

The time-course of the application of binding constraints in reference resolution

Patrick Sturt*

Department of Psychology, University of Glasgow, 58 Hillhead Street, Glasgow G12 8QB, UK

Received 9 January 2002; revision received 5 August 2002

Abstract

We report two experiments which examined the role of binding theory in on-line sentence processing. Participants' eye movements were recorded while they read short texts which included anaphoric references with reflexive anaphors (*himself* or *herself*). In each of the experiments, two characters were introduced into the discourse before the anaphor, and only one of these characters was a grammatical antecedent for the anaphor in terms of binding theory. Both experiments showed that Principle A of the binding theory operates at the very earliest stages of processing; early eye-movement measures showed evidence of processing difficulty when the gender of the reflexive anaphor mismatched the stereotypical gender of the grammatical antecedent. However, the gender of the ungrammatical antecedent had no effect on early processing, although it affected processing during later stages in Experiment 1. An additional experiment showed that the gender of the ungrammatical antecedent also affected the likelihood of participants settling on an ungrammatical final interpretation. The results are interpreted in relation to the notions of bonding and resolution in reference processing.

© 2002 Elsevier Science (USA). All rights reserved.

Keywords: Anaphor resolution; Sentence processing; Eyetracking

One of the central issues in the study of human language processing is reference resolution; natural language is full of anaphors and other referring expressions, whose referents must be determined in order to arrive at a coherent interpretation of a text. Previous research in on-line human language processing has uncovered a large number of semantic factors which affect reference resolution (see Garrod & Sanford, 1994 for a review). The present paper considers the contribution of syntactic factors to reference resolution. In theoretical syntax, it is well known that reference resolution is affected by grammar, via the binding theory (Chomsky, 1981), which defines the syntactic constraints that allow the identification of grammatical antecedents to pronouns, reflexives and other referring expressions.

Chomsky's formulation of binding theory divides the constraints into three principles: Principle A, which we will be concerned with in this paper, defines constraints on the reference of anaphors such as reflexives and reciprocals, while Principle B defines constraints on pronouns, and Principle C defines constraints on other referring expressions. For example, in (1a), Principle A allows *Peter*, but not *John*, to be an antecedent for the reflexive anaphor *himself*, while in (1b), Principle B allows *John* but not *Peter* to be the antecedent of the pronominal *him*.

- (1a) John thinks that Peter hates himself.
(1b) John thinks that Peter hates him.

The usual explanation for contrasts such as these is syntactic: *Peter* and the anaphor *himself/him* are in the same local domain (in this case, a clause), and *Peter* is also in a position of syntactic prominence in relation to

* Fax: +44-141-330-4606.

E-mail address: patrick@psy.gla.ac.uk.

the anaphor [for example, in the Government and Binding framework (Chomsky, 1981), *Peter* is said to c-command the anaphor]. Under such conditions, Principle A allows the reflexive anaphor *himself* to refer to *Peter*, but the Principle B prevents *him* from referring to *Peter*. In the research reported in this paper, we will concentrate on Principle A.

Although binding theory has been intensively studied by theoretical syntacticians, very little work has considered its role in on-line processing. In this paper, we report two eye-tracking experiments which tackle this issue. There are two particularly interesting questions that can be asked with respect to the on-line processing of binding constraints. The first is a time-course question; to what extent is on-line processing affected by antecedents that are ungrammatical with respect to binding theory? This issue is well suited to examination by the eye-tracking methodology that we use in the present research, and it will be our main concern in this paper. The second question relates to interpretation; to what extent do binding principles act as a filter on the *final interpretation* of anaphoric reference—in other words, how often do people end up with a final interpretation of a sentence that violates binding constraints, as standardly defined? Some recent work has begun to address this second question using experimental methodologies in off-line tasks (Asudeh & Keller, 2002; Gordon & Hendrick, 1997). This, however, is not the main question of the present paper, although in the discussion of Experiment 1, we will report the results of a comprehension study, relating directly to the interpretation issue.

There are a number of different possible models of the time-course of binding constraints. Probably the simplest is what has been called the binding-as-initial-filter hypothesis (Nicol & Swinney, 1989). According to this hypothesis, binding theory is applied at the earliest possible stage in processing, and constrains all subsequent stages. Recall (1a) repeated below:

(1a) John thinks that Peter hates himself.

According to the binding-as-initial-filter account *Peter* is immediately chosen as the antecedent of *himself*, because it is the only legal antecedent in terms of binding theory, while *John*, which is not a permissible antecedent, is ruled out from the earliest stages of processing, and cannot subsequently be considered, even if *John* is highly focused in the discourse. In this paper, we will use the term *binding-accessible antecedent* to refer to an antecedent that is legal in terms of binding theory, and *binding-inaccessible antecedent* to refer to an antecedent that is not. When the meaning is clear, we will abbreviate these terms to *accessible antecedent* and *inaccessible antecedent*, respectively.

A second possible hypothesis is that binding theory is applied at at some delay. If so, the earlier stages of processing may include the consideration of antecedents

that will eventually be ruled out by by binding theory. We will call this second account the *binding-as-late-filter* account. This account would predict that there is an early stage in processing during which binding-inaccessible antecedents can have an affect on processing, but that these antecedents subsequently become unavailable as binding constraints are applied. This could occur if, during the initial processing of an anaphor, there was a stage of lexical access at which the syntactic information necessary for the computation of binding constraints was not yet available. Note that this would be consistent with a modular view of lexical access (Forster, 1979).

A third alternative account is that, as in the initial-filter model, the binding constraints are applied at the earliest stages of processing, but that, unlike in the initial-filter model, they may later be violated. Such processing patterns could occur for example, if the syntactic relations necessary to compute the binding constraints are available at early stages of processing, while certain conditions, such as discourse focus, that encourage reference to non-grammatical antecedents become available only later. According to such a model, binding constraints act as a *defeasible filter* on reference resolution, that is, a filter which can be violated during subsequent stages of processing.

A fourth possibility is that all relevant constraints, including binding theory, discourse focus, and others, are combined in a parallel, competition process from the earliest stages of processing. Thus, if binding-inaccessible antecedents are favored by alternative constraints such as discourse focus, these antecedents could affect processing. This type of model would fit with a *multiple constraints* view of language processing (see, e.g., MacDonald, Pearlmuter, & Seidenberg, 1994; Spivey & Tanenhaus, 1998). The *multiple constraints* framework could be realized by any number of actual implemented models, and the predictions of these models would depend on the relative strengths of the constraints and on the overall architecture of the model. If syntactic binding and discourse focus both correspond to strongly biasing constraints in the model, this might predict early effects of both of these constraints in processing. The predictions of such a constraint-based model would not coincide with those of any of the hypotheses outlined above. If, on the other hand, binding constraints are so strongly biasing as to overwhelm the constraints which favor binding-inaccessible antecedents, the predictions of a multiple constraints model might resemble those of the binding-as-initial filter model. Moreover, the fact that many constraints are available simultaneously does not necessarily mean that their effects on processing will also be simultaneous. Thus, if discourse constraints are relatively weakly biasing, then it is possible that their effects will emerge at some delay, in comparison to the effect of binding constraints. So the predictions of a constraint-based model may also resemble those of the

defeasible-filter model outlined above. In the absence of an implemented model making explicit and detailed predictions, it is likely to be very difficult to distinguish between the constraint-based model and the other hypotheses. The clearest support for a constraint-based view would come from a finding that both binding-accessible and binding-inaccessible antecedents could affect processing equally early, since this finding would effectively falsify the other hypotheses mentioned above. However, as should be clear from the above discussion, many other experimental outcomes, probably including many of those described below, could also be compatible with at least one possible implementation of a constraint-based model. This point should be remembered, although, in the interests of clarity of exposition, our discussion will concentrate on the initial filter, late filter and defeasible filter hypotheses, as these allow for relatively sharply defined empirical predictions.

There has been surprisingly little on-line experimental research on the time-course of binding constraints. Probably the first such study was by Hirst and Brill (1980), who examined what is now called Principle C of the binding theory using whole-sentence self-paced reading, combined with a comprehension task. An example item from their Experiment 2 is given in (2):¹

- (2) John stood watching. He ran for a doctor after Henry fell down some stairs.

The results showed that reading times for the second sentence differed as a function of the plausibility of the binding-inaccessible antecedent (e.g., *Henry*) as the referent of the pronoun, but comprehension judgements did not show any evidence for ungrammatical interpretation. Although the study shows that inaccessible antecedents can affect processing, the use of the whole-sentence self-paced reading technique does not allow us to consider the detailed time-course with which this effect emerged.

Initial support for the binding-as-initial-filter model came from a cross-modal priming study reported by Nicol and Swinney (1989), in which participants had to make a lexical decision to a visually presented word while listening to sentences like (3):

- (3) The boxer told the swimmer that the doctor for the team would blame him for the recent injury.

In (3), binding theory allows the pronoun *him* to refer to *the boxer* or *the swimmer*, but not *the doctor*. Thus, the binding-as-initial-filter account predicts that *the doctor* is immediately ruled out as an antecedent of the pronoun, and can never subsequently be considered in the interpretation. To test this, Nicol and Swinney (1989) pre-

sented a probe word at the pronoun offset in sentences like (3), and this probe word could be either related or unrelated to *boxer*, *swimmer*, or *doctor*. Lexical decision times showed a priming effect for probe words that were related to the two binding-accessible antecedents, but no such effect for probe words that were related to the inaccessible antecedent *doctor*. Nicol and Swinney (1989) interpreted this as evidence for the binding-as-initial-filter model. However, although the experiment gives us a “snapshot” view of the state of the language processor when the pronoun has just been processed, it does not tell us very much about how the binding constraints develop over time. For example, it might be the case that the processor begins to consider the binding-inaccessible antecedent at a point downstream of the offset of the pronoun, where probes were not presented.

Evidence for the binding-as-initial-filter model was also found by Clifton, Kennison, and Albrecht (1997) who recorded participants’ reading times as they read sentences like (4) in a self-paced reading task:

- (4a) The supervisors paid him yesterday to finish typing the manuscript.
 (4b) The supervisor paid him yesterday to finish typing the manuscript.
 (4c) The supervisors paid his assistant to finish typing the manuscript.
 (4d) The supervisor paid his assistant to finish typing the manuscript.

In (4c) and (4d), *the supervisor(s)* is a binding-accessible antecedent for the possessive pronoun *his*, but this antecedent matches the pronoun in grammatical number in (4d) only. In (4a) and (4b), in contrast, *the supervisor(s)* is not a binding-accessible antecedent for the accusative pronoun *him*. If binding theory acts as an initial filter on subsequent processing, then there should be a match/mismatch effect in (4c) and (4d), but not in (4a) and (4b). This is what Clifton et al. (1997) found. In the regions immediately following the word *his/him*, reading times were increased in (4c) in relation to the other three conditions. Therefore Clifton et al. (1997) argued for a mechanism in which binding theory acts as an initial filter. However, because this experiment was conducted using self-paced reading, which has a relatively low temporal resolution we do not know whether the result is informative about early or late processing. Note also that the effect was found in the regions immediately following the critical pronoun, but not at the pronoun itself. Thus, we do not know whether the delayed nature of the effect reflects the relatively low resolution technique of self-paced reading, or whether this was indeed the earliest point at which the processing of the relevant conditions emerged.

Evidence against the binding-as-initial-filter model was reported by Badecker and Straub (2002). In a word-by-word moving-window self-paced reading experiment

¹ Principle C prevents *Henry* from taking *He* as its antecedent in (2).

with a probe-recognition secondary task, they found that reading times at the two words following *himself* were faster in (5a) those at the same position in (5b):

- (5a) Jane thought that Bill owed himself another opportunity to solve the problem.
 (5b) John thought that Bill owed himself another opportunity to solve the problem.

The only difference between the two conditions was the gender of the initial name, which is not accessible as an antecedent of *himself*, in terms of binding theory. As processing nevertheless differed between the two conditions, Badecker and Straub (2002) argued that binding theory does not act as an initial filter on anaphor resolution. Instead, they argued that binding theory, discourse focus and other constraints are simultaneously available from the earliest stages when a referring expression is being processed, and that the various constraints compete with each other to determine the outcome of reference resolution. Again, it should be noted that the experiment is not fully informative about the detailed time-course of processing; we cannot be sure whether these results are the result of processes occurring at the earliest stages of reference processing, or at later integrative stages. As the interference effect was obtained at the two words following the reflexive anaphor, rather than at the anaphor itself, it could have been delayed in relation to the initial stages of reference resolution. Alternatively, the effect could have been immediate, but its slightly late appearance in the reaction-time data could simply have been an artifact of the comparatively low-resolution self-paced reading task, where effects in spill-over regions are often found.

Another series of studies by Runner, Sussman, and Tanenhaus (2002) (see also Runner, Sussman, & Tanenhaus, 2001) used eye-tracking techniques to study the processing of spoken passages such as (6):

- (6) Look at Ken. Have Ken touch Harry's picture of himself/him.

The phrase *Harry's picture of himself* is an example of a "picture noun phrase." If a picture noun phrase has a possessor in its specifier position (like *Harry's* in (6)), then, according to standard accounts of Binding Theory, this possessor plays the grammatical role of a subject in the picture noun phrase, making the noun phrase correspond to the local domain in which a reflexive must be bound, and a pronoun must not be bound. This means that, according to standard accounts, a reflexive in the syntactic position occupied by *himself* in (6) must bind to the possessor, but a standard pronoun in the same position (like *him* in (6)) must not. Thus, standard accounts of binding theory predict that reflexives and pronouns should appear in complementary distribution in picture noun phrases. The experiments of Runner et al. (2002) were designed to test this claim.

In the experiments, Runner et al. (2002) had participants manipulate dolls in visual world scenes while they listened to spoken instructions like (6). They found that, for pronouns (like *him*), participants nearly always made the doll touch the grammatically appropriate picture (e.g., the picture of *Ken* in (6)). Thus, the data for pronouns are compatible with a model in which Principle B acts as an absolute filter on the interpretation of pronouns inside picture noun phrases. However, for reflexives (like *himself*), participants chose the grammatically appropriate picture (e.g., the picture of *Harry* in (6)) on only about 75% of the trials.

Runner et al.'s experiments show convincingly that reflexives and pronouns are not in complementary distribution in picture noun phrases with possessors, contrary to the assumptions of most syntactic theories.² Further analysis of fixation data also showed that the preference for the grammatically appropriate antecedent is established very quickly following the anaphor, even though in the cases of reflexives, the preference is not absolute. As Runner et al. point out, one possible interpretation of their data is that reflexives inside picture noun phrases with possessors are interpreted *logophorically*—in other words, they are not constrained by Principle A. This would bring the treatment of such noun phrases in line with similar proposals that have already been made by Pollard and Sag (1994) and Reinhart and Reuland, 1993 for reflexives in picture noun phrases *without* possessives, (like *a picture of himself*).

For the purposes of the present paper, the interpretation of Runner et al.'s results with respect to the time-course of binding constraints depends on whether or not one assumes that reflexives inside picture noun phrases with possessives are in fact subject to the binding theory. If they are, then Runner et al.'s results should be interpreted as evidence against the binding-as-initial-filter model, since participants often ended up with "ungrammatical" interpretations. If reflexives in sentences like (6) are not in fact constrained by Principle A, then we cannot conclude for or against the binding-as-initial-filter view of the application of Principle A, simply because, by hypothesis, Principle A is not being applied here. Note that there is already some doubt in the literature about whether picture noun phrases with possessors should in fact be analysed as local binding domains for the purposes of Principle A (see Asudeh & Keller, 2002; Reinhart & Reuland, 1993, p. 683).

To conclude this section, previous work in this area has considered the question of whether binding theory acts as an initial filter, but has not provided a conclusive answer. In particular, previous research sheds little light

² Note that in picture noun phrases *without* possessors, it is relatively uncontroversial than not in complementary distribution.

on the detailed issues related to the time-course of processing. The exception to this is the study of Runner et al. (2002), but this study examines a construction that may not be processed using standard Binding Theory principles. The experiments reported in the present paper are intended to consider the time-course issue in some depth, using a sentence type which is generally agreed to involve the application of Principle A to the binding of a reflexive. We use an eye-tracking methodology to build up a detailed picture of how participants' patterns of eye fixations unfold over time as they read short texts. The eye-tracking methodology gives an extremely fine-grained and continuous picture of the time-course of processing, allowing us to determine the precise nature of the application of binding constraints.

Experiment 1

Experiment 1 was an eye-tracking experiment in which participants were required to read short texts. An example text is given in (7), showing the four experimental conditions which will be explained in detail below.

(7a) *Accessible-match/inaccessible-match*

Jonathan was pretty worried at the City Hospital. He remembered that the surgeon had pricked himself with a used syringe needle. There should be an investigation soon.

(7b) *Accessible-match/inaccessible-mismatch*

Jennifer was pretty worried at the City Hospital. She remembered that the surgeon had pricked himself with a used syringe needle. There should be an investigation soon.

(7c) *Accessible-mismatch/inaccessible-match*

Jonathan was pretty worried at the City Hospital. He remembered that the surgeon had pricked herself with a used syringe needle. There should be an investigation soon.

(7d) *Accessible-mismatch/inaccessible-mismatch*

Jennifer was pretty worried at the City Hospital. She remembered that the surgeon had pricked herself with a used syringe needle. There should be an investigation soon.

A named character (*Jonathan* or *Jennifer*) is introduced in the first sentence, and this character is subsequently referred to using a pronoun (*he* or *she*) in the second sentence. This puts the character strongly into discourse focus (Sanford & Garrod, 1988). The second sentence also introduces a second character *the surgeon*, and includes a reflexive anaphor. Although the first named character is in discourse focus, it is not a possible antecedent for the reflexive in terms of binding theory, while the second character (*the surgeon*) is a possible antecedent. We will call this second character the *bind-*

ing-accessible antecedent, and the first character the *binding-inaccessible* antecedent. The design orthogonally manipulated the gender agreement between the reflexive and the accessible and inaccessible antecedents, yielding two factors; *accessible antecedent* and *inaccessible antecedent*, each with two levels *match* and *mismatch*, and these factors were combined to create a 2×2 factorial design. This manipulation allows us to determine the point in processing at which the binding constraints are initially applied. A stereotypical gender manipulation was used for the binding-accessible antecedent (in (7), we use the stereotypically male *surgeon*, along with an alternation between *himself* and *herself*), in order to avoid exposing the experimental participants to ungrammatical sentences in cases where the reflexive mismatched the gender of the accessible antecedent.

It is expected that processing difficulty will occur when the stereotypical gender of the binding-accessible antecedent does not match the marking on the reflexive, in comparison to when it does. Previous work by Carreiras, Garnham, Oakhill, and Cain (1996) and Osterhout, Bersick, and McLaughlin (1997) shows that gender stereotype violations are quickly detected, and cause disruption when a mismatching anaphor is processed. The timing of this disruption is important, because, in the context of the current experiment, the earliest point at which it can be detected represents the earliest point at which we can assume that Principle A is being applied. We can judge how early this effect occurs by looking at where in the eye-movement record the relevant conditions begin to diverge. This point can then be used as a temporal "yardstick," against which the timing of any other effects can be measured.

In this experiment, we are particularly interested in effects of the binding-*inaccessible* antecedent. The experiment was deliberately designed to maximize the chances of finding such an effect. Thus, as mentioned above, the inaccessible antecedent was put into discourse focus by being introduced by a name and subsequently referred to with a pronoun. In addition, the second mention of the inaccessible antecedent (*he/she*) c-commanded the anaphor. Badecker and Straub (2002) consistently found effects of the binding inaccessible antecedent when this antecedent c-commanded the anaphor, but not when it did not. The main point of interest of the experiment was *when* any such effects of the inaccessible antecedent would be found, in relation to the effect of the binding-accessible antecedent.

If the initial filter model is correct, we should find no evidence of gender effects for the inaccessible antecedent at either late or early points in the eye-movement record, even though this antecedent is highly focused in the discourse. If the late filter model is correct, we should find effects of both accessible and inaccessible antecedents in early measures, but effects of only the accessible antecedent in late measures. Finally, if binding con-

straints act as a defeasible filter only on the early stages of processing we should find effects of the accessible antecedent at the earliest possible points in the eye-movement record, but no such evidence of an early influence of the inaccessible antecedent. However, effects of the inaccessible antecedent would be expected to emerge at later points in the eye-movement record.

Eyetracking is well suited to comparing these hypotheses, because of the number of eye-movement measures that can be used to find effects at different points in time. These include measures of very early processing, such as the time spent on the eye's first-fixation on some critical word, but also measures of delayed processing, such as the time taken in re-reading the word (see below). The eye-tracking technique also allows one to look at processing on the regions of text that occur downstream of the critical word, and this again, can be used as evidence for delayed processing effects.

Participants

Twenty-four native speakers of English from the Glasgow University community were paid to participate in the experiment.

Stimuli

Twenty-four stimuli were constructed in the form illustrated in (7) above (see Appendix A). The 24 materials included 12 that used stereotypically male nouns, and 12 that used stereotypical female nouns for the binding-accessible antecedent. These nouns were chosen partly from the norms published by Carreiras et al. (1996), and partly on the basis of intuition.³ For the binding-inaccessible antecedent, male/female pairs of first names were chosen, such that each pair shared the same length in terms of numbers of characters, and each pair was of roughly of the same frequency in the British National Corpus. Each item also included a third "wrap-up" sentence, that was included to make the discourses more cohesive, and also insulated the eye-movement record for the critical regions from noise in the data due to end-of trial events.

Procedure

The 24 stimuli were divided into four lists, such that each list contained exactly one condition of each item, and in any given list, each condition occurred the same number of times. Each list was combined with 40 filler discourses, and the lists were randomized for each participant such that no two experimental sentences appeared adjacent to each other.

³ The experimental results clearly demonstrate that this manipulation was effective.

The eye-tracker display allowed a maximum of 80 characters per line. The experimental materials were divided into lines of text according to the example below, where line breaks are marked with double slashes (//).

- (7) Jonathan was pretty worried at the City Hospital.
He remembered // that the surgeon had pricked himself with a used syringe needle. // There should be an investigation soon.

The first line included the first sentence, along with the pronoun and main verb of the second sentence. The second line included the remainder of the second sentence. The third line included only the final sentence. To aid subsequent fixation analysis, two blank lines were inserted between each actual line of text.

The experiment was run using a Generation 5.5 Fourward Technologies Dual Purkinje Image eyetracker. The tracker has angular resolution of 10' of arc. The tracker monitored only the right eye's gaze location, but viewing was binocular. A PC displayed items on a VDU 70cm from the participants' eyes. The VDU displayed about four characters per degree of visual angle. The tracker monitored participants' gaze location every millisecond, and the software sampled the tracker's output to establish the positions of eye fixations and their start and finish times.

Before the experiment started, the participant sat at the eyetracker, and three practice trials were presented to familiarize the participant with the experimental procedure. The experimenter then used bite bars and forehead restraints to immobilize the participant's head. Before each trial, the experimenter had the participant fixate a series of squares at various positions on the screen, to test accuracy of calibration. If this was inaccurate, the eyetracker was recalibrated. The last square fixated in this sequence was in the same position as the first character of the text. When the participant fixated this square, the experimenter pressed a button and the text was displayed. The participant pressed a button when he/she had finished reading the text, after which a question was displayed on the screen, with two possible answers displayed on the left and the right of the screen (e.g., *Did the incident take place at a medical institution? Yes/No*). The participant answered the comprehension question using the left or right buttons. Comprehension questions followed every trial. To avoid putting too much emphasis on the experimental manipulation, the questions never directly probed the referent of the anaphor. Mean comprehension accuracy for the current experiment was 85%, and all subjects scored more than 70%.

Data analysis

An automatic procedure pooled short contiguous fixations. The procedure incorporated fixations of less

than 80 ms into larger fixations within one character, and then deleted any remaining fixations of less than 80 ms. Readers do not extract much information during such short fixations (Rayner & Pollatsek, 1989). Extremely long fixations (greater than 1200 ms) were also removed, as these usually indicate tracker loss. We also removed 3.1% of the trials (evenly distributed among conditions) in which track-losses and participant blinks made it impossible to determine the course of fixations at and around the critical reflexive region.

The analysis only considered the critical second sentence in the discourse.

For purposes of analysis, the sentence was divided into the following regions:

1. He remembered that (*initial region*).
2. the surgeon had pricked (*pre-critical region*).
3. himself/herself (*reflexive region*).
4. with a (*spill-over region*)
5. used syringe (*pre-final region*)
6. needle. (*final region*).

We will report means and statistical analysis from the pre-critical region onwards. The pre-critical region consisted of the material between the complementizer and the reflexive (exclusive). The reflexive region consisted of the reflexive pronoun itself. The spill-over region consisted of the two words following the reflexive. The pre-final region consisted of all words between the spill-over region and the last word (exclusive), and the final region consisted of the last word of the sentence.

The fixation data for each region were analysed according to a number of different eye-movement measures, which give a range of information about the time-course of processing. *First-Fixation durations* refer to the duration of the first-fixation in a region. This measure is informative about the earliest processes that accompany and immediately follow lexical access. We will report first-fixation times only for the region corresponding to the reflexive pronoun, since this word represents the region of the sentence at which we can expect to find the earliest theoretically interesting processing differences between the conditions.

We will also report *First-Pass Reading Times*. This measure is calculated by summing the fixations in a region, between the time when the reader's eye-gaze first enters the region from the left, to the time when the region is first exited to either the right or left. First-pass reading times also record early processes associated with a region, but because even a single word can sometimes receive a number of consecutive fixations, means for first-pass reading times are generally longer than those of first-fixation times, thus allowing for processes that occur somewhat later. When the region consists of a single word, we will use

the term *gaze durations* interchangeably with first-pass reading times.

Regression Path Times (Brybaert & Mitchell, 1996; Duffy, Morris, & Rayner, 1988; Konieczny, 1996; Konieczny, Hemforth, Scheepers, & Strube, 1997; Liversedge, 1994; Traxler, Pickering, & Clifton, 1998) are the sum of fixations from the time when the reader first enters the region from the left, to the time when the region is first exited to the right. Note that regression path times always correspond to first-pass reading times if the region is first exited to the right. However, regression path times differ from first-pass reading times if the first exit from the region is a regression. In such cases, the regression path times include all fixations during that regression, plus any re-fixations on the critical region before the eye-gaze proceeds to subsequent regions. Thus, means for regression path times are generally longer than those for first-pass reading times, and can be seen to reflect slightly later processes, possibly including processes that accompany the integration of the critical word with the preceding context.

First-fixation durations, first-pass reading times and regression path times will collectively be referred to as *first-pass measures*. In the first-pass measures, if a region was skipped in first-pass reading (i.e., if subsequent regions were fixated before any fixation was made on the region of interest), the data point for that trial was treated as missing data, and not counted as zero (i.e., the mean was calculated from the other data points in the design cell). This is because in trials where the region was skipped, the reader presumably spent some unknown time processing the region parafoveally during fixations on the previous region. For the purposes of calculating first-pass reading times and first-fixation durations on the critical reflexive region, we attempted to take such parafoveal processing into consideration by using what we call the *leftward-shifting procedure*. If the critical reflexive region was skipped, we allowed any fixation up to a maximum of four characters to the left of the region boundary to count as a first-pass fixation on the reflexive. If there was still no fixation when the region boundary had been extended leftwards in this way, then the trial was treated as missing data.⁴ In the current experiment, the procedure increased the rate of first-pass fixation on the reflexive from 83 to 95%, with no significant differences among conditions. See Rayner and Sereno (1994) for a fuller description of first-pass analysis methods and the motivations behind them.

In addition to the first-pass measures described above, *Second-Pass Reading Times* are the sum of fixations

⁴ We used the *eyepoke* algorithm to calculate first-fixation times, and the *eyegaze* algorithm to calculate first-pass reading times for the reflexive region. Both of these procedures are defined in Chuck Clifton's PCEXPT software package. We are grateful to Chuck Clifton for making this software available.

tions made on a region after that region has already been exited (either to right or left) for the first time.⁵ As this measure excludes time spent during the initial reading of a region, it is informative about any processes that are somewhat delayed in relation to the first encounter of the critical region. For the second-pass reading times, trials where a region was not re-fixated contributed a value of 0 ms to the cell mean.

Results

Overview of results

Before describing the results in detail, we will give a brief overview of the main findings. Recall that the purpose of the accessible antecedent manipulation was to determine the time at which the binding constraints are initially applied. The results of the early measures on the reflexive (first-fixation and first-pass reading times) show that these constraints were applied extremely early. First-fixation and first-pass reading times were faster when the gender of the anaphor matched the stereotype of the accessible antecedent (e.g., *surgeon...himself*) than when they did not (e.g., *surgeon...herself*), but they did not differ reliably as a function of whether the inaccessible antecedent matched the anaphor. The first-fixation effect is shown in the top panel of Fig. 1.

This early effect shows that binding constraints were being applied extremely quickly following the initial reading of the reflexive. In contrast, the results for measures indicative of later processing show reliable influences of the inaccessible antecedent in addition to effects of the accessible antecedent. This effect can be seen in the bottom panel of Fig. 1, which shows the second-pass reading times in the pre-final region. The effect is driven by differences in the time spent during regressions into this region from later regions, and probably mainly from the last word in the sentence. Thus it can be seen as indicative of sentence-final wrap-up processes. There was a reliable difference between the two conditions in which the stereotypical gender of the accessible antecedent matched the gender of the reflexive (e.g., *surgeon...himself...).* In these conditions, second-pass reading times in the pre-final region were longer when the inaccessible antecedent mismatched the reflexive (*Jennifer...the surgeon...himself...)* than when it did not (*Jonathan...the surgeon...himself...).* This shows that antecedents which are ruled out by binding

theory can nevertheless affect processing, at a relatively late stage. To summarize, the results therefore support a model in which the binding constraints are applied at an extremely early stage, but where this early application of binding constraints does not act as a filter on all subsequent processes of interpretation.

Detailed results and discussion

Condition means for all the measures are given in Table 1. Analyses of variance were computed for each region, on the participant means collapsing over items (F_1), and on the item means collapsing over participants (F_2). The ANOVA design included accessible antecedent and inaccessible antecedent as the two within-participant (and within-item) factors. Each of these factors had two levels: match and mismatch.

We will divide the discussion of the results into two sections, beginning with results for early processing (the first-pass measures on the critical reflexive and spill-over regions). We will then go on to discuss the later processing effects (the results for both first- and second-pass measures on regions following the spill-over region, as well as results for second-pass reading time on the reflexive, pre-critical and spill-over regions).

There were no significant effects in the pre-critical region for the first-pass or regression path measures (all F 's < 2.4, all p 's > .14).

Early processing

We now turn to the reflexive region, which represents the first point in the sentence at which theoretically interesting results can be expected. In first-fixation durations (calculated using the leftward-shifting procedure described in the data analysis section above) a main effect of accessible antecedent showed that fixation times were reliably faster when the accessible antecedent matched the gender of the reflexive than when it did not ($F_1(1, 23) = 7.61, p < .05$; $F_2(1, 23) = 6.91, p < .05$). However, at this early stage in processing, the inaccessible antecedent had no effect on processing (main effect of inaccessible antecedent, and interaction; F 's < 1) (see Fig. 1, top panel).

This pattern was replicated in the first-pass (gaze duration) data, calculated using the leftward-shifting procedure described above. Again, analyses of variance showed a main effect of accessible antecedent ($F_1(1, 23) = 5.56, p < .05$; $F_2(1, 23) = 10.30, p < .01$), but no effect of inaccessible antecedent, and no interaction between the two factors (all F 's < 1).

As with the first-fixation and gaze duration data, regression path times showed a main effect of accessible antecedent ($F_1(1, 23) = 8.83, p < .01$; $F_2(1, 23) = 5.70, p < .05$), but no effect of inaccessible antecedent, or interaction between the two factors (F 's < 1).

In the spill-over region, the basic pattern found on the reflexive region in the first-pass measures was also

⁵ Due to a technical error, previous conference presentations of this research reported total reading times (sum of all fixations on the region, including re-fixations) as the measure of delayed processing, instead of second-pass reading times. Neither the main pattern of the results nor the theoretical conclusions are affected by different measure adopted. The results of all eye-movement measures reported here have been cross-verified with two separate analysis programs.

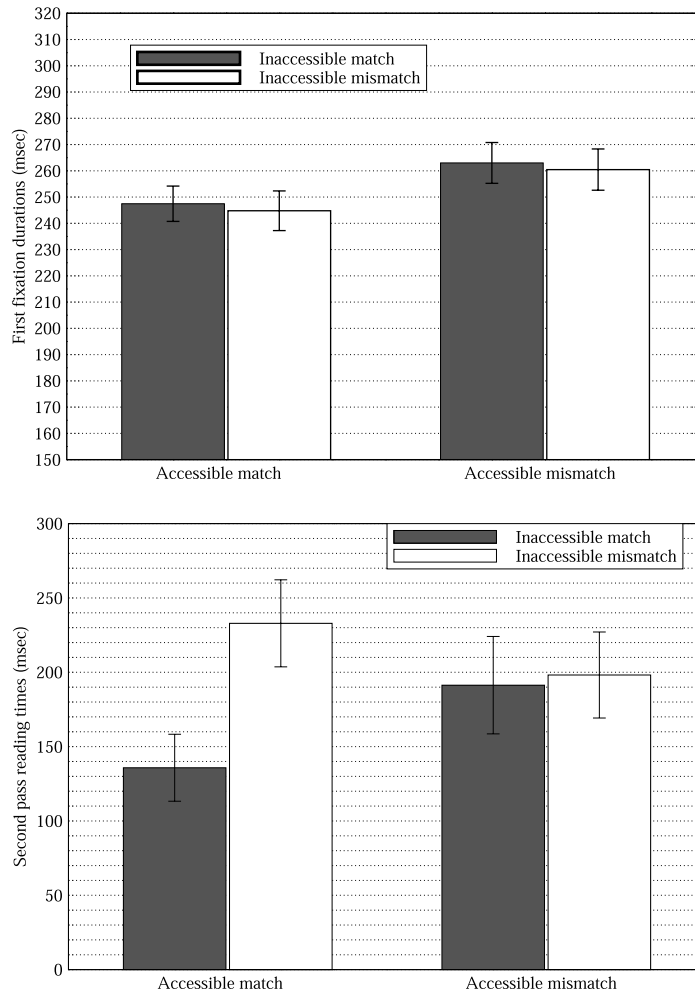


Fig. 1. Mean first-fixation durations on the reflexive (above), and second-pass reading times on the pre-final region (below) (with standard errors).

found in the regression path times (main effect of accessible antecedent; both p 's < .05, main effect of inaccessible antecedent and interaction, both F 's < 1). No effects approached significance in the first-pass reading times in the spill-over region. However, as can be seen from Table 1, there was a numerical difference of 39 ms in the first-pass times between the two accessible-match conditions, with the inaccessible-match condition taking numerically longer to read than the inaccessible-mismatch condition. A contrast analysis comparing only these two conditions showed that this difference was not significant (both p 's > .1). However, since crucial theoretical distinctions rely on the presence or absence of an effect of the inaccessible antecedent in the early measures, we should investigate this difference with some care. To rule out the possibility of a genuine first-pass effect of the inaccessible antecedent, we ran a number of follow-up analyses on a pooled region consisting of the

reflexive followed by the spill-over region. We analysed first-pass reading times and regression path times on this pooled region. We also ran an analysis of the proportion of first-pass regressions (defined as the proportion of trials on which the participant made a regression from the region before a subsequent region was fixated), both for the reflexive region and for the pooled region. Means for these analyses are given in Table 2.

Analyses of variance revealed significant, or nearly significant main effects of accessible antecedent in all of the analyses (all p 's < .06), but neither the interaction nor the main effect of inaccessible antecedent approached significance (all p 's > .1). In the first-pass reading times for the pooled region, the numerical difference between the two accessible-match conditions resembles that of the spill-over region. However, contrast analyses comparing these two conditions in the first-pass reading times showed that pooling the regions

Table 1
First-fixation, first-pass, regression-path, and second-pass times for Experiment 1

	Pre-critical the...pricked	Reflexive himself	Spill-over with a	Pre-final used syringe	Final needle
<i>First-fixation</i>					
Acc-match/inacc-match	—	247	—	—	—
Acc-match/inacc-mismatch	—	245	—	—	—
Acc-mismatch/inacc-match	—	263	—	—	—
Acc-mismatch/inacc-mismatch	—	260	—	—	—
<i>First-pass</i>					
Acc-match/inacc-match	710	268	344	395	300
Acc-match/inacc-mismatch	672	263	305	410	306
Acc-mismatch/inacc-match	668	298	333	405	330
Acc-mismatch/inacc-mismatch	715	291	320	371	307
<i>Regression-path</i>					
Acc-match/inacc-match	921	315	429	515	519
Acc-match/inacc-mismatch	944	344	428	577	669
Acc-mismatch/inacc-match	905	390	497	643	604
Acc-mismatch/inacc-mismatch	975	378	503	598	552
<i>Second-pass</i>					
Acc-match/inacc-match	329	92	78	136	45
Acc-match/inacc-mismatch	440	126	125	233	55
Acc-mismatch/inacc-match	456	188	138	191	42
Acc-mismatch/inacc-mismatch	397	151	151	198	50

First-fixation and first-pass times for the reflexive region are calculated using the leftward-shifting procedure described in the data analysis section. Acc = “accessible,” Inacc = “inaccessible.”

Table 2
Proportion of first-pass regressions for the reflexive region, and first-pass reading times, regression-path times and proportion of first-pass regressions for the pooled reflexive/spill-over region in Experiment 1

	First-pass (pooled)	Regression path (pooled)	Regress (%) (reflexive)	Regress (%) (pooled)
Acc-match/inacc-match	501	589	11	12
Acc-match/inacc-mismatch	454	596	15	18
Acc-mismatch/inacc-match	530	760	21	26
Acc-mismatch/inacc-mismatch	529	709	21	22

did not lead to a significant difference (again, both p 's > .1). The same two conditions in the proportion of regressions measure in the pooled region show the opposite numerical pattern, casting further doubt on the genuineness of the difference in the first-pass reading times. The regression-path measure, which effectively combines the first-pass reading times with the time spent during first-pass regressions, shows only a far-from significant 7 ms difference between the two accessible antecedent conditions; in other words, the numerical differences in first-pass reading and regressions cancelled each other out.

To summarize, the results of the first-pass measures in the reflexive and spill-over regions therefore show that the stereotypical gender of the binding-accessible antecedent has an effect at an extremely early stage in pro-

cessing. In fact, a difference in first-fixation durations represents the earliest detectable processing difference, given the eye-tracking methodology. It is also clear that, at this early stage, the binding-inaccessible antecedent has no effect. This suggests that the binding constraints are applied as a filter at the very earliest stages of processing the anaphor.

Later processing

We will now look at the results which are informative about processing which occurs at some delay after reading the reflexive, concentrating particularly on the second-pass reading times. We will begin with the second-pass times on the reflexive region, illustrated in Fig. 2:

As with the results of the early measures, there was an effect of the accessible antecedent, with longer

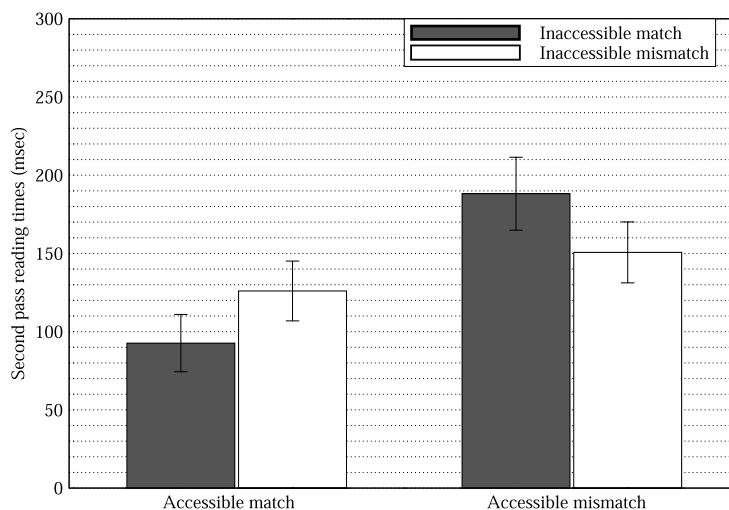


Fig. 2. Mean second-pass reading times in the reflexive region, with standard errors.

re-reading times when the stereotypical gender of the accessible antecedent mismatched the reflexive than when it matched ($F_1(1, 23) = 10.56, p < .01$; $F_2(1, 23) = 8.83, p < .01$). This effect was modulated by an interaction between the two experimental factors, which was significant in the participants analysis and marginal in the items analysis ($F_1(1, 23) = 4.93, p < .05$; $F_2(1, 23) = 3.09, p = .09$). A similar interaction was found in the pre-critical region, immediately preceding the reflexive ($F_1(1, 23) = 5.52, p < .05$; $F_2(1, 23) = 3.37, p < .08$). Perhaps the clearest way of looking at this interaction is in terms of the match/mismatch effect of the accessible antecedent at the two different levels of the inaccessible antecedent factor. Contrast analyses for both regions showed that the match/mismatch effect of the accessible antecedent was significant for the inaccessible-match conditions (all F 's > 4.45 , all p 's $< .05$) but not for the inaccessible-mismatch conditions (all p 's $> .2$). This demonstrates that the inaccessible antecedent influenced re-reading both of the reflexive and the preceding region. No other effects approached significance in the second-pass reading times for the reflexive or pre-critical regions (all p 's $> .1$).

The second-pass reading times for the spill-over region again showed a main effect of the accessible antecedent, with longer re-reading times when the stereotypical gender of the accessible antecedent mismatched the gender of the reflexive than when it matched ($F_1(1, 23) = 4.52, p < .05$; $F_2(1, 23) = 8.41, p < .01$). There were no other significant effects in the spill-over region for second-pass reading times (all p 's $> .05$).

In the pre-final region, there was a main effect of inaccessible antecedent in the second-pass reading times, which was significant by participants, but not by items, with second-pass reading taking longer when the inaccessible antecedent mismatched than when it matched

($F_1(1, 23) = 6.78, p < .05$; $F_2(1, 23) = 2.86, p = .10$). If we look at the means (see Fig. 1, lower panel) it is clear that this effect is driven entirely by the difference between the two accessible-match conditions. Contrast analyses confirmed this; the two accessible-match conditions differed reliably, by both participants and items ($F_1(1, 23) = 7.23, p < .05$; $F_2(1, 23) = 6.69, p < .05$), while the two inaccessible-mismatch conditions did not (both F 's < 1). This pattern resulted in a marginal interaction between the two experimental factors, though this approached significance only on the analysis by participants ($F_1(1, 23) = 3.72, p < .07$; $F_2(1, 23) = 2.50, p = .12$). A subsidiary analysis of second-pass reading times was carried out for this region, including only re-reading after the participant's gaze had exited the region to the right for the first time. In other words, this analysis measured time spent during regressions from subsequent regions.⁶ In this analysis, the interaction was robust for both subjects and items ($F_1(1, 23) = 5.86, p < .05$; $F_2(1, 23) = 4.38, p < .05$), suggesting that the effect mainly reflects differences in the time spent during regressions from subsequent regions, and probably from the final region.

In the first-pass times, also in the pre-final region, there was also an interaction between the two experimental factors, reliable only in the analysis by items ($F_1(1, 23) = 2.10, p = .16$; $F_2(1, 23) = 4.83, p < .05$). Contrast analysis suggested that this effect was most likely to have been due to the difference between the two accessible-mismatch conditions, where, in the items analysis, first-pass reading times were shorter when the

⁶ The means for this analysis were: accessible-match/inaccessible-match: 96 ms; accessible-match/inaccessible-mismatch: 176 ms; accessible-mismatch/inaccessible-match: 126 ms; accessible-mismatch/inaccessible-mismatch: 108 ms.

inaccessible antecedent matched the gender of the reflexive than when it did not ($F_1(1, 23) = 2.31$, $p = .14$; $F_2(1, 23) = 6.02$, $p < .05$). There was no hint of a difference for the accessible-match conditions (both F 's < 1).

The only other effect to approach significance in the pre-final region was a main effect of accessible antecedent in the items analysis of the regression-path times ($F_1(1, 23) = 2.74$, $p = .11$; $F_2(1, 23) = 8.43$, $p < .05$), with gender-matching conditions being read faster than gender-mismatching conditions, as in previous regions.

In the final region, regression-path times showed an interaction that was significant in the items analysis, and marginal in the participants' analysis ($F_1(1, 23) = 3.00$, $p = .097$; $F_2(1, 23) = 6.44$, $p < .05$), suggesting that regressions from the end of the sentence took particularly long in the accessible-match/inaccessible-mismatch condition. This was the same condition in which second-pass times were longer in the pre-final region. If the immediately preceding region was often the target of these regressions, regression-path times in the final region would correlate with second-pass times in the pre-final region. Thus, the effect appears to be a sentence wrap-up effect, which manifests itself in regressions from the last word in the sentence. There were no other significant effects in the final region.

To summarize the results of the experiment, early effects were found for the accessible antecedent only. This result is compatible with the binding-as-initial-filter account, but rules out the late-filter account. However, the experiment also showed that the inaccessible antecedent could affect processing at some delay. This is compatible with the defeasible filter account, in which binding constraints may be applied at an early stage in processing, but may later be violated, for example, if a binding-inaccessible antecedent is highly focused in the discourse, as it was in this experiment.

Given the finding that inaccessible antecedents can affect processing, it is important to consider whether or not these antecedents can also affect the final interpretation of the anaphoric reference, a question which cannot be answered with reference to eye-tracking data alone. Recall that the comprehension questions for the eye-tracking experiment did not directly probe for the antecedent of the reflexive. This was done to avoid focusing attention on the crucial manipulation of the experiment. To gain an idea of the final interpretation of the sentences, we ran a follow-up study, described below, where a sentence-by-sentence self-paced reading task was followed by a question that directly probed for the antecedent of the reflexive.

Follow-up experiment

The follow-up experiment used the context and target sentences of Experiment 1. The final wrap-up sentence

was not used, as we wanted to measure the participants' interpretations of the critical sentences immediately after reading them. The experimental design was the same as Experiment 1. The procedure was sentence-by-sentence self-paced reading. A trial began with an asterisk presented on a computer monitor, and the participant pressed a button to see the first sentence. When the participant pressed the button again, the first sentence disappeared from view, and was replaced by the second sentence. When the participant pressed the button again, the second sentence was replaced by a wh-question, with two possible answers, which appeared on the left and the right of the screen. The subject had to answer the question by pressing a button on the left or on the right, as appropriate. The question never included a reflexive pronoun, and never included any morphological gender clues. The questions for the follow-up experiment can be found in the Appendix A. The sequence of displays for a typical item was as follows:

(Display 1:) Jonathan was pretty worried at the City Hospital.

(Display 2:) He remembered that the surgeon had pricked himself with a used syringe needle.

(Display 3:) Who had been pricked with a used needle?

Jonathan/The surgeon

In Display 3, the appearance of the accessible vs. inaccessible antecedent on the left or the right of the screen was counterbalanced. The materials were divided into four lists, such that each list contained exactly one condition of each experimental item. The experimental items were combined in a random order with 40 filler items, adapted from those of Experiment 1, each of which was also followed by a wh-question. The dependent variable was the proportion of experimental trials in which the participant gave a response that was inconsistent with binding theory (e.g., in the example item given above, a response indicating *Jonathan* as the individual who had been pricked).

Twenty-four participants from the same student population as Experiment 1 volunteered to take part.⁷

The comprehension results are given in Table 3.

Analyses of variance on the percentages revealed a marginal main effect of accessible antecedent, with more ungrammatical responses in the accessible-mismatch than the accessible-match conditions ($F_1(1, 23) = 4.10$, $p < .06$, $F_2(1, 23) = 6.62$, $p < .08$). There was also a significant main effect of inaccessible antecedent, with more ungrammatical responses in the inaccessible-match conditions than in the inaccessible-mismatch conditions ($F_1(1, 23) = 17.81$, $p < .001$, $F_2(1, 23) = 11.50$, $p <$

⁷ This follow-up experiment was run by five undergraduate students as part of a third-year student project.

Table 3
Comprehension results: follow-up experiment

	Percentage of ungrammatical interpretations
Accessible-match/inaccessible-match	17
Accessible-match/inaccessible-mismatch	9
Accessible-mismatch/inaccessible-match	31
Accessible-mismatch/inaccessible-mismatch	6

.001). These two main effects were moderated by a significant interaction between the two factors ($F_1(1, 23) = 8.90, p < .01$; $F_2(1, 23) = 11.50, p < .01$).

From looking at the means, the interaction appears to have occurred because the effect of inaccessible antecedent was larger in the accessible-mismatch conditions than in the accessible-match conditions, although the relevant contrasts were significant, or nearly so, at both levels of the accessible antecedent factor (accessible-match: $F_1(1, 23) = 3.42, p < .08$, $F_2(1, 23) = 5.81, p < .05$; accessible-mismatch: $F_1(1, 23) = 21.76, p < .001$; $F_2(1, 23) = 29.38, p < .001$).

This pattern of results suggests that Principle A did not act as an absolute filter on the final interpretation of the anaphoric reference. Participants were significantly more likely to choose the binding-inaccessible antecedent when this antecedent matched the gender of the reflexive, compared with when it did not, and this effect was particularly strong when the binding-accessible antecedent did not match the anaphor in stereotypical gender.

Incidentally, these results appear to be at odds with findings of Gordon and Hendrick (1997), who ran a number of off-line acceptability-judgement studies examining the co-reference preferences of naïve English native speakers. They found no evidence to suggest that Principle A was ever violated in the participants' interpretations. However, their results are not strictly speaking comparable with those presented here, because our study presented the anaphoric sentences in a discourse context, while those of Gordon and Hendrick (1997) did not. Thus, Gordon and Hendrick (1997) might have found their participants to be more willing to accept binding-violations if these involved the anaphor referring to a focussed participant in the discourse, as ours were in the present experiment, while the present experiment might have resulted in fewer ungrammatical interpretations if the sentences were presented out of context.

With the interpretation results in hand, we are now in a position to propose a more detailed interpretation of the eye-tracking data, and in particular, the results for later processing, where evidence was found for an influence of the inaccessible antecedent. This influence resulted in interactions in the pre-critical, reflexive and pre-final regions in the second-pass reading times. Perhaps the clearest results are for the pre-final region (recall Fig. 1, lower panel) where second-pass reading times

differed between the two accessible-match conditions. For these conditions, re-reading was longer when the inaccessible antecedent mismatched the gender of the reflexive than when it matched. This result is likely to reflect sentence-final wrap-up effects, manifested in regressions from the last word of the sentence (recall that the second-pass effect was enhanced when only regressions from subsequent regions were considered). It could well be the case that some of the discourse preferences relevant to anaphor interpretation begin to have an effect during sentence wrap-up, and in Experiment 1, participants may sometimes have attempted to bind the reflexive to the binding-inaccessible antecedent during this phase of processing, as this antecedent was focused in the discourse. When the gender of the binding-inaccessible antecedent mismatched the gender of the reflexive, any such attempt to make an anaphoric link would have produced a feature clash, which could explain the elevated second-pass reading times in this condition for the pre-final region. Conversely, when the gender matched, participants may sometimes actually have re-interpreted the reflexive to refer to the inaccessible antecedent, as is suggested by the fact that there were more ungrammatical interpretations in this condition in the follow-up experiment.

An interesting question is why a similar wrap-up effect was not found for the accessible-mismatch conditions, which did not differ reliably in the second-pass times in the pre-final region (recall Fig. 1, lower panel). In fact, if anything, in the accessible-mismatch conditions, later processing appeared to take longer for the inaccessible-match condition rather than the inaccessible-mismatch condition, as is suggested by the numerical pattern of means in the second-pass times on the reflexive region (see Fig. 2), and by the first-pass results on the pre-final region. There are good reasons to expect processing differences between the accessible-mismatch and accessible-match conditions. In the accessible-mismatch conditions, incompatibility between the stereotypical gender of the accessible antecedent and the morphological gender of the reflexive resulted in immediate processing difficulty, as we can see from the results of the early measures. We can assume that participants were forced into recovery mode at this point, and it is likely that the inaccessible antecedent began to be considered earlier in these conditions than in the accessible-match conditions. Recovery could proceed in

two ways. Either the critical noun (e.g., *surgeon*) could be re-interpreted to refer to the non-stereotypical gender, or a new antecedent must be found for the reflexive, and the only other alternative is the binding-inaccessible antecedent. Far from slowing processing down, the gender clash between a mismatching inaccessible antecedent and the reflexive might actually have aided this recovery process. When the gender of the inaccessible antecedent does not match the reflexive, the choice of recovery strategy is easy to make—*surgeon* must be re-interpreted so that it denotes a female. The comprehension data support this, as the accessible-mismatch/inaccessible-mismatch condition had the lowest proportion of non-grammatical interpretations. In contrast, when the gender of the inaccessible antecedent does match the gender of the reflexive, there is no morphological cue to aid the decision process, and the comprehension data show that in about 30% of the trials, participants actually ended up interpreting the reflexive to refer to the inaccessible antecedent in this condition.

To summarize, the results of Experiment 1 show that the gender mismatch effects can occur very early. The fact that the effect in the first-fixation data was found for binding-accessible character, but not for the binding-inaccessible character suggests that the binding theory constraints are operative at or immediately following lexical access. The second pass results on the reflexive, and on the pre-final region show that this early application of the binding constraints did not prevent the inaccessible antecedent from having an effect during processing, but this effect was considerably delayed, in relation to the effect of the accessible character. In addition, the follow-up experiment suggests that although Principle A was applied early, it did not act as an absolute filter on the final interpretation of the sentence.

Experiment 2

The purpose of Experiment 2 was to consider an alternative possible explanation for the early effects found in Experiment 1. It is possible that the early mismatch effect was not driven by binding-theory configurational accessibility at all, but simply by linear position in the string. Recall that the binding-accessible antecedent was introduced later in the string than the alternative antecedent, and was therefore closer to the reflexive. Moreover, the binding accessible antecedent appeared on the same line of text as the reflexive, but the inaccessible antecedent appeared on the previous line. Thus, the mismatch effect could simply reflect lexical priming of a recently processed and perceptually close constituent. In other words, *the surgeon* could have caused facilitation for the masculine *himself*, or suppression for the feminine *herself*, or both. Conversely, the binding-inaccessible antecedent could have failed to cause such priming

effects because it was too distant from the reflexive. Importantly, these effects could have occurred for reasons that are completely unrelated to the syntactic configurations that are relevant to binding theory.

Experiment 2 was therefore designed to test whether the early first-fixation and gaze duration effects are robust against reversing the relative linear positions of the antecedents. Thus, we examined the processing of passages such as (8), which shows an experimental passage in the four conditions:

(8a) *Accessible-match/inaccessible-match*

Jonathan was pretty worried at the City Hospital. The surgeon who treated Jonathan had pricked himself with a used syringe needle. There should be an investigation soon.

(8b) *Accessible-match/inaccessible-mismatch*

Jennifer was pretty worried at the City Hospital. The surgeon who treated Jennifer had pricked himself with a used syringe needle. There should be an investigation soon.

(8c) *Accessible-mismatch/inaccessible-match*

Jonathan was pretty worried at the City Hospital. The surgeon who treated Jonathan had pricked herself with a used syringe needle. There should be an investigation soon.

(8d) *Accessible-mismatch/inaccessible-mismatch*

Jennifer was pretty worried at the City Hospital. The surgeon who treated Jennifer had pricked herself with a used syringe needle. There should be an investigation soon.

In (8), the linear positions of the binding accessible and inaccessible antecedents are reversed in relation to Experiment 1, while their accessibility with respect to the binding theory is kept constant. Thus, if the gender-mismatch effect is driven purely by linear string position, we should find effects of the gender of the inaccessible character (*Jennifer/Jonathan*) on the processing of the reflexive. Incidentally, such effects would also occur if binding constraints are applied to analyses of partial strings in the preceding input, as an analysis of the substring *Jennifer/Jonathan had pricked himself/herself* allows *Jennifer/Jonathan* to be an accessible antecedent, even though it is not accessible in the analysis global sentence. In contrast, if the mismatch effect is driven by binding-theoretic configurations based on the grammatical analysis of the complete preceding substring, we should find effects of the binding-accessible character's gender. And if the binding constraints act as a filter on this early process, as suggested by Experiment 1, there should be no effects of the binding-inaccessible character in early measures.

The questions which Experiment 2 seeks to answer are similar to those considered by Morris (1994). In her Experiment 2, Morris (1994) tested sentences like (9) as part of a larger design:

- (9a) The gardener talked as the barber trimmed the mustache after lunch.
 (9b) The gardener talked to the barber and trimmed the mustache after lunch.

Morris found that, relative to appropriate controls, gaze duration on *mustache* was shorter in (9a), where *mustache* is a co-argument of the related *barber*, than in (9b), where *mustache* is a co-argument of the unrelated *gardener*. As the relative positions of the two nouns in the sentence do not differ, this shows that priming effects in the initial stages of word processing are constrained partly by the syntactic and thematic structure of the sentence, and do not depend exclusively on linear positions.

Experiment 2 was not intended to examine effects of the inaccessible antecedent. In contrast to Experiment 1, the inaccessible antecedent did not c-command the reflexive. Badecker and Straub (2002) found effects of a inaccessible antecedents on the processing of reflexives only when the antecedent c-commanded the reflexive. Therefore, in contrast to Experiment 1, we do not expect to find effects of the inaccessible antecedent in Experiment 2. The purpose of the experiment is simply to rule out the linear order explanation of Experiment 1.

Participants

Twenty-four native speakers of English from the Glasgow University community were paid to participate in the experiment.

Stimuli

The stimuli for Experiment 1 were adapted on the model of (8) (see Appendix A). The noun denoting the binding-accessible character was always the head of the initial noun phrase. This noun phrase was always modified by a relative clause with a coindexed subject gap. The binding-inaccessible antecedent always appeared as the object of the main predicate in the relative clause. The predicate of the relative clause was never a verb that could take a tensed clause as its complement. This was to avoid the possibility of participants interpreting the second verb of the sentence as the head of a complement clause, as in (10):

- (10) The surgeon who *knew* (*that*) Jennifer had pricked herself. . .

Procedure

As in Experiment 1, four lists were created. The line-breaks in the experimental materials were as follows:

- (8d) Jennifer was pretty worried at the City Hospital.
 The surgeon // who treated Jennifer had pricked

himself with a used syringe needle. // There should be an investigation soon.

The first line included the first sentence and the first two words of the second (determiner followed by noun). The second line contained the remainder of the second sentence, and the third line contained the third sentence.

The experimental materials were combined with 48 items from an experiment on plural reference (example item: *John and Mary painted the flat. They really liked the colour. . .*), as well as 24 items from an experiment on the processing of main and subordinate clauses (example item *There was a murder committed at a hotel recently. After the porter phoned the authorities, the private detective investigated the scene of the crime. Everyone was very upset by the whole thing.*). The lists were randomized such that no two items from the current experiment appeared adjacent to each other. As in Experiment 1, comprehension questions followed every trial, and these questions did not probe the antecedent of the anaphor. Mean comprehension accuracy for the current experiment was 87%, and all subjects scored more than 70%. Other aspects of the procedure were identical to Experiment 1.

Results and discussion

Data analysis

Eye-movement data were prepared for analysis as in Experiment 1. As in Experiment 1, we removed trials in which track losses made it impossible to determine the fixation patterns at and around the critical reflexive region. This procedure affected 3.6% of the trials (evenly distributed among conditions).

The regions were divided as follows:

1. The surgeon who treated (*initial region*).
2. Jonathan had pricked (*pre-critical region*).
3. himself/herself (*reflexive region*).
4. with a (*spill-over region*)
5. used syringe (*pre-final region*)
6. needle (*final region*).

Overview of results

The main result of Experiment 2 was a replication of the early effect of the accessible antecedent in the first-fixation and gaze duration measures on the reflexive. In contrast to Experiment 1, however, there was no evidence for an influence of the inaccessible antecedent, either in early or in late measures. Because of the simpler pattern of results, we will not divide the discussion into earlier and later processes.

Details of results

Condition means for the various measures are presented in Table 4.

Table 4
First-fixation, first-pass, regression-path and second-pass times for Experiment 2

	Pre-critical J..pricked	Reflexive himself	Spill-over with a	Pre-final used syringe	Final needle
<i>First-fixation</i>					
Acc-match/inacc-match	—	253	—	—	—
Acc-match/inacc-mismatch	—	259	—	—	—
Acc-mismatch/inacc-match	—	271	—	—	—
Acc-mismatch/inacc-mismatch	—	272	—	—	—
<i>First-pass</i>					
Acc-match/inacc-match	502	268	327	421	276
Acc-match/inacc-mismatch	513	280	317	403	302
Acc-mismatch/inacc-match	531	292	326	383	295
Acc-mismatch/inacc-mismatch	492	307	340	398	287
<i>Regression-path</i>					
Acc-match/inacc-match	722	359	358	514	540
Acc-match/inacc-mismatch	737	341	397	514	490
Acc-mismatch/inacc-match	687	406	481	644	603
Acc-mismatch/inacc-mismatch	677	386	529	574	545
<i>Second-pass</i>					
Acc-match/inacc-match	278	88	93	164	50
Acc-match/inacc-mismatch	223	69	88	166	15
Acc-mismatch/inacc-match	269	153	120	176	55
Acc-mismatch/inacc-mismatch	251	124	137	203	44

First-fixation and first-pass times for the reflexive region are calculated using the leftward-shifting procedure described in the data analysis section of Experiment 1. Acc = “accessible,” Inacc = “inaccessible.”

There were no significant effects in the first-pass, regression path or second-pass times in the pre-critical region (all F 's < 2.1, all p 's > .15).

As in Experiment 1, first-fixation and first-pass times for the critical reflexive were calculated using the leftward-shifting procedure described in the data-analysis section of Experiment 1. This procedure raised the rate of first-pass fixation on the reflexive from 87% to 95%, with no significant differences among conditions.

We begin with the first-fixation results. As in Experiment 1, ANOVAs demonstrated a main effect of the accessible antecedent's stereotypical gender, such that first-fixations for the matching conditions were shorter than those for the mismatching conditions ($F_1(1, 23) = 5.74, p < .05; F_2(1, 23) = 5.90, p < .05$). Also as in Experiment 1, there was no effect of the inaccessible antecedent's gender, and the two factors did not interact (all F 's < 1).

The gaze durations on the critical reflexive, again calculated using the leftward-shifting procedure, revealed a similar pattern. There was a main effect of the accessible antecedent's stereotypical gender ($F_1(1, 23) = 7.47, p < .05; F_2(1, 23) = 9.18, p < .01$). In this measure, however, there was also a weak trend towards a main effect of the inaccessible antecedent's gender, with mismatching antecedents taking marginally longer to read than matching antecedents, but this difference only approached significance in the analysis by participants

($F_1(1, 23) = 3.14, p < .09; F_2(1, 23) = 2.74, p > .1$). There was no significant interaction between the two factors (both F 's < 1).

The regression-path measure replicated the main effect of stereotypical gender of the accessible antecedent, with mismatching antecedents causing longer reading times than matching ones ($F_1(1, 23) = 4.34, p < .05; F_2(1, 23) = 4.49, p < .05$). In the regression path times, there was, however, no hint of an effect of the inaccessible character's gender, nor did the two factors interact (all F 's < 1).

In second-pass reading times there was also a main effect of the accessible character's stereotypical gender ($F_1(1, 23) = 18.00, p < .001; F_2(1, 23) = 15.34, p < .001$), but unlike in Experiment 1 there was no trace of an effect of the inaccessible character's gender, or of an interaction between the two factors (all p 's > .1).

In the spill-over region, the main effect of accessible antecedent was replicated for regression path and second-pass times (all F 's > 6.0, all p 's < .05). The main effect of inaccessible antecedent and the interaction did not approach significance in these measures (all F 's < 1). The first-pass reading times showed a very weak trend towards a main effect of accessible antecedent, with mismatching conditions taking longer to read than matching conditions. However, this approached significance only in the analysis by items ($F_1 < 1; F_2(1, 23) = 3.22, p < .09$).

There were no significant effects in the three measures for the pre-final or final regions (all all p 's > .05).

To summarize the results, as in Experiment 1, we found that reading times on the reflexive were slowed when its morphological gender marking did not match the binding accessible antecedent's stereotypical gender. Again, this effect was found in the earliest measures, from first-fixation onwards, and again, the early measures did not show any evidence for an influence of the inaccessible antecedent. The one possible exception to this was the marginal trend towards an effect of inaccessible antecedent in gaze duration on the reflexive. However, this effect was probably spurious, since it only approached significance in the participants analysis, and was completely absent in both the other first-pass measures in this region (First-fixation, Regression path; F 's < 1). Thus it is probably safe to conclude that the early effect of accessible antecedent is not simply a priming reflex based on linear position of the antecedents in the preceding string, but that the configurational information relevant to computing the binding constraints is available at the earliest stages of processing.

The lack of an influence of the inaccessible antecedent contrasts with the second-pass results of Experiment 1. As we have pointed out, the current experiment was not set up to look for effects of the inaccessible antecedent, while Experiment 1 was. In addition to the fact that the inaccessible antecedent c-commanded the reflexive in Experiment 1 but not in Experiment 2, there are other reasons why the inaccessible antecedent might have been less prominent in Experiment 2 than in Experiment 1. Although the inaccessible antecedent was introduced with a proper name in both experiments, it was subsequently referred to with a pronoun only in Experiment 1 (see Sanford & Garrod, 1988). Secondly, the pronoun referring to this antecedent was a subject in Experiment 1, while the second mention of the inaccessible antecedent was not a subject in Experiment 2. The purpose of Experiment 2 was simply to determine whether early processing of the reflexive was sensitive to the relative order of different antecedents, or to configurational properties of the syntactic context, and the results conclusively support the latter conclusion.

General discussion

Both Experiment 1 and Experiment 2 showed evidence for processing difficulty when the gender of a reflexive anaphor mismatched the stereotypical gender of an antecedent that was accessible in terms of binding theory. These gender-mismatch effects were found at the very earliest points in processing in both experiments. In Experiment 1, it was also found that processing could be influenced by the gender of an antecedent that was inaccessible in terms of binding theory. However, these

effects occurred substantially later than effects of the binding-accessible antecedent, and they appeared to be related to recovery strategies and wrap-up effects, rather than processes related to the initial interpretation of the anaphor. In addition, the follow-up to Experiment 1 showed that participants often made ungrammatical interpretations of the anaphoric reference, in cases where the gender of the inaccessible antecedent matched that of the anaphor. The results of the eye-tracking experiments suggest that these ungrammatical interpretations were established relatively late in processing.

In the introduction to this paper, we discussed a number of possible models of the relation between binding constraints and on-line processing. The binding-as-initial-filter model predicts that binding theory is operative from the earliest stages of processing, and that it constrains all subsequent stages of interpretation following the first encounter of the anaphor. Although the first-pass and first fixation results of both Experiments support a model in which binding constraints are applied extremely early, the second-pass results of Experiment 1, and the comprehension results of the follow-up study, suggest that this does not constrain all subsequent stages of processing, and does not fully constrain the final interpretation. In contrast, the late filter model predicts that there is an initial stage in which binding constraints are not applied. This model is ruled out by the results of the two experiments reported here, where effects of binding-accessible antecedent, and the lack of such effects for the binding-inaccessible antecedent, were found in the earliest measures of processing. Thus, the combined results of Experiment 1 show that processing can indeed be affected both by a binding-accessible antecedent, and by a (binding-inaccessible) discourse-focused antecedent. The detailed results of the eye-movement measures strongly suggest that the relevant constraints become operative at temporally distinct stages, and are not both simultaneously available at the earliest point in processing, in other words, the results support a model in which Principle A acts as an early but defeasible filter. In the following paragraphs, we will discuss the notion of the defeasible filter in relation to a two-stage account of reference resolution.

The conclusion that reference resolution is a two-stage process is consistent with earlier claims that reference processing consists of distinct *bonding* and *resolution* processes (Garrod, 1994; Garrod & Sanford, 1994; Garrod & Terras, 2000; Sanford, Garrod, Lucas, & Henderson, 1983). According to these accounts, in an initial *bonding* stage, a link is made between the referring expression and one or more candidate antecedents in the discourse context, on the basis of superficial information. In a subsequent *resolution* phase, the link made in the bonding process is then evaluated, re-computed if necessary, and integrated into the semantic interpretation.

The notion of bonding was first discussed in relation to examples like (11), by Sanford et al. (1983):

Contexts

- (11a) Ronald parted his long hair.
 (11b) Ronald parted his long hair with a comb.

Target sentence

It was twisted with many teeth missing.

In a sentence-by-sentence self-paced reading study, Sanford et al., 1983 reported longer reading times for the target sentence following context (a) than context (b). When the target follows context (a), people often report a reading in which *it* refers to Ronald's hair, rather than the unmentioned comb. Sanford et al., 1983 argued that when a pronoun such as *it* is encountered, a *bonding* process is immediately triggered in which the preceding context is searched for a noun phrase that is congruent in number and gender (like *his long hair* in (11a)). The results of the experiments reported in this paper suggest that this initial search considers only those antecedents which are legal in terms of binding theory. Once this link has been established, it may be evaluated semantically, in a subsequent *resolution* phase. In the case of context (a), the link of *it* to *his long hair* will eventually be evaluated as implausible, resulting in longer reading times as the link is re-computed. Thus, the resolution phase may involve not only semantic evaluation of the link established in the bonding phase, but may also involve re-computation of the link if it is found to be inappropriate. Importantly, bonding and resolution are argued to be two independent processes, which need not take place simultaneously (Sanford et al., 1983).

More recently, Garrod and Terras, 2000 used eye-tracking to study bonding and resolution in the processing of passages like (12):

Contexts

- (12a) The teacher was writing a letter.
 (12b) The teacher was writing on the blackboard.

Target Sentence

...the pen/chalk dropped to the floor.

Garrod and Terras (2000) found that, in the target sentence, early processing of the phrase *the pen* was not affected by contextual appropriateness (i.e., first-pass reading was not slowed when *the pen* appeared in the context of writing on a blackboard). Instead, the results of the first-pass measures were dominated by the lexical facilitation of *pen* in relation to *chalk* as the dominant implicit instrument of *write*, regardless of whether the contextual situation involved writing on a blackboard or writing a letter (see Garrod & Terras, 2000, for full details of the experimental design, including controls). The effect of contextual appropriateness was found, but it emerged later, in the second-pass reading times.

Thus the results reported by Garrod and Terras (2000) support a two-stage model of reference resolution in which the initial *bonding* phase involves relatively superficial information (in this case, lexical association preferences) and the later *resolution* phase involves a much richer integration of the interpretation with contextual information. Interpreted within such a framework, the results of the present paper show that the processor does not only consider low-level superficial information during the bonding phase, but that syntactic configurations are also relevant. The results of Experiment 1 suggest further, that the resolution phase may involve a re-alignment of previously computed co-reference relations, and may yield an interpretation which is strictly speaking ungrammatical.

One question that should be considered is why binding-inaccessible antecedents are ever considered at all, given the existence of binding theory as a prior filter.

One possible answer is suggested by sentences where a reflexive requires a non-standard interpretation. For example, Zribi Hertz, 1989 discusses many examples of Principle A violations, including example (13) [p. 724], where many speakers find it acceptable for *himself* to refer to *John*, even though this antecedent is outside the local domain required by Principle A of the binding theory.

- (13) John hopes that Jane is speaking only to himself.

Clearly, a processor which irredeemably ruled out binding-inaccessible antecedents to anaphors would not be able to support an interpretation for sentences like (13), and it is therefore likely that the language processing system has a way of considering binding-inaccessible antecedents. As pointed out by Zribi Hertz, 1989 see also Reinhart and Reuland, 1993, successful interpretation of examples like (13) often depends on certain discourse properties. In the case of (13), co-reference between *himself* and *John* is made much more natural by the presence of the focus particle *only*. It is possible that the relevant discourse relations that encourage such interpretations become available only at a relatively late stage in processing, and if so, this would explain the lack of any discernible effect of inaccessible antecedent in the early measures in the present experiments, but the presence of such effects in later processing, and in the interpretation data.

A second conclusion that can be taken from the current experiments concerns the very early nature of the effect of stereotypical gender information. Based on self-paced reading experiments conducted in English and Spanish, Carreiras et al. (1996) argued that when a word like *surgeon* is read, an elaborative inference is quickly made to assign a gender to the discourse entity. This can explain early mismatch effects when a subsequent anaphor does not agree with the gender assigned in the inference. The results of the present experiment, where

gender-stereotype effects were found as early as the first fixation on the anaphor, suggest that this effect may actually be lexically based. On this view, a word like *surgeon* would be lexically marked for (perhaps default) gender, and a subsequent clash with a word like *herself* would be a clash between two sets of lexical features rather than a clash between one word's lexical features and those of an elaborate situation model. This would explain the presence of the gender stereotype effect in the first-fixation durations.

Acknowledgments

I would like to acknowledge the contribution of Lorna Morrow, who collected the eye-movement data for Experiment 2, and of Peter Ward, Julia Weller, Allison Welsh, Kate Williams, and Eddie White, who collected the data for the follow-up to Experiment 1 as part of an undergraduate project. I would like to thank Don Mitchell, Simon Garrod, Tony Sanford, Matt Traxler, and two anonymous reviewers for helpful comments on earlier versions of this paper. Remaining errors are my own. Various portions of this research were presented at the 2001 AMLaP conference in Saarbrücken, Germany, and the 2002 CUNY conference in New York, NY. I thank the audiences of these conferences for their comments. The research was supported by a British Academy Postdoctoral Fellowship.

Appendix A. Experimental materials

The experimental materials are given below. All the materials are given in the accessible-match/inaccessible-match condition. The alternative names (and the alternative pronouns, for Experiment 1) for the inaccessible antecedent manipulation are given in parentheses. For the second sentence, we give the two alternatives for the respective sentences in square brackets, separated by a forward slash (/). At the end of each item, we give the question that was asked in the follow-up to Experiment 1, with the alternative name for the inaccessible-mismatch conditions in parentheses.

1. John (Mary) didn't enjoy the pleasure flight at all. [He (She) claimed that the pilot injured himself quite badly during the journey./ The pilot who scared John (Mary) injured himself quite badly during the journey.] It's enough to make anyone nervous. Who got injured? John (Mary) / the pilot
2. Henry (Jenny) was shaken up after the accident at the factory. [He (She) mentioned that the firefighter had cut himself on a piece of broken glass./ The firefighter who saved Henry (Jenny) had cut himself on a piece of broken glass.] Luckily there were no fatalities. Who got cut? Henry (Jenny) / the firefighter
3. Jonathan (Jennifer) was pretty worried at the City Hospital. [He (She) remembered that the surgeon had pricked himself with a used syringe needle./ The surgeon who treated Jonathan (Jennifer) had pricked himself with a used syringe needle.] There should be an investigation soon. Who had been pricked with a used needle? Jonathan (Jennifer) / the surgeon
4. Tony (Anne) found the funeral parlour quite scary at night. [He (She) noticed that the undertaker whispered to himself all the time with a grimace./ The undertaker who worked with Tony whispered to himself all the time with a grimace.] Not an ideal holiday job. Who did the undertaker whisper to? Tony (Anne) / the undertaker
5. Harry (Helen) really enjoyed walking in the local sports ground. [He (She) was glad that the footballer had taught himself to paint the lines on the pitch./ The footballer who talked to Harry (Helen) had taught himself to paint the lines on the pitch.] It was a nice location. Who learned to paint the lines? Harry (Helen) / the footballer
6. Tommy (Sally) felt very tired after arriving at the station. [He (She) was sure that the porter would upset himself quite a lot about all the luggage./ The porter who met Tommy (Sally) had upset himself quite a lot about all the luggage.] Travelling isn't always fun. Who would be upset about the luggage? the porter / Tommy (Sally)
7. Bruce (Julia) was finding the court case a new experience. [He (She) reported that the judge disciplined himself very carefully all the time./ The judge who impressed Bruce (Julia) disciplined himself very carefully all the time.] The wig looked pretty ridiculous, though. Who was disciplined? the judge / Bruce (Julia)
8. Peter (Nancy) was having a lot of trouble with his boiler. [He (She) said that the engineer convinced himself that the valve was faulty./ The engineer who visited Peter convinced himself that the valve was faulty.] Heating systems can be quite frustrating. Who got convinced about the valve? the engineer / Peter (Nancy)
9. Gordon (Rachel) found the visit to the building site frustrating. [He (She) realized that the bricklayer lied about himself damaging the tools in the shed./ The bricklayer who annoyed Gordon lied about himself damaging the tools in the shed.] This was going to become complicated. Who did the bricklayer lie about? the bricklayer / Gordon (Rachel)
10. Timothy (Miranda) wanted to renew the pre-war wiring in the flat. [He (She) felt that the electrician owed himself another attempt to solve the problem./ The electrician who called on Timothy owed himself another attempt to solve the problem.] A patient attitude was all that was needed. Who was owed another attempt? the electrician / Timothy (Miranda)
11. Andrew (Sheila) had a good time at the pub by the army barracks. [He (She) was amused that the paratrooper had awarded himself the top prize in the pub quiz./ The paratrooper who drank with Andrew (Sheila) had awarded himself the top prize in the pub quiz.] The alcohol flowed freely. Who got the top prize in the pub quiz? the paratrooper / Andrew (Sheila)
12. Donald (Louise) moved to the countryside to escape the city. [He (She) was aware that the farmer had paid for himself to join the local fox hunt./ The farmer who spoke to Donald had paid for himself to join the local fox hunt.] Rural life takes a bit of getting used to. Who was going to join the foxhunt? the farmer / Donald (Louise)

13. Alice (Roger) found the surgery very busy that morning. [She (He) was surprised that the nurse criticized herself for being late for the appointment./ The nurse who interviewed Alice (Roger) criticized herself for being late for the appointment.] The traffic can be terrible these days. Who did the nurse criticize? Alice (Roger) / the nurse
14. Laura (James) had some strange stories about the seaside hotel. [She (He) confirmed that the receptionist almost killed herself in early May for no reason./ The receptionist who phoned Laura (James) almost killed herself in early May for no reason.] Tourism suffered after that. Who was almost killed? Laura (James) / the receptionist
15. Julia (Barry) would never forget visiting the council offices. [She (He) recalled that the secretary had treated herself to several cakes with pink icing./ The secretary who amused Julia (Barry) had treated herself to several cakes with pink icing.] Not very good for the teeth apparently. Who ate the cakes? Julia (Barry) / the secretary
16. Kate (Greg) really enjoyed the show business party on Saturday. [She (He) revealed that the ballet dancer introduced herself to all the stars from Hollywood./ The ballet dancer who invited Kate (Greg) introduced herself to all the stars from Hollywood.] Fame can be very seductive. Who met all the famous people? Kate (Greg) / the ballet dancer
17. Victoria (Nicholas) was finding life quite tough as a single parent. [She (He) knew that the babysitter would blame herself for the children's bad diet./ The babysitter who cheated Victoria (Nicholas) would blame herself for the children's bad diet.] Health is so important. Who would be blamed for the children's bad diet? Victoria (Nicholas) / the babysitter
18. Jessica (Richard) was having staff problems at the fashion salon. [She (He) accepted that the beautician disliked herself quite a lot and was unhappy./ The beautician who mistrusted Jessica (Richard) disliked herself quite a lot and was unhappy.] If only there was an easy solution. Who was disliked? Jessica (Richard) / the beautician
19. Sarah (Frank) really wanted to do well in the clothing business. [She (He) saw that the dressmaker familiarised herself with all the modern styles./ The dressmaker who employed Sarah (Frank) familiarised herself with all the modern styles.] It's important to be up to date. Who was familiarised with the styles? the dressmaker / Sarah (Frank)
20. Polly (Jason) found the police station quite an eye-opener. [She (He) was appalled that the prostitute had cut herself with a rusty razor blade./ The prostitute who scared Polly (Jason) had cut herself with a rusty razor blade.] These things are very worrying. Who was cut with the razor blade? the prostitute / Polly (Jason)
21. Alison (Trevor) was concerned at the American football match. [She (He) thought that the cheerleader had made a fool of herself in front of the players./ The cheerleader who stood near Alison (Trevor) had made a fool of herself in front of the players.] Sport can cause strong emotion. Who was made to look foolish? the cheerleader / Alison (Trevor)
22. Angela (George) managed to inspire good morale at the school. [She (He) was happy that the cleaner was proud of herself and really liked doing the job./ The cleaner who liked Angela (George) was proud of herself and really liked doing the job.] The atmosphere was good. Who was the cleaner proud of? the cleaner / Angela (George)
23. Barbara (Michael) employed some servants after winning the lottery. [She (He) found that the nanny was comfortable with herself in the large remote house./ The nanny who helped Barbara (Michael) was comfortable with herself in the large remote house.] The moors can be beautiful sometimes. Who did the nanny feel comfortable with? the nanny / Barbara (Michael)
24. Maggie (Thomas) tried to get on with everybody at the office. [She (He) remarked that the typist didn't trust herself very much with the new photocopier./ The typist who insulted Thomas didn't trust herself very much with the new photocopier.] Working environments can be quite delicate. Who was not trusted with the new machine? the typist / Maggie (Thomas)

References

- Asudeh, A., Keller, F. (2002). Experimental evidence for a predication-based binding theory. In *Papers from the 37th meeting of the Chicago Linguistics Society, Vol. 1*. Chicago Linguistics Society.
- Badecker, W., & Straub, K. (2002). The processing role of structural constraints on the interpretation of pronouns and anaphora. *Journal of Experimental Psychology: Learning Memory and Cognition*, 28, 748–769.
- Brysbaert, M., & Mitchell, D. C. (1996). Modifier attachment in sentence parsing: Evidence from dutch. *Quarterly Journal of Experimental Psychology, Section A: Human Experimental Psychology*, 49, 664–695.
- Carreiras, M., Garnham, A., Oakhill, J., & Cain, K. (1996). The use of stereotypical gender information in constructing a mental model: Evidence from english and spanish. *Quarterly Journal of Experimental Psychology*, 49, 639–663.
- Chomsky, N. (1981). Lectures on government and binding. Foris.
- Clifton, C., Kennison, S. M., & Albrecht, J. E. (1997). Reading the words *him* and *her*: Implications for parsing principles based on frequency and on structure. *Journal of Memory and Language*, 36, 276–292.
- Duffy, S. A., Morris, R. K., & Rayner, K. (1988). Lexical ambiguity and fixation times in reading. *Journal of Memory and Language*, 27, 429–446.
- Forster, K. I. (1979). Levels of processing and the structure of the language processor. In W. Cooper & E. Walker (Eds.), *Sentence processing: Psycholinguistic studies presented to Merrill Garrett* (pp. 27–85). Englewood, NJ: Erlbaum.
- Garrod, S. (1994). Resolving pronouns and other anaphoric devices. In C. Clifton, L. Frazier, & K. Rayner (Eds.), *Perspectives on sentence processing* (pp. 339–359). Englewood, NJ: Erlbaum.
- Garrod, S., & Sanford, A. J. (1994). Resolving sentences in a discourse context: How discourse representation affects language understanding. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 675–698). San Diego, CA: Academic Press.
- Garrod, S., & Terras, M. (2000). The contribution of lexical and situational knowledge to resolving discourse roles: Bonding and resolution. *Journal of Memory and Language*, 42, 526–544.

- Gordon, P. C., & Hendrick, R. (1997). Intuitive knowledge of linguistic co-reference. *Cognition*, 62, 325–370.
- Hirst, W., & Brill, G. A. (1980). Contextual aspects of pronoun assignment. *Journal of Verbal Learning and Verbal Behavior*, 19, 168–175.
- Konieczny, L. (1996). Human sentence processing: A semantics-oriented approach. Unpublished doctoral dissertation, University of Freiburg, Germany.
- Konieczny, L., Hemforth, B., Scheepers, C., & Strube, G. (1997). The role of lexical heads in parsing: Evidence from German. *Language and Cognitive Processes*, 12, 307–348.
- Liversedge, S. P. (1994). Referential contexts, relative clauses, and syntactic parsing. Unpublished doctoral dissertation, University of Dundee.
- MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). Lexical nature of syntactic ambiguity resolution. *Psychological Review*, 101, 676–703.
- Morris, R. K. (1994). Lexical and message-level sentence context effects on fixation times in reading. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 20, 92–103.
- Nicol, J., & Swinney, D. (1989). The role of structure in coreference assignment during sentence comprehension. *Journal of Psycholinguistic Research*, 18, 5–19.
- Osterhout, L., Bersick, M., & McLaughlin, J. (1997). Brain potentials reflect violations of gender stereotypes. *Memory & Cognition*, 25, 273–285.
- Pollard, C., & Sag, I. (1994). *Head-driven phrase structure grammar*. Stanford, CA and Chicago, IL: CSLI and University of Chicago Press.
- Rayner, K., & Pollatsek, A. (1989). *The psychology of reading*. Englewood Cliffs, NJ: Prentice-Hall.
- Rayner, K., & Sereno, S. C. (1994). Eye-movements in reading: Psycholinguistic studies. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 57–81). San Diego, CA: Academic Press.
- Reinhart, T., & Reuland, E. (1993). Reflexivity. *Linguistic Inquiry*, 24, 657–720.
- Runner, J. T., Sussman, R. S., Tanenhaus, M. K. (2001). The role of binding theory in the on-line reference resolution of reflexives. Poster presented at the fourteenth Annual CUNY sentence-processing conference.
- Runner, J. T., Sussman, R. S., Tanenhaus, M. K. (2002). Assignment of reference to reflexives and pronouns in picture noun phrases: Evidence from eye movements, in progress.
- Sanford, A., Garrod, S., Lucas, A., & Henderson, R. (1983). Pronouns without explicit antecedents. *Journal of Semantics*, 303–318.
- Sanford, A., & Garrod, S. C. (1988). Proper names as controllers of discourse focus. *Language and Speech*, 31, 43–56.
- Spivey, M. J., & Tanenhaus, M. K. (1998). Syntactic ambiguity resolution in discourse: Modeling the effects of referential context and lexical frequency. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 24, 1521–1543.
- Traxler, M. J., Pickering, M. J., & Clifton, C. (1998). Adjunct attachment is not a form of lexical ambiguity resolution. *Journal of Memory and Language*, 39, 558–592.
- Zribi Hertz, A. (1989). Anaphor binding and narrative point of view: English reflexive pronouns in sentence and discourse. *Language*, 65, 695–727.