Cognition 127 (2013) 361–374

Contents lists available at SciVerse ScienceDirect

Cognition

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

The case of the missing pronouns: Does mentally simulated perspective play a functional role in the comprehension of person?

Manami Sato [a,*], Benjamin K. Bergen [b]

[a] Graduate School of Education, Hiroshima University, Japan
[b] Department of Cognitive Science, University of California, San Diego, United States

ARTICLE INFO

Article history:
Received 3 September 2010
Revised 6 February 2013
Accepted 7 February 2013

Keywords:
Sentence comprehension
Perspective
Mental simulation
Personal pronouns
Human experimentation
Japanese

ABSTRACT

Language comprehenders can mentally simulate perceptual and motor features of scenes they hear or read about (Barsalou, 1999; Glenberg & Kaschak, 2002; Zwaan, Stanfield, & Yaxley, 2002). Recent research shows that these simulations adopt a particular perspective (Borghi, Glenberg, & Kaschak, 2004; Brunyé, Ditman, Mahoney, Augustyn, & Taylor, 2009). Moreover, features of utterances influence the perspective that comprehenders are led to adopt. For instance, language about *you* primes a participant visual perspective, while third person *he* and *she* prime an observer perspective. But what role does perspectival mental simulation play in the comprehension of person? On the one hand, the different perspectives adopted during language understanding could be necessary for successfully determining the meaning of an utterance. However, current empirical evidence is also compatible with the possibility that adopting a perspective in mental simulation is not essential to comprehending who did what to whom. If the latter is the case, then we should be able to find cases where language comprehenders understand who did what to whom without measurably performing mental simulation from a particular perspective. A candidate language that might display such a case is Japanese, where grammatical subject pronouns can be omitted when the subject is inferable from context. We replicated a previously used method for assessing perspectival mental simulation during language comprehension, but tailored it to Japanese. The results showed that when pronouns were present, like in English, sentences facilitated identification of an image matching the proposed perspective associated with the mentioned pronoun. This replicated the previous finding for English. But when the subject pronoun was omitted, so that the sentence did not explicitly mention the subject, there was no such effect. Nonetheless, native comprehenders of Japanese automatically and easily tracked who the subjects of the sentences with omitted subjects were. Together, these findings suggest that while grammatical person modulates visual perspective in mental simulation, visual perspective is not necessary for successful identification and representation of event participants.

1. Introduction

Nearly every sentence in English (and many other languages) specifies a subject—the entity, often a person, primarily involved in the described event. The subject can be any entity, but languages encode specific classes of subjects in their grammars. One key dimension along which grammars distinguish subjects is their *person*: first (*I*, *we*), second (*you*, *y'all*), or third (*it*, *he*, *she*, *they*). The person of the subject is among the more critical components of the meaning of a sentence. The facts described by *He's sup-*

* Corresponding author.
E-mail addresses: msato@hiroshima-u.ac.jp, manamisato@gmail.com (M. Sato).

0010-0277/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.cognition.2013.02.004

posed to do the dishes tonight and *You're supposed to do the dishes tonight* are quite different, as are the inferences one can reasonably make on the basis of them, and the actions that would be appropriate to take. But relatively little is known about how comprehenders understand who the subject is, and what consequences that has for the mental representations they construct to understand the meaning of the sentence or the inferences they draw. What is different about the processes one engages in when reading sentences about things *you* are doing versus sentences about things *he* or *she* is doing?

One possibility, suggested in the literature, is that language about *you* and language about *he* or *she* might lead to representations of the described events that adopt different perspectives (Bergen & Chang, 2005; MacWhinney, 2005). Several lines of recent work have provided evidence supporting this view. For instance, Brunyé, Ditman, Mahoney, Augustyn, and Taylor (2009) presented English speakers with sentences that differed only in whether the subject was *I*, *you*, or *he*. After each sentence, participants saw an image that depicted the event either from the perspective of a person performing the action or someone observing it. For instance, after a sentence like *You are slicing the tomato* or *She is slicing the tomato*, participants might see an image of tomato-slicing from the perspective of the slicer, or from a perspective of someone observing the slicer. They found that people were faster to respond to images when the depicted perspective matched the perspective of the sentence; that is, people were faster to respond to a picture from the slicer's perspective after a sentence about *you* slicing than after a sentence about *he* slicing, and vice versa for a picture from an observer's perspective. This finding is compatible with other results, suggesting that comprehenders have an easier time accessing mentioned objects when they would be closer to *you* in the scene (Borghi et al., 2004).

These results suggest that people tend to adopt perspectives appropriate to the subject of the sentence. But does this perspective adoption play a role in comprehension? The strongest position arises from the large body of recent research exploring the hypothesis that comprehension involves the construction of mental simulations—i.e., mental (re)creations of real-world perceptual and motor experiences acquired through daily interactions (Barsalou, 1999). On this view, people engage their motor or perceptual systems to internally simulate described actions or percepts, and this process allows them to update their beliefs, make inferences, and respond appropriately, in much the same way that real perception or motor control experiences would. A number of experimental studies have revealed that in the process of comprehending language, comprehenders unconsciously simulate various implied perceptual and motor details, including object shape (Stanfield & Zwaan, 2001), orientation (Zwaan, Stanfield, & Yaxley, 2002), color (Connell, 2007; Connell & Lynott, 2007), visibility (Yaxley & Zwaan, 2007), and direction of motion (Bergen, Lindsay, Matlock, & Narayanan, 2007; Zwaan, Yaxley, & Aveyard, 2004), as well as direction of hand motion (Glenberg & Kaschak, 2002), detailed hand shape (Wheeler & Bergen, 2010), number of hands involved (Setti, Borghi, & Tessari, 2009), manual affordances elicited by objects (Glenberg & Kaschak, 2002; Tucker & Ellis, 1998, 2004), and even higher level aspects of events, including what part of the event to focus on (Bergen & Wheeler, 2010; Madden & Zwaan, 2003).

One very typical interpretation of the mental-simulation-based view is that the different mental simulations that comprehenders construct while understanding sentences about *you* or about *(s)he* contribute to the differences in meaning that they extract from these sentences. If mental simulation is somehow constitutive of the meaning one extracts from an utterance, this could explain how representations are different for the two types of sentence. It could also explain how different inferences are generated (since the different simulated experiences would lead to different inferences), and how different responses would be generated (since the different representations would lead to different appropriate actions).

This account is appealing in its parsimony and explanatory potential. However, current evidence does not allow us to conclude that perspective in simulation is responsible for understanding who events are about. The existing evidence that people adopt pronoun-induced perspectives in sentence comprehension (similar to a good deal of work on mental simulation during language processing more broadly) shows facilitation of image processing when an image adopts the perspective suggested by a sentence. But this does not entail that differences in simulated perspective, induced by different subject pronouns, play a functional role in comprehension of person.

Instead, it's possible that taking a particular perspective is not part of understanding who the event participants are or to make appropriate inferences. Mental simulations do tend to take perspectives suggested by explicitly mentioned persons, such as those referred to by subject pronouns *you* or *she*, as shown by previous studies, but this perspective may be irrelevant to comprehension of who did what to whom. It could be that the processes by which comprehenders understand who the event participants are and what roles they play are distinct from the simulation of an event from a particular, linguistically indicated perspective.

These two positions offer substantively different accounts of how understanding works. The first claims that mental simulation—and perspective adopted therein in particular—is critical to comprehension of who did what to whom. The second proposes that other comprehension processes, perhaps algorithmic, symbolic ones, are responsible for this component of comprehension, and that perspective in mental simulation is epiphenomenal. These two positions are specific instantiations of opposing views in the more general debate currently underway in cognitive science, addressing the role of embodied or modal knowledge and processes in language and other higher cognitive capacities (Bergen, 2012; Chatterjee, 2010; Dale, Dietrich, & Chemero, 2009; Dove, 2009; Mahon & Caramazza, 2008).

Fortunately, we can extract different, testable predictions from these two positions. On the one hand, if simulation from a given perspective is a necessary component of person comprehension, then whenever people successfully understand who did what to whom, we should also ob-

serve measurable perspective effects in their mental simulations. On the other hand, if perspectival simulation is superfluous, then we should find cases where comprehenders systematically understand who did what to whom without performing mental simulations from measurably different perspectives for different persons.

At first blush, it might seem hard to find cases in which different linguistic person does not drive differences in simulated perspective, since the current literature reports such perspective differences with a variety of language stimuli in different tasks. But the world's languages display remarkable diversity in the ways that they express person, and even in whether or not they express it at all. Some languages, so-called *pro-drop* languages, allow subjects and other arguments to be omitted if they are inferable from context. It could be that people using a pro-drop language—when there is no explicit mention of the subject in some sentences—understand who did what to whom, but do so without performing mental simulation from the perspective of the subject. If this is the case, then it would serve as evidence against the strong version of the simulationist view that perspective in simulation is responsible for understanding person.

We can find such subject omission in pro-drop languages like Japanese, which allow subjects and other arguments to be omitted if they are inferable from context. Pro-drop languages can look to outsiders as though they are leaving out critical information, but speakers of pro-drop languages routinely track who the most likely filler is for each role, using prominence in context or the entity that a discourse most centrally concerns to fill in unmentioned material (Okuma & Tamura, 1996; Yeh & Chen, 2007). With specific reference to Japanese, Fujisawa, Masuyama, and Naito (1993) found in a corpus study that 90.9% of omitted entities were subjects and 87.6% of the antecedents appeared in the previous or current sentence. This can be thought of as merely a more extreme version of what pronouns do in English. We might start a story in English with *I know this hipster named Peter*. Then we can continue *He wears suspenders to the beach*. Where in English we would continue with *He also eats nothing but liverwurst*, a Japanese speaker might conventionally just say the equivalent of *Eats nothing but liverwurst*. In the same way that *he* is more reduced than repeating *Peter* in each sentence, so saying nothing at all in Japanese is more reduced than repeating a pronoun.

Japanese (like other pro-drop languages) therefore provides a promising case to test for how necessary perspective is to person comprehension. When pronouns are omitted, but the subject (and its person) is obvious from context, like in the *Peter* example above, do comprehenders still adopt the appropriate perspective in mental simulation? Or do they understand the utterance, including who the subject is, without displaying perspective effects in simulation?

We investigated whether person and simulation perspective are separable, using a modified version of Brunyé et al.'s (2009) methodology in Japanese. First, we needed to determine whether Japanese is like English—whether sentences using a pronoun to indicate the person of the subject would lead Japanese speakers to respond faster to pictures with a compatible perspective. This was the aim of Experiment 1 (Section 2), below. Experiment 2 (Section 3) introduced the key test case by omitting the explicit subjects from critical sentences. When subject pronouns are omitted, the two accounts outlined above make different predictions. On the strong simulationist view, whenever Japanese speakers understand who the subject of a sentence is, even when the subject is omitted, they should be faster to respond to perspective-matching images. However, if Japanese comprehenders are able to understand the person of the omitted subject without displaying any perspectival facilitation effect, this only follows from the second view, that perspective is not essential to the comprehension of person. We subsequently conducted two small experiments to ensure that the difference between Experiments 1 and 2 was not due to differences in timing when a subject is omitted (Section 5) and then finally replicated the explicit–implicit subject difference observed in Experiments 1 and 2 in a within-participants design, in Experiment 4 (Section 6).

2. Experiment 1

The first experiment aimed to determine whether Japanese native speakers processing Japanese sentences with explicit subjects automatically and flexibly incorporate into their mental simulations the perspectives evoked by subject pronouns, in the same way that English native speakers do when processing English.

The method we used was adapted from the experiments conducted by Brunyé et al. (2009). Native speakers of Japanese read sequences of three sentences in Japanese, all containing subject pronouns (see (1) below for an example), which had either Second or Third Person subjects. This was the first factor of interest. Each critical set of three sentences was followed by a picture depicting the action from either an internal perspective (as if the participant were a participant of the action) or external perspective (as if the participant were an observer of the action performed by another person). This was the second factor of interest. Examples of these images are in Fig. 1, below. Filler sets of sentences were followed by images depicting an unrelated action. Half of these filler sets used second-person and half used third-person subjects, and orthogonally, half of the filler images adopted an internal and half an external perspective. However, we were only interested in the critical cases, in which the depicted action had been mentioned by the preceding language. The participant's task was to decide whether the depicted action had been mentioned in the preceding language or not.

2.1. Method

2.1.1. Sentence materials

Twenty-four sets of critical sentences were created, with both second-person and third-person versions of each sentence. In addition, 24 sets of filler sentences (with expected "no" responses) and 12 sets of practice sentences were created, half in the second person and half in the third person, similar to the criticals. In order to reinforce a particular

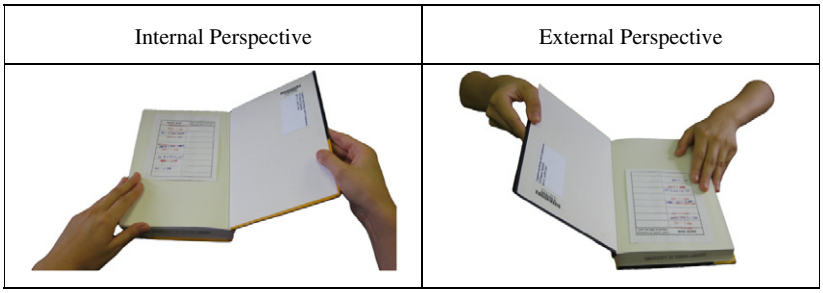


Fig. 1. Internal vs. external perspective images.

person as subject, the three sentences of each set included: (1a) a description of the protagonist, (1b) a description of a general event type that the protagonist is engaged in, and (1c) a description of an action with a temporal marker indicating that the event is ongoing. Each set of sentences set up a discourse context using exclusively second-person or third-person subjects. In the second-person condition, each sentence began with the second-person pronoun *anata* 'you'. In the third-person condition, the first sentence began with a Japanese first name (like *Taiki*) and the next two began with the third-person pronoun *kare* 'he'.

(1) a. Anata/Taiki-wa toshokan-de hataraite-imasu
You/Taiki-top library-loc work-prog
'You/Taiki are/is a librarian (working at a library).'

b. Anata/ Kare-wa hon-no kashidashibi-o shirabete-imasu
You/Taiki-top book-gen due dates-acc check-prog
'You/He are/is checking due dates.'

c. Anata/Kare-wa choodo ima hon-o hiraiteiru-tokoro-desu
You/He-top right now book-acc open-prog-cop
'You/He are/is opening the book right now.'

The complete list of materials can be found in Appendix A.

2.1.2. Picture materials:

We employed the same specifications for picture stimuli used by Brunyé et al. (2009). Images were created specifically for these experiments, with photographs taken at a viewing distance of about 40 inches and at a 35 degree downward angle. The experiment used 84 pictures, one for each sentence set/perspective combination. Specifically, for each of the 24 sets of critical sentences, pictures of a corresponding event depicted from both internal (i.e., performer's viewpoint) and external perspective (i.e., observer's viewpoint) were created, producing a total of 48 pictures (Fig. 1). Another 24 sets of filler sentences were followed by unrelated pictures depicted from an internal or an external perspective, with 12 from each perspective. An additional 12 pictures (six internal and six external perspective) were created for the practice session. Filler and practice pictures were developed in a manner similar to the criticals.

A norming study verified that each of the 48 critical sentences clearly conveyed the intended event depictions. This norming study used four native speakers of Japanese who did not participate in the main experiment. Each picture was presented for 1200 ms on the computer screen. The participant then gave a description of the picture. Pictures were selected if at least three participants gave a response that matched the intended event description.

2.1.3. Procedure

Participants were tested individually. The experiment began with a set of six practice trials. To confirm that participants understood their task, during the practice session participants received feedback on the picture verification and the comprehension questions. The practice session continued until participants correctly answered six consecutive trials. Failure before the sixth consecutive correct answer resulted in another six trials being randomly selected from the 12 practice items. After successful completion of the practice session came the experimental session, which consisted of 24 criticals (requiring "yes" responses) randomly interspersed with 24 fillers (requiring "no" responses). Half of the critical and half of the filler items were in the second-person condition, and half of each were in the third-person, and as a result participants encountered an equal number of second-person and third-person sentences in both "yes" and "no" trials in the experimental session.

For each trial, a fixation cross appeared on-screen for 500 ms, followed by the first sentence in the middle of the screen. Each of the three sentences in the set remained on the screen for 1500 ms before being replaced by the next sentence. After the third and final sentence, another fixation cross was displayed for 500 ms, followed by a picture depicting the action described in the final sentence, from either an internal or external perspective. Participants then decided if the pictured event was mentioned in the prior set of sentences, as quickly and accurately as possible, and indicated their response by pressing a button ("l" for *yes* and "a" for *no*). Critically, we expected participants to answer "yes" if the depicted event was mentioned, regardless of the pictured perspective. Although the instructions did not explicitly mention perspective, participants implicitly complied with this expectation, systematically ignoring perspective variations. Finally, to ensure that participants were attentive to all presented sentences (and not merely the third and final sentence of

each trial), one third of all trials were followed by comprehension questions that addressed the first, second, or third sentence of the given set, in equal proportions. Responses and reaction times for picture verification and comprehension questions were recorded by the experimental software.

Critical sentences and corresponding pairs of internal and external perspective pictures were fully crossed so that matching and mismatching pictures for each item were presented across participants. The four conditions—two sentence pronouns (*anata* 'you', *kare* 'he') and the two picture perspectives (internal, external)—were distributed across four lists in a Latin-square design. Each set of experimental sentences appeared in only one of these four conditions for each participant, counterbalanced across participants, with each participant receiving an equal number of experimental items in each condition.

2.1.4. Participants

Sixty-four native speakers of Japanese, all of whom were students at the University of Hawai'i or Hiroshima University, participated in exchange for credit in an introductory linguistics course or for a small amount of monetary compensation. All participants reported normal or corrected-to normal hearing and vision.

2.1.5. Predictions

We hypothesized that if personal pronouns drive Japanese comprehenders to adopt specific perspectives in mental simulation, as they do in English, then participants should respond faster to pictures whose perspective matched the perspective elicited by the pronouns in the preceding sentences. That is, an internal perspective picture should be recognized faster after second-person sentences than after third-person sentences. Likewise, an external perspective picture should have shorter recognition time after a third-person sentence than after its second-person counterpart.

2.2. Results

No subjects or items were excluded. Incorrect responses and exceedingly slow responses (those over 3000 ms) as well as responses that were more than 2.5 *sd* above or below the mean response time for each participant or item were removed. This resulted in less than 3.6% of the data being eliminated. Accuracy in picture responses for the items averaged 99.1%; accuracy for each condition is given below Fig. 2. Two-way Repeated-Measures ANOVAs revealed a marginal main effect of picture Perspective in the subject analysis ($F_1(1,63) = 3.2$, $p = 0.08$, $\eta_p^2 = 0.05$) and in the item analysis ($F_2(1,23) = 2.7$, $p = 0.1$, $\eta_p^2 = 0.11$). Pronoun produced no significant main effect ($F_1(1,63) = 0.08$, $p = 0.8$, $\eta_p^2 = 0.001$; $F_2(1,23) = 0.4$, $p = 0.6$, $\eta_p^2 = 0.02$). Importantly, however, we observed a large and significant interaction between Pronoun and picture Perspective ($F_1(1,63) = 10.4$, $p < 0.01$, $\eta_p^2 = 0.14$; $F_2(1,23) = 31.5$, $p < 0.01$, $\eta_p^2 = 0.6$). Planned pairwise *t*-tests showed that after participants read sentences marked with third-person pronouns, external-perspective pictures were verified significantly faster than internal-perspective pictures ($t_1 = 4.1$, $p = 0.0001$; $t_2 = 5.0$, $p < 0.01$). Likewise, after sentences marked with second-person pronouns, internal-perspective pictures were processed quantitatively faster than external-perspective ones, though this difference was only significant in the items analysis ($t_1 = 1.4$, $p = 0.16$; $t_2 = 2.4$, $p = 0.02$). As seen in Fig. 2 below, internal-perspective pictures were verified faster after *you* sentences than *he* sentences ($t_1 = 2.78$, $p < 0.01$; $t_2 = 3.8$, $p < 0.01$), while the reverse was true for external-perspective pictures ($t_1 = 2.5$, $p = 0.014$; $t_2 = 4.2$, $p < 0.01$). In sum, similar to previous studies on English, responses were significantly faster when the picture matched the perspective implied by the explicit pronoun—second-person pronouns facilitated internal perspective, while third-person pronouns facilitated external perspective.

In order to assess whether there could be a speed-accuracy tradeoff present in these data, we conducted an error analysis on the data after again excluding responses over 3000 ms. We ran Repeated-Measures ANOVAs with accuracy as the dependent measure and Person and picture Perspective as the independent measures. These revealed no significant main effects of Person (Fs < 1) or Perspective ($F_1 < 1$; $F_2(1,23) = 1.39$, $p = 0.25$, $\eta_p^2 = 0.06$), nor an interaction effect ($F_1 < 1$; $F_2(1,23) = 1.32$, $p = 0.26$, $\eta_p^2 = 0.05$).

2.3. Discussion

These results show that sentences about *you* drive faster responses to images taken from an internal perspective

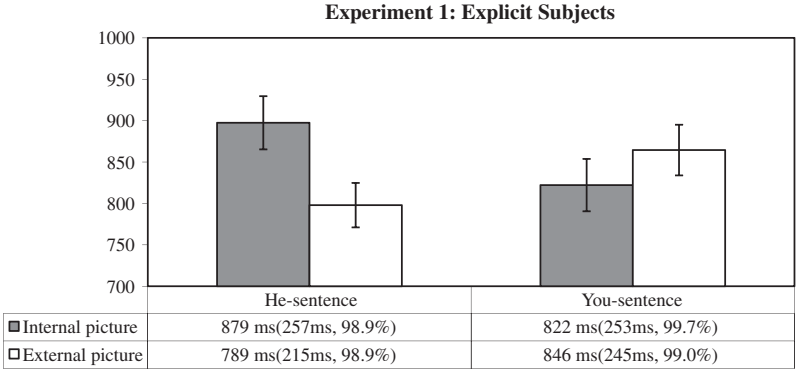


Fig. 2. Mean response times (ms), standard deviation, and accuracy (%) for verification of pictures. Error bars indicate standard error. Mean RTs for picture verification demonstrated an interaction effect between picture Perspective (internal vs. external) and Person (2nd vs. 3rd) with explicit pronouns.

while sentences about a third person facilitate responses to images from an external perspective. In line with previous research using similar methodologies (Brunyé et al., 2009; Zwaan et al., 2002), this in turn allows us to infer that Japanese language comprehenders not only mentally simulate the actions involved in described events, but also specifically simulate them from an internal or external perspective that is modulated by second- or third-person language. This replicates findings from English in Japanese, an unrelated, culturally distinct, and typologically different language.

A further consequence of these results has to do with the robustness of perspective effects. We observed perspective effects despite a difference in the methods from previous work; Brunyé et al. (2009) used no-response (that is, filler) images that depicted the scene described by the preceding set of sentences but with the action not being performed (for instance, for a sentence about *slicing a tomato*, it showed hands merely posed with a knife over a tomato, but not actively slicing it). This design feature could in principle have increased the degree of attention that participants paid to the performance of the action itself, and thus augmented the level of detail in simulation beyond what would normally be observed in sentence processing. A side effect could have been the inclusion of perspective in mental simulations. Instead, in the current work, we used no-response images that depicted unrelated objects and actions, which had less potential to artificially inflate the degree of detail in mental simulations. And yet, we measured the same sort of effect of person on perspective in mental simulation. With this approach, it could not be that the detailed nature of the yes–no decision forced participants to attend to the performance of the action. This suggests that Brunyé et al.'s (2009) finding is not driven by strategic attention to detail driven by the character of the task.

While these results confirm that Japanese speakers, like English speakers, engage mental simulations from perspectives indicated by the explicit subjects of sentences, they do not demonstrate that this perspective in mental simulation is a necessary prerequisite to understanding who performed the described action. In other words, they are consistent with the possibility that language comprehenders could in principle successfully identify and represent the action executer without adopting a particular perspective in their mental simulation of the described scene. To address this issue, we conducted a second experiment, in which we removed subject pronouns from the critical sentences.

3. Experiment 2

While sentences in English almost always include an explicit subject in each sentence, the same is not true for all languages. Japanese speakers, for instance, omit subjects regularly, provided that the implied subject is recoverable from context. Japanese comprehenders are perfectly able to understand these subjectless sentences, including who the event participants are and what roles they play. However, it isn't currently known whether Japanese sentences without subject pronouns drive mental simulation to incorporate a particular perspective, as sentences with subjects do. If they do, this would be compatible with both the view that perspectival mental simulation is essential to comprehending person and the view that it is not. However, an absence of perspective in mental simulation in response to subjectless sentences would be compatible with only the latter view—assuming that comprehenders are in fact understanding who did what to whom. Said differently, if dropping subjects also drops perspective from mental simulation, then this would suggest that identifying the subject in understanding a sentence does not require mental simulation from the corresponding perspective.

We investigated this issue through an experimental design and methodology almost identical to Experiment 1. The images and the first two sentences in each set were exactly the same as before; they had explicit subjects. The only difference was that subject pronouns were removed from the third sentence of each trial (see Section 3.1). Since the method and stimuli were identical to those in Experiment 1, with this single exception, similar effects observed in the two studies would indicate that comprehenders adopt a particular simulation perspective regardless of the presence or absence of explicit subjects. However, if the perspective-compatibility effect from Experiment 1 were to disappear when subjects are omitted in Experiment 2, this would imply that explicit subjects and implicit subjects do not have the same effects on the perspective that comprehenders adopt when simulating described scenes. This in turn would suggest that visual perspective in mental simulation is not necessary for understanding the who-did-what-to-whom of a sentence.

3.1. Method

3.1.1. Sentence materials

Sets of critical sentences were identical to those used in Experiment 1, but the subject pronoun was removed from the third sentence of each trial, as in (2)

(2) a. Anata/Naoya-wa hikkoshiyasan-desu
You/Naoya-top mover-cop
'You/Naoya are/is a mover.'

b. Anata/Kare-wa heya-no katazuke-o shite-imasu
You/He-top room-gen clean-up-acc do-prog
'You/He are/is cleaning up the room.'

c. Ima danboorubako-ni teepu-o hatteiru-tokoro-desu
Now cardboard box-loc tape-acc stick-prog-cop
'(You/He are/is) taping the cardboard box now.'

It was critical that we ascertain whether native Japanese speakers were able to determine who the subjects of these final sentences were. If not, then these stimuli would not allow us to test whether comprehenders successfully understand person independent of perspective in mental simulation. To this end, we conducted a norming study

with four native speakers of Japanese. Each norming participant read all 24 sets of critical sentences and was asked to identify the subject of the final sentence. Each participant identified the implied subject of every final sentence with 100% accuracy.

3.1.2. Picture materials:

Picture stimuli used in Experiment 2 were identical to those used in Experiment 1.

3.1.3. Procedure

The procedure was identical to the one described in Section 2.1.

3.1.4. Participants

A new group of 36 native Japanese speakers, all of whom reported normal or corrected-to normal hearing and vision, participated in the experiment in exchange for credit in introductory linguistics courses at the University of Hawai'i.

3.2. Results

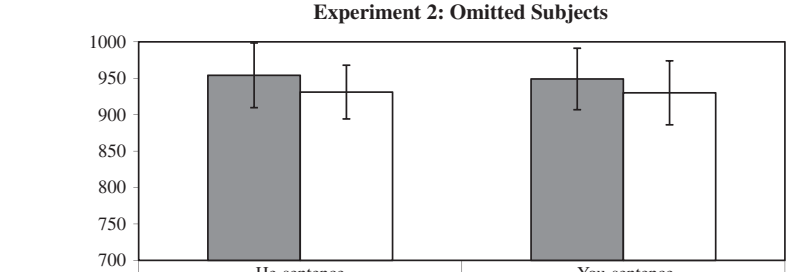
The data were trimmed as in Experiment 1, resulting in less than 3.8% of the data being removed, and no subjects or items being eliminated. Accuracy in picture responses for the items averaged 99.4%; accuracy for each condition is given below Fig. 3.

As Fig. 3 shows, unlike in Experiment 1, there was no interaction between Person and Perspective. Though the subjects of sentences were retrievable, in the absence of pronouns, the compatibility effect seen in Experiment 1 disappeared (in Repeated-Measures ANOVAs by participants and items, both Fs < 1). It also did not trend numerically in the direction of the interaction shown in Experiment 1. Moreover, while events depicted from an external perspective were verified slightly faster than those depicted from an internal perspective, this difference was not significant, nor was the main effect of Pronoun (see Fig. 3).

To determine whether error rates differed significantly across conditions, we conducted Repeated-Measures ANOVAs with accuracy as the dependent measure (after eliminating responses slower than 3000 ms). These revealed no significant main effects of Person (Fs < 1) or Picture ($F_1(1,35) = 1.84$, $p = 0.18$, $\eta_p^2 = 0.05$; $F_2 < 1$), nor an interaction effect ($F_1(1,35) = 1.84$, $p = 0.18$, $\eta_p^2 = 0.05$; $F_2(1.23) = 3.29$, $p = 0.08$, $\eta_p^2 = 0.13$).

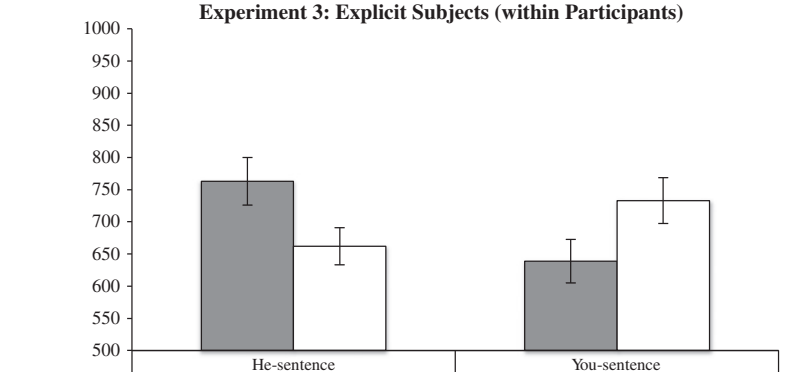
4. Combined analysis of Experiments 1 and 2

Experiment 1 used Japanese stimuli that were much like the English ones used previously in studies of



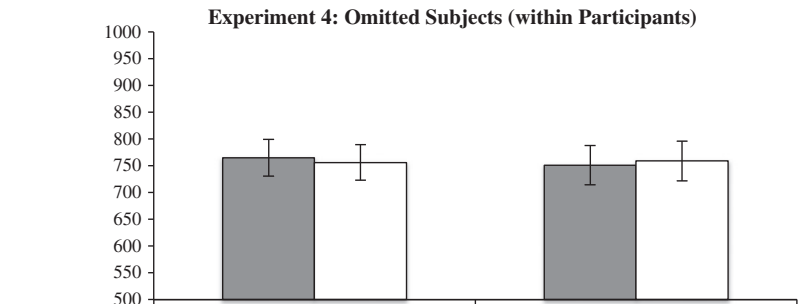
	He-sentence	You-sentence
Internal picture	945 ms(266ms, 98.6%)	949 ms(253ms, 99.5%)
External picture	913 ms(222ms, 100%)	930 ms(263ms, 99.5%)

Fig. 3. Mean response times (ms), standard deviation, and accuracy (%) for verification of pictures. Mean RTs for picture verification with omitted pronouns demonstrated no interaction between picture Perspective (internal vs. external) and Person (2nd vs. 3rd). Error bars indicate standard error.



	He-sentence	You-sentence
Internal picture	763 ms(233ms, 99.2%)	639 ms(215ms, 99.2%)
External picture	662 ms(183ms, 99.2%)	733 ms(225ms, 98.8%)

Fig. 4. Mean response times (ms), standard deviation, and accuracy (%) for verification of pictures. Error bars indicate standard error. Mean RTs for picture verification demonstrated an interaction effect between picture Perspective (internal vs. external) and Person (2nd vs. 3rd) with explicit pronouns.



	He-sentence	You-sentence
Internal picture	765 ms(217ms, 97.9%)	751 ms(210ms, 98.8%)
External picture	756 ms(231ms, 99.2%)	759 ms(235ms, 99.4%)

Fig. 5. Mean response times (ms), standard deviation, and accuracy (%) for verification of pictures. Mean RTs for picture verification with omitted pronouns demonstrated no interaction between picture Perspective (internal vs. external) and Person (2nd vs. 3rd). Error bars indicate standard error.

perspective, and showed that second-person subjects facilitate processing of internal-perspective images, while third-person subjects facilitate processing of external-perspective images. Experiment 2 showed no such effect; not only was the effect non-significant, there was no quantitative trend in the direction of the effect exhibited in Experiment 1. In order to examine whether the presence (Experiment 1) or absence (Experiment 2) of explicit subject pronouns affected the presence of perspective in mental simulation, we combined the data from the two experiments for analysis. We conducted two Repeated-Measures ANOVAs with three independent variables: Person, Perspective, and Experiment, one taking participants as random factors, and the other taking items as random factors. These revealed a significant main effect of picture Perspective only in the subject analysis, but not in the item analysis ($F_1(1,98) = 4.1$, $p = 0.045$, $\eta_p^2 = 0.04$; $F_2(1,23) = 1.2$, $p = 0.29$, $\eta_p^2 = 0.05$), where external-perspective images were recognized faster than internal-perspective ones (Mean: 867 vs. 893, respectively). The only significant two-way interaction was between Pronoun and picture Perspective ($F_1(1,98) = 4.9$, $p = 0.03$, $\eta_p^2 = 0.05$; $F_2(1,23) = 7.1$, $p = 0.01$, $\eta_p^2 = 0.24$). This effect appears to have been predominantly carried by the results from the first experiment. Evidence for this comes from the critical three-way interaction among Person, picture Perspective, and Experiment, which was significant by both participants and items ($F_1(1,98) = 4.5$, $p = 0.04$, $\eta_p^2 = 0.04$; $F_2(1,23) = 17.1$, $p < 0.01$, $\eta_p^2 = 0.43$). Combined with the presence of a significant interaction between Person and Perspective in Experiment 1 and the absence of a significant interaction effect in Experiment 2, this three-way interaction suggests that the presence or absence of explicit subject pronouns had different effects on whether language comprehenders adopted particular perspectives in response to person.

5. Discussion: The absence of perspective effects when subjects are omitted

The absence of perspective effects in simulation in Experiment 2 and the significant three-way interaction in the combined analysis above seem to suggest that people do not adopt a perspective in mental simulation when subjects are not lexically marked, even though they understand who did what to whom. However, before reaching this conclusion, we have to consider two other potential explanations for the finding, pertaining to the time course of simulation.

Sentences with and without explicit subjects (such as the critical sentences in the two experiments, respectively) might well be processed differently from each other. Specifically, null subject sentences (as in Experiment 2) might take longer to process, due to the fact that their subjects have to be retrieved from somewhere other than the sentences themselves. This longer processing time might push back the point when mental simulation from a particular perspective can begin, so much that—in the task used in the experiments above—comprehenders still have not begun simulation when the image is presented. If they have not yet constructed a mental simulation when prompted with the image, no interaction between Person and picture Perspective should be observed.

The literature does not precisely indicate whether, in Japanese or other pro-drop languages, sentences with omitted subjects are processed more slowly than sentences with subjects, but from our own data, we can see suggestive evidence for such an effect. Response times to images were slower overall in the null-subject sentences of Experiment 2 (average RT: 936 ms) compared to their explicit counterparts from Experiment 1 (average RT: 846 ms). Because the presence or absence of pronouns was manipulated between participants, slower image processing times in Experiment 2 could very well be due to factors other than the absence of the pronoun. However, they could also be due to later mental simulations when people process sentences with subjects, in the following way. If people take longer to construct mental simulations while processing sentences without explicit subjects, then when an image is presented relatively soon (500 ms) after the offset of the sentence, as in the experiments described above, comprehenders might have a harder time matching images to preceding sentences, simply because they have not yet had time to access the perceptual details appropriate to the described scene.

We tested this explanation in a follow-up experiment, in which we provided comprehenders with more time to retrieve the null subject referents and potentially to incorporate the implied perspective into eventual simulations. In the follow-up experiment, we increased the delay before the picture verification task from 500 ms (in Experiment 2) to 750 ms (in Experiment 3a). This increase of 250 ms, we thought, would more than compensate for the potentially longer processing time for subjectless sentences (which as noted above took 90 ms longer to process than their explicit-subject counterparts). Sixteen native speakers of Japanese participated in exchange for a small amount of monetary compensation at Hiroshima University. The collected data were systematically trimmed as in Experiments 1 and 2, resulting in less than 5% of the data being eliminated, and no participants or items were removed. No main effect of picture Perspective was observed (average RT for Internal Perspective was 716 ms (*sd* = 201), and for External Perspective, was 730 ms (*sd* = 188)). Nor was a significant interaction of picture Perspective and Person (Average RT for Perspective-Person matching conditions was 719 ms (*sd* = 182), and for mismatching conditions was 727 ms (*sd* = 207)). A significant main effect of Person was observed in the subject analysis, but was not significant by items: $F_1(1,15) = 5.7$, $p = 0.03$, $\eta_p^2 = 0.28$; $F_2(1,23) = 0.8$, $p = 0.38$, $\eta_p^2 = 0.03$ (average RT for Second-person sentences was 712 ms (*sd* = 197), and for Third-person was 734 ms (*sd* = 188)).

The absence of an interaction effect with a longer sentence-image latency in this follow-up experiment suggests that the absence of an interaction effect with omitted subject pronouns is not solely due to participants having insufficient time to construct mental simulations of described scenes with the appropriate perspective. But another time-based hypothesis is possible. Perhaps instead of taking longer to construct a mental simulation in the absence of an explicit subject pronoun, comprehenders might actually have processed the subjectless sentences faster, as has been shown in Spanish (Gelormini & Almor, 2011) or processed them as fast as overt pronoun sentences, as has been shown in Chinese (Yang, Gordon, Hendrick, & Wu, 1999). After all, subjectless sentences are shorter because they have no subject. According to accessibility theory, among referring expressions in discourse, null pronouns are ranked highest in the hierarchy of accessibility markers. This signals that the omitted subject is highly accessible and retrievable from a mental representation of the developed context or a discourse (Ariel, 1991), which might lead comprehenders to adopt a context-induced perspective quickly when performing mental simulation. As a result, construction of mental simulations in Experiment 2 might have already been completed by the time the image was presented. Quick completion of simulation would provide extra time for the comprehender to activate other information associated with the described event, information which might be irrelevant to perspective, and which may have obscured any perspective effects. This line of reasoning could explain the lack of match-mismatch perspective effects in Experiment 2, and could also elucidate the overall slower RTs found in Experiment 2; if the image prompt was presented too long after simulation terminated, then image responses might not have been facilitated by a simulation process.

This hypothesis led us to design a second follow-up experiment, where we decreased the interval between the offset of the final sentence and the picture presentation from 500 ms (as in Experiment 2) to 200 ms (Experiment 3b). Twenty native speakers of Japanese participated in this study in exchange for a small amount of monetary compensation at Hiroshima University. The same trimming criteria applied to the other experiments resulted in less than 5% of the data being removed, and no subjects or items were eliminated. No main effects were observed (*F*s < 1; average RT for Second-person sentences was 864 ms (*sd* = 200), and for Third-person sentences was 836 ms (*sd* = 182); average RT for Internal Perspective was 845 ms (*sd* = 170), and for External Perspective was 855 ms (*sd* = 214)). Nor were any interaction effects observed (average RT for Perspective-Person matching conditions was 862 ms (*sd* = 210), and for mismatch conditions was 838 ms (*sd* = 164)).

Results from these two follow-up experiments did not support the time-based explanations for the absence of perspective effects in subjectless sentences. That is, it appears not to be the case that perspective effects disappear due to insufficient time to process a sentence, or due to early completion of simulation.

6. Experiment 4

Although Experiment 1 (with overt pronouns) provided results consistent with a simulated perspective, the three experiments with null pronouns did not. Still, because these were separate studies, conducted with different samples of participants, these differences could in principle be caused by differences across the samples, not by simulation effects induced by the explicit and implicit pronouns. To ensure that the observed effects of perspectival pronouns is robust, we conducted another experiment in which we combined the critical manipulations from Experiments 1 and 2—that is, presence or absence of subject pronoun—as a within-subjects factor.

A new group of 44 native Japanese speakers, who reported normal or corrected-to normal hearing and vision, participated in this experiment in exchange for a small amount of monetary compensation. In this design, half of the participants did the complete task of Experiment 1 with overt pronouns, and then did the complete task of Experiment 2 with null pronouns. The other half participated in the experiments in the reversed order, Experiment 2 first and then Experiment 1. In either case, the participants were exposed to the same set of contexts and picture materials twice. To avoid a situation where participants would receive the same set of contexts with the same pronoun and the same perspectival pictures in both the first and the second experiment, participants were assigned to different lists. Moreover, each trial was followed by different comprehension questions in the two experiments.

6.1. Results

The data were trimmed as in Experiments 1 and 2, resulting in less than 4.2% of the data being removed in

the Explicit Experiment while 4.3% was removed in the Null Experiment. No item, but one subject was eliminated due to low accuracy in the picture verification task (lower than 75% accuracy) and an additional three subjects were eliminated to have an equal number of participants in each list. This resulted in forty participants—19 participants participated in the Explicit Experiment first while 21 participants did the Null Experiment first. Accuracy in picture responses for the items averaged 99.1% for the Explicit Experiment (accuracy for each condition: 'He' followed by an external picture (He-Ext) = 99.2%, 'He' followed by an internal picture (He-Int) = 99.2%, 'You' followed by an external picture (You-Ext) = 98.8%, and 'You' followed by an internal picture (You-Int) = 99.2%) while the average accuracy was 98.8% for the Null Experiment (accuracy for each condition: He-Ext = 99.2%, He-Int = 97.9%, You-Ext = 99.4%, You-Int = 98.8%).

First, we conducted three-way Repeated-Measures ANOVAs with Person, Perspective, and Explicitness as the independent variables and response times to pictures as the dependent measure. No significant main effect of picture Perspective (Fs < 1) or Pronoun ($F_1(1,78) = 2.5$, $p = 0.1$, $\eta_p^2 = 0.03$; $F_2(1,46) = 2.6$, $p = 0.1$, $\eta_p^2 = 0.05$) was observed. We found a significant two-way interaction between Pronoun and picture Perspective ($F_1(1,78) = 10.8$, $p = 0.002$, $\eta_p^2 = 0.1$; $F_2(1,46) = 14.7$, $p < 0.001$, $\eta_p^2 = 0.24$). Importantly, the three-way interaction among Person, picture Perspective, and Explicitness was significant by both participants and items ($F_1(1,78) = 7.7$, $p = 0.007$, $\eta_p^2 = 0.09$; $F_2(1,46) = 13.9$, $p = 0.001$, $\eta_p^2 = 0.23$). This three-way interaction suggests that whether or not the subject was explicitly marked affected the perspective that language comprehenders adopted in simulating the described event.

Second, we conducted two Repeated-Measures ANOVAs (one each for the Explicit half and the Null half) with Pronoun and picture Perspective as the independent variables and response times to pictures as the dependent measure (Fig. 4). When subject was Explicit, there was a main effect of picture Perspective in the subject analysis ($F_1(1,39) = 4.2$, $p = 0.048$, $\eta_p^2 = 0.1$), but the effect was marginal in the item analysis ($F_2(1,23) = 2.6$, $p = 0.1$, $\eta_p^2 = 0.1$). Pronoun produced no significant main effect ($F_1(1,39) = 0.03$, $p = 0.9$, $\eta_p^2 = 0.001$; $F_2(1,23) = 0.06$, $p = 0.8$, $\eta_p^2 = 0.002$). Importantly, we observed a large and significant interaction between Pronoun and picture Perspective ($F_1(1,39) = 12.5$, $p < 0.01$, $\eta_p^2 = 0.24$; $F_2(1,23) = 32.6$, $p < 0.01$, $\eta_p^2 = 0.6$).

Planned pairwise t-tests revealed that external-perspective pictures were verified significantly faster than internal-perspective pictures ($t_1 = 3.5$, $p = 0.001$; $t_2 = 4.1$, $p < 0.001$) after participants read sentences marked with third-person pronouns. The opposite was also true; internal-perspective pictures were processed faster than external-perspective ones ($t_1 = 2.5$, $p = 0.02$; $t_2 = 3.8$, $p = 0.001$), after sentences marked with second-person pronouns. Moreover, internal-perspective pictures were verified faster after *you* sentences than after *he* sentences ($t_1 = 2.5$, $p = 0.02$; $t_2 = 2.8$, $p = 0.01$), while the reverse was true for external-perspective pictures ($t_1 = 3.9$, $p < 0.001$; $t_2 = 4.6$, $p < 0.001$). These results replicated Experiment 1 with overt pronouns and showed that second-person pronouns facilitated internal perspective, while third-person pronouns facilitated external perspective.

In order to check whether error rates were significantly different across conditions, we conducted Repeated-Measures ANOVAs with accuracy as the dependent measure (after eliminating responses slower than 3000 ms). These revealed no significant main effects of Person (Fs < 1) or Picture (Fs < 1), nor an interaction effect (Fs < 1).

In the Null half, we observed no main effect of picture Perspective (average RT for Internal Perspective was 759 ms (sd = 221), and for External Perspective, 762 ms (sd = 200)) and no main effect of Person (average RT for Second-person sentences was 753 ms (sd = 214), and for Third-person was 768 ms (sd = 212)), all Fs < 1 (Fig. 5). No significant interaction of picture Perspective and Person was found (Average RT for Perspective-Person matching conditions was 758 ms (sd = 207), and for mismatching conditions was 763 ms (sd = 221)), Fs < 1. These results also replicated Experiments 2 with null pronouns.

To test whether error rates differed across conditions, Repeated-Measures ANOVAs with accuracy as the dependent measure (after eliminating responses slower than 3000 ms) were conducted. These revealed no significant main effects of Person (Fs < 1) or Picture ($F_1(1,39) = 3.4$, $p = 0.07$, $\eta_p^2 = 0.08$; F_2 (1,23) = 2.9, $p = 0.1$, $\eta_p^2 = 0.1$), nor an interaction effect ($F_1(1,39) = 1.47$, $p = 0.23$, $\eta_p^2 = 0.04$; $F_2(1.23) = 1.0$, $p = 0.33$, $\eta_p^2 = 0.04$).

In sum, perspective effects were driven by the presence of explicit subject pronouns, even when pronoun manipulations were introduced as a within-subjects factor.

7. General discussion

Experiment 1, like other research in the same paradigm (Brunyé et al., 2009), demonstrates that language comprehenders adopt different visual perspectives when processing language about actions *you* perform and actions *he* or *she* performs. The results from Experiment 1 extend previous findings by showing that this is not a unique property of English—personal pronouns have similar effects in Japanese, a linguistically and culturally unrelated and typologically different language.

We investigated the notion that mental simulation of described scenes plays a functional role in language comprehension—that running perceptual and motor simulations of described content provides the cognitive material that generates the mechanical substrate for generating inferences, updating beliefs appropriately, and preparing an appropriate response (Barsalou, 1999; Feldman & Narayanan, 2004; Zwaan, 1999). Given that previous work, as well as Experiments 1 and 4, find second-person language to prime processing of images of described scenes as shown from the subject's perspective, and third-person language to prime processing of images shown from an outside observer's perspective, it's tempting to conclude that the way a comprehender understands the difference between second- and third-person language is a result of differences in the perspective adopted in the mental simulations he or she constructs during processing. Language about *you* leads a comprehender to simulate him or herself

as the protagonist, and to make appropriate inferences, update beliefs, and so on, and likewise for third-person sentences, with an external perspective on events in simulation.

But our findings in Experiments 2, 3, and 4 with null subjects suggest that it's possible to construct a mental representation of who an utterance is about—who its protagonist is—without exhibiting behavior demonstrative of performing visual simulation of the described scene from that person's perspective. By manipulating the presence or absence of a subject pronoun as a within-subjects factor, Experiment 4 confirms that these perspective effects were unlikely to be due to uncontrolled differences across participant samples. Japanese comprehenders are able to process the meanings of subjectless sentences when the subject is clear from context, fully knowing who did what to whom, and at the same time exhibit no priming of event perception from that person's perspective. This forces us to reconsider the role that perspectival mental simulations play during language processing. While explicit grammatical person may drive comprehenders to adopt one perspective or another, this process may not be necessary for their construction of an accurate representation of the utterance's meaning, or for the relevant inferences and other components of comprehension.

How to make sense of this? It's possible that mental simulation may still adopt one perspective or another even when person is not linguistically expressed, but just not to the extent that perspective strongly modulates participant performance. In essence, our null finding in the null expression condition could be due to an instrument that was too blunt or inappropriate. This interpretation is consistent with previous work that suggests that a range of strength and richness of perceptual and motoric information can be simulated depending on which perspective is adopted and what type of information (e.g., action-related or object-related information) is simulated in understanding the event (Ditman, Brunyé, Mahoney, & Taylor, 2010). This is also analogous to an enactment effect; that is, that overt performances of activities during the activities' encoding are retained longer in memory than covert/imagined performances or silent-verbal encodings (Hornstein & Mulligan, 2004; Masumoto et al., 2006; Nilsson et al., 2000).

Another possibility is that understanding Japanese sentences with explicit pronouns, but not those with null pronouns, drives the pronoun-perspective effects because repeating pronouns in a pro-drop language is an infrequent and marked situation, which signals contrastive focus towards an unexpected referent (Blackwell, 2001). Such repetition of pronouns might therefore be salient to comprehenders and the perspective effects observed in the current study could be due to this specific task. The fact that Japanese comprehenders do not display perspective effects in the more natural pro-drop contexts might indicate that perspective adoption is less common overall by Japanese than English users.

Or perhaps we were looking in the wrong modality. Adopting a particular perspective modulates not just the visual features we tested, but also what motor behaviors are appropriate. It's possible that language comprehenders perform mental simulations from a given perspective even when person is covert, but that this involves simulation of actions rather than vision.

The strongest possible interpretation of these results is that mental simulation does not play a functional role in understanding at all. It could be that comprehension of person—in the case of the utterances used in the current paradigm, who the agent is—can proceed entirely without mental simulation taking the appropriate perspective. That is, the perspective effects we and others have observed might be epiphenomenal—they might be activated collaterally or subsequent to whatever meaning processes are actually taking place, and serve no critical function in comprehension per se (see, for instance, arguments put forward by Mahon & Caramazza, 2008). Perhaps knowing who the agent is achieved by binding some abstract—amodal—representation of the agent, whether it be some third party or the comprehender, to a representation of the appropriate role in the described action.

The data we've reported do not discriminate among these possibilities. For that matter, it's also possible that perspective in mental simulation is integral to understanding, but that it's measurably different when subjects are explicit and when they are implicit.

All the foregoing discussion has been predicated on the assumption that what we learn about Japanese informs what comprehenders of other languages are doing. But this isn't necessarily true. It's possible that perspective in simulation plays different roles for comprehension, depending on the language. Typological differences across languages, for instance, whether a language allows subjects to be omitted liberally (like Japanese) or not (like English), could produce systematic differences in the processes in which speakers of different languages engage. Speakers of pro-drop languages like Japanese, where sentences regularly lack subjects, might be less systematic in their adoption of perspective in general than their non-pro-drop counterparts. Of course, English also does use null anaphora, such as in *You are washing dishes and looking for a cloth*. However, in English, which is categorized as a "sentence-oriented" or "subject prominent" language, these elliptical expressions are highly syntactically constrained so that their references are identified within the sentence structure (Huang, 1984; Li & Thompson, 1976). Therefore, the present null effects of implicit pronominal perspective on mental simulation in Japanese might or might not replicate with zero anaphora in English. On the other hand, in Japanese, which is a "discourse-oriented" or "topic prominent" language, referents of zero pronouns can be identified by the pragmatic or discourse context more generally (Huang, 1984; Li & Thompson, 1976). Therefore, speakers of pro-drop languages and those of non-pro-drop languages might need to recruit different strategies for representing differences in events that correspond to the person of their participants. These different routines in different languages might be entrenched as different patterns of thinking-for-speaking (Slobin, 1996). Differences in routine cognition like this could be part of what it means to think differently in different languages; the presence or absence of perspective in mental simulation could certainly be a consciously accessible subjective experience that people

notice and that contributes to what it feels like to be a speaker of one language or another.

On this account, Japanese comprehenders might adopt one of several strategies when processing subjectless sentences. It could be that they simulate content without adopting a specific or a single perspective, or simulations may involve activations of parallel representations from different perspectives (Connell & Lynott, 2007), when no linguistic cues specify perspective. In Japanese, when no particular perspective is linguistically prompted, participants might simulate the elements of the described scene not from a single perspective, but without adopting any perspective. Of course, it could be that, when confronted with subjectless sentences, Japanese comprehenders prefer to perform simulations from one particular perspective or the other. In Experiment 2, we observed that events depicted from an observer's perspective were responded to faster overall (931 ms) than those depicted from a participant's perspective (952 ms), though this difference was not significant. An overall preference for an external perspective may be a reflection of the Japanese language, which is often described as "Become-language" or "Theme-oriented language" (Ikegami, 1981; Suzuki, 2001), but, again, a tendency in this direction does not reach significance in our data.

While further work will be required to pin down what the exact functions of mental simulation in language understanding are (if any, and in response to what sorts of language), the current work has clear implications for sentence processing research. The difference in results between Experiments 1 and 2 (and in these same conditions within Experiment 4) demonstrates that the presence or absence of an explicit mention of person in a sentence matters. That is, sentences with no subject are not the same for the purposes of processing, at least as measured by mental simulation, as sentences with explicitly present subjects that indicate person. This is a non-trivial finding. Since native speakers of pro-drop languages are able to access the intended referents similarly whether or not a referent is explicitly expressed, it's quite reasonable to posit that their comprehension processes in the two cases could be similar.

These results also carry implications about the brain mechanisms involved in sentence processing. Representations in the vision system start by encoding viewpoint-specific representations of objects, but some later regions, such as the lateral occipital complex (Grill-Spector, Kourtzi, & Kanwisher, 2001) encode viewpoint-invariant representations. It could be that the presence or absence of explicit subjects affects which parts of the visual system are engaged for simulation—later areas that encode viewpoint-invariant representations when there is no subject, but earlier areas that encode viewpoint-specific representations when there is an explicit subject.

In summary, results from a first experiment confirmed that the perspective component of mental simulation is triggered by explicit subjects in Japanese as it is in English (Experiments 1 and 4). This corroborates the core claim of theories of grounded cognition (Barsalou, 2008; Zwaan & Madden, 2005). That is, understanding language engages mental simulation. However, even though perspective in simulation is a robust effect of sentences of a certain type, as observed in typologically distinct languages, it is only systematically varied when person is explicitly encoded by overt linguistic markers (Experiments 2, 3, and 4). These results lead us to conclude that contextual cues alone are not sufficient to generate perspective in simulation, at least not to the same extent as explicitly and locally marked perspective cues.

The lack of perspective effects with null subjects in Experiments 2, 3, and 4 suggests that even when the referent of a null subject is rapidly and naturally retrieved from the discourse context in conversation, the mental content of the null subject may not be integrated into mental simulation in the same way as it would be if explicitly mentioned. This disparity in mental representations may provide fertile ground for future investigation into the functional role of mental simulation, which appears to hinge both on linguistic details and on the information available in context during language comprehension.

Appendix A

Only the explicit subject version marked with a second-person pronoun is shown below. The null subject version was identical except that subject pronouns were removed from the third sentence of each trial.

Criticals	Fillers
1 You are a waiter at a restaurant You are serving a customer Right now, you are pouring milk (into the glass)	1 You are a policeman You are patrolling the neighborhood Right now, you are shining a flashlight
2 You are a coffee shop customer You are preparing to eat your meal Right now, you are stirring coffee	2 You are a doctor You are checking patients Right now, you are checking a patient's blood pressure
3 You are a postal worker You are mailing letters Right now, you are putting on a stamp (on the envelope)	3 You are a novelist You are writing a new book Right now, you are typing the first line
4 You are a kindergarten teacher You are teaching kids to draw a snowman Right now, you are drawing a circle (on the paper)	4 You are a bus driver You are stopping at the bus stop Right now, you are opening the door
5 You are a librarian You are checking due dates Right now, you are opening the book	5 You are a taxi driver You are dropping off riders Right now, you are giving them change
	6 You are a photographer

Appendix A (*continued*)

Criticals	Fillers
6 You are a kitchen helper You are cleaning up the dining room Right now, you are wiping the table	You are taking pictures of cities Right now, you are changing the lens
7 You are a grocery owner You are preparing to open the store Right now, you are opening a box	7 You are a pianist You are preparing for a concert Right now, you are tuning the piano
8 You are a mathematician You are calculating formulas Right now, you are punching a number on the calculator	8 You are a scientist You are performing an experiment Right now, you are checking your instruments
9 You are a grade-school teacher You are teaching kids to make paper animals Right now, you are folding colored paper	9 You are a security guard You are on night duty Right now, you are locking the gate
10 You are a house keeper You are cleaning the floor Right now, you are squeezing the towel	10 You are a fireman You are practicing a fire drill Right now, you are rolling up the fire hose
11 You are a baker You are making apple pie Right now, you are cutting an apple	11 You are a lawyer You are researching for a case Right now, you are reading documents
12 You are a bartender You are taking customer orders Right now, you are opening a bottle	12 You are a barber You have just finished work for the day Right now, you are sweeping the floor
13 You are a florist You are arranging a bouquet Right now, you are holding a vase	13 You are a baseball player You are playing in a tournament Right now, you are pitching the ball
14 You are a tailor You are making a skirt Right now, you are cutting some cloth (with scissors)	14 You are a veterinarian You are checking a sick cat Right now, you are giving the cat some medicine
15 You are a dishwasher You are cleaning the lunch dishes	15 You are a chef You are preparing for dinner Right now, you are cutting carrots

Appendix A (*continued*)

Criticals	Fillers
Right now, you are wiping a plate	
16 You are a mover You are cleaning up the room Right now, you are taping a box	16 You are a tour guide You are giving a tour of the city Right now, you are giving the brochures (to the tourists)
17 You are a secretary You are filing documents Right now, you are stapling the report	17 You are the captain of a ship You are about to dock the ship Right now, you are holding the steering wheel
18 You are a card player You are playing poker Right now, you are holding the cards	18 You are a mountain climber You are about to climb Mt. Everest Right now, you are holding (up) your boots
19 You are a cooking teacher You are preparing banana bread Right now, you are peeling a banana	19 You are a life guard You are working at the beach Right now, you are looking through the binoculars
20 You are a watch repair person You are fixing a watch Right now, you are resetting the watch	20 You are a TV reporter You are covering a news story Right now, you are holding a microphone
21 You are a dentist You are teaching a patient/someone how to brush teeth Right now, you are grabbing the toothbrush	21 You are a fisherman You are fishing at the lake Right now, you are putting bait on your line
22 You are a professor You are lecturing on chapter 3 from the textbook Right now, you are turning the page	22 You are an auto mechanic You are working on a truck Right now, you are changing a tire
23 You are a carpenter You are building a shelf Right now, you are pounding a nail	23 You are a magician You are performing on stage Right now, you are holding a dove
24 You are a waiter at a café You are setting the table Right now, you are folding a napkin	24 You are a construction worker You are building a bridge Right now, you are

(*continued on next page*)

Appendix A (*continued*)

Criticals	Fillers
	mixing some concrete

References

Ariel, M. (1991). The function of accessibility in a theory of grammar. *Journal of Pragmatics, 16*, 443–463.

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

Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology, 59*, 617–645.

Bergen, B. (2012). *Louder than words: The new science of how the mind makes meaning*. New York: Basic Books.

Bergen, B., & Chang, N. (2005). Embodied Construction Grammar in simulation-based language understanding. In J.-O. Östman & M. Fried (Eds.), *Construction grammars: Cognitive grounding and theoretical extensions* (pp. 147–190). Amsterdam: John Benjamins.

Bergen, B., Lindsay, S., Matlock, T., & Narayanan, S. (2007). Spatial and linguistic aspects of visual imagery in sentence comprehension. *Cognitive Science, 31*, 733–764.

Bergen, B., & Wheeler, K. (2010). Grammatical aspect and mental simulation. *Brain and Language, 112*, 150–158.

Blackwell, S. E. (2001). Testing the neo-Gricean pragmatic theory of anaphora: The influence of consistency constraints on interpretations of coreference in Spanish. *Journal of Pragmatics, 33*(6), 901–941.

Borghi, A. M., Glenberg, A. M., & Kaschak, M. P. (2004). Putting words in perspective. *Memory and Cognition, 32*(6), 863–873.

Brunyé, T. T., Ditman, T., Mahoney, C. R., Augustyn, J. S., & Taylor, H. A. (2009). When you and I share perspectives: Pronouns modulate perspective taking during narrative comprehension. *Psychological Science, 20*, 27–32.

Chatterjee, A. (2010). Disembodying cognition. *Language and Cognition, 2*, 79–116.

Connell, L. (2007). Representing object colour in language comprehension. *Cognition, 102*, 476–485.

Connell, L., & Lynott, D. (2007). Is a bear white in the woods? Parallel representations of implied object color during language comprehension. *Psychonomic Bulletin & Review, 16*(3), 573–577.

Dale, R., Dietrich, E., & Chemero, A. (2009). Explanatory pluralism in cognitive science. *Cognitive Science, 33*, 739–742.

Ditman, T., Brunyé, T., Mahoney, C., & Taylor, H. (2010). Simulating an enactment effect: Pro-nouns guide action simulation during narrative comprehension. *Cognition, 115*, 172–178.

Dove, G. (2009). Beyond perceptual symbols: A call for representational pluralism. *Cognition, 110*(3), 412–431.

Feldman, J., & Narayanan, S. (2004). Embodied meaning in a neural theory of language. *Brain and Language, 89*, 385–392.

Fujisawa, S., Masuyama, S., & Naito, S. (1993). A basic study on ellipsis and anaphora in Japanese sentences. *Proceedings of Information Processing Society of Japan, 34*(9), 1909–1918.

Gelormini, C., & Almor, A. (2011). Repeated names, overt pronouns, and null pronouns in Spanish. *Language and Cognitive Processes, 26*, 437–454.

Glenberg, A. M., & Kaschak, M. P. (2002). Grounding language in action. *Psychonomic Bulletin & Review, 9*, 558–565.

Grill-Spector, K., Kourtzi, Z., & Kanwisher, N. (2001). The lateral occipital complex and its role in object recognition. *Vision Research, 41*(10/11), 1409–1422.

Hornstein, S. L., & Mulligan, N. W. (2004). Memory for actions: Enactment and source memory. *Psychonomic Bulletin & Review, 11*(2), 367–372.

Huang, C. T. (1984). On the distribution and reference of empty pronouns. *Linguistic Inquiry, 15*(4), 531–574.

Ikegami, Y. (1981). *"Suru" to "naru" no gengogaku: Gengo to bunka no taiporojii e no shiron: Nihongo sosho*. Tokyo: Taishukan Shoten.

Li, C. N., & Thompson, S. A. (1976). Subject and topic: A new typology of language. In C. N. Li (Ed.), *Subject and topic* (pp. 457–489). New York: Academic Press.

MacWhinney, B. (2005). The emergence of linguistic form in time. *Connection Science, 17*, 191–211.

Madden, C. J., & Zwaan, R. A. (2003). How does verb aspect constrain event representations? *Memory and Cognition, 31*(5), 663–672.

Mahon, B. Z., & Caramazza, A. (2008). A critical look at the embodied cognition hypothesis and a new proposal for grounding conceptual content. *Journal of Physiology-Paris, 102*, 59–70.

Masumoto, K., Yamaguchi, M., Sutani, K., Tsuneto, S., Fujita, A., & Tonoike, M. (2006). Reactivation of physical motor information in the memory of action events. *Brain Research, 1101*(1), 102–109.

Nilsson, L. G., Nyberg, L., Klingberg, T., Aberg, C., Persson, J., & Roland, E. (2000). Activity in motor areas while remembering action events. *Neuroreport, 11*, 2199–2201.

Okuma, M. & Tamura, K. (1996). Zero pronoun resolution in Japanese discourse based on centering theory. In *Proceedings of the 16th international conference on, computational linguistics* (pp. 871–876).

Setti, A., Borghi, M. A., & Tessari, A. (2009). Moving hands, moving entities. *Brain and Cognition, 70*, 253–258.

Slobin, D. (1996). From "thought and language" to "thinking for speaking". In J. Gumperz & S. Levinson (Eds.), *Rethinking linguistic relativity* (pp. 70–96). Cambridge, UK: Cambridge University Press.

Stanfield, R. A., & Zwaan, R. A. (2001). The effect of implied orientation derived from verbal context on picture recognition. *Psychological Science, 12*, 153–156.

Suzuki, T. (2001). *Words in context: A Japanese perspective on language and culture (A. Miura, Trans.)*. Tokyo: Kodansha International.

Tucker, M., & Ellis, R. (1998). On the relations between seen objects and components of potential actions. *Journal of Experimental Psychology: Human Perception and Performance, 24*, 830–846.

Tucker, M., & Ellis, R. (2004). Action priming by briefly presented objects. *Acta Psychologica, 116*, 185–203.

Wheeler, K., & Bergen, B. (2010). Meaning in the palm of your hand. In S. Rice & J. Newman (Eds.), *Empirical and experimental methods in conceptual structure, discourse, and language*. Stanford, CA: CSLI.

Yang, C. L., Gordon, P. C., Hendrick, R., & Wu, J. T. (1999). Comprehension of referring expressions in Chinese. *Language and Cognitive Processes, 14*, 715–743.

Yaxley, R. H., & Zwaan, R. A. (2007). Simulating visibility during language comprehension. *Cognition, 105*, 229–238.

Yeh, C.-L., & Chen, Y.-C. (2007). Zero anaphora resolution in Chinese with shallow parsing. *Journal of Chinese Language and Computing, 17*(1), 41–56.

Zwaan, R. A. (1999). Embodied cognition, perceptual symbols, and situated models. *Discourse Processes, 28*, 81–88.

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 thought* (pp. 224–245). New York: Cambridge University Press.

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

Zwaan, R. A., Yaxley, C. J., & Aveyard, M. E. (2004). Moving words: Dynamic mental representations in language comprehension. *Cognitive Science, 28*, 611–619.