Consciousness and Cognition 22 (2013) 562–571

Contents lists available at SciVerse ScienceDirect

Consciousness and Cognition

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

Language-guided visual processing affects reasoning: The role of referential and spatial anchoring

Magda L. Dumitru [a,*], Gitte H. Joergensen [b], Alice G. Cruickshank [b,c], Gerry T.M. Altmann [b]

[a] Macquarie University, Australia
[b] University of York, United Kingdom
[c] University of Bradford, United Kingdom

Text

ARTICLE INFO

Article history:
Received 16 October 2012
Available online xxxx

Keywords:
Visual-world paradigm
Coordination
Anchoring
Reasoning
Language-meaning verification process
Attention
Speech stream
Visual stream

ABSTRACT

Language is more than a source of information for accessing higher-order conceptual knowledge. Indeed, language may determine how people perceive and interpret visual stimuli. Visual processing in linguistic contexts, for instance, mirrors language processing and happens incrementally, rather than through variously-oriented fixations over a particular scene. The consequences of this atypical visual processing are yet to be determined. Here, we investigated the integration of visual and linguistic input during a reasoning task. Participants listened to sentences containing conjunctions or disjunctions (*Nancy examined an ant **and**/**or** a cloud*) and looked at visual scenes containing two pictures that either matched or mismatched the nouns. Degree of match between nouns and pictures (referential anchoring) and between their expected and actual spatial positions (spatial anchoring) affected fixations as well as judgments. We conclude that language induces incremental processing of visual scenes, which in turn becomes susceptible to reasoning errors during the language-meaning verification process.

© 2013 Elsevier Inc. All rights reserved.

1. Introduction

1.1. Visual information processing

Language and vision are the primary means for accessing higher-order conceptual knowledge. We may capture the idea that bees collect pollen, for instance, either from hearing that *Bees collect pollen* or from seeing a picture of bees collecting pollen. Naturally, conceptual knowledge is retrieved more easily when linguistic and visual cues agree than when they are contrasting or competing (e.g., Kashak & Glenberg, 2000; Richardson, Spivey, Barsalou, & McRae, 2003; Stanfield & Zwaan, 2001), although the mechanism by which cues combine is poorly understood, as there are important differences in the way linguistic and visual information is represented and processed. In particular, language information is captured incrementally as words gradually build up a sentence, whereas visual information is captured by fixations in various directions over a particular scene. Nevertheless, research using the visual-world paradigm (Allopenna, Magnuson, & Tanenhaus, 1998; Cooper, 1974; Kamide, Altmann, & Haywood, 2003; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995) has determined that visual information presented together with linguistic information is processed incrementally such that, when a spoken word refers to an object in a visual display, attention is rapidly and automatically directed toward that object. These findings support the idea that language guides visual processing and have fuelled research into how information from various sources i.e. verbal, visual and world knowledge affects language comprehension and visual recognition. However, the consequences

* Corresponding author. Address: Department of Cognitive Science, Macquarie University, Sydney, NSW 2109, Australia. Fax: +61 2 9850 6059.
E-mail address: magda.dumitru@gmail.com (M.L. Dumitru).

1053-8100/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.concog.2013.03.008

of this atypical i.e. incremental processing of visual information in linguistic contexts have never been directly investigated. In the current study, we explore whether incremental visual processing is vulnerable to unconscious cognitive biases that may trigger reasoning errors.

1.2. Consequences of incremental processing: successful anticipation

There is a growing body of empirical evidence supporting the hypothesis that language comprehension involves dynamic perceptual simulations (e.g., Barsalou, 1999; Glenberg & Kaschak, 2003; Kaschak et al., 2005; Zwaan, 2004) that can directly affect visual perception. Dils and Boroditsky (2010), for instance, have shown that people's interpretation of an ambiguous figure e.g., a hawk flying downward vs. a goose flying upward is affected by previously hearing stories that describe physical motion directed upwards or downwards. Further evidence that language guides visual processing comes from studies using the visual-world paradigm. It was shown, in these studies, that visual processing in linguistic contexts mirrors language processing and happens incrementally thus leading to several surprising effects, to which we now briefly turn.

First, people are ready to instantly fill in missing information in visual displays to make them fit accompanying verbal descriptions. As shown in Matlock (2004) and in Richardson and Matlock (2007), the eyes may scan the same drawing e.g., a road amid palm trees differently, according to whether the accompanying sentence featured a fictive-motion verb (*The road goes through the desert*) or a static verb (*The road is in the desert*) and according to whether the terrain had been previously described as being easy (*The desert is flat*) or difficult (*The desert is hilly*). Second, incremental processing may lead people to anticipate a match between words and visual stimuli and identify a relevant target before being mentioned. As shown in Altmann and Kamide (1999), participants who viewed a scene depicting a boy, a cake, and several toys were faster to fixate on the cake when hearing *The boy will **eat** the cake* rather than *The boy will **move** the cake* because the verb rapidly evoked the knowledge that humans are likely to eat something edible. Finally, people are able to look at nothing in order to confirm a match e.g., they fixate on a blank location if a relevant visual stimulus had previously occupied that region of space (Altmann, 2004; Demarais & Cohen, 1998; Johansson, Holsanova, & Holmqvist, 2006; Spivey & Geng, 2001).

Taken together, these findings suggest that anticipation, which is the landmark of incremental processing, is more likely to affect visual cognition in linguistic contexts than otherwise. But exactly what are the effects of anticipation and thereby of incremental processing? Research so far has been investigating cases where anticipation is successful i.e. cases in which visual and linguistic information match or at least are compatible with each other, as when the word 'goose' is used to describe a hawk-like figure. The role of anticipation in examples like those above is to enrich or help disambiguate visual stimuli. However, it is conceivable that incremental visual processing may not always benefit cognitive processes. Anticipation might fail in a series of mismatches between visual and linguistic information, as when the word 'goose' follows a mismatch and is used to describe a chest of drawers, for instance. In the current paper, we explore the consequences of incremental visual processing for language comprehension and basic reasoning by looking at both successful and unsuccessful cases of visual stimuli anticipation.

1.3. Anticipation failure: the referential-anchoring hypothesis

Incremental processing in the visual-world paradigm relies on a successive series of reference types (matches or mismatches) between words and visual stimuli. Reference types are bound to influence each other for better or for worse e.g., the anticipation of a match may succeed or fail. In particular, according to the anchoring hypothesis (e.g., Tversky & Kahneman, 1974), people are biased towards first creating a reference point i.e. an 'anchor' and then adjusting to this reference when evaluating subsequent information. We carried out an eye-movements study of binary expressions i.e. coordinate nouns to monitor the influence of the first reference type on the second reference type. In particular, we recorded participants' eye movements as they watched two pictures and listened to sentences featuring conjunctions (*Nancy examined an ant **and** a cloud*) and disjunctions (*Nancy examined an ant **or** a cloud*). Their task was to decide whether conjunctions and disjunctions are true or false. The anchoring hypothesis predicts that participants should sometimes fail to anticipate the second reference type when different from the anchor, that is, the first reference type. For example, when hearing *and ant or/and a cloud*, people may anticipate a match following a match (e.g., if they see an ant, they next expect to see a cloud) or a mismatch following a mismatch (e.g., if they see something other than an ant, they next expect to see something other than a cloud). As a result, they may fail to correctly identify the second visual stimulus and base their decision about the accuracy of conjunctions and disjunctions on misperceptions.

1.4. Reasoning with conjunctions and disjunctions

Let us now briefly review the requirements for correctly reasoning with conjunctions and disjunctions. In order to determine that a conjunction is true e.g., that it adequately describes a binary visual display, there must be a double match between conjuncts and visual stimuli. For example, *Nancy examined an ant and a cloud* is a true description if the visual display features both an ant and a cloud. Nevertheless, anchoring effects may lead participants to unduly validate single-match conjunctions. Likewise, disjunction is true when at least one disjunct matches a visual stimulus. For example, *Nancy examined an ant or a cloud* is true if the visual display features an ant, a cloud, or both. However, anchoring effects may lead participants to unduly invalidate single-match disjunctions.

1.5. Further consequences of incremental processing: spatial anchoring

The main characteristic of speech is its irreversibility in time as it flows uninterrupted from the past towards the future. Similarly, writing in the Latin alphabet proceeds in a unique direction, from left to right. Moreover, the structures of spoken and written language are very similar and consist of a single and continuous string of discrete elements such that, within a word, each phoneme or letter precedes and follows a single phoneme or letter. We may therefore hypothesise that incremental visual processing that is, visual scanning of a series of pictures in linguistic contexts or in other contexts that encourage incremental processing may also proceed from left to right. Specifically, we advance the spatial-anchoring hypothesis according to which incremental processing in the visual-world paradigm relies on a series of mappings of each word in the speech stream onto a particular position in the visual stream, as follows. The first-mentioned noun in the speech stream is mapped onto the leftmost picture in a series of two or more, whereas the second-mentioned noun all through the last-mentioned noun are mapped onto a position to the right of the leftmost position up to the rightmost position.

The prerequisite of mapping the speech stream, which proceeds from the past towards the future, onto the visual stream, which proceeds from left to right, is that the two streams be made compatible to each other. One way of achieving compatibility is by converting the speech stream, which relies on temporal order, to the writing stream, which relies on spatial order. Another way of achieving compatibility is through direct analogy, by taking the visual stream as the source and the speech stream as the target. Careful investigation of the mechanisms by which compatibility is achieved, however, lies beyond the scope of the present paper and remains a matter for future research. The assumption we can make at this point, however, is that the mapping of the two streams onto each other is required by processing limitations. It should be easier to keep track of two parallel-running streams when they are being linked to each other than when they are being kept separate. Specifically, we predict that participants should start scanning the two pictures from left to right and thus match the first noun in the sentence to the leftmost picture and the second noun in the sentence to the rightmost picture. Processing difficulties and even reasoning errors can be expected in cases where the order in which the pictures are mentioned mismatches the order in which the pictures are visually scanned – for example, in cases where the first noun in the sentence matches the rightmost picture in the visual scene.

1.6. Does incremental processing affect meaning retrieval?

The consequences of incremental processing of visual stimuli are unlikely to target the process of meaning retrieval that is, participants should have no difficulty accessing the meaning of conjunctions and disjunctions as coordination sentences unfold. Meaning retrieval is an essentially predictive process and therefore participants are likely to gain access to the meaning of the two nouns and to the relationship between them in coordination sentences before identifying the pictures and thus before attempting to map this information onto the visual input. The language-primacy assumption in the visual-world paradigm has been repeatedly verified in experimental studies reporting anticipation of linguistic and visual matching effects, as detailed earlier. In order to determine whether incremental visual processing indeed allows access to language meaning e.g., to the rules governing the correct use of conjunctions and disjunctions and thus whether it only affects the rule-implementation process, we varied the coordinator linking the two nouns to be either a conjunction (*and*) or a disjunction (*or*). We expect to determine through a whole set of measurements i.e. response latency, response accuracy, probability of fixating the two pictures while hearing the second noun, and first-saccade latency to the second noun, that there are differences in processing conjunction vs. disjunction trials that are orthogonal to accuracy rates obtained for same vs. different-anchor trials or for left-to-right vs. right-to-left trials.

To summarize, we expect to observe two consequences of language-induced incremental processing in our experimental study. The first consequence concerns the degree of match between visual input and the identity of the two nouns in the sentence. We are particularly interested in conditions where one of the nouns mismatches one of the pictures to determine whether measurements at the second noun are influenced by whether the first noun had been a match or a mismatch. The referential-anchoring hypothesis predicts that attention will target a particular picture while hearing the second noun only in cases where the first noun had already matched one of the two pictures in the binary visual display. For example, a (mis)match following a match should be fixated on significantly more than a (mis)match following a mismatch. This may result in lower response-accuracy rates or in longer response latencies. The second consequence concerns the degree of match between visual input and the spatial mapping of the two nouns in the sentence. We expect more fixations on the picture corresponding to the second noun when it follows a first noun matching a picture to the right of the visual display than when it follows a first noun matching a left-side picture. Based on our spatial-anchoring hypothesis, according to which pictures are mapped onto a visual stream running from left to right, we anticipate that participants should start visual scanning with the leftmost picture and expect the second picture to be displayed towards the right. We thus predict more fixations on misplaced pictures when the second noun matches a picture to the left than a picture to the right. Misplacements may also affect response-accuracy or response-latency results. We further predict a difference in saccade-latency to the second noun in conjunction vs. disjunction trials, which would indicate that participants can gain early access to the meanings of the two coordinators.

1.7. Method

1.7.1. Participants

Twenty-four volunteering university students participated in the experiment in exchange for 15 AUD. They were all native speakers of English and reported normal or corrected-to-normal vision and normal hearing.

1.7.2. Apparatus and stimuli

Eye movements were recorded by an Eye Link 1000 remote eye-tracker sampling at 500 Hz from the right eye. Visual stimuli consisted of four sets of 24 picture pairs each = 96 picture pairs selected from monochrome versions of images developed by Rossion and Pourtois (2004) based on a standardized set of line drawings of familiar objects (Snodgrass & Vanderwart, 1980), as seen in Fig. 1. Each set of visual stimuli corresponded to one of four experimental conditions. There were thus 24 different visual displays (i.e. both pictures in the display were presented only one time during the whole 96 trials) for each of the four possible combinations of matching and mismatching correspondences between words and visual stimuli, that is, the match + match (*MM*), the mismatch + match (*mM*), the match + mismatch (*Mm*) and the mismatch + mismatch (*mm*) condition. Each set of 24 picture pairs was associated to two sets of spoken stimuli presented through stereo loudspeakers, one featuring conjunctions, and the other disjunctions e.g., *Nancy examined an ant and (or) a cloud*. Further, we controlled the order of the pictures in the display such that half of the visual stimuli matched the nouns mentioned in the sentence in the left-to-right order and the other half in the right-to-left order. Nouns in spoken sentences across the four conditions were matched to each other in length (number of phonemes) and in lexical frequency estimated from log-transformed frequency counts reported in the $SUBTLEX_{US}$ corpus (Brysbaert & New, 2009). Pictures corresponding to nouns across the four conditions were also matched to each other along three dimensions i.e. image agreement, familiarity, and complexity (cf. Snodgrass & Vanderwart, 1980).

Sentences were recorded in randomized order by a male native speaker of English. Markers were placed in each speech file, allowing us to identify the onset and offset of the subject, verb, first determiner, first noun, connective, second determiner, and second noun as follows: *Nancy |examined |an|ant |or|a|cloud|*. We ensured that the duration of the second noun (e.g., *cloud*) in speech files was similar across conditions (all p values >0.05).

1.7.3. Design and procedure

The experiment consisted of six practice trials followed by 96 experimental trials for each of the two lists we constructed. Each participant saw a visual display once and heard either a conjunction or a disjunction sentence (Latin-square design). The presentation order of the stimuli was randomized in the same fashion across both lists. Participants were informed that two matches between nouns and pictures were required for the validation of conjunction trials and that a single match was sufficient for the validation of disjunction trials. They were further informed that they were free to decide whether to validate or invalidate disjunction trials featuring two matches. We thereby attempted to balance the experimental design and have four conditions in which participants may give a 'yes' answer (*MM* conjunction trials and *MM*, *Mm*, and *mM*, disjunction trials) and four conditions in which participants may give a 'no' answer (*mM*, *Mm*, and *mm* conjunction trials and *mm* disjunction trials), without imposing constraints on *MM* disjunction interpretation that might influence the output. Indeed, previous reports in literature have identified two robust groups of respondents, namely those who spontaneously validate

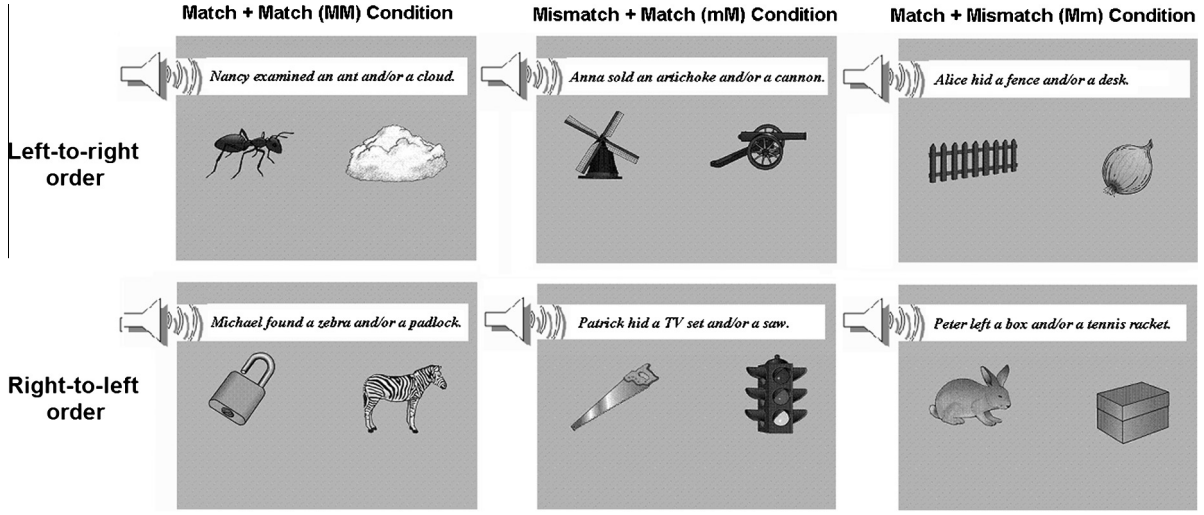

Fig. 1. Example stimuli displayed in the left-to-right order (upper panels) and in the right-to-left order (lower panels) for the *MM*, *mM*, and *Mm* conditions. Participants viewed a visual scene (e.g., containing an ant and a cloud) and heard *Nancy examined an ant and a cloud* or *Nancy examined an ant or a cloud*. Visual scenes in the *mm* condition (not shown) featured two pictures, neither of which matched the nouns mentioned in the sentence.

MM disjunction trials, and those who spontaneously invalidate them. Further comparing 'yes' and 'no' response times may appear unconventional, especially that different answers are expected for a particular condition (e.g., *mM* or *Mm*) in conjunction vs. disjunction trials. Nevertheless, trials will not be identical, as sentences accompanying a visual display containing only one match, for instance, will feature either a conjunction or a disjunction word. Moreover, participants need not only determine whether the nouns mentioned are matching any of the pictures in the visual scene, but further attend to the meaning of the coordinator linking the nouns and thus perform a reasoning task.

Participants were seated at a comfortable distance from the computer screen. At the start of the experiment, we performed a nine-point calibration and validation procedure, which was monitored and adjusted as necessary after the training phase. Participants were instructed to attend carefully to the stimuli, and decide whether each sentence could match the visual display by pressing the right button for a 'yes' answer, and the left button for a 'no' answer (counterbalanced). The structure of a trial was as follows. First, a fixation dot appeared in the centre of the screen, followed by a pair of images. After 1500 ms, a spoken sentence was presented. The objects remained onscreen until response.

2. Results and discussion

We recorded response accuracy and response latency for all trials. The distinction between the first and the second object, however, does not apply in *mm* trials, and hence we reported the probability of fixating the pictures while hearing the second noun and the first-saccade latency to the second picture after second-noun onset only for the remaining trials (*MM*, *mM*, *Mm*). We also included 'order' as a supplementary factor when analysing results for each condition under each measurement in these trials. We excluded responses for which saccade latencies were shorter than 80 ms or longer than 1500 ms (less than 1% of the trials).

2.1. Response accuracy

We coded correct responses as '1' and incorrect responses as '0'. By convention, 'yes' responses in *MM* disjunction trials were coded as 'correct', although both responses are logically valid. The left panel in Fig. 2 shows the validation patterns across trials. We carried out a 2 (connective: *and* vs. *or*) × 4 (condition: *MM* vs. *mM* vs. *Mm* vs. *mm*) ANOVA and found a main effect of condition, $F(3, 21) = 5.43$, $p = .006$, $\eta_P^2 = .437$, showing that participants validated significantly more trials in the *mm* condition than in any of the remaining conditions (*MM*, *mM*, and *Mm*). This finding is particularly important and supports the idea that the *mm* condition is a suitable criterion by which to measure performance in the remaining conditions. We also found an effect of connective, $F(1, 23) = 3.59$, $p = .007$, $\eta_P^2 = .278$, further qualified by an interaction between condition and connective, $F(3, 21) = 3.59$, $p = .031$, $\eta_P^2 = .339$. In particular, there were significantly higher rates of accuracy in conjunction trials than in disjunction trials in the *MM* condition, $F(1, 23) = 4.97$, $p = .036$, $\eta_P^2 = .178$, which is not surprising, considering that some participants interpreted disjunction exclusively and hence invalidated *MM* disjunction trials.

We further carried out analyses for each of the four conditions, including 'order' as a factor. A 2 (connective: *and* vs. *or*) × 2 (order: *left-to-right* vs. *right-to-left*) ANOVA over *MM* trials showed no effect of order ($p = .664$) and no interaction between factors ($p = .185$). Similar results obtained for *Mm* trials; there was no effect of order ($p = .660$) and no interaction between factors ($p = .403$). The 2 × 2 ANOVA over *mM* trials, however, revealed a main effect of order, $F(1, 23) = 5.86$, $p = .024$, $\eta_P^2 = .203$, with higher validation rates of left-to-right trials compared to right-to-left trials. We may infer that the *mM* condition is the most susceptible to incrementality effects, both anticipation-related and mapping-related.

2.2. Response latency

As seen in the right panel of Fig. 2, the response latency data agree with the response accuracy data. The 2 (connective: *and* vs. *or*) × 4 (condition: *MM* vs. *mM* vs. *Mm* vs. *mm*) ANOVA revealed a main effect of condition, $F(3, 21) = 4.02$, $p = .011$,

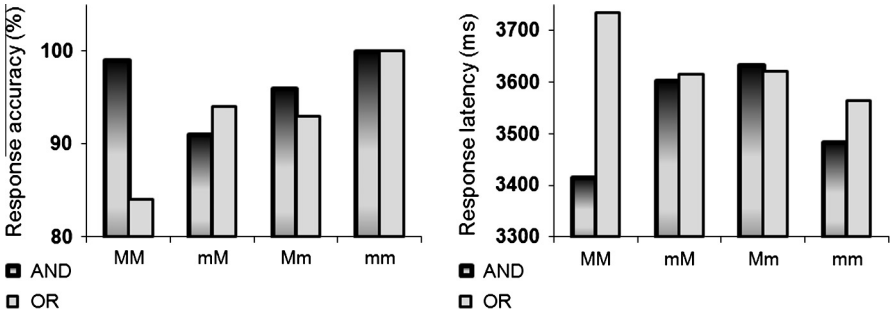

Fig. 2. The left panel shows the percentage of correct answers as a function of connective type (*and*, *or*) and condition (*MM*, *Mm*, *mM*, *mm*). By convention, we coded 'yes' responses as 'correct' and 'no' responses as 'incorrect' in the *MM* disjunction condition. The right panel shows response latencies across conditions (ms). Both panels indicate show that performance in mixed conditions (*mM* and *Mm* trials) is impaired compared to performance in the *mm* condition, which is the fastest and most accurate.

$\eta_P^2 = .149$, with faster responses in the *mm* condition compared to either the *mM* or the *Mm* condition. We also found an effect of connective, $F(1, 23) = 24.28$, $p < .001$, $\eta_P^2 = .514$, with faster responses in conjunction trials than in disjunction trials. The effect was further qualified by an interaction between condition and connective, $F(3, 21) = 11.21$, $p < .001$, $\eta_P^2 = .616$, such that processing was significantly faster in conjunction trials than in disjunction trials in the *MM* condition, $F(1, 23) = 45.22$, $p < .001$, $\eta_P^2 = .663$, and in the *mm* condition, $F(1, 23) = 11.72$, $p = .002$, $\eta_P^2 = .338$. The findings suggest that participants were trying to decide on whether to accept an inclusive interpretation of disjunction or not, and hence whether to validate or invalidate disjunction trials where both pictures mentioned in the spoken sentences were present.

A 2 (connective: *and* vs. *or*) × 2 (order: *left-to-right* vs. *right-to-left*) ANOVA over *MM* trials revealed an interaction between factors, $F(1, 23) = 19.49$, $p < .001$, $\eta_P^2 = .459$, such that faster responses obtained for left-to-right than for right-to-left conjunction trials ($F(1, 23) = 8.20$, $p = .009$, $\eta_P^2 = .263$) and for right-to-left than for left-to-right disjunction trials ($F(1, 23) = 11.68$, $p < .002$, $\eta_P^2 = .337$). These results seem surprising but can be readily explained if we considered the possibility that some participants may have stopped processing the second matching picture and decided that sufficient evidence was available (i.e., the first match) for them to respond in disjunction trials. This explanation is further supported by response-accuracy data showing that a significant proportion of participants invalidated *MM* trials. The 2 × 2 ANOVA over *mM* trials showed no effect of order ($p = .479$) and no interaction between factors ($p = .267$). Similarly, the 2 × 2 ANOVA over *Mm* trials showed no effect of order ($p = .977$) and no interaction ($p = .990$).

2.3. Probability of fixating the pictures while hearing the second noun

The upper panels in Fig. 3 show the probability of fixating the pictures while hearing the second noun. We compared the probability of fixating the first and second picture in a 2 (connective: *and* vs. *or*) × 3 (condition: *MM* vs. *mM* vs. *Mm*) × 2 (picture: first vs. second) ANOVA. We observed a main effect of condition, $F(2, 22) = 7.24$, $p = .004$, $\eta_P^2 = .397$, with significantly more fixations in the *mM* condition than in either the *MM* or the *Mm* conditions. Overall, the second picture was fixated on more than the first picture, $F(1, 23) = 222.39$, $p < .001$, $\eta_P^2 = .908$, which would seem to confirm previous findings in literature that the eyes target the object that is being mentioned. If this explanation may hold for results during the second noun in *MM* and *mM* trials, it must be abandoned once we consider the results obtained during the second noun in *Mm* trials. We found that the eyes fixated on the 'other' picture even when it mismatched the second noun. We interpret these results as evidence in support of the anchoring hypothesis: Participants anticipated another match to follow the first match in *Mm* trials, as they did in *MM* trials, and hence they were prepared to fixate on the matching picture. The effect of picture was further qualified by an interaction between factors, $F(2, 22) = 29.18$, $p < .001$, $\eta_P^2 = .726$. Planned comparisons showed that there were more fixations on the second picture in *MM* trials than in *mM* trials or in *Mm* trials and in *Mm* than in *mM* trials. Con-

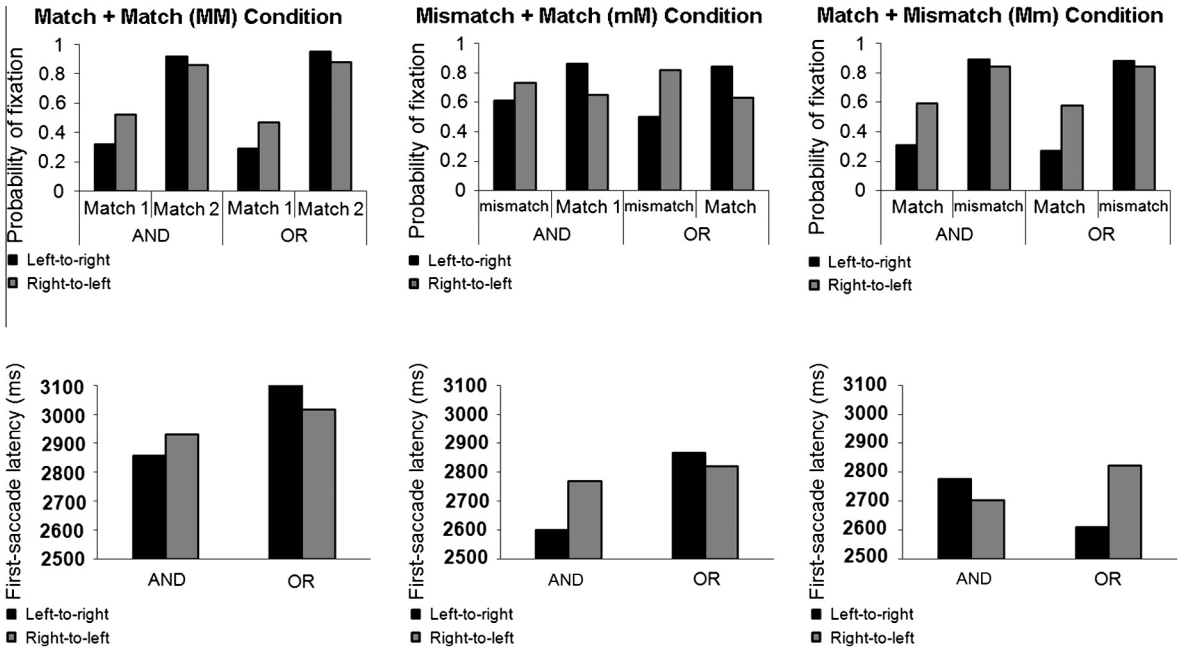

Fig. 3. The upper panels show the probability of fixating the two pictures while hearing the second noun. Note that the second picture is fixated on more than the first picture in *MM* and *Mm* trials, but not in *mM* trials, as predicted by the anchoring hypothesis. The lower panels show the time, from second-noun onset, of launching a first saccade to the second picture. Examples illustrate the *MM*, *mM*, and *Mm* conditions; the notions of (mis)matching pictures and spatial order of the pictures do not apply in the *mm* condition.

versely, there were more fixations on the first picture while hearing the second noun in *mM* trials than in *MM* trials and in *mM* than in *Mm* trials. These findings further support the anchoring hypothesis.

A 2 (connective: *and* vs. *or*) × 2 (picture: first vs. second) × 2 (order: *left-to-right* vs. *right-to-left*) ANOVA over *MM* trials revealed a main effect of order, $F(1, 23) = 6.18$, $p = .021$, $\eta_P^2 = .212$, with more fixations in right-to-left than in left-to-right trials while hearing the second noun. We also observed an interaction between order and picture, $F(1, 23) = 7.24$, $p = .013$, $\eta_P^2 = .240$, with more fixations on the first picture in right-to-left than in left-to-right trials while hearing the second noun, $F(1, 23) = 8.01$, $p = .009$, $\eta_P^2 = .258$; order did not influence significantly fixations on the second picture ($p = .080$). These results support the spatial-order hypothesis according to which participants expect matching objects to be placed in the left-to-right order in a visual scene. Participants would tend to move their eyes towards the right when matching the second noun to the second picture. Their fixating on the first picture placed towards the right while hearing the second noun suggests that their expectations for the second matching object to be placed in that position have not been met.

The 2 × 2 × 2 ANOVA over *mM* trials revealed an interaction between picture and order, $F(1, 23) = 20.85$, $p < .001$, $\eta_P^2 = .476$, as follows. In left-to-right trials (that is, when the matching picture was placed to the right), the second (matching) picture was fixated on more than the first (mismatching) picture, whereas in right-to-left trials, the first picture was fixated on more than the second while hearing the second noun. The results suggest that, whereas participants had no difficulty in fixating on the matching picture when placed in the anticipated position (to the right in left-to-right trials), they failed to fixate on the matching picture when misplaced (to the left in right-to-left trials) and fixated on the mismatching picture instead because it was placed in the anticipated position for verifying the second reference type, namely to the right.

As for *Mm* trials, the 2 × 2 × 2 ANOVA revealed a main effect of order, $F(1, 23) = 36.73$, $p < .001$, $\eta_P^2 = .615$, with more fixations in right-to-left trials than in left-to-right trials. There was also an interaction between order and picture, $F(1, 23) = 18.71$, $p < .001$, $\eta_P^2 = .449$, such that there were more fixations on the first picture in right-to-left trials than in left-to-right trials, but there was no difference between the two spatial orders for the second picture ($p = .158$). The results can be accounted for by the spatial-order hypothesis as reflecting participants' failed expectations to start visual scanning at the left and hence find the first (matching) picture at the left side of the visual display.

2.4. First-saccade latency to the second picture following second-noun onset

The lower panels in Fig. 3 show the latencies of first saccades to the second picture following the onset of the second noun. A three (condition: *MM* vs. *mM* vs. *Mm*) × 2 (connective: *and* vs. *or*) revealed a main effect of condition, $F(2, 22) = 10.48$, $p = .001$, $\eta_P^2 = .583$, with first saccades to the second object launched later in the *MM* condition than in either the *mM* or the *Mm* conditions, and in the *Mm* than in the *mM* condition. We may attribute these effects to participants' difficulty with constructing overall sentence meaning by finding relevant links between matching nouns. Discarding the meaning of mismatching nouns appears to be an effortless process – a result confirmed by optimal accuracy ratings and overall latency results obtained for *mm* trials, as detailed above. We also found an effect of connective, $F(1, 23) = 20.168$, $p < .001$, $\eta_P^2 = .558$, with saccades launched later in disjunction trials than in conjunction trials. These results are unsurprising, considering that disjunction is arguably more complex a concept than conjunction is. They also agree with our prediction that participants would be able to gain early access to the meaning of the two coordinators (i.e. of conjunction- vs. disjunction words) and process the input accordingly.

A two (connective: *and* vs. *or*) × 2 (order: *left-to-right* vs. *right-to-left*) ANOVA over *MM* trials showed no effect of order ($p = .931$) and no interaction between factors ($p = .254$). The same analysis over *mM* trials revealed a marginally significant interaction between connective and order, $F(1, 23) = 3.90$, $p = .061$, $\eta_P^2 = .157$, with first saccades to the second picture launched later in right-to-left than in left-to-right conjunction trials, $F(1, 23) = 4.39$, $p = .048$, $\eta_P^2 = .173$; there was no significant difference in disjunction trials ($p = .535$). The effect may be explained based on spatial order as well as on the assumption that participants are able to gain early access to coordinator meaning, as follows. Since, after identifying a mismatch, participants anticipated conjunction to be invalid, they were less motivated to rapidly find a match and thus shift their gaze towards the second picture by fighting their way against the left-to-right visual stream. In contrast, the effort was well warranted in disjunction trials where one match is sufficient for validating the trials. The 2 × 2 ANOVA over *Mm* trials only revealed a marginally significant interaction between connective and order, $F(1, 23) = 3.97$, $p = .069$, $\eta_P^2 = .249$, such that saccades to the second picture were launched later in right-to left than in left-to-right disjunction trials, $F(1, 23) = 5.88$, $p = .032$, $\eta_P^2 = .329$; there was no significant difference between the two orders in conjunction trials ($p = .534$). As for the previous condition, we may assume that participants were satisfied with a single match in disjunction trials and hence were less motivated to go against the visual stream in launching saccades to the second picture to find another match. In contrast, the effort was absolutely necessary in conjunction trials, as validation was pending.

3. Sentence-plausibility study

We carried out a calibration study to control for the possibility that sentence plausibility might have affected the results. Thirty volunteering university students participated in the experiment, in return for course credit. They were presented with the same conjunction and disjunction sentences used in the eye-tracking study in randomized order. We asked participants

to judge the likelihood of the action involving the two objects mentioned in each sentence using a rating scale from 1 to 7 (1 being "very unlikely" and 7 being "very likely). Answers from 7 participants were discarded, as they reported not having grown up in an English-speaking region. For conjunction trials, the mean ratings were 4.49 (SD = 1.10) in the Mm condition, 4.29 (SD = 1.31) in the mM condition, 4.06 (SD = 1.21) in the MM condition, and 4.06 (SD = 1.27) in the mm condition. For disjunction trials, the mean ratings were 4.56 (SD = 1.07) in the Mm condition, 4.37 (SD = 1.20) in the mM condition, 4.14 (SD = 1.11) in the MM condition, and 4.25 (SD = 1.01) in the mm condition. We carried out a regression analysis of the results across items from the calibration study together with the results across items from the eye-tracking study. Table 1 presents a summary of the main statistics. We were unable to analyse response accuracy results for mm disjunction trials because responses were constant (at ceiling). We can observe that none of the remaining p-values reached significance, suggesting that sentence plausibility did not affect the effects reported in our eye-tracking study (i.e., response accuracy, response latency, probability of fixating the pictures, and first-saccade latency to the second picture).

4. General discussion

In the current study, we showed that language induces incremental visual processing, which in turn triggers reasoning errors during the language-meaning verification process. We identified two consequences of incremental processing, both of which are amenable to an explanation in terms of anchoring effects (cf. Tversky & Kahneman, 1974).

The first consequence relates to anticipating similar reference types, thus anticipating matches between linguistic and visual stimuli to follow matches and mismatches between these stimuli to follow mismatches. We found that participants anchored their expectations relating to upcoming reference types in the most recent reference type. Indeed, there were lower rates of response accuracy and higher rates of response latency in mixed trials (*mM* and *Mm*) compared to simple trials (*mm*), effects that we attribute not to a lack of knowledge or proper access to relevant language/reasoning rules, but to an unconscious anticipation failure of the evidence available that is, the visual stimuli needed to verify the rules. Further effects in mixed trials will be discussed directly.

The second consequence of incremental processing of visual stimuli relates to anticipating a particular mapping of the speech stream onto the visual stream. Specifically, the first of two nouns in a coordination sentence should correspond to a left-side picture in a visual scene and the second noun to a right-side picture. We found that participants anchored each correspondence type between linguistic and visual stimuli to a particular position in the visual scene. Spatial-anchoring effects targeted all conditions containing at least one match between linguistic and visual stimuli and hence mixed trials (*mM* and *Mm*) as well as *MM* trials. Indeed, we found lower rates of response accuracy in right-to-left than in left-to-right *mM* trials and greater response latencies in right-to-left than in left-to-right *MM* trials.

The visual-world paradigm is particularly suited to testing the consequences of incremental visual processing, as it allowed us to observe a link between lower-accuracy responses in mixed conditions and high fixation probabilities for

Table 1
Regression-analysis results using sentence-plausibility ratings as predictors of performance across conditions. Non-significant *p*-values indicate that sentence plausibility did not affect performance.

	R^2		$F(1, 22)$		*p*-Value	
	AND	OR	AND	OR	AND	OR
Response accuracy						
mm	.004	–	.083	–	.776	–
MM	.032	.020	.732	.442	.401	.513
mM	.066	.063	1.54	.089	.226	.769
Mm	.006	.005	.124	.106	.728	.748
Response latency						
mm	.004	.008	.086	.173	.772	.681
MM	.001	.026	.023	.577	.881	.455
mM	.005	.215	.117	1.06	.735	.313
Mm	.009	.000	.193	.006	.664	.940
First-saccade latency (object 2)						
MM	.074	.072	1.68	1.71	.208	.204
mM	.020	.113	.450	2.79	.509	.109
Mm	.000	.000	.002	.007	.967	.935
Fixation probability (object 1)						
MM	.002	.000	.034	.002	.855	.964
mM	.043	.033	.979	.752	.333	.395
Mm	.047	.028	1.08	.638	.308	.433
Fixation probability (object 2)						
MM	.057	.002	1.32	.047	.263	.831
mM	.000	.150	.005	3.86	.943	.062
Mm	.159	.008	4.16	.170	.053	.684

the mismatching stimulus along with low fixation probabilities for the matching stimulus in mixed trials. These patterns go against previous findings in literature that the eyes should fixate on matching stimuli more than on mismatching stimuli, but are in accord with the referential-anchoring hypothesis that participants should not fixate on a picture (whether a match or a mismatch) unless it follows a match. As for the match between pictures and spatial order, the visual-world paradigm allowed us to clearly establish that visual scanning is oriented from left to right. Spatial anchoring resulted in more fixations on pictures in the right-to-left order than in the left-to-right order both in *MM* as well as in *mM* and *Mm* conditions.

Importantly, we found that incremental visual processing spared meaning retrieval, allowing us to conclude that anchoring affects the sentence-meaning verification process. In particular, saccades to the second picture were launched later in disjunction trials than in conjunction trials, especially in the *MM* condition. The results mirror the response-latency data i.e. participants took longer to process disjunction trials than conjunction trials and it is interesting to observe that participants were able to make the distinction early on i.e. before second-noun offset. We may hypothesise that people's representation of conjunctions and disjunctions is based on their experience with coordination situations. For example, because people are often prompted to select both items mentioned when hearing a conjunction (e.g., *Have a cup of coffee and a biscuit!*), they are likely to represent conjunction as a single cognitive object (cf. Link, 1983). Likewise, because people are often prompted to choose between alternatives when hearing a disjunction (cf. Braine & Rumain, 1981; Fillenbaum, 1974; Simons, 2001) they are likely to represent disjunction (e.g., *Have a biscuit or a fruit!*) as two cognitive objects. Further, if we assume an analogy with findings from the visual-processing literature that there is no cost involved in shifting attention between object parts but that there is a cost in shifting attention between objects (Egly, Driver, & Rafal, 1994; Lamy & Egeth, 2002), we may see why the eyes switch easier that is, earlier from the first to the second picture in conjunction trials than in disjunction trials. Our assumption is warranted by previous findings by Matlock (2004) and by Richardson and Matlock (2007) among others, that eye movements are modulated by the mental representations which are currently active, thus triggered by language descriptions.

Our findings further shed light on results that have been occasionally reported in previous studies on coordination, especially in the developmental literature, but have received only tentative explanations. Paris (1973) as well as Braine and Rumain (1981), for instance, investigated disjunction sentences and found very low rates of correct responses to *mM* and *Mm* trials in children and adults. More recently, in a statement-picture verification task with adults, Chevallier et al. (2008) reported a moderate drop in correct percentages for mixed conditions (*Mm* and *mM*) compared to *mm* conditions by at most 9% for conjunction- and 10% for disjunction trials. These results are comparable with our own - a drop of up to 9% in conjunction trials and of 8% in disjunction trials. However, in a study with typically-developing and autistic children, Chevallier, Wilson, Happé, and Noveck (2010) reported a dramatic drop in 'correct' responses (up to 58%) for disjunction trials but only a very modest drop (1%) for conjunction trials. We therefore suggest that there are two mechanisms at work when reasoning with coordinators, of which one concerns the meaning-assignment process that is, disjunction interpretation, and the other the meaning-verification process that is, anchoring. We further believe that task details are important for triggering rather one or both mechanisms. Recall that we only found a significant difference in response accuracy between the *MM* and either the *Mm* or the *mM* condition for conjunction trials. However, we maintained that anchoring affected all trials because accuracy rates in the *MM* condition were significantly lower for disjunction trials compared to conjunction trials. The difference may be due to interpretation factors independent of cognitive-biasing effects.

Finally, we would like to consider the larger implications of our main finding that incremental processing is a source of error for domains other than language. We may hypothesise that speech and spelling errors are easily detectable because language offers a straightforward categorization mechanism. For instance, a variety of cars of various brands, colours, and dimensions can be simply labelled as 'car', whereas the visual system needs to pack a lot of category-related information to conceptualize a percept. Although visual perception happens much faster than language processing, the cognitive load associated with visual cognition is certainly greater. Even simple cognitive operations over visual percepts e.g., sequence formation or comparisons require some form of categorization or tagging, making the visual system particularly susceptible to errors. This may be the reason why language has long been the preferred medium of expressing thought-related processes.

5. Conclusions

Language-guided visual processing happens incrementally and therefore is susceptible both to anticipation success and to anticipation failure. The present study is the first to directly investigate the effects of anticipation on language comprehension and basic reasoning. We determined that, whereas people could easily access language meaning (of conjunctions vs. disjunctions) when presented with both visual- and linguistic information, their success rates with verifying this meaning cannot be optimal unless information types match, or at least are compatible with each other. The upshot of our finding is that anticipation itself is a form of cognitive bias that, like other unconscious phenomena, benefits information processing when it operates with extremely stable labels.

References

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

Altmann, G. M. T. (2004). Language-mediated eye movements in the absence of a visual world: The blank screen paradigm. *Cognition, 93*, 79–87.

Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition, 73*, 247–264.
Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences, 22*, 577–660.
Braine, M., & Rumain, B. (1981). Development of comprehension of 'Or': Evidence for a sequence of competencies. *Journal of Experimental Child Psychology, 31*(1), 46–70.
Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved frequency measure for American English. *Behavior Research Methods, 41*(4), 977–990.
Chevallier, C., Noveck, I. A., Nazir, T., Bott, L., Lanzetti, V., & Sperber, D. (2008). Making disjunctions exclusive. *The Quarterly Journal of Experimental Psychology, 61*(11), 1741–1760.
Chevallier, C., Wilson, D., Happé, F., & Noveck, I. (2010). Scalar inferences in autism spectrum disorders. *Journal of Autism and Developmental Disorders*. http://dx.doi.org/10.1007/s10803-010-0960-8.
Cooper, R. M. (1974). The control of eye fixation by the meaning of spoken language. A new methodology for the real-time investigation of speech perception, memory, and language processing. *Cognitive Psychology, 6*(1), 84–107.
Demarais, A., & Cohen, B. H. (1998). Evidence for image-scanning eye movements during transitive inference. *Biological Psychology, 49*, 229–247.
Dils, A. T., & Boroditsky, L. (2010). Processing unrelated language can change what you see. *Psychonomic Bulletin and Review, 17*(6), 882–888.
Egly, R., Driver, J., & Rafal, R. D. (1994). Shifting visual attention between objects and locations: Evidence from normal and parietal lesion subjects. *Journal of Experimental Psychology: General, 123*, 161–176.
Fillenbaum, S. (1974). Or: Some uses. *Journal of Experimental Psychology, 103*(5), 913–921.
Glenberg, A. M., & Kaschak, M. P. (2003). The body's contribution to language. In B. H. Ross (Ed.), *The Psychology of Learning and Motivation. Advances in Research and Theory* (Vol. 43, pp 93–126). New York: Academic Press.
Johansson, R., Holsanova, J., & Holmqvist, K. (2006). Pictures and spoken descriptions elicit similar eye movements during mental imagery, both in light and in complete darkness. *Cognitive Science, 30*(6), 1053–1080.
Kamide, Y., Altmann, G. M. T., & Haywood, S. L. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language, 49*, 133–159.
Kashak, M. P., & Glenberg, A. M. (2000). Constructing meaning: The role of affordances and grammatical constructions in sentence comprehension. *Journal of Memory & Language, 43*, 508–529.
Kaschak, M. P., Madden, C. J., Therriault, D. J., Yaxley, R. H., Aveyard, M., Blanchard, A. A., & Zwaan, R. A. (2005). Perception of motion affects language processing. *Cognition, 94*, B79–B89.
Lamy, D., & Egeth, H. (2002). Object-based selection: The role of attentional shifts. *Perception and Psychophysics, 64*, 52–66.
Link, G. (1983). The logical analysis of plurals and mass terms: A lattice-theoretical approach. In R. Baeuerle et al. (Eds.), *Meaning, use, and interpretation of language* (pp. 302–323). Berlin: de Gruyter.
Matlock, T. (2004). Fictive motion as cognitive simulation. *Memory and Cognition, 32*, 1389–1400.
Paris, S. G. (1973). Comprehension of language connectives and propositional logical relationships. *Journal of Experimental Child Psychology, 16*(2), 278–291.
Richardson, D., & Matlock, T. (2007). The integration of figurative language and static depictions: An eye movement study of fictive motion. *Cognition, 102*, 129–138.
Richardson, D. C., Spivey, M. J., Barsalou, L. W., & McRae, K. (2003). Spatial representations activated during real-time comprehension of verbs. *Cognitive Science, 27*, 767–780.
Rossion, B., & Pourtois, G. (2004). Revisiting Snodgrass and Vanderwart's object pictorial set: The role of surface detail in basic-level object recognition. *Perception, 33*(2), 217–236.
Simons, M. (2001). Disjunction and alternativeness. *Linguistics and Philosophy, 24*, 597–619.
Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory, 6*(2), 174–215.
Spivey, M. J., & Geng, J. J. (2001). Oculomotor mechanisms activated by imagery and memory: Eye movements to absent object. *Psychological Research, 65*, 235–241.
Stanfield, R. A., & Zwaan, R. A. (2001). The effect of implied orientation derived from verbal context on picture recognition. *Psychological Science, 12*(2), 153–156.
Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science, 268*, 1632–1634.
Tversky, A., & Kahneman, D. (1974). Judgment under uncertainly: Heuristics and biases. *Science, 185*, 1124–1130.
Zwaan, R. A. (2004). The immersed experience. Toward an embodied theory of language comprehension. In B. H. Ross (Ed.), *The psychology of learning and motivation* (vol. 44, pp. 35–62). New York: Academic Press.