

# Short article

# "They" as a gender-unspecified singular pronoun: Eye tracking reveals a processing cost

# Anthony J. Sanford and Ruth Filik

University of Glasgow, Glasgow, UK

The plural pronouns *they* and *them* are used to refer to individuals with unknown gender and when a random allocation of gender is undesirable. Despite this apparently felicitous usage, "singular they/them" should raise processing problems under the theory that pronouns seek gender- and number-matched antecedents. Using eye-tracking, we investigated whether there was any processing cost associated with using singular they/them. There was a clear cost of number incompatibility for they/them. Thus, although singular they/them is in current usage, it does not appear that they/them is immediately tolerant of a plural antecedent, though such may be rapidly accommodated. The data are consistent with the search account of pronoun resolution and preserve the semantics of they/them as denoting plurality.

Referential pronominal anaphors are used to refer back to entities introduced in a discourse. Usually anaphors refer to entities that were recently explicitly mentioned (Erku & Gundel, 1987; Sanford, Garrod, Lucas, & Henderson, 1983), although there are some exceptions (Greene, Gerrig, McKoon, & Ratcliff, 1994; Yule, 1982). Furthermore, mentioned entities normally match the pronoun in number and gender. However, although simple agreement is usual, in vernacular English it is possible to refer to singular individuals using a plural pronoun, as in (1):

 I was looking for the railway station when I saw someone on the other side of the street. I asked them if they knew where it was. In (1), a specific person of unspecified gender is introduced, and referred to by means of the plural pronoun *they*. It has been suggested that *they* is being used in a situation where there is either no wish or no need for the speaker to reveal the gender of the individual concerned (MacKay, 1980; Meyers, 1990). Thus *they* seems to be the ideal candidate for a gender-free singular pronoun (Bodine, 1975; Mackay, 1980; Martyna, 1978). In examples such as (1), *they* and *them* are being used as gender-free singular pronouns, and, intuitively, such usage appears to cause no difficulties.

In a study of this use of *they*, Foertsch and Gernsbacher (1997) showed that when the

Correspondence should be addressed to A. J. Sanford, Department of Psychology, University of Glasgow, Glasgow, G12 8QQ, UK. E-mail: tony@psy.gla.ac.uk

This work was supported by EPSRC Grant GR/S76434/01 and AHRC Grant B/RG/AN8799/APN19525. Materials available from the authors on request.

gender of an antecedent could be recovered as a default, reading times were faster for a gender-matching singular pronoun than for *they*:

2. My nurse was able to explain how my medication would affect me, even though she/they had no say in prescribing it.

It is known that the default gender interpretation of *nurse* is a female, and that anaphoric violations of this default increase processing time (e.g., Carreiras, Garnham, Oakhill, & Cain, 1996), and the Foertsch and Gernsbacher study shows that the processing of *they* under this circumstance was nearly as disadvantaged as a singular pronoun of the wrong gender.

More crucially for the present paper, Foertsch and Gernsbacher (1997) found that when the gender of a specific antecedent is not recoverable by default, using *they* resulted in the same reading times as *he* or *she*:

3. A runner I knew always ate lots of pasta the night before a race, even though he/she/they would've rather had a steak.

This result is consistent with the idea that they can be used as a genderless singular and is consistent with the idea that it carries no processing cost. Models of pronoun resolution that assume a search for an antecedent with a number and gender match (e.g., Sanford, 1985) are compromised by this observation. One possibility is that the pronouns they and them carry the option of being gender-neutral singular as part of specification, so that when an antecedent is sought, a gender-neutral singular is perfectly acceptable from a processing point of view. This is plausible because *they* is now commonly used as a singular pronoun in vernacular English. Of course, were this to be true, it would have to be the case that singular antecedents were restricted to underspecified examples where gender cannot be recovered. (e.g., a person, someone, somebody). This view we call the unspecified-singular-match account: On encountering the pronouns they or them, an explicit antecedent is sought that could be either a genderless singular or a plural entity.

However, there is a second, simpler possibility. It could be that they initiates a search for antecedents, but is not tolerant of singular antecedents, with gender-unspecified singulars being initially unacceptable as potential antecedents. Instead, when the search fails, gender-unspecified singulars have to be later accommodated in some way. It may be the case that the global selfpaced reading paradigm used by Foertsch and Gernsbacher was too coarse a measure to pick up the disruption caused by a mismatch. Were this to be true, the view that they basically initiates a procedure to test for plural antecedents could be maintained. The experiment reported here tests this idea. We compared the processing of singular and plural pronouns as they were related to singular (e.g., a person) or plural (e.g., some people) antecedents, as in (4):

4. Mr Jones was looking for the station. He saw [someone/some people] on the other side of the road, so he crossed over and asked [them/her] politely....

It is uncontentious that with a plural antecedent, there will be a mismatch effect for the singular her. Such mismatches should serve as a baseline against which to evaluate mismatches for the plural them. If the unspecified-singular-match account is correct, then there should be no disruption for the pronoun them. In contrast, if this account is wrong, then the use of them following a singular antecedent should produce a disruption in the tracking record. Foertsch and Gernsbacher were not concerned with this comparison, and did not make it, so we argue that their experiments were not optimal for examining our basic question (though they were successful at showing the conditions under which they might most readily be used as a singular).

# Method

#### **Participants**

A total of 36 native English speakers from the University of Glasgow community, with normal vision, were paid or given course credit to participate.

# Materials and design

The design was 2 (singular antecedent/plural antecedent) × 2 (plural/singular pronoun). A total of 24 experimental items were produced in four versions; an example is:

5. Mr Jones was looking for the station. He saw [someone/some people] on the other side of the road, so he crossed over and/ asked [them/her]<sub>1</sub>/ politely<sub>2</sub>/ where the station was.<sub>3</sub>/ It was in a different part of town./

(Slashes mark analysis boundaries)

The first sentence of each item was a scenesetting sentence. The second sentence introduced either a single person of unknown gender (e.g., a person, someone), or a plural referent (e.g., some people). Where a person was introduced, it was clearly a specific person, and not a generic. The second sentence also included a pronoun, either singular (him/her) or plural (them), which was intended to refer to the earlier mentioned antecedent. The third sentence was introduced so that the test sentence was not the last one. Comprehension questions were provided on 50% of trials, half requiring "yes" and half "no" responses. For (5), the question was Was the station in a different part of town? (yes).

The materials were in four files, with each item appearing in only one of its four versions in a given file. Over the four files, it appeared in all versions. A given file comprised six materials in each of the four conditions. Each file also included 86 filler items of a similar length to the test materials, but otherwise unrelated to the present study. Texts were presented as three or four written lines, with two blank lines between each line of text to aid fixation analysis. The pronoun region was always two words long, the adverb region was always a single word, and both were positioned towards the centre of a line.

#### Procedure

We used a Generation 5.5 Fourward Technologies Dual Purkinje Image eye-tracker (angular

resolution 10 minutes of arc), with binocular viewing and the tracker monitoring the right eye's gaze. Items appeared on a monitor approximately 80 cm from the participant's eyes, with ~4 characters per degree of visual angle. The participant's gaze location was monitored every millisecond, and custom software was used to establish the timing and positions of fixations. Before the start of the experiment, the eye-tracking procedure was explained, and participants were instructed to read normally for comprehension. Participants were seated at the eye-tracker, and a bite-bar and forehead rests were used to minimize head movements. They then completed a calibration procedure. Before the start of each trial, calibration accuracy was checked and was recalibrated if needed. This process ended with participants fixating a box in the upper left half of the screen, at which point the experimenter prompted the computer to present a target text, with the first character of the text replacing the fixation box. Once participants had completed reading each text, they pressed a key, and the computer displayed a question following 50% of total trials. Participants responded by pressing one of two keys (for "yes" and "no") and took breaks as required.

#### Regions

The critical sentence of each material was divided into three regions for analysis, indicated by the slashes in (5). Region 1 contained the pronoun and previous word, Region 2 the adverb following the pronoun, and Region 3 the remaining four words of the sentence.

#### Analysis

An automatic procedure pooled short contiguous fixations. Fixations less than 80 ms were incorporated into larger adjacent fixations within one character, and fixations of less than 40 ms that were not within three characters of another fixation were deleted. Fixations over 1,200 ms were truncated. Prior to analysing the data trials where participants failed to read the sentence or there had been tracker loss were eliminated. We removed those trials where two

or more adjacent regions had zero first-pass reading times, which accounted for 1.04% of the data. We report three measures of early processing: first-pass reading time, the sum of all the fixations made in a region until the point of fixation leaves the region either to the left or the right; and regression path reading time (also known as go-past reading time), the sum of fixations from the time that a region was first entered from the left to the time that the region was first exited to the right. This measure includes fixations made to reinspect earlier portions of text and is usually interpreted as providing an indication of early processing difficulty along with time spent reinspecting the sentence in order to recover from such difficulty. We also report the incidence of first-pass regressions. These are regressions out of a region to the left from the first time that the region is entered from the left. We report one later processing measure, total reading time, which sums the duration of all fixations made within a region and provides a measure of overall comprehension difficulty at this region. Ms/character reading times are also reported for Region 1, in order to correct for length differences at this region. In cases where the region of interest was skipped in first-pass and regression path reading times, the relevant point was excluded from the analysis, and means were calculated from the remaining data points in the design cell.

#### Results

Error rates for comprehension questions were very low, at 6%, indicating that participants read adequately for comprehension.

Eye-tracking data for each region were subjected to two 2 (singular/plural antecedent)  $\times$  2 (singular/plural pronoun) analyses of variance (ANOVAs) treating participants ( $F_1$ ) and items ( $F_2$ ) as random variables. Table 1 shows mean first-pass reading times, first-pass regressions

out, regression path, and total reading times for Regions 1–3.

# First-pass reading time

There was only one result to emerge from this analysis. At the pronoun region, Region 1, following conversion to ms/character, there was a main effect of pronoun, with longer reading times for singular than for plural pronouns,  $F_1(1, 35) = 15.48$ , p < .01;  $F_2(1, 23) = 12.79$ , p < .01. There was no main effect of antecedent type ( $F_8 < 1$ ) and no reliable interaction between antecedent type and pronoun type,  $F_1(1, 35) = 1.04$ ;  $F_2(1, 23) = 2.92$ . There were no significant effects in Regions 2 or 3 ( $F_8 < 2$ ).

## First-pass regressions out

With this measure, evidence for problems in processing was observed when a singular pronoun was used to refer to a plural antecedent. At Region 1, there were no significant effects. At Region 2, the main effect of antecedent type was not significant (Fs < 1.9). However, there was a main effect of pronoun,  $F_1(1, 35) = 7.92, p < .01; F_2(1, 23) =$ 4.58, p < .05, with more regressions out for singular than plural pronouns. More importantly, there was also a significant interaction,  $F_1(1, 35) =$ 6.92, p < .05;  $F_2(1, 23) = 5.34$ , p < .05. Simple effects analyses revealed that for sentences containing singular pronouns, there were more regressions when there was a plural than a singular antecedent,  $F_1(1, 35) = 8.58$ , p < .01;  $F_2(1, 23) =$ 5.98, p < .05. In contrast, for sentences containing plural pronouns, there were no differences (Fs < 1.4). For sentences with singular antecedents, there was no difference in regressions for sentences containing singular pronouns and those containing plural pronouns (Fs < 1). However, for sentences with plural antecedents, there were more regressions for sentences containing singular than for those containing plural pronouns,  $F_1(1, 35) =$ 19.27, p < .01;  $F_2(1, 23) = 7.24$ , p < .05). Thus, in terms of regressions out of Region 2, there is

<sup>&</sup>lt;sup>1</sup> An analysis of skipped trials showed only one systematic effect. In Region 1, there was a main effect of pronoun ( $F_1 = 4.93$ , p < .05;  $F_2 = 4.95$ , p < .05), with singular pronouns being skipped more often (at .68 of trials on average) than plural pronouns (at .40 of trials on average). This numerically small effect is attributable to differences in lengths of these pronouns.

Table 1. Mean first-pass reading times, first-pass regressions out, regression path reading time, and total reading times for Regions 1-3 of the critical sentence, for sentences containing a singular or a plural pronoun, and either a singular or plural antecedent for this pronoun

	Region	Singular antecedent		Plural antecedent	
		Him/Her	Them	Him/Her	Them
First-pass reading time <sup>a</sup>	1	264 (13.2)	296 (10.9)	290 (13.7)	274 (11.7)
	$1^{\mathrm{b}}$	31 (1.2)	29 (1.1)	34 (1.7)	28 (1.0)
	2	290 (8.2)	283 (7.5)	295 (13.0)	282 (10.5)
	3	599 (32.0)	619 (27.1)	642 (22.4)	637 (29.0)
First-pass regressions out	1	8.5	9.3	8.2	10.8
	2	9.4	10.6	18.0	7.1
	3	17.3	10.6	16.4	11.8
Regression path reading time <sup>a</sup>	1	359 (20.9)	382 (20.6)	370 (16.5)	378 (18.5)
	$1^{\mathrm{b}}$	37 (2.2)	35 (1.9)	38 (1.9)	35 (1.7)
	2	342 (17.3)	334 (16.1)	427 (28.7)	326 (23.0)
	3	796 (51.2)	764 (40.8)	906 (67.3)	740 (35.2)
Total reading times <sup>a</sup>	1	340 (20.4)	390 (22.0)	431 (27.5)	331 (14.6)
	$1^{\mathrm{b}}$	35 (2.1)	36 (2.0)	45 (2.8)	31 (1.4)
	2	320 (15.8)	309 (14.2)	326 (20.1)	293 (17.8)
	3	720 (38.6)	715 (30.9)	764 (32.0)	691 (36.2)

Note: Standard errors in parentheses.

evidence that the processing of singular pronouns is adversely affected by following a plural antecedent, but a similar effect is not observed for plural pronouns following a singular antecedent.

#### Regression path

The regression path data conform to the same pattern as the proportion of regressions out. In the pronoun region, Region 1, using raw times, there were no reliable effects (Fs < 1), a result that held after conversion to ms/character (Fs < 2.0). At the postpronoun region, Region 2, the main effect of antecedent was not significant,  $F_1(1, 35) = 2.39$ , p > .05;  $F_2(1, 23) =$ 1.35, p > .05, but there was a main effect of pronoun,  $F_1(1, 35) = 10.16$ , p < .01;  $F_2(1, 23)$ = 5.46, p < .05, with longer reading times for sentences containing singular than those containing plural pronouns. There was also an interaction between pronoun type and antecedent type,  $F_1(1,$ 35) = 6.73, p < .05;  $F_2(1, 23) = 3.93, <math>p$  = .06. Simple effects analyses showed that for sentences with a singular antecedent, there were no differences in reading time between singular and

plural pronoun conditions (Fs < 1). However, for sentences with a plural antecedent, there were longer reading times for the singular pronoun conditions than for plural pronoun conditions,  $F_1(1, 35) = 12.35$ , p < .01;  $F_2(1, 23) = 8.02$ , p < .01. At the end-of-sentence region, Region 3, there was no main effect of antecedent (Fs < 1.4). There was a main effect of pronoun,  $F_1(1, 35) = 7.20$ , p < .05;  $F_2(1, 23) = 7.65$ , p < .05, with longer reading times for sentences containing singular than those containing plural pronouns. There was no interaction (Fs < 2).

Thus, on this measure, as with proportion of regressions out, while processing the singular pronoun appeared to be adversely affected by following a plural antecedent, processing a plural pronoun was not reliably affected following a singular antecedent.

## Total reading times

With this measure, good evidence was found not only for singular pronouns working best following singular antecedents, but also for plural pronouns working best following plural antecedents. At the

<sup>&</sup>lt;sup>a</sup>In ms. <sup>b</sup>In ms/character.

pronoun region, Region 1, raw reading times showed no main effects (Fs < 2.1). However, there was a significant interaction between antecedent type and pronoun type,  $F_1(1, 35) = 24.89$ ,  $p < .01; F_2(1, 23) = 20.81, p < .01.$  The interaction is shown in Figure 1. For sentences containing singular pronouns, there were longer reading times when there was a plural rather than a singular antecedent,  $F_1(1, 35) = 12.07$ , p < .01;  $F_2(1, 23) = 12.69$ , p < .01. In contrast, for sentences containing plural pronouns, there were longer reading times when there was a singular rather than a plural antecedent,  $F_1(1, 35) =$ 11.33, p < .01;  $F_2(1, 23) = 8.74$ , p < .01. Thus number mismatch effects occurred for both singular and plural pronouns. Other analyses confirm this picture. Thus, for sentences with singular antecedents, there were longer reading times for plural than singular pronouns,  $F_1(1, 35) = 7.06$ , p < .05;  $F_2(1, 23) = 5.15$ , p < .05. For sentences with plural antecedents there were longer reading times for singular than plural pronouns,  $F_1(1,$  $(35) = 13.5, p < .01; F_2(1, 23) = 14.55, p < .01.$ A similar pattern occurred when the measure was corrected to ms/character. For the pronoun region there was no main effect of antecedent

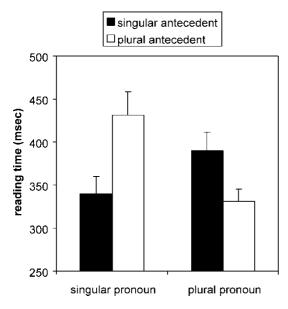


Figure 1. Total reading times at the pronoun region, Region 1.

( $F_s < 1.6$ ). Reading times were longer for singular than for plural pronouns,  $F_1(1, 35) = 13.21$ ,  $\rho <$ .01;  $F_2(1, 23) = 10.61$ , p < .01. There was a significant interaction,  $F_1(1, 35) = 24.50, p < .01;$  $F_2(1, 23) = 21.15, p < .01$ . For sentences with singular pronouns, reading times were longer when there was a plural rather than a singular antecedent,  $F_1(1, 35) = 12.65$ , p < .01;  $F_2(1, 5)$ 23) = 11.95, p < .01. For sentences with plural pronouns, there were longer reading times for sentences with singular than plural antecedents,  $F_1(1,$ 35) = 10.64, p < .01;  $F_2(1, 23) = 8.18$ , p < .01. For sentences with singular antecedents, there was no difference in reading time between singular and plural pronouns ( $F_s < 1$ ). For sentences with plural antecedents there were longer reading times for singular than plural pronouns,  $F_1(1, 35) =$ 26.32, p < .01;  $F_2(1, 23) = 23.32$ , p < .01. At Region 2 there was no main effect of antecedent and no interaction (Fs < 1). The main effect of pronoun was not significant,  $F_1(1, 35) = 3.90, p$ > .05;  $F_2(1, 23) = 4.15$ , p > .05. At Region 3 there was no main effect of antecedent and no interaction (Fs < 1.7). Again, the main effect of pronoun was not robustly significant,  $F_1(1, 35)$  $= 2.72, p > .05; F_2 = 4.61, p < .05.$  There were no effects in other regions.

#### Discussion

While the use of they as a genderless "singular" referential pronoun in certain contexts certainly occurs and does not seem to cause problems of felicity from the point of view of casual observation, some processing difficulties were nevertheless observed in our eye-tracking study. Earlier researchers (Foertsch & Gernsbacher, 1997) found that with neutral, apparently genderless, antecedents like someone, or a runner subsequent clauses referring to that individual by he or she, or they, revealed no reliable disadvantage in the case of using they. On the surface, this might be taken as compatible with the position that there is indeed no processing disadvantage to using they as a genderless singular. In the present experiment, we increased the sensitivity of the design in two ways. First, we used continuous eye-tracking, enabling more subtle measurement of any possible patterns of disruption. Secondly, we compared the effects of genderless referential plural antecedents with that of genderless singular referential antecedents. On total time for the pronoun region, we observed a strong, conventional, number-mismatch effect, such that plural pronouns created less processing disruption in the context of plural antecedents than in the context of genderless singular antecedents.

This result is compatible with the view that after encountering a plural pronoun (they, them), a search is initiated for a plural antecedent in the mental representation of the discourse and not for one that could be either plural or singular. So where does this leave the singular use of they/them? Since it is in common use, we suggest that although it gives rise to a mismatch, it is rapidly accommodated as an acceptable deviation. This is quite unlike the case with singular pronouns in the context of plural antecedents, because these are not in common use and, we claim, do not make sense without making an inference like "he or she refers to just one of the plurality in the antecedent".

An additional finding was that the number-mismatch effect manifests differently in the singular and plural cases. For singulars, it appeared early in processing with more regressions out, and longer regression-path times, as well as in greater total time at Region 1. In the case of plural mismatches, the effect only appeared in total time in Region 1. So singular pronouns showed an earlier sensitivity to number mismatches than did plurals.

Of course, this in no way weakens the finding in total time that there is a mismatch effect for the plural pronoun. However, the difference between the singular and plural cases is noteworthy and is not an isolated case. In the present study, we observed that there were longer first-pass reading times for singulars than plurals, regardless of whether the condition was match or mismatch. This was followed by an earlier response to a mismatch in the singular case. So it is possible that singulars are associated with a greater processing effort at matching a potential antecedent immediately than is the case with plurals. A consistent pattern was observed by Moxey et al. (2004; Exp.

2), who found that under conditions that did not favour plural pronoun usage, the mismatch effect was delayed in comparison to a similar mismatch effect for singular pronouns. So, evidence is emerging that plural pronouns may be more delayed in checking the fit of a potential antecedent than are singulars. We conjecture that such a delay results from the range of antecedent types that plurals can take (e.g., sets, groups, "split" antecedents like Jack and Jill). Support for the idea that encountering he or she may result in more immediate resource-consuming effort to test the fit of an antecedent than does the plural pronoun comes from an event-related potential (ERP) study (Filik, Sanford, Emmott, Morrow, & Leuthold, 2006). On encountering he/she, there was a greater N400 than that to they, suggesting a greater integration effort with he/she.

In sum, although *they* and *them* are used as genderless, singular pronouns and do not appear greatly infelicitous when used under circumstances where the gender of an antecedent is not recoverable, eye-tracking measures reveal a mismatch effect. This offers support for the position that plural pronouns are not in fact mentally represented as hybrid plural–singular pronouns. Rather, the singular use, restricted to special circumstances, requires accommodation following a detection of a true number mismatch by the comprehension system. The present study also adds to a growing literature suggesting that *he/she* and *they* induce different patterns of processing.

Original manuscript received 20 June 2006 Accepted revision received 22 August 2006 First published online 30 October 2006

#### REFERENCES

Bodine, A. (1975). Androcentrism in prescriptive grammar: Singular "they", sex-indefinite "he", and "he" or "she". *Language in Society*, 4, 129–146.

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.

- Erku, F., & Gundel, J. K. (1987). The pragmatics of indirect anaphors. In J. Verschueren & M. Bertuccelli-Papi (Eds.), *The pragmatic perspective: Selected papers from the 1985 international pragmatics conference* (pp. 533–545). Amsterdam: John Benjamin.
- Filik, R., Sanford, A. J., Emmott, C., Morrow, L., & Leuthold, H. (2006, March). Pronouns without antecedents: The processing cost of "Institutional They". Poster presented at CUNY sentence processing conference, New York, USA.
- Foertsch, J., & Gernsbacher, M. A. (1997). In search of gender neutrality: Is singular They a cognitively efficient substitute for generic He? *Psychological Science*, 8, 106–111.
- Greene, S. B., Gerrig, R. J., McKoon, G., & Ratcliff, R. (1994). Unheralded pronouns and management by common ground. *Journal of Memory and Language*, 33, 511–526.
- MacKay, D. G. (1980). On the goals, principles and procedures for prescriptive grammar. *Language in Society*, *9*, 349–367.

- Martyna, W. (1978). What does "he" mean? *Journal of Communication*, 28, 131–138.
- Meyers, M. (1990). Current generic pronoun usage: An empirical study. *American Speech*, 65, 228–237.
- Moxey, L. M., Sanford, A. J., Sturt, P., & Morrow, L. I. (2004). Constraints on the formation of plural reference objects: The influence of role, conjunction, and type of description. *Journal of Memory and Language*, 51, 346–364.
- Sanford, A. J. (1985). Aspects of pronoun interpretation: Evaluation of search formulations of inference. In G. Rickheit & H. Strohner (Eds.), *Advances in psychology, 29: Inferences in text processing*. Amsterdam: North Holland.
- Sanford, A. J., Garrod, S. C., Lucas, A., & Henderson, R. J. (1983). Pronouns without explicit antecedents? *Journal of Semantics*, 2, 303–318.
- Yule, G. (1982). Interpreting anaphora without identifying reference. *Journal of Semantics*, 1, 315–322.

Copyright of Quarterly Journal of Experimental Psychology is the property of Psychology Press (UK) and its content may not be copied or emailed to multiple sites or posted to a listsery without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.