

Use of colour for language processing during reading

Albrecht W. Inhoff¹, Bradley A. Seymour¹, and
Ralph Radach²

¹Department of Psychology, Binghamton University, Binghamton,
NY, USA

²Department of General and Biological Psychology, Wuppertal
University, Wuppertal, Germany

The study examined whether literal correspondence is necessary for the use of visual features during word recognition and text comprehension. Eye movements were recorded during reading and used to change the colour of dialogue when it was fixated. In symbolically congruent colour conditions, dialogue of female and male characters was shown in orchid and blue, respectively. The reversed assignment was used in incongruent conditions, and no colouring was applied in a control condition. Analyses of oculomotor activity revealed Stroop-type congruency effects during dialogue reading, with shorter viewing durations in congruent than incongruent conditions. Colour influenced oculomotor measures that index the recognition and integration of words, indicating that it influenced multiple stages of language processing.

Keywords: Colour; Eye movements; Reading; Visual imagery; Visual symbols.

Recent work has shown a close linkage between sensory perception and the representation of linguistic content (see Barsalou, 1999; Fischer & Zwaan, 2008; Zwaan & Radvansky, 1998, for reviews). The gender of an auditorily perceived spoken sentence endowed, for instance, the semantic representation of gender-neutral constituent words with male or female attributes (Geiselman, 1979; Geiselman & Bellezza, 1977), and it generated role-specific expectancies for sentence content (Van Berkum, van den Brink, Tesink, Kos, & Hagoort, 2008). Similarly, comprehension of sentence content can include representation of matching visual features. After

Please address all correspondence to Albrecht Inhoff, Department of Psychology, Binghamton University, Binghamton, NY 13902, USA. E-mail: inhoff@binghamton.edu

We thank Meghan Dembrosky, Kaitlyn Pittman, Chris Cotrone, and Vanessa Kay for their help with stimulus preparation and data collection, and the reviewers for insightful comments.

© 2012 Psychology Press, an imprint of the Taylor & Francis Group, an Informa business
<http://www.psypress.com/viscog> <http://dx.doi.org/10.1080/13506285.2012.745917>

sentences with object-specific content were read in Zwaan, Stanfield, and Yaxley's (2002) study (e.g., "The ranger saw the eagle in the sky"), participants distinguished a picture of the sentence object more effectively from a novel object when the depicted shape matched the implied shape of the sentence object (e.g., an eagle with outstretched wings vs. an eagle sitting in a tree). Other studies showed that classification of visual objects with a particular spatial orientation, location, or shape was influenced by previously read sentence content that expressed a related spatial orientation (e.g., Richardson, Spivey, Barsalou, & McRae, 2003; Yaxley & Zwaan, 2007), shape (Zwaan et al., 2002), or location (Bergen, Lindsay, Matlock, & Narayanan, 2007; Yaxley & Zwaan, 2007). Together, these studies show that language representation and perception are closely linked when there is a direct, literal correspondence.

In the absence of a literal relationship, perception may not interact with language representation. In Bergen et al. (2007), metaphoric sentence content (e.g., "the stock market soared") did not influence performance in a subsequent spatial detection task, presumably because there was no direct, literal correspondence between the representation of sentence meaning and visuospatial features. However, contrary to the literal-relationship assumption, nonliteral visual features influenced the representation of meaning when the experimental task involved the processing of isolated words (Maier, Robinson, & Clore, 2004; Sherman & Clore, 2009; Weger, Maier, Robinson, & Inhoff, 2007). In Weger et al. (2007), performance on a visuospatial classification task was modulated by the affective content of a previously viewed word. Visual targets appearing in the upper half of a monitor screen were classified more effectively after the recognition of words with a positive affective connotation (e.g., "kind"), whereas targets appearing in the lower half were classified more effectively after the processing of negative words ("hostile"). Similarly, Sherman and Clore's (2009) Stroop-type task showed that individually presented words with positive and negative moral value (e.g., "honesty" and "greed") were classified more effectively when shown in white and black, respectively.

The current study examined whether literal correspondence is necessary for the use of visual features during language processing. Similar to Sherman and Clore, word colour was manipulated to obtain Stroop-type congruency effects when individual words were viewed. However, instead of the classification of individually presented words, the experimental task required sentence comprehension, similar to prior studies of visual imagery for sentence content (e.g., Bergen et al., 2007; Zwaan et al., 2002). Other novel aspects of the work were that eye movements—an integral part of normal reading—rather than extraneous manual responses were used to assess

language processing, and that different oculomotor measures were computed to determine the use of symbolic colour during the recognition and sentence integration of words.

In the experiment, participants read a play with male–female dialogue, shown in blue or orchid, respectively, to create symbolically congruent colour conditions (Picariello, Greenberg, & Pillemer, 1990; Pomerleau, Bolduc, Malcuit, & Cossette, 1990). The colouring of dialogue was reversed to create incongruent conditions, and no distinct colouring was applied in a control condition. Stroop-type congruency effects should emerge, if the (symbolic) meaning of colour influenced word recognition and/or the integration of words into sentence context.

METHOD

Participants

Eighteen students at Binghamton University participated in the study. All were native speakers of English and naïve regarding the purpose of the experiment.

Apparatus

Text was shown on an Iiyama Vision Master Pro 514 CRT monitor with a resolution of 1024×768 pixels. Characters were shown in 12 point Courier font in black against a light grey background, with each character occupying a 15 (vertical) \times 10 (horizontal) pixel area. Head position was fixed at a distance of 65 cm, and each character subtended approximately 0.3 degrees of visual angle. An Eye-link 1000 tracker was used to monitor eye movements at a rate of 1000 Hz and a tracking error of 0.1° degrees or less.

Materials

The materials consisted of a play, written in English, with three acts and a total of 1843 words. The first act developed the plot and involved direct spoken exchanges between a father and his daughter. The second act elaborated the plot primarily through telephone-based dialogue between the daughter and a male friend and direct dialogue between the male friend and his mother, and the third act closed with a humorous resolution of a puzzle through direct and telephone-based dialogue.

The text of individual trials extended across one to three lines. Line 1 contained 10 to 16 words, and Lines 2 and 3 generally contained fewer words

... one block (dialogue changing to orchid)

Kelly: "Of course not. It was probably just a prank. My guy friends at school do stuff like that." She was curious.

Kelly's Dad: "We should look into this. As a single dad, I have to do twice the parenting." He frowned, looking down.

... another block (dialogue changing to blue)

Kelly: "Really, it's not that serious." She got up from the table and headed back to her room.

Kelly's Dad: "It used to be easier raising kids when I was young", he mumbled to himself.

Figure 1. A colour-coded dialogue. To view this figure in colour, please see the online issue of the Journal.

that wrapped up a trial. Each trial started with a one- or two-word identifier that specified the character's name or a gender-specific role (father, mother). This was followed by dialogue spoken by the character and a descriptor segment that typically described an action (see Figure 1). On average, dialogue contained 9.5 words for male and female characters, and the descriptor segment contained 6 words. All lines were presented near the vertical screen centre and words were left-adjusted with a margin width of five characters. Character gender alternated across successive trials, so that each act (block) contained an approximately equal number of female and male dialogue.

The beginning and closing acts contained 27 trials, the more complex second act contained 42 trials. The mean log frequency of words was 3.8 ($SD = 4.0$) (Brysbaert & New, 2009), and mean word length was 4.2 letters ($SD = 2.0$). The two indexes were negatively correlated, $r = -.47$, which is typical for natural text.

Procedure

To exclude individuals with poor colour perception and with atypical or weak gender-specific colour schemes,¹ the experiment started with the colour rating of 10 words, six that denoted objects with unambiguous colours (e.g., “lemon”), two with ambiguous colours (“dress”, “car”), and a male and a female name. Individuals were asked to select one of six colours for each word, and to rate the confidence on a 7-point scale. To be included, individuals had to select correct colour preferences for unambiguous objects, and they had to associate blue and orchid with the male and female name, respectively, with a confidence of >3.5 .

Eyetracker calibration was followed by the presentation of 10 practice sentences. Eye-movement-contingent display changes were used to change the colour of the dialogue segment when the eyes moved from the gender identifier onto dialogue (see Figure 2). Participants were told that the colour of some text segments would be changed during reading, and that the study sought to determine the ease with which coloured text was read. None of the participants reported being aware of the symbolic colour value after the experiment was completed, although pretesting could have influenced the use of colour.

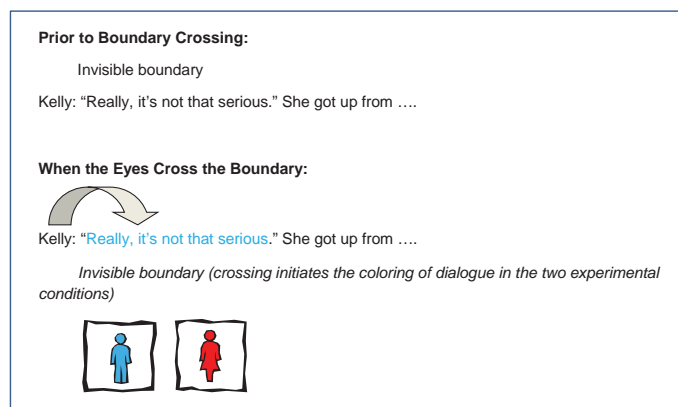


Figure 2. The contingent speech technique. An eye movement onto dialogue initiated a change of colour in the experimental conditions. To view this figure in colour, please see the online issue of the Journal.

¹The names were included to select individuals with internalized gender-specific colour schemes, as these schemes are arbitrary, subject to cultural differences, and amenable to change. At the beginning of the twentieth century, pink was considered acceptable for boys and blue for girls. “The reason is that pink being a more decided and stronger colour is more suitable for the boy, while blue, which is more delicate and dainty, is prettier for the girl” (*Ladies Home Journal*, June 1918).

The experimental manipulation involved the colour coding of dialogue during the reading of two of the three acts/blocks of the play. Throughout the reading of a complete block, all dialogue was shown in just one colour: Orchid, blue, or black (no colour change occurred in the “black” baseline condition), and the colour-to-block assignment was counterbalanced over subjects. Since male- and female-character dialogue alternated within a block, approximately half of the trials of colour-coded blocks were thus either gender congruent or gender incongruent. Reading of the play was followed by a questionnaire that probed memory for content with 10 multiple-choice questions. Across readers, accuracy was 80% or higher.

Data selection and analyses

Fewer than 1% of the trials were removed because of track losses, and this left 16,223 words of dialogue and 9542 descriptor words for the analyses of fixation probabilities (skipping rates). Three viewing measures were computed for words that received at least one fixation between 75 ms and 800 ms ($n = 13,324$): (1) The duration of the first fixation (FFD), (2) cumulated viewing duration until another word was viewed (first pass gaze duration, GD), and (3) repair time. First fixation duration and gaze duration are generally assumed to index the ease of visual word identification (Rayner, 1998). Repair time (RPT) was defined as the interval between the onset of the first fixation on a word and the offset of its last fixation. That is, the interval included not only time on the word but also the time off a word when it was reread. If, for instance, a reader viewed Word A during first pass reading, and then reread it after having viewed Words B, C, and D, then A's RPT would include the time viewing A[1], B, C, D, and A[2]. RPT sought to capture postrecognition processing, and it was based on the assumptions that a word is recognized during first pass reading, that subsequent rereading is associated with a sentence integration difficulty (Reichle, Warren, & McConnell, 2009), and that a word's representation maintained a distinct status in working memory until the integration difficulty was resolved through rereading (see also Daneman & Stainton, 1993).

All data were analysed using linear mixed models (LMM), as implemented in the lme4 package of the R system for statistical computing (Bates & Maechler, 2011; R Core Team, 2012). The model included two fixed factors, dialogue colour (orchid, blue, black) and talker (male vs. female), and two crossed random factors, subjects and words. Helmert contrast coding was used to establish two orthogonal colour contrasts. A *baseline* contrast that compared the baseline [black, i.e., no colour change] with the mean of the two colour conditions, and a *colour* contrast, that compared the two colours. The frequency distributions of the three viewing duration measures were positively skewed, and this was corrected

though log transformations; LMM effect sizes (b), and standard errors (SE) are thus reported in log units for these data. The significance of pairwise numeric contrasts was determined with the `pvals` function from the `languageR` package (Baayen, 2008). Skipping rates were analysed using a binomial link function, and effect sizes and SE s are reported in logits.

RESULTS

Dialogue reading

Skipping rates are shown as a function of dialogue colour and story gender in Table 1. Neither colour nor talker gender influenced skipping rate, and the critical interaction of dialogue colour with talker gender was negligible (all p -values $>.35$). However, all three viewing duration measures, shown in Table 1 and Figure 3, yielded robust Stroop-type congruency effects, as readers spent less time viewing words of dialogue when colour and talker gender matched than when they mismatched: FFD, $b = -.104$, $SE = 0.016$, $t = -6.53$; GD, $b = -.10$, $SE = 0.019$, $t = -5.17$; RPT, $b = -.16$, $SE = 0.037$, $t = -4.27$, all $ps < .001$. There was also a robust baseline–colour effect, with shorter durations when dialogue remained black than when it was coloured: FFD, $b = -.055$, $SE = 0.0071$, $t = -7.75$; GD, $b = -.051$, $SE = 0.0085$, $t = -6.02$; RPT, $b = -.11$, $SE = 0.016$, $t = -6.47$, all $ps < .001$. No other effect approached significance (all $ps > .15$).

Dialogue colour could have interacted with talker gender because it influenced the attribution of a to-be-read dialogue segment to a particular character or because it influenced the ease with which dialogue content was comprehended. The two accounts make somewhat different predictions regarding the time course of colour usage. The instantiation of a pertinent character should occur at the very onset of dialogue reading; comprehension should occur throughout dialogue reading.

To distinguish these accounts, the words of a dialogue segment were separated into three approximately equal-size groups comprising beginning, middle, and ending words,² and a supplementary analysis with the factors colour, talker, and position (beginning, middle, end) was applied to the gaze data. Colour congruency influenced the reading of all three dialogue segments (Figure 4), and the significant two-way interaction of colour

²Position effects were applied to Line 1. Fewer than half of the dialogue segments extended into Line 2, and Line 2 segments were relatively small.

TABLE 1
Oculomotor activity during dialogue and descriptor reading as a function of dialogue colour and talker gender

Dialogue colour	Female talker			Male talker		
	Black	Orchid	Blue	Black	Orchid	Blue
Skipping rate						
Dialogue	0.44 (0.5)	0.43 (0.5)	0.42 (0.5)	0.44 (0.5)	0.42 (0.49)	0.43(0.5)
Description	0.46 (0.5)	0.44 (0.5)	0.44 (0.5)	0.49 (0.5)	0.46 (0.5)	0.47 (0.5)
First fixation duration (in ms)						
Dialogue	205 (68)	217 (72)	218 (72)	205 (66)	226 (72)	210 (70)
Description	213 (74)	218 (77)	213 (75)	222 (78)	228 (77)	216 (79)
Gaze duration						
Dialogue	226 (97)	240 (105)	241 (104)	229 (98)	249 (105)	232 (104)
Description	236 (107)	248 (118)	247 (119)	248 (109)	259 (115)	243 (114)
Repair time (in ms)						
Dialogue	349 (455)	399 (554)	440 (672)	395 (609)	465 (710)	391 (525)
Description	403 (507)	430 (462)	386 (581)	393 (430)	393 (475)	425 (539)

Mean values were computed from nontransformed data. Standard deviations for these data are shown in parentheses.

with talker gender: GD, $b = -.09$, $SE = 0.021$, $t = -4.31$, $p < .001$, was not modulated by position ($p > .3$).

Descriptor reading

Skipping rates did not show a robust effect pattern (all $ps > .2$), and there was no consistent pattern across viewing durations. All three measures showed longer durations for male than female descriptor segments, although the effect was not reliable for RPT: FFD and gaze, $b = .031$, $SE = 0.011$, $t = 2.81$, and $b = .033$, $SE = 0.015$, $t = 2.22$, both $ps < .05$; RPT, $b = .03$, $SE = 0.025$, $t = 1.25$, $p > .2$. All three measures also yielded numerically shorter viewing durations after the reading of blue- than orchid-coloured dialogue, but now only the RPT difference was reliable: FFD, $b = -.02$, $SE = 0.011$, $t = -1.83$, $p < .1$; GD, $b = -.018$, $SE = 0.014$, $t = -1.32$, $p > .15$; RPT, $b = -.034$, $SE = 0.016$, $t = -2.21$, $p < .05$. Numeric—though not significant—congruency effects that emerged for first fixation duration and gaze duration are also inconclusive, as they could have been compromised by subtle differences between male and female descriptor text and effects of prior dialogue colour: FFD, $b = -.029$, $SE = 0.021$, $t = -1.32$, $p > .15$; GD, $b = -.044$, $SE = 0.026$, $t = -1.66$, $p < .1$; RPT, $b = -.038$, $SE = 0.046$, $t = -1.04$, $p > .2$.

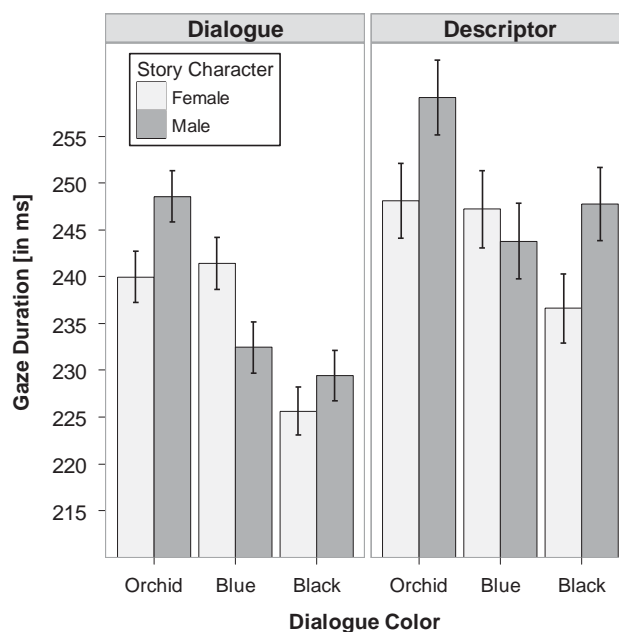


Figure 3. Nontransformed gaze duration as a function of colour and talker gender. Standard errors were computed from LMM residuals.

DISCUSSION

The current study revealed robust Stroop-type colour congruency effects during dialogue reading. First fixation duration, gaze duration, and RPT were shorter when the symbolic colour of fixated dialogue and the gender of the talking character matched than when they mismatched. These findings extend prior work in two theoretically significant aspects: The nonliteral meaning of colour was automatically used for the representation of language meaning, and colour congruency effects emerged in oculomotor measures that indexed the recognition and integration of words. This implies that literal correspondence is not a necessary condition for the use of visual features during language representation.

When dialogue was read, the viewing and recognition of a word entailed perception of colour. Effects of colour congruency on all three dialogue viewing duration measures indicate that perceived colour assumed a symbolic function and that it endowed word and sentence meaning with gender-specific properties. Auditory and visual images for spoken dialogue could, for instance, have had a higher pitch and less volume—and the talker could have been represented as more delicate—when dialogue was coloured orchid than when it was coloured blue.

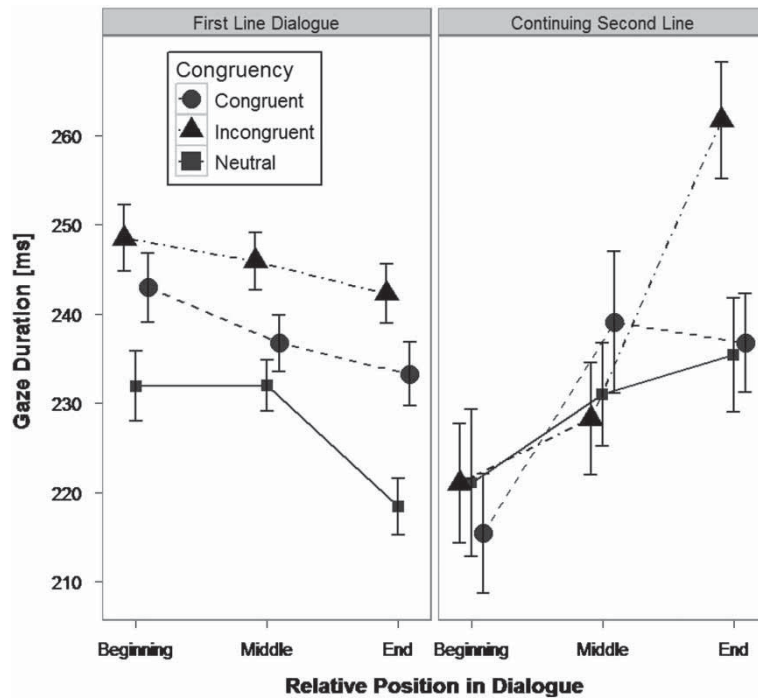


Figure 4. Gaze duration for the beginning, middle, and ending line dialogue segment as a function of colour congruency. Standard errors were computed from LMM residuals.

A subset of dialogue words was reread ($n=1190$), presumably due to integration difficulties. In these cases, RPT included first pass gaze duration, i.e., time that is attributed to word identification, time spent viewing other words, and rereading time, i.e., time components that can be attributed to sentence integration. To distinguish effects of colour congruency on recognition and integration, first pass gaze duration and the remaining “integration time” (RPT minus gaze duration) were analysed separately. The results revealed reliable congruency effects for both indexes, $b = -.10$, $SE = 0.05$, $SE = -1.98$, and $b = -.19$, $SE = 0.09$, $SE = -2.01$, respectively, both $ps < .05$, thus indicating that the symbolic value of colour influenced word- and sentence-level processing.

Viewing durations in the congruent conditions were not shorter, however, than viewing durations for the baseline. The baseline was used to determine dialogue reading in the absence of a colour change, as such changes could have attracted attention. Black also yielded a higher contrast against a light grey background than colour, and reduced text-to-background contrast has been shown to increase viewing durations (Reingold & Rayner, 2006).

Consistent with this, all viewing duration measures revealed longer durations for coloured than for black dialogue. Potential benefits conveyed by congruency could thus have been offset by other factors. Critically, the congruency effect during dialogue reading was not influenced by colour and talker-gender differences. The two factors yielded highly reliable crossover interactions and negligible main effects for all three measures of dialogue viewing duration.

However, two of the viewing duration measures, first fixation duration and gaze duration, yielded negligible congruency effects for female talkers, although both genders yielded sizable congruency effects with RPT. A broader range of colours might thus be concordant with the female gender, and the symbolic value of colour may have contributed to the representation of female-character dialogue only when integration difficulties were encountered. Even with this caveat, the current results have clear implications: Nonliteral visual features automatically contributed to multiple stages of language processing, and this occurred under relatively natural task conditions.

REFERENCES

- Baayen, R. H. (2011). *languageR: Data sets and functions with "Analyzing linguistic data: A practical introduction to statistics"* (R package version 1.1) [Computer software]. R Foundation for Statistical Computing, Vienna, Austria.
- Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, 22(4), 577–660.
- Bates, D., & Maechler, M. (2011). *lme4: linear mixed-effects models using Eigen and syntax* (R package version 0.999375-39) [Computer software]. R Foundation for Statistical Computing, Vienna, Austria. Available from <http://www.r-project.org>
- Bergen, B., Lindsay, S., Matlock, T., & Narayanan, S. (2007). Spatial and linguistic aspects of visual imagery in sentence comprehension. *Cognitive Science*, 31, 733–764.
- Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, 41, 977–990.
- Daneman, M., & Stainton, M. (1991). Phonological recoding in silent reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 17, 618–632.
- 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.
- Geiselman, R. E. (1979). Inhibition of the automatic storage of speaker's voice. *Memory and Cognition*, 7, 201–204.
- Geiselman, R. E., & Belezza, F. (1977). Incidental retention of speaker's voice. *Memory and Cognition*, 5, 658–665.
- Meier, B. P., Robinson, M. D., & Clore, G. L. (2004). Why good guys wear white: Automatic inferences about stimulus valence based on brightness. *Psychological Science*, 15, 82–87.
- Picariello, M. L., Greenberg, D. N., & Pillemer, D. B. (1990). Children's sex-related stereotyping of colors. *Child Development*, 61, 1453–1460.

- Pomerleau, A., Bolduc, D., Malcuit, G., & Cossette, L. (1990). Pink or blue: Environmental gender stereotypes in the first two years of life. *Sex Roles, 22*, 359–367.
- Rayner, K. (1998). Eye movements in reading and information processing. *Psychological Bulletin, 124*, 372–422.
- R Core Team. (2012). A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria. ISBN 3-900051-07-0. <http://www.R-Project.org>.
- Reichle, E. D., Warren, T., & McConnell, K. (2009). Using E-Z Reader to model the effects of higher-level language processing on eye movements during reading. *Psychonomic Bulletin and Review, 16*, 1–21.
- Reingold, E. M., & Rayner, K. (2006). Examining the word identification stages hypothesized by the E-Z Reader model. *Psychological Science, 17*, 742–746.
- Richardson, D. C., Spivey, M. J., Barsalou, L. W., & McRae, K. (2003). Spatial representations activated during real-time comprehension of verbs. *Cognitive Science, 27*, 767–780.
- Sherman, G. D., & Clore, G. L. (2009). The color of sin: White and black are perceptual symbols of moral purity and pollution. *Psychological Science, 20*, 1019–1025.
- Van Berkum, J. J. A., van den Brink, D., Tesink, C. M., Kos, M., & Hagoort, P. (2008). The neural integration of speaker and message. *Journal of Cognitive Neuroscience, 20*, 580–591.
- Weger, U. W., Meier, B. P., Robinson, M. D., & Inhoff, A. W. (2007). Things are sounding up: Affective influences on auditory tone perception. *Psychonomic Bulletin and Review, 14*, 517–521.
- 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 working memory. *Psychological Bulletin, 123*, 162–185.
- Zwaan, R. A., Stanfield, R. A., & Yaxley, R. H. (2002). Language comprehenders mentally represent the shapes of objects. *Psychological Science, 13*, 168–171.

Manuscript received May 2012
Manuscript accepted October 2012
First published online December 2012