Theory of Mind and embedding of perspective

A psychological test of a literary "sweet spot"

D. H. Whalen, Lisa Zunshine and Michael Holquist
Graduate Center, City University of New York and Haskins Laboratories/University of Kentucky, Lexington/Yale University and Columbia University

Theory of Mind (ToM) has been proposed to explain social interactions, with real people but also with fictional characters, by interpreting their mind as well as our own. "Perspective embedding" exploits ToM by placing events in characters' minds (e.g., "he remembered she was home"). Three levels of embedment, common in literature, may be a "sweet spot" that provides enough information about a character's motivation, but not a confusing over-abundance. Here, we use short vignettes with 1 or 3 characters and 0–5 levels of perspective embedding in two reading studies to see whether these preferences might be related to processing ease. Self-paced readers were fastest with one level of embedment, increasingly slower as embedment increased; vignettes without embedment were approximately as slow as level 4. With both self-paced and imposed timing, error rates on probe questions increased only at the fifth level. Readers seem to prefer literary texts in which ToM operations are obvious due to embedding of perspectives within the narrative but still somewhat challenging.

Keywords: theory of mind, reading, perspective embedding, literature

Why do we read literature? Part of the reason seems to be that we are biologically predisposed to interpret the intent of others, even those who exist only in fiction. In order to participate fully in human society, we must all be "mind readers" (e.g., Atran, 2002; Boyer, 2001; Carruthers & Smith, 1996; Singer, 2006): We have to know what intentions others have, or we will not be able to understand their actions, especially those that involve us. We need to have a "Theory of Mind" (ToM) about other people so that their actions make sense in terms of the mental states (i.e., thoughts, feelings, and goals) that we can reasonably attribute to them. The terms "Theory of Mind" and "mind reading" are not ideal; the latter is particularly inapt. Given how many of our attributions and interpretations of thoughts, feelings, and intentions are wrong or only approximately correct, they might as well

Scientific Study of Literature 2:2 (2012), 301–315. DOI 10.1075/ssol.2.2.06wha
ISSN 2210–4372 / E-ISSN 2210–4380

call it "mind misreading." But since evolution doesn't deal in perfection, this less than perfect "reading of minds" allows us to function as the highly social creatures we are.

Having a ToM allows us to attribute mental states to fictional characters as well as to people in our immediate presence. This feature is related to displacement: the use of language to refer to events not immediately present, that is, either at a distance or in a different time. This feature is common to all human languages (Hockett, 1960) (although it has controversially been claimed not to be universal: Everett, 2005). Displacement that allows us to know what others in our everyday lives intend and therefore to relate positively or negatively to them is the same ability that allows us to react to people we have only heard about. Once we make such a transition, i.e., to remove those "people" from the immediately present world, we can begin to empathize with, and even care deeply about, people who exist only in literary texts. ToM has been shown to activate areas of the brain that are also active in story comprehension (Mar, 2011). This has been found for nonverbal as well as more typical verbal stories (Gallagher, et al., 2000). Such processes may have substantial overlap with other activities, such as self projection or navigation (e.g., Spreng, Mar, & Kim, 2009), but the patterns of activation do provide us with a basis for exploring these processes further.

Works of fiction exhibit "sociocognitive complexity," which has been defined as the depiction of a mental state embedded within another mental state (Zunshine, 2011). (This term has also been used in management science (e.g., Ginsberg, 1990), but Zunshine takes a different perspective.) "I am sad" is less sociocognitively complex than "He knew she was sad," which in turn is less complex than "Surprisingly, he knew that she was sad," because "surprisingly" implies someone else's mind — perhaps the narrator's? — contemplating a mental state of one character who is aware of the mental state of another character. A succession of scenes featuring third-level complexity — a mind within a mind within a mind, as in the above case of "Surprisingly, he knew ..." — we speculate is the baseline for most fiction. It seems that some authors/genres/works (Proust, Henry James' later works) routinely operate on the fourth level, and some reach to the fifth and even sixth levels. In contrast, encyclopedia entries never rise to the third level, unless they deal with subjects that come with their own higher sociocognitive complexity (e.g., a Wikipedia entry featuring the plot synopsis of a novel or a movie) (Zunshine, 2011: 119–20).

The proliferation of persons (including fictional characters) in our thinking leads to another aspect of cognition, perspective embedding, or "embedment." The term embedment emphasizes that perspective embedding is distinct from embedding one narrative within another, or narrative levels (e.g., Herman, 2006). Language allows us to describe facts, but it further allows us to describe the mental states of actors associated with those facts. "I went to the store" is simply a fact (or, at least, a putative fact), but "I remembered I went to the store" is embedded by a level of perspective. "I rememberd" adds a mental state to the fact, namely that I remember its occurrence. "Jane knows I remember that I went to the store" has two levels of embedment, plus a new actor, Jane. Embedments can, in theory, continue indefinitely. In practice, however, when they exceed five in written or spoken communication, they tend to be incomprehensible (Dunbar, 1996, 2000). When embedment is achieved primarily by means of body language and unconscious inference as opposed to explicitly articulated propositions (e.g., "she wants him to know that she believes X"), the effects of increasing the level of such "metarepresentational sophistication" are somewhat less clear (Sperber, 2000).

Similarly, it is important to note that perspective embedding (or just embedment, in this article) is not the same as syntactic embedding (e.g., Chomsky, 1965). Syntactic embedding can occur without any increase in perspective embedding ("The rat that the cat chased was grey"). It is more difficult to attain perspective embedding without some additional syntactic complexity, but the two are nonetheless distinct. Here, we will use embedment only to refer to perspective embedding, whether or not syntactic embedding occurs as well.

Each level of embedment can have a single actor or different actors ("I thought I knew" vs. "Jane thought I knew"), but there is a sense in which an actor is added even when it is the same person. The "I" of "I thought I knew" has to take two perspectives, and perspectives can really only be taken by people. Because the perspectives are different, there is a sense in which the actors must be different, even if they inhabit the same body. Arguments can be raised on either side of the issue of whether there are one or two "I"s in "I thought I knew," (a problem that Aristotle attempted to sort out in his logic and that developmental psychologists work on today (Keil, Rozenblit, & Mills, 2004)). Greater differences arise when there are multiple actors. For example, "I knew that Ivan and Katya went to the store" has three virtual actors, while "Ivan, Katya and I remembered that we went to the store together" has six (three in each embedment).

As a preliminary step toward understanding the role of this kind perspectival embedment in literary complexity, we devised an empirical test to begin to determine whether the number of actors and the level of embedment may have an impact on reading time. It has been found, not surprisingly, that readers spend more time on difficult passages (e.g., Britton, Westbrook, & Holdredge, 1978). Here, two behavioral tests had participants reading short vignettes, to determine whether certain levels of perspective embedment were easier to process, and whether it mattered whether there was one actor or more in the story. We tested vignettes with either one or three actors, and with one of six maximum levels of embedment (0–5). For the higher levels, it is nearly impossible to have every sentence in a story at the maximum level of embedment. Sentences with lower levels of embedment would presumably dilute the effect of embedment, but the maximum level indicates the approximate value we could expect. As a further exploration, we tested whether adding "virtual actors" (as with the second "I" of "I thought I knew") provided a better explanation of the results. When texts are difficult, readers often backtrack ("regressions", e.g., Rayner, 1977). Reading times cannot be directly compared if they contain different numbers of regressions. Therefore, we used a moving window presentation that did not allow any regressions.

We expected that the greater levels of embedment would be more difficult, with the exception that the zero-level (no indication of ToM) would also be difficult. With greater levels of embedment, the relationships that must be stored in memory to understand a vignette increase, which should increase reading time. However, for the zero level, there is no perspective to attach the facts to. Our assumption is that humans engage in stories about humans and are therefore more inclined to organize information in that way. Stories that lack such a perspective will then be more difficult than the "zero" label would indicate.

Different levels of difficulty were measured by differences in total reading time. It was impossible to balance individual sentences within vignettes, so that only total reading time made sense as a measure of difficulty. We further expected that there would be no difference between the one and three actor stories. Because the actor of the one actor stories would have multiple "versions" by virtue of the embedded perspectives, it seemed likely that adding a perspective would be equivalent to adding an actor. However, it did seem possible that an explicit encoding of the number of virtual actors might explain more of the variance than number of actors alone.

Experiment 1: Self-timed reading

The first experiment was designed to use reading time as a proxy for difficulty of processing the vignette. We also presented a probe question after each vignette, which both ensured that the participants were reading for understanding and allowed us to explore difficulty a second way, via accuracy on the probe questions.

Method

Materials and design

We wanted to test reading time in a somewhat controlled way, because there are so many factors that influence reading time with connected text. We controlled for the direction of reading by presenting only one segment of text at a time. This technique prevents readers from backtracking when the text is difficult (for a review, see Rayner, 2009), which can complicate measures of overall reading time. We used the "moving window" technique (Just, Carpenter, & Woolley, 1982; McConkie & Rayner, 1975) to allow only a portion of the text to appear at a time. The reader could move forward at a self-selected pace but could not go back. The total amount of material was indicated by hyphens; there was a space indicating where the segments (portions of text that appeared at once) were. The hyphens were replaced by text (letters, punctuation or spaces) when the reader hit a key, and that material was again replaced with hyphens when the next key press came (see Figure 1).

There were 84 vignettes, 7 at each combination of number of actors (1 or 3) and levels of embedment (0–5). These were presented in a different random order for each reader.

We set the number of characters (including the alphabetic ones and punctuation, but excluding spaces) in each vignette to 350. They averaged 426.1 characters with spaces (ranging from 412 to 436).

The number of lines was always 7. The last line was always about half as long as the other lines. The number of punctuation marks averaged 6.2 (ranging from 5 to 9). The number of sentences averaged 4.8 (ranging from 2 to 10).

There were always four segments in the first six lines and two in the last. The end of a sentence always coincided with the end of a segment. Sometimes this left only a single character in the next segment (the word "I").

After receiving instructions about the procedure, readers were given a sample vignette to gain familiarity with the mechanics of the experiment. At the beginning of a vignette, the screen showed only the hyphens and the spaces between segments. When the "8" key was entered, the first segment of the first sentence was displayed. When the "8" key was entered again, the first segment reverted to

```
------------ how to get from Surf Avenue ----------------------

----------------------------------------------------------------

------------------------------------------------------------

--------------------------------------------------------------

----------------------------------------------------------------

--------------------------------------------------------------

------------------------------
```

Figure 1. A sample vignette, as seen by the readers. The second segment of the vignette is displayed.

hyphens, while the second segment was displayed (as in Figure 1). This procedure continued until all the segments of the vignette had been displayed. The reading time for individual portions varied greatly, given that they varied from 1 to 14 characters in length. The total reading time across the 26 segments that constituted each vignette, however, could be compared across vignettes.

After each vignette had been read, a comprehension question appeared on the screen. The questions were of the yes/no variety, and were deliberately varied in difficulty. We believed that if all the questions were difficult, it would discourage the readers. If they were all easy, they would fail in their main purpose, which was to ensure that the readers were paying attention to the content of the vignettes (rather than just hitting the button to speed the experiment along). Some of the questions were quite difficult and did not have a completely clear correct answer. It was hard to devise short, yes/no questions for the fourth and fifth level of embedment that could not be argued about.

Vignettes were presented on a computer screen via the ePrime program (Psychology Software Tools).

Participants

Readers were 16 young adult native speakers of English with no reported speech, language or reading difficulties. Ten were female, and six male. They came from various regions of the United States, primarily the Northeast. All were college students or college educated. They were paid for their participation.

Results

We analyzed the reading time of all vignettes for all readers regardless of their responses to the probe questions. Because the probes varied in difficulty, we expected some mistakes (or disagreements), even among those readers who had thoroughly read the text. Thus an incorrect answer did not indicate that the reading time was not meaningful.

Overall accuracy on the probe questions was 83.8% (with participants' values ranging from 74 to 93%). Probe questions for individual vignettes ranged from 21% to 100% accurate. There were four vignettes that had less than 50% accuracy, and four more at 50%. These were questions that could have been argued one way or another; we thought we were right, but we could see how someone could disagree. Given that the questions appear to have ensured that the readers paid attention to the meaning of the vignettes, we did not make any further interpretation of the absolute rates.

Figure 2 shows the mean accuracy on the problem questions at each level of embedment for 1 actor and 3 actor vignettes. Embedment had an effect on accuracy (F(5,168) = 7.94, p < .001). Neither the number of actors nor the interaction of the two main effects were significant (both Fs less than 1). Post-hoc tests indicated that only the fifth level differed from the others.

As seen in Figure 3, there was no speed-accuracy trade-off in these results. If anything, readers took longer with the vignettes for which they had lower accuracy.

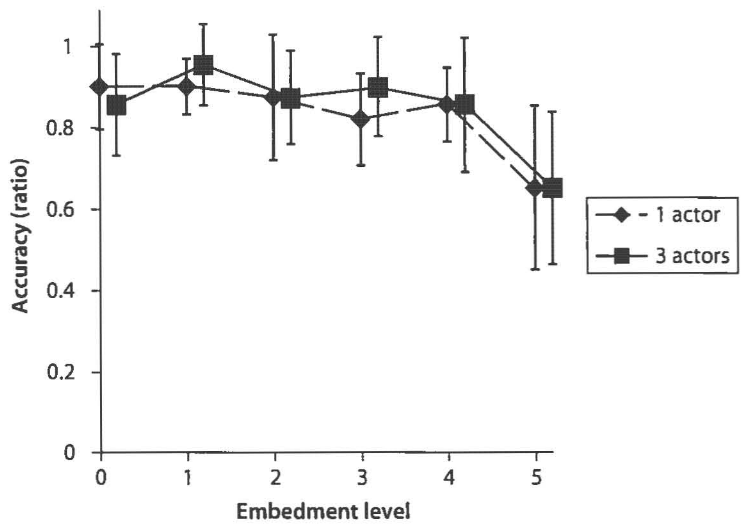

Figure 2. Mean proportion correct on probe questions. Error bars indicate plus or minus one standard deviation.

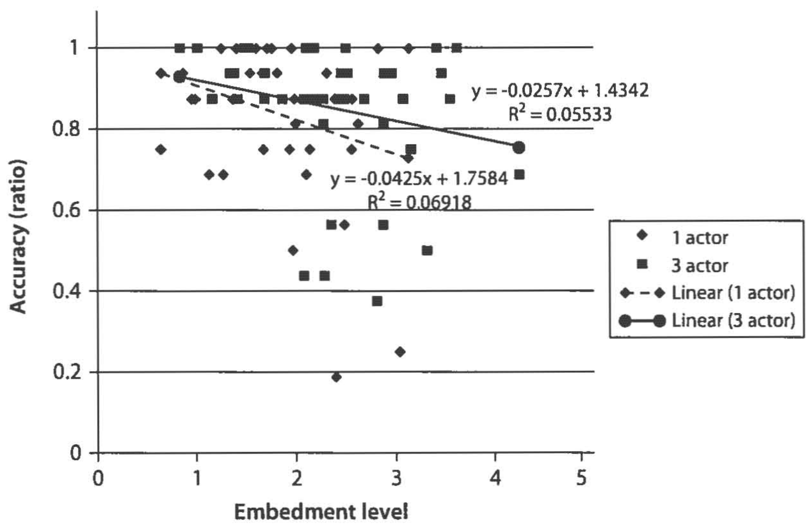

Figure 3. Speed vs. accuracy, Experiment 1. Values are collapsed across all participants; all 84 vignettes are represented.

Across the 16 readers, embedment level 1 garnered the fastest reading time, level 5 the slowest, and level 0 a relatively slow rate (see Figure 4).

A repeated measures Analysis of Variance was performed with the factors Number of Actors (2 levels) and Embedment (6 levels) with participant as a repeated measure. Both main effects were significant (Actors: $F(1,72) = 14.14$, $p < .001$; Embedment: $F(5,72) = 4.08$, $p < .05$). Vignettes with three characters took an average of 1.03 seconds longer to read than those with one character. Level of embedment had the v-shaped pattern apparent in the figure. The interaction was not significant ($F(5,72) < 1$), indicating that the changes due to embedment were statistically identical for both numbers of actors. This is the case despite the apparent divergences at level 4 for the one actor case.

Vignettes with one actor took about a second less to read than those with three actors (1.031 s difference, 21.739 s for one actor vignettes, 22.770 s for three actor ones.) The vignettes with one level of embedment were the fastest, with fairly steady increases as embedment increased. The vignettes without embedment were read at a rate around that of the fourth level of embedment.

To explore possible effects of virtual actors as well as nominal actors, we calculated the maximum number of virtual actors for each vignette. The rationale for taking the maximum was the same as for taking the maximum embedment level; it should differentiate the vignettes most directly. Virtual actors were assumed when they were embedded. Thus "I knew Anne and Tom went to the store" has three actors and three virtual actors. "Anne, Tom and I knew we went to the store" has three actors and six virtual actors. For the 1 actor vignettes, the number of virtual actors was always the embedment level plus 1. For the 3 actor vignettes, the number varied from 3 to 10. Because only the 3 actor vignettes differed along this dimension, we limited our analysis to those vignettes. We performed pairwise correlations of

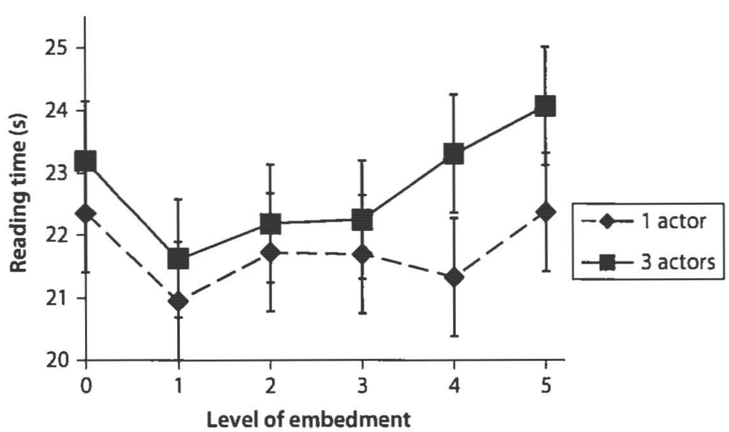

Figure 4. Total reading time averaged across 16 readers, plotted separately for each level of embedment and number of actors. Error bars represent the 95% confidence interval.

embedment, number of virtual actors and the total reading time. With all six levels of embedment, there was a slightly better correlation with virtual characters than with embedment ($r = .33$ ($p < .05$) and .30 ($p < .10$) respectively). When we removed the 0 level (which is nonlinear with respect to the other levels of embedment), the correlations are stronger, but now the embedment provides a moderately larger correlation than the virtual characters (embedment: $r = .55$, $p < .001$; virtual: .45, $p < .01$). In short, there seems to be little difference in coding the complexity of the vignettes in terms of embedment or of number of virtual characters.

Discussion

The error rates were essentially the same for all levels except for the fifth. Although there is likely to be a processing reason attached to this result, it is also the case that the questions for the fifth level struck us as more difficult, both in constructing them and in answering them ourselves. We could have constructed unambiguous level zero questions, such as "Was the mitten green?", but the questions for the higher levels needed, at least on occasion, to force the reader to interpret the greater degree of embedment. It is difficult to make such questions short as well as clearly true or false. Thus the questions for the higher levels, particularly level five, were expected to have lower success rates. Indeed, some of the answers were debatable. We retained these questions in an effort to ensure deep processing of the vignettes, at the expense of strict comparability of error rates. It is possible that the lower levels were at ceiling, and that using harder questions in general would show a more nuanced pattern. However, our goal with the probe questions was to ensure thorough reading, and we believed that having all the question be hard ones would be generally discouraging to the participants. Thus the error rates should be interpreted with caution.

The reading time results partially confirm our predictions. Embedment does indeed facilitate reading, as measured by reading time, when we compare the "0" level with most of the others. The vignettes with the first level of embedment were read the fastest, followed by the various further levels of embedment. When there was no embedment, reading times were relatively long. The vignettes with a single actor, however, were read more quickly than those with three actors. Even so, the levels of embedment behaved quite similarly, so the added "actors" implied by the embedment did not, at first glance, add to the reading time in the same way. Embedment seemed to be fairly equivalent to the number of virtual agents in the vignette in terms of its effect on reading time.

It is impossible to match total "complexity" or "amount of information" across these vignettes. There is currently no metric for these types of meanings that could give us an unambiguous basis for making such a measure. The vignettes had similar levels of vocabulary and syntax, though the more embedded versions necessarily had more syntactic complexity as well. So it is possible that some aspect of information and/or complexity other than those we are focusing on are responsible for the differences in reading time. To control for at least some aspects of information, we avoided, for example, describing social situations that would recruit reasoning along the lines of "cheater detection" (Cosmides, 1989). However, equating the vignettes for reading time would necessarily make them reverse just the distinctions we are trying to examine, by making the hard cases easier and the easy cases harder. We can only report that we attempted to make all the vignettes as equally accessible as possible. It may be that the fourth level of embedment for the single actor vignettes was somewhat less complex than the equivalent stories with three actors, thus leading to the dip in the function (see Figure 4). Similarly, there is no agreed-upon measure of information of this kind of discourse, so it is possible that there is more information in the zero level embedment vignettes than in the higher level ones, which could also explain the longer reading times. Alternatively, it may be that readers find it difficult to engage with vignettes that do not have any social content, i.e., any indication of another, engaged human (implied by embedment) and therefore take longer to read them.

Experiment 2: Reading with preset timing

The results of the first experiment suggest that readers engage in texts differently depending on the degree of embedment in the story. In the second experiment, we tested whether this effect would be seen in the error rates if we enforced the timing of the presentations rather than allowing the reader to select the pace. To that end, we used the average time spent on each segment by the readers in Experiment 1 to determine the display in Experiment 2. Thus each segment was presented for a time that was proportional to the average reader's self timing. Further, the total reading time was set to 23 seconds for all vignettes, primarily in order to prepare for a brain imaging study with these stimuli. This value was somewhat longer than the average reading time from the first study, and the longest time we could fit into our proposed brain imaging protocol. The easy vignettes would, on average, seem somewhat leisurely, while the harder ones should seem somewhat hurried.

Method

The same procedure as in the first study was used, except for the changes in the timing of the presentation mentioned before.

Participants

Ten new readers (who had not participated in the first study) were recruited as before. Six were female and four were male.

Results

Only accuracy on the probe question could vary, so that is all that is reported. Readers on average gave correct responses to the probe questions 86% of the time. This did not differ from the previous experiment either by readers ($F(1,22) < 1$, n.s.) or by items ($F(1,83) < 1$, n.s.). There was no clear pattern in relationship to number of actors nor of embedment (see Figure 5).

Discussion

The pattern of errors was very similar to that of the first study. Although we might have expected that the easier vignettes would have been even easier, and the harder more difficult, there was, in fact, no effect of imposing a single timing format to the vignettes. For the harder vignettes, this might have due to an increase in overall attention that would be available. With the imposed timing, the reader was freed from the task of advancing the text. Although not terribly demanding, this task nonetheless can be assumed to take some attentional resources. These resources would presumably have been available for the task of reading the text, which would perhaps compensate for the lesser amount of time available. For the easier

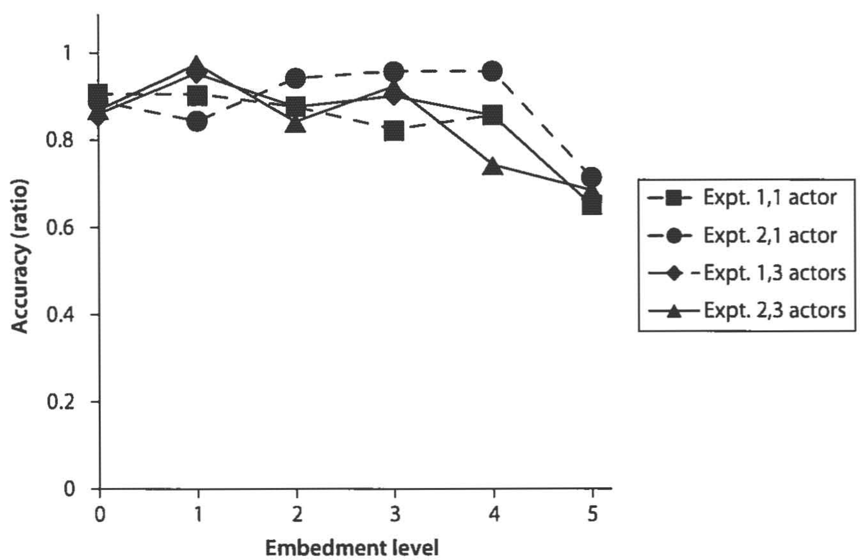

Figure 5. Comparison of error rates across the two experiments.

vignettes, it may be that there is a ceiling effect, and so there is no improvement even though the reading times were longer on average than in the first experiment. In general, the accuracies are remarkably similar across the two experiments.

General discussion

When readers are faced with increasing levels of cognitive embedment, they take progressively longer to read passages of the same length, with one notable exception: When there is no embedment, they take as long as when there is a high level of embedment. It would seem that having no engagement with social and cognitive relations makes a text harder to read. It is conceivable that the lack of engagement makes reading slow just because there is less of a compelling reason to read on, but the task demands of the experiment did not lend themselves to that interpretation. Participants were trying to read as quickly as possible and still be able to answer the comprehension question. It seems likelier that the zero levels were simply more difficult. The difficulty may come from the social factors directly (the basis for our prediction of slow reading times for level zero): Our mechanisms for understanding narrative may preferentially make use of social content, which is lacking in the level zero vignettes. Alternatively, it may be that the level zero vignettes had more "information" and thus needed more reading time on that count. There is no currently available metric for determining the amount of information in a vignette, so we are not able to dissociate these interpretations. The vignettes with zero-level mental embedment approximate real-life texts, such as instruction manuals, that eschew fictionality (if we define fictionality as presence of complex mental embedments (Zunshine, 2011)) and as such require more laborious processing on the part of the readers. Further experiments should shed some light on this issue.

The slower reading times for the fourth and fifth levels do not necessarily imply that the harder texts were less rewarding; it may be that the added effort put into the texts would, on the contrary, make their interpretation more satisfying. However, given the instructions and the relative lack of literary merit in our vignettes, it seems most likely that reading time primarily tracked difficulty of interpretation. If we were assessing literary merit, finding texts that elicited the fastest reading times would not be a goal.

We are aware, of course, that the kind of reading we do when perusing a literary text differs from the reading done by our subjects in a number of ways. To identify only one: when reading a novel, each segment of the reading is influenced by the total structure of the text, a frame absent from our vignettes. Reading a literary text is a cumulative experience, whereas subjects in our experiment were confronted with a number of disconnected vignettes. It is the elaborateness of a literary text's total structure — often reflected in the depth and complexity of the mental states it portrays — that comprises its complexity. But such complexity can be realized in many different ways. The complexity of Kafka is not that of Proust. Thus finding a quantifiable stand-in for literary complexity is difficult, but we think that embedment of the kind found in this study is at least a beginning.

One way to further pursue this result is to examine areas of brain activation during the reading of the vignettes. Then, these areas can be compare to those for ToM (e.g., Powell, Lewis, Dunbar, Garcia-Finana, & Robert, 2010; Saxe, 2006) and for the mental representation of numbers, or numerosity (e.g., Heim, et al., 2012; Piazza, Izard, Pinel, Le Bihan, & Dehaene, 2004). Such a study has been initiated, and we expect the results to further elucidate the reading time indicators found here.

Readers, apparently, do not seek the easiest text when they read fiction. If they did, writers would concentrate on the first level of embedment, and we would still find a Dick-and-Jane-see-Spot-run level of complexity attractive. Instead, readers prefer to go deeper, spending more time at the third level. Although it might appear that including even higher levels of embedment would be rewarding, it seems that the challenge may be too great. As Zunshine (2008), in a somewhat different context, says, "There seems to be a fine line here between 'good, because quite challenging' and 'not that good, because way too challenging' " (p. 36). Our preliminary results indicate that the third level of embedment is this "sweet spot" for engaging readers in literature.

Acknowledgments

This research was supported by a grant from the Teagle Foundation to Yale University and Haskins Laboratories, and by NIH grant HD-01994 to Haskins Laboratories. We thank members of the Teagle consortium who contributed vignettes: Robert F. Barsky, Peter Steiner, and Eugenia Kelbert. We thank Josh Coppola for running the studies, Peter Molfese for technical assistance, and Kenneth R. Pugh, Stephen J. Frost, Einar Mencl, Philip E. Rubin and Robert F. Barsky for helpful comments.

References

Atran, S. (2002). *In gods we trust: The evolutionary landscape of religion.* Oxford: Oxford University Press.

Boyer, P. (2001). *Religion explained: The evolutionary foundations of religious belief.* New York: Basic Books.

Britton, B. K., Westbrook, R. D., & Holdredge, T. S. (1978). Reading and cognitive capacity usage: Effects of text difficulty. *Journal of Experimental Psychology: Human Learning and Memory, 4*, 582–591.

Carruthers, P., & Smith, P. K. (Eds.). (1996). *Theories of theory of mind*. Cambridge: Cambridge University Press.

Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.

Cosmides, L. (1989). The logic of social exchange: Has natural selection shaped how humans reason? Studies with the Wason selection task. *Cognition, 31*, 187–276.

Dunbar, R. I. (1996). *Grooming, gossip, and the evolution of language*. Cambridge, MA: Harvard University Press.

Dunbar, R. I. (2000). On the origin of the human mind. In P. Carruthers & A. Chamberlain (Eds.), *Evolution and the human mind: Modularity, language, and meta-cognition* (pp. 238–253). Cambridge: Cambridge University Press.

Everett, D. L. (2005). Cultural constraints on grammar and cognition in Piraha: Another look at the design features of human language. *Current Anthropology, 46*, 621–646.

Gallagher, H. L., Happé, F., Brunswick, N., Fletcher, P. C., Frith, U., & Frith, C. D. (2000). Reading the mind in cartoons and stories: An fMRI study of 'theory of mind' in verbal and nonverbal tasks. *Neuropsychologia, 38*, 11–21.

Ginsberg, A. (1990). Connecting diversification to performance: A sociocognitive approach. *The Academy of Management Review, 15*, 514–535.

Heim, S., Amunts, K., Drai, D., Eickhoff, S., Hautvast, S., & Grodzinsky, Y. (2012). The language-number interface in the brain: a complex parametric study of quantifiers and quantities. [Original Research]. *Frontiers in Evolutionary Neuroscience, 4*(4), 1–12.

Herman, D. (2006). Genette meets Vygotsky: Narrative embedding and distributed intelligence. *Language and Literature, 15*, 357–380.

Hockett, C. F. (1960). The origin of speech. *Scientific American, 203*(3), 89–96.

Just, M. A., Carpenter, P. A., & Woolley, J. D. (1982). Paradigms and processes in reading comprehension. *Journal of Experimental Psychology: General, 111*, 228–238.

Keil, F. C., Rozenblit, L. R., & Mills, C. (2004). What lies beneath? Understanding the limits of understanding. In D. T. Levin (Ed.), *Thinking and seeing: Visual metacognition in adults and children* (pp. 227–249). Cambridge, MA: MIT Press.

Mar, R. (2011). The neural bases of social cognition and story comprehension. *Annual Review of Psychology, 62*, 103–134.

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

Piazza, M., Izard, V., Pinel, P., Le Bihan, D., & Dehaene, S. (2004). Tuning curves for approximate numerosity in the human intraparietal sulcus. *Neuron, 44*, 547–555.

Powell, J. L., Lewis, P. A., Dunbar, R. I., Garcia-Finana, M., & Robert, N. (2010). Orbital prefrontal cortex volume correlates with social cognitive competence. *Neuropsychologia, 48*, 3554–3562.

Rayner, K. (1977). Visual attention in reading: Eye movements reflect cognitive processes. *Memory & Cognition, 5*, 443–448.

Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *Quarterly Journal of Experimental Psychology, 62*, 1457–1506.

Saxe, R. (2006). Why and how to study Theory of Mind with fMRI. *Brain Research, 1079*, 57–65.

Singer, T. (2006). The neuronal basis and ontogeny of empathy and mind reading: Review of literature and implications for future research. *Neuroscience and Biobehavioral Reviews, 30*, 855–863.

Sperber, D. (2000). Metarepresentations in an evolutionary perspective. In D. Sperber (Ed.), *Metarepresentations: A multidisciplinary perspective* (pp. 117–137). Oxford: Oxford University Press.

Spreng, R. N., Mar, R. A., & Kim, A. S. N. (2009). The common neural basis of autobiographical memory, prospection, navigation, Theory of Mind, and the default mode: A quantitative meta-analysis. *Journal of Cognitive Neuroscience, 21*, 489–510.

Zunshine, L. (2008). *Strange concepts and the stories they make possible.* Baltimore: Johns Hopkins University Press.

Zunshine, L. (2011). What to expect when you pick up a graphic novel. *SubStance, 40*, 114–134.

Corresponding author

D. H. Whalen
Haskins Laboratories
300 George St., Suite 900
New Haven, CT 06511
203-865-6163, ext. 234
fax: 203-865-8963

whalen@haskins.yale.edu