Pupillary responses reflect ambiguity resolution in pronoun processing

Margreet Vogelzang\textsuperscript{a,b}, Petra Hendriks\textsuperscript{a} and Hedderik van Rijn\textsuperscript{b}

\textsuperscript{a}Center for Language and Cognition Groningen, University of Groningen, Groningen, The Netherlands; \textsuperscript{b}Department of Experimental Psychology, University of Groningen, Groningen, The Netherlands

\section*{ABSTRACT}
The resolution of ambiguous pronouns is influenced by the preceding linguistic discourse. This raises the question whether the processing of an object pronoun is also influenced by the preceding sentential subject. In an experiment with Dutch adults, we recorded pupil dilation as a measure of the cognitive effort involved in resolving pronominal versus full noun phrase (NP) subjects and pronominal versus reflexive objects. Our results indicate that more effort is needed to resolve a pronominal subject or object compared to a less ambiguous full NP subject or reflexive object. These results support the hypothesis that the ambiguity of a referring expression influences processing. Contrary to our expectations, no evidence was found that the form of the subject influences the processing of a subsequent pronominal object. We conclude that pupillary responses reflect ambiguity resolution in pronoun processing, and that the process to resolve pronouns commences as soon as the pronoun is encountered.

\section*{Introduction}
Pronouns are a highly flexible means in language to refer to individuals that have been mentioned earlier in the linguistic discourse. Although they introduce potential ambiguity, listeners usually resolve pronouns correctly on the basis of available information. Several semantic and discourse factors have been found to influence pronoun resolution, including grammatical role parallelism (Branigan, Pickering, Liversedge, Stewart, & Urbach, 1995; Frazier, Taft, Roeper, Clifton, & Ehrlich, 1984; Smyth, 1994), gender information (Arnold, Eisenband, Brown-Schmidt, & Trueswell, 2000; Badecker & Straub, 2002; Ehrlich, 1980), antecedent prominence, accessibility, and topicality (Arnold et al., 2000; Chafe, 1976; Cunnings, Patterson, & Felser, 2014; Givón, 1983; Järvi-Kivi, Pyykkönen-Klauck, Schimke, Colonna, & Hemforth, 2014; Spener, Smits, & Hendriks, 2009; van Rijn, Van Rijn, & Hendriks, 2013), and interference of prominent competitor antecedents (Badecker & Straub, 2002; Clackson, Felser, & Claessen, 2011). In addition, syntactic constraints play a role (Hendriks & Spener, 2006; Reinhart & Reuland, 1993). These factors interact in a complex manner and suggest that pronoun resolution is a cognitively effortful process.

In this paper, we investigate the cognitive effort involved in pronoun processing. Previous studies have typically investigated subject or object pronoun processing in isolation. However, as the linguistic discourse can influence the processing of both subject pronouns and object pronouns, it is conceivable that information provided by the sentential subject contributes to the resolution of a pronominal object, influencing the cognitive effort needed to resolve this object.

Because syntactic, semantic, and discourse factors might influence pronoun processing without affecting the final interpretation, online measures of cognitive effort are useful for investigating pronoun resolution (Kaiser, 2011). For example, prominence of the antecedent in the discourse (Van Rijn, 2012) and gender information (Arnold et al., 2000; Clackson et al., 2011) has been found to affect pronoun processing, while leaving the final interpretation unaffected. Pupillary dilation measures are a way to investigate cognitive effort in cognitive processing non-invasively and online, although slightly delayed (e.g. Van Rijn, Dalenberg, Borst, & Spener, 2012). Measures of pupil dilation reliably reflect cognitive effort in language processing (Engelhardt, Ferreira, & Patsenko, 2010; Just & Carpenter, 1993), and have been shown to reveal subtle differences in effort in processes that are hard to distinguish by other means (e.g. Wierda, van Rijn, Taatgen, & Martens, 2012) and in processes that do not feel effortful to listeners (e.g. Wolff, Scholz, Akyürek, & van Rijn, 2015), such as pronoun resolution.

To investigate the cognitive effort involved in pronoun processing, we performed an eye-tracking...
experiment with a referent selection task. While recording pupil dilation, we examined whether the specific form of the sentential subject (a pronoun such as he, compared to a full noun phrase (NP) such as the hedgehog) influences the interpretation and processing of a pronominal object (a pronoun such as him, compared to the reflexive himself).

The experiment tests three main hypotheses. Firstly, we predict that ambiguity as defined by the number of potential referents influences pronoun processing, and that processing less ambiguous referring expressions requires less cognitive effort than processing more ambiguous referring expressions. Whereas full definite NPs typically unambiguously refer to a particular referent, pronouns can refer to various individuals that are salient in the context. Likewise, whereas reflexive objects unambiguously refer to the sentential subject, pronominal objects can refer to various salient individuals that are not the sentential subject. Although these two constraints on reflexive and pronominal objects are both considered to be part of the grammar (e.g. Chomsky, 1981; Reinhart & Reuland, 1993), the potential ambiguity of pronominal objects compared to reflexive objects – even when there are only two potential referents – can be witnessed by English- and Dutch-speaking children’s interpretations of these forms (e.g. Chien & Waxler, 1990), the initial stage of adults’ online sentence processing (e.g. Badecker & Straub, 2002), and adults’ offline interpretations in specific discourse contexts (Grodzinsky & Reinhart, 1993). So, both subject and object pronouns can have multiple potential antecedents and are therefore more ambiguous than NPs and reflexives. When a pronoun has multiple potential antecedents, resolution is increasingly difficult and error prone (Ariel, 1990; see also Clackson et al., 2011). We expect that more cognitive effort is needed when processing a pronominal subject or object compared to a non-ambiguous subject NP or reflexive object (cf., e.g. Hendriks, Van Rijn, & Valkenier, 2007; Van Rij, Van Rijn, & Hendriks, 2010; Van Rij et al., 2013), a difference that should be reflected in fewer topic-continuation interpretations in referent selection and in longer response times (RTs).

Our second hypothesis is that the discussed differences between less and more ambiguous referring expressions are reflected in pupil dilation. We will examine whether the relatively novel technique of measuring pupillary responses can serve as a useful tool in psycholinguistic research on processing effort. Specifically, we predict that the cognitive effort needed for ambiguity resolution in pronoun processing is indexed by pupillary responses, with more ambiguous referring expressions eliciting increased pupillary responses compared to less ambiguous referring expressions.

Finally, we hypothesise that the form of the sentential subject influences the processing of a pronominal object. There is some evidence that pronouns are processed quickly and online (a.o. Arnold et al., 2000; Badecker & Straub, 2002). If a subject pronoun is completely resolved when encountered, the form of the subject should not influence the processing of a subsequent pronominal object. If, on the other hand, a pronominal subject is only preliminarily resolved at first encounter and final resolution is delayed (cf. McDonald & MacWhinney, 1995; Millis & Just, 1994; Stewart, Holler, & Kidd, 2007), this pronominal subject still has multiple potential antecedents. In that case, a subsequent object pronoun also has multiple potential antecedents and cannot be resolved though a simple process of elimination. If this holds, the delayed resolution of the subject pronoun may interfere with the resolution of a subsequent object pronoun. We predict that pronouns are not completely resolved at first encounter, and that this delayed processing will influence the processing of a subsequent pronominal object. Thus, we predict that a pronominal object is more effortful to resolve when following a subject pronoun than when following a subject NP. This effect should be reflected by interactions between the subject form and the object form in the RTs and the pupillary responses.

Methods

Participants

Forty undergraduate psychology students from the University of Groningen (16 men; mean age 19.7; SD = 1.4) participated in the experiment for course credits. All participants were native speakers of Dutch and had normal or corrected-to-normal vision and hearing. The ethical committee of the Psychology Department of the University of Groningen approved the experiment (code 12171-N).

Materials

The experiment used 120 Dutch stories consisting of three clauses each (for an example story, see Table 1). The third clause was presented in four combinations: the subject was either a full NP (e.g. de egel “the hedgehog”) or the pronoun hij “he” and the object was either the reflexive zich “himself” or the pronoun hem “him”.

Each story had the same structure and started with the introduction of a first referent. In the second clause, this referent was again mentioned as the subject, thus establishing its status as the discourse topic, and a second referent was introduced in a less
prominent position. In a third clause introduced by the temporal conjunction terwijl ("while", "during the time that"), two anaphoric expressions appeared as the subject and the object of the clause, followed by a finite verb. Because the verb followed the anaphoric subject and object, its influence on the online processing of the anaphoric expressions was limited. The Appendix lists the transitive and reflexive verbs used in the third clause of the stories.

After each story a referent selection question was presented. Two types of questions were used: subject questions and object questions. Subject questions probed participants’ interpretation of a referring expression in subject position, and object questions probed participants’ interpretation of a referring expression in object position. Example questions for the different conditions in Table 1 are shown in Table 2, including the target responses. For subject questions, target responses took the most prominent referent (the discourse topic, i.e. the hedgehog) as the antecedent of the subject and of the reflexive object. As grammatical constraints on pronominal binding disallow object pronouns to refer to the local subject, the object pronoun is predicted to be resolved as the referent in object position in the second clause (i.e. the mouse). The target response of object questions in stories with a reflexive is also the referent in object position in the second clause (the mouse). Note that these latter questions refer to the second clause of the story, whereas all other questions refer to the third clause.

### Design

The experiment used a 2 × 2 within-subjects design with the conditions subject form (full NP vs. pronoun) and object form (reflexive vs. pronoun). The experiment consisted of four blocks of 30 trials each, and took approximately 60 minutes.

The combinations of the conditions and the two question types resulted in 2 × 2 × 2 = 8 lists. Condition, question type, and the location of the target response on the screen (left or right) were counterbalanced across lists. The order of presentation of the stories within a list was randomised for each participant.

### Procedure

Participants were tested individually using an EyeLink 1000 (SR Research) eye-tracker to measure pupil dilation by means of the pupil area and a Microsoft Sidewinder gamepad to record responses. Participants heard the stories and questions through speakers and were instructed to listen carefully, blink between trials, and blink as little as possible during trials.

### Table 1. Example story in Dutch (with English glosses).

<table>
<thead>
<tr>
<th>Clause</th>
<th>Subject form</th>
<th>Object form</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. De egel heeft een boomhut gebouwd</td>
<td>de egel</td>
<td>de egel</td>
</tr>
<tr>
<td>2. Afgelopen dinsdag liep de egel met de muis door het bos naar huis,</td>
<td>de muis</td>
<td>de muis</td>
</tr>
</tbody>
</table>

Note: The complete list of stimuli used in the experiment can be found on [http://let.webhosting.rug.nl/~vogelzang/experiments.html](http://let.webhosting.rug.nl/~vogelzang/experiments.html) or by contacting one of the authors. The mean sentence length of the test sentence from the onset of the subject until the end of the sentence varied between conditions. The mean length of sentences with a full NP in subject position and a reflexive in object position was 2821 ms (SD = 423), with a full NP in subject position and a pronoun in object position 2770 ms (SD = 458), with a pronoun in subject position and a reflexive in object position 2456 ms (SD = 392), and with a pronoun in subject position and a pronoun in object position 2392 ms (SD = 451).

### Table 2. Questions following the story in Table 1 in Dutch (with English glosses).

<table>
<thead>
<tr>
<th>Object form</th>
<th>Question type</th>
<th>Question</th>
<th>Target response</th>
<th>Story condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>reflexive</td>
<td>subject</td>
<td>Wie haastte zich?</td>
<td>de egel</td>
<td>3a, 3c</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Who hurried himself?</td>
<td>the hedgehog</td>
<td></td>
</tr>
<tr>
<td></td>
<td>object</td>
<td>Met wie liep iemand door het bos naar huis?</td>
<td>de muis</td>
<td>3a, 3c</td>
</tr>
<tr>
<td></td>
<td></td>
<td>With whom walked someone through the forest to home?</td>
<td>the hedgehog</td>
<td></td>
</tr>
<tr>
<td>pronoun</td>
<td>subject</td>
<td>Wie volgde iemand?</td>
<td>de egel</td>
<td>3b, 3d</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Who followed someone?</td>
<td>the hedgehog</td>
<td></td>
</tr>
<tr>
<td></td>
<td>object</td>
<td>Wie werd gevolgd?</td>
<td>de muis</td>
<td>3b, 3d</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Who was being followed?</td>
<td>the mouse</td>
<td></td>
</tr>
</tbody>
</table>

For the different subject forms, the same questions could be used. For the different object forms, different questions had to be used. The story condition indicates with which story condition in Table 1 the question could be used.
Figure 1. The structure of presentation of the stimuli during one trial. Each trial started by showing a fixation cross for 1500 ms, before the referents appeared on the screen. The story started 1500 ms after presentation of the referents. The story was followed by a 1100 ms pause, followed by the question. After the question, the participant could respond by selecting one of two referents on the screen. The selected referent was highlighted with a black frame for 1000 ms. Finally, the participant was given the opportunity to blink for 3000 ms before the next trial started.

Figure 1 shows the trial structure. During a practice block of six trials and during the main experiment, the two referents in each story were presented as pictures on the screen; one on the left side and the other on the right. Participants could answer the questions by selecting one of the referents on the screen through the left or right button on the gamepad. No feedback on the correctness of the response was provided; only a black frame was displayed around the selected referent to indicate that the response was recorded.

Results

Trials that contained more than 25% missing data after blink removal in the relevant time window of the pupillary data were removed. The missing data points of the remaining trials were linearly interpolated. No additional calculations or modifications were performed on the pupil dilation data. One trial was excluded because the RT was longer than 10 sec. In total, 618 trials (12.9%) were excluded, and 4182 trials remained for analysis. All analyses (responses, RTs, pupillary data) were performed on the same data.

Responses

In all conditions, the target response was indeed the most frequently selected referent. Figure 2 shows the percentages of target responses for each subject/object combination.

The responses were analysed using a binomial generalised linear mixed-effect-based model with, based on the theoretical predictions, subject form, object form, and question type as fixed effects. Based on model comparisons (cf. Baayen, Davidson, & Bates, 2008), random deviation coding was used with SubjectForm:NP, ObjectForm:Reflexive, and QuestionType:Subject coded as −0.5 and SubjectForm:Pronoun, ObjectForm:Pronoun, and QuestionType:Object coded as 0.5.

Table 3. An overview of the estimates (β), the z-scores, and the p-values of the fixed factors entered in the generalised linear mixed-effect model for the analysis of the responses.

<table>
<thead>
<tr>
<th>Predictors</th>
<th>Estimate (β)</th>
<th>SE</th>
<th>z</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>3.925</td>
<td>0.226</td>
<td>17.406</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>SubjectForm</td>
<td>−1.607</td>
<td>0.292</td>
<td>−5.501</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>ObjectForm</td>
<td>−0.735</td>
<td>0.304</td>
<td>−2.42</td>
<td>.016</td>
</tr>
<tr>
<td>QuestionType</td>
<td>−0.645</td>
<td>0.292</td>
<td>−2.212</td>
<td>.027</td>
</tr>
<tr>
<td>SubjectForm × ObjectForm</td>
<td>0.679</td>
<td>0.583</td>
<td>1.164</td>
<td>.245</td>
</tr>
<tr>
<td>SubjectForm × QuestionType</td>
<td>1.649</td>
<td>0.584</td>
<td>2.825</td>
<td>.005</td>
</tr>
<tr>
<td>ObjectForm × QuestionType</td>
<td>1.610</td>
<td>0.584</td>
<td>2.759</td>
<td>.006</td>
</tr>
<tr>
<td>SubjectForm × ObjectForm × QuestionType</td>
<td>−3.267</td>
<td>1.168</td>
<td>−2.798</td>
<td>.005</td>
</tr>
</tbody>
</table>
intercepts for participants and items, and random slopes for object form within items were included. The results of the model are presented in Table 3.

The results show the predicted main effects of subject and object form, indicating that participants interpreted subject NPs more often as referring to the discourse topic than subject pronouns, and interpreted reflexive objects more often as referring to the discourse topic than pronominal objects.

A main effect of question type was found, as well as interactions between subject form and question type and between object form and question type. Post hoc tests indicate that the effect of question type ($\beta = 0.17; z = 0.73; p = .467$) and the interaction with subject form ($\beta = 0.01; z = 0.01; p = .999$) disappear when only object pronouns (i.e. omitting the reflexive objects) are included. Thus, the effects of question type are caused by the deviating object questions with a reflexive, as shown in Figure 2. No interaction was found between subject form and object form.

Lastly, a three-way interaction between subject form, object form, and question type was found. Figure 3 displays this three-way interaction. The figure visually shows that three conditions have similar effects, with one condition, depicted by the solid line in the right graph, clearly deviating. These results suggest that the observed three-way interaction is caused by the deviating object questions about stories with a reflexive.

Figure 3. Illustration of the reported three-way interaction between subject form, object form, and question type in the analysis of the responses. Note that object questions about stories with a reflexive, indicated by the solid line in the right graph, referred to the second clause of the story and therefore care should be taken when comparing the results from these questions to the other questions, which referred to the third (i.e. final) clause of the story. See main text for further discussion.

Figure 4. The mean RTs for each subject/object combination, per question type. Error bars are based on a 95% two-sided non-parametric confidence interval estimated through bootstrapping. Object questions about stories with a reflexive, indicated with diagonal lines, referred to the second clause of the story and cannot directly be compared to the other questions, which referred to the third (i.e. final) clause of the story.

Table 4. An overview of the estimates ($\beta$), the t- and $p$-values of the fixed factors entered in the linear mixed-effect model for the analysis of the RTs.

<table>
<thead>
<tr>
<th>Predictors</th>
<th>Estimate ($\beta$)</th>
<th>SE</th>
<th>df</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>6.190</td>
<td>0.057</td>
<td>44</td>
<td>108.326 &lt;.001</td>
<td></td>
</tr>
<tr>
<td>SubjectForm</td>
<td>0.147</td>
<td>0.037</td>
<td>36</td>
<td>3.987 &lt;.001</td>
<td></td>
</tr>
<tr>
<td>ObjectForm</td>
<td>0.142</td>
<td>0.041</td>
<td>119</td>
<td>3.462 &lt;.001</td>
<td></td>
</tr>
<tr>
<td>QuestionType</td>
<td>0.164</td>
<td>0.027</td>
<td>3894</td>
<td>5.972 &lt;.001</td>
<td></td>
</tr>
<tr>
<td>SubjectForm × ObjectForm</td>
<td>0.036</td>
<td>0.055</td>
<td>3892</td>
<td>0.651 .515</td>
<td></td>
</tr>
<tr>
<td>ObjectForm × QuestionType</td>
<td>-0.129</td>
<td>0.055</td>
<td>3890</td>
<td>-2.343 .019</td>
<td></td>
</tr>
<tr>
<td>SubjectForm × QuestionType</td>
<td>0.009</td>
<td>0.055</td>
<td>3895</td>
<td>0.171 .864</td>
<td></td>
</tr>
<tr>
<td>ObjectForm × QuestionType</td>
<td>0.279</td>
<td>0.110</td>
<td>3891</td>
<td>2.539 .011</td>
<td></td>
</tr>
</tbody>
</table>

Deviation coding was used with SubjectForm:NP, Objectform:Refl exive, and QuestionType:Subject coded as −0.5 and SubjectForm:Pronoun, Objectform:Pronoun, and QuestionType:Object coded as 0.5.
Importantly, post hoc tests show that the effects of subject form ($\beta = -2.42; z = -4.53; p < .001$) and object form ($\beta = -1.54; z = -2.89; p = .004$) persist when only subject questions are included, so these effects are independent of the deviating object questions.

**Response times**

The RTs were measured from the end of the question to the button press. Figure 4 shows the mean RTs for each subject/object combination. We analysed the RTs using a linear mixed-effect-based analysis of the log-transformed RTs with, based on the theoretical predictions, subject form, object form, and question type as fixed effects. Based on model comparisons (cf. Baayen et al., 2008), random intercepts for participants and items, and random slopes for subject form within participants and for object form within items were included. The degrees of freedom were approximated using the Satterthwaite method (as recommended by Bates, Maechler, Bolker, & Walker, 2014). The results of the model are presented in Table 4.

Main effects of subject and object form were found, indicating that RTs were shorter for stories with subject NPs than with subject pronouns, and shorter for stories with reflexive objects than with pronominal objects. Furthermore, subject questions were answered faster than object questions, an effect that persists when only object pronouns are included (i.e. omitting the reflexive objects, $\beta = 0.17; t(1898) = 4.32; p < .0001$). An interaction between subject form and question type was found, but post hoc tests indicate that this effect disappears when only object pronouns are included ($\beta = 0.02; t(117) = 0.21; p = .83$); it is thus influenced by the deviating object questions with reflexives.

Lastly, a three-way interaction between subject form, object form, and question type was found (see Figure 5). The results of the model are presented in Table 4.

We analysed the RTs using a linear mixed-effect-based analysis of the log-transformed RTs with, based on the theoretical predictions, subject form, object form, and question type as fixed effects. Based on model comparisons (cf. Baayen et al., 2008), random intercepts for participants and items, and random slopes for subject form within participants and for object form within items were included. The degrees of freedom were approximated using the Satterthwaite method (as recommended by Bates, Maechler, Bolker, & Walker, 2014). The results of the model are presented in Table 4.
are included, so these effects are independent of the deviating object questions.

**Pupil dilation data**

For the analysis of pupillary data, we focus on the time window from 300 ms before the onset of the subject in the third clause until the onset of the question that follows the third clause. Pupil dilation is aligned to the onset of the subject and is expressed as proportional change compared to a 100 ms baseline window, calculated for each trial separately, immediately before the onset of the subject. Figure 6 plots the proportional pupil dilation for each subject/object combination over the course of a trial, averaged over all trials.

The pupil dilation data were analysed using generalised additive models (GAMs; Wood, 2006). GAMs estimate the relationship between predictors and the dependent variable with a non-linear smooth function. The appropriate degrees of freedom for the smooth function are determined automatically through cross validation. GAMs allow for an analysis of pupil dilation data over time for different conditions within one analysis.

The model fitted to the pupil dilation data contained, based on the theoretical predictions, fixed effects for subject form, object form, and the interaction between subject and object form. Based on model comparisons, random slopes and intercepts for participants and items, an effect of trial order (centred), and an interaction between time within a trial and trial order was included. The model fits are plotted in Figure 7. The results of the model are presented in Table 5.

The results indicate that, in subject position, pronouns influence pupil dilation when compared to NPs. The left graph in Figure 7 shows a higher proportional pupil dilation with subject pronouns than with subject NPs, indicating increased cognitive effort. In object position, pronouns influence pupil dilation when compared to reflexives. Inspection of the right graph in Figure 7 shows that pronominal objects elicit an increased proportional pupillary response compared to reflexive objects. No interaction is plotted as no interaction was observed between subject pronoun and object pronoun.

**Discussion**

In this experiment, the effects of subject form and object form on processing effort were investigated through responses to comprehension questions and their associated RTs. Additionally, we investigated whether pupillary
responses are a viable measure of processing effort in ambiguity resolution.

The experiment tested three hypotheses. The first hypothesis was that less ambiguous referring expressions require less processing effort than more ambiguous referring expressions. We formulated two predictions on the basis of this hypothesis. Firstly, we predicted for sentential subjects that full NPs are less effortful to process than pronouns because, in contrast to pronouns, full NPs often have just one potential referent in the discourse. The results of the pupillary analysis support this prediction by showing increased dilation in sentences with a pronoun. Additionally, the responses to the interpretation questions and their associated RTs show that participants gave fewer target responses and were slower to respond to sentences with subject pronouns, compared to subject NPs. This suggests that, even though the referent of the subject was taken to be the expected discourse topic in the story, a subject pronoun is still more effortful to process than a subject NP. Interestingly, although a full subject NP in the test sentence resulted in the use of this form three times in a row, we found no evidence of a repeated name penalty (Gordon, Grosz, & Gilliom, 1993). Importantly, these findings were obtained in a discourse context featuring two referents. As pronoun resolution is influenced by the presence of competitor antecedents (Badecker & Straub, 2002; Clackson et al., 2011), the number of referents in the discourse might modulate the resolution effort of pronouns.

The same hypothesis also predicted that reflexive objects are less effortful to process than more ambiguous pronominal objects, because reflexive objects are required by grammatical constraints to refer to the sentential subject, whereas pronominal objects can have multiple potential antecedents. This prediction was supported by the results of the pupil dilation analysis, as this analysis demonstrated that the pronoun hem “him” elicits more pupil dilation than the reflexive sich “herself”. Moreover, in line with Clackson et al. (2011), the analyses of the responses and their RTs demonstrate that participants gave fewer target responses and were slower to respond to sentences with a pronoun, compared to sentences with a reflexive.

Our second hypothesis predicted that the cognitive effort needed for ambiguity resolution in pronoun processing is indexed by pupillary responses. The results support this hypothesis by showing that pupil dilation is a viable measure of processing effort in ambiguity resolution. Effects in the predicted direction were found for both subject form and object form, and these effects were consistent with the results found in the responses and the RTs. This shows that the relatively novel technique of measuring pupillary responses can serve as a useful tool in research on processing effort.

Finally, we hypothesised that the form of the sentential subject influences the processing of a pronominal object. The results of the pupil dilation measures, as well as the results of the responses and the RTs, did not show an effect of subject form on the processing of the pronominal object. These results do not support our hypothesis, and thereby do not provide support for the idea that a pronominal subject is only resolved preliminarily at first encounter. Rather, the results can be explained by the idea that the subject has already been interpreted by the time the object pronoun is encountered. Thus, the results suggest that the process to resolve pronouns commences as soon as the pronoun is encountered, and finishes in relatively little time (cf. a.o. Arnold et al., 2000; Badecker & Straub, 2002).

Overall, the results support our prediction that ambiguity as defined by the number of potential referents influences pronoun processing. Contrary to a secondary prediction, no indication was found that the form of the subject influences the processing of a subsequent object. However, as predicted, pupil dilation proved a successful tool for measuring online processing effort in language processing.

**Disclosure statement**

No potential conflict of interest was reported by the authors.

**References**


Appendix

The verbs used in the experiment are presented in Table A1. The verbs satisfy a number of criteria. Because of a planned parallel experiment in Italian, the verbs had to be preferably used either transitively or reflexively in both Dutch and Italian and had to be able to occur with an animate subject and an animate object in these languages. Because of a planned experiment with children, simple (easily intelligible) verbs were chosen; this was checked through Age of Acquisition (AoA) and frequency data (Kuperman, Stadthagen-Gonzalez, & Brysbaert, 2012 (English); Moors et al., 2013 (Dutch); Marconi, Ott, Pesenti, Ratti, & Tavella, 1993 (Italian)). The potential appropriateness of the verbs for children aged 5 and up was checked by at least two native speakers of each language.

Table A1. The transitive and reflexive verbs that were used in the third clause of the stories.

<table>
<thead>
<tr>
<th>Transitive verbs</th>
<th>Reflexive verbs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dutch</td>
<td>Italian</td>
</tr>
<tr>
<td>bedanken</td>
<td>ringraziare</td>
</tr>
<tr>
<td>bekijken</td>
<td>guardare</td>
</tr>
<tr>
<td>duwen</td>
<td>spingere</td>
</tr>
<tr>
<td>helpen</td>
<td>aiutare</td>
</tr>
<tr>
<td>kiezen</td>
<td>scegliere</td>
</tr>
<tr>
<td>roepen</td>
<td>chiamare</td>
</tr>
<tr>
<td>slaan</td>
<td>colpire</td>
</tr>
<tr>
<td>tekenen</td>
<td>ritrarre</td>
</tr>
<tr>
<td>volgen</td>
<td>seguire</td>
</tr>
<tr>
<td>zien</td>
<td>vedere</td>
</tr>
</tbody>
</table>