New arguments in favour of an automatic gender pronominal process

François Rigalleau a, David Caplan b & Vanessa Baudiffier a

a Université de Poitiers, Poitiers, France
b Massachusetts General Hospital, Boston, MA, USA

Published online: 13 May 2010.


To link to this article: http://dx.doi.org/10.1080/02724980343000549

PLEASE SCROLL DOWN FOR ARTICLE
New arguments in favour of an automatic gender pronominal process

François Rigalleau
Université de Poitiers, Poitiers, France

David Caplan
Massachusetts General Hospital, Boston, MA, USA

Vanessa Baudiffier
Université de Poitiers, Poitiers, France

This paper examines the automatic and strategic use of gender information in pronominal processing. Experiments 1 and 2 used short sentences where a pronoun was preceded by two potential antecedents. Results showed that even when adult readers did not use pronominal gender to strategically accelerate pronominal resolution, they remained sensitive to a gender disagreement between the pronoun and its potential referents. This gender sensitivity was further explored in Experiments 3 and 4. These experiments used longer texts where only one of the two potential referents was highly accessible when the pronoun was encountered. A gender disagreement between the pronoun and this antecedent induced longer reading times. The four experiments confirm the existence of a nonstrategic gender coindexation process between a pronoun and the entity in the focus of a discourse.

IMMEDIATE EFFECTS OF GENDER MARKING IN PRONOMINAL COINDEXATION

Introduction

Anaphoric processing is the process by which a linguistic expression derives its meaning from an earlier, or antecedent, linguistic expression. This process is needed to construct a coherent discourse representation.

Some recent theories of anaphoric processing assign a crucial importance to the accessibility of the anaphoric antecedent (Gerrig & McKoon, 1998; Greene, Gerrig, McKoon, &...
Ratcliff, 1994; Greene, McKoon, & Ratcliff; 1992; McKoon, Gerrig, & Greene, 1996; Myers & O’Brien, 1998). Because many corpus analyses show that pronouns are usually used to refer to the most highly accessible entity in the discourse (e.g., Ariel, 1988, 1990; Brennan, 1995; Fox, 1987), Greene and his collaborators suggested that antecedent accessibility may be the central factor explaining how pronouns are processed. Pronouns share less information with their antecedents (i.e., gender, number, person) than do anaphoric noun phrases, which usually share some more specific semantic properties with their antecedent (e.g., the criminal referring to a burglar). Therefore, these authors argue, the appropriate use of pronouns requires that their antecedents be easily accessible for the hearer or the reader. In natural discourse, the discourse preceding the pronoun usually places the antecedent in the focus of attention (e.g., Chafe, 1974; Fletcher, 1984). This observation can explain why Greene et al. (1992) claimed that the most relevant question about pronoun processing is not to find out how the processing system of comprehenders uses a pronoun to find its referent, but how discourse processing makes the referent highly accessible.

This view assigns a crucial role to the accessibility of an antecedent in pronoun processing; the contribution of bottom-up processes triggered by the pronoun becomes less important. In particular, gender information carried on a pronoun is said to play little role in initial comprehension processes. For instance, Greene et al. (1992) suggest that the gender of a pronoun is not automatically used to reduce the search space for a referent of a pronoun when several referents are equally accessible. In such conditions, the use of gender information is one of several processes that are employed strategically to select a unique referent (Garnham, Oakhill, & Cruttenden, 1992; MacDonald & MacWhinney, 1990).

Our goal in this paper is to investigate automatic processing of gender information in pronouns. We begin with a brief review of studies that establish that this information is processed strategically. These studies have at times led researchers to suggest that gender information in pronouns is only subject to strategic processing—that is, that it is not processed automatically. Not all researchers who have shown strategic processing of pronouns have come to this stronger conclusion (see, for instance, Garnham et al., 1992), but it is represented in writings such as Greene et al. (1992), who say: “When the discourse has only one entity in the focus of attention at the time the pronoun is encountered, then it may be that essentially no processing is required for the pronoun. It may be that information predicated of the pronoun is attached to the focused entity by means of an attachment process that is simple, automatic, and demanding of little processing capacity” (p. 281, italics ours). Note that it is “information predicated of the pronoun” and not information carried on the pronoun that is automatically attached to the focused entity, on this view. The same researchers have, however, also suggested that an incorrect pronoun (i.e., a pronoun that is not in agreement with the most accessible referent) would be expected to disrupt processing (Greene et al., 1994), indicating that uncertainty exists about the possibility that information carried on a pronoun might be attached to the focused entity, and about whether any such attachment is automatic. The precise contribution of pronominal gender to automatic aspects of pronominal processing remains to be established. A related question concerns the relation between these two types of gender process. If an automatic gender process exists, what is its scope, and what relation does it have to strategic pronominal processing? Our focus in this paper is on automatic processing of pronominal gender, but we will comment on the relationship of the automatic processing we envisage to other aspects of processing pronouns.
In the first section of our introduction, we review the main arguments in favour of the strategic nonautomatic processing of gender information and against automatic processing of gender information of pronouns. These arguments come mainly from studies that used a probe recognition task to investigate the effect of reading a pronoun on the accessibility of its antecedent, but some reading time data are also consistent with this view. In the second section, we summarize some of our recent experimental results that are not consistent with this view (Rigalleau & Caplan, 2000). Although we accept that part of pronominal gender process is strategic, we maintain that an automatic pronominal gender coindexation process exists. This process immediately checks whether the gender of a pronoun is in agreement with information in the discourse focus (i.e., the highly accessible information that constitutes the most probable potential antecedent for the pronoun). The four experiments reported in this paper tested this “automatic coindexation hypothesis”.

Probe recognition and pronominal processing

The probe recognition task is often used to assess the effect of an anaphor on the level of activation of its antecedent. For instance, Gernsbacher (1989, Exp. 5) asked participants to read sentences, like (1) where a pronoun was preceded by two potential antecedents (two nouns agreeing in gender with the pronoun).

1. Ann predicted that Pam would lose the track race, but she came in first very easily.

Sentences were presented at a relatively slow rate (approximately 300 ms per word). The reading was interrupted by the presentation of a probe word corresponding either to the antecedent (i.e., Pam) or to the nonantecedent (i.e., Ann). In (1), locations of the probes are indicated by the symbol $\Delta$. Participants decided rapidly whether the probe had occurred in the sentence they were reading. When the probe was presented either just before or immediately after the pronoun, no recognition time advantage was noticed for the pronominal antecedent over the other preceding noun. However, at the end of the clause starting with the pronoun, there was a significant advantage of accessibility for the pronominal antecedent. This clause-final advantage was greater when pronominal gender disambiguated the pronoun (i.e., the non-referent in (1) was replaced by a male noun), but in this condition, an antecedent advantage was still not observed immediately after the pronoun. This result suggested that pronominal gender was not immediately processed to select the appropriate referent.

Greene et al. (1992) questioned the automaticity of pronominal gender processing. In four experiments, they presented longer texts (Text 2 is an example) where the position of the pronoun was less predictable. They also increased the presentation rate of the text (to 250 ms per word) to make the occurrence of a controlled or strategic process less probable. Finally, they did not ask readers any comprehension questions requiring the identification of the pronominal antecedent.

2. Mary and John were doing the dishes after dinner.
   One of them was washing while the other dried.
   Mary accidentally scratched John with a knife
   and then she dropped it on the counter.
In such conditions, they failed to observe any advantage of the probe corresponding to the pronominal antecedent, even at the end of the text. This null result was not observed when the authors used slower presentation texts, shorter texts (only the final sentence of Text 2 was presented) and asked comprehension questions about the pronouns; in these conditions, the usual passage-final advantage for the antecedent was re-obtained (Greene et al., 1992, Exp. 5). These results suggested that the identification of a single referent for a pronoun is a strategic process, requiring time and attention.

It is possible to question the sensitivity of the probe recognition task in the new conditions created by Greene et al. (1992). However the authors rejected this view because they noticed that, even with their fast presentation rate, another type of anaphor (i.e., an anaphoric noun phrase) induced a specific activation of its antecedent. The authors suggested that their long texts with pronouns tended to place the two potential antecedents equally in the focus of attention, and that in conditions where a strategic use of gender information was made difficult, a single referent was not selected using gender cues. The final conclusion of their paper rejected an automatic use of pronominal gender information: They claimed that the pronoun had no effect on the relative accessibility of the two potential referents. This null effect was central to supporting a “no processing” thesis. It was consistent with the idea that readers did not process the gender of the pronoun.

In more recent papers (Gerrig & McKoon, 1998; Gerrig & McKoon, 2001; Greene et al., 1994; McKoon et al., 1996), the same authors reported more direct evidence in favour of their view. They used longer texts where the pronoun was encountered at a point where its single antecedent was the highly activated referent. The texts were long (13 to 16 sentences) and they contained an “unheralded pronoun”. Unheralded pronouns had very distant antecedents (5 to 6 sentences between the last explicit mention of the antecedent and the pronoun). However, they remained appropriate because, just before the pronoun, there was a sentence that could evoke (i.e., reestablish in working memory) the appropriate referent. Using the probe recognition task, Greene et al. (1994) observed that the pronominal antecedent was indeed highly activated just before the pronoun, and its level of activation remained equally high after the sentence containing the pronoun. In other experiments using the same texts, McKoon et al. (1996) found the same type of results when the sentence containing the pronoun (e.g., Gloria asked Jane: “Did she make the evening unbearable?” where in the previous part of the text, a cousin named Marilyn received Jane during the evening) was replaced by an “allusion sentence” with no unheralded pronoun (e.g., “Gloria asked Jane: “Was the evening unbearable?”). So, it seems that when the antecedent is the only highly accessible potential antecedent, the pronoun has no effect on its accessibility. This result was seen as evidence in favour of a strong “no processing” thesis. Because the unheralded pronoun did not seem crucial in determining the final level of activation of its antecedent, McKoon et al. entitled their paper “Pronoun resolution without pronouns”.

However, if there is really no processing of the pronoun, and more specifically of its morphosyntactic agreement with its antecedent, we should expect that readers would accept an unheralded pronoun that disagrees in gender with its antecedent (e.g., in the context above, Gloria asked Jane: “Did he make the evening unbearable?”). The authors did not make this strong prediction. They claimed that pronouns “confirmed accessibility”. This phrase is consistent with two different views of pronominal processing. On the first view, the occurrence of a pronoun (whatever its gender) confirms that the most highly activated referent
should be attached to the information predicated of the pronoun. According to the second view, morphosyntactic agreement is relevant to and used in this confirmation function. This last version could be consistent with the results reported by Greene et al. (1992) if we assume that gender agreement checks that an element in the focus of attention is in agreement with the pronoun but has no automatic effect on the level of activation of either this or other elements in the discourse. This view, however, would require revising the strong version of the “no processing” thesis. Experiments 3 and 4 examine these two options.

**Reading time studies**

Garnham et al. (1992) reported an experiment in which they varied two factors influencing pronominal comprehension: the congruency of the pronominal clause with the implicit causality bias of a verb mentioned in the previous clause, and the relevance of the pronominal gender cue to resolving the reference of a pronoun.

Some transitive interpersonal verbs show an implicit causality bias (Garvey & Caramazza, 1974). For instance, when subjects are asked to continue a fragment like (3), they tend to use “she” because they think that the implicit cause of the admiration is usually the object of the admiration (Garnham, 2001). The verb “to admire” is said to have a bias in favour of its object (N2). On the other hand, for a fragment with the verb “to confess”, like (4), subjects tend to use “he”, referring to N1.

3. Peter admired Ann because ...
4. Peter confessed to Ann because ...

The effect of the implicit causality of verbs can also be observed in comprehension. With gender-ambiguous pronouns, two continuations can be constructed: one congruent with the bias of the verb, as in (5), the other incongruent, as in (6).

5. Peter confessed to Bill / because he wanted a reduced sentence.
6. Peter confessed to Bill / because he offered a reduced sentence.

Reading time for the second clause is longer when the ending is incongruent (Garnham et al., 1992; Leonard, Waters, & Caplan, 1997).

Garnham et al. (1992) reported that this congruency effect was not reduced when pronominal gender was sufficient to disambiguate the pronoun. The occurrence of a relevant gender cue reduced the time to read the subordinate clause. This suggested that readers made use of the gender cue to accelerate the comprehension of this clause. This gender cue effect was observed when all sentences were followed by a question about the identity of the pronominal referent. In their Experiment 4, Garnham et al. (1992), compared this condition with a condition where filler items had a subordinate clause without a pronoun, as in (7).

7. Vicky lied to Alan because the truth was too dreadful.

This new condition was intended to discourage subjects from seeing their task as one in which it was necessary to resolve the reference of pronouns. In this last condition, the authors did not
observe an effect of the gender cue on the reading time of subordinate clauses for experimental items. They concluded that the gender cue effect in the previous condition was strategic. It seemed that readers used the relevant gender cue to accelerate their reading only when they systematically had to resolve the reference of the pronouns.

This result suggested that readers use gender to determine a pronoun’s antecedent only under conditions in which attention is likely to be drawn to the pronoun. However, Garnham et al. (1992) pointed out that such a conclusion “does not mean that subjects take no notice of the gender marking when they do not always have to resolve pronouns to answer questions” (p. 249). “People readily notice when the wrong pronoun is used, so they must be extracting gender information” (p. 253). In our view, Garnham et al. articulated the major counterintuitive empirical phenomenon that is predicted by the hypothesis that gender processing is only strategic.

The probe recognition and reading time studies that we have reviewed converge to suggest that the use of gender to process the pronoun is only strategic. More precisely, the effects of gender information on probe recognition times for antecedents of pronouns or on reading times of sentences containing pronouns seem restricted to conditions where readers: (1) are questioned about the identity of the pronouns; (2) have enough time to develop a strategic process; and (3) can easily guess the position of the pronoun. The requirement to “resolve” the reference of a pronoun (i.e., to determine who is referred to by the pronoun among several candidates) seems to be an important aspect of the conditions that elicit this strategic gender process.

Arguments in favour of automatic pronominal gender coindexation

In contrast to this view, Rigalleau and Caplan (2000) argued that pronominal gender coindexation is automatic. They reported three experiments in which they showed that participants were sensitive to a disagreement between a pronoun and the noun in the focus of attention even under nonstrategic task conditions. These experiments used a cross-modal naming paradigm: Subjects listened to the first part of a sentence, and a probe word was presented in the middle of a screen; subjects had to name this probe word, and their naming latency was measured, after which they listened to the end of the sentence. In their second experiment, Rigalleau and Caplan presented sentences where the probe word corresponded to the pronoun (the italicized word in the examples 8–11). Sentences could contain two names of the same gender or different genders, and the final clause could be congruent or incongruent with the implicit causality of the verbal bias, as illustrated in examples 8–11.

Different gender, congruent ending:
8. Tom questioned Sue because he wanted to learn the truth.
Different gender, incongruent ending:
9. Tom questioned Sue because she had not told the truth.
Same gender, congruent ending:
10. Tom questioned Sam because he wanted to learn the truth.
Same gender, incongruent ending:
11. Tom questioned Sam because he had not told the truth.
In this experiment, sentences were sometimes followed by a question about the identity of the pronominal referent (12.5% of the presented sentences). The results showed that the naming latency of pronouns in the different gender (Examples 8 and 9) condition was longer than that in the same gender condition (Examples 10 and 11). There was no significant effect of congruency or interaction between gender and congruency. The effect of gender suggested that participants made immediate use of the gender cue when it was relevant to pronoun resolution. Rigalleau and Caplan (2000) suggested that the longer naming latency reflected a disengagement process: The participant disengaged activation from the referent that disagreed with the pronoun. To examine whether this effect was strategic, Rigalleau and Caplan (2000, Exp. 3) used another condition where the speech rate of the sentences was accelerated and where no questions were asked about pronouns. As in Garnham et al. (1992, Exp. 5), there were many filler items where the second clause did not contain a pronoun. For these sentences, subjects had to name another word, like “of” in (12):

12. Wayne accused Brian because of a fingerprint on a piece of furniture.

In this condition, the gender effect noticed in the previous condition was no longer significant. This result was consistent with the observations of Garnham et al. (1992): In conditions where participants were not focused on pronominal resolution, and nonstrategic processing of a pronoun was discouraged, there was no trace of an immediate use of gender cue.

To test whether subjects processed pronominal gender under these nonstrategic conditions, Rigalleau and Caplan (2000, Exp. 3) included a minority of sentences (11%) where the pronoun was either in agreement or in disagreement with the only antecedent mentioned in the sentence, as in (13) and (14).

13. Logan lied out of spite; he had no sense of honor.

14. Nancy paid without being asked; he had a sense of honor.

The naming latency of pronouns in disagreement was longer than the naming latency of pronouns in agreement. This result confirmed the intuition of Garnham et al. (1992): Even when participants did not strategically process the gender cue to select one referent among several, they remained automatically sensitive to a gender disagreement. So at least one part of the gender process seemed automatic: Checking the gender agreement between pronoun and a name in focus.

A preliminary inventory of components of pronominal gender processing

This review of the tasks and conditions where automatic or strategic pronominal gender processing occurs allows us to propose a model that makes a distinction between two processes that utilize gender information. First, we postulate that there is an automatic coindexation process, which consists in matching gender (and possibly other features carried on pronouns) between a pronoun and the most accessible referent(s) in working memory. This process notes matches and mismatches between the gender of a pronoun and the gender of these nouns. It is related to the “confirmation” function that Greene et al. (1994) assigned to pronouns, in the
second sense of “confirmation” that we discussed above. Second, we posit a strategic disengagement process. This disengagement process requires that the automatic coindexation process has occurred (since it operates on its output), and it requires that there be a motivation to identify a unique referent (or as small a set of referents as permitted by the discourse) for the pronoun (e.g., comprehension questions about the pronouns in an experiment). This second process reduces the activation level of highly accessible referents that do not agree with the pronoun.

This distinction is consistent with the probe recognition results reported by Gernsbacher (1989) and Greene et al. (1992). These authors reported that pronouns induced an advantage of recognition time for their referent (relative to a nonreferent). But this advantage only became significant at the end of the pronominal sentence (Gernsbacher, 1989) and when readers had special motivation to identify the pronominal referent (Greene et al., 1992). Moreover, this advantage seems to involve an inhibition of the nonreferent (Gernsbacher, 1989). Thus, this advantage could reflect what we call a strategic disengagement process.

The distinction is also consistent with the reading time and naming latency studies. These studies found on-line differences between conditions with two same gender nouns and two different gender nouns, but only when participants were questioned about the identity of the unique referent. Under these conditions, a pronoun that disagrees with one of two potential referents has a longer naming latency (Rigalleau & Caplan, 2000, Exp. 2), and there is a shorter reading time for the pronominal sentence when a pronoun agrees with only one name (Garnham et al., 1992). Rigalleau and Caplan suggested that these effects are both results of the postulated strategic disengagement process. Disengagement of the link between a pronoun and a noun phrase with which it disagrees in gender, though strategic, occurs immediately upon encountering the pronoun, leading to longer naming latencies for a pronoun when disengagement can occur (the mixed gender conditions). When the predicate following the pronoun is then processed, its processing benefits from the resources made available because the disengagement process has eliminated one possible attachment of the predicate, and a more rapid semantic integration of this predicate with the previous discourse is possible. Thus this model can account for both these results.

The present study

Our model claims that an automatic coindexation process is triggered by the pronoun. However, the evidence for this process is not strong. In the present work, we wanted to achieve a more complete testing of the automatic gender coindexation hypothesis. Three goals were pursued.

The first was to address methodological issues associated with the studies reported in Rigalleau and Caplan (2000). First, we used a different paradigm in order to avoid problems associated with naming latency studies of pronouns. In the cross-modal naming paradigm, participants have to process an isolated pronoun presented on a screen. Pronouns belong to the closed-class vocabulary (i.e., function words and grammatical inflexions). They are rarely fixated by readers of texts (Ehrlich & Rayner, 1983). Kennison and Gordon (1997) reported an eye fixation study in which readers skipped approximately 76% of pronouns referring to grammatical subjects. Of course, this does not mean that pronouns are not processed parafoveally, but it could be suggested that requiring subjects to fixate a pronoun is sufficient
to induce a controlled process that involved its gender information. Therefore, in the experiments reported here, we decided to present pronouns in combination with the words that followed them. Subjects were not forced to fixate pronouns or to name pronouns because the reading time of the pronominal sentence was measured in a self-paced reading paradigm where the complete pronominal sentence was presented on the screen. Second, Rigalleau and Caplan used very short passages (one sentence). Greene et al. (1992) used quite long texts when they did not find an effect of pronominal gender in probe recognition tasks. It could be suggested that longer texts are a better condition to avoid any strategic process. In Experiments 3 and 4 reported in this paper, we used texts containing six sentences (more than in the texts used by Greene et al., 1992, which contained four sentences).

Our second, related, goal was to extend the evidential basis for the automatic coindexation hypothesis to the reading task used by Garnham et al. (1992). These authors found that, in a strategic but not in a nonstrategic condition, readers had shorter reading times for the complete clause containing a pronoun when pronominal gender was relevant to the selection of an antecedent than when it was not. We attempted to replicate this result and extended the set of stimuli to include sentences where two potential nouns were mentioned before a pronoun that disagreed with both of them in gender. We made two predictions. First, we predicted that we would replicate Garnham et al.’s results: The gender cue effect observed in the strategic condition (longer reading times for pronouns agreeing with two potential referents) would not be significant in the nonstrategic condition. Second, in this nonstrategic condition, we predicted that readers would remain sensitive to a disagreement between a pronoun and the two potential referents in the context.

The third goal of our experiments was to test the following hypothesis: The automatic gender coindexation process concerns the most accessible referent (the referent that is in the focus of attention). In the experiments reported by Rigalleau and Caplan (2000), the pronominal context was very short and contained only one potential referent (see Examples 13–14). In Experiments 3 and 4 of the current paper, we used long texts with two potential referents for the pronoun. The textual distance between the last mention of each referent and the pronoun was varied: It was 35 words for the distant referent and only 6 words for the closer referent. So the closer antecedent should be more activated in working memory when the pronoun was encountered. This prediction was checked in Experiment 3 using the probe recognition paradigm. In Experiment 4, we measured the reading time of the pronominal sentence when the pronoun agreed with the distant referent or with the closer referent, and no comprehension questions were asked about the pronominal sentence. Our prediction was that readers would still be sensitive to a gender disagreement between the pronoun and the most activated referent in this condition. We also used a condition where the pronominal anaphora was replaced by a repeated noun. The predictions about this specific condition will be formulated in the introduction of Experiment 4.

In summary, the four experiments reported in this paper tested the hypothesis that automatic gender coindexation of a pronoun occurs even when strategic use of gender information is not engaged (Experiments 1 and 2), and that it involves a process that checks gender agreement between a pronoun and the most highly activated potential referent in working memory (Experiments 3 and 4).

We emphasize that our hypothesis concerns the existence of a nonstrategic pronominal gender coindexation process, and not its contribution to pronominal resolution. However, if
such a process exists, it is relevant to pronominal resolution, as our sketch of a model above indicates. We note, moreover, that our review of the current literature on the topic shows that arguments in favour of an automatic sensitivity to a gender mismatch are rare compared to the arguments in favour of a strategic use of the gender cue to select a referent.

EXPERIMENT 1

In this first experiment, our goal was to replicate the main results obtained by Garnham et al. (1992, Exp. 4) in their “strategic condition”. These authors used a clause-by-clause self-paced reading procedure. They asked participants to read sentences where a pronoun appeared in a subordinate causal clause. They varied the implicit causality bias of the verb in the main clause, the congruity of the subordinate clause relative to the bias of the verb, and the relevance of the pronominal gender cue. All the sentences were followed by questions about the identity of the pronominal referent. Their main result was that participants showed a shorter reading time when the gender cue could help pronoun resolution. This result suggested that readers made use of the gender cue to accelerate their resolution of the pronouns.

Method

Materials

Twenty verbs with an implicit causality bias were used to construct the verbal material. Of these, 10 verbs had a bias in favour of their grammatical subject (VN1), and 10 had a bias in favour of their grammatical object (VN2). The existence and the strength of the bias was checked in a pilot experiment involving 20 subjects (16 female, 4 male, mean age: 23.45 years, \(SD = 3.36\)). The participants were presented fragments containing two different-gender first names around a verb in the past tense, followed by the causal connective (e.g., “Jean a suivi Anne parce que . . .” [“John followed Ann because . . .”]). Half of the fragments began with a masculine first name, and different pairs of first names were used for each fragment. Forty verbs were used to construct the material, chosen on the basis of the material used in several English language experiments on implicit causality (in particular: Garnham, Traxler, Oakhill, & Gernsbacher, 1996; Greene & McKoon, 1995; Rigalleau & Caplan, 2000); French translations of the verbs were used. Each fragment was presented on a sheet of paper, and participants had to write a continuation for each fragment starting with a pronoun “he” or “she” referring to one of the two characters mentioned in the fragment. After examination of the results, 10 VN1 verbs were selected in which the pronoun referred to the grammatical subject in at least 80% of the participants; 10 VN2 verbs were selected because the grammatical object was referred by the pronoun for at least 65% of the subjects. The VN1 verbs were: interroger (question), suivre (follow), fatiguer (make somebody tired), vendre (sell), mentir (lie), perdre (lose), se confier (confess), décevoir (disappoint), gagner (win), and agacer (irritate). The mean percentage of N1 choice for these verbs was 86.5% \(SD = 11.36\). The VN2 verbs were: avoir confiance (trust), punir (punish), récompenser (reward), critiquer (criticize), craindre (fear), envier (envy), engueuler (blame), congratuler (congratulate), féliciter (congratulate), and croire (believe). The mean percentage of N2 choice for these verbs was 87.5% \(SD = 11.08\).

For each of the verbs, four versions of a sentence were constructed. Sentences 15 and 16 illustrate the two first versions. Their first clause contained two first names of different gender with the same number of letters, one before the verb (N1) and the second after the verb (N2). For each type of verb, there were five verbs with a feminine subject name and five verbs with a masculine subject name; different names were used for each verb. For one version (15), the subordinate clause introduced by “parce que” (“because”) started with a pronoun referring to N1. These pronouns were congruent with the
implicit causality bias of VN1 verbs (as in Example 15) but it was incongruent for VN2 verbs. For the second versions illustrated in (16), the pronoun referred to N2, was not congruent with the verbal bias of VN1 verbs, and was congruent with the bias of VN2. The predicate of the pronoun was constructed to be consistent with the pronoun, so that the incongruent subordinate clause was not implausible. As illustrated by (15) and (16), the difference in length (in words) between the congruent and the incongruent versions was kept to a minimum.

Different gender, congruent ending:
15. Catherine a suivi Ferdinand parce qu'elle ignorait le chemin.
Mary followed John because she did not know the way.

Different gender, incongruent ending:
16. Catherine a suivi Ferdinand parce qu'il connaissait le chemin.
Mary followed John because he knew the way.

The two other versions of a sentence were constructed by replacing the second name by a name of same length but with opposite gender, so that the two names in the main clause had same gender. The pronoun was in agreement with the two names, as illustrated in (17) and (18):

Same gender, congruent ending:
17. Catherine a suivi Véronique parce qu'elle ignorait le chemin.
Mary followed Lucy because she did not know the way.

Same gender, incongruent ending:
18. Catherine a suivi Véronique parce qu'elle connaissait le chemin.
Mary followed Lucy because she knew the way.

The mean length of the subordinate clause (without counting the connective and the pronoun) was 3.9 words for VN1 verb congruent sentences, 3.7 for VN1 verb incongruent sentences, 4 for VN2 verb congruent sentences, and 4.2 for VN2 verb incongruent sentences. The mean length of the main clauses was identical for sentences with VN1 verbs and VN2 verbs. During the experiment, the sentences were presented in two parts, first the main clause, second the subordinate clause (starting with “parce que” or “because”). Each sentence was followed by a yes/no comprehension question about the identity of the pronominal referent (e.g., “Was N1 ignorant of the way?”). The correct answer was “yes” for half of the questions.

Procedure

Subjects were tested individually in a quiet room. Successive stimuli appeared in the centre of a Macintosh computer screen with a white background. Stimuli were presented, and data were collected in a Psycscope script (Cohen, MacWhinney, Flatt, & Provost, 1993). The two first clauses were presented in black characters (Geneva 24-point bold characters). Responses were collected via a button box with three buttons. A first press on the middle yellow button triggered the appearance of the first clause. A second press on the same button made the first clause disappear, and it was replaced by the second clause which appeared in the same location as that of the first clause. A third press on the yellow middle button made the second clause disappear and the comprehension question appear (red Geneva 24-point characters). The participant responded by using the right green button for “yes” or the left green button for “no”. The question remained on the screen until the participant responded. This answer was followed by feedback about the correct expected answer (blue Geneva 24-point characters). A final press on the yellow button made the message “Next Trial” appear.
The delay between the first yellow button press and the second was taken as the reading time of the first clause, and the delay between the second and the third was taken as the reading time of the second clause. The time to answer the question was also measured, along with the nature of the answer.

Participants were instructed to read sentences “at their own normal speed” and to respond “as quickly as possible without making mistakes”.

**Design and subjects**

There were three variables in the design: the causality bias of the verb in the main clause (VN1 vs. VN2); the congruency of the subordinate clause relative to this bias (congruent vs. incongruent); and the relevance of the gender of the two names in the main clause (same vs. different gender). The verb bias variable was a between-sentence factor; the two other variables were within-sentence factors. The three variables were within-subject factors.

The 80 experimental sentences were all presented to each participant. This aspect of the experiment was also partly present in Garnham et al.’s (1992) study who presented the congruent and incongruent versions in separate blocks. Four blocks of 20 sentences were constructed, each with one different version of a sentence with each verb. The order of presentation of the blocks was systematically counterbalanced between the subjects. For each block, the order of presentation of the sentences was random (using the function “random presentation” of Psyscope).

A total of 22 subjects participated in this experiment (19 females, 3 males, mean age: 22.3 years; $SD = 5.4$). They were students of the University of Poitiers.

**Results**

The goal of this experiment was mainly to see whether our verbal materials allowed us to replicate the reading times results obtained by Garnham at al. (1992) demonstrating a strategic use of gender information. These results concerned the reading time of the subordinate clause under conditions where gender was relevant or irrelevant to resolve the pronoun. Therefore, the result section focuses on the results concerning the subordinate clause.

For all the collected times (reading times of the subordinate clauses, and response time to the question), times that were more than 10,000 ms were replaced by the longest time of the subject in the condition, leading to replacement of one response. The times for passages where an incorrect answer was given were replaced by the mean time of the subject in the condition.

Table 1 reports mean proportion of errors and mean correct response times for the question about pronouns in all conditions.

**Table 1**

<table>
<thead>
<tr>
<th>NP1 verb</th>
<th>NP2 verb</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Congruent</td>
</tr>
<tr>
<td>Gender</td>
<td>% error</td>
</tr>
<tr>
<td>Same</td>
<td>18.6%</td>
</tr>
<tr>
<td>Different</td>
<td>4.1%</td>
</tr>
<tr>
<td>Overall mean</td>
<td>11.3%</td>
</tr>
</tbody>
</table>

$^a$RT in ms.
These data were analysed in two analyses of variance (ANOVAs) for each dependant variable. The first ANOVA used subjects as a random factor \(F_1\) and involved three within-subject factors: verb bias (VN1 vs. VN2); congruency of the final clause with the verb bias (congruent vs. incongruent), and gender (same vs. different gender of nouns preceding the pronoun). The second ANOVA used items as a random factor \(F_2\): it involved the same factors, but congruency and gender were within-item factors, and verb bias was a between-items factor.

For the response time measure, the gender factor had a significant main effect, \(F_1(1, 21) = 27.55, p < .01; F_2(1, 18) = 61.4, p < .01\): readers answered more rapidly when gender was a relevant cue to identify the pronominal referent. There was also a main effect of congruency, \(F_1(1, 21) = 39.7, p < .01; F_2(1, 18) = 101.9, p < .01\), due to longer times to answer questions in the incongruent condition. The main effect of verb bias was not significant, \(F_1(1, 21) = 0.71; F_2(1, 18) = 0.63\), but there was a significant two-way interaction between verb bias and congruency, \(F_1(1, 21) = 8.71, p < .01; F_2(1, 18) = 9.22, p < .01\). Although the effect of congruency was significant for the two types of verb, VN1, \(F_1(1, 21) = 16.01, p < .01; F_2(1, 9) = 37.1, p < .01\); VN2, \(F_1(1, 21) = 41.6, p < .01; F_2(1, 9) = 64.9, p < .01\), an effect of verbal bias was significant by items only in the incongruent condition, \(F_1(1, 21) = 4.02, p > .05; F_2(1, 18) = 4.37, p < .05\): In this condition readers had longer response times in the VN2 condition than in the VN1 condition. The interaction could reflect somewhat more incongruent endings for VN2 verbs than for VN1 verbs. All other two-way interactions were not significant at the level .05 by subjects and by items.

The ANOVAs on error proportions gave similar results. The main effect of gender was significant, \(F_1(1, 21) = 81.9, p < .001; F_2(1, 18) = 60.7, p < .001\), with more errors when the two potential referents had same gender. The main effect of congruency was also significant, \(F_1(1, 21) = 25.2, p < .01; F_2(1, 18) = 10.25, p < .01\): Readers made more errors in the incongruent condition. The main effect of verb bias was not significant, \(F_1(1, 21) = 3.4, p > .05; F_2(1, 18) = 0.8\). For this measure, the interaction between verb bias and congruency was not significant, \(F_1\) and \(F_2 < 1\). However, there was a significant two-way interaction between gender and congruency, \(F_1(1, 21) = 11.26, p < .01; F_2(1, 18) = 7.21, p < .05\). This interaction reflected a clear effect of congruency when no gender cue allowed readers to correctly answer, \(F_1(1, 21) = 34.37, p < .01; F_2(1, 18) = 10.96, p < .01\): Incongruent endings induced more errors in this condition. However, when gender cue was relevant, congruency had no significant effect, \(F_1(1, 21) = 0.81; F_2(1, 18) = 1.01, p > .05\). A tendency to emphasize speed over accuracy could explain the fact that errors were more sensitive to this interaction: When no gender cue was present, subjects could prefer to respond rapidly at the expense of an incorrect answer. No other interactions were significant at level .05 by subjects or by items.

Finally, and more importantly for our hypothesis, we conducted two ANOVAs with the same factors on the reading times of the second clause. Table 2 reports the mean reading times in each condition.

---

1. In this experiment and in Experiment 2, ANOVAs were also performed for reading times measured in milliseconds/character. The results were identical with respect to the effect of gender; they showed minor differences with respect to the significance of some effects of congruency and verb bias in some of the analyses by subjects. Since these results are secondary to the main focus of this paper, we report the analyses of entire sentence reading times to allow for a direct comparison of our results with those of Garnham et al. (1992).
Each of the three factors had significant main effects: Second clauses were read more rapidly when they were preceded by two nouns of different gender, $F_1(1, 21) = 14.5, p < .01; F_2(1, 18) = 39.7, p < .01$, reproducing the facilitatory effect of the gender cue observed by Garnham et al. (1992) in the strategic condition of their Experiment 4. The congruent second clauses were read more rapidly than incongruent second clauses, $F_1(1, 21) = 13.44, p < .01; F_2(1, 18) = 10.34, p < .01$. Finally, participants read the subordinate clauses more rapidly when they followed VN1 verbs than when they followed VN2 verbs, but this effect was only significant in the subject analysis, $F_1(1, 21) = 11.7, p < .01; F_2(1, 18) = 3.44, p > .05$. This effect could reflect more difficult endings for VN2 verb sentences. No two-way or three-way interactions were significant at level .05 by subject or by item.

**Discussion**

The results of this experiment showed that readers benefited from a relevant gender cue at two levels: when they read the pronominal sentence, and when they answered questions about the pronominal referent. The facilitatory effect of the gender cue was mainly additive with the effect of congruency, as in the experiments reported by Garnham et al. (1992). The only measure that showed this interaction was the proportion of errors to the question of comprehension. We suggested that it could reflect a tendency to value speed over accuracy when a question followed a sentence where the two nouns had same gender. In this condition, readers made more errors with an incongruent ending, but they were not significantly slower to answer. The main result of this experiment is that readers integrated the second clauses more rapidly and more accurately when they contained a pronoun that agreed with only one of the mentioned referents.

Garnham et al. (1992) suggested that this gender effect reflected a strategic use of the gender cue because it was not observed in a condition where readers were not systematically questioned about pronouns. Experiment 2 attempts to replicate this second effect, with additional material with agreement errors introduced to see if the readers would still be sensitive to pronominal gender.

<table>
<thead>
<tr>
<th>TABLE 2</th>
<th>Mean reading times* for the subordinate clauses in Experiment 1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NP1 verb</td>
</tr>
<tr>
<td></td>
<td>Congruent</td>
</tr>
<tr>
<td>Gender</td>
<td>$M$</td>
</tr>
<tr>
<td>Same</td>
<td>2,024.6</td>
</tr>
<tr>
<td>Different</td>
<td>1,939.2</td>
</tr>
<tr>
<td>Overall mean</td>
<td>1,981.9</td>
</tr>
</tbody>
</table>

*aIn ms.*
EXPERIMENT 2

In Experiment 1, a strong effect of gender cue was observed on the reading times of pronominal clauses. Garnham et al. (1992, Exp. 4) reported that this effect became nonsignificant when they added filler items that did not require the resolution of a pronoun. Our goal in Experiment 2 was to replicate this aspect of their results. But we also wanted to check whether participants who did not strategically accelerate their reading when the pronominal gender cue was relevant would still be sensitive to a disagreement between a pronoun and the two same-gender nouns mentioned in the previous clause. Therefore we made a slight modification in Garnham et al.’s experiment: In their “nonstrategic” condition, the authors presented filler items without pronouns, and they retained the questions requiring pronoun resolution for the experimental items. This aspect of their experiment could be problematic for our experiment, because we wanted to include some sentences where the pronoun would not be correct (4.55% of the sentences presented to each participant). We suspected that the occurrence of pronominal questions on half of the material could make the participants sensitive to the disagreement mistakes. So a more stringent criterion was used concerning the comprehension questions: The questions about pronouns were eliminated and were replaced by questions about the first clause of the experimental material. However, half of the filler items (without pronouns) were followed by questions about the second clause, to ensure that participants read this clause.

Method

Materials

The same 80 sentences as those used in Experiment 1 were presented in Experiment 2. For each sentence, the yes/no comprehension question was modified to focus on the first clause. This question took the form of the first clause with either the same order of N1 and N2 (answer “yes”) or the reverse order (answer “no”).

Sixteen new sentences were constructed to test the sensitivity of readers to a gender disagreement. Eight sentences contained a VN1 verb, and eight sentences a VN2 verb. The verbs were selected from the set of 40 verbs that served in the pre-experiment reported in the Materials section of Experiment 1. They were different from the verbs used in the materials of Experiment 1, but they were selected to have a clear implicit causality bias. The main clause was followed by a causal subordinate clause introduced by “parce que”, as in the other 80 verbal materials. We chose to systematically construct a subordinate clause that was congruent with the verbal bias because this condition seemed more likely to avoid focusing attention on the subordinate clause of these sentences.

Eight VN1 verbs were selected because the grammatical subject was referred to by the pronoun for at least 65% of the participants for each verb; eight VN2 verbs were selected because the grammatical object was referred to by the pronoun for at least 60% of the subjects. The VN1 verbs were: énerver (annoy), tromper (mislead), inquiéter (worry), surprendre (surprise), déconcerter (disconcert), étonner (amaze), réjouir (delight), and amuser (amuse); the mean percentage of N1 choice for these verbs was 88.125% (SD = 11.08). The VN2 verbs were: réconforter (comfort), disputer (scold), consoler (console), dénoncer (denounce), complimenter (compliment), gifler (slap), secouer (shake up), and aider (help); the mean percentage of N2 choice for these verbs was 88.13% (SD = 14.32).

For each of the 16 selected verbs, a sentence was constructed where the main clause contained two first names of same gender (four masculine and four feminine) for the 8 VN1 verbs and for the 8 VN2 verbs. The first names used for these materials were different from those used for the other 80
experimental sentences. Sentences 19 and 20 are two examples of such sentences (with a VN1 verb and a VN2 verb respectively).

19. Franck a énervé Samuel parce qu’il refusait de travailler.
   Franck annoyed Samuel because he refused to work.
20. Géraldine a complimenté Charlotte parce qu’elle avait fait un effort.
   Wendy complimented Nancy because she made an effort.

The mean length of the subordinate clause (without counting the connective and the pronoun) was 3.87 words for both N1 and N2 verb sentences. For the experiment, the 16 sentences were divided into two blocks of eight sentences (two sentences with VN1 and “she”, two sentences with VN1 and “he”, two sentences with VN2 and “he”, and two sentences with VN2 and “she”). For each participant, the pronoun in one of the blocks was replaced by an incorrect pronoun (e.g., “il” – “he” in Sentence 19 was replaced by “elle” – “she”). Sentences 19 and 20 are examples of the verbal material used to test the disagreement sensitivity. In its disagreement version, Sentence 19 became: “Franck a énervé Samuel parce qu’elle refusait de travailler” (Franck annoyed Samuel because she refused to work). The comprehension question associated with these materials concerned the main clause (e.g., Did Franck annoy Samuel ?). A “yes” response was correct for 8 of the 16 sentences.

There were 80 filler items in this experiment. As in Garnham et al.’s (1992) experiment, these items were constructed with the same structure as that of the experimental sentences, but with a subordinate clause that did not contain a pronoun referring to one of the two characters mentioned in the main clause. The verbs appearing in main clauses were different from the verbs used in the experimental sentences, but we tried to use verbs that could intuitively have a causality bias (by using translations of the usual implicit causality verbs found in the English literature on this topic). Half of the filler items contained two same-gender names in the main clause, and the other half contained two different-gender names. Half of the filler items started with a masculine first name, and the other half started with a female first name. Sentences 21 and 22 are examples of these filler items.

21. Isabelle a admiré Thibault parce que le trophée était difficile à obtenir.
    Patricia admired Nicholas because the reward was difficult to obtain.
22. Sandrine a effrayé Marianne parce que la cascade était dangereuse.
    Virginia frightened Michelle because the stunt was dangerous.

The mean length of the second clause of these filler items (without counting the connective) was 4.7 words. Each filler sentence was associated with a yes/no comprehension question about the second clause for half of the filler items. The correct answer was “yes” for half of the questions.

Because each participant read only eight sentences with an incorrect pronoun, the global proportion of sentences containing an incorrect pronoun was 8/176 sentences (4.55%). If only sentences containing pronouns were taken into account, the proportion of sentences with incorrect pronouns was of course greater (8/96 or 8.33%), but it remained small.

Procedure

The procedure was similar to the procedure used in Experiment 1.

Design and subjects

The three variables used in Experiment 1 were again present in the design: the causality bias of the verb in the main clause (VN1 vs. VN2); the congruency of the subordinate clause relative to this bias
(congruent vs. incongruent); and the relevance of the gender of the two names in the main clause (same vs. different gender). The verb bias variable was a between-sentence factor; the two other variables were within-sentence factors. The three variables were within subject factors.

An additional variable was the agreement variable tested on 16 sentences. These 16 sentences were divided in two blocks of 8, as explained in the previous Materials section. Half of the subjects saw one block with a correct pronoun and the other block with an incorrect pronoun. The other half of the subjects met the reversed assignment. This design allowed the agreement variable to be varied within subjects and within sentences.

The 176 sentences were all presented to each participant. The four blocks of 20 sentences established for the Experiment 1 were preserved, and 20 filler items were added to each block. For the 16 items used to test the agreement sensitivity, four sets of 4 items were established so that 2 items in each sentence presented to each subject contained an incorrect pronoun. Each of these four sets was added to one of the four blocks.

The order of presentation of the blocks was systematically counterbalanced between the subjects. For each block, the order of presentation of the sentences was random (using the function “random presentation” of PsyScope).

A total of 22 subjects participated in this experiment (19 females, 3 males, mean age: 21.7 years; $SD = 2.27$). They were all French native speakers, students of the University of Poitiers. None of them had participated in Experiment 1.

Results

For the 80 filler items without pronouns, the mean percentage of errors to the comprehension questions was 5.51% ($SD = 3.07$), suggesting that participants carefully read the filler items.

Again, response times that were more than 10,000 ms were replaced by longest value of the subject in the condition. This affected only two response times and one reading time of a second clause. The times for which incorrect answers were given were replaced by the mean of the subject in the condition.

Results for the material in common with Experiment 1

The first results reported concern the part of the Experiment that was common to Experiment 1. Proportions of errors and response times are reported in Table 3.

<table>
<thead>
<tr>
<th>Gender</th>
<th>NP1 verb</th>
<th></th>
<th></th>
<th></th>
<th>NP2 verb</th>
<th></th>
<th></th>
<th>Overall mean</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Congruent</td>
<td>% error</td>
<td>RT</td>
<td>Incongruent</td>
<td>% error</td>
<td>RT</td>
<td>Congruent</td>
<td>% error</td>
<td>RT</td>
<td>Incongruent</td>
</tr>
<tr>
<td>Same</td>
<td>8.2%</td>
<td>1,482.1</td>
<td>9.1%</td>
<td>1,427.7</td>
<td>5.5%</td>
<td>1,373.9</td>
<td>7.2%</td>
<td>1,440.7</td>
<td>7.5%</td>
<td>1,431.1</td>
</tr>
<tr>
<td>Different</td>
<td>3.2%</td>
<td>1,336.2</td>
<td>3.2%</td>
<td>1,406.0</td>
<td>4.1%</td>
<td>1,357.8</td>
<td>2.7%</td>
<td>1,353.3</td>
<td>3.3%</td>
<td>1,363.3</td>
</tr>
<tr>
<td>Overall mean</td>
<td>5.7%</td>
<td>1,409.2</td>
<td>6.1%</td>
<td>1,416.8</td>
<td>4.8%</td>
<td>1,365.8</td>
<td>4.9%</td>
<td>1,397.0</td>
<td>4.8%</td>
<td>1,365.8</td>
</tr>
</tbody>
</table>

$^a$RT in ms.
As in Experiment 1, two ANOVAs were performed for the two measures. For response times, a main effect of gender was significant by subjects, $F_1(1, 21) = 7.59, p < .05; F_2(1, 18) = 4.01, p < .05$, indicating longer times to answer questions that followed sentences when the two nouns had the same gender than when they had two different-gender nouns. This effect could reflect the fact that questions concerned the main clause, where the two potential referents were introduced. Garnham et al. (1992), who found a similar effect, suggested that two people of different sexes are represented more distinctly than two people of the same sex. So readers could have fewer difficulties in answering a question when their mental representation allowed a clearer distinction among the referents in this representation. No other main effect or interaction effects were significant at level .05 by items or by subjects.

For the proportions of errors, the ANOVAs showed the same effect of gender, $F_1(1, 21) = 20.81, p < .01; F_2(1, 18) = 14.24, p < .01$. Readers made more errors when the question concerned a main clause where the two referents had the same gender. As in the previous experiment, we note that the accuracy measure seemed more sensitive than the time measure, perhaps reflecting a speed–accuracy trade-off favouring speed. No other main effect or interactions were significant at the .05 level by subjects or by items.

Two final ANOVAs were done on the reading times of the subordinate clauses. The mean reading times in each condition (with standard deviations by subjects) are reported in Table 4.

Contrary to Experiment 1, the main effect of gender was not significant, $F_1(1, 21) = 0.05; F_2(1, 18) = 0.08$. It is interesting to note that the main effect of congruency remained significant by subjects, $F_1(1, 21) = 6.26, p < .05; F_2(1, 18) = 2.3, p > .05$, with longer reading times for incongruent endings. We return to this effect in an analysis combining Experiments 1 and 2. However, Garnham et al. (1992) did not report a significant reduction of the congruency effect in their condition with “no resolution” fillers (similar to ours) relative to their condition requiring the resolution of pronouns. So this aspect of our results is consistent with their experiment. The main effect of verb bias was significant by subjects, $F_1(1, 21) = 23.1, p < .01; F_2(1, 18) = 2.44, p > .05$, as in the previous experiment, suggesting that subordinate clauses following VN2 were more difficult to integrate than subordinate clauses following VN1 verbs. No interaction effects were significant by subjects or by items at the .05 level.

Because our main result was a nonsignificant effect of the gender cue in Experiment 2, it was important to directly compare Experiments 1 and 2. This comparison was made in two ANOVAs involving the same factors as those in the previous ANOVA, but with Experiment

### Table 4

Mean reading times$^a$ for the subordinate clauses in Experiment 2

| NP1 verb | | NP2 verb | |
|----------|---------------------------------|---------------------------------|
|          | Congruent | Incongruent | Congruent | Incongruent | Congruent | Incongruent | Overall mean |
| Gender   | M | SD | M | SD | M | SD | M | SD | M | SD | Overall mean |
| Same     | 1,668.9 | 298.6 | 1,642.2 | 312.1 | 1,742.2 | 268.7 | 1,818.0 | 389.5 | 1,717.8 |
| Different| 1,652.2 | 331.4 | 1,692.6 | 289.4 | 1,673.7 | 306.3 | 1,829.3 | 352.2 | 1,711.9 |
| Overall mean | 1,660.3 | 1,667.4 | 1,707.9 | 1,823.6 |

$^a$In ms.
(1 and 2) as an additional factor (between subjects, but within sentences). The two ANOVAs concerned the reading times of the subordinate clauses. There was a main effect of Experiment, $F_1(1, 42) = 9.6, p < .01$; $F_2(1, 18) = 300.2, p < .001$. The reading times of the second clause were shorter in Experiment 2 ($M = 1714.9$ ms) than in Experiment 1 ($M = 2101.8$ ms). This effect could partly reflect the fact that subjects in Experiment 2 were more rarely questioned about the subordinate clause, so they read this part of the sentence more rapidly. The two-way interaction between the experiment and the gender factor was significant, $F_1(1, 42) = 9.15, p < .01$; $F_2(1, 18) = 18.7, p < .01$, confirming that the gender cue effect was significantly different in two experiments. Indeed, it was highly significant in Experiment 1, and it was nonsignificant at the .05 level in Experiment 2. So, as in Garnham et al.’s experiment, we found that the facilitation effect of the gender cue was limited to the condition where subjects were questioned about pronouns. It should be noticed that the effect of experiments was reliable for the two conditions of gender: in the same gender condition: $F_1(1, 42) = 11.47, p < .01$; $F_2(1, 18) = 217.73, p < .001$; in the different-gender condition: $F_1(1, 42) = 7.13, p < .05$; $F_2(1, 18) = 176.36, p < .001$. Therefore the global acceleration of reading times in Experiment 2 was not specific to the same-gender condition. It can be concluded that gender cue really accelerated reading times in Experiment 1, and that the absence of any questions about the second clause further accelerated reading times in Experiment 2. The two-way interaction between experiments and verb bias was not significant, $F_1(1, 42) = 0.84$; $F_2(1, 18) = 0.94$: The specific difficulty observed in Experiment 1 in reading the subordinate clauses following VN2 verbs was not influenced by the contextual modification between the two experiments.

Finally, the two-way interaction between experiments and congruency was only significant by items, $F_1(1, 42) = 3.47, p > .06$; $F_2(1, 18) = 5.21, p < .05$, suggesting a more important effect of congruency in Experiment 1 (mean difference between incongruent and congruent endings in Experiment 1: 600.3 ms; in Experiment 2, 61.3 ms). Garnham et al. (1992) did not report a significant interaction between their variation of context and the congruency effect. However a close examination of their data (cf. Garnham et al., 1992, p. 247) shows that the same trend existed, with a numerically greater effect of congruency in their condition involving “resolution fillers” (150 ms) than in their condition without such fillers (96 ms).

This result suggests that part of the congruency persisted when pronominal resolution was not required. At least two interpretations could be given for this result. First, an attempt to integrate the two clauses semantically persisted in Experiment 2. This attempt might only consider the content of the predicate in the second clause as a cause for the action or state mentioned in the main clause, perhaps without considering the identity of the participants. Second, despite our efforts to make congruent and incongruent endings as similar as possible, slight modifications of a few lexical items in incongruent endings could be sufficient to affect reading times, even without subjects attempting to integrate the two clauses. Garnham (2001, p. 137) suggested that part of the congruency effects observed in the literature could reflect “an artifact of the different materials used in the consistent and inconsistent sentences”. Although we agree with this conclusion, we note that the reduction of the congruency effect between our two experiments suggested that the congruency effect in Experiment 1 partly reflected the difficulty in integrating the two clauses.

Our main focus was on gender cue effect, for which the two experiments showed the clearest results (as in Garnham et al., 1992). The experiment factor did not enter in any three-way or four-way significant interactions at the .05 level (by subjects or by items).
To complete this comparison between the two experiments, two ANOVAs were also undertaken on the reading times of the main clauses. A main effect of Experiment was also found on this measure, $F_1(1, 42) = 6.59, p < .05; F_2(1, 18) = 493.0, p < .001$, showing longer reading times in Experiment 1 ($M = 1991.3$ ms) than in Experiment 2 ($M = 1,676.7$ ms). Garnham et al. (1992) reported a similar significant effect. They suggested that it was further evidence for strategic processing when readers were required to resolve pronouns: Subjects read the clause mentioning the two potential referents (and the implicit causality bias) more carefully. We note that the authors did not report a similar significant effect for reading times of subordinate clauses. But in their paper, each of the eight mean values reported for this measure was longer in their condition inducing a strategic effect of gender cue than in their condition that failed to induce this effect (cf. Garnham et al., 1992, Exp. 4, p. 247). The mean difference between the two conditions in their experiment was $230$ ms, suggesting that the effect found in our experiment was also present in theirs. This aspect of the results reported by Garnham et al. and of our own results suggests that the reduction of the gender cue effect occurs concomitantly with a general reduction of reading times.

Our results showed a replication of the main results reported by Garnham et al. (1992, Exp. 4) in favour of a strategic use of the pronominal gender cue. This cue speeded the reading time of the pronominal sentence only when readers could anticipate that the sentence would be followed by a question regarding the identity of the pronominal referent. This result is consistent with their not processing the pronominal gender cue in Experiment 2, which did not require identifying the referent of the pronoun. We now turn to the results of the materials in Experiment 2 where incorrect pronouns occurred.

Results for the sentences testing the sensitivity to a gender disagreement

Table 5 reports mean reading times for the subordinate clauses in the condition where the pronoun agreed and where the pronoun disagreed with the two mentioned nouns. In Table 5, the eight subordinate clauses presented to each participant are distinguished. We return to this aspect of the results in the subsection devoted to the evolution of the effects along Experiments 1 and 2.

The ANOVAs for the four measures used the factor agreement with two levels (agreement vs. disagreement); this factor was within subjects and within items. For the reading time of the first clause (agreement, $M = 1,634.9$ ms, $SD = 313.5$; disagreement, $M = 1,665.3$ ms, $SD = 353.1$), the main effect of agreement was not significant, $F_1(1, 21) = 0.75; F_2(1, 15) = 0.31$. For the reading time of the subordinate clause, the main effect of agreement was significant, $F_1(1, 21) = 40.95, p < .001; F_2(1, 15) = 56.87, p < .001$: The mean reading time for subordinate clauses containing an incorrect pronoun was approximately $430$ ms longer than the mean reading time of the same clauses containing a correct pronoun. Because the design used the same subordinate clauses in which only the pronoun was changed, this effect reflected sensitivity to the disagreement between the pronoun and the two nouns mentioned in the main clause. The proportions of errors to the comprehension questions (agreement, $M = 4.5\%, SD = 7.2$; disagreement, $M = 6.2\%, SD = 10.7$) did not significantly differ in the two conditions, $F_1(1, 21) = 0.46; F_2(1, 15) = 0.45$; these questions did not concern the clause containing the pronoun. The mean correct answer time was longer, $F_1(1, 21) = 18.9, p < .01; F_2(1, 18) = 22.6,$
\[ M = 1,792.5 \text{ ms}, \ SD = 608.9 \] than in the agreement condition \( (M = 1,338.7 \text{ ms}, \ SD = 245.7) \). This effect could reflect a spillover effect: Maybe readers who were disrupted by the subordinate clause with a pronoun in disagreement took more time to refocus their attention on the content of the following question.

These results showed a clear sensitivity to the gender disagreement, contrary to the non-significant gender effect noticed for the other materials. The 16 sentences used in this part of the study were different from the sentences used to test the strategic use of pronominal gender cue. Although we took care to construct very similar sentences in the two materials, it was possible that readers noticed new verbs in the first clauses of the 16 sentences and then decided to be more attentive to the subordinate clause of these sentences. To check this possibility, we conducted more specific ANOVAs on the reading times. The 16 sentences used to test the sensitivity to a gender disagreement were sentences with two nouns of same gender, with a VN1 or a VN2 verb in the main clause, and with a congruent subordinate clause. We compared their reading times with the reading times for the 48 sentences in the two conditions corresponding to this condition (the sentences with VN1 and VN2 verbs in the main clause, with two nouns of same gender, and a congruent ending). These reading times were compared with the reading times of the 16 sentences used to test gender sensitivity, but only in the agreement condition. The ANOVAs used a one-factor design with the three following levels: agreement condition for the 16 sentences used to test agreement sensitivity; congruent ending with same gender nouns and VN1 verbs for the 10 sentences used to test gender cue effect; and congruent ending with same gender nouns and VN2 verbs for the 10 sentences used to test gender cue effect. This factor was within subjects and between sentences. For the main clauses, the mean reading times that were compared were \( 1,634.9 \text{ ms} (SD = 313.5), 1,712.8 \text{ ms} (SD = 400.4), \) and \( 1,693.4 \text{ ms} (SD = 417.8), \) respectively. The main effect was not significant, \( F(1, 42) = 1.08, p > .05; F(2, 33) = 1.12, p > .05, \) suggesting that readers did not spend more time on the first clauses for the 16 sentences used to test agreement sensitivity than on the other comparable materials used to test the strategic use of gender cue. For the subordinate clauses, the mean reading times under comparison were: \( 1,693.7 \text{ ms} (SD = 307.7), 1,668.9 \text{ ms} (SD =

<table>
<thead>
<tr>
<th>Sentence</th>
<th>Agreement M</th>
<th>Agreement SD</th>
<th>Disagreement M</th>
<th>Disagreement SD</th>
<th>Overall mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2,023</td>
<td>704.4</td>
<td>2,540</td>
<td>906.6</td>
<td>2,281</td>
</tr>
<tr>
<td>2</td>
<td>1,787</td>
<td>435.5</td>
<td>2,321</td>
<td>711.1</td>
<td>2,054</td>
</tr>
<tr>
<td>3</td>
<td>1,568</td>
<td>436.5</td>
<td>2,134</td>
<td>1,085.0</td>
<td>1,851</td>
</tr>
<tr>
<td>4</td>
<td>1,683</td>
<td>384.7</td>
<td>2,222</td>
<td>709.7</td>
<td>1,952</td>
</tr>
<tr>
<td>5</td>
<td>1,743</td>
<td>508.9</td>
<td>2,223</td>
<td>945.2</td>
<td>1,983</td>
</tr>
<tr>
<td>6</td>
<td>1,619</td>
<td>541.9</td>
<td>1,928</td>
<td>512.2</td>
<td>1,774</td>
</tr>
<tr>
<td>7</td>
<td>1,557</td>
<td>405.5</td>
<td>1,800</td>
<td>490.5</td>
<td>1,679</td>
</tr>
<tr>
<td>8</td>
<td>1,569</td>
<td>453.0</td>
<td>1,829</td>
<td>932.0</td>
<td>1,699</td>
</tr>
<tr>
<td>Overall mean</td>
<td>1,693.7</td>
<td>2,124.6</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(^{a}\text{In ms.}\)
331.4), and 1,742.1 ms (SD = 306.3), respectively. Again, the main effect was not significant, $F_1(2, 42) = 1.48, p > .05$; $F_2(2, 33) = 0.48$. So it seems that the sentences used to test gender agreement were not significantly different from the other comparable materials used in Experiment 2.

Evolution of the effects in Experiments 1 and 2

An anonymous reviewer invited us to consider an interesting possibility about the disagreement sensitivity noted in Experiment 2. The reviewer suggested that strategic factors might develop with repeated exposure to gender disagreement and that readers might not be sensitive to the disagreement on the first instance of an incorrect pronoun. Therefore, we decided to examine whether there were differences between the first and subsequent encounters with a disagreement. For each participant, the eight successive instances of a pronoun in agreement or in disagreement were extracted. The mean reading times of the second clause of these sentences are reported in Table 5. An ANOVA using subjects as a random variable was performed on these data. The successive eight positions and the agreement factor were both within-subjects factors. There was no significant two-way interaction between the agreement factor and the successive instances, $F_1(7, 147) = 0.74$; this two-way interaction was still not significant when the design was restricted to the two first instances. Moreover, the agreement effect was significant for the first instance, $F_1(1, 21) = 7.86, p < .05$, suggesting that readers were immediately sensitive to the disagreement. This result was consistent with an automatic gender sensitivity.

The same reviewer suggested that the comprehension questions used in Experiment 2 could give rise to a gender strategy. Because most of these questions were about the first clause, and the relative order of the two names in this clause, the readers could strategically decide to remember this clause as “female verbed male” when it mentioned two different gender people. This strategy would not be logically relevant to the items testing gender sensitivity because the two names had same gender, but it could induce a special way of processing gender information, even in this nonstrategic condition. It is difficult to establish good objective evidence regarding the occurrence of such a strategy. However, it could result in faster reading times for the second clause, and more accurate and faster answers, in the different gender condition towards the end of Experiment 2. To test this prediction, for each participant, we calculated the mean proportions of errors and response times for each of the four blocks of the materials. These means were calculated for the experimental materials used to test the effect of the gender factor (i.e., same vs. different). Each block contained one quarter of the verbal materials, so the means involved 10 items in each block. The analysis by blocks was also applied to Experiment 1, in which we and Garnham et al. (1992) have argued that a strategic effect of gender is found. Table 6 reports the means for each block in the two experiments.

Three dependent variables were considered: error proportion, response times to the comprehension questions, and reading times of the second clause. For each dependent variable, an ANOVA by subjects was computed, with two factors: gender of the names in the first clause (same vs. different), and blocks (Block 1, 2, 3, and 4). The two factors were within subjects. The ANOVA by items could not be computed because the group size (22 participants) did not allow each of the four quarters of the materials to be equally represented in each of the four blocks.
The interaction of gender and blocks was not significant in any of these measures in Experiment 2. For Experiment 1, a significant two-way interaction, $F_{1}(3, 63) = 5.46, p < .01$, was obtained on reading times of the second clause. The effect of the gender factor was not significant for Block 1, $F_{1}(1, 21) = .05$. However, reading times became significantly longer in the same-gender condition than in the different-gender condition for the three following blocks: Block 2, $F_{1}(1, 21) = 5.19, p < .05$; Block 3, $F_{1}(1, 21) = 28.05, p < .01$; Block 4, $F_{1}(1, 21) = 5.48, p < .05$. This result confirmed the emergence of the gender effect over time in Experiment 1, consistent with its strategic nature. For the two other measures, the two-way interaction between blocks and agreement failed to be significant at level .05: response time, $F_{1}(3, 63) = 0.2$; error proportion, $F_{1}(3, 63) = 2.61$. For this last measure, we should notice that error rate was significantly higher in the same-gender condition for the four blocks, but there was some evidence in favour of a greater effect in Block 1, $F_{1}(1, 21) = 41.05, p < .001$, than in the other blocks: Block 2, $F_{1}(1, 21) = 12.17, p < .01$; Block 3, $F_{1}(1, 21) = 11.03, p < .01$; Block 4, $F_{1}(1, 21) = 19.05, p < .01$. This trend could suggest that readers devoted more attention to the semantic implications of the second clause (in the same-gender condition) after the Block 1. For Experiment 2, the interaction between blocks and gender was not found for the three measures: response time, $F_{1}(3, 63) = 0.47$; error proportion, $F_{1}(3, 63) = 0.04$; reading time of clause 2, $F_{1}(3, 63) = 1.55, p > .05$.

In summary, the gender effect observed in Experiment 1 became significant only when readers had read one quarter of the material, which is consistent with an effect reflecting a strategic process. We could find no trace of a gender strategy emerging along the blocks in Experiment 2. The gender disagreement effect in Experiment 2 was significant at the first instance of

### Table 6

Mean reading times for the subordinate clause, percentage of errors, and correct response times as a function of the four successive blocks for the materials shared by Experiments 1 and 2.

<table>
<thead>
<tr>
<th>Exp. Gender</th>
<th>Block</th>
<th>M</th>
<th>SD</th>
<th>M</th>
<th>SD</th>
<th>M</th>
<th>SD</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Same</td>
<td>1</td>
<td>2,248.6</td>
<td>563.8</td>
<td>2,198.4</td>
<td>619.4</td>
<td>2,164.6</td>
<td>546.6</td>
<td>2,089.1</td>
<td>631.9</td>
</tr>
<tr>
<td>Different</td>
<td></td>
<td>2,261.7</td>
<td>463.6</td>
<td>2,041.7</td>
<td>518.1</td>
<td>1,861.7</td>
<td>493.7</td>
<td>1,948.1</td>
<td>557.1</td>
</tr>
<tr>
<td>Same</td>
<td>2</td>
<td>1,924.8</td>
<td>370.3</td>
<td>1,687.2</td>
<td>314.5</td>
<td>1,724.5</td>
<td>354.7</td>
<td>1,534.6</td>
<td>337.5</td>
</tr>
<tr>
<td>Different</td>
<td></td>
<td>1,914.5</td>
<td>319.8</td>
<td>1,714.2</td>
<td>332.9</td>
<td>1,636.1</td>
<td>301.6</td>
<td>1,582.7</td>
<td>404.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1,565.8</td>
<td>344.7</td>
<td>1,456.1</td>
<td>365.8</td>
<td>1,371.7</td>
<td>273.9</td>
<td>1,338.4</td>
<td>289.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1,457.7</td>
<td>325.1</td>
<td>1,404.1</td>
<td>327.8</td>
<td>1,307.0</td>
<td>269.2</td>
<td>1,284.3</td>
<td>348.1</td>
</tr>
</tbody>
</table>

*aIn ms.

*bIn percentages.
an incorrect pronoun, consistent with an immediate and automatic sensitivity to pronominal gender agreement with antecedents in the context.

Discussion

Experiment 2 replicated the main result reported by Garnham et al. (1992, Exp. 4). In a condition where subjects were not required to systematically resolve the reference of pronouns, they did not show longer reading times for subordinate clauses where the pronoun agreed with the two potential referents than for subordinate clauses where the pronoun agreed with only one of the potential referents. This contrasted with the gender cue effect observed in the Experiment 1 where subjects were systematically questioned about pronouns. This result confirmed that part of the pronominal gender effect in Experiment 1 was strategic. Readers took advantage of the gender cue to accelerate their reading time when this cue was sufficient to resolve the pronoun. The analysis suggested by an anonymous reviewer allowed us to find a new argument in favour of the strategic status of this acceleration: It emerged only after the readers met one quarter of the verbal materials. However, in Experiment 2, we found that the readers were still sensitive to a gender disagreement between the pronoun and the potential referents mentioned in the previous clause, even to the first occurrence of this disagreement. This result is consistent with the intuition of Garnham et al. (1992) who suggested that even in a condition where a gender cue was not strategically used, readers could still be sensitive to the gender information contained in the pronoun.

This result is also consistent with the study reported by Rigalleau and Caplan (2000), because these authors found an immediate sensitivity to a gender disagreement in a condition where the subjects had no questions about pronouns. It extends their observations in two ways. First, Rigalleau and Caplan measured the naming latency of the pronoun. This measure, taken immediately on the pronoun, could have an undesirable effect. By forcing participants to focus on pronouns, it could induce controlled processing of the pronoun. In our Experiment 2, participants read the entire pronominal clause, so they were not forced to read pronouns aloud or in isolation. The sensitivity to a gender disagreement was still observed. Second, Rigalleau and Caplan (2000, Exp. 2) found a strategic effect on pronominal naming latency, which was reversed relative to the strategic effect noticed by Garnham et al. (1992). Naming latency was longer when the pronoun agreed with only one noun in the main clause than when it agreed with two nouns. Here, we observed a strategic gender cue effect, which had the same direction as Garnham et al.’s effect: The reading time of the subordinate clause was shorter when it contained a pronoun agreeing with only one noun in the main clause.

To explain the different directions of the effects on the two measures, we propose that the strategic gender cue effect on naming latency reflects subjects strategically disengaging their attention from a pronoun that disagrees in gender from a possible antecedent. This process is important in a context where the two potential referents are highly activated in working memory. In a probe recognition task, Garnham et al. (1996) failed to show a shorter recognition time for the implicit cause when probes were tested before the pronoun. Rigalleau and Caplan (2000) showed that the naming latency of a pronoun referring to the implicit cause was not shorter than the naming latency of a pronoun referring to the other potential referent in the context. These results show that under these conditions, both nouns are equally highly accessible; therefore they both can be candidates for the pronominal reference. The disengagement
process is important in order to reduce the activation of the nonreferent, an essential operation when readers know that they will be questioned about the identity of the pronominal referent.

We suggest that this disengagement process allows subjects to devote more effort to the semantic integration of the words following the pronoun. This is the reason why the reading time of the entire pronominal clause is shorter when gender cue is relevant. The disengagement process is strategic and does not occur when subjects are not questioned about the identity of the pronoun. Experiment 2 confirmed this prediction: Without any questions about the pronouns, the relevant gender cue did not significantly accelerate the reading time of the subordinate clause (relative to a condition where gender was not relevant for pronoun resolution).

An anonymous reviewer asked why the relevant gender cue did not accelerate reading times in Experiment 2 relative to Experiment 1. As noted above, the presence of questions about pronouns in Experiment 1 may have induced participants to take into account the specific actors and themes in each clause when considering the likelihood of the second clause serving as the cause of the first in Experiment 1, and the absence of questions about pronouns in Experiment 2 may have induced participants to only consider whether the activity reported in the second clause could, in general, be a cause of the activity reported in the first clause in that experiment. Evidence that subjects processed the second clause more deeply in Experiment 1 than in Experiment 2 is that the congruency effect was less marked in Experiment 2 than in Experiment 1 (at least in the item analysis). This additional semantic processing would lead to overall longer reading times for the second clause in Experiment 1. Though the disengagement of the pronoun from the gender-mismatched possible antecedent would accelerate second-clause reading times in Experiment 1 in the different compared to the same gender condition, it would not obviate this additional semantic processing of the second clause in Experiment 1 compared to Experiment 2.

To return to the main theme, Experiments 1 and 2 showed that readers automatically check whether there is an agreement between a pronoun and at least one of the nouns mentioned in the previous clause. In the Introduction, we suggested an additional property of this automatic gender coindexation: It checks the agreement between the pronoun and the most accessible entities in the discourse representation. This property could not be tested in the materials of the two previous experiments, because the two potential referents were probably equally activated in working memory when the pronoun was encountered. Experiments 3 and 4 use longer texts where the relative accessibility of the two potential referents was easier to vary.

EXPERIMENT 3

Greene et al. (1992) suggested that when a discourse has only one entity in the focus of attention at the time a pronoun is encountered, then it might be that essentially no processing is required for the pronoun. Information that follows the pronoun could be automatically attached to the focused entity. As mentioned in the Introduction, this hypothesis was consistent with many of their probe recognition task results because they showed that the level of activation of the pronominal referent was not modified after encountering the pronoun (Greene et al., 1994). Furthermore, replacing the pronominal sentence by an “allusion” sentence with approximately the same semantic content but without a pronoun did not change the activation of the pronominal referent. This view of automatic pronominal processing must
be tested under the stringent conditions that Greene et al. (1992) imposed to avoid any strategic processing of the pronoun. First, readers should not have any special motivation to pay attention to pronouns; no comprehension question should be asked that could require the identification of the intended referent for the pronoun. Second, Greene et al. (1992) estimated that some probe recognition experiments showing a pronominal effect used slower than normal reading time presentation for the words in the texts; they emphasized that such slow presentation times could allow readers to apply slow strategic processes. Third, texts should be long enough to avoid readers paying specific attention to the pronominal information.

In the two next experiments, all the conditions to avoid strategic processing of pronouns were fulfilled. The following sample illustrates the experimental materials used in Experiment 3:

French:
Maxime et Carole rentrent de leurs vacances en Suisse.
Comme toujours, les services de douane vont vérifier les papiers.
Les marchandises de valeur doivent être déclarées.
Avant de passer, Maxime décide de dissimuler des bijoux en or.
Il doit faire preuve de sang froid.
Carole propose à Maxime de mettre les bijoux dans son chignon.

Referent probe: Maxime
Nonreferent probe: Carole

Question de compréhension: Maxime et Carole reviennent de Belgique?

Translation:
John and Mary are returning from their vacation in Switzerland.
As usual, the customs department will check the identity papers.
Articles of value must be declared.
Before going through customs, John decides to hide gold jewellery.
He must keep calm.
Mary suggests that John should put the jewels in her hair.

Referent probe: John
Nonreferent probe: Mary

Comprehension question: Have John and Mary come back from Belgium?

The two probe positions used in Experiment 3 are marked with numbers. For the gender factor variation, “John” in the fourth sentence was replaced by “Mary” (or “Maxime” by “Carole” in the French material), and correspondingly “he” was replaced by “she” (or “il” by “elle” in the French material); the two probes were reversed in this case.

Every experimental text contained six sentences, and the pronoun appeared in the fifth sentence for experimental texts (many filler texts did not contain pronouns in this position). Each line of the text was presented in succession. Participants pushed a button to make the next line appear (and the previous line disappear). Readers self-regulated the speed of presentation. There were no questions about the sentence (fifth sentence) containing the pronoun for the experimental texts. For filler texts, there were some questions about the fifth sentence, but this sentence did not contain pronouns. For all the texts, there were no questions about the identity of a pronominal referent in the texts.
Finally, Greene et al. (1992) suggested that “normal” pronouns occur when there is only one potential referent in the focus of attention. As illustrated in the previous sample, there were two potential referents mentioned in our text, and they both could be related to the predicate attached to the pronoun (i.e., “to keep calm”). However only one potential referent was mentioned in the sentence that preceded the pronoun; the other referent was distant from the pronoun. We predicted that only the closer potential referent would be highly activated in working memory when the pronoun was encountered. This prediction was checked in Experiment 3 with a probe recognition task, with two probe positions: just before the pronoun, and after the pronominal sentence. These two positions are indicated with a digit in the sample.

In Experiment 3, three factors were varied: the probe (referent vs. nonreferent), the position of the probe (immediately before pronoun vs. after the pronominal sentence), and the gender of the pronoun (feminine vs. masculine). This last factor was varied by replacing the name mentioned in the fifth sentence by the opposite referent (i.e., “John” was replaced by “Mary”), and also changing the gender of the pronoun. This last variation was possible because the predicate following the pronoun could concern a woman or a man. It was a way to obtain probe recognition times for the same words either as a distant referent or as a close referent. This factor should have no significant effect. The main predictions were that (1) the close referent should be more activated than the distant referent before encountering the pronoun, and (2) the activation advantage of the closer referent should not be changed in the second probe position. This second prediction is consistent with the no-pronoun-processing thesis: The pronoun should not modify the level of activation of the most highly activated referent.

**Method**

**Materials**

The experimental material consisted in 48 texts composed of six sentences. (See sample sentences.) The first sentence of each text mentioned two characters that were combined in a coordinate subject noun phrase. They were introduced as two first names: One was masculine, and the other was feminine. The two first names were familiar, clearly marked in gender, and had the same number of letters (mean 6.6 letters, \(SD = 1.17\), min. 4, max. 9). The order of the two names was masculine–feminine for 24 texts, and feminine–masculine for 24 texts. This introductory sentence was followed by two filler sentences that did not mention or refer to the two characters. The filler sentences described some aspects of the situation. The fourth sentence referred to only one of the two characters by means of a repeated noun in subject position. This character was systematically varied as a factor of the experiment; it could be the feminine name or the masculine name. The only variation between the two versions of this sentence concerned the first name used in the sentence. The fifth sentence (the “pronominal sentence”) always started with a subject pronoun (“il” or “elle”, “he” or “she”) that was in agreement with the first name mentioned in the previous sentence. Besides the first name in the fourth sentence, there were no names of characters in the four first sentences that could be taken as antecedent of the referent of the pronoun. The pronominal sentence contained between six and eight words (mean = 7; \(SD = 0.35\)). The first word following the pronoun was always “doit” (“must”), and the clause following this modal verb was semantically consistent with either of the two characters mentioned in the text. In particular, it could be semantically attached either to a woman or to a man. Besides the pronoun, there were no other words in the pronominal sentence that referred to one of the two characters, and agreement inflection on the words (for example adjectives or verbs) had no gender value that could inform readers of the gender of the
pronominal subject. A final sentence re-mentioned both characters and informed readers of the conclusion of the text. In this final sentence, the two characters were mentioned by their first names. This sentence was mainly included to allow comprehension questions about a sentence following the pronominal sentence. The reason for this last sentence was that we wanted readers to distribute their attention to the entire text, avoiding the potential strategy of not reading the end of the text because no question was asked about it.

The mean distance between the pronominal referent and the pronoun was 6.52 words (SD = 1.01, min. 5, max. 8). The mean distance between the nonreferent and the pronoun was 35.5 words (SD = 1.4, min. 33, max. 38).

Each experimental text was associated with two probe nouns corresponding to the referent of the pronoun (i.e., the closer first name) and to the nonreferent (the distant first name). These probes were presented either immediately before the pronoun or immediately after the pronominal sentence. Each participant saw only one condition among the four resulting from crossing the two variables (probe noun and probe position). So for the experimental texts, each participant saw only one probe. Each text was also associated with one yes/no comprehension question. Eight texts had a question about the first sentence, eight about the second sentence, eight about the third, and eight about the fourth sentence, and the 16 remaining questions concerned the sixth (final) sentence. None of the exact answers required identifying the correct referent of the pronoun in the fifth sentence. The correct answer was “yes” for 24 questions and “no” for 24 questions. This equal partitioning of yes and no answers was also applied for the questions concerning each sentence.

There were 48 filler texts, each with six sentences. A total of 24 texts had approximately the same structure. Among them, 12 mentioned two same-gender first names in the first sentence, and 12 mentioned two different-gender first names in this sentence. The fourth sentence re-mentioned only one of the two names (as in the experimental texts). However, the fifth sentence did not contain an anaphoric pronoun referring to one of the two characters, but an expletive pronoun that could not refer to any previous noun in the text. The final sentence mentioned the two names (as in the experimental texts). The remaining 24 filler texts had a different structure. They did not contain characters introduced by first names, and they did not contain any anaphoric pronoun. A yes/no comprehension question was associated with each filler text. As for experimental texts, this question concerned the first sentence for eight texts, the second for eight texts, the third for eight texts, and the fourth sentence for eight texts; the 16 remaining questions concerned the fifth sentence, which did not contain an anaphoric pronoun.

Each filler text was associated with two probes, for a total of 96 probes. For the filler texts, the two probes were systematically presented during the reading of the text. A total of 30 probes were positive (corresponding to a common noun mentioned in the text). A total of 66 probes were negative (absent from the texts), 26 were common nouns, and 40 were first names. So globally the proportion of positive probes in the experiment was 78/144 = 54.17%. The probes in the filler texts had fixed positions of presentation; 24 probes were presented just after the third sentence, 24 after the fourth sentence, 24 probes after the fifth sentence, and 24 after the sixth sentence. Taking into account experimental and filler texts, the proportion of positive probes after the third sentence was 11/24, after the fourth sentence, 28/48, after the fifth sentence, 28/48, and after the sixth sentence 11/24.

Procedure

All the texts and probes were presented on the screen of a Macintosh Computer. The experiment ran on Psyscope (Cohen, MacWhinney, Flatt, & Provost, 1993). Response time data were obtained with a button box composed of three buttons. The participant first pressed the middle yellow button to trigger the appearance of the first sentence. The participant pressed on the same button to see the next sentence. This press made the first sentence disappear, and the second sentence appeared in the same position as that of the previous sentence. The time of appearance of the sentence was measured: It was called the
“reading time” of this sentence. The display of each experimental text was interrupted once for presentation of a probe word. This interruption occurred either immediately before the pronominal sentence, or immediately after the pronominal sentence. The display of each filler item was interrupted twice for presentation of probe words. When a probe word was presented, it appeared alone on the screen in all uppercase red letters. The probe word remained on the screen until a response button was pressed (right green button for “yes” the word had appeared in the text, and left red button for “no” the word had not appeared in the text). After the final sentence of a text, or after the probe following this sentence for 24 filler texts, the comprehension question associated with the text was presented on the screen. Participants used the right green button or the left red button to answer yes or no to this question. The question disappeared when participants answered, and it was replaced by feedback about the expected correct answer. This feedback appeared in blue letters. Pressing on the middle yellow button made this feedback disappear, the message “Next trial” was presented in the middle of the screen, and the reader could begin to read the next text. The order of presentation of the experimental texts and the filler texts was randomized for each subject, using the random order option of Psychophysics software.

Subjects were tested individually in a quiet room. Participants were instructed to read at their normal speed and to respond as quickly as they could without making mistakes. The requirement of a rapid answer was emphasized for the probe recognition task.

**Design and subjects**

There were three variables for the experimental texts: the probe word (pronominal referent vs. nonreferent); the location of the probe word (immediately before the pronominal sentence vs. immediately after the pronominal sentence); and the gender of the pronoun (masculine vs. feminine). The eight conditions formed by crossing these three variables were combined in a Latin square design with eight sets of texts (six per set) and with eight groups of subjects (two in each group). A total of 16 subjects (14 females, 2 males; mean age 28.1 years, $SD = 6.9$) participated in this experiment. They were French native speakers, and none of them had participated in the previous experiments.

**Results**

The mean proportion of errors to the comprehension questions was 7.8% for experimental texts and 8.3% for the filler texts, indicating a good comprehension of the texts. For the probe task, for the positive filler items, mean recognition times and error rates were 1,038.8 ms and 8.9%, respectively, and for the negative filler items they were 959.8 ms and 2.36%, respectively.

Means were calculated for each subject and each item in each condition. When the experimental probes were incorrectly rejected as not mentioned in the text, the recognition time was replaced by the mean recognition time of the subject in the same condition. Table 7 reports the means of the recognition times, with means of the error rates.

ANOVA were conducted on both subjects ($F_1$) and item ($F_2$) means. The three factors were within subjects and within items. For response times, there was a main effect of probe, $F_1(1, 15) = 29.1, p < .01; F_2(1, 47) = 52.9, p < .01$; probes corresponding to the nonreferent (i.e., the more distant referent) inducing longer response times than probes corresponding to the referent. No other main effect or interaction effects were significant at the .05 level by subjects or by items. This effect of probe was also reflected in error rates, $F_1(1, 15) = 22.4, p < .01; F_2(1, 47) = 49.6, p < .01$, with more errors on the most distant probe. The mean error rate for this distant referent was 12.5%, suggesting that it was not completely forgotten by the subjects, even if it was clearly less accessible than the pronoun referent. Again, no other main
effect or interaction effects were significant at the .05 level for error rates. We also analysed the reading times of all the sentences in the texts, using the same factors. No effects were significant at the .05 level for the seven sentences.

Discussion

The results of Experiment 3 showed that our texts placed the closer referent in the focus of attention. This referent was more activated than the distant referent before the pronoun was encountered, and this advantage of accessibility persisted after the pronominal sentence. The interaction between probe position and probe word was not significant \( (F_s < 1) \). Therefore we can estimate that these texts are good instances of texts where no processing of a pronoun should occur, according to model presented by Greene et al. (1992): One referent is more accessible than the other, and reading the pronominal sentence does not change this advantage of accessibility.

In Experiment 4, we used the same texts to test the hypothesis that readers could be sensitive to a gender disagreement between a pronoun and the most accessible antecedent in working memory.

### EXPERIMENT 4

Experiment 3 showed that our texts made one of the two potential referents more active than the other at the time the pronoun was encountered. So these texts show one of the crucial properties that should permit “no processing for the pronoun” according to Greene et al. (1992, p. 281). The nonsignificant interaction between probes and probe positions is further reason to claim that these texts are good candidates to test the “no processing” hypothesis. The levels of activation of the two potential referents were not changed after reading the pronominal sentence, suggesting that the pronoun had no effect on this measure.

According to the “no processing” hypothesis, in conditions that do not induce strategic processing of a pronoun, readers directly attach the predicate of the pronoun to the entity in the focus of attention. If the entity in the focus does not agree in gender with the pronoun, readers would not be sensitive to this disagreement. The hypothesis of an automatic gender
coindexation process makes a different prediction. It predicts that even in such conditions, the gender of the pronoun is automatically matched with the gender of the entity in the focus of attention. Readers should have a longer reading time if the pronominal sentence contains a pronoun that disagrees with the entity in focus. So Experiment 4 compared a condition where the pronoun agreed with the referent in focus and a condition where it agreed with the least accessible of the two possible referents.

Another condition was included in Experiment 4, as illustrated in the following sample:

French:
Maxime et Carole rentrent de leurs vacances en Suisse.
Comme toujours, les services de douane vont vérifier les papiers.
Les marchandises de valeur doivent être déclarées.
Avant de passer, Maxime/Carole décide de dissimuler des bijoux en or.
Il/Elle/Maxime/Carole doit faire preuve de sang froid.
Carole propose à Maxime de mettre les bijoux dans son chignon.

Question de compréhension: Maxime et Carole reviennent de Belgique?

Translation:
John and Mary are returning from their vacation in Switzerland.
As usual, the customs department will check the identity papers.
Articles of value must be declared.
Before going through customs, John/Mary decides to hide gold jewellery.
He/She/John/Mary must keep calm.
Mary suggests that John should put the jewels in her hair.

Comprehension question: Have John and Mary come back from Belgium?

Words in bold characters correspond to the different options tested in Experiment 4. The eight resulting versions were all tested.

The pronoun was replaced by a repeated noun corresponding to the referent that was in gender agreement with it. This condition allowed a further test of the accessibility difference between the two potential referents. Gordon, Grosz, and Gilliom (1993) observed that when the antecedent is highly accessible (e.g., when it is the grammatical subject of the sentence preceding the pronoun), there is a “repeated noun penalty”: Reading time of an anaphoric sentence referring to this antecedent is longer when the anaphor is a repeated noun than when it is a pronoun. This repeated noun penalty is not observed when the referent is less accessible.

The experiments reported by Gordon et al. used short passages where the least accessible referent was mentioned in one of the two preceding sentences. So their variation of accessibility could be weaker than ours: The least accessible referent in the sample text is mentioned four sentences before the anaphora. In such a condition, we predicted that the repeated noun could be more easily processed than a pronoun. In other words, the repeated noun penalty would be replaced by a pronoun penalty for the least accessible referent.
Method

Materials

The experimental material consisted of the same 48 texts as those used in Experiment 3. The gender of the referent in the fourth sentence was systematically varied (as in Experiment 3). The anaphoric form starting the fifth sentence was varied: It could be either one of the first names mentioned in the text or a pronoun. The last variable concerned the distance between the anaphor and its antecedent (i.e., the referent in gender agreement for pronominal anaphora). In Experiment 3, this last factor was not varied; the pronoun systematically referred to the closer referent. In Experiment 4, the anaphor could refer either to the closer referent (mean distance: 6.52 words) or to the distant referent (mean distance: 35.5 words). The 48 filler items used in Experiment 3 were again used in Experiment 4.

The comprehension questions used in Experiment 4 were the same as those in Experiment 3. There was no probe recognition task in Experiment 4, so no probes were associated with the experimental and filler texts.

Procedure

The procedure was similar to that of Experiment 3, except that no probe recognition task was required: No probe appeared on the screen. Only reading times of the sentences and answer to the questions were collected. As in Experiment 3, participants were instructed to read at their own normal speed and to respond to questions “as quickly as possible, without making mistakes”.

Design and subjects

There were three variables for the experimental texts: the anaphoric form (pronoun vs. repeated noun); the gender of the first name, which was close to the anaphor (masculine vs. feminine); and the distance between the last mention of the anaphoric referent and the anaphor (close vs. distant). The eight conditions formed by crossing these three variables were combined in a Latin square design with eight sets of texts (six per set) and with eight groups of subjects (three in each group). A total of 24 subjects (20 females, 4 males; mean age 26.5 years, SD = 3.9) participated in this experiment. They were all native French speakers. None of them had participated in the previous experiments.

Results

For comprehension questions, the mean error rate was 8.3% (8.07% for filler texts). The reading times for texts with incorrect answers were not excluded (contrary to Experiments 1 and 2). Because texts were long, it seemed inappropriate to decide that an error on a question about a fact reflected a complete failure of the text comprehension.

We only report the ANOVAs on the reading times of the anaphoric sentence (fifth sentence) in the experimental texts. Table 8 reports the mean reading times of this sentence in the eight conditions that were compared.

The main effect of distance was significant, $F_1(1, 23) = 64.9, p < .001; F_2(1, 47) = 67.1, p < .001$. The reading time of the anaphoric sentence was longer when the anaphor referred to the distant referent than when it referred to the close referent. However, there was a significant two-way interaction between the anaphoric form and distance, $F_1(1, 23) = 23.06, p < .01; F_2(1, 47) = 36.5, p < .01$. Examination of the terms of this interaction showed that the distance effect was not significant for repeated nouns, $F_1(1, 23) = 1.67, p > .05; F_2(1, 47) = 2.11, p > .05$, but
that there was a significant effect of distance for pronouns, $F_1(1, 23) = 68.2, p < .001; F_2(1, 47) = 92.7, p < .001$. This last effect reflects the fact that when pronouns did not agree with the most accessible antecedent, readers had longer reading times for the sentence ($M = 2,400.5$ ms) than when the pronoun agreed with it ($M = 1,937.5$ ms). In addition, the effect of the anaphoric form differed as a function of the distance factor. When the anaphor referred to the close referent, reading time was shorter for a pronoun than for the repeated noun, $F_1(1, 23) = 19.6, p < .01; F_2(1, 47) = 18.08, p < .01$: This effect corresponds to a repeated noun penalty. On the other hand, when the anaphor referred to the distant referent, the reading time was longer for a pronoun than for a repeated noun, $F_1(1, 23) = 8.02, p < .01; F_2(1, 47) = 15.56, p < .01$. This last effect is difficult to interpret because it is contaminated by the sensitivity to the gender disagreement between the close referent and the pronoun. However, it is consistent with the corpus data on production that reveal a preference for repeated nouns referring to distant referents (cf. Marslen-Wilson, Levy, & Tyler, 1982). No other main effects or interaction effects were significant at the .05 level by subjects or by items. The results of this ANOVA did not change when reading times were expressed as milliseconds/character, to take into account the fact that names were usually longer than pronouns.

As in the previous experiment, we also analysed the reading times of the other sentences in the experimental texts. There were no significant effects for the four first sentences.

As with Experiment 2, we examined the potential evolution of the agreement effect according to the successive instances of pronouns during the Experiment 4. An ANOVA with subjects as a random variable was computed: The instance factor had 12 levels and was within subjects, the distance factor (distance of the referent) had two levels and was within subjects. This ANOVA failed to show a significant interaction between instances and distance, $F_1(11, 253) = 0.75$. The distance effect was significant, $F_1(1, 23) = 21.07, p < .01$, when the mean reading time of the first distant pronoun ($M = 2,874.4$ ms; $SD = 530.9$) was compared with that of the first closest pronoun ($M = 2,258.8$ ms; $SD = 549.6$). Therefore, the sensitivity to the gender agreement was obtained immediately in Experiment 4.

### TABLE 8

Results of Experiment 4: Mean reading times of the anaphoric sentence when the anaphora was a pronoun or a repeated noun referring either to the close or the distant antecedent

<table>
<thead>
<tr>
<th>Referent</th>
<th>Time $^b$</th>
<th>Speed $^c$</th>
<th>Time $^b$</th>
<th>Speed $^c$</th>
<th>Overall mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Close</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pronoun</td>
<td>1,924.2</td>
<td>347.4</td>
<td>1,950.9</td>
<td>400.1</td>
<td>1,937.5</td>
</tr>
<tr>
<td>Repeated noun</td>
<td>2,127.7</td>
<td>439.5</td>
<td>2,191.0</td>
<td>392.0</td>
<td>2,159.3</td>
</tr>
<tr>
<td>Distant</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pronoun</td>
<td>2,385.9</td>
<td>497.5</td>
<td>2,415.2</td>
<td>437.5</td>
<td>2,400.5</td>
</tr>
<tr>
<td>Repeated noun</td>
<td>2,194.2</td>
<td>374.5</td>
<td>2,252.3</td>
<td>362.7</td>
<td>2,223.2</td>
</tr>
</tbody>
</table>

$^a$Gender of close referent.

$^b$Reading time in ms.

$^c$Reading speed in ms/character.
Discussion

In Experiment 4, using the labels “close referent” and “distant referent” to report the results may be misleading. For pronouns, the only cue that indicated that the referent was not the close referent was the gender of the pronoun. Without taking it into account, readers would have directly attached the predicate of the pronoun to the most accessible referent, which could be a man or a woman in different versions of the same text. The main result was that normal readers were sensitive to pronominal gender, with longer reading times when the pronoun disagreed with the most accessible antecedent.

It could be argued that our texts had pronominal predicates that were more consistent with either a man or a woman, so that changing the gender of the most accessible referent could disrupt the direct attachment of the pronominal predicate in only one of the two versions. This interpretation would be consistent with the no processing view defended by Greene et al. (1992). However, it is ruled out by the design of our experimental materials.

We therefore conclude that readers in Experiment 4 were sensitive to a disagreement between the pronoun and the most accessible noun in working memory, even if they were never questioned about the pronominal sentence. The lack of questions concerning pronominal reference leads to a question about pronominal comprehension in our texts. Did readers resolve the pronouns that were in disagreement with the close antecedent? Of course, we have no information about their final interpretation of the pronominal sentence. However, we can consider several issues that arise regarding assignment of pronominal reference in these sentences.

To explore this possibility, consider a text that was not used in our experiments, but that would be consistent with this view. The text in (23) corresponds to a hypothetical sequence of two sentences (fourth and fifth sentences in our texts) with a pronoun referring to the close noun.

23. John came into the bathroom. He took out a razor to shave.

Here, if we replace “John” by “Mary” as in (24), the reading time of the pronominal sentence should be longer, even if “he” is not processed.

24. Mary came into the bathroom. He took out a razor to shave.

This effect could be taken as a disagreement effect but it would not imply that the pronoun is processed: The predicate of the pronoun is pragmatically less likely to be associated with a female noun. So a “pseudo-disagreement” effect could occur even with no processing of the pronoun. However, in Greene et al.’s (1992) theory, such an effect would also be obtained with “she” as the pronoun if “Mary” replaces “John”, as in (25).

25. Mary came into the bathroom. She took out a razor to shave.

Furthermore, if we take the “no processing” hypothesis in its strongest sense, there would be no reason to observe a processing difficulty for Sentence 26 because the predicate is pragmatically consistent with a male noun.

26. John came into the bathroom. She took out a razor to shave.

In summary, the no processing view could be consistent with a disruption of the integration of the pronominal sentence in a disagreement condition, if only two versions of our texts were used, those that corresponded to hypothetical texts (23) and (24). However, in Experiment 4, all the four versions were used. Under the “no pronominal processing” view, reading times for (24) should be longer than those for (23), but they should also be longer for (25) than for (26). The mean value for the “gender disagreement conditions” collapses the reading times for (24), which is pragmatically less plausible if pronominal gender is ignored, and (26), which is pragmatically more plausible if pronominal gender is ignored, and the mean value for the “gender agreement condition” collapses reading times for (23), which is pragmatically more plausible if pronominal gender is ignored. Therefore, a difference between the “gender agreement” and “gender disagreement” conditions is not attributable to greater pragmatic plausibility of the attachment of the predicates following the pronouns to the entities in the focus of the discourse.
The results of Experiment 3 indicate that, when a pronoun was encountered, the most distant referent was still present in working memory. Readers failed to recognize this distant referent in only 12.5% of the cases. It could be suggested that part of the long reading time for pronouns referring to a distant referent could reflect the time to search back for this referent in working memory. In our view, this search must be provoked by the sensitivity to the disagreement with the close antecedent: If this disagreement was not detected, there was a good accessible candidate for pronominal reference. It is interesting to note that the reading time for pronouns referring to distant referents was approximately 460 ms longer than that for pronouns referring to close antecedents. This difference is very similar to the gender disagreement effect observed in Experiment 2 (430 ms), and this numerical similarity could suggest that the same process was operating in the two conditions. If this is the case, it is possible that readers in Experiment 4 were sensitive to the disagreement but did not try to find a possible referent, or that their capacity to resolve pronouns by retrieving the distant referent depended on their working memory capacity (cf. Daneman & Carpenter, 1980). This aspect of pronominal processing requires more investigation. The important result for our hypothesis was that readers were sensitive to the gender agreement in Experiment 4.

Finally, our results showed a clear repeated noun penalty when the anaphor referred to the subject of the previous sentence. This was consistent with the high accessibility of the close referent. However, there was a pronoun penalty when the anaphor referred to a distant antecedent. This effect could partly reflect the difficulty of processing pronouns that disagreed with the most accessible entity. In particular, we would note that our results do not show that repeated nouns are anaphoric forms that are specialized for reference to a very distant noun. We found no evidence of an advantage for repeated nouns referring to a distant noun relative to a repeated noun referring to a close noun. This is consistent with a recent version of centring theory (Gordon & Hendrick, 1998). The essential function of proper names is to introduce new referents and not to refer to previously mentioned referents.

GENERAL DISCUSSION

In Experiments 1 and 2, pronouns were encountered after a clause containing two potential referents mentioned as subject and object of an implicit causality verb. In Experiment 1, participants were systematically questioned about the identity of the pronominal referent. We replicated the effect of gender cue observed by Garnham et al. (1992): The reading time of the pronominal clause was shorter when pronominal gender was sufficient to resolve the pronoun than when pronominal gender was in agreement with the two nouns mentioned in the main clause. This facilitation effect of the gender cue was attributed to a strategic disengagement process required to reduce the level of activation of the nonreferent, an essential operation when readers know that they will be questioned about the identity of the pronominal referent. The late occurrence of this facilitation effect in Experiment 1 is a further argument in favour of its strategic status. We suggest that this disengagement process allows subjects to stop trying to semantically integrate the words following the pronoun with one possible referent and to concentrate this integration process on just one entity. This is the reason why the reading time of the entire pronominal clause is shorter when the gender cue is relevant. In this view, the disengagement process is strategic, and it should not occur when subjects are not questioned about the identity of the pronoun. Experiment 2 confirmed this...
prediction: Without any question about the pronouns, the relevant gender cue did not significantly accelerate the reading time of the subordinate clause. This result was a replication of Garnham et al.'s results.

However, in Experiment 2, readers had longer reading times for pronominal clauses when the pronoun was in disagreement with both of the two mentioned names than when it was in agreement with both names. This result confirmed the intuition of Garnham et al. (1992): Even in a condition where readers do not strategically use the gender cue to select the referent, they remain sensitive to pronominal gender. This aspect of pronominal gender processing seems automatic because it occurs even in a condition where a strategic use of the gender cue cannot be observed. The fact that it arose on even the first occurrence of a mismatched pronoun provides additional evidence for its automaticity. We call it “automatic gender coindexation”. We return to the relation between this automatic gender coindexation and the strategic disengagement process later in the discussion.

Experiments 3 and 4 examined a property of automatic gender coindexation that could not be studied with the short sentences used in Experiment 2: the fact that gender coindexation concerns the most highly accessible entity in working memory at the time the pronoun is encountered. As mentioned in the previous section, some studies suggest that the two nouns surrounding an implicit causality verb are equally activated when the pronoun is encountered. To create an accessibility advantage for one of two potential referents, Experiments 3 and 4 used longer texts where one of the potential referents was mentioned 35 words before the pronoun, and the other potential referent was mentioned 6 words before the pronoun. This large difference was expected to be sufficient to place only the closer referent in the focus of attention when the pronoun was encountered. The probe recognition task used in Experiment 3 confirmed the accessibility difference between the two potential referents. It also showed that reading the pronominal sentence did not significantly change the activation of the two potential referents. These findings are consistent with the results reported by Greene et al. (1992, 1994) and McKoon et al. (1996) because they show that when a pronoun is encountered in conditions where strategic processes are not involved, probe recognition times of potential referents are not significantly modified by pronominal processing. Following the logic of these authors, it could be concluded that pronouns were not processed in such conditions—at least for their agreement properties. To check this prediction, Experiment 4 used the same texts, but in some cases the pronouns were in disagreement with the most highly accessible potential referent. Readers were sensitive to this disagreement, with longer reading times for sentences with pronouns that disagreed with the most accessible antecedent.

Our results thus confirm the intuition of Garnham (2001, p. 132): “The use of gender cue to speed comprehension is under strategic control. It does not follow, however, that people ignore gender information when it is not speeding their comprehension. It is likely that they would baulk if a main clause with two male names was followed by a subordinate clause with the pronoun ‘she’.” This automatic gender sensitivity has important implications for models of pronominal processing. It rules out a model that maintains that automatic pronominal processing is only a null process that directly attaches the predicate of the pronoun to the most highly activated referent.

Automatic pronominal processing may show language specific effects. Some languages (e.g., Spanish, Italian) allow for nonimperative finite verbs without any explicit subject. These are called “pro-drop” languages because they allow pronominal subjects to be
dropped in finite clause (Huang, 2000). For instance, in Italian, explicit pronouns exist, but they only occur when the antecedent is not easily accessible, or when the speaker wants to focus on one referent among several highly accessible referents. If the referent is easily accessible, then no explicit subject is required. This situation corresponds exactly to the “no processing thesis” suggested by Greene et al. (1992), because when the pronoun is not necessary to identify its referent, it does not occur. It may be that, in such languages, the comprehension system is prepared to deal with sentences in which no processing of a pronoun is required when a single entity is in focus. However, in non-pro-drop languages like English or French, a pronoun must be produced to refer to the most highly accessible entity. Because of this constraint, and because pronouns do contain agreement features, a highly automatic agreement process seems the most useful option for the speech comprehension system in non-pro-drop languages with overt gender marking on pronouns, like English or French. In these languages, pronouns are automatically processed, and the experiments reported in this paper show that they are not only processed as empty words: Their agreement features are matched with the features of their antecedent (what we called the “gender coindexation process”).

One important aspect of this process that remains obscure is the exact nature of the features that are matched. There are two options that cannot be easily distinguished in English. The first option consists of automatic coindexation using conceptual gender (i.e., sex). Because the first name “Peter” refers to a male referent, the gender coindexation process of pronoun “he” could consist of checking that the most highly accessible entity in the conceptual representation of the text is male. This was also the case for our French verbal material because we used human referents. However, as Garnham (2001) noted, in some languages, the agreement process is based on a purely grammatical gender feature. For instance, “the table” is feminine in French (“la table”) and the same feminine pronoun (“elle”) that is used to refer to female names, as in (27), is also used to refer to an inanimate feminine noun, as in (28).

27. La femme est arrivée, elle (*il) est dans l’entrée.
   The woman is arrived, she is in the hall.
28. La table est arrivée, elle (*il) est dans l’entrée.
   The table is arrived, it is in the hall.

This phenomenon suggests that the gender coindexation process involves purely morphosyntactic features, at least in languages where gender can be a purely grammatical feature (Cacciari, Carreiras, & Cionini, 1997; Garnham, Oakhill, Ehrlich, & Carreiras, 1995). This aspect of the coindexation process is important to investigate.

A related issue concerns the number feature of pronouns. Gernsbacher (1991) showed that in some cases, a plural pronoun could be used when the explicit antecedent is singular, as in (29).

29. I need a plate. Where do you keep them?

It is possible that number feature is not affected by an automatic coindexation process, maybe because it has a direct conceptual representation (see also Albrecht & Clifton, 1998, for data suggesting that automatic number coindexation does not exist). Anecdotally, we note that
“plate” is a feminine name in the French language (i.e., “une assiette”). The conceptual number agreement is also perfectly acceptable in the French translation of Sentence 29 given in Sentence 30. However, if the gender of the plural pronoun becomes masculine, the sentence becomes incorrect.

30. J’ai besoin d’une assiette. Où sont-elles (*ils) ?

Though Greene and his collaborators (1992) did not consider cross-linguistic issues, they could object to our theory on the grounds that our data were collected in experiments with French sentences. The above-mentioned facts about this language could be specific to a language where every noun has a gender feature. Gender coindexation could be more automatic in such language. We believe that this potential counterargument is not directly relevant. First, Rigalleau and Caplan (2000) reported data obtained in English that were consistent with the current data. Second, the formulations of Greene et al. (1992) were not presented as specific to English. However, both languages in which nouns are associated with grammatical gender and those in which they are not need to be studied.

Another objection could concern the effect of presenting some abnormal sentences in these studies. For instance, it could be said that the occurrence of (even rare) gender errors in our material was sufficient to induce a strategic gender process. In other words, the conditions of normal reading were not fulfilled in Experiments 2 and 4. Three arguments can be advanced against this objection. First, the fact that readers were sensitive to the gender errors when no judgement or comprehension questions about pronouns were presented indicates that they processed gender even when no reason was present to process it. Detecting an error only serves to disturb and slow down aspects of comprehension demanded by the experiment, which a strategy would be expected to enhance, not make more difficult to accomplish. Furthermore, if readers did neglect the pronoun and only attached the following predicate to the most accessible entity, the text could be easily understood. If gender is only strategically processed, how could readers be sensitive to errors when there is no utilitarian basis for a strategy that processes gender to develop? Second, the disagreement sensitivity in Experiments 2 and 4 was obtained immediately on the first instance where it occurred, and it remained on the following instances. Third, in Experiment 2, we showed that the occurrence of sensitivity to gender mismatches did not induce the usual effect associated with a strategic gender process (i.e., shorter reading times for pronominal clauses when they are preceded by two names of different gender than by two names of same gender). So the occurrence of some rare abnormal sentences was not sufficient to induce a well-documented strategic process involving pronominal gender.

We conclude with a short discussion of the relation between automatic and strategic gender processing. We begin this discussion with a summary of the relevant data.

Probe recognition studies using material with two highly accessible potential referents show an advantage for the referent, which agrees in gender when three conditions are met: (1) The probe is presented at the end of the pronominal clause; (2) reading presentation is slow; and (3) there are questions about the identity of the referent. On the other hand, they fail to show such an advantage when the reading presentation is fast, with longer texts, and when no question are asked about pronouns (Gernsbacher, 1989; Greene et al., 1992). With the same conditions of presentation, but with materials where there is only one referent, which is highly
accessible, the referent, advantage remains the same before and after the pronoun (Greene et al., 1994). Experiments 3 and 4 reported that this type of material does not induce a disappearance of the sensitivity to a gender disagreement between the pronoun and the most highly accessible referent.

Reading time studies have mainly used verbal materials where two potential referents were highly accessible at the time the pronoun is encountered. If readers are questioned about pronouns, two different gender referents lead to an acceleration of the reading time of the pronominal sentence (relative to a condition where both referents have same gender). When no questions are asked about pronouns, this acceleration is not observed (Garnham et al., 1992). We replicated both results in Experiments 1 and 2. But we also noticed that a gender disagreement sensitivity could be observed even when readers were not questioned about pronouns.

Naming latencies studies (where pronouns must be named) used the same material as reading times studies. When there were questions about pronouns, naming latency was longer if the pronoun agreed with only one referent than when it agreed with both referents. When these questions were removed, this difference was not significant. However, participants remained sensitive to a pronoun that was in disagreement with both referents (Rigalleau & Caplan, 2000).

Our model presented in the Introduction section is consistent with these results. The automatic coindexation process is assumed to occur in all conditions. It can explain why readers are systematically sensitive to a gender disagreement. On the other hand, the disengagement process would be strategically triggered when subjects have frequent questions about the pronoun. This process would elicit a longer naming latency of a pronoun when one of the potential referents disagrees with the pronoun. However, it would also induce a probe advantage in favour of the antecedent of the pronoun, and a shorter reading time of the predicate when this process allows the reader to disengage resources from integrating the predicate from the disagreeing referent, and to engage more resources in the integration of the predicate with the remaining referent.

We would like to emphasize the relation between this model and the results reported by Greene et al. (1992). In their “nonstrategic” conditions, Greene et al. (Exp. 1–4) did not find any recognition time advantage for a pronominal referent relative to a pronominal non-referent. This is consistent with the idea that activation remains engaged on the two potential referents in this condition. In their strategic conditions, Greene et al. (Exp. 5) found that the probe recognition time of the nonreferent was longer after the pronominal clause (1,067 ms) than before the pronoun (933 ms). This “inhibition” effect was not present for the pronominal referent (1,043 ms before the pronoun and 1,054 ms after the pronoun). This pattern is consistent with the occurrence of a disengagement process (disengagement of activation from the nonreferent). We suggest that the disengagement process requires two conditions to be met: First it considers the output of the coindexation process (agreement or disagreement) and therefore requires coindexation to have occurred; second it is triggered by the motivation of the comprehender (it occurs when the reader’s goal is to resolve the pronoun). When only one of the two conditions is present, the two consequences of the disengagement process are not detected: First the processing time of the pronominal is not significantly affected by the relevance of the pronominal gender cue, and, second, the probe recognition times for the potential referents are not affected by reading the pronoun.
We think that the results reported by Greene et al. (1992) showed that the probe recognition task is sensitive to the strategic disengagement of activation from a referent in disagreement with a pronoun. However, this task is not sensitive to the automatic coindexation process. But the nonsensitivity of a task to a process does not mean that this process does not exist (cf. Gordon, Hendrick, & Ledoux Foster, 2000). Our conclusions contrast with the provocative formulations of Greene et al. (1992, “no processing required for the pronouns”) and of McKoon et al. (1996, “Pronoun resolution without pronouns”). These formulations make strong predictions that were not directly tested by the authors. These researchers used only conditions involving correct pronouns, and they measured probe recognition times. However their formulations require testing conditions involving pronouns in disagreement to check whether the pronouns are really not processed.

Finally, we think that the existence of an automatic gender process is consistent with the hypothesis that the main function of a pronoun is not to create accessibility but to confirm accessibility. This confirmation function is ensured by checking the gender agreement between a noun in focus and the pronoun (cf. Sanford & Garrod, 1989, for a similar process named “immediate pronominal bonding”). This checking process is required for the pronoun even if the pronoun does not systematically induce a complete identification of a single referent. The exact relation between the checking process and pronominal resolution remains to be clarified. Our work was mainly aimed at show that this checking process is automatic and can involve gender when the pronoun has this information. We hope that our distinction between an automatic gender coindexation process and a strategic disengagement process is a step in the construction of a model that can account for all aspects of establishing the reference of pronouns.

REFERENCES


