

Anaphoric reference to mereological entities
Derya Çokal*, Ruth Filik, Patrick Sturt, Massimo Poesio

Dr. Derya Çokal
Department of German Language and Literature I–Linguistics, University of Cologne, Germany
Email: dcokal@uni-koeln.de
<https://orcid.org/0000-0002-5653-1412>

Dr. Ruth Filik
Department of Psychology, The University of Nottingham, Nottingham, United Kingdom
Email: ruth.filik@nottingham.ac.uk
<https://orcid.org/0000-0001-9022-8364>

Dr. Patrick Sturt
Department of Psychology, The University of Edinburgh, Edinburgh, United Kingdom
E-mail: patrick.sturt@ed.ac.uk
<https://orcid.org/0000-0002-2055-6933>

Prof. Dr. Massimo Poesio
School of Electronic Engineering and Computer Science, Queen Mary University of London,
London, United Kingdom
E-mail: m.poesio@qmul.ac.uk
<https://orcid.org/0000-0001-8469-2072>

*corresponding author

Anaphoric reference to mereological entities

Abstract

Corpus evidence suggests that in contexts in which the presence of multiple antecedents might favour plural reference, the disadvantage observed for singular reference may disappear if the potential antecedents are combined in a group-like plural entity. We examined the relative salience of antecedents in conditions where the context either made a group interpretation available (i.e., mereological entity) (e.g., *The engineer hooked up the engine to the boxcar...*, where group = *train*), or not (e.g., *The engineer detached the engine from the boxcar...*). Results from three experiments in which participants were asked to identify referents for singular versus plural pronouns (Experiment 1), to confirm the referents of pronouns in a sentence completion task (Experiment 2), and to provide paraphrases for given texts (Experiment 3), collectively provided evidence that the creation of a group makes that entity (i) a possible referent for singular anaphoric reference and (ii) more salient than its constituents.

Key words: plural pronouns, plural object reference, atomic interpretation, complex anaphors, groups, underspecification, singular pronoun

Anaphoric expressions are pervasive in spoken and written language and play a crucial role in the establishment of discourse coherence. Experimental studies in psycholinguistics have yielded useful information about the processing of a range of anaphor types, including pronouns like “he”, definite descriptions like “the girl”, and demonstratives like “this” (Çokal & Poesio, 2022; Kaiser & Fedele, 2019; Poesio et al., 2016). Such studies are informative about the processes of resolving an anaphor and the mental representation of the antecedent.

Most studies on anaphora, however, have focused on the relatively simple situation in which an anaphor refers to an atomic (i.e., singular) discourse entity, such as a named individual (e.g., “John”) or a referent introduced with a singular noun phrase (e.g., “the actor”). There has also been some research on plural discourse entities, in particular when evoked by a conjoined phrase such as “John and Mary” (Albrecht & Clifton, 1998; Garrod & Sanford, 1982; Gordon et al., 1999; Moxey et al., 2004). The similar situation also applies in other areas of cognitive science—e.g., computational linguistics—where anaphor resolution algorithms are largely limited to simple cases involving a singular reference (Poesio et al., 2023; Yu et al., 2021).

However, natural spoken and written language includes a wide range of more complex types of reference, which have been less explored in psycholinguistics. One of these complex types of reference is to discourse entities that we will call *mereological entities*. The term *mereology* is used in philosophy of language and in semantics to refer to the study of the relations between a whole and its parts and vice versa (Varzi, 2019). In this paper, we use the term mereological entities to refer to the special case of plural entities where the constituents (e.g., “engine and goods carriage” or “ham and bread”) are individual parts of a whole object, which has its own name (“train” and “sandwich”, respectively). Corpus evidence suggests that such mereological entities can become available to readers when a joining verb (e.g., “hook up”) is used to connect

the constituents, as in “the engine was hooked up to the goods carriage” (see below for further explanation of the definition and previous findings on mereological entities). The main issues raised by this type of discourse entity include: (a) when are such mereological entities available antecedents; (b) whether sums/individual parts are more salient than mereological entities/groups; (c) how different types of anaphoric expressions play a role in such contexts; and (d) whether anaphoric plural reference is available when there is no syntactic plural unit available in the previous discourse. To be able to address such issues, we need to widen our domain of enquiry to achieve a full understanding of the representations and processes involved in reference.

A small number of psycholinguistic studies have explored the processing of plural reference (e.g., Arnold et al., 2021; Albrecht & Clifton 1998; Çokal & Sturt, 2017; Garrod & Sanford, 1982; Gordon et al., 1999; Koh & Clifton, 2002; Moxey et al., 2004; Sanford & Lockhart, 1990). The aim of such studies is to examine how groups of individuals are mentally represented, and how linguistic input influences these groupings. It has been claimed that plural pronouns require a *plural entity* to be present in a discourse model for reference to be possible (Eschenbach et al., 1989). Consider (1) for example:

1. John and Mary were baking a cake. They/He wanted to finish before the party.

It is a very robust finding that *he* is more difficult to process than *they* in contexts such as (1), a phenomenon which has been called *conjunction cost* (see Albrecht & Clifton, 1998; Garrod & Sanford, 1982; Gordon et al., 1999; Moxey et al., 2004). This finding has been explained in terms of plural entities: If the use of a conjoined noun phrase leads to the formation of a plural entity in the discourse model, this will facilitate the subsequent use of a plural pronoun. A singular pronoun will be more difficult to process, either due to the difficulty of individuating the

two relevant distinct atomic individuals (Moxey et al, 2004; Sanford & Moxey, 1995), or because the plural entity becomes more salient than each atomic entity (Gordon et al., 1999). Another factor that is claimed to influence the formation of a plural entity is *role mapping* (c.f., Gelormini-Lezama, & Almor, 2014; Moxey et al., 2011; Moxey et al., 2004; Sanford & Moxey, 1995). If two protagonists play a common role in an action or episode, this constitutes grounds for using a plural pronoun to refer to those individuals as they function as a plural entity. Another factor that affects the creation of complex antecedent with plural pronoun is previous discourse. Recent studies tested the effect of quantifiers on the interpretation of plural pronouns (Filik et al., 2021; Heinat, & Klingvall, 2020; Ingram & Ferguson, 2018; Upadhyay et al., 2019).

In addition to the conjunction cost for singular pronouns, there is some evidence that when the use of a conjunction leads to the formation of a plural entity, as in (1), plural pronouns are processed more readily than when no plural entity has already been formed, as in (2).

2. John was baking a cake with Mary. They/He wanted to finish before the party.

Presumably this is because in this case a plural entity would need to be constructed “on the fly” at the point when the plural pronoun is processed (Moxey et al., 2004, but also see Clifton & Ferreira, 1987; Eschenbach et al, 1989; Hielscher & Müssler, 1990; Kaup et al, 2002). The contrast between conjunction cases (where a plural entity is introduced by a single [albeit complex] noun phrase) and cases where no syntactic unit corresponds to the plural referent needs to be investigated empirically. The novel contribution of the current study is an empirical investigation of an example of this latter case in which the plural antecedent is not introduced by a single syntactic constituent.

A relatively straightforward way to semantically interpret plural entities is to extend the domain of interpretation to include *plural individuals*. A well-known version of this approach proposed in theoretical semantics is the theory developed by Link (1983), who proposed a lattice theoretic treatment of plurality in which both plural and singular individuals are defined within a universe of interpretation that models both atoms and their sums. *John* and *Mary* are *atomic individuals (atoms)*, which do not dominate other individuals in the lattice. The object $John \cup Mary$, a plural individual, dominates the two atomic individuals. In Link's terminology, this plural individual is called the *sum* of *John* and *Mary*.¹ A number of equivalent versions of this treatment have been proposed which differ from Link's proposal in how the lattice is defined (see, e.g., Landman, 1996; Winter & Scha, 2019).

However, there is also evidence that the picture is more complex than suggested by the experiments on the conjunction cost. In this paper, we explore the contrast between individual and plural references to plural entities in the contexts studied by Poesio and Reyle (2001) and Poesio et al. (2006). Poesio and Reyle (2001) analysed task-oriented conversations from the TRAINS corpus collected at the University of Rochester and found a number of cases of anaphoric reference in which the following occurred: (i) more than one potential antecedent of an anaphoric expression matched it in gender and number, and (ii) no focusing principle of which the authors were aware made one of the interpretations preferred, yet (iii) the recipient of the utterance seemed able to proceed without signalling a problem. The assumption here is that in task-oriented conversations participants need to signal when they do not understand something (Garrod & Anderson, 1987; Micklos et al., 2020). Crucially for this paper, many such examples

¹ We are using in this paper a slightly different notation from Link's to avoid confusion with the notation for mereological entities used in Poesio and Reyle (2001). Link uses the \oplus symbol to indicate sum, so in his notation the sum of John and Mary would be indicated as $John \oplus Mary$.

involved singular pronouns used in the context of entities that had been joined together, to form what Poesio et al. (2006) called “*complex mereological objects*” and we are going to call here *mereological entities*. Consider the sample dialogue below:

- 3.1 M: can we...kindly hook up
3.2 : uh
3.3 : engine E2 to the boxcar at...Elmira
4.1 S: ok
5.1 M: and send *it* to Corning
5.2 : as soon as possible please

It is ambiguous whether the pronoun *it* used by speaker M in line 5.1 refers to the engine, the boxcar, both, or is unspecified; yet somehow this does not seem to matter to the participant marked as S. This raises several questions. A preliminary question is whether this is indeed the case (i.e., whether indeed there is no ambiguity cost in such contexts). This question was addressed by Poesio and Reyle (2001), as discussed below. A subsequent question would be: “Why is a singular pronoun used in a context where it is apparently ambiguous and where an introduced plural entity does not cause any comprehension difficulty for S?” Might this perhaps be because if the engine and the boxcar have been joined together, it does not matter how *it* is interpreted, since both the engine and the boxcar will be moved to Corning? The answer to this second question proposed by Poesio and Reyle (2001) is presented next, and possible issues with this answer are raised after that. However, in a study with a visual-world paradigm, Brown-Schmidt, et al. (2005) showed that demonstrative reference with *that* to the newly formed composite was more likely than singular anaphoric reference (e.g., “Put the cup on the saucer. Now put *it/that* next to the candle”). This leads us to a third list of general questions: (a) How are such pronouns interpreted? (b) Has a particular interpretation been chosen, and if so, which one? and (c) Is the pronoun’s interpretation underspecified in some way? In this paper, we

sought to answer this third set of general questions, which remains unanswered within the context of anaphoric plural and singular reference.

Poesio and Reyle (2001) proposed an *underspecification*-based account of their findings, according to which it is sometimes possible for the referent of a pronoun like *it* to remain undetermined in contexts such as the one above. Specifically, Poesio and Reyle (2001) hypothesized that the referent(s) of the pronoun *it* may be underspecified, resulting in ambiguous anaphoric expressions that are not perceived as infelicitous, in cases where:

1. Two potential antecedents are explicitly mentioned, but these are elements of an underlying mereological individual—an individual which has constituents like a sum, but is related to its constituents by a stronger, *part-of* relation in addition to an *element-of* relation, for which we will use here the notation $a \oplus b^2$ —that has been explicitly constructed (and made salient) in the dialogue. In the example above, the mereological individual would be the train formed by the engine \oplus the boxcar. As already mentioned, we will use here the term *mereological entity* to refer to the representation in the discourse model of such mereological individuals.
2. The presence in the discourse model of this mereological entity makes an *underspecified interpretation* possible in which the anaphoric expression is interpreted as denoting an underspecified entity z . This entity (z) is a disjunctive underspecified representation that ‘covers’ all possible interpretations (that is, the engine (e), the boxcar (b), and the mereological entity formed by combining the two, i.e., the train ($e \oplus b$)).

² As already mentioned above, Poesio and Reyle (2001) use the \oplus symbol differently from Link (1983): Link used the symbol to denote sums, whereas Poesio and Reyle use it to denote mereological entities.

3. All possible interpretations (z, e, b, e \oplus b) are equivalent for the purposes of the plan (i.e., in this case after the two explicitly mentioned potential antecedents are joined, if one of them is moved the other one must be moved as well).

Using an offline *Magnitude Estimation* (ME) task (Bard et al., 1996), Poesio and Reyle (2001) tested the hypothesis that pronouns satisfying these conditions are not subject to ambiguity penalties. In a ME task, participants are asked to assign a magnitude (an arbitrary number) to a reference sentence, and then judge the acceptability of other sentences relative to the reference magnitude. In Poesio and Reyle's (2001) study, participants judged whether sentences such as (3a), in which the engine and the boxcar are joined together (that we will call from now on *mereology constructing contexts*), are 'more acceptable' than sentences such as (3b), in which the engine and the boxcar are not joined (*non-mereological contexts*):

3.
 - a. The engineer hooked up the engine to the boxcar and sent it to London.
 - b. The engineer separated the engine from the boxcar and sent it to London.

Poesio and Reyle (2001) found that mereology contexts were significantly more acceptable than non-mereology contexts. They interpreted this difference in acceptability in terms of perceived ambiguity; that is, they assumed that non-mereological context (3b) was perceived to be unacceptable because it was ambiguous, while no ambiguity was perceived in the mereological context (3a). As a control, participants were also asked to judge the acceptability of the initial parts of the sentences (see [4]), to ensure that any differences in acceptability were not due to features of the sentences before the anaphoric section:

4.
 - a. The engineer hooked up the engine to the boxcar.
 - b. The engineer separated the engine from the boxcar.

Poesio and Reyle found that shorter control sentences (4a) and (4b) did not differ in acceptability, suggesting that the differences between (3a) and (3b) were indeed due to the sentences' anaphoric content. Thus, since (3a) was not perceived as ambiguous, Poesio and Reyle (2001) interpreted these findings as providing evidence for underspecification.

Poesio and Reyle's (2001) offline study left several questions unanswered, however. One such question is that there are potentially other explanations of the difference in acceptability between (3a) and (3b), besides readers not committing to an interpretation. One alternative explanation is suggested by theories about the semantics of plurals in which a distinction is made between two types of plural individuals: *sums* and *groups* (e.g., Barker, 1992; Champollion, 2020; Landman, 1989a, 1989b; 1996, Link, 1984; Winter and Scha, 2019). According to such theories, if Committee A currently consists of *John* and *Mary*, there are (at least) two interpretations of the sentence *Committee A is old*. According to the *sum* interpretation ($\text{John} \cup \text{Mary}$),³ the sentence tells us something about John and Mary, namely that they are old. According to the *group* interpretation, indicated by Landman as $\uparrow(\text{John} \cup \text{Mary})$, the sentence tells us that *Committee A* was formed a long time ago. In other words, the properties predicated of the *sum* interpretation of a plural term such as *committee* apply to its members distributively, whereas the properties ascribed to a plural term under the *group* interpretation are predicated of the plural individual (the committee) but may not hold of its members (Barker, 1992).

The uses of singular anaphoric reference in plural contexts studied by Poesio et al. (2001) could also be explained in terms of this distinction between groups and sums. Consider the slightly simplified (5):

³ A reminder that in this paper we use the notation $a \cup b$ to indicate the sum of a and b, reserving the symbol \oplus to indicate mereological entities.

5. The engineer hooked up the engine and the boxcar, and sent it/them to the station.

Sentence (5) introduces two noun phrases in the conjoined phrase *the engine and the boxcar*. Based on previous work (Moxey et al., 2004), we expect this construction to encourage the formation of a plural entity with two constituent atomic entities. A subsequent plural pronoun *them* will preferentially lead to a sum interpretation for this plural entity, $engine \cup boxcar$. The interesting case here is the pronoun *it*. In the context of (5), *it* could potentially refer to either of the two conjuncts *the engine* or *the boxcar*. However, because of the semantics of *hooked up*, it could also refer to a train, the mereological entity $engine \oplus boxcar$ of which both engine and boxcar are parts. The key point is that this latter interpretation of the referent of *it* as a single mereological entity, contrasts with the interpretation of a plural pronoun as a sum in a way similar to the *group/sum* contrast discussed above. Intuitively, the *train* ($engine \oplus boxcar$) interpretation of *it* appears to involve a single individual, the train, that is more than the sum of its parts, just like references to groups like “the committee”. In contrast, the interpretation of *them* involves reference to the sum of its two constituent entities, $engine \cup boxcar$. So, it is possible that what licenses the use of singular *it* in the contexts studied by Poesio et al. (2001) is that readers/writers made a strong commitment to an interpretation of *it* referring to the train in (3a), while this interpretation was not available for (3b) because of the lack of a mereology-forming predicate. The relative unacceptability of (3b) might thus have been due to the difficulty of choosing between two atomic entities that are roughly equal in salience (i.e., *the engine* and *the boxcar*), while the relative acceptability of (3a) might have been due to the ease of referring to the salient (although unmentioned) train. This alternative explanation differs from Poesio and Reyle’s explanation because they hypothesize that the interpretation of the pronoun is left underspecified between x , y or the mereological entity $x \oplus y$, whereas the alternative group explanation just discussed assumes

a fully specified commitment to one alternative, the group / mereological entity $x \oplus y$. Therefore, it is important to establish the interpretations that people are making in what we are going to call *mereological contexts*.

We now repeat the questions that are the foci of this paper: “How are such singular and plural pronouns interpreted? Has a particular interpretation been chosen (and if so, which one?), or is the singular pronoun’s interpretation underspecified in some way?”. To answer these questions, we probe the mental representation of mereological individuals in mereological contexts such as:

6. The engineer hooked up the engine to the boxcar and sent it/them to the station.

In this sentence a mereological (i.e., part-whole) relationship is established in the context clause between two or more constituent atomic entities e and b and a mereological entity ($e \oplus b$). In this paper we investigated the following questions regarding how *it* is ultimately interpreted in mereological contexts:

1. Are the constituent entities of the mereological entity accessible for reference? Or is only the mereological entity (the train) accessible?
2. If the latter is preferred, is the interpretation sum-like, or group-like?
3. Finally, how do these part-whole relationships interact with what we already know about mereological entities and the processing of anaphors?

We report three experiments. Experiment 1 was a forced-choice task and investigated whether when the two atomic entities were combined to form a mereological entity, which referents (i.e., constituents vs. plural/mereological entities) were identified as the referents of *it* and *them*.

Experiments 2 and 3 probed readers’ interpretations using a sentence completion task and a paraphrase task.

Experiment 1

In Experiment 1 we asked participants to read a sentence introducing a pair of objects in a mereological or non-mereological context, and then asked them to (1) provide, if possible, a name for the pair of objects and (2) choose an interpretation of the pronoun *it* or *them* from a list of possibilities. The Queen Mary Ethics of Research Committee at the Queen Mary University of London approved the current study (QMREC2327a).

Method.

Participants. Participants were recruited via Mturk. They ($n = 40$) were native English-speakers living in the United States who were not proficient in other languages. To learn whether participants were proficient in other languages, we asked questions at the end of the experiment (i.e., Do you speak other languages in addition to English? If yes, please specify. When did you learn these languages? Could you please rate your proficiency in each language?). Five participants who identified themselves as “very proficient/proficient” in another language or who stated they learned another language when they were growing up, were excluded from the database.

Additional requirements for recruitment were a US Graduate Degree and a HIT Approval Rate (HIT) greater than 99%. HIT represents the proportion of completed tasks approved by Requesters. This means that only workers who consistently produced high quality tasks were able to access our experiment. In addition, they were not allowed to participate in the current study more than once.

Four participants were excluded from our database because they skipped 30% or more items. All participants were paid after they submitted their responses. Participant’s rejection/data exclusion was conducted after each participant submitted their responses.

Materials. There were 40 experimental items, 60 fillers and eight practice items. All sentences were of similar length. In filler items, we used plural pronouns. Texts were presented in 3 or 4 lines. Each participant saw all the fillers and practice items. They saw only one condition of each experimental item.

This experiment utilized a 2×2 within-subject design, crossing mereology (mereology/joined verb and non-mereology/disjoined verb) with anaphora (*it* and *them*). The four experimental conditions are illustrated in Example (7) below:

7a. Pronoun *it* in the mereology/joined verb condition:

On Sunday afternoon, there were many delays at London King's Cross station. The railway man hooked up the engine to the goods wagon and sent it quickly to the central station. He hoped things would improve soon.

7b. Pronoun *them* in the mereology/joined verb condition:

On Sunday afternoon, there were many delays at London King's Cross station. The railway man hooked up the engine to the goods wagon and sent them quickly to the central station. He hoped things would improve soon.

7c. Pronoun *it* in the non-mereology/disjoined verb condition:

On Sunday afternoon, there were many delays at London King's Cross station. The railway man detached the engine from the goods wagon and sent it quickly to the central station. He hoped things would improve soon.

7d. Pronoun *them* in the non-mereology/disjoined verb condition:

On Sunday afternoon, there were many delays at London King's Cross station. The railway man detached the engine from the goods wagon and sent them quickly to the central station. He hoped things would improve soon.

Predictions. There are three potential referents in the mereology condition with *it* (7a). Firstly, *it* could refer to one of the two atomic entities: the engine or the goods wagon. However, the semantics of the verb ‘hook up’ could signal that those two parts (i.e., ‘engine’ and ‘goods wagon’) were joined and thus the mereological entity ‘train’ was created. In this instance, participants could prefer either one of the separate parts of a train and/or the group interpretation of the mereological entity (the train) as an antecedent of *it*. Alternatively, participants might prefer the sum interpretation (i.e., the plural entity ‘engine + goods wagon’) and thus prefer a plural pronoun reference to this sum interpretation.

In contrast, due to the semantics of the verb ‘detaching’ in the non-mereology contexts (as in 7c), the group interpretation (i.e., train) would not be salient, and participants would prefer to refer to the atomic entities (the engine or the goods wagon) with *it*. In (7d), participants would prefer to refer to the sum interpretation with *them* rather than an atomic entity reference.

The following is an example for our filler items:

8. The Fusion Music Festival line-up was extremely eclectic. The bands had rehearsed consistently. The builders had extensively checked the stage. They had prepared carefully and were very happy to be part of the festival.

Procedure. Using Amazon Mechanical Turk, participants were firstly asked to consent at the beginning of the online task. The consent form included a description of the study procedure, risks, benefits of participation, confidentiality of records, and information about voluntary participation. After the participants consented, the task instructions appeared on the screen. The instructions on the screen were as follows:

“Below is a series of descriptions in which two objects are combined in some way. Your task is to indicate whether the resultant object is a recognisable, everyday object and if

so, to indicate whether this resultant object has a name. Please write the name of the whole object. If the resultant object formed a conventional whole that had no obvious name, please write ‘has no obvious name’. After each item, please press the next button. In the next stage, you need to identify what *it* or *them* means for you. We have made this stage easier for you and provided some options. Please only choose one answer. You will have two practice items. When you are ready, you can start”.

As seen, the participants were asked to do the following: (a) write the name of the whole object formed when two objects were joined, and (b) identify and write the referents of *it* and *them* for each item: parts (i.e., atomic), whole object (i.e., group), two objects (i.e., sum), and not sure. A reference to parts of an object indicates a reference to an atomic entity (e.g., a reference to either an engine or a goods wagon). To provide interpretations for item (b), we provided a drop-down menu from which participants selected their response as in the following:

(a) Engine (b) Goods wagon (c) Train (d) Engine and goods wagon (e) Not sure

In all experiments, participants read instructions, requirements for participation, informed consent, and compensation information prior to beginning the experiment.

Data analysis.

We conducted two logistic mixed effects regressions, incorporating all fixed effects and interactions in a single step (Bates et al., 2015; Barr et al., 2013, Jaeger, 2008). We used the following packages: lme4 to run logistic mixed effects regressions models; Sjplot to calculate odds ratios, random effects; library (emmeans) to calculate standard errors and confidence Intervals; ggplot2 to graph estimate proportions of each condition. Below we ran the model LME4 and reported odd ratios.⁴

⁴ We ran the packages on RStudio 2021.09.1 Build 372. R version was 4.1.2 and lme version 1.1.27.1

In the first model (model 1), we compared participants' unsure responses for the referents of *it* and *them* across mereology and non-mereology conditions. These unsure responses were coded as *other* in the second and third models, and were excluded from the data analyses. In the second model (model 2), we compared referent choices for an atomic entity with a second category that combined sum (two object) and group (whole object) references in a single response category. In the third model (model 3), we excluded atomic entity references and separated sum and group references into two categories. On the drop-down menu that participants selected, the sum references correspond to conjoined two NPs (i.e., engine and goods wagon), whereas group references correspond to the mereological entity references (i.e., train). To separate these two categories, we followed participants' responses on the drop-down menu.

Factor labels were transformed into numerical values and centered prior to analysis with a mean of zero and a range of 1. In model 1 we coded the factor "pronoun" in a way that an Odds ratio > 1 indicates more unsure cases for factor level *it* and fewer for *them*. Factor "verb" was coded so that an Odds ratio > 1 indicates more unsure cases for a factor-level mereology verb and fewer for a non-mereology verb. In model 2, we coded the factor "pronoun" in a way that an Odds ratio > 1 indicates more group/sum references for factor-level *them* and fewer for *it*. Factor "verb" was coded so that an Odds ratio > 1 indicates more group/sum references for a factor-level mereology verb and fewer for a non-mereology verb. In model 3 we coded the factor "pronoun" in a way that an Odds ratio > 1 indicates more group references for factor level *it* and fewer for *them*. Factor "verb" was coded so that an Odds ratio > 1 indicates more group references for factor-level mereology verb and fewer for non-mereology verb. Random slope parameters corresponding to the two experimental factors and their interactions were included in the maximal model for both participants and items. All analyses reported below incorporated

crossed random intercepts for participants and items. For cases where the full model did not converge, we reduced the random effect structure until convergence was reached. The resulting model for unsure responses (model 1) is as follows: (pronoun + verb type + 1|Participant) + (pronoun + verb + 1|Item) + pronoun * verb type). The resulting model for model 2 and model 3 is as follows: (pronoun + verb type + 1||Participant) + (pronoun + verb type + 1||Item) + pronoun * verb type). The same model structure was used for all three analyses, with different choices of data.

The standard errors and z-values for each fixed effect and interaction were reported in the Supplementary materials as well. Data, scripts, and stimuli for all experiments are available at https://osf.io/ua4hc/?view_only=e54c08faa2114c4eaa0927ab364c2ffe.

Results.

For the question in which participants could write the name of the mereological entity/group (i.e., *train*), there was an average agreement of 88% across each set of 40 items. In 16 % of the cases, participants said they were not sure of the referents of *it* and *them* (percentages of unsure cases: mereology: *it* = 7%, *them* = 3%; non-mereology: *it*: 2%, *them*: 4%). In the first model (model I), there was a main effect of pronoun, with more unsure referents for *it* than *them*, $OR = 67.1$, $SE = 24.06$, $p < .001$. There was a main effect of verb, with more unsure references in the mereology verbs than the non-mereology verbs, $OR = 0.23$, $SE = 0.08$, $p < .001$. The interaction between verb type and pronoun was not significant, $OR = 0.38$, $SE = 0.20$, $p = .069$. These results indicate that in the mereology context *it* could refer both to one of the atomic constituent entities and/or to the mereological entity / group, and therefore we had higher percentages of unsure referents with *it* than *them*, $it = 9\%$, $them = 6\%$.

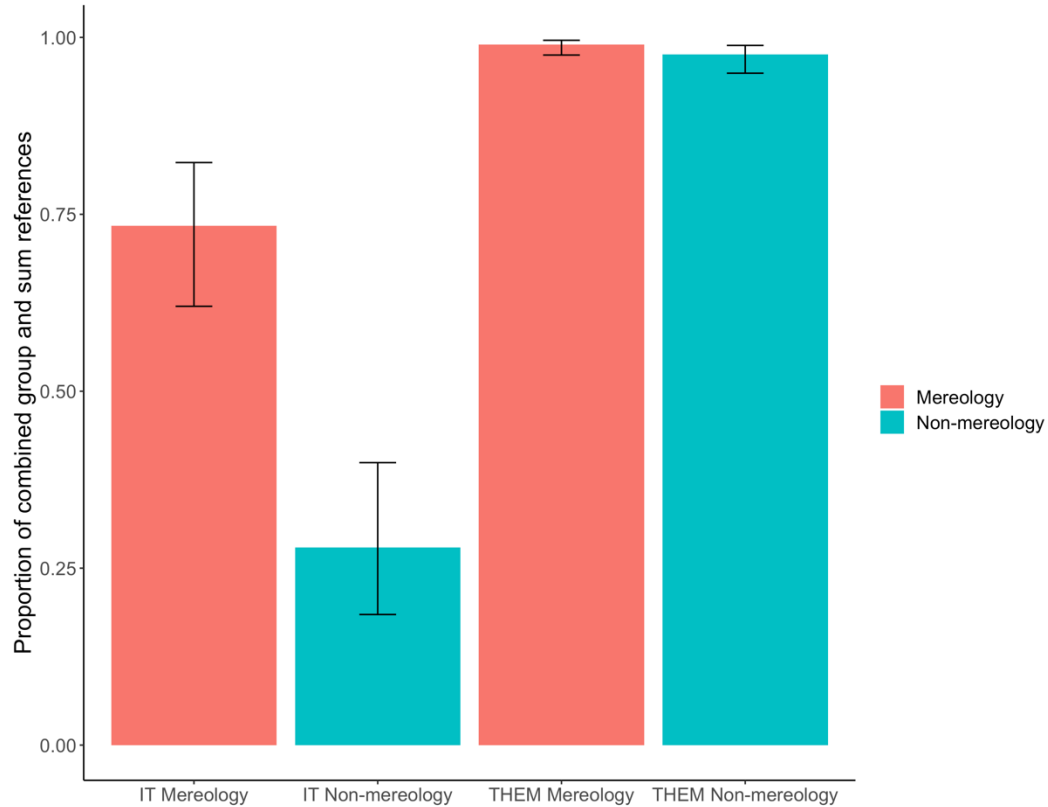


Figure 1. Estimated proportions of combined group and sum references out of all references (group/sum references and atomic entity references) in mereology and non-mereology contexts with *it* and *them*. Error bars show the 95% confidence intervals.

The second model, model 2, compared referent choices for an atomic entity with a second category that combined group and sum references into a single response category (the resulting estimated proportions are charted in Figure 1) (see Tables S1 & S2). There was a main effect of pronoun, such that the sum/group reference was chosen more often with the plural than with the singular pronoun, $OR = 59.10$, $SE = 21.73$, $p < .001$. There was also a main effect of verb type, with more sum/group interpretations for the mereology verbs than the non-mereology verbs, $OR = 4.19$, $SE = 1.29$, $p < .001$. The interaction between pronoun and verb types was significant, $OR = 0.34$, $SE = 0.16$, $p = .024$, such that, for the plural pronouns, the preference for sum/group reference over atomic reference was similar across verb-types, while for the singular pronouns,

this preference was reversed across verb-types; this is clearly visualized in Figure 1. References to sums/groups with *it* in the mereology context were higher than those with *it* in non-mereology context, $OR = 7.36, SE = 2.27, p < .001$.

Our results for the non-mereology condition with *it* suggest that the group interpretation (i.e., train) was not strongly observed in this condition, presumably due to the semantics of the verb ‘detaching’ (as in 7c). The results with the non-mereology condition with *them* (as in 7d) suggest that since the non-mereology verb phrase ‘detach the engine from goods wagon’ highlights two separate (atomic) entities, a plural reference with *them* was preferred to one with *it*.

In our third model (model 3), we excluded the atomic entity responses, and separated the sum and group references into two categories. We coded the sum references as 0 and group references as 1 (see Figure 2, Tables S1 and S2):

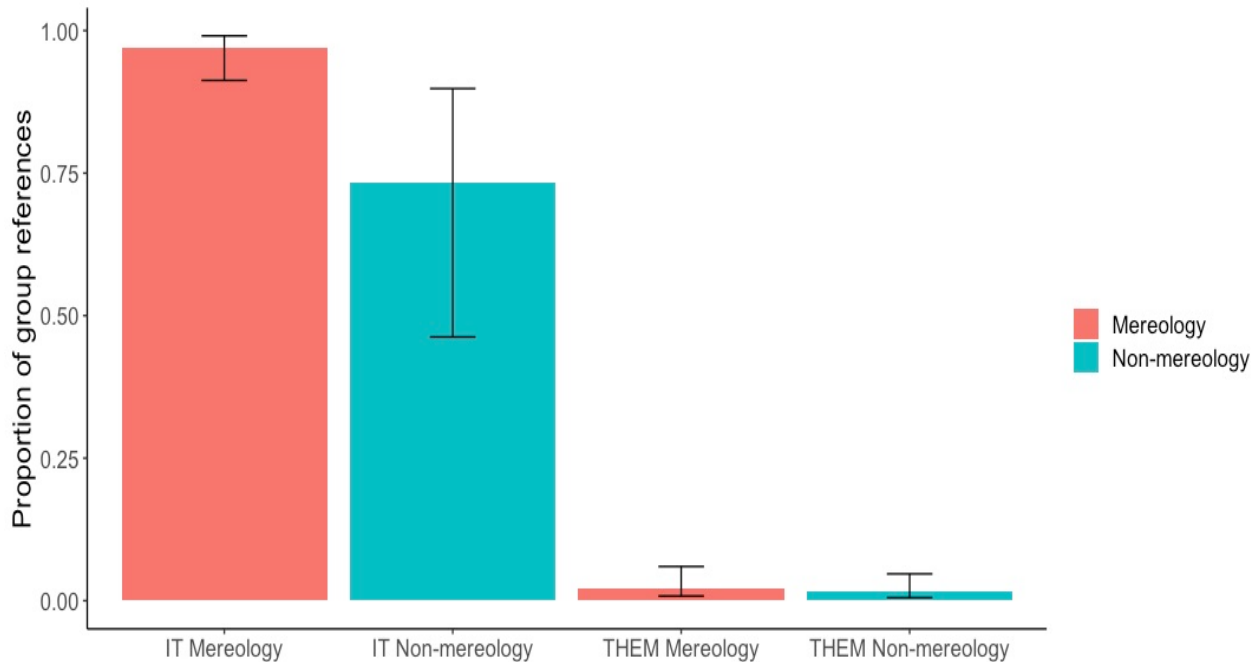


Figure 2. Estimated proportions of group references out of group and sum references in mereology and non-mereology contexts with *it* and *them*. Error bars show the 95% confidence intervals.

There was a main effect of verb, with more mereological entity / group interpretations for the mereology verbs than the non-mereology verbs, and with more sum interpretations for the non-mereology verbs than mereology verbs, $OR = 4.24$, $SE = 1.64$, $p < .001$. The main effect of pronoun indicates more references to group interpretations with *it* than *them*, $OR = 522.61$, $SE = 352.20$, $p < .001$. The interaction between pronoun and verb types was significant, $OR = 8.53$, $SE = 6.12$, $p = .003$, such that, for the plural pronouns, the sum interpretation was preferred over the mereological entity / group interpretation across verb types, while for singular pronouns, there was an overall preference for the group interpretation, which was stronger for the mereology than for the non-mereology verbs (see Figure 2.). Model 3 suggests that in the mereology condition, *it* was preferred over *them* when referring to the group /mereological entity (train = engine \oplus goods wagon), $OR = 882.50$, $SE = 601.74$, $p < .001$. In the non-mereology condition, participants preferred *them* over *it* when viewing the two entities as forming a plural entity/sum ‘engine + goods wagon’, $OR = 1382.16$, $SE = 2622.63$, $p < .001$.

Conclusion. Our Experiment 1 provides further evidence for the hypothesis that the mereological contexts affect the preferences for singular and plural pronoun interpretation. In the mereology condition, we observe a clear preference for *it* to be interpreted as referring to the group / mereological entity (the train); by contrast, in the non-mereology condition, we observe a strong preference for *it* to refer to the atomic entities (the engine or the goods wagon).

The experiment also showed that in mereological contexts, in which the two atomic constituents were combined, participants had a mereological entity (group) interpretation for *it* and had a sum interpretation for *them*. Such different preferences provide some evidence that the contrast between groups and sums does play a role in these cases. These preferences were further investigated in Experiment 2.

Experiment 2

Experiment 2 tested participants' antecedent preferences using a sentence completion method. The Queen Mary Ethics of Research Committee at the Queen Mary University of London approved the study (QMREC2327a). In addition, since some participants were from University of Edinburgh, Psychology, the Philosophy, Psychology and Language Sciences (PPLS) Research Ethics Committee at the University of Edinburgh also approved the current study (Ref No: 257-2021/1).

Method.

Participants. Experiment participants ($n = 32$) were native English-speakers aged 21-24 from Queen Mary University of London (QMUL) or the University of Edinburgh. All were unaware of the study's purpose, and none had participated in Experiment 1. We sent our fliers to the Departments of Humanities and Social Sciences at QMUL and published our advertisement on a job announcement web-page at the University of Edinburgh. Our main selection criteria were being a native speaker of English, not being a bilingual (i.e., not growing up in an environment where more than one language was spoken), or not being of a native-like level in another language. The participants, who fit these criteria, were invited to our Experiment via Zoom or Skype.

Materials. We used the same sentences as in Experiment 1. Unlike in Experiment 1, each participant was provided with an initial sentence/context and asked to provide a completion answer for the sentence fragment ending with *it* or *them* in a manner consistent with the previous text. One experimental item with its conditions is given below:

9a. Pronoun *it* in the mereology/joined verb condition:

It was a Sunday family brunch in Edinburgh. Sarah demonstrated a lack of table manners.
She pressed the ham onto the bread and admired it...

9b. Pronoun *them* in the mereology/joined verb condition:

It was a Sunday family brunch in Edinburgh. Sarah demonstrated a lack of table manners.
She pressed the ham onto the bread and admired them...

9c. Pronoun *it* in the non-mereology/disjoined verb condition:

It was a Sunday family brunch in Edinburgh. Sarah demonstrated a lack of table manners.
She separated the ham from the bread and admired it...

9d. Pronoun *them* in the non-mereology/disjoined verb condition:

It was a Sunday family brunch in Edinburgh. Sarah demonstrated a lack of table manners.
She separated the ham from the bread and admired them...

Participants finished all sentence completions and sent their completions to the experimenter.

Immediately after they sent their completions, they were asked to go over their completions and underline what *it* or *them* referred to. The filler sentences included other referential expressions (e.g., they, he, this) and thus participants also underlined their referents as well.

This two-block data collection process was previously used in the studies of plurals (Çokal & Sturt, 2017; Koh & Clifton, 2002). There were 40 experimental and 60 filler sentences. There were four experimental conditions, one of which included the word *it* or *them* as the final

word before the blank (as in the sample above). In addition, one of the conditions included a joining / attachment verb such as *pressed* and the other one included a detachment verb such as *separated*. Four versions of each sentence and four files were constructed, using Latin Square counterbalancing. In each file, each sentence appeared in only one condition, with an equal number of items from each condition. Sentences were presented in a Word document in a fixed random order.

Procedure. Each participant made an appointment for a Zoom or Skype session. Both sessions of the experiment were completed in the same appointment session. Participants kept their camera on during the experiment. Participants read instructions, requirements for participation, informed consent, and compensation information prior to beginning the experiment. The requirements for Experiment 2 were the same as for Experiment 1. Each participant was provided one hour thirty minutes to complete the Word document online. After each participant finished the task, we asked them to underline what *it* or *them* refers to in each sentence. No feedback was given to the participants. The filler sentences had plural pronouns and participants underlined those cases as well (see a sample filler below.).

10. It was a hot and humid day in Rio de Janeiro. The Summer Olympics were very spectacular. The track athletes had worked very hard. The journalists had worked incredibly hard. They...

Data analysis. We used the following continuation codings and samples for *it* and *them*. The data analysis below was based on participants' underlined interpretations. It should be noted that due to the nature of the sentence completion experiment, participants could underline one atomic constituent of an object (e.g., the engine or the goods wagon) or two objects (e.g., engine, goods

wagon) but could not differentiate the sum interpretation from the group interpretation (i.e., train).

- If *it* or *them* referred to one atomic part of an object, then its referent was coded as a single object referent, as in 11 and 12 below.

11. It was a complex operation at Whitechapel Hospital in London. The orthopedic surgeon removed the artificial joint from the bone and examined *it* before deciding if a replacement was needed.

12. It was two hours before John's 50th birthday meal. Preparations were going badly. Emma laid the meat sauce onto the pasta sheet and put *them* in the oven with little time to spare.

- If *it* or *them* referred to two objects, then its referent was coded as a two/whole object referent (i.e., a sum or a group interpretation), as in 13-16 below.

13. It was an art lesson at Lavender Leonardo's children's art studio in London. The little child put the colorful paper onto the cardboard and admired *it* from afar before handing it to his mother for safe keeping.

14. It was a complex operation at Whitechapel Hospital in London. The orthopedic surgeon fused the artificial joint to the bone and examined it rigorously to ensure there was no chance of a fault.

15. It was a Sunday family brunch in Edinburgh. Sarah demonstrated a lack of table manners. She pressed the ham onto the bread and admired it before stabbing it with a fork and nibbling at the sandwich like it was an ice cream.

16. On Sunday, the woodcutter found a good tree in the Forest of Dean in Gloucestershire.

He removed the sharp head from the handle and checked them over before preparing to start cutting down the tree.

Unclear sentence completions (10.52%) were excluded from the logistic mixed effects regression analysis. Our unclear completions were cases in which the referents of *it* or *them* were underlined in the first sentence instead of the entities in the verb phrase (see 17 below.).

17. The furniture that Jane had ordered had to be dispatched early in the morning. The carpenter detached the last leg from the wooden top and checked it before packing it up and placing the box in the lorry.

Other unclear completions involved cases where a new discourse focus was introduced after *it* or *them* (see 18 below).

18. It was a sunny day at London's Walthamstow Reservoir. There were no fish to be caught. The fisherman attached the reel to the rod and inspected them. "Ere John!" said Steve, "Why do we always fish here when we know we'll never catch anything?". "It's better than fishing in the Thames and catching Hep C!" replied John.

Results.

We used the same statistical packages as in Experiment 1: logistic mixed effects regression, taking pronouns (*it* vs. *them*) and type of verb (joining verb vs. detaching verb) as the fixed effects, and including crossed random intercepts and slopes for participants and items.

Since the full model did not converge, we reduced random effect structure until the convergence was reached:

$$\text{response} \sim (\text{pronoun} * \text{verb type} + 1 \mid \mid \text{Participant}) + (\text{pronoun} * \text{verb type} + 1 \mid \mid \text{Item}) + \text{pronoun} * \text{verb type}$$

In the logistic mixed effects regression, we coded references to (atomic) parts of an object as 0 and references to two objects (sums) as 1. Factor labels were transformed into numerical values, and centered prior to analysis, with a mean of 0 and a range of 1. We used “scale contrast” for both factors (i.e., pronouns and verb) to have main effects. Figure 3 shows the estimated proportions of references to one or two objects with *it* and *them* in participants’ completions.

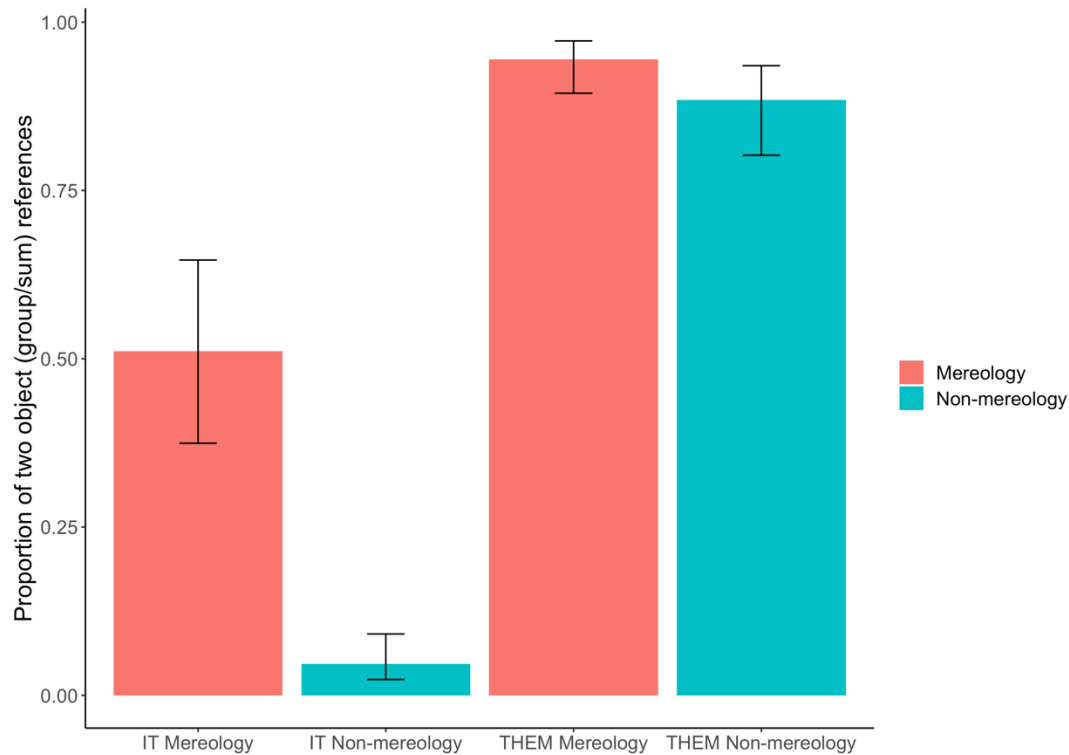


Figure 3. Estimated proportion of two object (group/sum) reference out of both one and two object references in mereology and non-mereology contexts with *it* and *them*. Error bars show the 95% confidence intervals.

The analysis revealed main effects of pronoun and verb type, with more two-object references (sum/group interpretations) with *them* than *it* in both mereology and non-mereology contexts, pronoun: $OR = 49.16$, $SE = 15.84$, $p < .001$ and verb type, $OR = 6.97$, $SE = 1.78$, $p <$

.001 (see Table S3 and S4). The interaction between pronoun and verb type was also significant, $OR = 0.11$, $SE = 0.05$, $p < .001$.

In the mereology condition, two-object references were more likely with *them* than with *it*, $OR = 17.71$, $SE = 7.25$, $p < .001$. However, *it* referred to two-objects more often in the mereology condition than in the non-mereology condition, $OR = 18.29$, $SE = 9.01$, $p < .001$, indicating sum or group interpretation. This suggests that in the mereology condition, participants preferred more group or sum interpretations with *it* than in the non-mereology condition.

The difference between *them* referring to two-objects (group or sum interpretation) in the mereological and non-mereological contexts was also significant: $OR = 2.42$, $SE = 0.95$, $p = .024$. In the non-mereology condition (e.g., detach the engine from the goods wagon), *it* referred to a single object more often than *them*, $OR = 117.33$, $SE = 60.61$, $p < .001$. In both mereology and non-mereology conditions, *them* referred to two-objects more often than *it* did, Mereology: $OR = 29.11$, $SE = 12.70$, $p < .001$, Non-mereology: $OR = 132.75$, $SE = 67.41$, $p < .001$.

Conclusion. Participants preferred *it* referring to both entities (i.e., interpreted the pronoun as having either a sum or group antecedent) in the mereological contexts, compared to *it* in non-mereological contexts, where *it* was more frequently interpreted as referring to one of the entities only. The fact that we found fewer references to NP1 in mereological contexts provides further evidence of a mereology effect for *it*—namely, that a mereological entity also became salient.

The results with *them* in the mereological contexts, where the two entities were arguments of a verb of junction / attachment (e.g., *hooked the engine to the goods wagon*), in particular the fact that participants preferred *them* when referring to both component objects, also suggests the formation of a mereological entity / group (e.g., engine \oplus goods wagon).

Interestingly, in the non-mereological context participants preferred *it* when referring to either of the two atomic entities (e.g., the engine or the goods wagon). In the same context, participants once again preferably interpreted *them* as referring to the sum (e.g., engine +goods wagon) since they formed a plural reference representation. Our results indicate that the formation of a plural reference as the interpretation for *them* took place in both mereological and non-mereological contexts, whereas this effect was more pronounced for *it* in mereological contexts than in non-mereological contexts

In order to determine how readers represented referent(s) of *it* and *them* in mereology and non-mereology contexts, and whether they obtained a group interpretation, we needed a task in which we could indirectly probe these representations. In Experiment 3, we used to a paraphrase task in which participants were asked to rewrite (or paraphrase) mereology and non-mereology constructing sentences.

Experiment 3

Method.

Participants. Participants ($n = 50$) were native English-speakers aged 21-24 from University of South Carolina, USA. One participant's responses were excluded from the analysis because they were not paraphrases of the given sentences. All participants were unaware of the purpose of the study and did not participate in Experiments 1 and 2. They earned course-credit for their participation.

Materials. Thirty-two stimuli from Experiment 1 were used. To prevent participants from adopting response strategies, each participant only saw eight items. To do this, we divided the 32 stimuli into subsets of eight and then run Latin Square over each subset. Each participant only saw one version of each experimental stimulus within a given subset. Sixteen

Word documents were prepared, and sentences presented in a fixed random order. In each file, there were two examples of each condition. Participants completed the Word document online.

Each participant was asked to paraphrase/rewrite given sentences (see examples 19a to 19d below.):

19a. Pronoun *it* in the mereology/joining verb condition:

The shoemaker glued the leather to the sole and inspected it meticulously with his eyepiece.

19b. Pronoun *them* in the mereology/joining verb condition:

The shoemaker glued the leather to the sole and inspected them meticulously with his eyepiece.

19c. Pronoun *it* in the non-mereology/detaching verb condition:

The shoemaker stripped the leather from the sole and inspected it meticulously with his eyepiece.

19d. Pronoun *them* in the non-mereology/detaching verb condition:

The shoemaker stripped the leather from the sole and inspected them meticulously with his eyepiece.

Procedure. After participants signed a consent form, a Word document including paraphrase sentences was sent to them via e-mail. They were asked to paraphrase the given sentences. At the beginning of the experiment, each participant was provided with two examples of paraphrase sentences (see below items 20 and 21).

20. Original sentence: Giraffes like Acacia leaves and hay, and they can consume 75 pounds of food a day.

Paraphrase: A giraffe can eat up to 75 pounds of Acacia leaves and hay daily.

21. Original sentence: Symptoms of influenza include fever and nasal congestion.

Paraphrase: A stuffy nose and elevated temperature are signs you may have the flu.

It should be noted that in the instructions, participants were NOT asked to replace full NPs with anaphoric expressions in their paraphrases. Each participant was allowed 40 minutes to complete the experiment and was asked to send their responses in a Word document to the first author's email address.

Data analysis.

We classified participants' responses as follows.

1. If *it* or *them* referred to a single atomic constituent (e.g., the engine, artificial joint, bone, canvas), then its referent was coded as a single-object referent (see examples 22 and 23 below).

Single object references do not include a group interpretation.

22. Original sentence: The orthopedic surgeon fused the **artificial joint** to the bone and examined it carefully through a deep incision.

Paraphrase: The surgeon carefully examined the artificial joint through a deep incision.

23. Original sentence: The restorer removed the **canvas** from the frame and sent them carefully to the museum.

Paraphrase: The restorer carefully sent the canvas to the museum.

2. If *it* or *them* referred to two objects, then its referent was coded as a sum referent ($a + b$) or a group / mereological entity ($a \oplus b$). We differentiated the group and sum interpretations in the coding as follows: The group references were the whole object references, which were not mentioned explicitly in the previous context (e.g., table) and participants constructed such

entities from the meaning of the sentence. On the other hand, a sum referent required two objects conjoined with *and* construction (e.g., the suspension + chassis).

24. Original sentence: The carpenter attached the last leg to the wooden top and checked it carefully for defects.

Paraphrase: The carpenter checked the table for defects after the last leg was attached.

25. Original sentence: She pressed the ham onto the bread and admired them hungrily while standing in the kitchen.

Paraphrase: She admired the ham sandwich she made in the kitchen.

26. Original sentence: The hardworking experienced autoworker unbolted the suspension from the chassis and sent them quickly to the next station.

Paraphrase: The suspension and chassis were sent quickly to the next station after the experienced autoworker unbolted them.

We had unclear cases in which participants' paraphrased sentences were ungrammatical (6.4% = 41 cases out of 637) or ambiguous (38% = 242 cases out of 637). 44.4% of the data (283 cases) were coded as "unclear cases" and excluded from the analysis. An independent rater, not involved in data collection or preparation of the classification criteria, and unfamiliar with the aims of the study, was trained and subsequently annotated 30% of the total sample. Classification reliability was calculated for this subsample of the data: percent disagreements over annotations were 1.8%. In initial consensus meetings, these disagreements were resolved by checking criteria.

We conducted the same statistical analyses and used the same packages as in Experiments 1 and 2. Two models were conducted. In model 1, the two-object (sum) references were combined with the whole-object (group) references and coded as 1, whereas an atomic

entity reference (i.e., one part of an object) was coded as 0. The full logistic mixed-effects regressions for model 1 including crossed random intercepts and slopes for participants and items converged:

response ~ (pronoun * verb types + 1 | Participant) + (pronoun * verb types + 1 | Item) + pronoun * verb types

In model 2, we excluded single object references and unclear references and then contrasted a sum interpretation ($a + b$) with a group (mereological entity) reference ($a \oplus b$). In model 2, we coded the sum interpretations as 0 and the group references as 1. Since the full model did not converge, we reduced the random effect structures until convergence was reached. We also checked the random effects with least variance:

response ~ (pronoun + verb types + 1 | | Participant) + (1 | Item) + pronoun * verb types

Results.

There was a main effect of verb, with more sum/group references for the mereology-constructing verbs than the non-mereology verbs, (see Tables S5 and S6) $OR = 13.69$, $SE = 5.46$, $p < .001$. The analysis also revealed a main effect of pronoun, $OR = 3.36$, $SE = 1.45$, $p = .005$. The interaction between pronoun and mereology was not significant, $OR = 0.33$, $SE = 0.26$, $p = .154$ (see Figure 4).

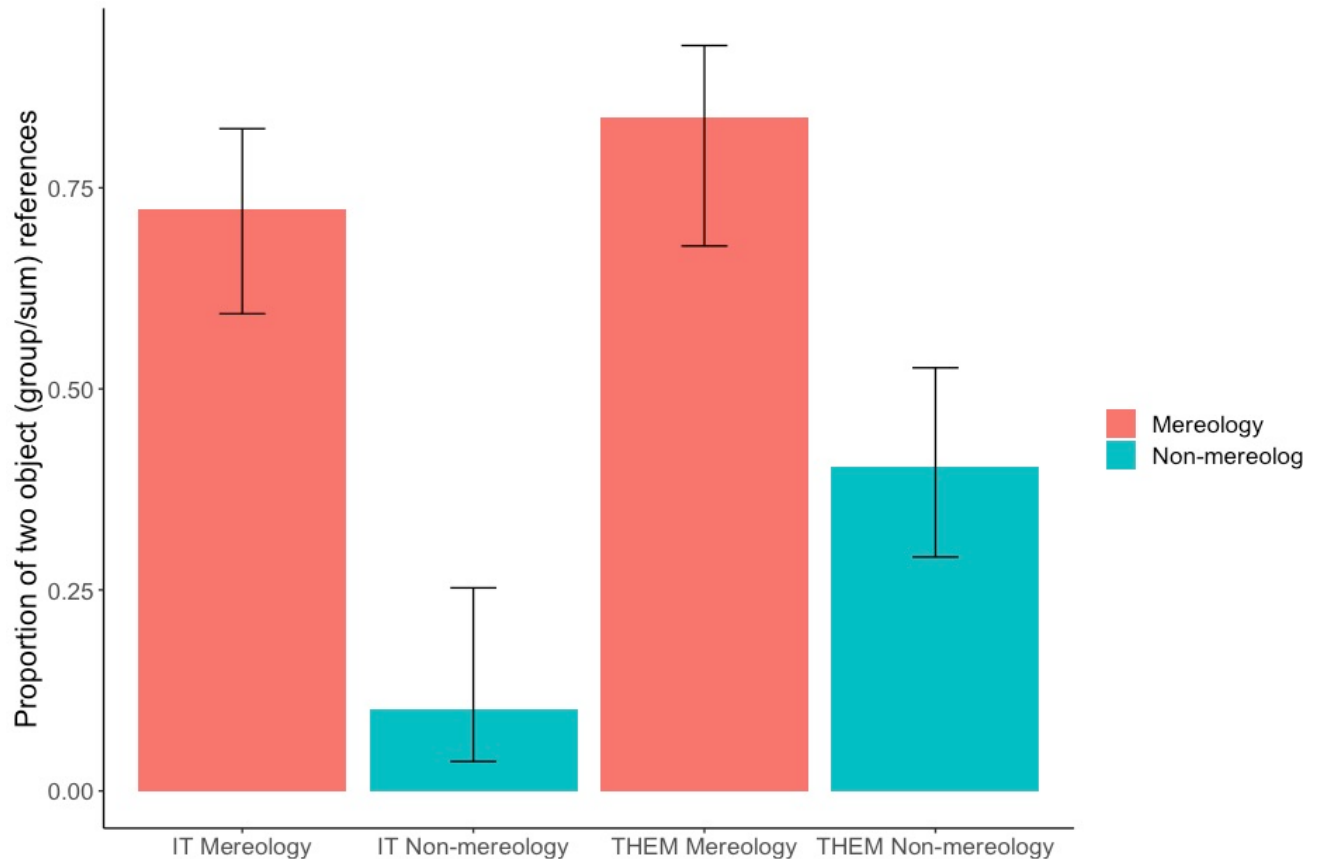


Figure 4. Estimated proportions of group and sum references out of the total number of references (combined group and sum references and an atomic entity reference) in mereology and non-mereology contexts with *it* and *them*. Error bars show the 95% confidence intervals.

In mereological contexts, the sum/group references were more frequent than references to an atomic part of an object, irrespective of anaphoric expression type. Thus, there were no significant differences between *it* and *them* in mereology contexts, $OR = 1.45$, $SE = 0.50$, $p = .276$. The proportions of *it* referring to sums/groups in mereology contexts were higher than those of *it* in non-mereological contexts, $OR = 14.73$, $SE = 6.47$, $p < .001$. In non-mereology cases, the proportions of sums/groups with *them* were greater than of those with *it*, $OR = 3.81$, $SE = 1.41$, $p < .001$.

Single object references and unsure cases were excluded from model 2 (see Figure 5), and the model compared distributions of references to sums ($a + b$) and groups ($a + b = z$) (see Tables S5 and S6). There was a main effect of pronoun, with more references to groups / mereological entities with *it* than *them* in both mereology and non-mereology contexts—pronoun: $OR = 3.77, SE = 1.80, p = .006$. There was no significant main effect of verb type, $OR = 1.01, SE = 0.50, p = .981$, or interaction between the two factors, $OR = 0.15, SE = 0.18, p = .112$. Note that the number of observations this model (and the results in Figure 5) is based upon are somewhat lower than in the other analyses (here $N = 185$). The latter might be the reason for why there seems to be an interaction pattern in the proportions (cf. Figure 5) but no significant interaction effect.

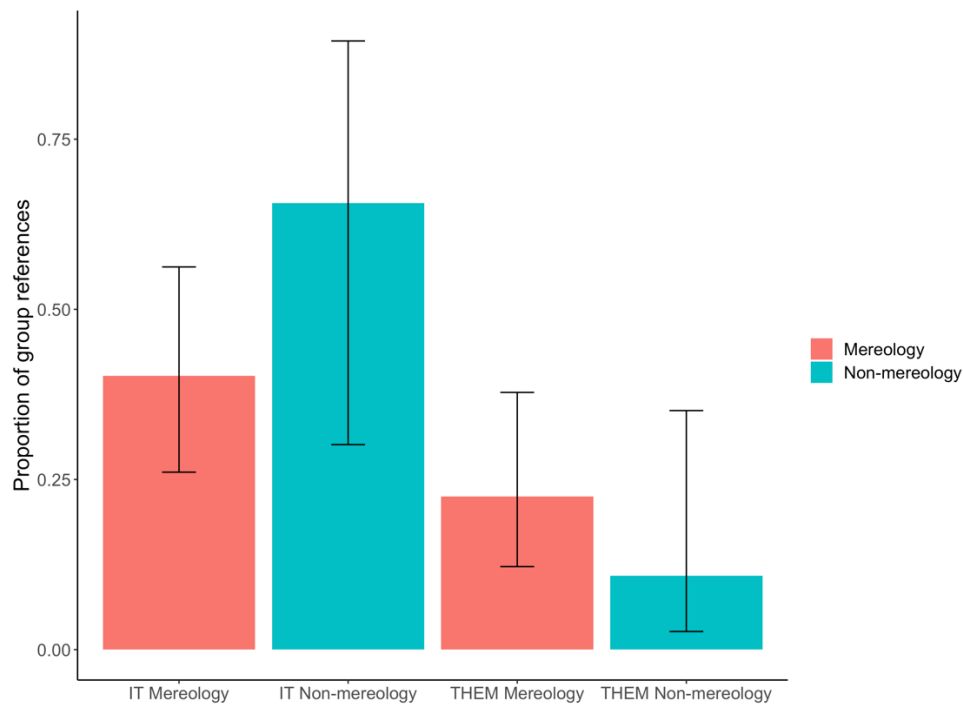


Figure 5. Estimated proportions of group reference of total (sum/group) references in mereology and non-mereology contexts with *it* and *them*. Error bars show the 95% confidence intervals.

Conclusion.

Model I revealed that in mereological contexts, the sentences with *it* and *them* were more likely to be paraphrased as references to a sum or group. The preference for references to one atomic constituent of an object with *them* in a non-mereology condition indicated that participants had a split antecedent representation (i.e., they considered 'engine' and 'goods wagon' as separate entities in their mental representation). Model 2 showed that in both mereology and non-mereology contexts, participants preferred a group/mereological entity representation with *it* to one with *them*.

General Discussion

In this paper, we investigated (a) whether constituents of a mereological entity (i.e., engine or boxcar) or the whole entity (i.e., train) are more accessible to a singular pronoun *it*; and (b) whether participants favor a sum-like or a group-like interpretation. Lastly, we explored how the part-whole relationship interacts with mereological entity references and processing of anaphors.

Our experiments showed that in a context in which mereological entities are created, there is a clear advantage for singular reference over plural reference. Experiments 1 and 3 demonstrated that using *it* to refer to a sum entity ($a \cup b$) or a group entity (i.e., the whole mereological entity, the train) was more likely in a mereological context than in a non-mereological context. Experiment 2 showed that in a sentence completion task, participants preferred continuations in which *it* refers to the atomic parts of an object; but even in this experiment, group / mereological entity references with *it* in a mereological context were significantly more numerous than those of *it* in a non-mereological context.

These findings are compatible with a view in which a mereology-forming predicate (i.e., *hooked up the engine to the boxcar*) immediately causes a new singular reference object, what we called here the group or mereological entity (i.e., train) to be added to a discourse representation, which becomes the preferred referent for a singular pronoun, even when the resulting entity (train) is not explicitly mentioned in the context.

Experiment 1 showed that participants preferred the sum interpretation for *them* – i.e., the interpretation referring to a plural entity consisting of both parts of an object – in both mereological (*hooked up the engine + the boxcar*) and non-mereological contexts (*detached the engine from the boxcar*). Experiments 2 and 3 showed the same pattern. On the other hand, in a mereological context with *it*, participants preferred the group interpretation referring to a mereological entity (i.e., the train). These findings shed light on our third research question of whether participants had a sum-like or group-like interpretation for the pronouns. The patterns in Experiments 1 and Experiment 3 clearly suggest a preference for a sum interpretation for *them* and a mereological entity (group) interpretation for *it*. In other words, Experiments 1 and 3 suggest that in a mereology condition, participants preferred interpreting *it* as the mereological entity 'train' ($\text{engine} \oplus \text{good wagons}$) and *them* as referring to a sum of two objects ($\text{engine} \cup \text{good wagons}$). In Experiment 2, while we observed the use of *it* referring to a single object (i.e., one part of an object) in the non-mereological contexts and *them* referring to two-objects (i.e., two parts of an objects) in the mereological and non-mereological contexts, the higher means of *it* referring to both entities in the mereological context indicate a group interpretation for *it*.

Our findings can be linked with those of other studies that have examined the use of singular pronouns in plural contexts (see Arnold et al., 2021; Albrecht & Clifton, 1998; Garrod & Sanford, 1982; Gelormini-Lezama, & Almor, 2014; Gordon et al., 1999; Moxey et al., 2004).

Previous studies showed the use of a conjunction led to the formation of a plural entity; thus, in such contexts, the use of a singular pronoun led to processing difficulty compared to the use of a plural pronoun (the conjunction cost). Here, it is worth making a clear distinction between cases where there is one conjoined NP representing a syntactic unit referring to a plural entity and cases where it is necessary to construct such an entity from the meaning of the sentence. Such conjunction difficulty was claimed to be due to either (a) individuating the two atomic elements of the sum (Moxey et al, 2004; Sanford & Moxey, 1995), (b) the plural entity / sum becoming more salient than each individual atomic referent (Gordon et al., 1999), or (c) role mapping. Even though we do not have online reading experiments to examine online processing difficulty, the results of the current study show that even when there is no conjunction, the verb semantics (i.e., hook up) can lead to the formation of a complex object. When the complex object can be interpreted as an entity in its own, a group, participants use a singular pronoun *it* in mereological context. This would seem to be particularly consistent with hypothesis (b)—the predicate makes the mereological entity / group (the train) relevant, but we do not exclude other explanations.

It is also relevant here to discuss the connection between our findings and Poesio et al.'s (2006) proposal regarding how potential antecedents of *it* may be represented. They proposed that in mereology constructing cases, *it* may refer to (i) the engine I, (ii) the boxcar (b), (iii) the engine and boxcar combined ($e \oplus b$), or (iv) an underspecified element z , which ranges over all these alternatives (e, b, and $e \oplus b$). The results of Experiments 1 and 3 suggest that in mereological contexts, whatever the initial interpretation may be, ultimately participants do not interpret *it* as referring to one of the individual elements (e.g., engine or boxcar), suggesting that (iii) or (iv) are more likely. While the cases in which the interpretation of *it* in mereological

contexts is more uncertain may be evidence for an underspecified interpretation, other explanations are also possible (e.g., Experiment 1).

In terms of Poesio and Reyle's (2001) theory, our findings suggest antecedent representation ranges over two different possible interpretations of the combined object $e \oplus b$. One of these interpretations is similar to the sum interpretation of plural entities. Intuitively, this interpretation preserves the properties of the individual component entities and is felicitously referred to with a plural pronoun. The second interpretation is like the group interpretation of plural entities. This interpretation involves the representation of the combined object as an individual, which is more than just the sum of its parts. It is felicitously referred to with a singular pronoun. Thus, the interpretation is underspecified between these two possible interpretations.

A final link that might be worth exploring is the connection between the mereological contexts studied here and the cases of so-called *context change accommodation* (Dale, 1992; Webber & Baldwin, 1992). The mereological entities discussed in this paper are cases of anaphoric reference to objects created as a result of actions described in a text. They are particularly common in instructional text such as the following recipes, where continuations 27a, 27b and 27c refer to entities created as a result of the action in the first sentence.

27. Mix the flour, butter and water.

27a. Knead the dough until smooth and shiny.

27b. Spread the paste over the blueberries.

27c. Stir the batter until all lumps are gone.

Another relevant direction for further research would be to investigate demonstratives and singular pronouns in mereological contexts and/or context change accommodation as in the

Brown-Schmidt et al.'s study (2005) that showed participants preferred *that* when referring to a newly formed composite entity (e.g., “the cup on the saucer” = that). If this is the case, participants would prefer *that/this* when referring to a mereological entity (engine + goods wagon) (e.g., The railway man hooked up the engine to the goods wagon and send *it/this/that* to the station.). Such a study would provide further information as to whether participants would keep the group interpretation with *it* or would prefer demonstratives in such contexts.

In the current study, one of our limitations is not having reading time data to explore a processing penalty of a singular pronoun when referring to one of the parts of a plural object. In addition, there would be a penalty for using a singular pronoun to refer to one of the parts of such an entity, even though only the parts were explicitly introduced. However, our offline experiments in the current study suggest that there may be no such penalty for using a singular pronoun to refer to a plural entity under a group interpretation. To investigate these potential processing difficulties and have reliable conclusions, online reading experiments (e.g., eye-tracking reading experiment) need to be conducted to pin down the costs of referring to mereological entities.

In the future, this line of research may also help us to understand the distinction between inferential and referential processing, as mereological interpretations are a simple case of the more general phenomenon of referents being introduced in discourse through inference processes. A more complex example is shown in 27 and 27a, where a new entity, ‘the dough,’ is obtained by mixing flour, butter, water, not introduced using an explicit mention. To interpret the subsequent reference to that entity (i.e., Knead the dough), a reader needs to make inferences from the discourse and establish a referential chain between 27 and 27a. To our knowledge, there has been very limited work on such cases (see Poesio et al., 2023). We hope our findings will

contribute to move the state-of-the-art in psychological and computational modelling of anaphora resolution forward beyond the simplest form of anaphora to cases requiring such inferences.

In conclusion, the results of the current study support the findings from previous work that the frequently observed preferences on singular versus plural reference in plural contexts are modified in mereological contexts (Poesio & Reyle 2001, Poesio et al., 2006) and provide new evidence regarding the interpretation of singular and plural pronouns in such contexts. Specifically, we found that the higher acceptability in (3a) than (3b) [repeated below] found by Poesio and Reyle (2001) is most likely due to the whole object (i.e., train) becoming available as an antecedent for *it* in mereological context.

- a. The engineer hooked up the engine to the boxcar and sent it to London.
- b. The engineer separated the engine from the boxcar and sent it to London.

REFERENCES

- Albrecht, J., & Clifton, C. (1998). Accessing singular antecedents in conjoined phrases. *Memory & Cognition*, 26, 599-610.
- Arnold, J. E., Mayo, H. C., & Dong, L. (2021). My pronouns are they/them: Talking about pronouns changes how pronouns are understood. *Psychonomic bulletin & review*, 28(5), 1688–1697. <https://doi.org/10.3758/s13423-021-01905-0>
- Bard, E. G., Robertson, D., & Sorace, A. (1996). Magnitude estimation of linguistic acceptability. *Language*, 72, 32-68.
- Barker, C. (1992). Group terms in English: Representing groups as atoms. *Journal of Semantics*, 9, 69-93.

- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67, 1–48.
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68, 10.
- Brown-Schmidt, S., Byron, D.K., & Tanenhaus, M.K. (2005). Beyond salience: Interpretation of personal and demonstrative pronouns. *Journal of Memory and Language*, 53, 292-313.
- Carrieras, M. (1997). Plural pronouns and the representation of their antecedents. *European Journal of Cognitive Psychology*, 9, 53-87.
- Champollion, L. (2020). Distributivity, collectivity and cumulativity. In L. Matthewson, C. Meier, H. Rullmann, & T. E. Zimmermann (Eds.), *Companion to Semantics* Wiley.
- Çokal, D. & Sturt, P. (2017). The effect of referring expression on antecedent-grouping choice in plural reference resolution. *Discourse Processes*, 55(2), 157-165.
- Clifton, C. & Ferreira, F. (1987). Discourse structure and anaphora: Some experimental results. In M. Coltheart (Ed.), *Attention and Performance* (Vol. X) (pp. 635-654). London: Erlbaum.
- Eschenbach, C., Habel, C., Herweg, M., & Renkamper, K. (1989). Remarks on plural anaphora. In Proceedings of 4th conference of European chapter of the association for computational linguistics (pp. 161-167), U Manchester.
- Garnham, A. (2001). *Mental models and the interpretation of anaphora*. Philadelphia, PA: Taylor and Francis.

- Garrod, S. C., & Sanford, A. J. (1982). The mental interpretation of discourse in a focused memory system: Implications for the interpretation of anaphoric noun-phrases. *Journal of semantics, 1*, 21-41.
- Gelormini-Lezama, C., & Almor, A. (2014). Singular and plural pronominal reference in Spanish. *Journal of Psycholinguistic Research, 43*(3), 299–313
- Gordon, P. C., Hendrik, R., Ledoux, K., & Yang, C. L. (1999). Processing reference and the structure of language: An analysis of complex noun phrases. *Language and Cognitive Processes, 14*, 353-379.
- Filik, R., Ingram, J., Moxey, L., & Leuthold, H. (2021) Irony as a Test of the Presupposition-Denial Account: An ERP Study. *J Psycholinguist Res, 50*, 1321–1335
- Heinat, F., & Klingvall, E. (2019). Anaphoric reference to quantified expressions in Swedish. *Journal of Psycholinguistic Research, 48*, 551–568.
- Hielscher, M. & Müssler, J. (1990). Anaphoric resolution of singular and plural pronouns: The reference to persons being introduced by different co-ordinating structures. *Journal of Semantics, 7*, 347-364.
- Jaeger, T.F., (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language, 59*, 434–446.
- Kaup, B. and Kelter, S. and Habel, C. (2002). Representing referents of plural expressions and resolving plural anaphors. *Language and Cognitive Processes, 17*, 405—450.
- Koh, S. & Clifton, C. (2002). Resolution of the antecedent of a plural pronoun: Ontological categories and predicate symmetry. *Journal of Memory and Language, 46*, 830-844.
- Koh, S., Sanford, A., Clifton, C. & Dawydiak, E. (2008). Good-enough representation in plural and singular pronominal reference: Modulating the conjunction cost. In J. Gundel and

- N. Hedberg (Eds.), *Interdisciplinary Perspectives on Reference Processing*. Oxford University Press.
- Ingram, J., & Ferguson, H.J. (2017). Complement Set Reference after Implicitly Small Quantities: An Event-Related Potentials Study. *Discourse Processes*, 55 (2), 1-11.
- Landman, F. (1989a). *Groups I. Linguistics and Philosophy* 12, 559-605.
- Landman, F. (1989b). *Groups II. Linguistics and Philosophy* 12, 723-744.
- Landman, F. (1996). Plurality. In Lappin, S. (Ed.), *The Handbook of Contemporary Semantic Theory* (pp. 425—457). Blackwell, London.
- Link, G. (1983). The logical analysis of plurals and mass terms: a lattice theoretical approach. In Bauerle, R., Schwarze, C., and von Stechow, A., (Eds.), *Meaning, Use and Interpretation of Language* (pp. 302-323). De Gruyter, Berlin.
- Link, G. (1984). Hydras. On the logic of relative clause construction with multiple heads. In F. Landman and F. Veltman (Eds.), *Varieties of Formal Semantics*. GRASS 3. Foris, Dordrecht.
- 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.
- Moxey, L. M., & Filik, R. (2010). The effects of character desire on focus patterns and pronominal reference following quantified statements. *Discourse Processes*, 47, 588–616.
- Moxey, L.M., Sanford, S.J., & Tonks, K. (2011). Representing characters in a scenario. What makes two individuals a set. *Discourse Processes*, 27 (9), 1405-1424.

- Micklos, A., Walker, B., & Fay, N. (2020). Are people sensitive to problems in communication. *Cognitive Science*, 44, 1-16.
- Yu, J., Moosavi, N.S., Paun, S., & Poesio, M. (2021). Stay together: A system for single and split-antecedent anaphora resolution. Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL, 4174-4184.
- Poesio, M. & Reyle, U. (2001). Underspecification in anaphoric reference. *Proceedings of 4th International Workshop on Computational Semantics*, Tilburg University.
- Poesio, M., Reyle, U. & Stevenson R. (2006). Justified sloppiness in anaphoric reference. In H. Bunt and R. Muskens (Eds.), *Computing Meaning 3* (pp.1-20). Kluwer, Amsterdam.
- Poesio, M., Yu, j., Paun, S., Aloraini, A., Lu, P., & Haber, J., Çokal, D. (2023). Computational Models of Anaphora. *Annual Review of Linguistics*.
- Sanford, A.J., Garrod, S.C, Lucas, A., and Henderson, R. (1983). Pronouns without explicit antecedents. *Journal of Semantics*, 2, 303-318.
- Sanford, A. J. & Lockhart, F. (1990). Description types and method of conjoining factors as influencing plural anaphora: A continuation study of focus. *Journal of Semantics*, 7, 365-378.
- Sanford, A. J. & Moxey, L. M. (1995). Notes on plural reference and the scenario mapping principle in comprehension. In C. Habel & G. Rickheit (Eds.), *Focus and cohesion in discourse* (pp. 18-34). Berlin: de Gruyter.
- Sanford, A. J., & Filik, R. (2007). “They” as a gender-unspecified singular pronoun: Eye tracking reveals a processing cost. *The Quarterly Journal of Experimental Psychology*, 60(2), 171–178.

Upadhyay, S. S. N., Houghton, K. J., & Klin, C. M. (2019). Is “few” always less than expected?;

The influence of story context on readers’ interpretation of natural language quantifiers.

Discourse Processes, 56(8), 708–727.

Varzi, A. (2019). Mereology. In E.N., Zalta (Ed.), *The Stanford Encyclopaedia of Philosophy*.

Metaphysics Research Lab, Stanford University

Winter, Y. and Scha, R. (2019). Plurals. In Lappin, S. and Fox, C. (Eds.), *The Handbook of*

Contemporary Semantic Theory, 2nd edition (pp. 77—113). Blackwell, London.

Table S1. Odds ratios, standard errors, and P-values across logistic mixed effects regressions models in Experiment 1

	Model 1			Model 2			Model 3		
Predictors	Odds Ratios	std. Error	P-Value	Odds Ratios	std. Error	P-Value	Odds Ratios	std. Error	P-Value
(Intercept)	0.31 (0.20 – 0.49)	0.07	<.001	7.92 (4.96 – 12.64)	1.89	<.001	0.48 (0.23 – 1.01)	0.18	.040
Pronoun	67.10 (33.23 – 135.51)	24.06	<.001	59.10 (28.75 – 121.49)	21.73	<.001	522.61 (139.49 – 1958.05)	352.20	<.001
Verb	0.23 (0.12 – 0.45)	0.08	<.001	4.19 (2.29 – 7.65)	1.29	<.001	4.24 (1.99 – 9.04)	1.64	<.001
Pronoun* verb	0.38 (0.13 – 1.08)	0.20	.069	0.34 (0.13 – 0.86)	0.16	.024	8.53 (2.09 – 34.78)	6.12	.003
	Random Effects			Random Effects			Random Effects		
σ^2	3.29			3.29			σ^2	3.29	
τ_{00} Par	0.90			0.95			τ_{00} Par	2.88	
τ_{00} Item	0.38			0.40			τ_{00} Item	1.29	
τ_{11} Participant.Pronoun	1.06			1.17			τ_{11} Participant.Pronoun	1.97	
τ_{11} Participant.verb	0.66			0.75			τ_{11} Participant.verb	0.07	
τ_{11} Participant.Pronoun:verb	0.00						τ_{11} Participant.Pronoun:verb		
τ_{11} Item.Pronoun	0.61			0.59			τ_{11} Item.Pronoun	1.57	

τ_{11} Item.verb.	1.54	0.63	τ_{11} Item.verb.	0.66
τ_{11} Item.Pronoun:verb			τ_{11} Item.Pronoun:verb	
ρ_{01}	0.30		ρ_{01}	
	0.55		ρ_{01}	
ICC	0.41	0.29	ICC	0.56
$N_{\text{Participant}}$	40	40	$N_{\text{Participant}}$	40
N_{Item}	40	40	N_{Item}	40
Observations	1579	1537	Observations	1007
Marginal R^2 / Conditional R^2	0.474 / 0.687	0.501/0.646	Marginal R^2 / Conditional R^2	0.571/0.811

Table S2. Coefficients, standard errors, and Z-values across logistic mixed effects regressions models in Experiment 1

	Model 1				Model 2				Model 3			
	β	SE	Z	p-value	β	SE	Z	p-value	β	SE	Z	p-value
Intercept	-1.630	0.228	-5.086	.001	2.068	0.238	8.670	.001	- 0.727	0.375	-1.938	.050
Pronoun	-4.206	0.358	11.731	.001	4.079	0.367	11.096	.001	-6.258	0.673	-9.287	.001
Verb	-1.469	0.344	-4.263	.001	1.432	0.307	4.659	.001	1.445	0.386	3.745	.001
Pronoun*Verb	-0.972	0.535	-1.817	.069	-1.086	0.479	- 2.264	.023	-2.143	0.717	-2.987	.002

Table S3. Odds ratios, standard errors, and P-values across logistic mixed effects regressions models in Experiment 2

Model I			
Predictors	Odds Ratios	std. Error	P-Value
(Intercept)	1.62 (1.02 – 2.59)	0.39	.042
Pronoun	49.16 (26.14 – 92.46)	15.84	<.001
Verb	6.97 (4.22 – 11.51)	1.78	<.001
Pronoun * Verb	0.11 (0.04 – 0.26)	0.05	<.001
Random Effects			
σ^2	3.29		
τ_{00} Par	0.09		
τ_{00} Item	1.39		
τ_{11} Item.Pronoun	1.08		
τ_{11} Item.verb.	0.23		
τ_{11} Item.Pronoun:verb	0.15		
τ_{11} Participant.Pronoun	0.08		
τ_{11} Participant.verb	0.50		
τ_{11} Participant.Pronoun:verb	1.03		
ρ_{01}			
ρ_{01}			
ICC	0.02		
N Participant	32		

N _{Item}	40
Observations	1234
Marginal R ² / Conditional R ²	0.602 / 0.611

Table S4. Coefficients, standard errors, and Z-values across logistic mixed effects regressions model in Experiment 2

	Model			
	β	SE	Z	p-value
Intercept	0.485	0.238	2.038	.004
Pronoun	3.895	0.322	12.08	.001
Verb	1.942	0.255	7.593	.001
Pronoun*Verb	-2.248	0.452	-4.968	.001

Table S5. Odds ratios, standard errors, and P-values across logistic mixed effects regressions models in Experiment 3

Predictors	Model I			Model II		
	Odds Ratios	std. Error	P-Value	Odds Ratios	std. Error	P-Value
(Intercept)	1.02 (0.67 – 1.56)	0.22	.391	0.42 (0.25 – 0.71)	0.11	.001
Pronoun	3.36 (1.44 – 7.82)	1.45	.005	3.77 (1.47 – 9.63)	1.80	.006
Verb	13.69 (6.26 – 29.92)	5.46	<.001	1.01 (0.39 – 2.66)	0.50	.981
Pronoun * Verb	0.33 (0.07 – 1.51)	0.26	.154	0.15 (0.01 – 1.56)	0.18	.112
Random Effects				Random Effects		
σ^2	3.29			3.29		
τ_{00} Par	0.02			0.45		
τ_{00} Item	0.11			0.05		
τ_{11} Participant.anaphora	2.17			1.33		
τ_{11} Participant.verb	0.58			0.55		
τ_{11} Participant.pronoun:verb	0.31					
τ_{11} Item.pronoun	0.00					
τ_{11} Item.verb.	0.01					
τ_{11} Item.pronoun:verb	0.55					
ρ_{01}	-0.10					
ρ_{01}				0.13		
ICC						

N _{Participant}	49	49
N _{Item}	8	8
Observations	354	185
Marginal R ² / Conditional R ²	0.396/NA	0.132/0.248

Table S6. Coefficients, standard errors, and Z-values across logistic mixed effects regressions model in Experiment 3

	Model 1				Model 2			
	β	SE	z	p-values	β	SE	z	p-values
Intercept	0.024	0.213	0.115	.908	-0.862	0.266	-3.237	.001
Pronoun	1.211	0.431	2.805	.005	1.326	0.478	2.769	.001
Verb	2.616	0.399	6.556	.001	0.011	0.492	0.024	.981
Pronoun*Verb	-1.100	0.771	-1.427	.153	-1.912	1.203	-1.589	.011

