

How Natural are Conceptual Anaphors?

Jane Oakhill, Alan Garnham

*Laboratory of Experimental Psychology, University of Sussex, Brighton,
East Sussex, U.K.*

Morton Ann Gernsbacher

*Department of Psychology, University of Oregon, Eugene, Oregon,
U.S.A.*

Kate Cain

*Laboratory of Experimental Psychology, University of Sussex, Brighton,
East Sussex, U.K.*

This paper reports three experiments on the interpretation of "conceptual" anaphors. These are anaphors that do not have an explicit linguistic antecedent, but one constructed from context. For instance, if one says "I need a knife. Where do you keep *them*?", *them* means something like "the knives that I presume you have in your house". In the first experiment, subjects rated sentences containing conceptual anaphors, of three different types, to be as natural as ones with a "linguistically correct" antecedent (e.g. "I need an iron. Where do you keep it?"), and as more natural than ones with neither a plausible conceptual antecedent nor a plausible linguistic one. In a second (self-paced) experiment, subjects judged whether the second sentence in such pairs was a sensible continuation from the first, and the time to make these judgements was measured. We found that acceptability judgements were high, and judgement times low, in just those sentences that were rated as more natural in the first experiment. These first two experiments showed that conceptual anaphors are quite easily understood. However, they did not show that such anaphors are processed without difficulty. In the third

Requests for reprints should be addressed to Jane Oakhill, Laboratory of Experimental Psychology, University of Sussex, Brighton, East Sussex BN1 9QG, U.K.

The research reported in this paper was supported by Economic and Social Science Research Council Grant COO 23 2439 to Garnham and Oakhill, and by NIH Research Career Development Award KO4 NS-01376 and NIH Research Grant RO1 NS-29926-01 to Morton Gernsbacher. Our collaboration began at the Sylvania Beach Language Comprehension Conference in 1989, sponsored by the Air Force Office of Scientific Research (AFOSR 89-0306) and the University of Oregon. We would also like to thank the two reviewers for many insightful comments which helped to clarify our thinking and writing.

experiment, we therefore compared conceptual anaphors (“plate . . . them”) with matched plural anaphors whose antecedents were explicit (“some plates . . . them”). The results were different for different types of anaphor: in one case (pronouns that referred to collective sets), the conceptual version followed by a plural pronoun was easier than the explicit plural version. For the other two types (references to generics and to implied multiple items), the explicit plurals were understood more rapidly than their conceptual counterparts.

INTRODUCTION

Although anaphors are widely used in texts and in conversation, and people have few problems understanding them, they pose many questions for psychologists and linguists trying to explain how they are understood.

There are certain constraints on the use of anaphors which are likely to guide comprehenders. First, anaphors and their antecedents usually agree in number, gender and case. Secondly, anaphors obey syntactic constraints, such as the C-command rule (Reinhart, 1983). Thirdly, thematic constraints, factors such as topic and focus, guide the search for antecedents. Fourthly, pragmatic constraints derived from knowledge of the world can guide processing. However, anaphors that violate the first constraint do not seem to pose major processing difficulties. A common example is the use of “they” to avoid a commitment to a particular gender: “I think I’ll ask a shop assistant for some help. *They* might know . . .”. Gernsbacher (1991) provides numerous other examples of such “linguistically illegal” pronouns.

In this paper, we will report three experiments on the interpretation of “conceptual” anaphors. These are anaphors that do not have an explicit linguistic antecedent, but one constructed from context. For instance, if one says “I need a knife. Where do you keep *them*?”, *them* means something like “the knives that I presume you have in your house”. Gernsbacher (1991) has investigated the understanding of such anaphors. She identified three distinct types of situations in which conceptual anaphors occur: (1) they are used to refer to frequently occurring events or multiply occurring items (as opposed to unique events and items); (2) they are used to refer to generic types (as opposed to specific tokens); and (3) they are used to refer to members of a collective set (as opposed to individual members of a set). In two experiments, Gernsbacher found that when sentences contained “illegal” plural pronouns that referred to multiple items or events, generic types or collective sets (i.e. pronouns that had no antecedent that matched in number), they were rated as more natural and comprehended more quickly than the same sentences with legal singular pronouns. In addition, they were rated as natural, and compre-

hended as rapidly, as legal singular pronouns that referred to unique items or events, specific tokens or individual members of a set.

Thus, in the above example, the second sentence of:

I need a plate. Where do you keep them?

was more acceptable than:

I need a plate. Where do you keep it?

and was as acceptable as the second sentence in:

I need an iron. Where do you keep it?

where the expectation is that, in a likely context, for example a person’s house, only *one* iron is likely to be available. Of course, other contexts are possible – in a shop, for example, where there are likely to be lots of irons, a plural pronoun may be more acceptable (“I need a new iron. Which aisle are they on?”). We will come back to this point later.

Gernsbacher’s results stress the important role that everyday knowledge plays in the interpretation of anaphora. However, Gernsbacher’s experiments did not address one important issue. Although she claimed that “conceptual, though technically illegal, anaphors *do not pose processing difficulties*”, this assertion was not tested explicitly in her experiments because she did not include a “legal” plural condition. Therefore, she did not compare a plural conceptual anaphor with a plural pronoun following an *explicit* plural antecedent. Thus, we do not know, for instance, whether “Where do you keep them” is equally easy whether it follows “I need a plate” or “I need some plates”. The inclusion of both conditions can address the issue of *when* processing occurs. Are several plates introduced into the mental representation when subjects read the first sentence, or does a retrospective inference need to be made when the pronoun is encountered? If final sentence processing times are shorter in the explicit plural condition, then we can infer that the conceptual cases do require some inferencing, and that conceptual uses of pronouns, though very acceptable, are not as readily comprehended as matched plural pronouns with explicit plural antecedents.

In order to test these hypotheses, we first needed to produce some revised and Anglicised materials for use with British English speakers. An examination of Gernsbacher’s (1991) materials led to the identification of some ambiguous stimuli. For example, in some of the sentences designed to permit either a generic or a specific reading of a subsequent pronoun (“I’m craving a diet coke” vs “I’m craving a diet coke with a twist of lime”), either the singular or the plural pronoun could have a generic interpretation following *either* sentence. In other sentences, which were designed to set up a context where many similar items or only one item

would be likely to exist, the singular pronouns could sometimes be interpreted generically. For example, in the (unique item) text: “Do you know where I can get a roll of dental floss? I never seem to have enough of them/it”. In this text, *them* can be interpreted as “rolls of dental floss”, whereas *it* is more likely to be interpreted as “dental floss in general”, rather than “the roll of floss that I need”. The first two experiments reported here are, therefore, essentially replications of Gernsbacher’s experiments with revised materials, and the third experiment tests the hypotheses outlined above. In the second and third experiments, we also asked subjects to state how they had interpreted the pronouns, to gain additional data about how they had been understood.

EXPERIMENT 1

Method

Subjects. The subjects were 48 volunteers from the student population of Sussex University, all of whom were naive as to the purpose of the experiment. They were paid for participating.

Materials. The materials were adapted from those used in Gernsbacher (1991). We changed the materials to make them comprehensible to speakers of British English, and rewrote any target sentence that contained more than one referring pronoun, for example, “*He* listened to *him* without saying a word”. In many cases, a simple adaptation of the original version was not possible, and many new sentence pairs were written specifically for this experiment. The materials were of the three types described above, with 16 materials of each type: multiple/unique items (“I need a plate/iron. Where do you keep it/them”); collective sets/individuals (“My sister went to work for IBM/the manager at IBM. They/he made her a very good offer”); and generic/specific (“Jimmy went to see a horror movie/the remake of *Dracula* yet again. He always has nightmares after seeing it/them”). It is important to note that the antecedent noun phrases in the multiple and generic context sentences did not have multiple or generic readings *in their own right*. Rather, they provided a context that licensed a multiple or generic reading for the plural pronouns in the second sentences. In the latter cases, the generics might have more appropriately been termed “non-specific” since, in those texts, the first sentence always contained a non-specific noun phrase. However, this terminology does not capture the essence of the texts – that they allowed a generic interpretation of a subsequent plural pronoun. We have, therefore, retained Gernsbacher’s (1991) terminology and referred to these texts as “generic”.

There were four variants of each sentence pair. The four variants were allocated to different lists, and each list contained equal numbers of

sentence pairs from the different conditions. The lists were randomised separately, and each list was presented to a given subject in one of two set random orders. At the beginning of each list there were two “lead in” materials (which were constant across lists), so that the subjects could see what was required of them. Under each sentence pair, there was a 5-point rating scale with the end-points tagged “not natural” and “very natural”. One of each type of text, in each of its four versions, is shown in Table 1.

TABLE 1
Examples of the Materials used in Experiment 1

<i>Multiple/unique items</i>	
1. Multiple item singular pronoun	I need a plate. Where do you keep it?
2. Multiple item plural pronoun	I need a plate. Where do you keep them?
3. Unique item singular pronoun	I need an iron. Where do you keep it?
4. Unique item plural pronoun	I need an iron. Where do you keep them?
<i>Generic types/specific tokens</i>	
1. Generic type singular pronoun	I was really frightened by a Doberman. It is a dangerous beast.
2. Generic type plural pronoun	I was really frightened by a Doberman. They are dangerous beasts.
3. Specific token singular pronoun	I was really frightened by my neighbour’s dog. It is a dangerous beast.
4. Specific token plural pronoun	I was really frightened by my neighbour’s dog. They are dangerous beasts.
<i>Collective sets/individual members</i>	
1. Collective set singular pronoun	Last night we went to hear a new jazz band. It played for nearly five hours.
2. Collective set plural pronoun	Last night we went to hear a new jazz band. They played for nearly five hours.
3. Individual member singular pronoun	Last night we went to hear a new jazz guitarist. He played for nearly five hours.
4. Individual member plural pronoun	Last night we went to hear a new jazz guitarist. They played for nearly five hours.

Procedure. The subjects were given written instructions which stated that they should read the sentence pairs and then rate how natural the second sentence sounded to them. They were told explicitly that they should not base their judgements on an assessment of grammaticality, but should decide on the basis of how easy it was to understand what the second sentence was about – who or what was being referred to. The subjects were told to indicate their rating by circling one of the numbers on the 5-point rating scale which appeared immediately below each sentence. They were tested in small groups in a quiet room.

Results

The mean ratings for the different versions of the sentences within each of the three groups are shown in Table 2. The three groups of sentences (multiple/unique items, collectives/individuals and generics/specifics) were analysed separately by analysis of variance.

Multiple vs Unique Items. In general, multiple item passages were rated more natural than unique item passages [$F(1,40) = 65.73, P < 0.001$; $F(1,15) = 9.99, P < 0.007$] and passages containing singular pronouns were rated more highly than those containing plural ones

TABLE 2
Mean "Naturalness" Ratings
for the Target Sentences for
the Three Types of Text (Mul-
tiple vs Unique Items, Gener-
ic Types vs Specific Tokens and
Collective Sets vs Individuals)

Texts	Pronouns	
	Plural	Singular
Multiple	4.32	3.78
Unique	2.63	4.48
Generic	4.42	3.48
Specific	3.05	4.48
Collective	4.44	3.06
Individual	2.94	4.31

Note: Ratings were on a scale of 1–5, where 1 = very unnatural and 5 = very natural.

[$F(1,40) = 55.95, P < 0.001$; $F(1,15) = 15.87, P < 0.002$]. However, as can be seen from Table 2, the most striking effect in these data is that the plural pronoun versions of the multiple item texts were rated more highly, whereas the singular pronoun versions of the unique item texts were preferred. These differences resulted in a significant interaction between the two factors [$F(1,40) = 199.09, P < 0.001$; $F(1,15) = 37.97, P < 0.001$].

Generic Types vs Specific Tokens. There was a tendency for generic pairs to be rated as more natural than specific ones [$F(1,40) = 5.62, P < 0.03$] and a tendency for sentences containing plural pronouns to be rated more highly than those containing singular ones [$F(1,40) = 17.05, P < 0.001$], though neither of these main effects was significant by materials. Again, the two factors interacted [$F(1,40) = 169.16, P < 0.001$; $F(1,15) = 70.53, P < 0.0001$]. As can be seen from Table 2, the plural pronouns were rated as more natural when they followed generic context sentences, and the singular ones when they followed specific contexts.

Collective Sets vs Individuals. Neither of the main effects was significant in either the analyses by subjects or by materials. As for the other two sets of materials, however, there was a highly significant interaction between the factors. The plural pronouns were rated more natural following collective sets, and the singular pronouns following individuals [$F(1,40) = 148.96, P < 0.001$; $F(1,15) = 129.17, P < 0.001$].

Discussion

These results replicate in essence the findings from Gernsbacher's first experiment. Indeed, in almost all cases, the mean ratings were very similar to those obtained in Gernsbacher's study. There was only one data point that stood out as being different: the relatively high naturalness rating accorded to multiple items followed by singular pronouns (e. g. "Is there a record I could listen to? Yes, it's on top of the hi-fi unit"). Although the interaction was highly significant for this group of materials, the precise nature of the interaction was rather different from that found with the other two types of material. We could not find any obvious explanation for this pattern of results. Nevertheless, even in this condition, the data relating to the "conceptual" case are clear: the multiple contexts followed by "illegal" plural pronouns were given very high average naturalness ratings (4.32) and their ratings were almost as high as those given to the unique items followed by ("legal") singular pronouns (4.48).

In the second experiment, we used the same materials in an on-line judgement task. The task was slightly different to that used by Gernsbacher in her second experiment – instead of paraphrasing the second (target) sentence, our subjects were asked to state explicitly how they interpreted the target pronoun after reading the target sentence. We anticipated that this task would serve both to focus their attention on the interpretation of the target pronouns, and would result in a higher proportion of data that could be used than would the paraphrase task.

EXPERIMENT 2

Method

Subjects. The subjects were 24 volunteers from the student population of Sussex University. They were paid to participate in this experiment and one other that was undertaken in the same experimental session.

Materials. The materials were the same as those used in Experiment 1, to which we made some very minor modifications for the following reason. In this experiment, we wanted to ask the subjects how they had interpreted the pronouns; therefore, in order to avoid confusion, we changed any sentences containing “dummy it” where the target pronoun was also *it*. So, for example, in the passage:

My husband always throws down his shirt/winter coat when he goes to bed.

It's a bore to keep picking it/them up.

we changed the second sentence to:

He always expects me to pick it/them up.

to avoid two occurrences of *it*.

Design. As in Experiment 1, four lists of materials were derived, so that each sentence pair occurred in one of its four versions in each list, and each list contained equal numbers of each version for each of the three types of materials. Each list was randomised separately, and there were two set random orders for each. The subjects were assigned randomly to one of the lists in one of its random orders.

Apparatus. The experiment was controlled on-line by a 6809-based microcomputer system. The sentences (and the third display containing a single pronoun) were presented in the centre of a TeleVideo TVI-912

visual display unit (VDU). There was a button box with a single button between the subject and the VDU.

Procedure. The subjects were tested individually in a small experimental room. Their task was to read the sentences in each text, which were displayed separately. The instructions emphasised that the subjects should read the sentences at their normal reading speed. The subjects advanced the display by pressing the response button with their dominant hand. A pronoun from the second sentence (which took its meaning from the first) was then shown in a third display. The pronoun was displayed in capital letters and appeared in the centre of the screen, as did the two previous sentences. The subjects were required to write down what the pronoun referred to. The pronoun remained on the screen until the subjects had finished their written response and were ready to move on to the next trial.

Before the 48 experimental trials, there were 8 practice trials to familiarise the subjects with the self-paced reading technique, and with the sorts of texts that they would be reading in the experiment. There was a 1-sec interval between sentence pairs, but the subjects were told that they could pause for longer if they wished, as long as they did so only when the \$\$\$NEXT TEXT\$\$\$ prompt was on the screen.

Results

Reading Time Data

The reading times for the second sentences were subjected to analysis of variance. Any data points that were more than 2.5 standard deviations from the mean for a particular subject or a particular item were replaced by the cut-off score. Altogether, 3% of the data points were replaced in this way. The mean reading times for the target sentences are shown in Table 3.

Multiple/Unique Items. As can be seen from Table 3, the sentences containing plural pronouns took a particularly long time to understand when they followed “unique items”, resulting in a significant interaction between multiple/unique context in the first sentence, and whether there was a singular or plural pronoun in the second sentence [$F(1,20) = 17.18$, $P < 0.001$; $F(1,15) = 14.14$, $P < 0.002$]. Using *t*-tests, it was confirmed that performance in the other three conditions was very similar – there was no significant difference in the reading times between the plural and singular pronouns when they followed a multiple context (“a plate . . . them” vs “a plate . . . it”), and the multiple plural condition was not significantly slower than the unique singular condition (“iron . . . it”) (all t s < 1).

TABLE 3
Mean Reading Times (msec)
for the Target Sentences Pre-
ceded by the Different Types
of Context Sentences (Multi-
ple vs Unique Items, Generic
vs Specific Contexts and
Collective Sets vs Individuals)

Texts	Pronouns	
	Plural	Singular
Multiple	2215	2207
Unique	2936	2116
Generic	2398	2711
Specific	2949	2410
Collective	2499	2887
Individual	3776	2713

In general, the second sentences of the multiple item texts were read faster than the unique item texts [$F(1,20) = 21.93, P < 0.001; F(2,15) = 6.61, P < 0.03$] and singular pronouns tended to be read faster than plural ones, though this effect was only significant in the analysis by subjects.

Generic Types vs Specific Tokens. As can be seen in Table 3, the generic contexts followed by a plural pronoun and the specific contexts followed by a singular pronoun were easier than the other two conditions, resulting in a highly significant interaction between the generic vs specific context in the first sentence and singular vs plural pronoun in the second [$F(1,20) = 13.92, P < 0.002; F(2,15) = 10.83, P < 0.005$]. This interaction arose because of the very long responses to the specific contexts followed by plural pronouns. Using *t*-tests, it was shown that there was no difference between the generic context conditions: in these contexts, singular and plural pronouns were equally easy to understand. Furthermore, specific contexts followed by a singular pronoun were not significantly easier than generic contexts followed by a plural. Neither of the main effects was significant in either the analyses by subjects or by items.

Collective Sets vs Individuals. In general, texts containing collective sets in the first sentence were read faster than those containing individuals [$F(1,20) = 17.12, P < 0.001; F(2,15) = 8.66, P < 0.01$]. Texts containing singular pronouns were read faster – the main effect of singular vs plural

pronoun in the second sentence was also significant [$F(1,20) = 8.67, P < 0.01; F(2,15) = 4.69, P < 0.05$]. However, there was a highly significant interaction between whether a collective set or an individual was introduced in the first sentence, and whether the second contained a singular or plural pronoun [$F(1,20) = 36.96, P < 0.001; F(2,15) = 31.12, P < 0.001$].

Using *t*-tests, it was shown that there was a significant difference between the type of pronoun referent used: collectives followed by a plural pronoun were read significantly faster than those referred to by a singular (though linguistically matching) pronoun [$t(23) = 3.09, P < 0.005; t(15) = 2.42, P < 0.03$]. In addition, collective terms followed by a plural pronoun were no more difficult than individual members followed by a singular pronoun. Indeed, the linguistically illicit plural pronouns were read *faster* in this case, as one would expect if readers were setting up a representation of the collective term as a *group of people*.

So, in general, the conceptual pronouns, although they had no matching linguistic antecedent, took no longer than singular pronouns in the same context, and no longer than the individual, specific or unique items followed by a singular pronoun. These results agree very well with those of Gernsbacher (1991). The only exception was the collective texts, where the “illegal” conceptual plural texts (“IBM . . . they”) were actually read *faster* than the collective singular ones (“IBM . . . it”).

Interpretation of Pronouns

If conceptual (plural) pronouns preceded by multiple items or generic types are interpreted conceptually, then when readers write down their interpretations, we would expect them to produce plural nouns, rather than the singular noun phrase antecedents that were specified in the text. In the case of collectives, subjects might “fill out” the collective term: “*the managers at IBM*”, “*the people at the phone company*”, etc. In contrast, if legal singular pronouns referring to unique items, specific types or individuals are interpreted literally, subjects would be expected to respond to those with singular noun phrases.

The responses were classified as either singular, plural or “other”. In many cases, scoring of the written responses was straightforward because the responses followed the exact form of the intended antecedent in the text. For responses that did not conform, two scoring criteria were adopted. First, “gist” responses were allowed. In many cases, these were indicative of the subjects forgetting the original wording. They included synonyms and near synonyms, e.g. “horror films” for “horror movies” and “the Dutch” for “Dutchmen”. Secondly, plausible interpretations of the text were permitted, e.g. “the noise from the Walkman” as an interpreta-

TABLE 4
Classes of Responses in Pronoun Interpretation Task: Experiment 2 (max = 96)

	Singular Context (Individual/Unique/Specific)			Conceptual Context (Collective/Multiple/Generic)											
	Singular Pronoun		Plural Pronoun	Singular Pronoun		Plural Pronoun									
	si.	pl.	oth.	co.	si.	pl.	oth.	co.	si.	pl.	oth.				
Collective vs individual	95	1	0	22	36	30	8	92	0	2	2	88	0	8	0
Multiple vs unique	96	0	0	—	32	61	3	—	93	3	0	—	22	73	1
Generic vs specific	96	0	0	—	29	65	2	—	90	3	3	—	24	72	0

Note: Responses = si., singular; pl., plural; oth., other; co., collective.

tion of *it* in "A lot of people wear a Walkman. Sometimes it is so loud it can be heard by other people" and "to buy a Mac computer" as an interpretation of *it* in "I'm thinking of buying a Macintosh computer. The only trouble is, it's so expensive".

In some cases, the subjects gave two alternative responses. These were permitted if both were singular or both plural, and both plausible (e.g. Dutchmen/women). Otherwise, they were classed as "other". Wrong antecedents from the text, and implausible antecedents that were not directly derived from the text, were also classed as "other".

The numbers of responses falling into each category are shown in Table 4. In the case of the collectives, there were a number of "collective" responses, especially in the collective set/plural pronoun and the specific/plural pronoun conditions. For the purposes of analysis, we classed these with the singular responses, since they are, strictly, linguistically singular. However, they are shown separately in Table 4. These responses were particularly interesting in the individual plus plural pronoun cases, where there was no collective term in the text, but in many cases (23%) the subjects constructed a collective term for themselves in an attempt to make sense of the text. Some examples are: "they" in "John Paul Getty . . . they" was interpreted as "The John Paul Getty Association". Similarly, "The borough architect . . . they" was interpreted as "The borough architect's office".

In the case of collectives, we had expected that subjects might respond to the conceptual plurals with an expanded version of the collective term. In fact, though, they simply repeated the collective term a high proportion of the time - another confirmation of its acceptability as the antecedent for a plural pronoun. Almost all of the responses in the conceptual contexts were collectives, and there was no difference depending on whether the conceptual was followed by a singular or a plural pronoun (both $t_s < 1.08$). Neither was there any significant difference between the number of collective responses in the collective plus plural case and the number of singular responses in the individual plus singular pronoun case, though there tended to be more of the latter type of response, and this contrast was marginally significant by items [$t_1(23) = 1.39, P = 0.19$; $t_2(15) = 1.9, P = 0.07$].

The other two types of materials produced a very similar pattern of data to one another. The conceptual plus plural pronoun conditions resulted in the production of large numbers of plurals. In the case of the multiple items, the number of plural responses was higher in the conceptual (multiple) plus plural case than in the unique plus plural condition, though the effect did not reach significance by items [$t_1(23) = 2.27, P < 0.02$; $t_2(15) = 1.77, P = 0.097$]. In the case of the generic materials, though a large number of plural responses was produced in the conceptual plus

plural condition, the difference between this condition and the specific plus plural condition was not significant (both t s < 1.27). These weak effects may have arisen because subjects were forced to write down something and, given an (anomalous) plural pronoun in the unique/specific plus plural cases, they really had no option but to produce a plural response. In the next experiment, we included “nothing sensible” as an option, and the results in the highly infelicitous condition were rather different.

The differences in responses that we found in the multiple/unique and generic/specific materials suggest that the conceptual pronouns were not as acceptable as their corresponding linguistically matched pronouns for those two types of material. The multiple (conceptual) plus plural condition resulted in fewer plural responses than the unique plus singular condition resulted in singular responses [$t(23) = 3.44, P < 0.002; t(15) = 5.97, P < 0.001$]. Similarly, the generic (conceptual) plus plural condition resulted in fewer plural responses than the specific plus singular resulted in singular responses [$t(23) = 3.46, P < 0.002; t(15) = 5.20, P < 0.001$].

Discussion

The reading time results broadly mirror those of Gernsbacher (1991). As in the “naturalness rating” data (Experiment 1), there was one discrepant cell mean for the multiple *vs* unique items. Although the factors multiple/unique and plural/singular pronoun interacted, in these materials the interaction arose solely because the unique items followed by plural pronouns were particularly difficult – as in the rating study, the multiple items followed by singular pronouns did not show any evidence of causing the subjects difficulty.

Otherwise, we found that the conceptual pronouns were understood very readily, though it was only in the collective materials that the conceptual versions showed any signs of being *more* acceptable than the corresponding linguistic matches.

EXPERIMENT 3

In this third experiment, we compared the conceptual anaphors with matched *explicit* plural versions, in order to test the hypothesis that conceptual pronouns do not cause *any* processing difficulties. For example, “I need a plate. Where do you keep them?” was compared with “I need some plates. Where do you keep them?” Although even in the plural context condition (“some plates”) the pronoun does not actually mean “the plates that I need”, and some inferential work is required, similar work is required in the comparable conceptual case. What is of interest is whether the superficial number match aids processing.

Method

Subjects. The subjects were 24 volunteers from the student population of the University of Sussex, who had not participated in either of the previous experiments on conceptual anaphors. They were paid to participate in this experiment and one other.

Materials. The 48 materials used in this experiment were derived from those in Experiment 2. We changed the first sentences, so that, in one of the four versions derived from each material, the first sentence contained an explicit plural. Thus, for each material, we could compare an explicit plural reading with a conceptual plural reading. In the case of the multiple items, this was achieved by changing, for example, “I need a plate” to “I need some plates”. In the case of collectives, the multi-person nature of the collective was made explicit: “the class”, for example, was changed to “the students in the class”. It was necessary, however, to treat the generic sentences in rather a different way. Because, in this case, plurals could still take a generic meaning, we would not know in which sense a plural pronoun following them was being interpreted. For instance, if “Jimmy went to see a horror movie . . .” is changed to “Jimmy went to see some horror movies” and this sentence is then followed by “He always has nightmares after seeing them”, the “them” could mean *either* “the particular horror movies that he saw” or “horror movies in general” and, in fact, the more likely reading still seems to be the *generic* one. Thus, changing the first sentence to a plural form in the generic texts does not necessarily change the reading of the plural pronoun from a generic to an explicit plural one. To get round this problem, we changed the *specific* versions of the texts instead. So, for instance, the specific singular “Every Thursday, Carla watches Dallas” was changed to “On Thursday evening, Carla watches Dallas and Eastenders”. Thus, we could compare the generic singular “Every evening, Carla watches a soap opera” with the explicit specific plural, when each was followed by “they”.

One of each type of text, in each of its four versions, is shown in Table 5. For all texts, the third display – a single pronoun – was the same as in the previous experiment, and the subjects were required to write down their interpretation of that pronoun.

Design and Procedure. The design of the experiment, and the apparatus used, were the same as in Experiment 2. The procedure was also the same as that in Experiment 2, except that when the subjects were asked to write down a meaning for the pronoun in the final display, an extra response option was included. If the subjects decided that there was no sensible interpretation of the pronoun, i.e. there was nothing in the first

sentence that it could reasonably take as its antecedent, then they were given the option to write N/S (for "nothing sensible") as their response. This response option was included because the explicit plural/singular pronoun texts were highly infelicitous and it allowed the subjects to reject texts rather than forcing them to try to find an interpretation for a pronoun. The 48 experimental trials were preceded by 8 practice trials to

TABLE 5
Examples of the Materials used in Experiment 3

<i>Multiple/unique items</i>	
1. Multiple item singular pronoun	I need a plate. Where do you keep it?
2. Multiple item plural pronoun	I need a plate. Where do you keep them?
3. Explicit plural singular pronoun	I need some plates. Where do you keep it?
4. Explicit plural plural pronoun	I need some plates. Where do you keep them?
<i>Generic types/specific tokens</i>	
1. Generic type singular pronoun	Every evening, Carla watches a soap opera. She'd watch it all day, if she could.
2. Generic type plural pronoun	Every evening, Carla watches a soap opera. She'd watch them all day, if she could.
3. Explicit plural singular pronoun	On Thursday evening, Carla watches Dallas and Eastenders. She'd watch it all day, if she could.
4. Explicit plural plural pronoun	On Thursday evening, Carla watches Dallas and Eastenders. She'd watch them all day, if she could.
<i>Collective sets/individual members</i>	
1. Collective set singular pronoun	Last night we went to hear a new jazz band. It played for nearly five hours.
2. Collective set plural pronoun	Last night we went to hear a new jazz band. They played for nearly five hours.
3. Explicit plural singular pronoun	Last night we went to hear our friends play in a new jazz band. It played for nearly five hours.
4. Explicit plural plural pronoun	Last night we went to hear our friends play in a new jazz band. They played for nearly five hours.

familiarise the subjects with the experimental procedure. The practice trials consisted of two examples of each of the four types of experimental text.

Results

Reading Time Data

As in Experiment 2, any reading times that deviated from the mean by more than 2.5 standard deviations were replaced by the cut-off score. In this way, 2.5% of the data points were replaced. The data were then analysed by analysis of variance, with separate analyses for each type of sentence. The mean reading times for the target sentences are shown in Table 6.

Multiple Items. Overall, the target sentences took longer to read when the first sentence provided an explicit, rather than a conceptual context [$F(1,20) = 15.22, P < 0.001$; $F(1,15) = 7.69, P < 0.02$]. They also took longer when the pronoun in the target sentence was singular rather than plural [$F(1,20) = 34.78, P < 0.001$; $F(1,15) = 19.02, P < 0.001$]. However, the plural pronouns in the explicit plural contexts were particularly easy, and the singular pronouns in these contexts particularly difficult, resulting in a significant interaction between whether the first sentence contained a conceptual or an explicit plural and whether there was a singular or plural pronoun in the second sentence [$F(1,20) = 44.52, P < 0.001$; $F(1,15) = 13.92, P < 0.002$].

TABLE 6
Mean Reading Times (msec) for the
Target Sentences Preceded by Concep-
tual vs Explicit Plurals

Texts	Pronouns	
	Plural	Singular
Conceptual (multiple)	1988	2137
Explicit plural	1787	3086
Conceptual (generic)	2284	2303
Explicit plural	1984	2935
Conceptual (collective)	2481	2774
Explicit plural	2820	3356

As in the previous experiments, we tested specific contrasts. There was no difference in reading time between the singular and plural pronoun sentences in the multiple item (conceptual) texts. There was, however, a tendency for plural pronouns in explicit plural contexts to be read faster than those in conceptual contexts. The difference between the plural pronouns in the explicit and in the conceptual contexts was not significant by subjects, although it was marginally significant by items [$t(15) = 2.02, P = 0.061$].

Generic Types vs Specific Tokens. Once again, target sentences containing plural pronouns were read faster overall [$F(1,20) = 14.96, P < 0.001$; $F(1,15) = 17.03, P < 0.001$], but there was no main effect of conceptual vs explicit initial sentences. As in the multiple vs unique item texts, the plural pronouns in the explicit contexts were particularly easy, and the singular pronouns in these contexts were particularly difficult. There was a highly significant interaction between whether the first sentence contained a conceptual or an explicit plural, and whether there was a singular or plural pronoun in the second [$F(1,20) = 21.51, P < 0.001$; $F(1,15) = 7.36, P < 0.02$].

There was no difference between the singular and plural pronouns following the generic contexts. As for multiple items, there was a tendency for the target sentences following explicit plurals to be read faster than those following conceptual plurals: When we compared the generic (conceptual) plurals with the explicit plurals, there was a marginally significant effect by subjects [$t(23) = 1.98, P = 0.06$], but not by items.

Collective Sets vs Individuals. The analysis of these data gave a completely different pattern of results to the other two. Both of the main effects were significant. Target sentences were read faster following a conceptual plural than following an explicit plural – although this effect was only marginally significant by items [$F(1,20) = 23.74, P < 0.001$; $F(1,15) = 4.00, P < 0.07$] – and faster when they contained a plural rather than a singular pronoun [$F(1,20) = 11.85, P < 0.003$; $F(1,15) = 11.16, P < 0.005$]. The interaction between conceptual vs explicit plurals and type of pronoun in the target sentence did not approach significance ($F(1) = 1.04$; $F(2) = 0.62$). This lack of interaction was because the plural pronoun was easier following *both the explicit and the conceptual cases*. Indeed, the collective set followed by a plural pronoun (the conceptual case) was read faster than the same condition followed by a singular pronoun, although this effect was significant only by subjects [$t(23) = 2.03, P < 0.05$] and not by items ($t = 1.52$). In addition, there was a tendency for the conceptual case followed by a plural pronoun to be read *faster* than the explicit plural, though this contrast was only marginally significant by subjects [$t(23) = 1.87, P = 0.074$] and not by items.

Interpretation of Target Pronouns

The data were scored as described for Experiment 2, except that in this experiment, some responses fell into the N/S category. The data are shown in Table 7. It should be noted here that the explicit plural context followed by a singular pronoun produced a high proportion of N/S responses for all three types of material. Thus, it seems reasonable to assume that, had we provided this response option in Experiment 2, the subjects would have used it and we would have seen a concomitant reduction in the numbers of plural responses in the individual/unique/specific followed by plural pronoun conditions, where the pronouns were highly infelicitous.

Let us turn first to the collective sets. In the conceptual contexts, almost all responses were collectives, and very few were singular interpretations, regardless of whether the pronoun in the text was singular or plural, and there was no significant difference in the number of collective/singular responses following the collective plus singular pronoun or the collective plus plural pronoun (both t s < 0.40). In the explicit cases, in general, there were far fewer collective responses but there tended to be more such responses when the explicit plural was followed by a singular pronoun, presumably because the subjects were trying hard to find an interpretation for these infelicitous pronouns. Notably, there were many more collective/singular responses to plural pronouns following the conceptual context than to plural pronouns following the explicit plural context, and this difference was highly significant [$t(23) = 11.52, P < 0.001$; $t(15) = 9.55, P < 0.001$]. Interestingly, too, more than 11% of the responses in the perfectly legitimate collective plus singular pronoun condition (“IBM . . . it”) were “nothing sensible”, a further indication that subjects prefer plural pronouns in such contexts.

In the case of both multiple items and generics, the plural pronouns in the conceptual contexts resulted in large numbers of plural responses, though not as many as in the explicit plural contexts and, for each type of material, the difference between the explicit and conceptual contexts was significant [multiple items: $t(23) = 3.19, P < 0.004$; $t(15) = 4.70, P < 0.001$; generics: $t(23) = 2.92, P < 0.008$; $t(15) = 4.14, P < 0.001$].

Thus, the pronoun interpretation data from the multiple and generic sentences provide evidence that the explicit plurals are more naturally followed by a plural pronoun than are the conceptual plurals – an indication that the conceptual interpretation of the plural pronoun is not the only one that is considered. As in Experiment 2, and in the reading time data from this experiment, the collective sets behaved rather differently. In the conceptual context, there was an overwhelming tendency to interpret the pronoun – whether singular or plural – as referring to the collective set. There was even some tendency to try to impose a collective reading on the pronouns in the explicit plural contexts.

TABLE 7
Classes of Responses in Pronoun Interpretation Task: Experiment 3 (max = 96)

	Explicit Plurals										Conceptual Plurals									
	Singular Pronoun					Plural Pronoun					Singular Pronoun					Plural Pronoun				
	co.	si.	pl.	N/S	oth.	co.	si.	pl.	N/S	oth.	co.	si.	pl.	N/S	oth.	co.	si.	pl.	N/S	oth.
Collective vs individual	48	1	3	35	9	20	2	67	1	6	80	0	0	11	5	80	2	11	1	2
Multiple vs unique	—	26	27	40	3	—	0	96	0	0	—	90	2	3	1	—	16	78	1	1
Generic vs specific	—	52	11	29	4	—	4	92	0	0	—	86	7	3	0	—	22	74	0	0

Note: Responses = si., singular; pl., plural; oth., other; co., collective; N/S, "nothing sensible".

Discussion

We will consider the result of each type of text in turn. In the multiple item texts, the conceptual plus plural pronoun cases ("A plate . . . them") had a very slight advantage over the conceptual plus singular pronoun cases ("A plate . . . it"), as in the previous experiment. However, the explicit plural followed by a plural pronoun ("Some plates . . . they") was the fastest condition of all, and was more than 200 msec faster than the conceptual plus plural case (though this difference was only marginally significant).

Similarly, in the case of the generics, the conceptual plural ("a horror movie . . . they") was read slightly faster than the conceptual plus singular pronoun, but the explicit plural ("The remakes of Dracula and Frankenstein . . . they") was read 300 msec faster than the conceptual case (though, again, this difference was only marginally significant). In both these cases, then, the plural pronouns were very acceptable in the conceptual plural versions, but there was some evidence that they were not as readily interpreted as plural pronouns following explicitly plural antecedents. It would seem, therefore, that although natural sounding, these types of conceptual pronouns do require some work for their interpretation – multiple entities or generic terms must be derived that allow the plural pronoun to be interpreted when there is no explicit plural antecedent for it to refer to. This conclusion is supported by the data from the pronoun interpretation task: the plural pronouns were less frequently given plural interpretations when they occurred in the conceptual plural, rather than the explicit plural condition. This conclusion might seem slightly at odds with the fact that singular pronouns were generally interpreted *more slowly* than plural ones in the conceptual cases for these two types of material. However, the comparison of singular vs plural pronoun within the conceptual condition is complicated because, although the singular pronouns result in a superficial linguistic match, the passages are slightly anomalous (and are not comparable to the specific/unique plus singular pronoun passages in Experiments 1 and 2). This anomaly arises because the first sentences are naturally interpreted as making non-specific references ("a soap opera", "a plate"), whereas the subsequent definite pronoun is naturally interpreted as referring to a *specific* entity.

The texts containing collectives behaved rather differently. In these texts, the conceptual plurals followed by a plural pronoun ("My sister went to work for IBM . . . they") were easiest overall. They were interpreted more rapidly than either the conceptual singular or the explicit plural texts. These results suggest that collective sets behave in a rather special way. Perhaps because they can *only* be thought of as collections of people, this attribute is immediately brought to mind when they are initially encountered in a text, and explicitly spelling out the collective ("the students in

the class . . .”) does not help processing – in fact, it seems to hinder it slightly. Indeed, Quirk and Greenbaum (1973) state that collectives can be followed by either singular or plural pronouns and verbs: “singular and plural verbs are more or less interchangeable in these contexts, the choice is based, if on anything, on whether the group is being considered as a single undivided body, or as a collection of individuals” (p. 177). These observations concur with the present finding that collectives followed by plural pronouns were very acceptable. Interestingly, Quirk and Greenbaum point out that the use of plural verb forms and pronouns following collective nouns is less common in American than in British English, so the results that we have obtained suggesting that collectives have a special status may be particular to British English (indeed, the data from Gernsbacher’s American subjects *did not* suggest that collective terms behaved differently from the other two classes of conceptual anaphors). The finding that the explicit versions of the collectives (“The students in the class . . . they”) were slightly *harder* than the conceptual plural versions must be treated with caution, however, because in these materials there was some unavoidable flexibility in how the plural pronouns could be interpreted in the explicit cases. *They* in the example above, could be interpreted as “the students” (plural), “the students in the class” (plural) or “the class” (conceptual plural). In practice, the meaning is always the same, but this flexibility of interpretation may have increased processing times in this condition.

In all three types of passage, the explicit plurals followed by a singular pronoun were, as expected, difficult to interpret – the pronoun cannot find an antecedent either at the conceptual or at the explicit (surface) level.

CONCLUSIONS

Our results broadly replicate those of Gernsbacher (1991) with a different population and different tasks: conceptual pronouns are rated as very natural and are readily understood. However, the answer to the question of whether they are *as easy* as plural pronouns with an explicit plural antecedent is not clear cut. In one case – the collective sets – the conceptual pronouns were the most readily interpreted overall, and they even tended to be read *faster* than plural pronouns following explicit plurals. The data from the interpretation task support this bias towards conceptual readings – the conceptual plurals produced a high proportion of collective responses. These data strongly suggest that when collective terms are encountered, their collective nature is encoded in the representation of the text, i.e. they introduce immediately into the mental model a representation of the members of a collection, which would naturally be referred to

using a plural pronoun, rather than just a representation of a collection of members, which would naturally be referred to with a singular pronoun.

In the two other types of materials, the results were less striking – there was a tendency for the conceptual plurals to take longer than the explicit plurals to interpret, suggesting that some inferencing is required. However, in both the multiple items and the generics, the effect was only marginally significant (though, of course, this was a relatively small-scale experiment, with only 24 subjects and 16 materials). Furthermore, this interpretation is supported by the results of the pronoun interpretation task. These results also suggest that a plural antecedent is not so unambiguously available in the case of the generic and multiple materials. Although the category “nothing sensible” was chosen very rarely for the conceptual cases, there was, nevertheless, quite a high proportion of singular responses. The combination of the reaction time results and the interpretation results suggest that subjects initially introduce only individuals into their mental model, and introduce the “corresponding” sets only when they are necessary for the interpretation of a plural pronoun.

Indeed, these results for the multiple items and generic types texts are not surprising when one considers that the initial sentences do not give the subjects any further information about what sort of model they should construct – there is no relevant context for their interpretation. In the “unique item” texts, for instance, it is possible, as we showed earlier, to construct contexts that favour a “multiple item” model (e.g. looking for an iron in a shop). These may not be the contexts that immediately come to mind but, since they are possibilities, subjects may suspend any elaborative processes and wait until they reach the target pronoun before looking for a suitable interpretation. Similarly, in the “generic” texts, as we pointed out, the texts only license a generic reading, and do not *rule out* a specific one (in the sense, for example, that “The lion is a dangerous beast” would). For this reason, the text could continue with a specific reference (“Jimmy went to see a horror movie. It made him have nightmares”). Our data suggest that collective sets behave rather differently from the other two types of materials – they may be automatically thought of as members of a collection, and explicitly mentioning those members does not further aid processing.

We have not, in these experiments, addressed the complex issue of exactly *how* the conceptual pronouns are linked to their antecedents. Our experiments have shown consistently that subjects have little difficulty in understanding conceptual uses of plural pronouns, but we have not considered all the complications that can arise in constructing a suitable discourse referent when a conceptual pronoun is encountered. Some idea of the complexity of the problem is evident from examples such as the following:

He made a considerable mark as Information Technology Minister, a post which ideally suited his interests as one who appreciated and became involved in the whole word processor revolution while he was still in opposition during the 1970s and long before most people had heard of *them*. (*Times Higher Education Supplement*, 30 May 1986)

Here, the pronoun *them* is used to refer to *word processors*. However, not only does the "antecedent trigger" *word processor revolution* fail to provide directly an antecedent that matches in number, but it is also the wrong "part of speech" – *word processor* is used adjectivally, whereas what is needed is a noun phrase. Moreover, there is an intervening potential antecedent that is far better matched syntactically: *the 1970s*. Despite these problems, the text still seems fairly readily comprehensible, but it is by no means straightforward to explain how subjects reject the linguistically matching antecedent (if, indeed, they ever consider it) on pragmatic grounds, and then go on to construct an antecedent from their representation of the preceding text. It is issues such as these that need to be addressed if we are to understand how conceptual pronouns and other forms of linguistically deviant anaphors are able to "find" antecedents.

Manuscript received September 1991
Revised manuscript received February 1992

REFERENCES

- Gernsbacher, M.A. (1991). Comprehending conceptual anaphors. *Language and Cognitive Processes*, 6, 81–105.
- Quirk, R. & Greenbaum, S. (1973). *A university grammar of English*. London: Longman.
- Reinhart, T. (1983). *Anaphora and semantic interpretation*. London: Croom Helm.