CALL #: P302 .D55
LOCATION: Washington State Holland

TYPE: Article
USER JOURNAL TITLE: Discourse Processes
OCLC JOURNAL TITLE: Discourse Processes
NTE CATALOG TITLE: Discourse processes.
ARTICLE TITLE: Kennison, Sheila: Comprehending referential expressions during reading: Evidence from eye tracking.
VOLUME: 24
NO: 2
COPYRIGHT: CCL
PAGES: 229-
ISSN: 0163-853X
OCLC #: 
VERIFIED:

PATRON: Kennison, Sheila
PATRON ID: 0
PATRON PHONE:
PATRON DEPT: Psychology
PATRON STATUS: Faculty 51416 jh
PATRON FAX:
PATRON ADDRESS:
PATRON E-MAIL:
PATRON NOTES:
Comprehending Referential Expressions During Reading: Evidence from Eye Tracking

SHELVIA M. KENNISON
University of Oklahoma

PETER C. GORDON
University of North Carolina, Chapel Hill

Two eye-tracking experiments investigated how referential expressions, consisting of proper names and pronouns, influence reading comprehension. Experiment 1 showed that repeated names caused elevated reading times compared to pronouns, a finding that has been called the repeated-name penalty in studies using self-paced reading (Gordon, Grosz, & Gilliom, 1993). Consistent with these previous studies, Experiment 1 also showed that the repeated-name penalty occurs for syntactic subjects that corefer with the subject of the preceding sentence, but does not occur for direct objects that corefer with an object of the preceding sentence. These results further serve to localize the repeated-name penalty within the sentence and to show that it can be associated with an increase in the frequency of regressive saccades out of the text region beyond the repeated name. Experiment 2 showed that the repeated-name penalty was modulated not only by syntactic factors within a sentence but also by the relationship between successive sentences in a discourse. The results of these experiments provide strong support for the centering theory of discourse and related approaches to the processing of referential and coreferential expressions.

During language comprehension, referential expressions (e.g., proper names, pronouns, definite nouns, etc.) provide readers and listeners with certain types of essential information. Such expressions introduce entities into the world of a discourse and can also link separate utterances together within a discourse world. Several theories of discourse structure have embraced these notions (Grosz, Joshi,
Theoretical Developments in Centering Theory and Discourse Representation Theory

Centering theory (Sidner, 1983; Grosz et al., 1983, 1995, Grosz & Sidner, 1986; Gordon et al., 1993 for a review) consists of the following central ideas. Discourses are composed of utterances that realize semantic entities as linguistic forms. Each utterance in a locally coherent discourse is linked to the immediately preceding utterance by a single entity, the “backward-looking center” or $Cb$. Each utterance also realizes a set of “forward-looking centers” or Cf’s that provide potential links to subsequent utterances. For example, consider passage (1). Utterance (a), being the initial utterance in a discourse segment, does not contain a backward-looking center because there is no preceding utterance. It contains the forward-looking centers Susan, Betsy, and some hamster. Utterance (b) contains the backward-looking center Susan and the forward-looking centers Susan, Betsy, and hamster, ranked in that order based on criteria such as grammatical role, surface position, realization as a pronoun and accent when speaking (for discussion see Gordon, 1993; Gordon et al., 1993).

(1a). Susan gave George a pet hamster.
   $Cf$ = {Susan, George, hamster}
(1b). She reminded him that such hamsters are quite shy.
   $Cb$ = Susan; $Cf$ = {Susan, George, hamster}
(1c). She asked him whether he liked the gift.
   $Cb$ = Susan; $Cf$ = {Susan, George, gift = hamster}
The construction rule for pronouns is triggered by a pronoun; it attempts to find a suitable antecedent (based on number, gender, animacy and fundamental principles of disjoint reference) among the entities already in the discourse model. (If that fails, the construction rule posits a new discourse entity.) The final construction rule for equivalence is triggered by the presence of two tokens of the same name predicated on different discourse entities; it adds a condition to the discourse representation equating the two entities.

Entities in the discourse model differ in prominence, a notion based on the idea of an ordered set of forward-looking centers developed in centering theory (Grosz et al., 1983, 1995). The operation of the construction rules for pronouns and equivalence, both of which are capable of establishing coreference, is differentially influenced by this prominence. The construction rule for pronouns searches the discourse entities from most prominent to least prominent, thereby facilitating pronominal reference to a prominent entity. This operation during comprehension reflects the use of pronouns or other reduced forms to refer prominent entities. In contrast, the construction rule for equivalence searches for a match from the least prominent entity to the most prominent entity, an operation that reflects the use of more complete forms to refer to non-prominent entities.

**Empirical Results on the Comprehension of Referring Expressions**

Experimental study of the comprehension of referring expressions is a very active domain of research that has yielded a large number of empirical results. The results that are most relevant to the approach taken here involve a comparison of the ease of understanding reduced referring expressions (e.g., pronouns) and full referring expressions (e.g., names) because centering theory (Grosz et al., 1983, 1995) and related approaches (Gordon & Hendrick, 1997, in press) specify circumstances under which reduced referring expressions are more easily understood than full referring expressions.

However, the idea that reduced referring expressions would ever be more easily understood than full referring expressions is counterintuitive. Reduced referring expressions seem to provide less information on which to identify their referents than do full referring expressions. This intuition is given theoretical expression by Gernsbacher (1989) who argues that principles of memory operation lead to the expectation that coreferential names will activate appropriate memory representations more effectively than do pronouns because the names provide a better match to the memory representations. Gernsbacher (1989) supports this position with the results of probe-word studies that show that repeated-name anaphors cause reduced response times to probe words consisting of that name, while pronoun anaphors do not. In contrast, Gordon et al. (1993) have argued that these results could reflect the relation between the anaphor and the probe rather than the relation between the anaphor and the memory representation to which it refers. They review the results of other work that indicates that pronouns are interpreted more readily than are repeated-name anaphors; they also present a series of self-paced reading time studies specifically designed to test predictions of centering theory.

Gordon et al. (1993) measured reading time for short passages containing utterances like 1(a) and 1(b) followed by a third utterance containing referential expressions in subject and direct object positions; the linguistic form of these two expressions was manipulated so as to create three conditions: Pronoun-Pronoun, Pronoun-Name and Name-Name. The theoretical analyses of Grosz et al. (1983, 1995) lead to the expectation that the first entity, the one in subject position, would be the backward-looking center. Reading time for the passages was significantly longer for the Name-Name condition than for the average of the Pronoun-Pronoun and Pronoun-Name conditions. There was no significant difference observed between the Pronoun-Pronoun and Pronoun-Name conditions. Gordon et al. (1993) called the elevated reading times observed when the backward-looking center was realized as a name, the “repeated-name penalty.” These results were consistent with the first rule of centering concerning the importance to discourse coherence of realizing the backward-looking center as a pronoun. The contention of centering theory that the center relations between successive utterances determines the contribution of pronominal reference to discourse coherence was supported by an experiment showing that the repeated-name penalty was reduced in an utterance that shifted the backward-looking center (e.g., 1.c) as compared to one that continued it (e.g., 1.c). Additional studies have shown that syntactic role rather than thematic role determines the backward-looking center of an utterance (Gordon & Chan, 1995) and that centering principles can account for the initial interpretation of syntactically ambiguous pronouns (Gordon & Searce, 1995).

The processing model of Gordon and Hendrick (1997, in press) readily accounts for the patterns observed in these reading time studies. The repeated-name penalty occurs for an expression in subject position when the entity it refers to was also referred to by the subject in the preceding utterance. In this situation, the expression is referring to an entity that has maximal prominence in the discourse model due to the syntactic status with which it was introduced to the model (Gordon & Hendrick, in press). Accordingly, this situation favors operation of the construction rule for pronouns and disfavors operation of the construction rule for names thereby yielding more rapid comprehension of a pronoun than a repeated name. In the situations where the repeated-name penalty is not observed, objects corefering with objects and subjects corefering with objects, pronominal reference is no longer favored because the entity referred to is no longer the most prominent in the discourse model due to the syntactic status of the antecedent.
expression. This reduced prominence decreases the efficiency of the construction rule for pronouns and increases the efficiency of the construction rule for equivalence. Thus, the Gordon and Hendrick (in press) analysis provides a straightforward account of the results observed in this current studies.

**Measures of the Time Course of Reading Comprehension**

The experiments reported by Gordon et al. (1993), Gordon and Chan (1995), and Gordon and Searce (1995), used a self-paced reading-time methodology. Readers viewed passages a phrase or a sentence at a time on a computer screen and responded with a manual key press when they had read the displayed region. In some experiments, sentences were divided into presentation regions at natural boundaries with each phrase containing between 2 and 6 words. In others, entire sentences were presented in a display. Each displayed phrase or sentence appeared at a central location on the computer screen. Readers could not look back at previously presented portions of the passage.

Self-paced reading methods are widely used by language comprehension researchers (see Ferreira & Henderson, 1990; Kennedy & Murray, 1984; Mitchell, 1984). However, the extent to which results gained using self-paced reading generalize to “normal” reading remains controversial. Self-paced reading differs from normal reading in two important ways. First, self-paced reading requires manual key presses for the reader to indicate that reading has been completed for a particular portion of text. The necessity of manual key presses may give rise to slower than normal reading rates when the displayed regions are short. Furthermore, some readers may adopt button-pressing strategies (Ferreira & Henderson, 1990; Mitchell, 1984), causing reading time measurements to reflect task-specific strategies. A second important difference between self-paced and normal reading is that during self-paced reading, readers may not be permitted to reread previously presented portions of text. During non-cumulative presentation, a technique widely used by comprehension researchers and the technique used by Gordon et al. 1993, Gordon and Chan (1995), and Gordon and Searce (1995), once a particular portion of the text is viewed and the button pressed, that portion of the text cannot be reaccessed for the purposes of rereading. However, this technique is preferred over cumulative presentation in which text, once read, remains in view until the entire sentence is completely read as one cannot conclude that reading time collected using the latter method stems from processing occurring on the newly presented region of text or from processing occurring both on the newly presented region of text and on previously processed regions of text (see Holmes, Stowe, & Cupples, 1989, for an example and see Ferreira & Henderson, 1990, for a critique).

In contrast, during eye tracking, passages are presented to readers in their entirety, enabling readers to read at a comfortable rate and to reread at any time.

Eye tracking provides a complete record of eye movement behavior during reading. It is possible to determine whether a word has been directly viewed or skipped, how long readers spend acquiring information from regions of the text, and when the eyes move back to a previously read portion of the text. In this way, initial processing on a word or region can be analyzed separately from later processing on a word or region (Rayner & Pollatsek, 1989; Rayner & Sereno, 1994). Eye tracking has been used to study processes involved in reading since the late 1880s (Huey, 1908; Dodge, 1900; Dodge, 1906). Since the 1970s, the number of researchers using eye tracking to investigate reading comprehension has steadily grown, despite the high cost of eye tracking devices. In response to the high cost of conducting eye tracking experiments, many researchers use self-paced reading as the initial methodology for addressing research questions and this methodology has provided promising results are obtained, use eye tracking as a means of more precisely exploring those questions.

Our study draws on the precision of the eye-tracking methodology to examine how type of referent (name vs. pronoun) influences the comprehension of short passages. In doing so, it examines whether results gained from eye tracking and self-paced reading provide similar pictures of the impact of repeated names on reading comprehension. The results of these experiments provide a more fine-grained view of the repeated-name penalty. Two experiments are described. These experiments measure reading time on short passages containing proper names and pronouns. The results of the experiments provide strong support for two of the central predictions of centering theory: (1) that using a repeated name to refer to the central entity of a discourse disrupts comprehension and (2) that this disruption interacts with whether there is a continue or shift relation between an utterance and the preceding utterances.

**EXPERIMENT 1**

This study used eye-tracking to examine the effects of varying type of referent (pronoun vs. name) in different grammatical roles (subject vs. object) in a short passage. The structure of the passages is illustrated in (2). The first sentence introduced a single named character in the subject position. The second sentence continued this character in subject position, realized as a pronoun. It also introduced a second named character as a postverbal object. The critical third sentence continued the subject character of the first two sentences and realized the second entity as the direct object. The type of referent was manipulated in this sentence for both the subject and direct object, yielding four experimental conditions: pronoun-pronoun; name-pronoun; pronoun-name; and name-name. The fourth (final) sentence of each passage was identical in all conditions.
2. Susan really likes animals. The other day she gave Betsy a pet hamster. She/Susan reminded her/Betsy that such hamsters are quite shy and need gentle handling. She told her that hamsters eat vegetable scraps and live about four years.

Previous self-paced reading time results have shown that using a name to realize an entity that is continued in subject position leads to elevated reading times, but that this elevation does not occur for direct objects if the subject of the sentence has been realized in the preceding sentence (Gordon et al., 1993; Gordon & Chan, 1995). The repeated-name penalty has been taken as a marker of the backward-looking center of an utterance. The repeated-name penalty provides evidence against Gernsbacher’s (1989) proposal that repeated name anaphors should be easier to process, and consequently should be read more quickly, than pronoun anaphors.

The repeated-name penalty is consistent with the Gordon and Hendrick (in press, 1997) model, as the construction rule for interpreting pronouns is most efficient when the antecedent is a syntactically prominent expression (e.g., the subject) whereas the construction rule for equivalence, which is used to establish coreference between names, is most efficient when the antecedent is not syntactically prominent. If eye-tracking and self-paced reading are influenced in the same way by the manner of realizing a referential expression, we would expect that reading time will be elevated when the subject of the third sentence is realized as a name, but not when the object of the third sentence is realized as a name.

Method

Participants. Twenty undergraduate and graduate students at the University of Massachusetts, Amherst, were recruited to participate in the experiment for course credit or for $5.00. All participants were fluent in American English, had normal or corrected vision (soft contact lenses only), and were naive to the purpose of the experiment.

Materials. Forty short passages (four versions each) were constructed. In these experimental passages, all proper names unambiguously indicated the gender of the entity. In nineteen of the passages, the two entities were of the same gender (e.g., “Susan” and “Betsy”). In the remaining twenty-one passages, the two entities were of different gender (e.g., “Sam” and “Lydia”). The average length of proper names was 4.9 characters having a range of between 2 and 9 characters. Eighty short passages were also included as fillers. Filler passages contained a variety of proper names and pronouns.

Apparatus. Eye movements were recorded by a Stanford Research Institute Dual Purkinje Eye tracker, which has a resolution of less than 10 min of arc. Viewing was binocular with eye position recorded from the right eye. The eye tracker was interfaced with an 80486 microcomputer which controlled the presentation of the sentences. Up to 80 character spaces per line were used. The letters were in lower-case, except where capital letters were called for (at the beginning of sentences and proper names). Subjects were seated 62 cm from the monitor and 4 letters equaled one degree of visual angle. The luminance from the monitor was adjusted to a comfortable brightness level for the subject and then held constant throughout the study, and the room was dark.

Procedure. For each participant, a bite-bar was constructed to minimize head movements during the experiment. The eye tracking system was then calibrated. This procedure required the participant to fixate nine markers sequentially, starting at the upper third of the screen and at the left corner, continuing on the same row at the middle of the computer screen, and continuing on the same row at the right corner of the computer screen, and continuing in the same manner for the lower third of the screen and the middle row of the computer screen. The voltage was recorded and interpolated for the intervening columns and rows. Before each trial, the participant was asked to serially fixate a row of five markers equally spaced across the computer screen in order to check this interpolation. If the interpolation deviated substantially from the markers, then the calibration was repeated.

At the beginning of each trial, the participant was asked to fixate on a box in the center of the screen and then on one at the left of the screen. When a marker that followed the eye indicated a successful fixation, the sentence appeared. After the participant finished reading the sentence, the participant looked away from the sentence and manually pressed a trigger. At that point, a comprehension question appeared. Comprehension questions appeared in the lower half of the computer screen. Under the question, the word “False” appeared at the left side of the computer screen and the word “True” appear at the right side of the computer screen. Participants answered appropriately by manually pressing the corresponding trigger: the right trigger for TRUE and the left trigger for FALSE. Each session began with a short practice of 16 passages. Each session lasted between 30–50 min.

Results and Discussion

Approximately 3% of trials were initially removed from the dataset due to track loss. Remaining trials were screened for false fixations, following the recommendations of Rayner, Sereno, Morris, Schmauder, and Clifton (1989). Fixations shorter than 80 ms in duration and only one character away from the prior or next fixation were merged with that fixation. Fixations shorter than 40 ms and less
than three characters away from the prior or next fixation were deleted as were any remaining fixations shorter than 50 ms.

Four primary measures of reading behavior were analyzed: (1) the percentage of trials on which the pronoun and proper names were skipped; (2) first pass reading time, which was defined as the sum of all fixations made in a region from the time that the eye first entered a region until the time that the eye first exited a region either to the left or to the right; (3) total reading time, which was defined as the sum of all fixations made in a region; and (4) percentage of passages on which a regressive eye movement was made from a region to a preceding part of the sentence. Further analyses were conducted to determine whether the results described in the preceding sections are solely due to differences in length of the region across conditions. All analyses of variances (ANOVARAs) and pairwise comparisons reported in this paper were performed for participants (F₁) and passages (F₂).

**Skipping Pronouns and Proper Names.** During normal reading, not every word of a sentence or passage receives a fixation; some words are skipped. Word length, printed frequency, and predictability from context are all factors that influence word skipping during reading (Rayner & Pollatsek, 1989). A widely accepted explanation for word skipping is that readers are able to fully identify the skipped word during the immediately preceding fixation. This claim is supported by the observation that the most commonly skipped words are short, high frequency words (Rayner & Pollatsek, 1989) and also by the finding that more skipping occurs on words when readers had available a preview of the word on the immediately preceding fixation than when preview of the word had been unavailable (Kennison & Clifton, 1995).

Table 1 displays the percentage of trials on which the pronoun or proper name was skipped for each condition. ANOVAs were conducted using type of referent and grammatical role as within-participant factors. Pronouns were skipped more frequently than proper names, resulting in a significant main effect of type of referent, F₁(1,19)=102.22, MSE = 151, p < .001, F₂(1,39) = 122.22, MSE = 252, p < .001. Repeat referents in object position, resulting in a significant main effect of grammatical role, F₁(1,19) = 40.29, MSE = 653, p < .001, F₂(1,39) = 66.86, MSE = 329, p < .001. The grammatical role by type of referent interaction approached but did not reach significance, F₁(1,19) = 2.88, MSE = 436, p < .11, F₂(1,39) = 3.83, MSE = 745, p < .06.

The finding that pronouns in subject position are skipped more frequently than those in object position does not appear to be solely due to differences in number of letters for those pronouns. When passages containing only the feminine pronouns “she” and “her” were analyzed separately from those passages containing the masculine pronouns “he” and “him,” a similar pattern of results was observed. “She” was skipped significantly more often than “her” (72% vs. 48%, respectively), F₁(1,19) = 13.61, MSE = 393, p < .002, F₂(1,39) = 24.25, MSE = 197, p < .001. However, for these passages names in subject position were not skipped significantly more often than names in object position (65% vs. 52%, respectively), Fs < 1.

The pattern of skipping the critical referring expressions provides some information about the processing difficulty incurred in reading pronouns and names in different syntactic positions, by suggesting that pronouns are identified more easily than names during a preceding fixation, and that subject pronouns are identified more easily than object pronouns. This could be interpreted as indicating that readers have strong expectations that pronouns will be used in subject position, and that this facilitates identification of subject pronouns during the preceding fixation. Such an interpretation would be consistent with the general findings that pronouns frequently refer to the subject of the preceding sentence (Frederiksen, 1981) and the more specific interpretation of centering theory realizing the backward-looking center as a pronoun is critical to the coherence of a discourse. Such an interpretation would be inconsistent with Germsbacher’s (1989) proposal that repeated name anaphors would involve less processing difficulty than pronoun anaphors. Nevertheless, this pattern of results is also open to plausible interpretations based on the characteristics of the referring expressions themselves rather than on how they fit into a discourse. Most obviously, the repeated names were of greater length than the pronouns, a factor that has a strong influence on the probability of fixating a word (Rayner & Pollatsek, 1989). Indeed, previous studies (Ehrlich, 1983; Ehrlich & Rayner, 1983) have shown that pronouns have a low probability of fixation, roughly on the order of what is observed here for pronouns in subject position, and have attributed this fact to the short length and high frequency of pronouns. The finding that subject pronouns were skipped more frequently than object pronouns of matched length (“she” vs. “her”) is less easily accounted for in terms of the characterististics of the expression itself. Both of these words have very high frequency (6039 and 3132, respectively, from Francis & Kucera, 1982), suggesting that neither length nor frequency is the operative fac-

<table>
<thead>
<tr>
<th>Grammatical Role</th>
<th>Type of Referent</th>
<th>Percentage Skipped</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subject</td>
<td>Pronoun</td>
<td>76</td>
</tr>
<tr>
<td></td>
<td>Name</td>
<td>42</td>
</tr>
<tr>
<td>Object</td>
<td>Pronoun</td>
<td>53</td>
</tr>
<tr>
<td></td>
<td>Name</td>
<td>30</td>
</tr>
</tbody>
</table>
tor. An additional difference exists between "she" and "her". The pronoun "her" has more than one possible syntactic category (noun, e.g., "I saw her," and speciﬁer, e.g., "I saw her purse."). and "she" has only one possible syntactic category (noun, e.g., "She is here."). However, it is unlikely that this level of ambiguity inﬂuenced the pattern of skipping in the experiment, as readers in an experiment reported by Cliffton, Kennison, and Albrecht (1997) did not show an effect of ambiguity when reading sentences containing the ambiguous pronoun "her" and the unambiguous pronouns "his" and "him". Rather, it is likely that this difference in the probability of ﬁxing the pronoun may reﬂect the central importance of pronouns in subject position to the comprehension of coherent discourse.

**Reading Time.** Reading time on passages was analyzed for three regions: (1) the "pre" region, which contained the first two sentences of each passage; (2) the "S-V-O" region, which contained the subject and object referring expressions; and (3) the "remainder" region, which contained the remainder of the passage. First pass and total reading time were analyzed both as raw reading time in milliseconds and in mean deviation from predicted first pass and total reading time in milliseconds using the length of the region as the predictor variable in regression equations computed for each subject (i.e., mean residual in milliseconds) (see Cliffton & Ferreira, 1987; Trueswell, Tanenhaus, & Garney, 1994). This latter method permits reading time on regions differing in length in characters to be directly compared. For the present experiment, this type of analysis is preferred as reading time is compared on regions containing pronouns and proper names, which differ in length.

Signiﬁcant results were observed only for the "S-V-O" region. Table 2 presents mean reading time results for the "S-V-O" region by condition. ANOVAs were conducted using type of referent in subject position and type of referent in object position as within-participant factors. Reading time was inﬂuenced by an interaction of type of referent in subject position and type of referent in object position, stemming from the fact that the reading time difference between proper name and pronoun conditions was larger for subjects than for objects. The interaction was signiﬁcant in mean ﬁrst pass reading time (by items), F(1,19) = 2.24, MSE = 13616, p < .16, F(1,39) = 4.42, MSE = 21254, p < .05, in mean deviation from predicted ﬁrst pass (by items), F(1,19) = 3.77, MSE = 7452, p < .07, F(1,39) = 7.12, MSE = 8983, p < .02, in mean total reading time, F(1,19) = 4.07, MSE = 15939, p < .06, F(1,39) = 5.82, MSE = 28557, p < .03, and in mean deviation from predicted total time, F(1,19) = 7.02, MSE = 9912, p < .02, F(1,39) = 14.80, MSE = 10085, p < .001. Reading time was signiﬁcantly longer for conditions in which the subject was a proper name than for conditions in which the subject was a pronoun, resulting in signiﬁcant main effects of type of referent in subject position in mean first pass reading time, F(1,19) = 17.84, MSE = 10480, p < .001, F(1,39) = 40.07, MSE = 10204, p < .001, in mean deviation from predicted first pass, F(1,19) = 16.59, MSE = 5921, p < .001, F(1,39) = 19.15, MSE = 10154, p < .001, in mean total reading time, F(1,19) = 20.10, MSE = 11910, p < .001, F(1,39) = 44.00, MSE = 11684, p < .001, and in mean deviation from predicted total time, F(1,19) = 19.29, MSE = 7336, p < .001, F(1,39) = 32.69, MSE = 8614, p < .001. Reading time was somewhat longer for conditions in which the object was a proper name than for conditions in which the object was a pronoun. However, the main effect of type of referent in object position was not signiﬁcant in any measure: first pass reading time, F(1,19) = 3.95, MSE = 17389, p < .07, F(1,39) = 2.06, MSE = 41212, p < .16, in mean deviation from predicted ﬁrst pass reading time, Fs < 1, in total time, F(1,19) = 1.43, MSE = 23199, p < .25, F(2, 1) < 1, and in mean deviation from predicted total reading time, Fs < 1.

This pattern of reading times shows that the repeated-name penalty, previously observed using self-paced reading time methodology, is also observed using eye-tracking methodology. In particular, Experiment 1 of Gordon et al. (1993) found that reading times were elevated when an entity repeated in subject position was realized with a name rather than a pronoun. In contrast, no such elevation was observed for entities in object position. This exact same pattern was observed in the present study using eye-tracking methodology. This consistency across methods provides some assurance that the results that we have observed previously (Gordon et al., 1993; Gordon & Chan, 1995; Gordon & Searce, 1995) are not due to strategies peculiar to self-paced reading methodology.

### Table 2

<table>
<thead>
<tr>
<th>Type of Referent in Subject position</th>
<th>Type of Referent in Object position</th>
<th>Pronoun</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Mean First pass Reading Time (Mean Residual)</td>
<td>627 (-275)</td>
<td>693 (-244)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>771 (-169)</td>
<td>822 (-152)</td>
<td></td>
</tr>
<tr>
<td>B. Mean Total Reading Time (Mean Residual)</td>
<td>682 (-265)</td>
<td>736 (-247)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>850 (-129)</td>
<td>878 (-145)</td>
<td></td>
</tr>
</tbody>
</table>
Regressive Eye Movements. During comprehension, processing difficulty is sometimes associated with rereading initiated when a regressive eye movement is made from a region of text to a preceding region of text (Rayner & Pollatsek, 1989). The pattern of regressive eye movements made following the "S-V-O" region provides further evidence that processing difficulty was greater in the subject-name condition than in subject-pronoun condition. As some of the passages contained short words immediately following the object name or pronoun which were likely to be occasionally skipped, we scored the two words following the object name or pronoun as a region. Table 3 displays the mean percentage of trials on which a regressive eye movement was made from the two-word region following the "S-V-O" region. Significantly more regressive eye movements were made out of this region when the subject had been a proper name than when the subject had been a pronoun, resulting in a significant main effect of type of referent in subject position, $F_{1}(1,19) = 5.39, MSE = 22, p < .04$, $F_{2}(1,39) = 4.33, MSE = 65, p < .05$. There was no significant difference in the percentage of regressive eye movements when the object had been a proper name versus when the object had been a pronoun, $F_{1}(1,19) = 1.80, MSE = 40, p < .20$, $F_{2}(1,39) = 1.25, MSE = 89, p < .28$. The sentence position by type of referent interaction was significant (by subjects), $F_{1}(1,19) = 7.14, MSE = 27, p < .02$, $F_{2}(1,39) = 3.83, MSE = 97, p < .06$.

This pattern of regressive eye movements confirms that reading comprehension is disrupted by the use of a name rather than a pronoun in subject position, but not in object position. The finding that realizing the subject as a name not only causes elevation of reading times in the immediate vicinity of the critical referring expression but also causes an increase in regressive fixations indicates that the type of referent has both immediate and long-term effects on the interpretation of referring expressions and on the integration of utterances into a coherent discourse.

Summary. The results of this experiment show that eye tracking provides a view of the disruptive effects of name repetition that is consistent with that which

<table>
<thead>
<tr>
<th>Type of Referent in Object Position</th>
<th>Type of Referent in Subject Position</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pronoun</td>
<td>Pronoun</td>
</tr>
<tr>
<td>3.45</td>
<td>4.40</td>
</tr>
<tr>
<td>Name</td>
<td>10.55</td>
</tr>
</tbody>
</table>

has been previously observed with self-paced reading (Gordon et al., 1993). A repeated name penalty, in the form of elevated reading times, was observed when the subject of the critical sentence was realized as a name rather than a pronoun, while no such repeated name penalty was observed for the direct object. In addition, eye tracking goes beyond self-paced reading by showing that the probability of fixating a referring expression varies as a function of its form and grammatical role, and that regressive eye fixations provide one component of the repeated name penalty.

EXPERIMENT 2

The goal of Experiment 2 was to obtain evidence that the repeated-name penalty observed in the previous experiment was due to how the sentence containing the critical reference fit into the discourse and was not simply due to the characteristics of the sentence itself. Gordon et al. (1993) showed this using self-paced reading by examining the effect of center transitions on the repeated-name penalty. Centering theory (Grosz et al., 1983, 1995) contrasts situations where the backward-looking center is continued from one utterance to the next with situations where the backward-looking center is shifted, as in (3). In the first case, the backward-looking center of an utterance is the same as that of the preceding utterance, in the second case it is different.

3. Initial Passage Fragment
Sue knew that Tom wanted the St. Bernard puppy in the pet store. She offered to buy it for him as a Christmas present.

Continue Condition
She/Sue said that St. Bernards generally make wonderful pets. However, taking care of an animal is a big responsibility.

Shift Condition
He/Tom said that St. Bernards generally make wonderful pets. However, taking care of an animal is a big responsibility.

Alternative Initial passage Fragment
Tom told Sue about the St. Bernard puppy in the pet store. He reminded her that Christmas was right around the corner.

Continue Condition
He/Tom said that St. Bernards generally make wonderful pets. However, taking care of an animal is a big responsibility.
Shift Condition
She/Sue said that St. Bernards generally make wonderful pets. However, taking care of an animal is a big responsibility.

According to centering theory, pronominal realization of the backward-looking center is a less effective cue to discourse coherence in the case of a shift than in the case of a continue. The Gordon and Hendrick model (1997, in press) accounts for this difference through the differential effects of syntactic prominence on the establishment of coreferential interpretation by the construction rules for pronouns and equivalence; a continue involves reference to a prominent entity while a shift involves reference to a non-prominent entity. Gordon et al. (1993) provided empirical support for these accounts by showing that the repeated-name penalty was reduced when the subject of a sentence realized an entity that was not the backward-looking center of the preceding sentence.

The experiment examined eye movements during the reading of such passages. The first sentence introduced two named characters, one female and one male, with one as the subject of the sentence and the other as a postverbal object. The second sentence realized each of these characters as a pronoun and placed them in the same grammatical roles that they had in the first sentence (subject or postverbal object). In the third sentence, the subject was the same as in the two preceding utterances (continue condition) or the subject was the character who had previously been realized as a postverbal object (shift condition). Type of referent (pronoun vs. name) was also manipulated so that the repeated-name penalty could be assessed in both the continue and shift conditions. The fourth sentence did not realize either of the characters and was the same in every condition. For counterbalancing purposes, alternative initial passage fragments were used so that the same critical third sentence could appear in either the continue or shift conditions depending on how it was paired with an initial passage fragment.

Method
Participants. Twenty-four undergraduate and graduate students at the University of Massachusetts, Amherst, were recruited to participate in the experiment for course credit or for $5.00. All participants were fluent in American English, had normal or corrected vision (soft contact lenses only), and were naive to the purpose of the experiment.

Apparatus. The same apparatus was used as in Experiment 1.

Procedure. The same procedure was used as in Experiment 1. Each session lasted between 30–60 minutes.

Materials. Forty experimental passages were constructed as described above. The average length of proper names was 3.9 characters, with a range of two to five characters. For the purposes of counterbalancing, four versions of each passage had the male name ordered before the female name, and the other four versions had the female name ordered before the male name. Eighty short passages were included as fillers. Filler passages contained a variety of proper names and pronouns.

Results and Discussion
Approximately one percent of trials were removed due to track loss. Remaining trials were screened for false fixations and analyzed as described in Experiment 1. ANOVA's were conducted using type of center transition (continue vs shift) and type of referent (pronoun vs name) as within-participant factors.

 Skipping Pronouns and Proper Names. As shown in Table 4, pronouns and proper names were again frequently skipped. Pronouns were skipped significantly more often than proper names, resulting in a significant main effect of type of referent, \( F_{(1,23)} = 85.73, MSE = 132, p < .001 \), \( F_{(1,39)} = 42.14, MSE = 455, p < .001 \). The main effect of type of center transition was not significant, \( F_{(1,23)} = 3.53, MSE = 168, p > .07 \), \( F_{(1,39)} = 2.21, MSE = 300, p > .14 \). The interaction between center transition and type of referent was also not significant, \( F_{(1,23)} < 1 \), \( F_{(1,39)} = 1.41, MSE = .264, p > .24 \).

The finding that pronouns are skipped more often than names is consistent with the results of Experiment 1; as we noted it is potentially due to the greater length and/or lower printed frequency of the names. The finding that pronouns were skipped at the same rate regardless of whether they continued or shifted the subject of the preceding sentence indicates that the probability of skipping a pronoun in subject position does not depend on its contribution to discourse coherence. This finding is interesting in comparison to the results of Experiment 1 where we found that subject pronouns were skipped more frequently than direct object pronouns. Together these findings lead us to the conclusion that the frequency of skipping pronouns in both subject and direct object position should be
explained in terms of their lexical characteristics or in terms of processes required for the comprehension of different syntactic constructions, not in terms of properties of discourse coherence.

**Reading Time.** Three analysis regions were defined: (1) the “pre” region, which contained the first two sentences of each passage; (2) the “subject+2” region, which contained the pronoun or proper name and the following two words; and (3) the “remainder” region, which contained the remainder of the passage. The “subject+2” region was scored in this way because the subject pronoun or name was sometimes followed by short, high frequency words that were likely to be occasionally skipped. Furthermore, this region was comparable in length to the “S-V-O” region analyzed in Experiment 1. First pass and total reading time was analyzed both as raw reading time in milliseconds and in mean deviation from predicted reading time using region length in characters as the predictor variable in regression equations computed for each subject (i.e., mean residual in milliseconds).

Significant results were not observed at the “pre” region. The primary results were observed at the “subject+2” region. Table 5 displays mean reading time results for the “subject+2” region by condition. Reading time in passages containing proper names in continue conditions took longer to read than those containing pronouns in continue conditions, while this difference was not observed for passages in the shift conditions. The statistical significance of this pattern is shown by the interaction between type of center transition and type of referent, which was significant in first pass reading time (by participants). $F_{1}(1,23) = 5.06, MSE = 10596, p < .04, F_{2}(1,39) = 2.36, MSE = 36751, p < .14$ and in mean deviation from predicted first pass, $F_{1}(1,23) = 4.52, MSE = 9110, p < .05, F_{2}(1,39) = 6.74, MSE = 20661, p < .02$. It was also significant in mean total reading time (by participants), $F_{1}(1,23) = 3.85, MSE = 7796, p < .07, F_{2}(1,39) = 1.11, MSE = 98198, p < .30$, but not in mean deviation from predicted total reading time, $F_{1}(1,23) = 3.10, MSE = 7447, p < .10, F_{2}(1,39) = 1.60, MSE = 41526, p < .22$. The main effects of type of referent was significant in first pass reading time, $F_{1}(1,23) = 3.30, MSE = 13657, p < .04, F_{2}(1,39) = 6.65, MSE = 17921, p < .02$, mean deviation from predicted first pass reading time, $F_{1}(1,23) = 3.77, MSE = 13602, p < .07, F_{2}(1,39) = 4.66, MSE = 27360, p < .05$, in total reading time, $F_{1}(1,23) = 4.45, MSE = 18220, p < .05, F_{2}(1,39) = 7.42, MSE = 21259, p < .01$, and in mean deviation from predicted total reading time, $F_{1}(1,23) = 3.49, MSE = 14638, p < .08, F_{2}(1,39) = 7.42, MSE = 21259, p < .01$. The main effect of type of center transition was not significant in any analysis, $F_{s} < 1$.

Reading time on the “remainder” region indicated significant results only in mean deviation from predicted total reading time. A significant main effect of type of referent was observed, as proper name conditions were generally read more slowly than pronoun conditions, $F_{1}(1,23) = 5.16, MSE = 24404, p < .04, F_{2}(1,39) = 5.96, MSE = 81068, p < .02$. For this measure, the main effect of type of referent and the interaction were not significant, $F_{s} < 1$.

This pattern of reading times again shows that the repeated-name penalty, as previously observed using self-paced reading time methodology, is also observed using eye-tracking. Experiment 4 of Gordon et al. (1993) showed that the elevation for reading times resulting from using a name rather than a pronoun (the repeated-name penalty) is reduced for entities in subject position that do not continue the subject entity of the preceding sentence. The current experiment showed the same pattern of reduction in repeated-name penalty as measured in the region of text including and immediately following the subject NP. This indicates that at least a portion of the repeated-name penalty occurs in the immediate vicinity of the critical referring expression and that this portion is influenced by discourse structure as it is determined by the relationship between the syntactic structures of successive sentences. These results contradict Gernsbacher’s (1989) prediction that repeated name anaphors should generally take less time to read than pronoun anaphors.

**Regressive Eye Movements.** Table 6 shows the mean percentage regressions out of the 2 word region following the “subject+2” region, which was four words to the right of the critical referring expression (i.e., the pronoun or name). This region was chosen because its distance from the subject referring expression roughly matches that of the region analyzed in Experiment 1. More regressive eye

<table>
<thead>
<tr>
<th>TABLE 5</th>
<th>Mean Reading Time in Milliseconds (and Mean Residual in Milliseconds) for the “subject+2” Region by Condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type of Center Transition</td>
<td>Type of Referent</td>
</tr>
<tr>
<td>A. Mean First Pass Reading Time (Mean Residual)</td>
<td>Continue</td>
</tr>
<tr>
<td></td>
<td>Shift</td>
</tr>
<tr>
<td>B. Mean Total Reading Time (Mean Residual)</td>
<td>Continue</td>
</tr>
<tr>
<td></td>
<td>Shift</td>
</tr>
</tbody>
</table>
movements were observed in the shift condition than the continue condition, resulting in a significant main effect of type of center transition, \( F(1,23) = 5.34, \text{MSE} = 162, p < .04; F(1,39) = 5.16, \text{MSE} = 90, p < .03 \). Neither the main effect of type of referent nor the interaction between type of referent and type of center transition was significant.

This pattern of regressive eye movements provides further insight into the effects of discourse structure on language comprehension. First, the finding that the type of referent does not influence the likelihood of regressive eye movements is surprising given the finding in Experiment 1 that the likelihood of regressive eye movements was influenced by the interaction of type of referent and syntactic role of the referring expression. The absence in the present experiment of an influence of type of referent on regressive eye movements suggests that readers may trade off reading time (which was influenced by type of referent) against the need for subsequent regressions. It is unclear what factors might influence readers' choices along such a tradeoff. Second, the finding that there were more regressive eye movements in the shift than in the continue condition is consistent with the hypothesis advanced within centering theory (Grosz et al., 1983, 1995) that shifting of discourse centers imposes a greater processing load than continuing discourse centers. However, the hypothesized general effect of center transitions has received only modest support in previous studies (Gordon et al., 1993). The finding in the present study that center transitions influence regressions but not reading times is consistent with the previous results indicating that the general effect of center transitions on comprehension is not powerful.

**GENERAL DISCUSSION**

The results of two eye-tracking experiments show that under specific circumstances repeating a name in a short passage disrupts reading comprehension as measured by reading time and also, in some instances, by regressive eye move-

ments. The pattern of disrupted comprehension observed here with eye tracking is similar to that observed previously using self-paced reading.

The present results add to a growing number of studies showing that similar results can be gained using eye tracking and self-paced reading (Clifton, Speer, & Abney, 1991; Trueswell, Tanenhaus, & Kello, 1993). As did the self-paced reading results of Gordon et al. (1993), the present results showed that the effects of referential processing predicted by centering theory can be observed during the initial processing of the referential expressions as well as during the processing of subsequent regions of passages. The undisputed advantage of eye tracking over self-paced reading stems from the fact that eye tracking can provide precise measurement of eye movement behavior during the normal reading process, unencumbered by manual key presses and the sequential and often non cumulative presentation of text. The results of Experiment 1 showed that the repeated name penalty for referents in subject position was observed during the initial processing of the referential expression, i.e., during first pass reading time, as well as during subsequent processing revealed by total reading time and when regressive eye movements were made from out of the region immediately following the referential expression to an earlier region of the passage. As predicted, no repeated name penalty was observed for referents in object position in any measure of reading behavior. In Experiment 2, the repeated name penalty was observed for referents in a continuing relation with referents in preceding sentences in both measures of initial and subsequent processing, i.e., first pass reading time and total reading time, respectively. As predicted, the repeated name penalty was not observed for referents in a Shifting Relation with referents in preceding sentences in any measure of reading behavior.

The present experiments also provided important new evidence regarding the pattern of skipping on referential expressions and how effects of discourse structure influence skipping. In Experiment 1, the percentage of passages on which referential expressions in subject position were skipped was significantly larger than the percentage of passages on which referential expressions in object position were skipped. These results suggest that either discourse factors or syntactic factors, in addition to word-level factors, such as word length, printed frequency, and predictability, may influence word skipping. The results of Experiment 2 show that the skipping of referential expressions in subject position was not influenced by whether the referential expression was in a continuing or shifting relation with referents in preceding sentences. This suggests that syntactic factors rather than discourse factors influence word skipping. The exact way in which syntactic and discourse factors influence word skipping merits further investigation.

The pattern of elevated reading times attributable to repeated names is consistent with characterizations provided by centering theory (Grosz et al., 1983, 1995) regarding what constitutes a well-formed discourse segment. This pattern can be
explained by Gordon and Hendrick's (1997, in press) processing model of the interpretation of reference and coreference in discourse. Because it provides a detailed account of the comprehension process, this model can be tested and refined using eye-tracking methodology in further efforts to gain an understanding of the nature of language comprehension.

Acknowledgement: This research was financially supported by Grants HD-07327 and HD-18708 awarded to the University of Massachusetts, Amherst, and Grant MH-19554 awarded to the University of Illinois, Urbana-Champaign, and by NSF Grant IRI-94-04756.

REFERENCES


Kennison, S.M. & Clifton, C. (1995). Determinants of parafoveal preview benefit in high and low working memory capacity readers: Implications for Eye Movement Con-
I propose that research on the role of common ground in language use is often insensitive to considerations of parsimony. By ignoring parsimony, our experiments might fall short of addressing common ground and instead the resulting data support a simpler theory which is subsumed within the theory of common ground. In order to address this inherent conceptual confound, this paper proposes the subsuming theory criterion for experiments on common ground and mutual knowledge. In practical terms, to demonstrate an effect of common ground the design must keep common information constant and only vary whether or not it is common. The paper demonstrates how consistent use of this necessary criterion will allow us to make stronger claims about when common ground plays a role and when it does not. To illustrate how experiments can be designed to satisfy the criterion, the paper evaluates some earlier studies which do and don’t satisfy it. In general, the study of language use could benefit from tightening the rules of evidence, allowing us to draw stronger conclusions about central concepts such as common ground.

The relationship between data and theory in science is fairly complex. Rarely do experiments provide data which uniquely support a particular theory, but instead data are typically interpretable in more than one way. In order to improve our ability to interpret data and select among potential explanatory theories, we rely on accepted “rules of evidence.” One of the most central rules of evidence is the principle of parsimony. If an experiment provides evidence which could support two alternative theories, and one theory is clearly more parsimonious than the other, then in general one should consider the experiment as supporting the more parsimonious theory. In other words, to the extent that experiments affect our beliefs, the experiment should not affect the strength of our belief in the less parsimonious theory. While parsimony is taken for granted in the general practice of

Direct all correspondence to: Boaz Keysar, Department of Psychology, The University of Chicago, 5848 South University Avenue, Chicago, IL 60637 <boaz@cep.uchicago.edu>