Research Report

When You and I Share Perspectives

Pronouns Modulate Perspective Taking During Narrative Comprehension

Tad T. Brunyé,[1,2] Tali Ditman,[2,3] Caroline R. Mahoney,[1,2] Jason S. Augustyn,[1] and Holly A. Taylor[2]

[1]U.S. Army Natick Soldier Research, Development and Engineering Center; [2]Tufts University; and [3]Massachusetts General Hospital/Harvard Medical School, Athinoula A. Martinos Center for Biomedical Imaging

ABSTRACT—*Readers mentally simulate the objects and events described in narratives. One common assumption is that readers mentally embody an actor's perspective; alternatively, readers might mentally simulate events from an external ''onlooker'' perspective. Two experiments examined the role of pronouns in modulating a reader's adopted perspective when comprehending simple event sentences. Experiment 1 demonstrated that readers embody an actor's perspective when the pronoun* you *or* I *is used, but take an external perspective when* he *is used. Experiment 2, however, found that a short discourse context preceding the event sentence led readers to adopt an external perspective with the pronoun* I*. These experiments demonstrate that pronoun variation and discourse context mediate the degree of embodiment experienced during narrative comprehension: In all cases, readers mentally simulate objects and events, but they embody an actor's perspective only when directly addressed as the subject of a sentence.*

> He lit a cigarette and tucked it between the remaining two fingers of his left hand, the one resting on the steering wheel. Keeping his black eyes on the road, he stooped forward, picked up the screwdriver lying between his feet, and handed it to me. I stuck it in the small hole in the door where the handle belonged and tucked it to roll down my window. (Hosseni, 2003, p. 228)

Tad T. Brunyé and Tali Ditman contributed equally to this work. Address correspondence to Tad T. Brunyé, U.S. Army NSRDEC, [redacted]

During reading, we encounter multiple characters with inherently different perspectives on described objects and events. Readers incorporate these pieces of information into their mental representation, or "situation model" (Zwaan & Radvansky, 1998). An emerging body of research suggests that readers embody described actions while incorporating them into a developing model; that is, readers mentally simulate a story's perceptual and motoric elements (Fischer & Zwaan, 2008; Glenberg, 2007). Inherent to this embodied approach is the assumption that described events are mentally performed from an actor's perspective (cf. Barsalou, 2008). However, it is unknown whether these mental simulations are consistently represented from a single character's perspective, or if readers dynamically adopt different perspectives. Indeed, readers of the above passage may take the first character's perspective and simulate finger movements with the cigarette and picking up the screwdriver from an internal embodied position. Alternatively, they may take an external perspective and simulate those same objects and events from an observer's position, perhaps imagining the man's actions from the passenger's seat; this explanation might be particularly parsimonious for narratives containing multiple character references. The present experiments were designed to evaluate the spatial perspectives characterizing the mental simulations readers develop while comprehending simple narrative events.

One theory is that readers simulate "the perspective on the situation that they would take if interacting with the object" (Barsalou, 2005, p. 643). This view implies that readers' mental simulations incorporate an actor's perspective, and this process is important for understanding described actions. Although there has been little work investigating perspective-taking during reading, there is some evidence that people adopt actors' perspectives when representing pictured events (e.g., Lozano,

Hard, & Tversky, 2007), and use deictic terms to differentially encode perspectives during reading (Black, Turner, & Bower, 1979). There is also evidence that objects associated with a protagonist's location are more accessible than those inconsistent with this location (e.g., Borghi, Glenberg, & Kaschak, 2004; Horton & Rapp, 2003; Morrow, Greenspan, & Bower, 1987). Additionally, readers seem to perform covert motor simulations of actions performed by a protagonist (Glenberg & Kaschak, 2002; Richardson, Spivey, McRae, & Barsalou, 2003). Finally, Tettamanti et al. (2005) found that listening to first-person action sentences activates the same neural motor circuits that subserve action execution. These studies provide preliminary support for the claim that readers perform mental simulations that spontaneously adopt a protagonist's perspective.

Alternatively, it is possible that the perspectives characterizing mental simulations may be contingent upon linguistic information within text (Bergen & Chang, 2005; Bergen, Lindsay, Matlock, & Narayanan, 2007; Black et al., 1979; Brunyé & Taylor, 2008). That is, even minor linguistic variations may be powerful mediators of the particular objects, actions, and perspectives comprising mental simulations. One potential linguistic mediator of perspective-taking during reading might be pronouns, as they explicitly identify the person performing an event. As in the above excerpt, a narrator might describe an action from his or her own perspective, the first person (*I stuck it in the small hole*), or describe a different character performing an action (*He lit a cigarette*; or even *You lit a cigarette*). Pronouns such as *I* or *you* may promote mental simulation from an internal (first-person) perspective, whereas third-person pronouns such as *he* may promote simulations from an external (third-person) perspective (see Ruby & Decety, 2001, for neuroimaging evidence that pronouns may modulate perspective). Although the previously mentioned studies may provide some support for the idea that readers embody an internal perspective during comprehension, no experiment to date has specifically examined whether manipulating pronouns can guide the perspectives characterizing mental simulations during reading. Two experiments tested the alternative possibilities that mental simulations during reading spontaneously take on actors' perspectives regardless of linguistic information and that perspectives characterizing mental simulations might be modulated by linguistic information such as pronouns. To investigate these alternatives, Experiment 1 used one-sentence event descriptions, and Experiment 2 used similar descriptions preceded by a larger discourse context.

EXPERIMENT 1

The first experiment manipulated whether first-, second-, or third-person pronouns were used in simple event sentences. Participants read about an event and then verified whether a displayed picture matched or mismatched the described event. These pictures were presented from an internal or external viewpoint; the internal viewpoint depicted events from an actor's perspective, whereas the external viewpoint depicted the same information from an observer's perspective. Critically, verification responses were orthogonal to the viewpoint; that is, participants need not monitor the picture viewpoint to perform the task. If readers automatically embody the perspective of an actor, then one would expect a faster response to internal relative to external perspective pictures following all three sentence types (*I*, *You*, or *He* sentences). However, if linguistic information influences the perspective adopted by a reader, then picture response times should be contingent upon the pronoun in the sentence.

Method

Participants and Design

Forty-eight native English speaking, right-handed Tufts University undergraduates (23 male, 25 female; mean age = 18.72 years) participated for monetary compensation. We used a 3 (description pronoun: *I*, *you*, *he*) × 2 (picture perspective: internal, external) repeated measures design to measure the influence of three pronouns (*I*, *you*, *he*) on the verification of two types of event pictures (internal, external).

Materials

Event Descriptions. Thirty-six (24 experimental, 12 practice) descriptions of simple events were constructed. Each description began with a pronoun subject (i.e., *I am*, *you are*, *he is*), continued with a verb (e.g., *slicing*, *taping*, *ironing*), and ended with a direct object (e.g., *the tomato*, *the package*, *the pants*, respectively).

Event Images. For each of the 24 experimental descriptions, four pictures were created by crossing whether the event was depicted from an internal or external perspective and whether it was or was not being performed (Fig. 1). Performing images depicted the event in midaction. Nonperforming images (depicting the objects but not the action) were included to equate the proportion of "yes" and "no" responses within each condition. All images depicted an approximate viewing distance of 40 in. at a 35° downward angle. Twelve additional images were similarly developed for the practice session, and 16 additional images served as filler events (never described; e.g., *breaking the pencil*).

Procedure

Participants read event descriptions and then verified whether a picture matched or mismatched the event described, regardless of the picture perspective. A practice session consisting of 12 trials confirmed that participants understood the instructions and did not simply match written and depicted perspectives. During practice, descriptions were presented in random order, each preceded by a 500-ms fixation cross and presented for 2 s (overall average presentation rate approximating 250 ms per word;

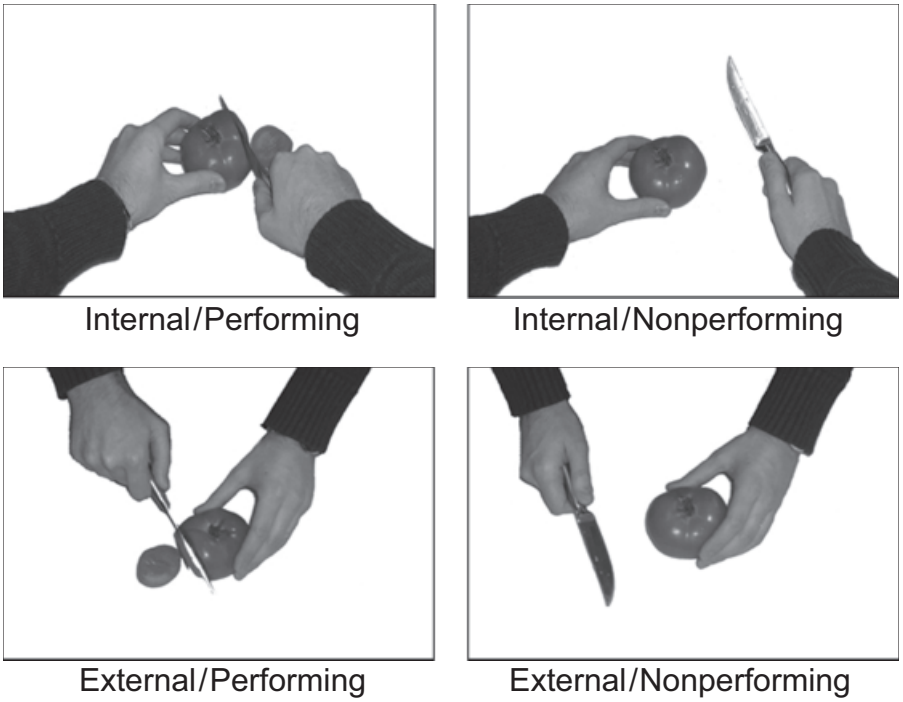

Fig. 1. Sample images (internal/external and performing/not performing) corresponding to Event Description 1 (slicing the tomato). Internal images were shot from a first-person perspective; external images were shot from a third-person perspective.

Rayner, 1998). An image was then presented in the center of the screen. Participants responded "yes" or "no" to each image, and trials timed out after 3 s without a response. Feedback was provided and the session was repeated until participants reached 100% accuracy (12 of 12 images correctly verified).

The experimental session followed the same procedures, but without feedback. Twenty-four trials were presented in random order, eight for each of the three description pronouns. Each participant saw 12 images from the internal and 12 from the external perspective. Of these 24 images, 18 depicted an event being performed and 6 depicted an event not being performed. Of the 18 images depicting an event being performed, 6 were fillers depicting never-described events. The fillers ensured that participants read the descriptions to perform image verifications rather than adopting a strategy of responding "yes" to all pictures depicting an action. Thus, half of the 24 trials were designed to elicit a "yes" response, and half were designed to elicit a "no" response. Across participants, descriptions were rotated through the three pronouns, and images were rotated through the four display types (internal, external, performed, not performed) in a Latin square design.

Results

Picture verification response times and accuracy for all correct *yes* trials were entered into 3 (description pronoun: first-person, second-person, third-person) × 2 (picture perspective: internal, external) repeated measures analyses of variance (ANOVAs). Response time data were $\log_{10}$-transformed, correcting for positive skew (Fisher, 1930; pretransform skewness = 1.614, posttransform skewness = 0.526). Analyses are reported by subjects (F_1, t_1) and items (F_2, t_2).

Response time analysis demonstrated no main effects of description pronoun ($F_1 < 1$, $F_2 < 1$), or picture perspective ($F_1 = 1.12$, $F_2 < 1$), but these two variables interacted, $F_1(2, 90) = 12.13$, $p_{rep} > .99$, $\eta^2 = .10$; $F_2(2, 46) = 4.83$, $p_{rep} = .95$, $\eta^2 = .17$ (see Fig. 2). Planned comparisons by subjects revealed faster verification of internal- relative to external-perspective pictures following the use of second-person pronouns, $t_1(46) = 3.09$, $p_{rep} = .99$, $d = 0.43$; $t_2(23) = 1.88$, $p_{rep} = .85$, $d = 0.38$, and first-person pronouns, $t_1(47) = 2.27$, $p_{rep} = .91$, $d = 0.24$, but this latter effect was not replicated in an items analysis ($t_2 < 1$). With third-person pronouns, we found faster verification of external- relative to internal-perspective pictures, $t_1(46) = 3.24$, $p_{rep} = .98$, $d = 0.48$; $t_2(23) = 1.94$, $p_{rep} = .86$, $d = 0.40$. Given that our pictures always depicted male hands, we entered male/female as a between-participants factor; these analyses confirmed that gender did not modulate any of our effects (all Fs < 1).

Accuracy was high overall ($M = .94$, $SE = .01$; see Table 1), demonstrating that participants understood the task. Accuracy did not vary as a function of description pronoun ($F_1 < 1$, $F_2 < 1$) or picture perspective ($F_1 = 2.88$, $F_2 < 1$); the two variables also did not interact ($F_1 < 1$, $F_2 < 1$), and there was no influence of gender (all Fs < 1).

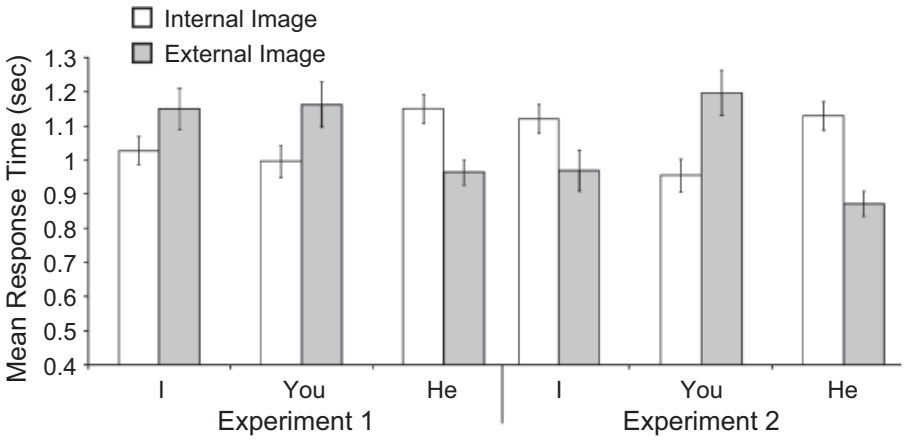

Fig. 2. Mean response times in Experiments 1 and 2, as a function of pronoun type (*I*, *you*, *he*) and image-verification type (internal, external). Internal images were shot from a first-person perspective; external images were shot from a third-person perspective. Error bars show standard errors.

EXPERIMENT 2

Experiment 1 demonstrated that readers do not always spontaneously mentally simulate described events from an actor's perspective; rather, they use linguistic cues to guide the nature of these simulations. A third-person pronoun cued an external perspective, away from the actor's perspective. As in previous studies using single-sentence stimuli (Borghi et al., 2004; Tettamanti et al., 2005), a second-person pronoun consistently (and first-person pronoun inconsistently) cued an embodied perspective.

Experiment 2 added short discourse contexts prior to event sentences to examine the possibility that more realistic and enriched descriptions of an actor may change the perspectives readers use while imagining events. If the partial lack of support for spontaneous embodiment in Experiment 1 (i.e., faster verification times to external- relative to internal-perspective pictures following *he*) was due only to ambiguity with regard to who "he" was, then one would expect that richer contexts would encourage readers to adopt a first-person perspective, leading to faster verification of pictures from the internal relative to external perspective.

TABLE 1

Mean Accuracy and Standard Errors for the Three Description Pronouns (*I, You, He*) and Two Picture Perspectives (Internal, External) in Experiments 1 and 2

	Picture perspective			
	Internal		External	
Pronoun	*M*	*SE*	*M*	*SE*
Experiment 1				
First person (*I*)	.92	.031	.96	.020
Second person (*you*)	.93	.026	.98	.015
Third person (*he*)	.91	.032	.94	.024
Experiment 2				
First person (*I*)	.92	.027	.94	.024
Second person (*you*)	.91	.028	.99	.010
Third person (*he*)	.95	.022	.94	.024

Method

Participants and Design

Forty-eight native-English-speaking right-handed Tufts University undergraduates (24 male, 24 female; mean age = 19.15 years) participated for monetary compensation; the design was identical to Experiment 1.

Event Descriptions

Two-sentence discourse contexts were developed for each event sentence from Experiment 1. The first sentence always provided descriptive information about the character (i.e., *[X] is a [Y]-year-old [Z]*; e.g., *I am a 30-year-old deli employee*), the second sentence reiterated *[X]* and stated a *[Z]*-appropriate goal (e.g., *I am making a vegetable wrap*), and the third sentence matched Experiment 1 with a consistent temporal marker (e.g., *Right now, I am slicing the tomato*).

Procedure

The procedure matched that used in Experiment 1 with the addition of two sentences, which were each presented for 3 s immediately before the event sentence (approximating a presentation rate of 250 ms per word, as in Experiment 1).

Results

Analyses were done as in Experiment 1; response time data were $\log_{10}$-transformed (pretransform skewness = 1.274, posttransform skewness = 0.508).

Response time analysis demonstrated no effect of description pronoun ($F_1 = 2.09$, $F_2 < 1$), but an effect of picture perspective, $F_1(2, 94) = 8.76$, $p_{rep} = .97$, $\eta^2 = .02$, showed faster

responses to external relative to internal pictures; this latter effect did not replicate in an items analysis ($F_2 = 1.21$). Description pronoun and picture perspective interacted, $F_1(1, 47) = 27.87, p_{rep} > .99, \eta^2 = .18$; $F_2(2, 46) = 18.59, p_{rep} > .99, \eta^2 = .45$ (see Fig. 2). As in Experiment 1, results revealed faster verification of internal- relative to external-perspective pictures following second-person pronouns (*You are . . .*), $t_1(47) = 3.95$, $p_{rep} > .99$, $d = 0.65$; $t_2(23) = 3.45$, $p_{rep} = .98$, $d = 0.71$, and faster verification of external- relative to internal-perspective pictures following third-person pronouns (*He is . . .*), $t_1(47) = 6.14, p_{rep} > .99, d = 0.84$; $t_2(23) = 3.35, p_{rep} = .97, d = 0.69$. In contrast to Experiment 1, participants responded faster to external- relative to internal-perspective pictures following first-person (*I am . . .*) pronouns, $t_1(47) = 4.47, p_{rep} > .99, d = 0.50$; $t_2(23) = 2.98$, $p_{rep} = .96$, $d = 0.61$. There were no effects of gender (all Fs < 1).

Accuracy was high overall ($M = .94$, $SE = .01$; see Table 1), and did not vary as a function of description pronoun ($F_1 < 1$, $F_2 < 1$) or picture perspective ($F_1 = 2.12$, $F_2 = 1.08$); the two variables also did not interact ($F_1 = 2.40$, $F_2 = 2.14$). There were no effects of gender (all Fs < 1).

GENERAL DISCUSSION

The primary difference between Experiments 1 and 2 is that response times for internal- versus external-perspective pictures reversed following first-person pronouns. With single sentences, such as *I am slicing the tomato*, the ambiguity with regard to the actor seems to lead readers to adopt internal perspectives on described events. This result supports earlier neuroimaging work using first-person action sentences (e.g., Tettamanti et al., 2005). However, when character identity is explicitly revealed through an extended discourse we find that readers are more likely to adopt an external perspective following first-person pronouns. It could be the case that reiterating pronouns in extended discourse helps readers disambiguate the actor from the observer and encourages them to play a role as one or the other. In any case, the present results suggest that to imagine oneself in "someone else's shoes" during narrative comprehension, the reader must be directly addressed as the subject of the sentence (see also Borghi et al., 2004). We continue to investigate the influence of pronouns on perspective-taking in relatively naturalistic reading environments.

These results are consistent with a situation-model framework (Zwaan & Radvansky, 1998) in demonstrating that comprehension goes beyond the lexico-semantic information inherent to the individual words making up a sentence or discourse, extending to mental representation of the described events. Embodied theories of language comprehension emphasize that an important component of rich situation models is the mental simulation of the text's perceptual and motoric elements (e.g., Fischer & Zwaan, 2008; Glenberg, 2007; Yaxley & Zwaan, 2007). The present work supports this notion and demonstrates that readers do not always assume the perspective of the person performing an event. In fact, when a realistic discourse was used in Experiment 2, we only found evidence for embodied language comprehension with the second-person pronoun (*You are . . .*). We suggest that the palpable sense of "being there" during reading does, in fact, involve the perceptual and motoric mental simulation of described discourse elements, but that actors' perspectives are not automatically embodied. In closing, it seems to be the case that embodying an actor's perspective is the exception, rather than the rule, during discourse comprehension; indeed, pronouns are one linguistic form that plays a powerful role in determining the perspectives characterizing mental simulations (cf. Bergen & Chang, 2005; Herman, 2002).

REFERENCES

Barsalou, L.W. (2005). Situated conceptualization. In H. Cohen & C. Lefebvre (Eds.), *Handbook of categorization in cognitive science* (pp. 619–650). St. Louis, MO: Elsevier.

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

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: 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.

Black, J.B., Turner, T.J., & Bower, G.H. (1979). Point of view in narrative comprehension, memory, and production. *Journal of Verbal Learning and Verbal Behavior*, *18*, 187–198.

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

Brunyé, T.T., & Taylor, H.A. (2008). Working memory in developing and applying mental models from spatial descriptions. *Journal of Memory and Language*, *58*, 701–729.

Fischer, M.H., & Zwaan, R.A. (2008). Embodied language: A review of the role of the motor system in language comprehension. *Quarterly Journal of Experimental Psychology*, *61*, 825–850.

Fisher, R.A. (1930). The moments of the distribution for normal samples of measures of departure from normality. *Proceedings of the Royal Society*, *130*, 16–28.

Glenberg, A.M. (2007). Language and action: Creating sensible combinations of ideas. In G. Gaskell (Ed.), *The Oxford handbook of psycholinguistics* (pp. 361–370). Oxford, England: Oxford University Press.

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

Herman, D. (2002). *Story logic: Problems and possibilities of narrative*. Lincoln: University of Nebraska Press.

Horton, W.S., & Rapp, D.N. (2003). Out of sight, out of mind: Occlusion and the accessibility of information in narrative comprehension. *Psychonomic Bulletin & Review*, *10*, 104–110.

Hosseni, K. (2003). *The kite runner*. New York: Riverhead Books.

Lozano, S.C., Hard, B.M., & Tversky, B. (2007). Putting action in perspective. *Cognition*, *103*, 480–490.

Morrow, D.G., Greenspan, S.L., & Bower, G.H. (1987). Accessibility and situation model narrative comprehension. *Journal of Memory and Language*, *26*, 165–187.

Rayner, K. (1998). Eye movements in reading and information processing: Twenty years of research. *Psychological Bulletin*, *124*, 372–422.

Richardson, D.C., Spivey, M.J., McRae, K., & Barsalou, L.W. (2003). Spatial representations activated during real-time comprehension of verbs. *Cognitive Science*, *27*, 767–780.

Ruby, P., & Decety, J. (2001). Effect of subjective perspective taking during simulation of action: A PET investigation of agency. *Nature Neuroscience*, *4*, 546–550.

Tettamanti, M., Buccino, G., Saccuman, M.C., Gallese, V., Danna, M., Scifo, P., et al. (2005). Listening to action-related sentences activates fronto-parietal motor circuits. *Journal of Cognitive Neuroscience*, *17*, 273–281.

Yaxley, R.H., & Zwaan, R.A. (2007). Simulating visibility during language comprehension. *Cognition*, *105*, 229–236.

Zwaan, R.A., & Radvansky, G.A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin*, *123*, 162–185.

(Received 4/24/08; Revision accepted 7/18/08)