Memory & Cognition
2005, 33 (6), 972-983

Immediate activation of stereotypical gender information

JANE OAKHILL, ALAN GARNHAM, and DAVID REYNOLDS
University of Sussex, Brighton, England

This article reports six experiments in which we explored whether gender stereotype information is typically invoked when certain role and profession terms are read and the extent to which the use of such information is under the reader's strategic control. All of the experiments used a design in which subjects had to decide whether two terms (one an occupation and one a kinship term) could refer to the same person (e.g., *surgeon–brother* or *surgeon–sister*). The presentation conditions and the instructions were varied from experiment to experiment, to try to encourage the subjects to respond strategically and to suppress their use of gender stereotypes when responding. The results support not only the hypothesis that information about the stereotypical gender associated with occupations and roles is typically incorporated into the reader's representation immediately, but also the hypothesis that such information is difficult or impossible to suppress. The implications of these findings for current theories of text processing and text representation are discussed.

The construction of a representation of the information in a text is influenced by background knowledge. Indeed, texts that do not have readily perceivable links with background knowledge are notoriously difficult to understand or appreciate—for example, the abstract description of washing clothes, used in an experiment by Bransford and Johnson (1972). However, the issue of how prior knowledge is used in the setting up of a text representation is contentious and difficult to investigate. It seems likely, however, that certain sorts of information derived from lexical items will be included in a representation if they are likely to make the subsequent text cohere. So, for instance, if the word *father* is encountered in a text, the information that the person is male is likely to be incorporated into the representation. More generally, if a reader encounters characters who have typically male or female names, they will build a representation in which those characters are male or female (in the absence of any information to the contrary, such as *Phil* turning out to be a short form of *Phyllis*; Carreiras, Garnham, & Oakhill, 1993; Chang, 1980; Garnham & Oakhill, 1985). Such filling in of *default values* probably results in cognitive economies, providing that the defaults assigned are not disconfirmed later in the text.

In this article, we address one aspect of background knowledge and how readily it is accessed—that of stereotypical gender information. That such information is used (even against the reader or listener's better judgment) is well illustrated by the following conundrum:

> A father and son are driving home one day, when they are involved in a serious accident. The father is killed outright, but the son is driven to hospital, where he is about to undergo an emergency operation. However, the surgeon refuses to operate, saying: "I can't operate on him: he's my son." The question is, how can this be?
>
> (Sanford, 1985, p. 311)

To investigate whether or not such passages cause problems, Reynolds, Garnham, and Oakhill (in press) conducted two experiments. In the first, the subjects read passages such as the above. In such passages, an incongruity arises in the final clause if readers have previously assigned a gender to a character introduced with a social role name. The subjects' confusion supported the contention that gender information had been incorporated into their mental representation at encoding of, say, *surgeon*. In a second experiment, online measures confirmed that readers were slower to read the final clause of such passages when there was a gender stereotype conflict. Even when questioned about their interpretations of such passages, many of the subjects did not immediately solve the inconsistency by inferring the appropriate gender for the role term. These findings provide strong evidence that gender activation has its locus at the time a role name is encoded—that is, that in the example above, the subjects encoded the sex of the surgeon as male. If they do not make that inference immediately on encountering the word *surgeon*, they must make it later on in the text. However, later in the text, the information that the boy's father is dead should make the inference that the surgeon is the boy's father problematic. One would think that the information that the father is dead would overrule the (presumably, rather weaker) information that a surgeon is likely to be male. However, the problems exhibited by the subjects

This research was supported by ESRC Grant R000 236481 (Mental models in text comprehension: Constraints on inference) to the first two authors. We thank Marcelle Crinean for help with collecting the data for Experiment 6. Correspondence concerning this article should be addressed to J. Oakhill, Laboratory of Experimental Psychology, University of Sussex, Brighton BN1 9QH, England (e-mail: j.oakhill@sussx.ac.uk).

Copyright 2005 Psychonomic Society, Inc.

in Reynolds et al.'s experiments indicate that stereotypical gender information is difficult to suppress.

Assumptions about gender stereotypes are common everyday occurrences. For instance, we often refer to a doctor or an architect of unspecified sex as *he* and might have to readjust our representation when we find that a particular architect being referred to is, in fact, a woman. We also often find the need to qualify terms when they go against cultural stereotypes. Thus, although it is perhaps becoming rarer for people to talk about a "lady doctor," they still often allude to *male nurses* and *male models*, even when it is irrelevant or unnecessary to do so (as in "my son is thinking of becoming a male model").

When a text does not provide specific information about the gender associated with an occupation or role term (e.g., *nurse*), prior knowledge—in this case, a gender stereotype—might be used to infer the more likely gender. In a text or discourse, pronouns are used frequently to refer to previously introduced referents. The mental models theory assumes that such pronominal reference is possible because the linguistic antecedent of a pronoun introduces a token representing an individual into the text representation (see, e.g., Garnham, 1987; Glenberg, Meyer, & Lindem, 1987; Johnson-Laird, 1983). To interpret a pronoun, the reader needs to recover a referent that matches the pronoun in number and gender. It is easy to imagine that if a reader encounters a character called *Ann,* it will be easier to resolve a subsequent anaphoric reference (*she* or *her*) if the reader has incorporated the information that Ann is female into the representation of the text. And indeed, it has been demonstrated empirically that pronouns that can be resolved on the basis of gender information are understood more rapidly than those that cannot (Carreiras et al., 1993; Garnham & Oakhill, 1985; Garnham, Oakhill, Ehrlich, & Carreiras, 1995; Gernsbacher, 1989; Kerr & Underwood, 1984).

Most terms for occupations and roles in English are not marked for gender, although many are subject to gender stereotypes. If a person is introduced into a discourse as *the surgeon* or *the secretary*, that phrase gives no specific information about gender, because the content refers to the job and a person of either sex can do either job. Thus, such terms could be followed by a pronoun of either gender. However, everyday knowledge might lead the reader to suppose that a surgeon is male and a secretary female (because, statistically, this is likely to be true). The surgeon riddle indicates that gender stereotype information is used immediately to encode gender, in much the same way as it would be used if the surgeon had been introduced as *Mr. Goodcut*. Thus, we suggest that a knowledge-based inference is made from the fact that a particular person is a surgeon to the fact that that person is (most probably) male and that such inferences are made when the gender-stereotyped term is encountered. Indeed, Banaji and Hardin (1996) have suggested that the activation of gender stereotypes is immediate and automatic, at least in some circumstances. However, we will reassess the evidence on which they made this claim below.

One issue that has received attention in the literature is whether the stereotyped information is activated immediately when a stereotyped term is encountered or only later—for example, when a pronoun needs to be resolved. Consideration of the Sanford (1985) riddle suggests that stereotyped gender information is invoked immediately. The studies of online processing, however, are not so clear-cut. A number of studies have indicated that gender stereotypes have an influence on how readily a reader is able to interpret a subsequent pronoun (consistent either with the idea of immediate inference or with the idea that stereotyped information comes into play only when it is needed to establish coherence). For instance, Carreiras, Garnham, Oakhill, and Cain (1996, Experiment 1) showed that it takes subjects longer to read sentences (in English) containing pronouns that conflict with the gender stereotype of previously introduced occupations or roles. There is also evidence that readers spend longer fixating such pronouns (Duffy & Keir, 2004, Experiment 1; Kerr & Underwood, 1984). In addition, Osterhout, Bersick, and McLaughlin (1997) measured event-related brain potentials when subjects read sentences containing reflexive pronouns that referred to a definitionally (e.g., *uncle* or *cowgirl*) or stereotypically (e.g., *bartender* or *babysitter*) male or female antecedent noun. Pronouns that conflicted with the gender definition or stereotype elicited a larger amplitude positive wave (P600) than did nonconflicting pronouns. Interestingly, the positive wave elicited by gender stereotype violation persisted even when the subjects judged those sentences to be acceptable. The data from the above studies do not enable us to determine precisely when the information about stereotypical gender came into effect; it may be that stereotyped gender information becomes available when a stereotyped term is first encountered or only when a reader attempts to retrieve an antecedent for a subsequent pronoun. However, previous data from Spanish (Carreiras et al., 1996, Experiments 2–4) suggest that stereotype information is activated early on. Spanish subjects spend longer encoding occupation and role terms when the gender stereotype conflicts with the article and with the morphological marking of the term (e.g., *la enfermera*/*el enfermero*) or even when it conflicts only with the article (*el*/*la futbolista*). Similarly, Duffy and Keir, in a second experiment, found that when prior discourse specified the gender of the role-named character, the later mismatch effect was eliminated. These results are consistent with the idea that the gender stereotype for a role name is accessed when that role name is first encountered.

It is difficult to reach definitive conclusions about the time course and automaticity of the activation of stereotypical information from reading time studies. However, other paradigms, in which subjects respond to single words, have addressed this issue. In particular, Banaji and Hardin (1996) suggested that the activation of gender-stereotyped information is automatic for some types of words and can occur despite subjects' intention to ignore gender-stereotyped prime words. They state that "gender

information imparted by words can automatically influence judgement" (p. 136). Those authors used three types of stimuli to assess the automaticity of gender stereotyping. Their experiments will be discussed in some detail here, since they are relevant to the issues under investigation, and we will argue that Banaji and Hardin's conclusions are not, in fact, supported by the data they present. In two experiments, the subjects first saw a prime word and then had to make a judgment about a second word. The three types of primes were those associated with a particular gender by definition (e.g., *mother* or *waitress*), those associated by normative base rates (i.e., stereotypes; e.g., *nurse* or *mechanic*), and those that were neutral with respect to gender (e.g., *citizen* or *cousin*). In their first experiment, the task was to judge whether a pronoun (*he* or *she*) was male or female. In their second experiment, the task was simply to judge a second word as either a pronoun or not a pronoun (e.g., *he* or *me* vs. *do* or *all*). Interestingly, neither judgment task *required* any linking between the first word and the word to be judged. In both experiments, the subjects responded more rapidly to male pronouns after male than after female primes and to female pronouns after female than after male primes. However, the effect was much more marked for primes linked to a gender by definition and was not nearly so striking in the case of primes linked to gender by normative base rates (and indeed, was entirely absent in this condition in the second experiment).[1] Thus, their data show clear effects (which they claim are automatic, because their task did not require any explicit linking between the prime and the target words) only for primes that are related to gender by definition (e.g., *mother*, *man*, or *waitress*). These effects can hardly be attributed to stereotypes; the term *mother* cannot be termed s*tereotypically female* when it is female by definition.

Since Banaji and Hardin (1996) failed to provide strong evidence for the activation of stereotype information (even though they claimed to do so), in the present article we make a further attempt to investigate gender stereotypes in the context of an attempt to address the extent to which a reader is able to bring information about stereotypical gender under strategic control. In the experiments that follow, our primary interest was in the extent to which readers invoke gender information for terms that do not include gender information in their definition (i.e., genuinely stereotyped terms). Thus, we were interested in the extent to which normative information is activated when role names that are stereotypically associated with one or the other gender are encountered. We hypothesized that even though Banaji and Hardin failed to find such effects with single words, a task in which the terms had to be deliberately linked in order to perform the task might show effects where theirs failed to do so (or showed only very small effects). We think that this is a reasonable hypothesis, because Banaji and Hardin obtained some effects in their first experiment, in which the decision task was gender related and, therefore, might have encouraged some linking of prime and target. This task contrasts with that used in their second experiment, in which the decision was not gender related and there were no effects on the normative base-rate materials. Although the task we used requires the subjects to deliberately link the two terms in a manner that Banaji and Hardin's tasks did not, it does not explicitly require them to invoke stereotyped information. In the experiments that follow, moreover, we investigated the automaticity of the processes involved in performing this task in a different way, by manipulating the instructions to the subjects and, thereby, their perceptions of the task. If there is an automatic component to the task, it should be very difficult or impossible to bring the pattern of responding completely under strategic control; that is, it should be very difficult to fully suppress the bias toward a stereotyped response. If, however, there is a strategic component to the task, the issue of interest is whether the effect is wholly strategic, in which case it should be possible to eliminate it.

In the task we devised, the subjects had to make a decision about whether or not two terms could possibly refer to the same person. This task cannot be done without linking the two terms, but we also expected that the task would encourage the subjects to think of the terms as applying to an individual (although not necessarily any particular person) and, therefore, to be more likely to make inferences about the probable gender of that person. Although our task, like Banaji and Hardin's (1996), does not require the reading of connected text, we argue that the results are relevant and applicable to such situations. Indeed, if effects of stereotypical gender are found using the present paradigm and those effects are difficult to suppress even when it is advantageous to do so, it is highly likely that such information is invoked automatically and that such effects would also occur when connected text is read. Following Banaji and Hardin, we distinguish between definitional and stereotypical gender. In order to perform optimally in our task, subjects need to take into account definitional gender (e.g., that dukes are necessarily male) but to disregard frequency information about stereotypical gender (e.g., that most surgeons are male). Our task was deliberately designed so that, in order to perform correctly, the subjects would need to suppress stereotypical gender but to use definitional gender.

In summary, in the following experiments, we were interested in whether gender-stereotyped information is invoked from single words and the extent to which subjects were able to strategically suppress such information. In previous experiments on gender stereotyping, the stereotyped terms have usually been inserted in short passages. Banaji and Hardin's (1996) study is an exception, in that the subjects were required to make a judgment about a single word, following a single-word prime. However, as was discussed above, their study provided little evidence for online gender stereotyping. When reading a passage, as opposed to single words, a reader may prefer to insert definite values (including information about the sex of the protagonists) into the mental model of the text. In the present experiments, we were interested to see whether

such information is invoked and how readily it can be brought under strategic control when only single words are presented for decision. As was mentioned above, we adopted a procedure that differed from that used by Banaji and Hardin, in that our subjects were required to state explicitly whether two words could refer to the same person, rather than making a decision on the second (supposedly unrelated) word. We assumed that this procedure would be more likely than that used by Banaji and Hardin to encourage the subjects to invoke gender information. Indeed, in some of our studies, the instructions we used specifically alerted the subjects to the possible interfering effects of gender stereotypes. In all of the experiments in this article, a design was used in which the subjects had to decide whether or not two terms (one an occupation and one a kinship term) could refer to the same person (e.g., *nurse–aunt* or *nurse–uncle*). They differed in the extent to which the nature of the materials and the possibility of strategic control over responding were made apparent to the subjects. If subjects were unable to completely suppress gender stereotype information, even under conditions in which it was made clear to them that it was advantageous to do so, we would have evidence that there is an automatic component to the evocation of such information.

EXPERIMENT 1

Method

Subjects. The subjects were 16 volunteers (8 of them male and 8 female) from the student population of Sussex University. They were paid £3 for taking part. The session lasted around 30 min.

Materials. A list of 12 stereotypical occupations/roles for each gender was constructed from the norms collected by Carreiras et al. (1996). The items that were rated as strongly biased toward male or female roles were chosen. Examples of strongly male-biased occupations are *footballer*, *bricklayer*, and *butcher*, and strongly female-biased occupations are *nurse*, *beautician*, and *secretary*. Twelve neutral ones were also selected, such as *artist*, *musician*, and *novelist*. The occupations in the three conditions were all single words and were matched for length as far as possible, so that the average length (in characters) was the same across the three congruence conditions. Six kinship terms (three pairs) were chosen. These were *father*, *mother*, *brother*, *sister*, *uncle*, and *aunt*.

Kinship and occupation/role terms were then paired to produce congruent, incongruent, and neutral pairs. In the congruent list, male and female stereotyped roles were paired with a kinship term that was congruent with the gender stereotype of the occupation/role—for example, *father–plumber* or *sister–typist*.In the incongruent condition, the pairings were reversed, so that the gender of the kinship term was incongruent with the stereotypical gender bias for that role—for example, *father–typist* or *sister–plumber*.

In the neutral condition, neutral occupations/roles were paired with the same set of kinship terms—for example, *father–novelist* or *sister–student*.

Each congruence condition comprised 72 pairs of kinship–occupation/role terms, obtained by pairing the 6 kinship terms (3 male and 3 female) with the corresponding 12 occupation/role terms for that condition. A full list of the sterotyped occupations is shown in the Appendix.

We also included 332 filler items, in which there was no gender ambiguity and to which the subjects could respond *yes* or *no* with certainty. These fillers were devised by matching the same six kinship terms with gender-marked occupations/roles (e.g., *stewardess* or *milkman*), with other terms that carry gender (e.g., *bachelor* or *spinster*) and official titles (e.g., *Lord* or *Lady*), and with each other.

In total, the subjects were presented with 548 pairings. Of these, 296 items, including all of the experimental items, required *yes* responses. The remaining 252 items required a *no* response. The pairs that required a *no* response were ones in which the two terms could not refer to the same person—for example,

Brother	Landlady	*Female gender marked*
Sister	Salesman	*Male gender marked*
Father	Aunt	*Two kinship terms*
Uncle	Countess	*Female title*
Aunt	Duke	*Male title*

Apparatus. The experiment was controlled online by a personal computer with an Advantech PCLabCard 812 to provide millisecond-accurate timing. The pairs of words for which the subject had to make a decision were presented one at a time in the center of the screen of a VDU. Between the VDU and the subject were two response buttons, one labeled YES–SAME PERSON, and the other NO–DIFFERENT PERSON.

Design. The kinship and occupation/role terms appeared on the screen together, with the kinship term always on the left. They stayed on the screen until the subject responded. There was a 1-sec interval between trials.

The main manipulation in the experiment was congruency or otherwise between the kinship and the occupation/role terms. Thus occupation/role terms that have a stereotypically female or male bias or are neutral were paired with either male or female kinship terms. This manipulation was within subjects but between materials, since we treated each pair as a separate item. Each subject made judgments about all the items, and presentation order was pseudo-randomized such that the repetition of a specific occupation/role term with different kinship terms was spread as evenly as possible throughout each experimental list.

Procedure. The subjects were tested individually in a small quiet laboratory. They were instructed to read each pair of words at their normal reading speed and to decide whether the two terms could apply to the same individual. They had to press the *yes* button if the two terms could apply to one individual and the *no* button if they could not. Once they had responded to a word pair, it disappeared from the display, and after 1 sec the next pair appeared.

The instructions, also presented via the VDU, provided examples of the types of kinship–occupation pairs that would appear. The instructions stressed that the items should be read at normal reading speed but that the subjects should avoid excessive deliberation over their responses. They were given two examples as part of the instructions, one of which required a *yes* and one a *no* response.

Prior to the experiment proper, there were 18 practice trials to familiarize the subjects with the experimental procedure. These included a representative sample of kinship–occupation pairs that the subjects would see in the experiment proper. The ratio of practice items for which we expected *yes* responses to those for which we expected *no* responses was approximately the same as that for the experimental items.

Results

The analyses we report were ones in which some items were excluded. The subjects were very resistant to accepting either *cheerleader* or *nanny* when these were matched with male kinship terms. These items resulted in only 48% *yes* responses, which was considerably lower than the mean response for any other items in the incongruent condition. In retrospect, *nanny* might have confused the participants,

since it is sometimes used as an affectionate term for a grandmother in British English. In addition, after the study had been completed, it was brought to our attention that in some American English dialects, there are special terms for male cheerleaders (although all our subjects were native speakers of British English and probably would not have known this fact). The corresponding length-matched male stereotype items *footballer* and *judge* and the neutral items *journalist* and *clerk* were also excluded.

All analyses, in all five experiments reported here, had the same factors. These were gender of stereotype (male, female, or neutral), gender of kinship term (male or female), and the particular kinship relation (aunt/uncle, sister/brother, or mother/father).[2]

Fillers. Performance on the filler items was uniformly high: about 95% correct.

Judgments. In this and all further analyses, judgments and response times were subjected to two ANOVAs: one in which subjects was treated as the random variable, and one in which items was treated as the random variable. The percentages of *yes* responses in the three congruence conditions are shown in Table 1.

The *yes* responses were analyzed by ANOVAs, in which subjects and materials were treated as the random variables. Only the main effect of congruity was significant [$F_1(2,30) = 8.1$, $p < .01$; $F_2(2,162) = 73.8$, $p < .001$]. As can be seen from Table 1, the incongruent pairs were rejected more frequently than either the congruent or the neutral pairs.

Response times. The mean response times to the pairs in the three conditions are also shown in Table 1.

From the complete data file containing 548 responses, the critical 216 (3×72) items were identified. All data points more than 2.5 standard deviations (*SD*s) above or below a subject's mean were replaced by the appropriate 2.5 *SD* cutoff score. In total, 95 data points out of a total of 3,456 (2.75%) were replaced for this reason. In addition, the data corresponding to the erroneous *no* responses were replaced (after the data trimming detailed above) by the appropriate subject $\times$ congruency condition or item $\times$ congruency condition means. A total of 192 data points were replaced for this reason.

Only the main effect of congruity was significant [$F_1(2,30) = 32.1$, $p < .001$; $F_2(2,162) = 74.8$, $p < .001$]. In these and all the following post hoc analyses, we compared the types of item, using *t* tests. Because the incongruent items were expected to be harder than the congruent ones, we adopted a one-tailed test for these comparisons. For the other comparisons, we adopted a two-tailed level of significance, because we did not make any particular predictions about the differences between neutral items and the other two types. Pairwise *t* tests revealed statistically significant differences between the incongruent and the congruent pairs [$t_1(15) = 6.4$, $p < .001$; $t_2(118) = 10.0$, $p < .001$]. There was also a significant difference between the incongruent and the neutral pairs [$t_1(15) = 5.8$, $p < .001$; $t_2(118) = 9.4$, $p < .001$], but the difference between the congruent and the neutral pairs was not significant (both *t*s < 1).

Discussion

The results of this first experiment provide clear evidence that people are biased by gender stereotypes when making judgments about the likely gender of people in certain occupations and roles. So, for example, our subjects were more likely to reject the *engineer–mother* pair as a possible same-person pair than they were to reject *engineer–father*. In addition, although in the majority of cases the subjects were able to overcome their gender biases and respond that a mother and an engineer could possibly be the same person, they took longer to make such decisions, whether compared with pairs that were congruent with a stereotype bias (*engineer–father*) or with gender-neutral terms. These results are somewhat surprising, since the experiment contained a large number of items and the subjects could probably readily discern the nature of the pairs they were being asked to judge (even during the practice trials). Thus, even though they were probably aware of the highly biased nature of some of the occupations, they could not entirely overcome their stereotype biases to accept such items or, at least, could not do so quickly.

In this first experiment, the kinship and the occupation terms were presented simultaneously. The stereotyped term was on the right, so normal reading patterns would suggest that it should be read after the kinship term. The subjects could, therefore, consider the (marked) kinship term first and then consider whether or not the other term could possibly refer to a person of the same sex. In the second experiment, the two terms to be judged were in separate displays, with the occupation/role term in the first display, so that the subjects were forced to read the occupation/role term first, before they saw and responded to the corresponding kinship term. Thus, if, say, they decided that an occupation could apply to either a male or a female, they would be able to respond immediately to the kinship term by pressing the *yes* button; the precise kinship term would be irrelevant. Similarly, if the first term could be applied only to males, the subject could prepare to respond *yes* only if one of the male kinship terms appeared; otherwise, he or she would respond *no*.

With this presentation format, the subjects could strategically focus their processing effort on the first term presented (the occupation/role) and then respond rapidly

Table 1
Percentages of *Yes* Judgments (%) and Mean Response Times (RTs, in Milliseconds) in Experiments 1–4

	Condition	%	RT
Experiment 1	Congruent	99.4	1,070
	Incongruent	88.4	1,264
	Neutral	99.1	1,082
Experiment 2	Congruent	97.7	685
	Incongruent	79.2	813
	Neutral	97.6	692
Experiment 3	Congruent	99.6	627
	Incongruent	93.0	670
	Neutral	99.8	591
Experiment 4	Congruent	98.9	600
	Incongruent	92.3	635
	Neutral	98.9	568

to the second (kinship) term. In the following experiments, we explored the effects of encouraging the subjects to adopt such a strategy, but in Experiment 2, we simply changed the presentation format so that the strategy of preparing a response to the kinship term (contingent on the nature of the occupation/role term) was an option for our subjects.

EXPERIMENT 2

Method

Subjects. The subjects were 16 volunteers from the student population of Sussex University (8 of them male and 8 female), who had not taken part in the previous experiment. They were paid £3 for taking part. The session lasted around 30 min.

Materials. The materials were identical to those in Experiment 1, but the presentation procedure was changed, as will be detailed below.

Design and Procedure. The design in Experiment 2 was identical to that in Experiment 1. However, in contrast to Experiment 1, where the kinship and the occupation terms appeared together on the screen, in Experiment 2 the occupation/role term of each pair was displayed centrally on the computer screen in isolation for 500 msec before the screen was cleared and the kinship term appeared (the interstimulus interval [ISI] was 0). The kinship term remained on the screen until the subject responded. When it appeared, the subject was required to make the *same*/*different* person judgment, as before. There was a 1-sec interval between trials. The instructions to the subjects were adjusted accordingly but were, in essence, the same as those in Experiment 1.

Results

The results were, broadly, similar to those from the first experiment.

Judgments. The percentages of *yes* judgments are shown in Table 1. There was a main effect of congruency [$F_1(2,30) = 23.0$, $p < .001$; $F_2(2,162) = 112.5$, $p < .001$]. As in Experiment 1, performance was near ceiling in the congruent and neutral conditions, but performance in the incongruent condition was consistently lower.

Response times. The mean response times to the pairs in the three conditions are also shown in Table 1. All data points more than 2.5 *SD*s above or below a subject's mean were replaced, as in Experiment 1. Eighty data points out of a total of 3,456 (2.3%) were replaced. In addition, the data corresponding to the erroneous *no* responses were replaced (after the data trimming detailed above) by the appropriate subject × congruency condition or item × congruency condition means. A total of 348 data points were replaced for this reason.

The mean response times to the kinship terms are also shown in Table 1. There was an overall decrease in reaction times, as compared with Experiment 1, but this is readily explicable in terms of the difference in the presentation of the stimuli: In Experiment 1, the times were for reading and responding to two words, rather than for one. As for the judgments, the main effect of congruity was highly significant [$F_1(2,30) = 9.9, p < .001$; $F_2(2,162) = 21.6, p < .001$].

Pairwise *t* tests revealed statistically significant differences between the incongruent and the congruent pairs [$t_1(15) = 3.0$, $p < .001$; $t_2(118) = 5.4$, $p < .001$]. There was also a significant difference between the incongruent and the neutral pairs [$t_1(15) = 4.2$, $p < .001$; $t_2(118) = 5.0$, $p < .001$], but there was no difference between the congruent and the neutral conditions (both ts < 1).

No other effects or interactions were significant.

Discussion

The results of Experiment 2 replicated those of Experiment 1. Even when the subjects were forced to read and consider the occupation term first, they still made a higher proportion of *no* responses to kinship terms that followed stereotypically mismatching occupation terms than to those following matching or neutral occupation terms. Once again, when the subjects did respond *yes* to the mismatching items, they took longer to do so. These data provide support for our hypothesis that information about the gender of a stereotyped role or occupation is activated and incorporated into the representation of a term.

Somewhat counterintuitively, in neither experiment was the congruent condition facilitated with respect to the neutral condition. This finding seems to be at odds with the literature on anaphor resolution, where an anaphor is easier to understand if there is a recent antecedent that matches in number and gender. One possibility is that gender information is not encoded when the occupation or role term is first encoded but that it is inferred only when the kinship term, which carries gender, has to be considered during the decision process. Perhaps the representation does not include information about gender until that point. However, as Carreiras et al. (1996) noted, such a strategy would be inefficient, since it would mean that other and, possibly, less reliable or conflicting information would have to be invoked to help assign gender. We will come back to this point in the Discussion section for Experiment 3, where this pattern of results was even more striking.

In the third experiment, we repeated the procedure used in Experiment 2, but the time for which the first (occupation or role) term was shown was increased, so that the subjects would have time to think about the possible genders implied by the occupation term and could, if they wished, prepare their response to the kinship term before it appeared and could inhibit any stereotype bias. Thus, the third experiment was an attempt to explore the extent to which the activation and suppression of gender information associated with stereotypically male or female occupations could be controlled by the subjects. We expected that an increase in the amount of time that the subjects had to consider the occupation terms would reduce or eliminate the stereotyping effects and, hence, that the difference between the congruent and the incongruent conditions, observed in Experiments 1 and 2, would be eliminated.

EXPERIMENT 3

Method

Subjects. The subjects were 16 volunteers from the student population of Sussex University (8 of them male and 8 female), who had not taken part in this series of experiments previously. They were paid £4 for taking part. The session lasted around 40 min.

Materials. The materials were the same as those in Experiment 2.

Design and Procedure. The design and procedure in Experiment 3 were the same as those in Experiment 2, except that the occupation/role term remained on the screen for 1,800 msec before it was replaced by the kinship term. The ISI between the occupation and the kinship terms was 0, and the kinship term remained on the screen until the subject responded. When it appeared, the subject was required to make the *same*/*different* person judgment, as before. There was a 1-sec interval between trials. All other features of the design and procedure remained the same as those in Experiment 2, except that following the experiment proper, the subjects were interviewed about their strategies. They were asked the following questions. (1) Can you describe as clearly as you can how you decided whether the word pair could be the same person or different? (2) Were you aware of any word pairs being more difficult than others? The subjects were sometimes probed for clarification or further information, as appropriate.

Results

Judgments. The percentages of *yes* judgments are shown in Table 1. Even though the proportion of *yes* responses was considerably elevated, by comparison with Experiments 1 and 2, the main effect of congruency was still highly significant [$F_1(2,30) = 8.4$, $p < .01$; $F_2(2,162) = 37.2$, $p < .001$].

Response times. The mean response times to the pairs in the three conditions are also shown in Table 1.

All data points more than 2.5 *SD*s above or below a subject's mean were replaced, as in Experiment 1. Seventy-eight data points were replaced (2.26%). In addition, the data corresponding to the erroneous *no* responses were replaced (after the data trimming detailed above) by the appropriate subject × congruency condition or item × congruency condition means. A total of 149 data points were replaced for this reason.

The mean correct response times are also shown in Table 1. Once again, there was a main effect of congruency [$F_1(2,30) = 7.2$, $p < .01$; $F_2(2,162) = 17.9$, $p < .001$].

In this case, pairwise *t* tests revealed statistically significant differences in all three comparisons. As in the previous experiments, there was a significant difference between the incongruent and the congruent pairs [$t_1(15) = 1.8$, $p < .05$; $t_2(118) = 3.2$, $p < .01$] and between the incongruent and the neutral pairs [$t_1(15) = 3.1$, $p < .01$; $t_2(118) = 5.2$, $p < .001$]. However, in contrast to the findings from the previous experiments, the pairs containing neutral terms were responded to more rapidly than the gender-congruent pairs [$t_1(15) = 3.0$, $p < .01$; $t_2(118) = 3.0$, $p < .01$].

Postexperiment interviews. The responses in the postexperiment interviews revealed that some subjects did reflect on their responses and clearly understood that their response to any pair whose first member was a stereotyped occupation/role term should be *yes*:

> If the first word of the pair was gender specific I matched the second word with it. If it wasn't gender specific then I didn't bother; I automatically said it was the same person. (Subject D.W.)

In order to test the hypothesis that the subjects' awareness of the nature of the materials may have led them to employ particular strategies, which could have reduced the incongruity effect, an additional analysis was performed. The subjects were divided into those who reported that they had used a strategy of automatically responding *yes* when they read a stereotype role ($n = 7$) and those who did not ($n = 9$). An ANOVA on the response times, with an additional between-subjects factor of strategy, showed a main effect of group; the subjects who reported a strategy were faster overall [$F(1,14) = 5.39$, $p < .05$]. There was also a significant interaction between strategy and congruency condition [$F_1(2,28) = 7.87$, $p < .01$]. The subjects who did not report that they had used a strategy demonstrated a pattern of means that was very similar to that seen in Experiments 1 and 2—that is, longer decision times for incongruent (849 msec) than for congruent (738 msec) and neutral (707 msec) pairs. This pattern of means contrasted with that from the subjects who reported using a strategy. These subjects showed a main effect of congruency [congruent = 542 msec, incongruent = 530 msec, neutral = 501 msec; $F_1(2,16) = 6.82$, $p < .001$]. It was shown by *t* tests that the effect was located in the difference between the neutral conditions and the others. In this group, there was a significant difference between the neutral and the congruent conditions [$t_1(15) = 3.52$, $p < .01$] and between the neutral and the incongruent conditions [$t_1(15) = 2.19$, $p < .05$]. The comparison between the congruent and the incongruent conditions did not reach significance [$t_1(15) = 1.40$].

In the strategy group, although decision accuracy was very high in all the congruency conditions (congruent = 99.6%, incongruent = 95.6%, neutral = 99.8%), the effect of congruency was highly reliable [$F_1(2,16) = 14.2$, $p < .001$]. The fact that the strategy group was more accurate than the no-strategy group in the incongruent condition indicated that the improved performance in the incongruent condition in this group cannot be explained in terms of a speed–accuracy trade-off.

Discussion

In Experiment 3, there was, once again, a congruity effect in the judgments, even though the greatly improved decision accuracy in the incongruent condition suggests that the delay did allow the subjects to inhibit the gender information or to anticipate that a kinship term of either gender could follow a stereotyped occupation term. The subjects were still less likely to accept occupation/kinship pairs when the stereotypical gender of the occupation conflicted with the gender of the kinship term. This result is surprising, since the subjects were given a considerable amount of time (1,800 msec) to consider the possible genders of the first item in each pair before the kinship term was presented to them. The response time findings in Experiment 3 were slightly different, in that the neutral items were responded to more rapidly than the congruent items. This was an unexpected result, since there is no reason to expect that subjects would respond more rapidly to kinship terms following neutral, rather than congruent, items. However, it is possible that even if the subjects prepared their responses (*yes* in both cases), they still were

more cautious and took longer in the case of stereotypical occupation and role terms, perhaps because they would not, or could not, suppress a gender-checking process when the kinship term was presented.

The results of the division of subjects into those who did and those who did not use a deliberate strategy indicate that, when subjects are aware of the influence that stereotypical gender information might have on their decision making, they can, to some extent, counteract the effects of this information, although not completely overcome it (judgment accuracy for the incongruent pairs was still significantly lower than that in the other congruency conditions for this group). In the fourth experiment, we tried a new modification in an effort to encourage the subjects to adopt the sorts of strategies naturally adopted by some of the subjects in Experiment 3. We adapted the instructions to make the subjects explicitly aware that some professions are not clearly marked for gender and that many jobs can now be held either by men or by women. We expected that such instructions would attenuate the effect of congruency condition on performance.

EXPERIMENT 4

Method

Subjects. The subjects were 16 volunteers from the student population of Sussex University (8 of them male and 8 female), who had not taken part in this series of experiments previously. They were paid £4 for taking part. The session lasted around 40 min.

Materials. The materials were the same as those in Experiments 2 and 3.

Design and Procedure. All features of the design and procedure remained the same as those in Experiment 3, with the exception that the instructions were enhanced to highlight the correct *yes* response appropriate to the incongruent items, in which the gender bias of the occupation/role term conflicted with that of the gender of the kinship term. The additional instructions were as follows:

> It is also worth remembering that some professions are not clearly marked for gender. That is, they don't have ". . . man," ". . . woman" or ". . . ess" to indicate gender in their title. Examples of these unmarked professions might include "Chauffeur" or "Receptionist," jobs which nowadays men and women both occupy.
>
> So the following are examples where a "Yes/Same Person" response is appropriate.
>
> Chauffeur
>
> Mother
>
> and
>
> Receptionist
>
> Father
>
> To facilitate your responses you should consider carefully whether the first word in each pair is an occupation that can only be filled by a man, only by a woman or by either gender.

Results

Judgments. The percentages of correct judgments are shown in Table 1. There was a main effect of congruency [$F_1(2,30) = 16.0, p < .001$; $F_2(2,162) = 21.7, p < .001$]. No other effects or interactions were significant.

Response times. The mean response times to the pairs in the three conditions are also shown in Table 1. All data points more than 2.5 *SD*s above or below a subject's mean were replaced, as in Experiment 1. Eighty-two data points (2.37%) were replaced. In addition, the data corresponding to the erroneous *no* responses were replaced (after the data trimming detailed above) by the appropriate subject $\times$ congruency or item $\times$ congruency means. A total of 175 data points were replaced for this reason.

The mean correct response times are also shown in Table 1. An ANOVA, with the same factors as those above, again showed a main effect of congruency [$F_1(2,30) = 7.7, p < .01$; $F_2(2,162) = 7.6, p < .001$] but no other effects or interactions.

As in the previous experiments, there was a significant difference between the incongruent and the congruent pairs [$t_1(15) = 2.2, p < .05$; $t_2(118) = 1.8, p < .05$] and between the incongruent and the neutral pairs [$t_1(15) = 3.4, p < .01$; $t_2(118) = 3.7, p < .01$]. As in Experiment 3, the pairs containing neutral terms were responded to more rapidly than the gender-congruent pairs [$t_1(15) = 2.0, p = .06$; $t_2(118) = 2.3, p < .05$].

Discussion

Despite the more detailed instructions, which were designed to "warn" the subjects against gender-stereotyped responses, the pattern of results was remarkably similar to that found in the previous experiments. Once again, the subjects made more errors in accepting the pairs that did not match gender stereotypes, and when they accepted such pairs, they took longer to do so. The response times again indicated a tendency for the subjects to take longer to accept the pairs with congruent genders than those with neutral terms (as was found in Experiment 3). Alerting the subjects to the relevance of the gender information did not seem to prevent its having an influence on the decision process. This finding is consistent with that in Banaji and Greenwald (1995, Experiment 4), in which they explored the effects of implicit gender stereotyping in judgments of fame. Those authors found that even when they introduced a manipulation to make the gender variable more salient to the subjects, this manipulation did not alter the degree of stereotyping they observed.

Experiments 3 and 4 support the conclusion that subjects who are given sufficient time to plan their response and are aware, or are encouraged to become aware, that some of the terms are stereotypically biased do improve the speed and accuracy with which they respond. However, this awareness by no means eliminates the incongruity effect—in either the response times or the judgments. It seems that the gender stereotype associated with certain roles and occupations is very strong and that subjects are not able to fully inhibit it and prevent its having an effect on their decisions in this task.

In the final experiments in this series, we required the subjects to make speeded responses, in an attempt to constrain their decision time. In Experiment 5, the occupation/role term was presented first, as in Experiments 2, 3, and 4, and the subjects were told that they had to make a very rapid response to the kinship term that followed. We expected that the subjects might be encouraged to be stra-

tegic and plan their response to the kinship term while viewing the occupation/role term, since they would know that the time they were allowed to respond to the second term was strictly limited. In this way, we gave the subjects maximal encouragement to plan their decision to the kinship term while they were viewing the occupation/role term, so that they would be able to make a rapid response. As was explained earlier in the article (introduction to Experiment 2), if an occupation can apply to either a male or a female, subjects should realize that they can respond *yes* to the kinship term, and should be able to respond rapidly. Similarly, if the first term could be applied only to, say, males, the subject could prepare to respond *yes* only if one of the male kinship terms appeared; otherwise, he or she should respond *no*.

We predicted that under such conditions, the activation of gender stereotype information would be expressed in more errors in the incongruent condition than had been seen previously, because there would be less time to suppress the activation of such information.

EXPERIMENT 5

Method

Subjects. The subjects were 16 volunteers from the student population of Sussex University (8 of them male and 8 female), who had not taken part in this series of experiments previously. They were paid £3 for taking part. The session lasted around 20 min.

Materials. The materials were the same as those in Experiment 2.

Design and Procedure. The design and procedure were identical to those in Experiment 2, except that presentation time of the kinship term was restricted. The occupation/role term remained on the screen for 500 msec before it was replaced by the kinship term. The ISI between the occupation and the kinship terms was 0, and the kinship term remained on the screen for only 600 msec or until the subject responded, whichever was first. When the kinship term appeared, the subjects were required to make the *same*/*different* person judgment, as before. Times in excess of 600 msec were not recorded, and a "Too Slow" message was displayed for 500 msec before the next item was shown. There was a 1-sec interval between trials. All other features of the design and procedure were the same as those in the previous experiments.

Results

Judgments. The percentages of correct responses made within 600 msec are shown in Table 2. The table also shows the percentages of incorrect responses made and the percentages of cases in which no response was made within 600 msec. As was expected, the overall percentage of correct responses was lower in this experiment, because the subjects were forced to make a rapid decision. However, the incongruity effect was still marked [$F_1(2,30) = 19.69$, $p < .001$; $F_2(2,162) = 54.8$, $p < .001$]. There were no other effects or interactions.

Because the subjects were under time pressure in this experiment, we also looked at the number of responses (whether correct or not) that were provided within the allotted time. There was a higher proportion of responses in the congruent (93%) and neutral (93%) conditions than in the incongruent condition (88%), and this effect was significant [$F_1(2,30) = 16.57, p < .001$; $F_2(2,162) = 38.02$, $p < .001$].

Response times. The mean response times to the pairs in the three conditions are also shown in Table 2, both for correct responses made within 600 msec and for all responses made within that time.

The mean correct response times are also shown in Table 2. Since a cutoff of 600 msec was applied, the 2.5 *SD* cutoff was not applied in this experiment. There was a main effect of congruency, as in the previous experiments [$F_1(2,30) = 8.21, p < .001$; $F_2(2,162) = 9.65$, $p < .001$]. However, pairwise *t* tests revealed that, in this experiment, the incongruent and the congruent conditions did not differ. There was a significant difference between the incongruent and the neutral conditions [$t_1(15) = 3.75$, $p < .01$; $t_2(118) = 2.89, p < .01$, two-tailed], and between the congruent and the neutral conditions [$t_1(15) = 3.22$, $p < .01$; $t_2(118) = 4.99, p < .001$].

It should be noted, however, that an analysis of times for all responses (whether correct or not) showed the more familiar difference between the congruent and the incongruent conditions. Thus, the failure to observe the congruity effect with the correct-only times seems to arise both

Table 2
Results of Experiments 5 and 6

Dependent Variable	Condition		
	Congruent	Incongruent	Neutral
Experiment 5			
Percentage of correct judgments within 600 msec	82.0	67.7	85.2
Percentage of incorrect judgments within 600 msec	10.9	21.0	7.4
Percentage of timeouts	7.1	11.3	7.4
Mean response times for correct answers within 600 msec	411	410	390
Mean response times for all answers within 600 msec	414	423	394
Experiment 6			
Percentage of correct judgments within 600 msec	77.6	56.5	75.9
Percentage of incorrect judgments within 600 msec	15.0	26.9	16.8
Percentage of timeouts	7.4	16.6	7.3
Mean response times for correct answers within 600 msec	460	472	469
Mean response times for all answers within 600 msec	460	472	466

because subjects made more incorrect responses in the incongruent condition when under time pressure and because they were more often timed out before the response had been made.

Discussion

The speeded response requirement seems to have forced the incongruity effect onto the judgment data. There was little effect in the response time data. The lack of a difference between the congruent and the incongruent conditions in this experiment might be attributable to the fast, but highly inaccurate, level of responses in the incongruent condition (see above).

Experiment 5 did not address the issue of what happens if subjects are required to respond very rapidly to the *stereotyped* term (in the experiment above, time pressure was on the kinship term, not on the stereotyped occupation term). If the terms were presented in the opposite order, accuracy might go down in general (because the subjects would respond to the stereotyped term on the basis of their initial impression). However, one possibility is that any effect of congruency condition might be reduced or eliminated. This pattern of results might be found if stereotype information is not activated immediately. In this case, subjects might be responding before stereotype information can have any effect. These possibilities were explored in the next experiment, in which the order of the terms was reversed, so that the kinship term always occurred first.

EXPERIMENT 6

Method

Subjects. The subjects were 16 volunteers from the student population of Sussex University (8 of them male and 8 female), who had not taken part in this series of experiments previously. They were paid £3 for taking part. The session lasted around 20 min.

Materials. The materials were identical to those in Experiment 5, except that the order of presentation of the occupation/role terms and the stereotyped terms was reversed. The subjects always saw the stereotyped term second.

Design and Procedure. The design and procedure were identical to those in Experiment 5, except that in this case, it was the presentation time of the occupation/role term that was restricted. The kinship term remained on the screen for 500 msec before it was replaced by the occupation/role term. The ISI between the kinship and the occupation terms was 0, and the occupation/role term remained on the screen for only 600 msec or until the subject responded, whichever was first. Times in excess of 600 msec were not recorded, and a "Too Slow" message was displayed for 500 msec before the next item was shown. There was a 1-sec interval between trials. All other features of the design and procedure were the same as those in the previous experiments.

Results

Judgments. The percentages of correct responses are shown in Table 2. The table also shows the percentages of incorrect responses made and the percentages of cases in which no response was made within 600 msec. As was expected, the overall percentage of correct responses was lower in this experiment than in Experiments 1–4 (and was also somewhat lower than that in Experiment 5), because the subjects were forced to make a rapid decision. However, the incongruity effect was still marked [$F_1(2,30) = 51.45$, $p < .001$; $F_2(2,162) = 48.79$, $p < .001$]. There were no other effects or interactions.

Because the subjects were under time pressure in this study, we also analyzed the number of responses (whether correct or not) that were provided within the allotted time. There was a higher proportion of responses in the congruent and neutral conditions than in the incongruent condition, and this effect was highly significant [$F_1(2,30) = 12.52$, $p < .001$; $F_2(2,162) = 19.63$, $p < .001$].

Response times. The mean response times to the pairs in the three congruency conditions are also shown in Table 2, both for correct responses made within 600 msec and for all the responses made within that time.

The mean correct response times are also shown in Table 2. As in Experiment 5, the 2.5 *SD* cutoff was not applied. There was a main effect of congruency, as in the previous experiments, although it was significant only by subjects [$F_1(2,30) = 5.82$, $p < .01$; $F_2(2,162) = 2.33$, n.s.]. However, pairwise *t* tests revealed that, in this experiment, the incongruent and the neutral conditions did not differ. There was a significant difference between the congruent and the incongruent conditions [$t_1(15) = 3.92$, $p < .001$; $t_2(118) = 1.68$, $p < .05$, two-tailed], and between the congruent and the neutral conditions [$t_1(15) = 2.96$, $p < .01$; $t_2(118) = 2.18$, $p < .05$].

Discussion

Once again, the speeded response requirement seems to have forced the incongruity effect onto the judgment data, probably because the response time data are curtailed: Most of the subjects took between 400 and 600 msec to respond, whatever the congruency condition, so there is less scope for clear differences between the conditions. There was only a relatively small effect in the response time data, and the effect was rather odd, in that the incongruent and the neutral conditions did not differ. As in Experiment 5, the lack of a difference between the neutral and the incongruent conditions (for correct responses) in this experiment might be attributable to the fast, but highly inaccurate, level of responses in the incongruent condition. The pattern of data was rather different from that found in Experiment 5. In Experiment 5, the times for the congruent and the incongruent conditions were similar (and responses in the neutral condition were faster), whereas in Experiment 6, responses in the incongruent condition were slower than those in the congruent condition. However, this pattern can readily be explained when one considers the difference between the tasks in these two experiments. In Experiment 5, in which the occupation term came first, the subjects could prepare a response (e.g., respond *yes* whatever term follows *doctor*, because a doctor can be a person of either gender), whereas in Experiment 6, they could not represent, say, *aunt* as possibly being of either gender and had to consider whether the occupation term could apply to a female, which might have increased the difficulty in the incongruent conditions (in-

deed, the task requirements in Experiment 6 seemed to increase the difficulty overall, relative to Experiment 5). Thus, although a strategy could be applied in both experiments, the strategy required was rather different in the two cases, and this difference was reflected in the pattern of response times.

However, positioning the occupation/role term second did have the expected effect of increasing errors, as compared with the previous experiment—in particular, in the incongruent condition, where they dropped to 55% correct. So it seems that if subjects are forced to make a rapid response to stereotyped items that do not conform to gender stereotypes, they are less able to ignore the stereotyped response and respond correctly.

GENERAL DISCUSSION

The six experiments reported here clearly demonstrated an effect of gender stereotyping in a task that required judgments about occupation and role terms. Both decision time and accuracy suffered when the stereotypical gender conflicted with the gender of a kinship term, as compared with conditions in which it was consistent or in which there was no gender bias. This effect appeared to be one of interference, rather than facilitation, since in no experiment did we find that decisions for the congruent pairs were facilitated relative to those on the pairs with gender-neutral occupation terms. Although the nature of the effect varied from experiment to experiment (it was sometimes more apparent in the response times, and sometimes in the error pattern), it was, nevertheless, strongly apparent in every experiment.

Contrary to the claims that Banaji and Hardin (1996) made for their own data, we argue that those authors did not provide good evidence for the automaticity of stereotyped responses in the case of the types of materials we have investigated. Banaji and Hardin's results provided only very weak evidence that stereotyped responses to role and occupation terms are automatic: The majority of the effects they found were limited to the materials that were gender-related by definition. Thus, Banaji and Hardin's data do not provide good evidence for the stereotyping effect itself, let alone evidence for the automaticity of this effect.

Our results provide much stronger evidence that there is an automated component to such responses, particularly because it is so difficult for subjects to bring these processes wholly under strategic control. This difference in results probably arises because of differences in the task demands between our experiments and those in Banaji and Hardin (1996). In our experiments, in which the subjects were required to decide whether or not a subsequent kinship term could refer the same person as a previous relation term, we found a much stronger, and persistent, effect of gender stereotyping. In Banaji and Hardin's task, the prime words (some of which were gender stereotyped) were, essentially, irrelevant to the decision for the target words that followed.

Our experiments strongly support the view that information about the stereotypical gender associated with occupations and roles is typically incorporated into a representation of an individual immediately (and probably automatically) as soon as an occupation or role name is read. We had expected that the task we chose would encourage the subjects to think of the terms as applying to an individual (although not necessarily any particular person) and, therefore, to be more likely to make inferences about the probable gender of that person. This is, perhaps, one of the reasons why the pattern of results we obtained was much stronger than that obtained by Banaji and Hardin (1996).

The conclusions from the present data are consonant with the findings from our earlier work in Spanish (Carreiras et al., 1996), in which we showed that occupation terms that are marked for gender (either morphologically or by the use of a definite article) that are incongruent with the gender stereotype (*el futbolista*) take longer to encode than those that are congruent (*la futbolista*). Those results indicated that Spanish readers expected to read about (for example) a male footballer and took longer to encode the representation when a female footballer was encountered.

We found that the effect of stereotypical gender was strong and persistent: In Experiments 3 and 4, the introduction of a delay between the occupation and the kinship terms and the instruction to the participants to pay deliberate attention to the possible stereotyping and to try not to be influenced by it attenuated, but did not entirely eliminate, the effect of incongruity. In those experiments, response times improved considerably, but a surprising finding was that the pairs with neutral occupations were now responded to *more* rapidly than those with occupations where the stereotypical gender and the gender of the kinship term were congruent. As was mentioned earlier, a possible explanation for this finding is that, even in the case of the stereotyped congruent items, where a *yes* response should have been facilitated by the congruency, the subjects showed a degree of caution or, perhaps, conducted an unnecessary gender check, because the occupation/role term was stereotyped. More generally, the results from the experiments in which we manipulated the instructions to the subjects show that information about gender stereotypes cannot be entirely suppressed. The effects can be diminished by instructions but were never completely eliminated in our experiments, even when the subjects were explicitly told that they should try not to be influenced by gender stereotypes or were encouraged to respond very rapidly.

We were surprised by the strength and persistence of the gender bias exhibited by our subjects. Even though we had previously observed such effects when we asked subjects to rate the likely gender of a number of occupations, it is surprising that the subjects in the present study persisted in evoking such stereotyped information even when it was counterproductive to their performance on the task to do so. It would appear that such information

is naturally invoked when a strongly stereotyped male or female occupation is encountered and that such information is difficult to suppress.

We touched on the reasons for the sorts of biases we have observed in these experiments in the introduction. It may be that they have an important role in guiding our processing and in reducing cognitive effort. We constantly make assumptions in our everyday encounters with people and things. We assume, for instance, that chairs have four legs and are sufficiently strong to support our weight, unless we receive evidence to the contrary, because that is what usually happens. Similarly, despite equality of opportunity, it is still the case statistically that the majority of bus drivers are men and that the majority of nurses are women. So it may be more cost effective in the processing of terms to make such assumptions, even if we have to undo them on the (perhaps relatively rare) occasions when they are wrong.

Of course, we cannot be sure that the results would generalize to discourse contexts. However, there is already evidence that gender stereotyping is commonly found in discourse contexts, and our view is that such results are more, rather than less, likely to occur in such contexts than in the type of task that we used here.

REFERENCES

Banaji, M. R., & Greenwald, A. G. (1995). Implicit gender stereotyping in judgments of fame. *Journal of Personality & Social Psychology,* **68**, 181-198.

Banaji, M. R., & Hardin, C. D. (1996). Automatic stereotyping. *Psychological Science,* **7**, 136-141.

Bransford, J. D., & Johnson, M. K. (1972). Contextual prerequisites for understanding: Some investigations of comprehension and recall. *Journal of Verbal Learning & Verbal Behavior,* **11**, 717-726.

Carreiras, M., Garnham, A., & Oakhill, J. V. (1993). The use of superficial and meaning-based representation in interpreting pronouns: Evidence from Spanish. *European Journal of Cognitive Psychology,* **5**, 93-116.

Carreiras, M., Garnham, A., Oakhill, J. V., & Cain, K. (1996). The use of stereotypical gender information in constructing a mental model: Evidence from English and Spanish. *Quarterly Journal of Experimental Psychology,* **49A**, 639-663.

Chang, F. R. (1980). Active memory processes in visual sentence comprehension: Clause effects and pronominal reference. *Memory & Cognition*, **8,** 58-64.

Duffy, S. A., & Keir, J. A. (2004). Violating stereotypes: Eye movements and comprehension processes when text conflicts with world knowledge. *Memory & Cognition*, **32**, 551-559.

Garnham, A. (1987). *Mental models as representations of discourse and text*. Chichester, U.K.: Ellis Horwood.

Garnham, A., & Oakhill, J. V. (1985). On-line resolution of anaphoric pronouns: Effects of inference making and verb semantics. *British Journal of Psychology,* **76,** 385-393.

Garnham, A., Oakhill, J. V., Ehrlich, M. F., & Carreiras, M. (1995). Representations and processes in the interpretation of pronouns: New evidence from Spanish and French. *Journal of Memory & Language,* **34,** 41-62.

Gernsbacher, M. A. (1989). Mechanisms that improve referential access. *Cognition*, **32,** 99-156.

Glenberg, A. M., Meyer, M., & Lindem, K. (1987). Mental models contribute to foregrounding during text comprehension. *Journal of Memory & Language,* **26,** 69-83.

Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness*. Cambridge: Cambridge University Press.

Kerr, J. S., & Underwood, G. (1984). Fixation time on anaphoric pronouns decreases with congruity of reference. In A. G. Gale & F. Johnson (Eds.), *Theoretical and applied aspects of eye movement research* (pp. 110-136). Amsterdam: Elsevier.

Osterhout, L., Bersick, M., & McLaughlin, J. (1997). Brain potentials reflect violations of gender stereotypes. *Memory & Cognition,* **25**, 273-285.

Reynolds, D., Garnham, A., & Oakhill, J. V. (in press). Evidence of immediate activation of gender information from a social role name. *Quarterly Journal of Experimental Psychology: A*.

Sanford, A. J. (1985). *Cognition and cognitive psychology.* London: Weidenfeld & Nicolson.

NOTES

1. In their first experiment, there was a highly significant three-way interaction between prime type (definition or normative base rate), prime gender (female or male), and target gender (female or male). The authors do not report whether there was a significant two-way interaction within the normative base rate materials, but the size of the mismatch effect was more than twice as large (36 msec) in the definitional than in the normative materials (15 msec). In addition, as the authors report, in the normative base rate conditions, the time taken to respond to male pronouns did not differ significantly depending on whether they followed a male or a female stereotyped term, so the small effect they observed in these materials held only for the female stereotypes (although it was still only 15 msec). The authors do not report the corresponding three-way interaction in the second experiment. However, the interaction between prime gender and target gender was much weaker than that in the first experiment and was entirely due to the primes related to gender by definition. It did not occur in those normatively related to gender; the mean size of the mismatch effect for these items was 1 msec.

2. We also performed analyses for this and all the subsequent experiments in which the sex of the subjects was included. There were no main effects of this factor and no interactions with experimental condition for either the percentages of correct judgments or the judgment times (all $ps > .1$).

APPENDIX
Stereotype Occupation/Role Terms Selected for the Incongruent, Congruent, and Neutral Congruency Conditions

Male Stereotype	Female Stereotype	Neutral to Stereotype
judge	nanny	clerk
carpenter	secretary	novelist
firefighter	telephonist	newsreader
butcher	florist	student
pilot	model	artist
politician	prostitute	astrologer
engineer	nurse	musician
paratrooper	housekeeper	interpreter
plumber	cleaner	swimmer
farmer	typist	singer
electrician	cheerleader	psychologist
footballer	beautician	journalist

(Manuscript received December 12, 2002;
revision accepted for publication September 26, 2004.)