

## Are inferences from stereotyped role names to characters' gender made elaboratively?

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

Two experiments provided evidence that gender stereotype inferences from role names—for example, that a surgeon is (probably) male—are made in a forward, elaborative, direction. We used sentences in which a person's gender was never made explicit, but was implied in two different ways. The two ways were by the use of a role name, and by mentioning an item of clothing (e.g., a bikini) or a biological characteristic (e.g., giving birth) that is typically associated with females or males. The two pieces of information (role name and clothing/biological characteristic) were presented in different orders in the two experiments. In both cases a mismatch between the associated genders slowed reading, showing that gender information has been activated. It is argued that if an inference about gender is made on the basis of the second piece of information, hence slowing comprehension, it is unlikely that the inference about gender based on the first piece of information was not made immediately.

Following the recognition by Bransford (e.g., Bransford, Barclay, & Franks, 1972; Bransford & Franks, 1971) and others of the importance of inference making in comprehension, it was also realized that inferences must be constrained (e.g., Corbett & Doshier, 1978; Singer, 1980; Thorndyke, 1976). A strong view, discussed by Garnham (1982, 1989), is that only those inferences necessary for a coherent interpretation of a text are routinely made, and that they are typically made in a backward direction. A related idea is found in the minimalist hypothesis of McKoon and Ratcliff (1992), which claims that only those inferences needed for local cohesion and those that are based on information readily available in memory are automatically made. The minimalist claim that knowledge-based inferences might be made suggests that inferences are probably not restricted solely to those needed for coherence. However, the specific claim in the minimalist hypothesis is problematic (e.g., Graesser, Singer, & Trabasso, 1994), largely because of the difficulty of establishing a criterion for what information is readily available that is independent of which inferences are made.

Garnham (1989) made a different suggestion about a set of inferences that are both unnecessary for coherence and possibly made routinely. His suggestion was that inferences that depend on the presence of a single lexical item in a text might be made. Garnham first made this suggestion in connection with the results of Garrod and

Sanford (1981). Garrod and Sanford studied a set of inferences that looked very like the bridging inferences studied by Haviland and Clark (1974) and others, and which were shown to be made in a backward direction. So, (3) was read more slowly after (2) than after (1):

- (1) We got some beer out of the trunk.
- (2) We checked the picnic supplies.
- (3) The beer was warm.

This result suggested that the inference that there was beer among the (Californian) picnic supplies was not made elaboratively when the picnic supplies were first mentioned, but only in order to understand the definite reference to beer in (3). Garrod and Sanford studied the apparently similar contrast of (4) or (5) followed by (6), but found that no extra time was required to interpret (6) following (5) in examples of this kind:

- (4) Mary put the clothes on the baby.
- (5) Mary dressed the baby.
- (6) The clothes were made of pink wool.

Garnham's (1989) suggestion was based on the fact that the definition of *dress* makes reference to clothes. The representation of (4), therefore, already contains a representation of a set of clothes to which the definite noun phrase "The clothes" in (6) can refer. By contrast, there is no single lexical item in (2) with a definition that includes beer. It is widely accepted that lexical entries have complex definitions. So, it is plausible that simply by using the meaning of *dress*, among other things, to construct the meaning of (5), clothes are automatically encoded into the representation. In (2), on the other hand, an additional inference would be required to encode a representation of beer into the representation of that sentence, and this work may well not be carried out.

---

Our research was supported by Economic and Social Research Council (U.K.) Grant RC00236481, "Mental Models in Text Comprehension: Constraints on Inference." We thank Tracey Crowther for help in testing subjects in Experiment 1. Correspondence should be addressed to A. Garnham, Laboratory of Experimental Psychology, School of Biological Sciences, University of Sussex, Brighton, BN1 9QG, U.K. (e-mail: alang@biols.susx.ac.uk).

— Accepted by previous editorial team

Even if this account of Garrod and Sanford's (1981) results is correct, it does not follow that all lexically based inferences are made elaboratively. The generality of Garnham's hypothesis can be determined only on a case-by-case basis. One type of lexically based inference for which there is some evidence that it is made elaboratively is that based on stereotyped role names, such as *judge* and *nurse*. The use of such a role name to describe a person could trigger an inference as to the (probable) sex of that person. Such an inference might be based purely on knowledge of the proportions of males and females in such occupations, or it might be based on other factors. From the point of view of this paper, it does not matter how such inferences arise.

A small number of published studies have shown that when a person is initially described using a stereotyped role name, a following pronoun that fails to match the stereotype is read more slowly than one that does match (Carreiras, Garnham, Oakhill, & Cain, 1996; Kerr & Underwood, 1984). This finding might be explained either in terms of elaborative inferences to the probable gender of the person when the role name is initially read, or in terms of the use of the same information when the pronoun, which is morphologically marked for gender, is linked to the role name. Note that this second type of use of the gender information would not be a backward inference in the usual sense. When (3) is linked to (2) by a backward inference, that inference (that the beer is part of the picnic supplies) is necessary to link the information in the two sentences. Linking, for example, *he* and *judge* requires only the information that judges can be male, not that they are usually male, though in a broader sense that information could be seen as relevant to making the link.

Carreiras et al. (1996, Experiments 2–4) attempted to provide evidence for the elaborative inference account. They used Spanish passages in which the gender of the person in the role was specified by a morphologically marked definite article (e.g., *el futbolista*, "the male soccer player," vs. *la futbolista*, "the female soccer player"). They found that the mismatch with the stereotype slowed reading of the part of the sentence containing the noun phrase *el/la futbolista*. However, unlike in English, no further difficulty was experienced in the gender mismatch case when a following coreferential definite pronoun was encountered (*la futbolista . . . ella . . .*, "the female soccer player . . . she . . .").

This finding provides suggestive evidence that the inference as to the probable gender of the role filler is made when the role name is read. However, as we have already noted, the evidence is not conclusive. More generally, it can be difficult to provide conclusive evidence of forward, elaborative, inferences (see, e.g., Keenan, Potts, Golding, & Jennings, 1990, and McKoon & Ratcliff, 1990). Keenan et al. noted that self-paced reading studies are more convincing if they include conditions in which the hypothetical inference is disconfirmed. Our studies of stereotyping include such conditions: those in

which there is a mismatch between the stereotype and other information, such as that provided by the gender of the pronoun. However, even an effect of mismatching does not prove that elaborative inferences are being made. More generally, the problems of interpretation noted by Keenan et al. and McKoon and Ratcliff have informed our procedure of looking for converging evidence from different paradigms to support the notion of elaborative inferences based on stereotypes.

Is there any reason to think that the inference about the gender of the person introduced by a role name is not inferred until it is linked to other information, other than the fact that it simply might be? We believe that there is at least one plausible argument, though it is not one that has been made in the previous literature. Consider an English example, such as (7) or (8) followed by (9), as used by Carreiras et al. (1996, Experiment 1):

(7) The electrician examined the light fitting.

(8) He needed a special attachment to fix it.

(9) She needed a special attachment to fix it.

The definite pronoun *He* in (8) or *She* in (9) provides (virtually) unambiguous information about the gender of its referent, which can be assumed to be incorporated into the mental representation of the passage. However, the role name *electrician* does not provide such unambiguous information. It is, therefore, plausible that the issue of gender is considered only (and with slight additional difficulty in the gender mismatch case) when the pronoun is read. To understand that someone is an electrician, it is necessary only to understand what work they do, not whether they are more likely to be male or female, so the inference need not be made when the role name is encountered. The Spanish case is similar. The definite article provides unambiguous information about gender, so the early use of stereotypical gender information in Spanish could be explained by the need to link the (unambiguous) information in the definite article to the information in the role name, and not to the mere occurrence of the role name.

Because of the difficulty of providing a single piece of unequivocal evidence for an inference based solely on the occurrence of a role name, we have adopted the tactic of trying to provide a body of evidence that makes this hypothesis increasingly plausible. The studies reported in this paper are one line of evidence pointing in this direction. They derive from the suggestion given above that stereotype gender information may be brought into play only when other, more definitive, gender information is also available (e.g., from a definite pronoun or a definite article). If the use of a stereotyped role name does not trigger an inference about the probable gender of a person, presumably because the information about gender is neither central to the meaning of a word such as *surgeon* nor particularly certain, then the use of another term with the same properties should not trigger an inference about gender either. So, if a reader has to

identify, say, a surgeon with someone described as wearing a bikini, and if describing someone as wearing a bikini does not, of itself, trigger an inference that the person is probably female, there is no reason why linking the *wearing a bikini* description to the *surgeon* description should either trigger a gender inference or cause any problems. On the other hand, if the *wearing a bikini* description, of itself, triggers a gender inference, and causes problems with integration with *surgeon*, then there is every reason to think that the *surgeon* description itself triggers an (elaborative) gender inference.

In the experiments reported in this paper, therefore, the role name clashes with information that is itself ambiguous. Thus, we consider cases in which gender is never made explicit, but is implied in two different ways. For example, in a sentence such as (10), referring to someone as a typist suggests that the person is probably female, whereas saying that the person put on a silk tie suggests that the person is probably male. Our goal is to show that the same effects occur as with pronouns and definite articles. However, if a role description, or a similarly ambiguous description, does not cause a gender inference of itself, and it is only the unambiguous gender information (in the pronoun or article) that causes the gender information in the stereotype to be activated, such mismatch effects should not be found in these studies.

- (10) The typist, who felt too casually dressed, went to the wardrobe and chose a favorite silk tie.

If a probabilistic inference about gender is made on the basis of the second piece of information (about the silk tie), it is likely that the (opposite) inference about gender based on the first piece of information (that the person is a typist) is also made. It is correspondingly unlikely that the inference based on *typist* is made only when the second piece of information is encountered, and not when the stereotyped role name is first read. More specifically, neither piece of information forces the inference that the person is female or male. So, if people are reluctant to commit themselves in such cases, they should do so neither when they read that the person is a typist nor when they read that the person put on a silk tie. The potential (probabilistic) clash between these two pieces of information should not affect comprehension. Indeed, a mismatch effect can be detected only if both inferences are made.

The same two pieces of information related to gender (a person is a typist; that person wears a silk tie) can also be presented in the other order, as in (11):

- (11) Choosing a favorite silk tie from the drawer, and tying it with a half Windsor knot, the typist wondered whether it was too formal.

Our case will be strengthened if we can demonstrate the same effects regardless of presentation order. The version in which the role name comes first more closely parallels our original study (*surgeon . . . she*; Carreiras et al., 1996, Experiment 1). The reversed version is of interest because, if stereotyped role names do not lead to

elaborative gender inferences, there is no reason to suppose that the typist in (11) will be represented as being (probably) female. There is thus no reason to suppose that identifying this person, by a backward link, with a person putting on a silk tie should cause problems. Note that neither gender inference is needed to create a coherent link between the information in the various parts of a sentence such as (11).

In Experiment 1, we used sentences such as (10), in which the role name information comes first, and in Experiment 2, we used sentences such as (11), in which it comes second. To provide the second indication of the probable gender of the person that the sentence is about, we used two different types of information. First, as in the example given above, we used items of clothing (e.g., silk tie) and makeup. Even though in some cases an item of clothing might be made for, say, females, it does not follow that someone who puts on such an item of clothing is female. This is merely a probable inference. The second type of information was about biological characteristics (e.g., having a beard or being pregnant) or kinship relations (e.g., being someone's uncle) that do provide stronger constraints on the gender of the person so described. Thus, only a male can be someone's uncle, and only a female can be pregnant. So, if inferential effects are dependent upon clear-cut information about a person's gender, the mismatch effect previously demonstrated with pronouns and definite articles should be replicated with the biological characteristic materials, but not with the clothing materials. Note, however, that definite pronouns and definite articles are morphologically and semantically unambiguous, whereas a characteristic such as having a beard is, in English at least, only semantically unambiguous. It is not morphologically marked for gender.

## EXPERIMENT 1

### Method

**Participants.** The participants were 24 volunteers from the student and staff populations of Sussex University. They were paid £3 for their participation in the experiment, which lasted for about 25 min.

**Materials.** Forty sentences were constructed as the experimental items for this experiment. Each sentence comprised three parts, which were presented separately. The first part introduced a character using a stereotyped role name taken from the norms collected by Carreiras et al. (1996). The other information in the first part of the sentence did not constrain the sex of the character. An example of the beginning of one of the sentences is (12a):

- (12a) The fortune teller, who had recently finished lunch,

The second part provided some further information about the character, and again did not constrain the character's sex. For example, the sentence above continued as in (12b):

- (12b) went to the washroom

The final part of the sentence did impose constraints on the sex of the character. There were two types of endings, one based on items of clothing and makeup that were either typically associated with females or males (e.g., bikini, lipstick, heavy size 12 boot, silk tie), and one based on biological characteristics or kinship relations

(e.g., gave birth, shouted “there’s my husband,” decided to grow a mustache, whispered “I’m your uncle”). The sentence above ended as in (12c):

(12c) and applied a fresh coat of lipstick.

An example of a sentence based on a biological characteristic is (13):

(13a) The soldier drove to the playgroup after work,

(13b) and picked up one of the children,

(13c) who said “Look what I did today daddy!”

Each sentence occurred in two versions. In one version the stereotype of the role name matched the gender associated with the item of clothing or the biological characteristic, as in the examples given so far. In the other version, there was a mismatch. The mismatched versions were created by swapping role names between sentences, so *fortune teller* in the first example became *plumber* and *soldier* in the second example became *housekeeper*. Example materials are shown in Appendix A.

In addition to the experimental items, 50 filler sentences were produced to provide a clear context in which the participants could be asked whether the third part of each sentence made sense as a continuation of the first two parts. Twenty fillers had neutral role names (from the Carreiras et al., 1996, norms) such as *singer*, *swimmer*, and *art historian*. Of these 20, 5 each were paired with female clothing, male clothing, female biological characteristics, and male biological characteristics. Ten fillers consisted of words with a meaning that was strictly either female (e.g., *heiress*) or male (e.g., *bridegroom*) paired with female or male clothing. Finally, 20 fillers were intended to produce “no” responses and were formed by pairing terms that were strictly male or female with clothing or biological characteristics of the opposite sex. These items also contained a pronoun or possessive in the final part of the sentence to ensure that there was a genuine clash between, for example, someone being described as a *workman* and putting on *her pink swimming costume*. Every participant saw the same 50 filler items.

**Design.** All participants saw both sentences based on items of clothing and sentences based on biological features, so this factor was within participants, but between materials. The same was true of sentences based on male and female stereotypes. Whether the clothing or biological feature matched or mismatched the stereotype varied both within participants and within items.

**Apparatus.** The experiment was carried out using a PC-compatible computer fitted with an Advantech PCLabCard, which provided millisecond timing of responses made via buttons attached to the card. The experiment was controlled by a version of the TSCOP software (Norris, 1984) written by A.G. The sentences were presented on the computer screen, and two buttons, one labeled “yes” and one labeled “no,” were placed between the participant and the screen.

**Procedure.** The participants were tested individually in a small, quiet experimental room. Their task was to read each sentence, which was displayed in three parts, and to judge whether the final part “follows on sensibly from the first two parts.” The participants controlled the display by using the two buttons, and the instructions emphasized that they should read at their normal reading speed, as though they were reading a book or a magazine. They were asked to base their judgments on their first impressions, and not on prolonged deliberation. Before each sentence, the prompt “\$\$ NEXT SENTENCE \$\$” appeared. A buttonpress caused the first part of the sentence to appear. A second buttonpress caused that part of the sentence to disappear and to be replaced by the second part. Another buttonpress caused this clause to be replaced by the final part of the sentence. Participants were asked to keep the index fingers of their dominant and nondominant hands on the “yes” and “no” buttons, respectively, and to use their dominant hand to advance the

display. To make the judgment they pressed either the “yes” button or the “no” button, as appropriate.

Before the main part of the experiment there were nine practice trials to familiarize the participants with the self-paced reading procedure and with the kinds of sentences that they would be reading in the main part of the experiment. Throughout the session, there was a 1-sec interval between the sentences, but the participants were told that they could pause for longer if they wished, as long as they did so only when the prompt “\$\$ NEXT SENTENCE \$\$” was on the screen.

## Results

**Proportion of positive judgments.** The proportion of positive judgments in the main experimental conditions is shown in Table 1. In this and all subsequent analyses, the data were subjected to two analyses of variance, one with participants as the random effect and one with sentences as the random effect.

There was a significant main effect of matching of the stereotype with the following gender-biased information [ $F_1(1,23) = 26.90, MS_e = 0.084, p < .001$ ;  $F_2(1,36) = 62.08, MS_e = 0.015, p < .001$ ], with more positive judgments following a match than a mismatch (.95 vs. .73). There was also a significant interaction of this factor with whether the subsequent information was about clothing or a biological characteristic [ $F_1(1,23) = 17.34, MS_e = 0.009, p < .001$ ;  $F_2(1,36) = 4.50, MS_e = 0.015, p < .05$ ], with the mismatch effect being larger for the clothing sentences than for the biological characteristic sentences (.28 vs. .16). Crucially for our hypothesis, separate analyses for the clothing and the biological characteristic conditions showed that the mismatch effect was significant for both clothing [ $F_1(1,23) = 34.64, MS_e = 0.052, p < .001$ ;  $F_2(1,18) = 30.34, MS_e = 0.025, p < .001$ ] and biological characteristics [ $F_1(1,23) = 14.75, MS_e = 0.041, p < .001$ ;  $F_2(1,18) = 47.09, MS_e = 0.005, p < .001$ ].

In the by-participants analysis, there were also main effects of the type of material [clothing vs. biological, .82 vs. .85;  $F_1(1,23) = 5.94, MS_e = 0.009, p < .05$ ] and of whether this information was biased toward females or males [.85 vs. .82;  $F_1(1,23) = 4.28, MS_e = 0.013, p = .05$ ]. However, these effects were not significant in the by-items analysis, in which they were between-sentence manipulations.

**Judgment Times.** The times for making positive judgments only were analyzed. All times greater than 2.5 *SD* from each participant’s mean were replaced by the 2.5 *SD*

**Table 1**  
Proportion of Positive Judgments in Experiment 1

Sentence Type	Genders	
	Match	Mismatch
Clothing		
Female	.98	.67
Male	.93	.70
Biological		
Female	.96	.81
Male	.91	.74

cutoff (2.6% of times were affected in this way). The mean times are shown in Table 2.

There was a significant main effect of matching of the stereotype with the following gender-biased information [ $F_1(1,23) = 11.86$ ,  $MS_e = 548,422$ ,  $p < .01$ ;  $F_2(1,36) = 22.65$ ,  $MS_e = 191,215$ ,  $p < .001$ ], with faster reading times associated with a match than a mismatch (2,119 vs. 2,488 msec). There was no interaction of this factor with type of material (clothing vs. biological characteristic), and separate analyses for the clothing and the biological characteristic conditions showed that the mismatch effect was significant for both: clothing [ $F_1(1,23) = 14.78$ ,  $MS_e = 260,017$ ,  $p < .001$ ;  $F_2(1,18) = 14.71$ ,  $MS_e = 209,260$ ,  $p < .01$ ]; biological characteristics [ $F_1(1,23) = 6.28$ ,  $MS_e = 432,403$ ,  $p < .05$ ;  $F_2(1,18) = 8.16$ ,  $MS_e = 173,170$ ,  $p < .05$ ].

In the by-participants analysis there were also main effects of the type of material [clothing vs. biological, 2,208 vs. 2,399 msec;  $F_1(1,23) = 15.97$ ,  $MS_e = 109,979$ ,  $p < .001$ ] and of whether this information was biased toward females or males [2,203 vs. 2,404 msec;  $F_1(1,23) = 7.17$ ,  $MS_e = 268,338$ ,  $p < .05$ ]. However, these effects were not significant in the by-items analysis, in which they were between-sentences manipulations. In addition, these reading time effects are for different sentences, which were not exactly matched for length, and so are difficult to interpret.<sup>1</sup>

## Discussion

The main result of this experiment is that the mismatch effect (when the stereotype for the role name clashes with the gender information implicit in the clothing or biological relation information) was demonstrated both in judgments and in times to make judgments. Furthermore, in both data sets, it was demonstrated for both clashes based on clothing information and clashes based on biological characteristics.

The most secure inference from this finding is that the mismatch effect does not depend on the provision of one definitive piece of information about gender. In previous studies (Carreiras et al., 1996), definite pronouns or definite articles provided such information, via their morphological markings. In this experiment, biological characteristics and kinship relations also provided such definitive information, though via semantics rather than morphology. However, information about wearing items

of clothing or makeup is not definitive in this way, even if those items are primarily intended for females or for males.

Our second conclusion is based on plausible reasoning. It is unlikely that people will infer that someone is female or male on the basis of a second piece of probabilistic information, if they have not already done so on the basis of an earlier piece of probabilistic information. If they do not conclude that someone who is a typist is likely to be female, they are unlikely to conclude that someone putting on a silk tie to dress formally is male. And if such inferences are not made, there is no possibility of a clash between the inferred probable genders of the same person, or of that clash affecting comprehension. Our results show that such clashes have strong effects and that the inferences are made. In particular, they suggest that when a stereotyped role name is applied to a person, an inference about the probable gender of that person is made. Thus, it is likely that when a surgeon is later referred to as "she," as in Carreiras et al. (1996, Experiment 1), the comprehension problem arises because someone who has already been represented as (probably) male turns out to be female, and not because the gender stereotype associated with *surgeon* is activated only when the pronoun is read.

To consolidate the findings from Experiment 1, we carried out a second experiment in which the clothing or biological characteristic information was presented first and the stereotyped role name information second. In this experiment we can be fairly sure that gender will be inferred early from the biological characteristics. So, if the items based on clothing behave in the same way as those based on biological characteristics, we can conclude that probabilistic inferences about gender based on items of clothing are made elaboratively. In relation to our hypothesis about inferences based on the occurrence of single words, it is worth noting that the names of items of clothing are typically lexical items.

## EXPERIMENT 2

### Method

**Participants.** The participants were 18 volunteers from the student and staff populations of Sussex University. They were paid £3 for their participation in the experiment, which lasted for about 25 min.

**Materials.** The materials were based on those of Experiment 1. As in that experiment, each sentence was about a character referred to by a stereotyped role name and introduced either a piece of clothing or makeup associated with females or males, or a biological characteristic or kinship relation associated with females or males. However, unlike in Experiment 1, the role name occurred in the final (third) part of the sentence, and the clothing or biological characteristic occurred in the first part. We originally intended to create the items for Experiment 2 by swapping the first and third parts of the sentences and making minor adjustments to make the sentences flow properly. In practice, many of the sentences so created sounded unnatural, particularly since we did not want to use a possessive (e.g., *his silk tie*) or a definite pronoun in the first part of the sentence. The materials for Experiment 2 were, therefore, not related in a one-

**Table 2**  
Judgment Times for Experiment 1

Sentence Type	Genders	
	Match	Mismatch
Clothing		
Female	1,817	2,205
Male	2,130	2,528
Biological		
Female	2,237	2,473
Male	2,247	2,569

to-one manner to those of Experiment 1. Nevertheless, the materials for this experiment conformed to the following specification.

The first part introduced an item of clothing or makeup, or a biological characteristic or kinship relation, for example (14a):

(14a) Wearing a blue garter

The second part provided some further information, but did not introduce any constraints on the character to be introduced in the third part of the sentence. For example, the sentence given above continued as in (14b):

(14b) was the only concession to tradition

The final part of the sentence introduced a character using a stereotyped role name taken from the Carreiras et al. (1996) norms. The other information in this part of the sentence did not constrain the sex of the character. The sentence above ended as in (14c):

(14c) for the dressmaker who was getting married.

An example of a sentence based on a biological characteristic is (15).

(15a) Giving birth less than 30 minutes after arrival

(15b) in the private hospital ward

(15c) the childminder still experienced a great deal of pain.

As in Experiment 1, each sentence occurred in two versions. In one version the stereotype of the role name matched the gender associated with the item of clothing or the biological characteristic. In the other version, there was a mismatch. The mismatched versions were created by swapping role names between sentences, so *dressmaker* in the example above became *politician*, and *childminder* became *paratrooper*. Furthermore, there were other passages in which *dressmaker* and *childminder* occurred in the mismatch condition and *politician* and *paratrooper* occurred in the match condition. Across the whole experiment, therefore, the lexical content of the match and mismatch items was completely balanced. Example materials are shown in Appendix B.

In fact the counterbalancing in this experiment was carried out by rotating stereotypically female, stereotypically male, and neutral role names through a set of 60 sentence frames. Participants saw 20 sentences in each of the three conditions (5 female clothing sentences, 5 female biological characteristic sentences, and 5 each of the corresponding male sentences). However, for comparability with Experiment 1, in which the neutral items were separate from the experimental items, we will treat the items with neutral role names as fillers in the analysis of this study.

In addition to the 40 experimental items and the 20 neutral filler sentences, there were an additional 28 filler items with gender-marked terms. Twenty of these items were designed to produce "no" responses to the following question: "Is the third part of the sentence a sensible continuation of the first two parts?" These items paired a marked term of one gender with an item of clothing or a biological characteristic of the other gender. There were 5 items in each of four categories (female clothing, female biological characteristic, male clothing, and male biological characteristic). The other 8 items were "yes" items, in which the genders matched, 2 in each of the four conditions.

**Design.** The design was the same as in Experiment 1.

**Apparatus.** The apparatus was the same as that used in Experiment 1.

**Procedure.** The procedure the same as in Experiment 1.

## Results

**Proportion of positive judgments.** The proportion of positive judgments in the main experimental conditions is shown in Table 3.

There was a significant main effect of matching the stereotype with the preceding gender-biased information [ $F_1(1,17) = 12.44$ ,  $MS_e = 0.026$ ,  $p < .01$ ;  $F_2(1,56) = 14.69$ ,  $MS_e = 0.018$ ,  $p < .001$ ], with more positive judgments following a match than a mismatch (.93 vs. .84). In the by-items analysis, there was an interaction between matching and the gender associated with the clothing or biological characteristic [ $F_2(1,56) = 4.12$ ,  $MS_e = 0.018$ ,  $p < .05$ ], but this effect was not significant by participants ( $p > .1$ ). The matching effect was larger for female clothing and biological characteristics than for male (.14 vs. .04). Separate analyses for the clothing and the biological characteristic conditions showed that the mismatch effect was significant for both: clothing [ $F_1(1,17) = 6.83$ ,  $MS_e = 0.016$ ,  $p < .05$ ;  $F_2(1,28) = 5.49$ ,  $MS_e = 0.017$ ,  $p < .05$ ]; biological characteristics [ $F_1(1,17) = 14.66$ ,  $MS_e = 0.015$ ,  $p < .01$ ;  $F_2(1,28) = 9.30$ ,  $MS_e = 0.020$ ,  $p < .01$ ].

In the by-participants analysis, there was an interaction of the type of material (clothing vs. biological) with the gender associated with the clothing or biological characteristic [ $F_1(1,17) = 9.04$ ,  $MS_e = 0.012$ ,  $p < .01$ ], which was marginal by materials [ $F_2(1,56) = 3.84$ ,  $MS_e = 0.024$ ,  $.1 > p > .05$ ]. There were more positive responses overall to female clothing items than female biological characteristics (.91 vs. .84), but the reverse was true for male traits (.87 vs. .91).

**Judgment times.** The times for making positive judgments only were analyzed. All times greater 2.5 *SD* from each participant's mean were replaced by the 2.5 *SD* cutoff (1.6% of times were affected in this way). The mean times are shown in Table 4.

There was a significant main effect of matching the stereotype with the preceding gender-biased information [ $F_1(1,17) = 18.04$ ,  $MS_e = 152,762$ ,  $p < .001$ ;  $F_2(1,56) = 5.28$ ,  $MS_e = 405,337$ ,  $p < .05$ ], with faster reading times associated with a match than a mismatch (2,607 vs. 2,884 msec). There was no interaction of this factor with type of material (clothing vs. biological characteristic), and separate analyses for the clothing and the biological characteristic conditions showed that the mismatch effect was significant for both, though only in the by-participants analysis for the items based on clothing: clothing [ $F_1(1,17) = 6.70$ ,  $MS_e = 141,895$ ,  $p < .05$ ;  $F_2(1,28) = 1.59$ ,  $MS_e = 495,917$ , n.s.]; biological characteristics [ $F_1(1,17) = 15.52$ ,  $MS_e = 121,506$ ,  $p < .01$ ;  $F_2(1,28) = 4.52$ ,  $MS_e =$

**Table 3**  
Proportion of Positive Judgments in Experiment 2

Sentence Type	Genders	
	Match	Mismatch
Clothing		
Female	.96	.87
Male	.90	.83
Biological		
Female	.94	.74
Male	.92	.90

**Table 4**  
**Judgment Times for Experiment 2**

Sentence Type	Genders	
	Match	Mismatch
Clothing		
Female	2,458	2,761
Male	2,750	2,907
Biological		
Female	2,616	2,915
Male	2,604	2,952

314,758,  $p < .05$ ]. No other effects or interactions were significant, either by subjects or by items.<sup>2</sup>

## Discussion

The results of this experiment were similar to those of Experiment 1. In particular, there were clear indications of a mismatch effect for both types of material: clothing and biological characteristics. There was some indication that the effects in this experiment were stronger for items based on biological characteristics. However, this pattern was not supported by the presence of a statistically significant interaction between the mismatch effect and the type of item. Furthermore, if anything, the results in Experiment 1 were stronger for items based on clothing than for items based on biological characteristics.

## GENERAL DISCUSSION

Inferences from stereotyped role names to the gender of the people to whom they are applied are one type of inference that can be based on the occurrence of a single lexical item in a text. Such inferences are elaborative in that they are not necessary to establish coherence at the point where the stereotyped name is used. Thus, if such inferences are made during comprehension, this fact would refute a strong theory about constraints on inference—namely, that only those inferences that are necessary for a coherent interpretation of a text are made as a matter of course.

It is plausible to suggest that these inferences are made because they depend on the occurrence of particular lexical items in texts. If the stereotyping information is strongly associated with those lexical items, at least for people in a particular culture, that information might be retrieved with the meanings of those lexical items and incorporated into the representation of the text.

Previous research, especially that of Carreiras et al. (1996), suggested that stereotype inferences were made in a forward, elaborative, direction. However, their results could also be explained in terms of inferences being made when gender was established for certain—for example, from the morphology and semantics of a definite pronoun or a (Spanish) definite article. By using sentences in which gender is never established for certain, but only by two separate (and possibly conflicting) pieces of probabilistic information, we have established

that definitive morphological or semantic information is not necessary for inferences based on stereotypes to be made. Furthermore, a plausible argument can be made that if an inference is based on one piece of probabilistic information (the second piece), then such an inference is probably also made when the first piece of probabilistic information is presented.

The studies presented in this paper constitute just one line of further evidence that we have been collecting for the immediate, and elaborative, making of inferences based on stereotyped role names. We will mention two other types of evidence here. First, Reynolds, Garnham, and Oakhill (2002) have documented the near irrevocable commitment that people make to the maleness of a surgeon in versions of the following riddle (Keenan, 1993; Sanford, 1985, 1987).

This morning a father and his son were driving along the motorway to work, when they were involved in a horrible accident. The father was killed and the son was quickly driven to hospital severely injured. When the boy was taken into the hospital a passing surgeon exclaimed: "Oh my God, that is my son!"

Reynolds et al. (2002) reasoned that if people do not initially take the surgeon to be male, they should not experience difficulties with the passage, since all the other information in it points to the surgeon being the boy's mother. The strong constraint that the boy's father is dead, and cannot be the surgeon, should override the weaker constraint that surgeons are typically male (but can also be female), and lead to the conclusion that the surgeon is the boy's mother.

Second, Oakhill, Garnham, and Reynolds (2000) have shown that the likely sex of characters described using role names cannot be suppressed. People were asked to judge whether two terms, such as *nurse* and *brother*, could apply to the same individual. The question was deliberately phrased as one of possibilities, rather than probabilities. Under a variety of conditions, although people were usually correct, they were slower and made more errors when the stereotype associated with the role name mismatched the sex of the kinship term (e.g., *nurse-uncle* or *engineer-mother*).

In view of this converging evidence, we believe that there is now strong evidence that inferences about the probable genders of people are drawn immediately when a stereotyped role name is used to describe them.

## REFERENCES

- BRANSFORD, J. D., BARCLAY, J. R., & FRANKS, J. J. (1972). Sentence memory: A constructive vs. interpretive approach. *Cognitive Psychology*, *3*, 193-209.
- BRANSFORD, J. D., & FRANKS, J. J. (1971). The abstraction of linguistic ideas. *Cognitive Psychology*, *2*, 331-350.
- 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.

- CORBETT, A. T., & DOSHER, B. A. (1978). Instrument inferences in sentence encoding. *Journal of Verbal Learning & Verbal Behavior*, **17**, 479-491.
- GARNHAM, A. (1982). Testing psychological theories about inference making. *Memory & Cognition*, **10**, 341-349.
- GARNHAM, A. (1989). Inference in language understanding: What, when, why and how. In R. Dietrich & C. F. Graumann (Eds.), *Language processing in social context* (pp. 153-172). Amsterdam: North-Holland.
- GARROD, S. C., & SANFORD, A. J. (1981). Bridging inferences and the extended domain of reference. In J. Long & A. Baddeley (Eds.), *Attention and performance IX* (pp. 331-346). Hillsdale, NJ: Erlbaum.
- GRAESSER, A. C., SINGER, M., & TRABASSO, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review*, **101**, 371-395.
- HAVILAND, S. E., & CLARK, H. H. (1974). What's new? Acquiring new information as a process in comprehension. *Journal of Verbal Learning & Verbal Behavior*, **13**, 512-521.
- KEENAN, J. M. (1993). Thoughts about the minimalist hypothesis: Commentary on Garnham on reading inference. *Psychology*, **4**(2).
- KEENAN, J. M., POTTS, G. R., GOLDING, J. M., & JENNINGS, T. M. (1990). Which elaborative inferences are drawn during reading? A question of methodologies. In D. A. Balota, G. B. Flores d'Arcais, & K. Rayner (Eds.), *Comprehension processes in reading* (pp. 377-402). Hillsdale, NJ: Erlbaum.
- 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. 195-202). Amsterdam: North-Holland.
- McKOOON, G., & RATCLIFF, R. (1990). Textual inferences: Models and measures. In D. A. Balota, G. B. Flores d'Arcais, & K. Rayner (Eds.), *Comprehension processes in reading* (pp. 403-422). Hillsdale, NJ: Erlbaum.
- McKOOON, G., & RATCLIFF, R. (1992). Inference during reading. *Psychological Review*, **99**, 440-466.
- NORRIS, D. G. (1984). A computer-based programmable tachistoscope for nonprogrammers. *Behavior Research Methods, Instruments, & Computers*, **16**, 25-27.
- OAKHILL, J. V., GARNHAM, A., & REYNOLDS, D. J. (2000). *Immediate activation of stereotypical gender information in reading*. Manuscript submitted for publication.
- REYNOLDS, D. J., GARNHAM, A., & OAKHILL, J. V. (2002). *Evidence of immediate activation of gender information from a social role name*. Unpublished manuscript, University of Sussex, Laboratory of Experimental Psychology.
- SANFORD, A. J. (1985). *Cognition and cognitive psychology*. London: Weidenfeld & Nicolson.
- SANFORD, A. J. (1987). *The mind of man: Models of human understanding*. Brighton, U.K.: Harvester.
- SINGER, M. (1980). The role of case-filling inferences in the comprehension of brief passages. *Discourse Processes*, **3**, 185-201.
- THORNDYKE, P. W. (1976). The role of inferences in discourse comprehension. *Journal of Verbal Learning & Verbal Behavior*, **15**, 437-446.

## NOTES

1. It might be asked whether the mismatch effects are restricted to the first items presented, and whether people adjust their reading strategies once they have read a number of items with mismatching stereotypes. To investigate this possibility we calculated, for each person in the experiment, the mean judgment times for the first 10 congruent and incongruent passages and compared them with the judgment times for the second 10 of each. There was a main effect of "half," with people speeding up as they went through the experiment [ $F(1,23) = 25.26, p < .001$ ]. However, there was no interaction between "half" and congruity [ $F(1,23) = 1.56, p = .21$ ].

2. As in Experiment 1, there was no indication of readers developing a strategy after reading a number of mismatching items (see Note 1). Again, there was a main effect of "half" of the experiment [ $F(1,17) = 14.96, p < .005$ ], but no interaction of that factor with stereotype matching [ $F(1,17) = .40, p = .54$ ].

## APPENDIX A

## Examples of Material Used in Experiment 1

## Clothing, Match

The midwife, who had just finished a hard day's work, went to get changed for swimming, and put on a striped bikini.

## Clothing, Mismatch

The plasterer, who had just finished a hard day's work, went to get changed for swimming, and put on a striped bikini.

## Biological Characteristic, Match

The housekeeper was rushed to hospital, taken to a private ward, and gave birth half an hour later.

## Biological Characteristic, Mismatch

The soldier was rushed to hospital, taken to a private ward, and gave birth half an hour later.

## APPENDIX B

## Examples of Material Used in Experiment 2

## Clothing, Match

After buying a new, comfortable strapless bra in the clothes store the dressmaker went to meet a friend.

## Clothing, Mismatch

After buying a new, comfortable strapless bra in the clothes store the politician went to meet a friend.

## Biological Characteristic, Match

Joining the other bass singers from the Welsh choir after leaving early from work the police officer was still late for choir practice.

## Biological Characteristic, Mismatch

Joining the other bass singers from the Welsh choir after leaving early from work the traffic warden was still late for choir practice.