barsalou, 1999; Prinz,

[7] It is worth pointing out that in their classic study Potter and Faulconer (1979) presented more naturalistic stimuli which would be difficult to implement in a fully orthogonal design. For instance, they used a cue sentence describing a "burning house", where participants were either shown a picture of a burning house (target) or upside-down airplane (foil), whereas in the single feature trials they were shown a picture of a house or an upright airplane. For a clean comparison between dual-feature vs. single feature trials, one would need to present a set of trials referring to and depicting a "burning house", "burning airplane", "upside-down house" and "upside-down airplane", as well as single feature trials using identical pictures but referring only to "burning", upside-down", "house" and "airplane". In order to avoid spurious RT effects between dual-feature vs single feature trials it would be important to avoid (a) feature imbalanced in terms of salience / discriminability, (b) non-equivalent instantiation of features across different pictures, and (c) strategic response effects due to differences in prior knowledge.

2002; Zwaan & Madden, 2005). Within this perspective, conceptual knowledge is grounded in perception through the way we interface and interact with the world, and linguistic representations create meaning by specifying the construction of perceptual representations (see also Lupyan & Bergen, 2016). It is through the combinatorial (re-)composition of linguistic representations that we can generate new perceptual representations by specifying sequential structure in terms of syntactic constraints. When parsing a sentence, surface syntax provides the instructions that are necessary for building perceptual representations (see Langacker, 1986).

To our knowledge, all previous accounts of perceptual simulation claim that language provides humans with the ability to control each other's representations in the absence of actual referents. We agree and would like to take this claim even one step further. Language provides us with the ability to generate our own representations and provides us with the necessary structure in order to regenerate and elaborate on successful perceptual representations that were generated by us in the past. In this sense, our store of linguistic representations function as a reservoir of verbal 'recipes' which we can use to reenact perceptual representations and recombine productively in order to create novel representations.

Within this perspective, conceptual knowledge is then jointly determined by the structure within and relations between of linguistic 'recipes' as well as the constraints and interactions that result from combining different perceptual 'ingredients' to meet the situational demands.

Author note

All raw data files for the experiments reported in this study have been made publicly available at Open Science Framework https://osf.io/gbuqm/. Original materials used to conduct the research will be made available to other researchers for purposes of replicating the procedure or reproducing the results.

CRediT authorship contribution statement

Bruno R. Bocanegra: Conceptualization, Methodology, Software, Formal analysis, Writing – original draft, Writing – review & editing, Visualization. **Fenna H. Poletiek:** Conceptualization, Writing – original draft, Writing – review & editing. **Rolf A. Zwaan:** Conceptualization, Writing – original draft.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences, 22*, 577–660.

Barsalou, L. W. (2012). The human conceptual system. In M. Spivey, K. McRae, & M. F. Joanisse (Eds.), *The Cambridge handbook of psycholinguistics* (pp. 239–258). New York, NY: Cambridge University Press.

Barsalou, L. W., & Prinz, J. J. (1997). Mundane creativity in perceptual symbol systems. In T. B. Ward, S. M. Smith, & J. Vaid (Eds.), *Creative Thought: An investigation of Conceptual Structures and Processes* (pp. 267–307). Washington, DC: American Psychological Association.

Barsalou, L. W., Santos, A., Simmons, W. K. & Wilson, C. D. (2008) Language and simulation in conceptual processing. In: *Symbols, embodiment, and meaning*, ed. M. De Vega, A. M. Glenberg & A. C. Graesser, pp. 245–83. Oxford University Press.

Bocanegra, B. R., & Hommel, B. (2014). When cognitive control is not adaptive. *Psychological Science, 25*, 1249–1255.

Borghi, A. M., & Binkofski, F. (2014). *Words as social tools: An embodied view on abstract concepts (Briefs in Cognition series)*. New York, NY: Springer.

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science, 6*, 3–5.

Carruthers, P. (1996). *Language, Thought, and Consciousness*. Cambridge: Cambridge University Press.

Craver-Lemley, C., Arterberry, M. E., & Reeves, A. (1999). "Illusory" illusory conjunctions: The conjoining of features of visual and imagined stimuli. *Journal of Experimental Psychology: Human Perception and Performance, 25*(4), 1036–1049.

Damasio, A. R. (1994). *Descartes' Error: Emotion, Reason, and the Human Brain*. New York: Putnam.

Edelman, G. M. (1992). *Bright Air, Brilliant Fire: On the Matter of Mind*. New York: Basic Books.

Feldman, J. (2010). Embodied language, best-fit analysis, and formal compositionality. *Physics of Life Reviews, 7*, 385–410.

Finke, R. A., Pinker, S., & Farah, M. J. (1989). Reinterpreting visual patterns in mental imagery. *Cognitive Science, 13*(1), 51–78.

Fodor, J., & Lepore, E. (1996). The red herring and the pet fish: Why concepts still can't be prototypes. *Cognition, 58*(2), 253–270.

Forbes, K., & Klein, R. M. (1996). The magnitude of the fixation offset effect with endogenously and exogenously controlled saccades. *Journal of Cognitive Neuroscience, 8*(4), 344–352.

Gallese, V., & Lakoff, G. (2005). The Brain's Concepts: The Role of the Sensory-motor System in Conceptual Knowledge. *Cognitive Neuropsychology, 22*, 455–479.

Goldstone, R. L., & Barsalou, L. W. (1998). Reuniting perception and conception. *Cognition, 65*(2), 231–262.

Gomila, A., Travieso, D., & Lobo, L. (2012). Wherein is Human Cognition Systematic? *Minds and Machines, 22*(2), 101–115.

Gordon, R. D., & Irwin, D. E. (1996). What's in an object file? Evidence from priming studies. *Perception & Psychophysics, 58*, 1260–1277.

Hommel, B. (2004). Event files: Feature binding in and across perception and action. *Trends in Cognitive Sciences, 8*, 494–500.

Kahneman, D., Treisman, A., & Gibbs, B. J. (1992). The reviewing of object files: Object-specific integration of information. *Cognitive Psychology, 24*, 175–219.

Kosslyn, S. M., Cave, C. B., Provost, D. A., & von Gierke, S. M. (1988). Sequential processes in image generation. *Cognitive Psychology, 20*, 319–343.

Langacker, R. W. (1986). An introduction to cognitive grammar. *Cognitive Science, 10*, 1–40.

Loftus, G. R., & Masson, M. E. (1994). Using confidence intervals in within-subject designs. *Psychonomic Bulletin & Review, 1*(4), 476–490.

Lupyan, G., & Bergen, B. (2016). How Language Programs the Mind. *Topics in Cognitive Science, 8*(2), 408–424.

Lupyan, G., & Ward, E. J. (2013). Language can boost otherwise unseen objects into visual awareness. *Proceedings of the National Academy of Sciences, 110*(35), 1419-201.

McConkie, G. W., & Rayner, K. (1975). The span of the effective stimulus during a fixation in reading. *Attention, Perception, & Psychophysics, 17*(6), 578–586.

Mooney, R. J. (1995). Encouraging experimental results on learning CNF. *Machine Learning, 19*(1), 79–92.

Ostarek, M., & Huettig, F. (2017). Spoken words can make the invisible visible – Testing the involvement of low-level visual representations in spoken word processing. *Journal of Experimental Psychology: Human Perception and Performance, 43*(3), 499–508.

Paivio, A. (1986). *Mental Representations: A Dual Coding Approach*. New York: Oxford University Press.

Paivio, A. (2007). *Mind and its evolution: A dual coding theoretical approach*. Mahwah, NJ: Erlbaum.

Potter, M. C., & Faulconer, B. A. (1979). Understanding noun phrases. *Journal of Verbal Learning and Verbal Behavior, 18*, 509–521.

Prinz, J. J. (2002). *Furnishing the Mind: Concepts and their Perceptual Basis*. Boston, MA: MIT Press.

Pulvermüller, F. (2005). Brain mechanisms linking language and action. *Nature Reviews Neuroscience, 6*, 576–582.

Rips, L. J. (1995). The current status of research on concept combination. *Mind & Language, 10*(1–2), 72–104.

Saiki, J. (2003). Feature binding in object-file representations of multiple moving items. *Journal of Vision, 3*, 6–21.

Santos, A., Chaigneau, S. E., Simmons, W. K., & Barsalou, L. W. (2011). Property generation reflects word association and situated simulation. *Language and Cognition, 3*(1), 83–119.

Spelke, E. (2003). What Makes Us Smart? Core Knowledge and Natural Language. In D. Gentner, & S. Goldin-Meadow (Eds.), *Language in Mind* (pp. 277–311). Cambridge, MA: MIT Press.

Spivey, M. J., Tyler, M. J., Eberhard, K. M., & Tanenhaus, M. K. (2001). Linguistically mediated visual search. *Psychological science, 12*(4), 282–286.

Trauzettel-Klosinski, S., & Dietz, K. (2012). Standardized Assessment of Reading Performance: The New International Reading Speed Texts IReSTStandardized Assessment of Reading Performance. *Investigative Ophthalmology & Visual Science, 53*(9), 5452–5461.

Wu, L. L., & Barsalou, L. W. (2009). Perceptual simulation in conceptual combination: Evidence from property generation. *Acta Psychologica, 132*, 173–189.

Zwaan, R. A., & Madden, C. J. (2005). Embodied sentence comprehension. In D. Pecher, & R. A. Zwaan (Eds.), *Grounding cognition: The role of perception and action in memory, language, and thinking* (pp. 224–245). New York: Cambridge University Press.

Zwaan, R. A., & Pecher, D. (2012). Revisiting mental simulation in language comprehension: Six replication attempts. *PLoS ONE, 7*, Article e51382.

Zwaan, R. A., Stanfield, R. A., & Yaxley, R. H. (2002). Language comprehenders mentally represent the shapes of objects. *Psychological Science, 13*, 168–171.