

UC Berkeley

UC Berkeley Previously Published Works

Title

Sortal concepts and pragmatic inference in children's early quantification of objects

Permalink

<https://escholarship.org/uc/item/0r4008pj>

Journal

Elsevier, 66(3)

Authors

Srinivasan, Mahesh
Chestnut, Eleanor
Li, Peggy
[et al.](#)

Publication Date

2013-01-20

Peer reviewed



ELSEVIER

Contents lists available at SciVerse ScienceDirect

Cognitive Psychology

journal homepage: www.elsevier.com/locate/cogpsych



Sortal concepts and pragmatic inference in children's early quantification of objects



Mahesh Srinivasan^{a,*}, Eleanor Chestnut^{a,c}, Peggy Li^b, David Barner^a

^aUniversity of California, San Diego, United States

^bHarvard University, United States

^cStanford University, United States

ARTICLE INFO

Article history:

Accepted 20 January 2013

Available online 28 February 2013

Keywords:

Sortal concepts

Quantification

Pragmatic inference

Lexical semantics

ABSTRACT

It is typically assumed that count nouns like *fork* act as logical sortals, specifying whether objects are countable units of a kind (e.g., that a whole fork counts as “one fork”) or not (e.g., that a piece of a fork does not count as “one fork”). In four experiments, we provide evidence from linguistic and conceptual development that nouns do not specify units of quantification, but include both whole objects and their arbitrary parts in their denotations. We argue that, to restrict quantification to whole objects, nominal concepts are enriched pragmatically, via contrast with concepts denoted by alternative descriptions: a piece of a fork is not counted as “one fork” because it is “one piece of fork.” Experiment 1 replicated previous findings that children count pieces of broken objects as whole objects (e.g., two pieces of fork as “two forks”), and showed that children also accept whole object labels as descriptions of object pieces (e.g., “two forks” to describe two pieces of fork). Experiment 2 showed that although children accept such descriptions in isolation, they prefer measure phrases (e.g., “two pieces of fork”) when they are explicitly presented as alternatives. Experiment 3 found that children were better at excluding pieces from their counts of whole objects when measure phrases were primed prior to counting, making them accessible as alternatives to whole object labels. Finally, Experiment 4 taught children names for novel objects, and found that they do not count parts that are given unique labels or that have non-linguistic properties that suggest they are members of distinct object kinds (e.g., unique functions or physical affordances). Together, our results suggest that for children and adults alike, nominal concepts do not provide necessary

* Corresponding author. Address: Department of Psychology, UCSD, 9500 Gilman Drive, La Jolla, CA 92093-0109, United States.

E-mail address: maheshs@ucsd.edu (M. Srinivasan).

and sufficient criteria for excluding parts from object kinds. To specify units of quantification – and do the work of sortals – concepts are contrasted with one another and enriched pragmatically.

© 2013 Elsevier Inc. All rights reserved.

1. Introduction

“Only a concept which isolates what falls under it in a definite manner, and which does not permit any arbitrary division of it into parts, can be a unit relative to a finite Number.”

– Gottlob Frege, *The Foundations of Arithmetic*

What do children learn when they acquire the meanings of nouns? Typically, it is assumed that common count nouns like *dog*, *table*, and *kite* are logical sortals: they not only distinguish between different kinds of things, but they also encode criteria for judging whether an object is a countable unit of a kind (see Frege, 1884/1980; Geach, 1962; Quine, 1960; Strawson, 1959; Wiggins, 1967). For example, although a request to “count the things in the room” is too vague to satisfy, a request to count “the books” or even “the pages” provides a clear specification of what should be counted. Thus, by most accounts, nouns specify units of quantification, guiding not only counting, but also the use of quantifiers and number marking in natural language. In this paper, we question whether nouns act as sortals, and thus, the role they play in quantification. Based on data from the counting behavior of 3- to 7-year-old children, we argue that the conceptual content of nouns does not alone explain how units of quantification are specified. Instead, we suggest that, to identify individual units, language users supplement noun meanings with a simple form of conversational inference, rooted in the pragmatics of lexical contrast (Clark, 1987, 1990).

There is widespread recognition that nouns play a critical role in specifying units of quantification (for discussion, see Carey, 2009; Macnamara, 1986; Xu, 2007). This is because, in the absence of conceptual constraints, almost anything can be considered an individual unit. Humans count objects ranging from planets and stars to blood cells and molecules. Even substances, events, and abstract entities can be counted, so long as they are first individuated. Nouns are important because they appear to specify which of the many candidate individuals to consider when choosing units for quantification. This is true not just in the case of counting, but also when interpreting linguistic forms such as quantifiers, determiners, and number agreement.¹

Despite a general consensus that nouns specify units of quantification, there is little agreement regarding how they might do so, and more generally, what form concepts take. Since Plato, discussions of concepts have begun with the intuition that they might be something like definitions, an idea sometimes called the “classical theory” of concepts (e.g., Carnap, 1932; Katz, 1972; Locke, 1690; see Clark (1973) for an example from developmental psychology). When we ask, in conversation, “What is a car?” we expect an answer that differentiates cars from all other things, and thus which provides necessary and sufficient conditions for reference. In simple cases, like *car*, the answer might include an appeal to an object’s physical constitution (wheels, seats, a protective structure), its function (rapid transportation of multiple people along streets and highways), or its origin (an artifact made by humans, either by hand or by machine; see Bloom, 1996; see also Malt, *in press*). These criteria offer an account of how a word like *car* might be applied to cars (but not to motorcycles), and whether an individual is the “same car” over time. Further, and most relevant to this paper, they offer an account of what counts as one whole car – e.g., why half of a car cannot be counted as “one car”. For example, if a car is an object that fulfills a particular function, and half a car fails to support this func-

¹ We should note that this conclusion has been largely overlooked by recent work on the approximate number system and “numerical perception”, where it is often assumed that the representation of numerosity is automatic and pre-attentive, rather than a computation over conceptually restricted sets of individuals (e.g., Burr & Ross, 2008; Cantlon & Brannon, 2007; Dehaene, 1997).

tion, then half a car does not count as “one car”. Thus, in theory, definitions specified by concepts explain how we use words to pick out individual units that belong to members of kinds.

A problem for the theory that nouns act as sortals, however, and thus for theories of individuation and counting, is that finding a good definition for just about any word is notoriously difficult (Dancy, 1985; Fodor, 1981, 1995; Kripke, 1972; Putnam, 1970; Wittgenstein, 1958). Most, if not all, candidate definitions are subject to exception. For example, Fodor (1981) argues that the verb *paint* cannot mean “cover with paint” since some acts of covering with paint are not acts of painting (e.g., the explosion of a paint factory). Many paradigm cases, including concepts like *bachelor*, *game*, *truth*, and *justice*, have failed to yield agreed upon definitions, even though philosophers have discussed them for centuries. Further, even when definitions can be articulated, they often invoke concepts that are no simpler than those they seek to explain. For example, in Jackendoff’s attempt to decompose the verb *kill*, he appeals to the concepts *cause* and *die* (i.e., to kill = to cause to die; see Jackendoff, 1990). The problem is how to decide when to stop: is *die* a primitive, or can it be decomposed further (e.g., cease to live)?

These problems, and others, have led some to abandon the classical theory in favor of accounts that propose statistical prototypes (e.g., Rosch & Mervis, 1975), intuitive theories (e.g., Carey, 1985; Gelman, 2003), or even a complete absence of conceptual structure (e.g., Fodor, 1998). However, these accounts also suffer from serious explanatory limitations, which are as serious, if not more serious than those faced by the classical theory (for review, see Laurence & Margolis, 1999). Thus, although existing theories of concepts offer useful frameworks for guiding empirical investigation, none currently succeed in providing a unified account of how words pick out individual kinds of things. As a result, although most researchers agree that nouns provide conceptual criteria that guide reference and individuation, there is currently no agreed upon account of exactly how this works, and thus how nouns encode sortal structure to guide counting and quantification.

In the present paper, we explore the idea that nouns do not, in fact, act as sortals in isolation, and thus do not fully specify units for quantification. This idea, which we develop below, has its origin in the history of semantics. In the early part of the 20th century, philosophers of language including Strawson and Carnap, as well as Wittgenstein and others after him, expressed skepticism that linguistic meaning could be understood using precise definitions – e.g., using the tools of formal logic. The primary reason for their doubt was the observation that language is often imprecise and ambiguous – e.g., allowing multiple meanings to be associated with a single form (for an introduction to this issue, see Gamut, 1990; Partee, 2011). However, a key observation, made most famously by Paul Grice (see, e.g., Grice, 1969), is that a logical approach to describing linguistic meaning is tenable if such meanings are supplemented by pragmatic knowledge of how language is used to communicate. In this way, the explanatory burden of semantics is lightened: the logical meanings of sentences are not expected to fully explain their interpretation, but are pragmatically enriched in conversation. Here, we draw on Grice’s idea, and argue that nouns may not have to do the work of sortals – and thus, the work of individuation – in isolation. The explanatory burden traditionally placed upon nouns may be lightened by an appeal to the pragmatics of communication.

To understand Grice’s idea, consider the statement in (1a), which contains the English quantifier *some*. According to classical logic, this statement would be literally true in a context in which the boy had eaten all of the cookies (since eating *all* of the cookies entails eating *some* of them). However, in conversation, this sentence typically implies that the boy did not eat all of the cookies. This is because, if the boy had eaten all of the cookies, a cooperative speaker would have said so (as in 1b).

- (1) a. The boy ate some of the cookies.
- b. The boy ate all of the cookies.

By most accounts, this inference requires that the listener (1) compute a logical meaning for the sentence (using a meaning of *some* that is compatible with all), (2) consider the stronger alternative statements that could have been uttered, and (3) negate these stronger alternatives, by reasoning that a cooperative speaker would have uttered them had they been true (i.e., because they are more informative). These alternative statements would be generated by replacing the word *some* with other words from the same semantic class or scale – e.g., quantifiers like *most* and *all*. For this reason, this inference is often labeled a “scalar implicature” (for discussion, see Horn, 1972).

The upshot of this idea, which has been applied not only to quantifiers, but also to logical connectives (*or*, *and*), number words (*one*, *two*, *three*), and a host of other cases, is that semantic theory can preserve the use of concepts and definitions from formal logic if those concepts are supplemented by pragmatic inference (for related discussions, see Hirschberg, 1985; Horn, 1972; Levinson, 1983). Listeners can infer meanings that go well beyond the logical meanings of words when they consider not only what the speaker literally said, but also what they chose *not* to say.

In the present study, we consider a similar approach to the case of nominal concepts, like *shoe* and *fork*, and argue that such concepts do only part of the work required of sortal theory – i.e., specifying units of quantification and counting. Although we do not endorse a classical theory of concepts, we argue that such an approach – or indeed any theory of concepts – may have a better chance of explaining word meaning if the explanatory burden is lightened by an appeal to pragmatics. For example, if words encode definitions, they may be only partial in nature, and may be “filled out” by considering other words that interlocutors could have used, but chose not to. To explore this idea, we investigated a case study in which children, in contrast to adults, do not appear to use nouns as sortals – i.e., when counting broken objects.

In a seminal study, Shipley and Shepperson (1990) presented children with small sets of objects (e.g., forks) in which one of the objects was cut into several pieces. Whereas an adult will count a fork as “one fork” (or sometimes “not a fork”) when it is sliced into three pieces, children as old as 7 count such broken forks as “three forks”, even when the pieces are lined up carefully and spaced only an inch apart (Brooks, Pogue, & Barner, 2011; Shipley & Shepperson, 1990; for related findings in event quantification, see Wagner & Carey, 2003). This behavior persists even when children are asked specifically to count “whole” objects (e.g., “count the whole shoes”), and does not differ from when they are merely asked to “count the things” in an array (Shipley & Shepperson, 1990; Sophian & Kalihwiwa, 1998; for a related phenomenon regarding the counting of collections, see Huntley-Fenner, 1995). In fact, children will count the pieces of a broken object if it is broken in front of them, and even if they counted it as “one” only seconds earlier (Brooks et al., 2011). Strikingly, children make these errors despite easily reporting that the objects are broken. Children continue to make such errors until at least age 7 (see Sophian & Kalihwiwa, 1998; see also Gutheil, Bloom, Valderrama, & Freedman, 2004).

Subsequent studies have provided evidence that these failures are not specific to the counting routine, but also arise when children interpret other linguistic forms (Brooks et al., 2011). For example, when asked, “Who has more shoes?” in a situation in which one character had two whole shoes and another had one shoe cut up into three pieces, 4-year-olds reliably judged that the character with the broken object had more shoes. Children also treated pieces of broken objects as individuals when producing plural morphology, and when interpreting quantity words like *some*, *both*, and *every* (e.g., in response to “Touch every shoe!”). Together, these studies of quantification suggest that children’s nouns do not initially specify units of quantification, and thus do not act as sortals – at least not in the way that they appear to for adults. For children, half of a shoe, despite lacking a shoe’s shape and function, still counts as a unit of *shoe* for the purposes of linguistic quantification.

Most significant to the present study, Brooks et al. also found conditions in which children did *not* count pieces of broken objects. Specifically, when 4-year-olds were shown broken objects whose parts had labels in English and had unique functions, they excluded those parts from their counts of whole objects. For example, when shown a bicycle with its wheels removed, children excluded the wheels when asked to count the bicycles. Similarly, they excluded ears when counting rabbits, arms when counting clowns, wings when counting butterflies, and handles when counting umbrellas.² Interestingly, children’s exclusion of these parts was predicted by their ability to label them in a subsequent task. Children who correctly labeled wheels, ears, wings, etc. when shown these parts in isolation were also less likely to include these parts in their counts of whole objects (for related data regarding the counting of parts, see Giral & Bloom, 2000).

This set of findings suggests a Gricean solution to how nouns might come to act as sortals. Specifically, children’s nominal concepts may not be deficient, and may not differ substantially from those of adults: they may specify precisely the same criteria of individuation. However, just as adults use prag-

² No rabbits, butterflies, or clowns were harmed in the process of conducting these studies.

matic inference to enrich the interpretation of quantifiers and logical connectives, they may also use inference to enrich their interpretation of nouns, and to derive clear units for quantification. Specifically, a wheel may not count as a *bicycle* because *wheel* is a better description of the object part than *bicycle*. Thus, when children are asked to “count bicycles”, they may exclude wheels because, if the speaker had meant for them to count wheels, then they should have said so. But although children know that words like *wheel* and *bicycle* contrast, and that *wheel* is a better description for wheels than bicycles, they may not spontaneously access better descriptions for arbitrary parts of objects. Thus, children may count half of a shoe as “one shoe” because *shoe* is the best available description of the object.

Here, we pursue this idea, and argue that nothing fundamental changes in children’s knowledge of noun concepts between ages 4 and 7. Like 4-year-olds, older children and adults may *not* require that nouns specify a set of necessary and sufficient conditions. Instead, all speakers may rely on noun meanings that provide only partial criteria for identifying units of quantification – whether these are partial definitions, prototypes, or something else. Critically, however, adults may consider a broader set of linguistic expressions against which they can contrast their use of nouns. In particular, when generating alternatives to whole object labels, adults may not only consider the names of parts, but may also consider measure words that apply to arbitrary parts of things, such as *piece*, *part*, *half*, *slice*, *bit*, and *portion*. For an adult, half of a shoe may not count as “one shoe” because a better description of it would be “one piece of a shoe”. Thus, although pieces of shoe may fall under the denotation of *shoe* and satisfy its conditions for kind membership even for adults, “piece of shoe” may be an even better description of such objects (for a related idea, see Hobbs, Stickel, Appelt, and Martin (1993) who treats conceptual interpretation as a form of conceptual abduction – i.e., inference to the best explanation). By this account, young children may include arbitrary broken parts in their counts of whole objects because they fail to access better, alternative descriptions of these parts – e.g., measure words like *piece* and *half*.

Interestingly, a similar account has been proposed to explain why children fail to compute scalar implicatures until age 6 or 7. An extensive literature indicates that young children fail to compute scalar implicatures for a variety of scalar contrasts including *some* vs. *all* (Huang, Snedeker, & Spelke, in press; Hurewitz, Papafragou, Gleitman, & Gelman, 2006; Musolino, 2004; Noveck, 2001; Papafragou & Musolino, 2003; Smith, 1980), *might* vs. *must* (Noveck, 2001), *a* vs. *some* (Barner, Inagaki, & Li, 2009), and the distinction between inclusive and exclusive *or* (Chierchia, Crain, Guasti, Gualmini, & Meroni, 2001; Gualmini, Crain, Meroni, Chierchia, & Guasti, 2001). For example, when shown three horses that all jump over a fence, children accept a sentence like “Some of the horses jumped over the fence” to describe the event, while adults do not. However, when children are explicitly presented with a more informative description as an alternative – e.g., in a forced choice – they easily recognize that a sentence like “All of the horses jumped over the fence” is a better description (see Foppolo, Guasti, & Chierchia, 2011; see also Chierchia et al., 2001). This evidence suggests that children have difficulty with forms of pragmatic inference, like scalar implicature, because they are unable to spontaneously activate contrasting, more informative alternatives (e.g., to spontaneously activate *all* when processing *some*; for evidence and discussion, see also Barner & Bachrach, 2010; Barner, Brooks, & Bale, 2011; Stiller, Goodman, & Frank, 2011).

Our proposal is that a similar difficulty explains children’s counting of broken objects. As in the case of quantifiers and other scalar items, children may learn descriptions for arbitrary parts of objects like *piece* and *half* relatively early in acquisition, but may not spontaneously activate these descriptions as contrasting alternatives when interpreting nouns for whole objects like *shoe* and *fork*. Thus, children may fail to exclude arbitrary parts of shoes from their counts of shoes because they fail to access a better label for these objects than *shoe*. Critically, we are not suggesting that the inference young children fail to make when they count broken objects is a kind of scalar implicature. Our proposal is simply that this inference, like scalar implicature, relies on an ability to access relevant alternative descriptions, and that in both cases this ability emerges late in acquisition, around the age of 6 or 7.

The present studies test this idea by investigating how knowledge and access to part names and measure words affects children’s counting behaviors. In Experiment 1 we replicate previous findings that children count arbitrary pieces of broken objects. Also, we extend this finding using a variation of the Truth Value Judgment task (see Crain & Thornton, 1998), which has been used frequently to assess

pragmatic competence in studies of young children. In Experiment 2, we test when children begin to comprehend measure words like *piece* and *half*, a prerequisite to treating them as alternatives to descriptions using whole object labels. Also, we ask whether children succeed at mapping whole object labels and measure phrase descriptions to their referents when these descriptions are explicitly contrasted in a forced choice. For example, we test whether young children prefer measure phrase descriptions (e.g., “two pieces of shoe”) over whole object descriptions (e.g., “two shoes”) when labeling parts of broken objects. In Experiment 3, we ask whether priming measure phrases, to make them accessible as alternatives to whole object labels, leads children to exclude parts of broken objects from their counts of whole objects. Finally, in Experiment 4, we revisit the finding that younger children exclude nameable parts from their counts (e.g., wheels when counting bicycles), and explore whether this performance is truly explained by contrasting linguistic alternatives, or instead by other non-linguistic information. To do so, we teach children labels for novel objects and manipulate whether their parts have unique labels and functions.

2. Experiment 1

Experiment 1 had two goals. First, we sought to replicate previous findings that children count parts of broken objects until at least age seven, using our own stimulus set (also used in Experiments 2 and 3, below; see Appendix A). Second, we sought to probe the relation between children’s counting behavior and previously attested failures with pragmatic inference. To accomplish these two goals, children first participated in a Felicity Judgment task (see Fig. 1a), followed by a Broken Object Counting task (see Fig. 1b).

The Felicity Judgment task, which resembles the Truth Value Judgment task (see Crain & Thornton, 1998), has been used in previous studies to test children’s ability to compute conversational implicatures (e.g., Chierchia et al., 2001; Papafragou & Musolino, 2003). For example, in previous studies, children were shown a scene – e.g., in which three out of three horses jumped over a fence – and were

(a) Felicity Judgment Task



Non-measure phrase trials

*Captain Blue says he
sees four socks.*

Measure phrase trials

*Captain Blue says he
sees four pieces of sock.*

Did he do a good job or a bad job?

(b) Broken Object Counting Task



Can you count the socks?

Fig. 1. Sketches of Experiment 1 stimuli and procedures.

then asked to judge whether a puppet's logically correct but under-informative description (e.g., "Some of the horses jumped over the fence") was "good" or "bad". By previous accounts, children's failure to reject such under-informative descriptions results from their inability to generate better, alternative descriptions – e.g., "All of the horses jumped over the fence" (see Barner & Bachrach, 2010; Barner et al., 2011; Chierchia et al., 2001; Foppolo et al., 2011).

In our Felicity Judgment task, children watched a character describe an array containing broken objects, and then decided whether the character did a "good job" or "bad job". For example, on one trial, children were shown two socks that had each been cut in half. They then heard a character describe the array as "four socks" and were asked whether he did a good or bad job (see Fig. 1a). Immediately after participating in this task, children participated in the Broken Object Counting task (see Fig. 1b). Here, we presented children with small sets of objects (e.g., socks) in which one of the objects had been cut into two or three pieces. We then asked them to enumerate the sets using whole object labels (e.g., "Can you count the socks?" see Fig. 1b).

We reasoned that if children's failure to exclude pieces of broken objects from their counts of whole objects is caused by factors analogous to their problems with scalar implicature – i.e., difficulty accessing more informative alternatives (e.g., "pieces of sock") – then their performance on the Felicity Judgment task should predict their performance on the Broken Objects Counting task. Experiments 2 and 3 provide a more direct test of this idea, by exploring children's acquisition of and access to measure word descriptions.

2.1. Method

2.1.1. Participants

The participants were 41 monolingual English-speaking children (20 boys) between the ages of 3;11 and 7;6 ($M = 5;5$; $SD = 12$ months). An additional nine children participated but were excluded because they didn't finish the task ($n = 1$), responded incorrectly to the two warm-up trials of the Felicity Judgment task ($n = 5$), or responded incorrectly to two or more of the control trials that were used to evaluate their understanding of that task ($n = 3$). For each of the experiments presented in this paper, we recruited and tested a distinct sample of children. Children were recruited either by phone and brought into the lab, or at daycares and museums in the San Diego area. All children received a token gift for participating. A group of 16 students attending the University of California, San Diego also participated.

2.1.2. Materials and procedure

The Felicity Judgment task was always presented first, and the Broken Objects Counting task was always presented second. For each task two item orders were created and counterbalanced across participants.

2.1.2.1. Felicity Judgment task. Participants were first introduced to a character named Captain Blue, and were told that he would describe the objects that were placed in front of him. They were told that they had to decide whether Captain Blue did "a good job" or "a bad job". Two warm-up trials ensured that participants could judge that Captain Blue did a good job when he provided an appropriate description (e.g., when he stated that he saw an apple and an apple was present) and that he did a bad job when he provided a poor description (e.g., when he stated that he saw a ball and only a cup was present). If participants responded incorrectly to either item, they were given feedback and another chance to respond.

Each participant completed eight critical trials and four control trials. The first four critical trials were non-measure phrase trials. On these trials, participants saw sets that included broken objects and then heard Captain Blue provide descriptions that incorrectly used whole object labels – e.g., describing two socks that had each been cut in half as "four socks" (see Fig. 1a). On two of the non-measure phrase trials there was one whole object and one broken object cut into two pieces. On another trial there was one broken object cut into two pieces. On the final trial there were two broken objects that had each been cut into two pieces. After the non-measure phrase trials, participants received four measure phrase trials. These items were always administered after the non-measure

phrase trials, because we did not want to make measure phrases (e.g., “piece of sock”) available to children as they were judging the non-measure phrase trials. On the measure phrase trials, Captain Blue correctly described sets containing broken objects by using the measure words *piece* or *half* – e.g., describing two socks that had each been cut into two pieces as “four pieces of sock” (see Fig. 1a). The word *piece* was used on two of the trials, each of which contained two broken objects that had each been cut into two pieces (e.g., “four pieces of sock”), and the word *half* was used on the other two trials, each of which contained one broken object that had been cut into two pieces (e.g., “two half plates”). For both measure phrase and non-measure phrase trials, sets consisted of socks, forks, plates, or shoes (see Appendix A). In each set, the pieces of the broken object were carefully aligned and separated by approximately 2.5 cm, so that the relationship between them would be clear.

In addition to the eight critical trials, participants also received four control trials, which were interspersed with the critical trials. On these trials, Captain Blue always provided an incorrect description of a set of whole objects. On two trials, Captain Blue incorrectly described the color of the objects (e.g., saying that a set of green blocks were red), and on the other two trials, he incorrectly described the objects’ location (e.g., saying that a set of oranges that are on a table were in a box). Items included toy cars, plastic oranges, plastic lemons, and blocks. Three children were excluded for incorrectly responding to more than one of the four control trials.

2.1.2.2. Broken Object Counting task. This task was modeled after Shipley and Shepperson (1990). On each trial, a set of objects was presented to the participant, consisting of two whole objects and a third object that had been cut into either two or three pieces. The experimenter then asked the participant to count the objects on the table, using the whole object label (e.g., “Can you count the [socks]?”; see Fig. 1b). Participants completed eight trials, with two trials for each of four object kinds. As in the Felicity Judgment task, the objects were forks, socks, plates, and shoes (see Appendix A). On four of these trials, one of the objects in the array had been cut into two pieces, and on the other four trials, the object had been cut into three pieces. As in the Felicity Judgment task, the pieces of the broken object were separated by approximately 2.5 cm and aligned with one another.

2.2. Results and discussion

The dependent measure for the Felicity Judgment task was the proportion of correct responses provided, averaging across the two types of critical trials – i.e., “bad job” responses on the four non-measure phrase trials (e.g., on which Captain Blue incorrectly described two socks that had each been cut in half as “four socks”) and “good job” responses on the four measure phrase trials (e.g., on which Captain Blue correctly described two socks that had each been cut in half as “four pieces of sock”).³

Adult participants responded as expected on the Felicity Judgment task, and gave correct responses on 95% of trials. They judged that Captain Blue did a good job on measure phrase trials on 92% of trials, and a bad job on non-measure phrase trials 97% of the time (both were significantly better than the chance level of 50%; measure trials: Wilcoxon Signed-Rank $T = 0$, $n = 16$, $p < .01$; non-measure trials: $T = 0$, $n = 15$, $p < .01$). Overall, the children also performed relatively well, responding correctly on 70% of trials, which was reliably more than chance ($T = 47$, $n = 32$, $p < .01$). They judged that Captain Blue did a good job on measure phrase trials 68% of the time, and a bad job on non-measure phrase trials 71% of the time (both were better than chance; measure trials: $T = 98.5$, $n = 36$, $p < .05$; non-measure trials: $T = 92$, $n = 32$, $p < .05$). However, children responded correctly significantly less often than adults (Mann–Whitney $U = 134.5$, $n = 57$, $p < .001$). Further, although children generally performed well when all ages were considered together, the youngest children performed poorly, often treating parts as equivalent to whole objects when making felicity judgments. Specifically, 4-year-olds ($n = 14$) responded correctly on only 58% of trials.⁴ In contrast, performance was higher in the older age groups: 5-year-olds ($n = 15$) responded correctly on 74% of trials, 6-year-olds ($n = 8$) on 75% of trials, and 7-year-

³ Preliminary analyses for this experiment and for the other experiments reported here did not find significant effects of gender. We have thus excluded this factor from our analyses.

⁴ We included the performance of one child, aged 3;11, in this analysis.

olds ($n = 4$) on 84% of trials. Overall, there was a significant correlation between age and performance on the Felicity Judgment task (Pearson $r(39) = .44, p < .005$).

The dependent measure for the Broken Object Counting task was the proportion of trials on which participants counted the pieces of a broken object as whole objects. Adults never counted pieces as whole objects: on 12% of trials they excluded pieces from their counts entirely (e.g., counting two whole socks and a third broken sock as “two socks”), and on 88% of trials they counted the pieces of a broken object as a single, whole object (e.g., counting two whole socks and a third broken sock as “three socks”). In contrast, children counted object pieces as whole objects on 57% of trials, which was significantly more often than adults ($U = 276, n = 49, p < .001$). On the 43% of trials in which children did not count pieces as whole objects, they counted like adults, and counted the pieces of a broken object as a single whole object on all trials. These results are similar to those of Shipley and Shepperson (1990), who found that 3- to 6-year-olds counted the pieces of broken objects on 56% of trials when pieces were aligned and closely spaced (as they were here).

A multiple regression explored whether children’s counting behavior was predicted by their performance on the Felicity Judgment task. The dependent variable was the proportion of trials on which children counted pieces of broken objects as whole objects. The independent variables were children’s proportion of correct trials on the Felicity Judgment task, and their age (in months). This analysis revealed that performance on the Felicity Judgment task reliably predicted the counting of parts as whole objects ($B = -.81, SE = .32, p < .05$), while age was not a significant predictor ($B = .002, SE = .007, ns$).

When counting, children occasionally made other types of counting errors. Specifically, on 2% of trials, children counted a whole object in the set multiple times (e.g., by counting its undetached parts), and on 1% of trials, children counted some of the pieces of the broken object, but not all of them. To ensure that our findings were unrelated to these errors, we repeated our analysis, excluding trials on which these behaviors were present. As before, performance on the Felicity Judgment task was a significant predictor of counting behavior ($B = -.81, SE = .32, p < .05$).

Together, these results replicate previous findings that children count parts of broken objects when asked to count whole objects, and showed that children’s felicity judgments of descriptions of broken objects were a significant predictor of adult-like counting. Critical to our hypothesis, successfully making felicity judgments requires not only an ability to correctly enumerate sets, but also an ability to comprehend measure words and recognize that they are appropriate descriptions of broken objects. In Experiment 2, we explored the acquisition of measure words, and whether access to these words might allow children to exclude pieces from their counts of whole objects.

3. Experiment 2

In Experiment 1, we replicated the previously reported finding that young children – at least until the age of 7 – count pieces of broken objects (e.g., pieces of a sock) when asked to count whole objects (e.g., to “count the socks”). We also provided evidence that children’s understanding of descriptions of broken objects develops between the ages of 4 and 7. Although 4-year-olds were not readily able to make appropriate felicity judgments – e.g., to recognize that “four socks” is a bad description of two socks that had each been cut in half, and that “four pieces of sock” is a good description – older children were better at making such judgments. Most importantly, we showed that children’s ability to make such judgments is a significant predictor of their counting of broken objects. This is consistent with the idea that children count broken pieces as whole objects because they fail to access better descriptions of them – e.g., measure words like *piece* and *half*.

Experiment 2 contrasted two possible reasons that children might fail to access measure words when interpreting whole object labels. First, children may fail to access measure words because they have not yet acquired their meanings – i.e., they may lack these words altogether, or fail to understand how they contrast with whole object labels. This, for example, could explain the relatively poor performance of 4-year-olds at making felicity judgments in Experiment 1. Alternatively, children may acquire measure words early in life – e.g., by age 4 – but fail to generate them as relevant alternatives to whole object labels. That is, children may count a piece of a shoe as “one shoe” because they fail to

(a) Semantic Forced Choice

Can you point to two shoes?

(b) Verbal Forced Choice

Farmer Brown says it's two shoes, and
Captain Blue says it's two pieces of shoe.
Who said it better?

(c) Measure Word Comprehension

Can you point to half an apple?

Fig. 2. Sketches of Experiment 2 stimuli and procedures.

spontaneously activate “one piece of shoe” as an alternative, contrasting description. To explore whether children have acquired measure words by age 4, we probed 3- and 4-year-old children's interpretations of these words in Experiment 2.

If children have acquired measure words but fail to generate them as alternatives to whole object labels, then they should perform similarly to adults when measure words are provided explicitly as alternatives to whole object labels (for related evidence from the pragmatics literature, see Chierchia et al., 2001; Gualmini et al., 2001), and should do so prior to the age at which they succeed at broken object counting tasks (i.e., around age 7; Sophian & Kalihiwa, 1998). To test these predictions, Experiment 2 probed 3- and 4-year-olds' comprehension of measure words in three tasks (see Fig. 2). In the Semantic Forced Choice task, we asked children to choose which of two sets was the referent of a single description – e.g., whether “two shoes” better applied to a shoe broken into two pieces or to two whole shoes. The Verbal Forced Choice task reversed this scenario, and asked children to choose between two alternative descriptions of a single set – e.g., whether “two shoes” or “two pieces of shoe” was a better description of a shoe broken in half. Finally, in the Measure Word Comprehension task, children were presented with four images, each of which included a different quantity or portion of an object (e.g., one quadrant depicted one whole apple, another half an apple, etc.). Children were asked to indicate which of the arrays was the best referent of a quantifying expression (e.g., “Point to an apple”, “Point to half an apple”, etc.).

3.1. Method

3.1.1. Participants

The participants were 42 monolingual English-speaking children (17 boys) between the ages of 3;0–5;0. There were 21 3-year-olds ($M = 3;4$, 3;0–3;10, 10 boys) and 21 4-year-olds ($M = 4;6$, 4;0–5;0, 7 boys).

3.1.2. Materials and procedure

The Semantic Forced Choice and Verbal Forced Choice tasks were always presented first, and their order was counterbalanced between subjects. The Measure Word Comprehension task was always presented last. Two versions of each task were constructed, which varied the order in which items were presented. Item order was counterbalanced across participants for each task.

3.1.2.1. Semantic Forced Choice. On each trial, the experimenter presented the child with two sets: two identical whole objects of a single kind (e.g., two shoes) and a single object of the same kind that had been cut in half (e.g., two pieces of a shoe). As in Experiment 1, four kinds of objects were used across the critical trials: socks, forks, cups, and shoes (see Appendix A). As before, the pieces of broken objects were closely aligned with one another.

After presenting the two sets, the experimenter asked the child to indicate which set was the referent of a target quantity phrase (e.g., “Can you point to two [shoes]?”; see Fig. 2a). Four target phrases were tested: “two [shoes]”, “two half [shoes]”, “two pieces of [shoe]”, and “two whole [shoes]”. Each child completed eight trials, and responded to two trials for each of the four target expressions. The side on which the correct set was presented was randomized across trials.

3.1.2.2. Verbal Forced Choice. On this task, children were presented with a single set and were asked to choose which of two different quantity expressions best described it. The sets were identical to those used in the Semantic Forced Choice task.

Prior to the task, the child was introduced to two characters: Farmer Brown and Captain Blue. The children were told that they would be shown some objects and then would have to decide whether Farmer Brown or Captain Blue described them better. On each trial, a set consisting of either two whole objects (e.g., two shoes) or one object cut in half (e.g., two half shoes) was placed in front of the child. The child then chose between a description of the set that used a whole object label and a description that used a measure phrase – e.g., “Farmer Brown says that it’s two [shoes] and Captain Blue says that it’s two pieces of [shoe]. Who said it better?” (see Fig. 2b).

Each child completed eight trials. Four trials contrasted *whole* expressions with *half* expressions (e.g., “two whole [shoes]” vs. “two half [shoes]”), and the other four trials contrasted expressions using whole object labels with *piece* expressions (e.g., “two [shoes]” vs. “two pieces of [shoe]”). We counter-balanced whether Farmer Brown or Captain Blue provided the correct description across the trials.

3.1.2.3. Measure Word Comprehension. On each trial, the children were shown a powerpoint slide including four images organized into quadrants (see Fig. 2c). Each image depicted a different quantity or portion of a particular object type: a whole object (e.g., an apple), a half of the same kind of object (e.g., half of an apple), a small piece of the same kind of object (e.g., a slice of an apple), and a plural set containing more than one instance of the whole object (e.g., two apples). On each trial, children were asked to indicate which image was the referent of a target quantity phrase (e.g., “Can you point to half an [apple]?”; see Fig. 2c). The four phrases used were: “half an [apple]”, “a whole [apple]”, “a piece of [apple]”, and “an [apple]”. Four object types were used: eggs, tables, envelopes, and apples (see Appendix B for images of all slides). There were 16 trials, which fully crossed object type with quantity phrase. Across trials, we randomized the quadrant in which the correct quantity was depicted.

3.2. Results and discussion

Our dependent measure for each of the three tasks was the proportion of children’s responses that were correct (see Fig. 3). For the Semantic Forced Choice and Verbal Forced Choice tasks, we defined chance responding as 50%, because there were always two possible responses on these tasks. For the Measure Word Comprehension task, we defined chance responding as 50% for the items that tested understanding of the *piece* phrase, because two of the four possible answers were acceptable for these items (i.e., *piece* can apply to half of an object or to a small piece of that object). However, for the rest of the items of this task, we defined chance as 25%, because only one of the four answers could be considered correct.

The 4-year-olds demonstrated a relatively sophisticated understanding of measure phrases across the three tasks. They responded correctly reliably more often than chance in the Semantic Forced Choice task ($M = .89$, $SE = .04$; $T = 0$, $n = 20$, $p < .01$), in the Verbal Forced Choice task ($M = .79$, $SE = .05$; $T = 2.5$, $n = 18$, $p < .01$), and in the Measure Word Comprehension task, for both the *piece* items ($M = .86$, $SE = .06$; $T = 10.5$, $n = 18$, $p < .01$) as well as the other items ($M = .78$, $SE = .03$; $T = 0$, $n = 21$, $p < .001$).

Our data indicate that even 3-year-olds have some understanding of measure phrases. The 3-year-olds were reliably above chance in the Semantic Forced Choice task ($M = .65$, $SE = .04$; $T = 7.5$, $n = 21$,

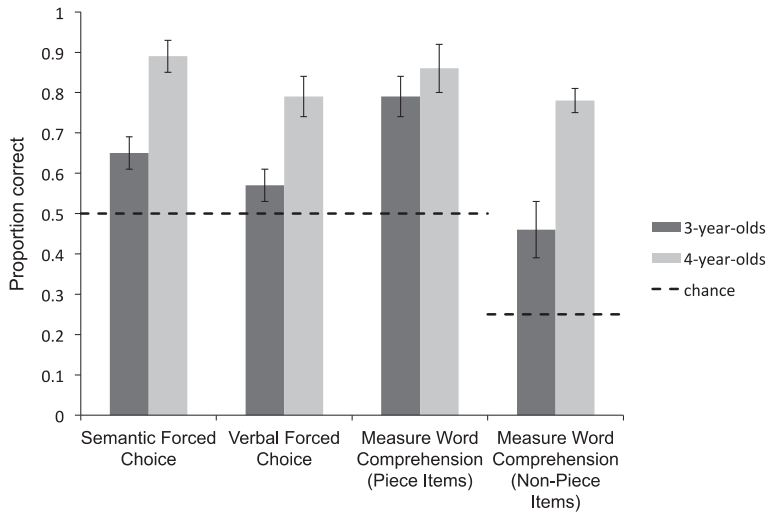


Fig. 3. Proportion of correct responses by 3- and 4-year-olds on the tasks of Experiment 2.

$p < .01$), and in the Measure Word Comprehension task—for both the *piece* items ($M = .79$, $SE = .05$; $T = 3.5$, $n = 16$, $p < .01$) as well as the other items ($M = .46$, $SE = .07$; $T = 28.5$, $n = 18$, $p < .01$). However, they did not perform reliably above chance in the Verbal Forced Choice task ($M = .57$, $SE = .04$; $T = 30$, $n = 14$, ns).

The results suggest that significant changes occur in children's understanding of measure phrases between the ages of 3 and 4. The 4-year-olds responded more accurately than the 3-year-olds in the Semantic Forced Choice task ($U = 363$, $n = 42$, $p < .001$), and while 13 out of 21 4-year-olds responded correctly to all of the items of this task, only 2 out of 21 3-year-olds did so, Chi-Square $\chi^2(1, n = 42) = 12.55$, $p < .001$. The 4-year olds also performed better than the 3-year-olds on the Verbal Forced Choice task ($U = 337.5$, $n = 42$, $p < .005$), and while 8 of 21 4-year-olds responded correctly to all of the items of this task, only 2 of 21 3-year-olds did so, $\chi^2(1, n = 42, p < .05) = 4.73$. Finally, although the 4-year-olds did not outperform the 3-year-olds on the *piece* items of the Measure Word Comprehension task ($U = 267.5$, $n = 42$, ns), they did on the other items of the task ($U = 347$, $n = 42$, $p < .001$).

These findings suggest that children's counting of parts as whole objects is not due to an ignorance of measure phrases or how they contrast with whole object labels. Four-year-olds clearly understand words like *piece* and *half* and have little difficulty mapping them to their referents. Further, when measure phrases are explicitly offered as alternative descriptions of broken objects, 4-year-olds strongly prefer them over whole object labels. Thus, although 4-year-olds have difficulty rejecting infelicitous descriptions when they are presented in isolation (e.g., to recognize that "three forks" is a bad description of three pieces of fork; see results of the Felicity Judgment task of Experiment 1), they choose better alternative descriptions when they are explicitly provided (e.g., "three pieces of fork"). This is consistent with the idea that the exclusion of broken objects from counts of whole objects depends on an ability to spontaneously access relevant, contrasting alternatives – an ability that, for both measure words and scalar words (e.g., *some* and *all*), might emerge at around the age of 7 (see Chierchia et al. (2001) for a similar pattern of results for scalar implicature). To further explore this idea, in Experiment 3 we asked whether 4- and 5-year-old children are more likely to exclude pieces from their counts of whole objects when measure word descriptions are provided in a prior task, thereby priming them for use when counting.

4. Experiment 3

Experiment 3 tested the idea that access to measure word descriptions for broken objects allows children under age 7 to count sets like adults. To explore this, we tested 4- and 5-year-old children

with a broken object counting task but preceded this with a task designed to facilitate children's access to measure phrase descriptions. Before each counting trial, children in the Forced Choice Priming condition were asked to make a series of forced choices regarding the referents of a whole object label and measure phrase (modeled after the Semantic Forced Choice task of Experiment 2). For example, children were presented with a sock and a piece of a sock, along with alternative descriptions of those objects (e.g., "Look! I have here a sock and a piece of a sock"). They were then asked, "Can you point to the sock?" and "Can you point to the piece of a sock?". After children responded to each question, they were shown a set of whole and broken socks and asked to count the socks (as in Experiment 1; see Fig. 4a). If children under age 7 fail to exclude parts of broken objects because they are unable to spontaneously consider better, alternative descriptions, then counting should become significantly more adult-like when alternative descriptions have been made salient in the context. To test this possibility we compared the counting of children in the Forced-Choice Priming condition to that of children in a Baseline condition, who did not receive the priming manipulation (see Fig. 4b).

4.1. Method

4.1.1. Participants

The participants were 64 monolingual English-speaking children (30 boys) between the ages of 4;0 and 5;11 ($M = 4;11$). A total of 32 children were tested in the Forced-Choice Priming condition ($M = 4;11$, 4;0–5;11, 14 boys), and 32 were tested in the Baseline condition ($M = 4;11$, 4;1–5;11, 16 boys). One additional child participated, but was excluded because he could not complete the task.

4.1.2. Materials and procedure

Children either participated in the Forced Choice Priming or Baseline conditions. Two item orders were constructed for each task and counterbalanced across children.

4.1.2.1. Forced Choice Priming condition. At the start of each trial, children were presented with a whole object (e.g., a sock) and an arbitrary piece of that kind of object, along with their alternative descriptions (e.g., "Look! I have here a [sock] and a piece of a [sock]"). Then, they were asked to choose which was the referent of the whole object label (e.g., "Can you point to the [sock]?"), and which was the referent of the measure phrase (e.g., "Can you point to the piece of a [sock]?"; see Fig. 4a). The order of these questions was counterbalanced across trials, and critically, children were not given feedback. After completing the forced choice, children were shown a set containing two whole objects of the same kind (e.g., socks), and a third object that was broken into two or three pieces (e.g., a sock cut

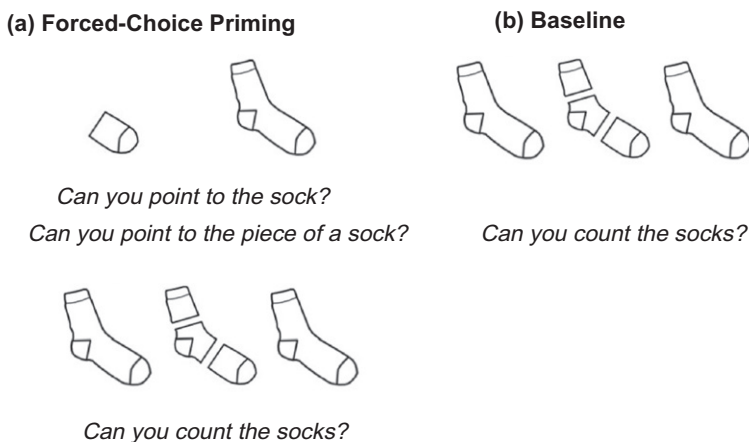


Fig. 4. Sketches of Experiment 3 stimuli and procedures.

in three). The children were then asked to count the set, using the whole object label (e.g., “Can you count the [socks]?”; see Fig. 4a). The objects were forks, shoes, and socks, and each object type was presented on two trials, for a total of six trials (see Appendix A).

4.1.2.2. Baseline condition. In this condition, children received only a counting task, which was identical to the Broken Object Counting task of Experiment 1 (see Fig. 4b). The trials were identical to those in the Forced Choice Priming condition, except that there were two additional trials, which tested sets of plates (see Appendix A). This difference arose because the two conditions were initially conceived as separate studies. Below, we report analyses both for the full sets of counting trials included in each condition, as well as for the subset of trials that were identical between conditions.

4.2. Results and discussion

At the start of the trials in the Forced Choice Priming condition – e.g., when children had to indicate the referents of “the sock” and of “the piece of a sock” – children correctly chose the referents of whole object labels 90% of the time, and of measure phrases 97% of the time. In each case, the average of correct responses was significantly higher than the chance level of 50% (whole object: $T = 19$, $n = 30$, $p < .01$; measure phrase: $T = 0$, $n = 32$, $p < .01$), consistent with the strong performance of 4-year-olds on the Semantic Forced Choice task of Experiment 2.

Our dependent measure for the counting task was the proportion of trials on which children counted the parts of a broken object as whole objects (see Fig. 5). A univariate ANOVA explored the effect of condition on part counting, with age (in months) as a covariate. Children counted parts of broken objects as whole objects on 63% of trials in the Baseline condition, but counted parts as whole objects significantly less often in the Forced Choice Priming condition, on only 37% of trials ($F(1,60) = 5.62$, $p < .05$). There was no effect of age ($F(1,60) = 2.29$, *ns*). A parallel analysis found identical results when the two additional trials in the Baseline condition were excluded from the analysis.

On the 63% of trials in which children in the Forced Choice Priming condition did not count pieces of broken objects as whole objects, they counted like adults, counting the pieces of a broken object as a single whole object on 69% of trials, and excluding pieces from their counts entirely on 31% of trials. When children in the Baseline condition did not count pieces as whole objects (on 37% of trials), they counted the pieces as a single whole object on 75% of trials, and excluded the pieces altogether on 25% of trials. Finally, just as in Experiment 1, we coded for whether children made other types of counting

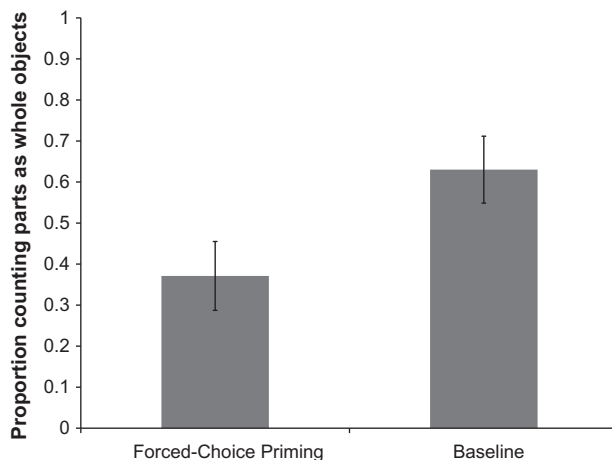


Fig. 5. Proportion of trials on which children counted pieces of broken objects as whole objects in Experiment 3.

errors during the counting task. Although children never counted a whole object multiple times, or failed to count an object altogether, children did count some but not all of the pieces of the broken object on 0.7% of trials. We repeated our analysis excluding these error trials and found the same results.

In sum, asking children to use their existing knowledge to categorize objects as either parts or wholes – and thus to access alternative labels for these things – was sufficient to significantly increase adult-like counting of sets in a subsequent task. These results are consistent with the idea that children count like adults, and exclude parts from their counts of whole objects, when they are better able to access alternative part descriptions. Still, it is worth noting that children who received our priming manipulation did not exclude parts nearly as consistently as adults do (e.g., adults in Experiment 1 never counted pieces as whole objects). Thus, although the forced choice task may have made both measure phrases and whole object labels accessible to children, some children may still not have treated these descriptions as relevant, contrasting alternatives. As noted by Brooks et al. (2011), simply accessing descriptions, in isolation, may not be sufficient for recognizing them as relevant, contrasting alternatives. This knowledge may also require experience hearing expressions contrasted, or may depend on a developing ability to attend to relationships between sentences in the discourse (e.g., their relevance relations; for discussion see Grice, 1969; Levinson, 1983).

Critically, our priming manipulation did not train children to respond correctly – i.e., children were not given feedback regarding which objects were pieces and which were wholes. Thus, it is unlikely that children's *conceptual* representations (e.g., of shoes, forks, socks, etc.) changed during the experiment. Instead, children's nominal concepts may be quite broad – so broad that they allow arbitrary parts of objects to be included as members of object kinds. Children may restrict their broad concepts and count like adults only when they access better, contrasting descriptions of parts. Consequently, our results raise the possibility that adult nominal concepts are also broad, and do not provide criteria for excluding arbitrary parts from object kinds in the absence of contrasting part labels. We return to this idea in the General Discussion.

5. Experiment 4

Together, Experiments 1 through 3 provide evidence that young children exclude arbitrary parts of broken objects from their counts of whole objects when they can access better alternative descriptions for them – i.e., measure phrases. Accessing part descriptions could be especially important for the exclusion of arbitrary parts, like the pieces of a shoe. Although arbitrary shoe pieces may not form a natural, cohesive, non-linguistic category, measure words allow any arbitrary grouping or portion of stuff to be labeled and counted as a kind of individual – e.g., as “pieces of shoe”. By individuating these units and categorizing them as distinct from whole objects, measure words may provide a mechanism for excluding parts from counts of whole things.

Although breaking objects into parts can result in arbitrary pieces, it can also result in non-arbitrary pieces that have their own labels in English (e.g., wheels, arms, ears, etc.). Recall that in the study by Brooks et al. (2011), 4-year-olds did not count nameable parts of familiar objects as whole objects (e.g., wheels as bicycles). On the face of it, these data are consistent with the idea that excluding a part depends on accessing a better description: children could have inferred that wheels shouldn't be included in a count of bicycles, because the best description of these things is *wheels* not *bicycles*. However, it's also possible that children excluded wheels independent of their labels, because they have a discernable function of their own and thus belong to their own artifact kind (see, e.g., Keil (1989), Kemler Nelson (1995, 1999) and Malt and Johnson (1992) for evidence that functional information plays a role in defining artifact kinds). Indeed, although Brooks et al. (2011) found a broad correlation between the labeling of parts and their exclusion from counts of whole objects, children correctly labeled parts less often than they excluded them. Thus, children may be more likely to learn names for parts – *and* exclude them from their counts of whole objects – when they have distinct functions.

Consistent with this idea, Kemler Nelson and colleagues have shown that when children as young as 2 learn a label for an artifact kind, they accept the use of that label to refer to an artifact whose func-

tion has been accidentally disabled (e.g., because the object is broken), but *not* to refer to an artifact whose function has been intentionally altered (Kemler Nelson, Herron, & Morris, 2002; Kemler Nelson, Holt, & Egan, 2004; but see Gutheil et al., 2004). Thus, although objects may remain category members when they are broken and dysfunctional, they are seen to belong to distinct artifact kinds when their functions are intentionally transformed. Together, these results, and those of Brooks et al., suggest that excluding parts from counts of whole objects may depend on recognizing that parts belong to different object kinds than whole objects – an inference that could be signaled either by accessing contrasting labels for parts (demonstrated in Experiments 1 through 3) or by recognizing their unique functions (see also Booth & Waxman, 2002; Waxman & Markow, 1995).

Experiment 4 explored this idea by teaching 3- to 5-year-olds labels for novel objects, and manipulating whether their parts were given their own labels and functions. After teaching children these words, we asked them to count using the novel whole object label, as in the preceding experiments (see Fig. 6). In the first two conditions, the functions of the novel objects were demonstrated, and the objects were then broken into functional parts. In one of these two conditions, the functional parts also received unique labels (the Labeled Functional Parts condition) and in the other they did not (the Unlabeled Functional Parts condition). In the third and fourth conditions, the functions of novel objects were not demonstrated, and these objects were broken into *arbitrary* parts, which either received their own labels (Labeled Arbitrary Parts condition) or did not (Unlabeled Arbitrary Parts condition). Finally, in the fifth condition, novel objects were broken into functional parts, but the functions of these objects were not explicitly demonstrated (Undemonstrated Functional Parts condition). This condition was motivated by previous research which suggests that the functions of objects can be inferred on the basis of their physical affordances alone. For example, Prasada, Ferez, and Haskell (2002) found that subjects are more likely to perceive novel stimuli as functional – and categorize them as kinds of objects – if they have complex, regular, structures (see also Barner et al., 2009; Li, Dunham, & Carey, 2009).



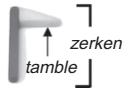







<u>Training</u>		<u>Counting</u>
(a) Unlabeled Functional Parts		<p><i>Look! This is a zerken! A zerken is for stirring juice [demonstrate]. Remember, this is a zerken!</i></p>  <p>Can you count the <i>zerkens</i>?</p>
(b) Labeled Functional Parts		<p><i>Look! This is a zerken! A zerken is for stirring juice [demonstrate]. And do you see this part? This is a tamble! Remember, this is a zerken, and this is a tamble!</i></p>  <p>Can you count the <i>zerkens</i>?</p>
(c) Unlabeled Arbitrary Parts		<p><i>Look! This is a zerken! I'm going to put the zerken right here. Remember, this is a zerken!</i></p>  <p>Can you count the <i>zerkens</i>?</p>
(d) Labeled Arbitrary Parts		<p><i>Look! This is a zerken! And do you see this part? This is a tamble! Remember, this is zerken, and this is a tamble!</i></p>  <p>Can you count the <i>zerkens</i>?</p>
(e) Undemonstrated Functional Parts		<p><i>Look! This is a zerken! I'm going to put the zerken right here. Remember, this is a zerken!</i></p>  <p>Can you count the <i>zerkens</i>?</p>

Fig. 6. Sketches of Experiment 4 stimuli and procedures.

5.1. Methods

5.1.1. Participants

The participants were 80 monolingual English-speaking children (33 boys) between the ages of 3;3–5;4 ($M = 4;3$). Of these, 16 children participated in the Labeled Functional Parts condition ($M = 4;4$, 3;6–5;3, 7 boys), 16 in the Unlabeled Functional Parts condition ($M = 4;1$, 3;3–5;4, 5 boys), 16 in the Labeled Arbitrary Parts condition ($M = 4;3$, 3;7–5;2, 3 boys), 16 in the Unlabeled Arbitrary Parts condition ($M = 4;4$, 3;4–5;4, 9 boys), and 16 in the Undemonstrated Functional Parts condition ($M = 4;4$, 3;4–5;2, 9 boys). An additional seven children participated, but did not complete the task.

Finally, because this experiment included a number of 3-year-olds, and because 3-year-olds often fail to understand the principles of counting (Wynn, 1990), we assessed children's counting abilities using Wynn's "Give-a-number" task (see Wynn, 1990) after they had completed the Broken Object Counting task. A total of 32 children (Mean age = 3;8) were excluded for failing to give the correct number of objects for requests of 5–8 objects. Thus, by Wynn's criteria, all of the children included in the study were "Counting Principle Knowers".

5.1.2. Materials and procedure

Children were tested in one of five conditions: (1) Unlabeled Functional Parts, (2) Labeled Functional Parts, (3) Unlabeled Arbitrary Parts, (4) Labeled Arbitrary Parts, or (5) Undemonstrated Functional Parts. The same novel objects were used in each condition (see Appendix C). There were six critical trials in each condition. There were two item orders.

5.1.2.1. Unlabeled Functional Parts condition. This condition tested whether children exclude functional parts from their counts of whole novel objects when they have not learned labels for those parts. Prior to each of the six trials, the child was presented with a novel object which was labeled by the experimenter, who then demonstrated its function – e.g., "Look! This is a *zerken*! A *zerken* is for stirring juice (while demonstrating with a cup of juice). Remember, this is a *zerken*!" Critically, the functions of each of the object's parts were evident in this demonstration – e.g., one part of the *zerken* extended into the juice, and the other part rested on top of the cup (see Fig. 6a; see Appendix C for information about the other items). The experimenter then laid out an array that included two or more instances of the novel object. One of those objects had been broken into two parts (on three trials) or three parts (on the other three trials), and each of those parts were functional (see Fig. 6a; Appendix C). The experimenter then asked the child to count the set, using the label for the whole novel object – e.g., "Can you count the *zerkens*?" (see Fig. 6a). Across the six trials, the novel labels used were *zerken*, *tupa*, *tibbit*, *modi*, *rapple*, and *blicket*.

5.1.2.2. Labeled Functional Parts condition. In this condition, we tested whether children would be more likely to exclude functional parts from their counts of whole novel objects when they had learned labels for those parts. The materials and procedure were identical to the Unlabeled Functional Parts condition, except that labels were assigned not only to the whole objects, but also to their functional parts. On each trial, after the experimenter labeled the novel object and demonstrated its function, she pointed to one of its parts and added – e.g., "And do you see this part? This is a *tamble*" (see Fig. 6b). The experimenter then repeated the whole object and part labels, while pointing at their referents – e.g., "Remember, this is a *zerken*, and this is a *tamble*." The novel labels used for parts were *tamble*, *zivvin*, *fengle*, *feppet*, *zuni*, and *toma*.

5.1.2.3. Unlabeled Arbitrary Parts condition. In this condition, we tested whether children would exclude arbitrary parts from their counts of whole novel objects when they had not learned labels for these parts or functions for the whole objects. At the start of each trial, the experimenter showed the child a novel object and provided a label for that object, but did not describe its function: e.g., "Look! This is a *zerken*! I'm going to put the *zerken* right here. Remember, this is a *zerken*!" The experimenter then laid out an array of whole objects and objects that had been broken into arbitrary parts and asked the child to quantify it (see Fig. 6c; Appendix C). Arbitrary parts were defined as parts that did not themselves have natural boundaries on unbroken objects and did not have their own apparent function. For example, in the case of a familiar object like a bicycle, cutting the object in half down the

middle would result in two arbitrary parts, whereas cutting off wheels, handlebars, the seat, etc., would result in functional parts.

5.1.2.4. Labeled Arbitrary Parts condition. In this condition, we tested whether children would exclude arbitrary parts when they had learned unique labels for those parts. The materials and procedure were identical to the Unlabeled Arbitrary Parts condition, except that labels were assigned not just to the whole novel objects, but also to one of their arbitrary parts (see Fig. 6d). On each trial, after the experimenter showed the novel object to the participant and labeled it, she pulled it apart and pointed to one of its arbitrary parts and labeled it – e.g., “Do you see this part? This is a *tamble*.” The experimenter then repeated the labels for the whole object and the arbitrary part while pointing to each – e.g., “Remember, this is a *zerken*, and this is a *tamble*.” The novel labels were the same as those used in the Labeled Functional Parts condition.

5.1.2.5. Undemonstrated Functional Parts condition. In this condition, we tested whether children would exclude functional parts from their counts of whole novel objects, even if they had not explicitly learned about the functions of the objects and their parts. The materials and procedure were identical to the Unlabeled Functional Parts condition, except that the experimenter did not describe or demonstrate the functions of the novel objects (see Fig. 6e). At the start of each trial, the experimenter showed the child a novel object and provided a label for that object, but did not describe its function: e.g., “Look! This is a *zerken*! I’m going to put the *zerken* right here. Remember, this is a *zerken*!” The experimenter then laid out an array of whole objects and objects that had been broken into functional parts, and asked the child to count it.

5.2. Results and discussion

Our dependent measure was the proportion of trials on which children counted the parts of a broken novel object as whole objects (see Fig. 7). A univariate ANOVA explored the effect of condition on part counting, with age as a covariate. The test yielded a significant main effect of condition ($F(4, 74) = 2.51, p < .05$), but no effect of age ($F(1, 74) = .96, ns$). Six planned contrasts further explored the effects of condition on part counting. Four of these contrasts treated the Unlabeled Arbitrary Parts condition as a baseline measure of part counting (e.g., because it is parallel to the paradigm case of

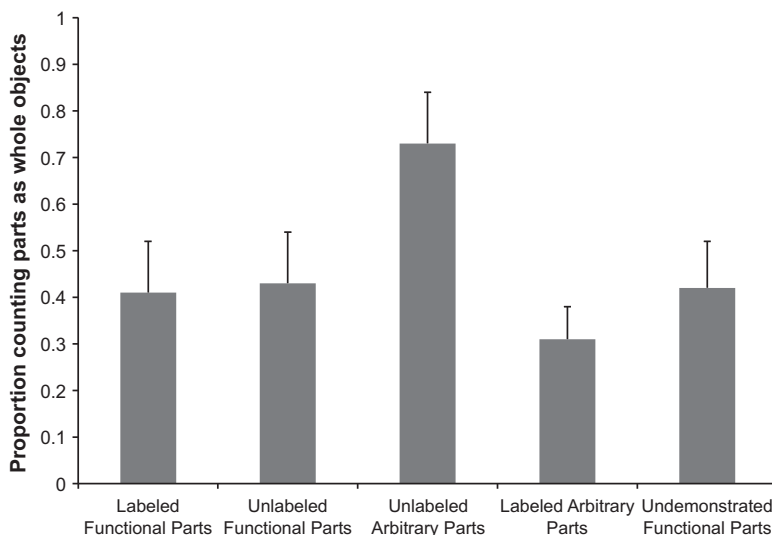


Fig. 7. Proportion of trials on which children counted pieces of broken objects as whole objects in Experiment 4.

counting a set of whole objects and arbitrary object parts), and compared counting behavior in that condition to counting in the other four conditions. The other two contrasts explored the role of labels and explicit functional knowledge in the exclusion of functional parts.

These analyses, described below, indicate that children more often exclude functional parts than arbitrary parts, and do so whether or not they have learned labels for functional parts. Children more often counted parts as whole objects in the Unlabeled Arbitrary Parts condition ($M = .73$, $SE = .11$), in which the parts were arbitrary and unlabeled, than in the Unlabeled Functional Parts condition ($M = .43$, $SE = .11$) or in the Labeled Functional Parts condition ($M = .41$, $SE = .11$), in which the parts were functional (Labeled Functional Parts vs. Unlabeled Arbitrary Parts: $t(74) = 1.96$, $p = .05$; Unlabeled Functional Parts vs. Unlabeled Arbitrary Parts: $t(74) = 2.25$, $p < .05$). These results are consistent with those of Brooks et al. (2011), who showed that children more often exclude functional, nameable parts of familiar objects (e.g., wheels) than arbitrary parts (e.g., pieces of a shoe).⁵ Our findings extend these previous results by showing that functional parts like wheels can be excluded whether or not they have received unique labels: children were no better at excluding functional parts when they had learned their labels than when they had not (Labeled Functional Parts vs. Unlabeled Functional Parts, $t(74) = .31$, *ns*). Thus, children may exclude parts that have labels more frequently not because they have labels, *per se*, but because they clearly belong to their own unique kinds. Parts that clearly belong to their own categories may be both more likely to be counted separately, and more likely to be lexicalized in language.

Our results also indicate that children exclude functional parts even when they have not explicitly learned their functions: children counted parts as whole objects more often in the Unlabeled Arbitrary Parts condition than in the Undemonstrated Functional Parts condition ($M = .42$, $SE = .10$; $t(74) = 2.20$, $p < .05$). Indeed, children were no better at excluding functional parts when they had explicitly learned their functions, compared to when they had not (Unlabeled Functional Parts vs. Undemonstrated Functional Parts: $t(74) = .06$, *ns*). This is consistent with previous evidence that parts can be deemed functional – and members of their own object kinds – on the basis of their physical affordances alone (see Prasada et al., 2002; see also Barner et al., 2009; Li et al., 2009). It is also consistent with previous findings that perceptual cues like object shape provide cues to kind membership (see, e.g., Bloom, 2000; Landau, Smith, & Jones, 1988).

Finally, our findings provide evidence that labels are sufficient for the exclusion of arbitrary parts: children counted parts as whole objects more often in the Unlabeled Arbitrary Parts condition than in the Labeled Arbitrary Parts condition ($M = .31$, $SE = .07$; $t(74) = 2.85$, $p < .01$). This is consistent with the results from Experiments 1 to 3, which suggest that generating a unique description for an arbitrary part (e.g., “three pieces of fork”) supports its exclusion.

Across the five conditions, when children counted like adults they either combined the pieces of an object to treat them as a whole (93% of trials), or excluded them from their counts entirely (7%). Children also made counting errors on 18% of trials, in which they counted a whole object multiple times (13%), failed to count a whole object altogether (4%), or counted some but not all pieces of the broken object (1%). These errors were more frequent than in Experiments 1 and 3, presumably because children were unsure how to individuate the novel stimuli. To ensure that our findings were not driven by these errors, all analyses reported above were repeated with these error trials excluded, and the identical patterns of significant and non-significant effects were found.

In sum, Experiment 4 suggests that children likely use multiple sources of information to exclude parts. In particular, while children may exclude parts when their labels contrast with those of whole objects – e.g., *piece of shoe vs. shoe*, *toma vs. blicket*, etc. – they can also exclude parts when they lack unique labels, but have obvious distinct functions. Although language may be one clue to category membership, functional information may also differentiate object kinds, even in absence of labels (see Booth & Waxman, 2002). Thus, the existence of part labels may restrict counting not because labels themselves are critical, but because they indicate the existence of distinct artifact kinds, which can be contrasted with the kinds denoted by whole object labels to define units of quantification.

⁵ In their study, children excluded functional parts at a higher proportion ($M = .88$) than did children here, probably due to greater experience with familiar objects like wheels.

6. General discussion

We began this study with Frege's observation that, for the purposes of counting and mathematical reasoning, units must be specified by concepts in a definite manner, and must not permit any "arbitrary division" into parts. Children, we noted, fail to respect Frege's principle, and often count arbitrary parts of objects as though they are instances of whole object kinds. For children as old as 7, a shoe cut into three pieces is "three shoes", and is therefore "more shoes" than two whole shoes (Brooks et al., 2011; Shipley & Shepperson, 1990). In our experiments, we sought to understand children's behaviors and their significance to sortal theory, and in particular, how nouns specify units for counting and quantification. Our hypothesis was that children's errors may not reflect immature nominal concepts, but may instead be the product of a less accessible lexical inventory. In particular, children may count a piece of a shoe as "one shoe" because they are unable to access better alternative descriptions of such objects, like "piece of shoe." Thus, calling a piece of a shoe "one shoe" may be consistent with the conceptual criteria provided by *shoe* even for adults, but may not be the best description available, given an adult lexicon that includes measure words like *piece* and *part*. Consequently, the intuition that nouns supply necessary and sufficient conditions for defining units, as expressed by Frege, may reflect the structure of the lexicon as a whole, rather than the conceptual content of individual nouns like "shoe" and "fork".

In support of this idea, we showed that although children count pieces of broken objects and accept whole object labels to describe them (Experiment 1), they strongly prefer measure word descriptions when they are explicitly provided (Experiment 2). For example, when shown an object broken into two pieces, 4-year-olds readily judged that "two pieces of shoe" was a better description than "two shoes" (Experiment 2), although they often judged that "two shoes" was a felicitous description in isolation (Experiment 1). This suggests that children's errors stem not from a misunderstanding of measure phrases but instead from an inability to generate them as alternatives to whole object labels. Indeed, children were significantly more likely to exclude pieces from their counts of whole objects when they had first judged the meanings of whole object and measure phrase descriptions – making such descriptions more accessible during counting (Experiment 3). Finally, by teaching children names for novel objects, we showed that although children often count unlabeled, arbitrary parts of objects as whole objects, they are less likely to count parts that have their own functions, or that have their own labels (Experiment 4). Part labels may allow children to exclude parts from counts of whole objects because they signal the presence of object kinds that are distinct from those that are encoded by whole object labels. Such categories, however, can also be inferred by observing a part's unique function, its physical affordances and shape, the intentions of its creator, and so on (see, e.g., Bloom, 1996, 2000; Booth & Waxman, 2002; Gutheil et al., 2004; Kemler Nelson et al., 2002, 2004; Landau et al., 1988; Prasada et al., 2002; Waxman & Markow, 1995).

Our studies help explain a long-standing puzzle in the counting literature, and begin to address fundamental questions regarding the nature of lexical concepts. Specifically, how do we reconcile our intuition, expressed by Frege, that lexical concepts should supply necessary and sufficient conditions for identifying units, with the widespread recognition that such conditions are almost always impossible to specify? Why is it so hard to specify the structure of concepts, given the strength of our intuitions – e.g., that half a shoe is surely not a shoe? One possibility is that concepts aren't alone responsible for these intuitions. But if they aren't, then what is? Our proposal, drawing on Grice's theory of pragmatics, is that concepts do not act as sortals – and thus, do the work of individuation – in isolation. The explanatory burden traditionally placed upon a theory of concepts – be it a classical theory, prototype theory, or intuitive theory – is lightened by an appeal to human pragmatic reasoning, and the ability consider not only what a speaker says, but also the alternative expressions that they could have said, but did not.

An important consequence of this claim, and of our findings, is that children's early counting behaviors cannot be explained by a change in nominal concepts, like *shoe* and *fork*. Adult concepts may not be so different from those of 4- and 5-year-old children: they may appear different only because adults interpret words in pragmatically sophisticated ways, and have access to a full inventory of lexical alternatives. The most direct evidence that children's concepts are adult-like is that their counting becomes more like that of adults when their access to alternative descriptions is facilitated. If children's early counting behaviors – i.e., the counting of pieces as whole objects – were explained

by deficient nominal concepts, we would not expect the results of Experiment 3, where priming alternative descriptions changed how children interpreted whole object labels. Rather, if adult-like counting required conceptual change, children should have needed extensive training to overcome their errors, which they did not receive.⁶ An appeal to deficient concepts also has difficulty explaining past findings. For example, when Brooks et al. (2011) presented 4-year-old children with broken objects and asked children if there was anything wrong with them, children almost always stated that they were broken. Children clearly understand that half a shoe is a deficient kind member, but nevertheless count it as “one shoe” when lacking a better alternative description.

A second consequence of our findings is that children’s reasoning about alternatives is importantly tied to – but not limited to – language. Although we couched our proposal in the Gricean theory of conversational implicature – which treats linguistic utterances as the alternatives over which reasoning is conducted – we were concerned primarily with the structure of sortal concepts, which need not be thought of as purely linguistic in nature. In particular, while sortals may be derived pragmatically by contrasting nominal descriptions with one another, they can also be derived by contrasting concepts for which children do not yet have labels. Thus, counting may be restricted not only by children’s ability to consider alternative linguistic descriptions, but also by their ability to categorize objects and their parts as members of distinct artifact kinds, on the basis of non-linguistic criteria. This conclusion is supported by Experiment 4, in which children excluded unlabeled parts from their counts of whole objects when those parts had distinct functions (or physical properties that suggested distinct functions). However, in keeping with previous studies (e.g., Waxman & Markow, 1995), Experiment 4 also found that labels invite children to infer conceptual categories. When arbitrary pieces of novel objects received unique labels, children excluded them from their counts of whole objects, despite failing to exclude them when they were unlabeled. In sum, whether suggested by labels or inferred non-linguistically, concepts are contrasted and enriched pragmatically to do the work of sortals and define units of quantification. Concepts may be contrasted to not only allow objects to be distinguished from their parts – e.g., by contrasting *shoe* and *piece of shoe* – but, by extension, to allow different kinds of objects to be distinguished from one another – e.g., by contrasting *bowl* and *plate*.

Left unsettled by this study, and indeed by studies of pragmatic development more generally, is why children’s ability to reason about alternatives changes as slowly as it does, and only becomes adult-like by the age of 6 or 7. Strikingly, children’s ability to count broken objects in an adult-like fashion emerges around the same time that they begin to reliably derive scalar implicatures – e.g., to recognize that “The boy ate some of the cookies” likely implies that he did *not* eat “all of the cookies” (see references in the Introduction). If, as we have argued, these two abilities are related, and each relies on computing inferences over alternative descriptions, the question remains as to why such inferences are difficult for young children, and so late to emerge in development.

One possibility is that children are able to reason about alternatives from an early age, but fail to do so for specific classes of words – e.g., for quantifiers like *some* and *all*, and for object descriptions like *piece of shoe* and *shoe* – because they fail to group these words into common classes of alternatives (see Barner & Bachrach, 2010; Barner et al., 2011). In support of this idea, previous evidence suggests that children are readily able to compute Gricean inferences over words they are likely to learn as being part of a common class. For instance, Wynn (1992) has argued that even 2- and 3-year-olds draw inferences about the meanings of number words by treating them as contrasting alternatives. In her study, children who knew the exact meaning of only *one* were shown one object (e.g., a balloon) next to a larger set (e.g., three balloons) and were asked to indicate the larger set – e.g., “Point at *three* balloons”. Despite not yet knowing the meaning of the larger number word, children were readily able to select the larger set, despite choosing randomly when asked to “Point at *blicket* balloons”. Thus, children may have restricted the meaning of *three* but not the meaning of *blicket* because they treated *three* – but not *blicket* – as an alternative to *one*.

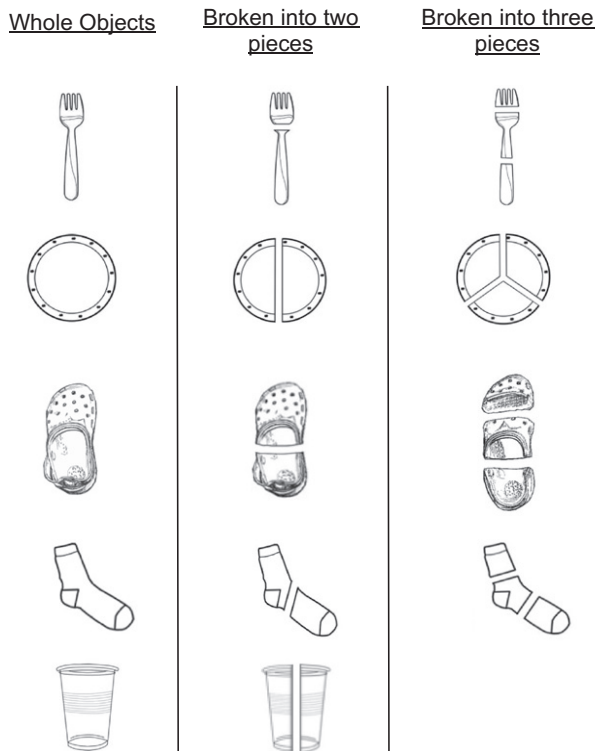
⁶ Although our results indicate that a conceptual change is not necessary to restrict reference to whole objects, they cannot rule out the possibility that such a change nonetheless occurs, and thus, that adult concepts are restricted to whole objects. We argue only that it is more parsimonious to assume that both children and adults’ concepts are broad – including whole objects and their arbitrary parts – and are only narrowed pragmatically.

Children may readily recognize that number words are alternatives because they learn, early in life, that such words are part of a common class – i.e., when they memorize the count list “1, 2, 3, 4 . . .” around age two (for additional evidence of this, see [Shatz \(1993\)](#), who argues that children group number, color, and time words into classes before acquiring meanings for each; see also [Tare, Shatz, & Gilbertson, 2008](#)). In contrast, children are never explicitly taught that quantifiers like *some* and *all* – or object descriptions like *shoe* and *piece of shoe* – are relevant alternatives. This could explain why children fail to compute inferences over such categories until later in life. Learning what counts as a relevant alternative to a particular description may involve not just learning the meanings of individual words and phrases, but also experience hearing alternatives used contrastively in conversation.

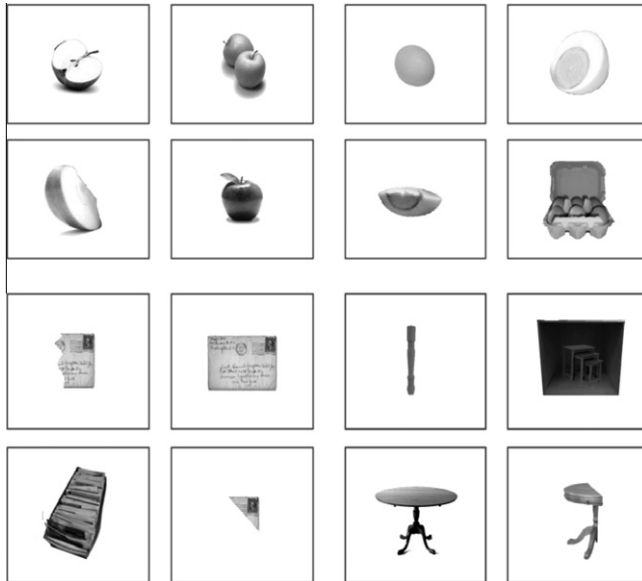
Acknowledgments

We thank Neon Brooks, Pierina Cheung, Katherine Kimura, Amanda Pogue, Jessica Sullivan, Nathan Winkler-Rhoades, members of the Language and Development Lab, and the reviewers and editor for discussion and comments on earlier drafts. We also thank Karen Bejar, Heather Halford, Lauren Havens, and Shirlene Wade for assistance with participant recruitment and data collection. Additionally, we thank Sean Barner for assistance with creating the images for our figures. We are grateful to the Reuben Fleet Science Center and to the families of children in the San Diego area. This work was supported by a grant from the James S. McDonnell Foundation to DB.

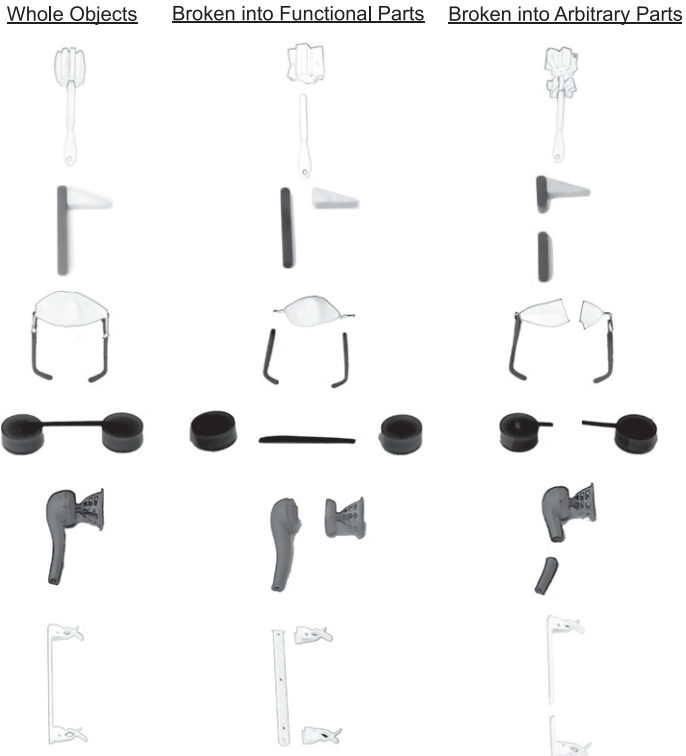
Appendix A. Sketches of stimuli used in Experiments 1, 2, and 3



Appendix B. The four sets of slides used in the Measure Word Comprehension task of Experiment 2



Appendix C. Sketches of stimuli used in Experiment 4



References

- Barner, D., & Bachrach, A. (2010). Inference and exact numerical representation in early language development. *Cognitive Psychology*, 60, 40–62.
- Barner, D., Brooks, N., & Bale, A. (2011). Accessing the unsaid: The role of scalar alternatives in children's pragmatic inference. *Cognition*, 188, 87–96.
- Barner, D., Inagaki, S., & Li, P. (2009). Language, thought, and real nouns. *Cognition*, 111, 329–344.
- Bloom, P. (1996). Intention, history, and artifact concepts. *Cognition*, 60, 1–29.
- Bloom, P. (2000). *How children learn the meanings of words*. Cambridge: MIT Press.
- Booth, A. E., & Waxman, S. (2002). Object names and object functions serve as cues to categories for infants. *Developmental Psychology*, 38(6), 948–957.
- Brooks, N., Pogue, A., & Barner, D. (2011). Piecing together numerical language: Children's use of default units in early counting and quantification. *Developmental Science*, 14, 44–57.
- Burr, D., & Ross, J. (2008). A visual sense of number. *Current Biology*, 6(18), 425–428.
- Cantlon, J. F., & Brannon, E. M. (2007). How much does number matter to a monkey? *Journal of Experimental Psychology: Animal Behavior Processes*, 33, 32–41.
- Carey, S. (1985). *Conceptual change in childhood*. Cambridge: MIT Press.
- Carey, S. (2009). *The origin of concepts*. Oxford: Oxford University Press.
- Carnap, R. (1932). The elimination of metaphysics through logical analysis of language. *Erkenntnis*, 60–81.
- Chierchia, G., Crain, S., Guasti, M. T., Gualmini, A., & Meroni, L. (2001). The acquisition of disjunction: Evidence for a grammatical view of scalar implicatures. In A. H.-J. Do, L. Domingues, & A. Johanesen (Eds.), *Proceedings of the 25th Boston University conference on language development* (pp. 157–168). Somerville, MA: Cascadia Press.
- Clark, E. V. (1990). On the pragmatics of contrast. *Journal of Child Language*, 17, 417–431.
- Clark, E. V. (1987). The principle of contrast: A constraint on language acquisition. In B. MacWhinney (Ed.), *Mechanisms of language acquisition* (pp. 1–33). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Clark, E. V. (1973). What's in a word? On the child's acquisition of semantics in his first language. In T. E. Moore (Ed.), *Cognitive development and the acquisition of language* (pp. 65–110). New York: Academic Press.
- Crain, S., & Thornton, R. (1998). *Investigations in universal grammar: A guide to experiments in the acquisition of syntax and semantics*. Cambridge, MA: MIT Press.
- Dancy, J. (1985). *Introduction to contemporary epistemology*. Oxford: Blackwell.
- Dehaene, S. (1997). *The number sense: How the mind creates mathematics*. New York: Oxford University Press.
- Fodor, J. A. (1981). The present status of the innateness controversy. In *Representations: Philosophical essays on the foundations of cognitive science*. Cambridge, MA: MIT Press.
- Fodor, J. A. (1998). *Concepts: Where cognitive science went wrong*. Oxford University Press.
- Foppolo, F., Guasti, M. T., & Chierchia, G. (2011). Scalar implicatures in child language: Give children a chance. *Language Learning and Development*.
- Frege, G. (1884/1980). *The foundations of arithmetic* (J. L. Austin, Trans.) (2nd revised ed.). Evanston, Illinois: Northwestern University Press.
- Gamut, L. T. F. (1990). *Logic, language and meaning*. Chicago: University of Chicago Press.
- Geach, P. T. (1962). *Reference and generality*. Ithaca, NY: Cornell University Press.
- Gelman, S. A. (2003). *The essential child: Origins of essentialism in everyday thought*. Oxford University Press.
- Giralt, N., & Bloom, P. (2000). How special are objects? Children's reasoning about objects, parts, and holes. *Psychological Science*, 11, 497–501.
- Grice, P. (1969). Utterer's meaning and intensions. *The Philosophical Review*, 68, 147–177.
- Gualmini, A., Crain, S., Meroni, L., Chierchia, G., & Guasti, M. T. (2001). At the semantics/pragmatics interface in child language. In *Proceedings of semantics and linguistic theory XI*. Ithaca, NY: CLC Publications, Department of Linguistics, Cornell University.
- Gutheil, G., Bloom, P., Valderrama, N., & Freedman, R. (2004). The role of historical intuitions in children's and adults' naming of artifacts. *Cognition*, 91, 23–42.
- Hirschberg, J. (1985). A theory of scalar implicature. Doctoral dissertation, University of Pennsylvania, Philadelphia, PA.
- Hobbs, J. R., Sticker, M. E., Appelt, D. E., & Martin, P. (1993). Interpretation as abduction. *Artificial Intelligence*, 63, 69–142.
- Horn, J. (1972). On the semantic properties of the logical operators in English. Doctoral dissertation, UCLA, Los Angeles, CA.
- Huang, Y., Snedeker, J., & Spelke, E. (in press). What exactly do numbers mean? *Language Learning and Development*.
- Huntley-Fenner, G. N. (1995). The effect of the whole object bias on preschoolers' understanding of collective nouns. In *Proceedings of the twenty-seventh child language research forum*, Stanford, CA.
- Hurewitz, F., Papafragou, A., Gleitman, L., & Gelman, R. (2006). Asymmetries in the acquisition of numbers and quantifiers. *Language Learning and Development*, 2, 77–96.
- Jackendoff, R. (1990). *Semantic structures*. Cambridge: MIT Press.
- Katz, J. (1972). *Semantic theory*. New York: Harper and Row.
- Keil, F. C. (1989). *Concepts, kinds, and cognitive development*. Cambridge, MA: MIT Press.
- Kemler Nelson, D. G. (1995). Principle-based inferences in young children's categorization: Revisiting the impact of function on the naming of artifacts. *Cognitive Development*, 10, 347–380.
- Kemler Nelson, D. G. (1999). Attention to functional properties in toddlers' naming and problem-solving. *Cognitive Development*, 14(1), 77–100.
- Kemler Nelson, D. G., Herron, L., & Morris, C. (2002). How children and adults name broken objects: Inferences and reasoning about design intentions in the categorization of artifacts. *Journal of Cognition and Development*, 3, 301–332.
- Kemler Nelson, D. G., Holt, M. B., & Egan, L. C. (2004). Two- and three-year-olds infer and reason about design intentions in order to categorize broken objects. *Developmental Science*, 7(5), 543–549.
- Kripke, S. (1972). Naming and necessity. In D. Davidson & G. Harman (Eds.), *Semantics of natural language*. Dordrecht: Reidel.
- Landau, B., Smith, L. B., & Jones, S. S. (1988). The importance of shape in early lexical learning. *Cognitive Development*, 3, 299–321.

- Laurence, S., & Margolis, E. (1999). Concepts and cognitive science. In E. Margolis & S. Laurence (Eds.), *Concepts: Core readings* (pp. 3–82). Cambridge: MIT Press.
- Levinson, S. (1983). *Pragmatics. Cambridge textbooks in linguistics*. Cambridge: Cambridge University press.
- Li, P., Dunham, Y., & Carey, S. (2009). Of substance: The nature of language effects on entity construal. *Cognitive Psychology*, 58(4), 487–524.
- Locke, J. (1690). An essay concerning human understanding. In P. H. Nidditch (Ed.), Oxford: Oxford University Press.
- Macnamara, J. (1986). *A border dispute: The place of logic in psychology*. Cambridge, MA: MIT Press.
- Malt, B. C. (in press). Naming artifacts: Patterns and processes. In B. Ross (Ed.), *The psychology of learning and motivation*. Elsevier.
- Malt, B. C., & Johnson, E. C. (1992). Do artifact concepts have cores? *Journal of Memory and Language*, 31, 195–217.
- Musolino, J. (2004). The semantics and acquisition of number words: Integrating linguistic and developmental perspectives. *Cognition*, 93(1), 1–41.
- Noveck, I. A. (2001). When children are more logical than adults: Experimental investigation of scalar implicatures. *Cognition*, 78, 165–188.
- Papafragou, A., & Musolino, J. (2003). Scalar implicatures: Experiments at the semantics-pragmatics interface. *Cognition*, 86, 253–282.
- Partee, B. H. (2011). Formal semantics: Origins, issues, early impact. In B. H. Partee, M. Glanzberg, & J. Skilters (Eds.), *The baltic yearbook of cognition, logic, and communication* (Vol. 6, pp. 1–52). Lawrence, KS: New Prairie Press.
- Prasada, S., Ferenz, K., & Haskell, T. (2002). Conceiving of entities as objects and stuff. *Cognition*, 83, 141–165.
- Putnam, H. (1970). Is semantics possible? In H. Keifer & M. Munitz (Eds.), *Language, belief, and metaphysics*. New York: State University of New York Press.
- Quine, W. V. (1960). *Word and object*. Cambridge, MA: The MIT Press.
- Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, 7, 573–603.
- Schatz, M. (1993). Ways to words: Early strategies for building a lexicon. *Paper presented Sixth International Congress for the Study of Child Language, Trieste, Italy*.
- Shiple, E., & Shepperson, B. (1990). Countable entities: Developmental changes. *Cognition*, 34, 109–136.
- Smith, C. L. (1980). Quantifiers and question answering in young children. *Journal of Experimental Child Psychology*, 30, 191–205.
- Sophian, C., & Kalihiwa, C. (1998). Units of counting: Developmental changes. *Cognitive Development*, 13(561–585), 3.
- Stiller, A., Goodman, N. D., & Frank, M. C. (2011). Ad-hoc scalar implicatures in adults and children. In *Proceedings of the 33rd annual meeting of the cognitive science society*.
- Strawson, P. F. (1959). *Individuals: An essay in descriptive metaphysics*. London: Methuen.
- Tare, M., Schatz, M., & Gilbertson, L. (2008). Maternal uses of non-object terms in child directed speech; Color, number, and time. *First Language*, 28, 87–100.
- Wagner, L., & Carey, S. (2003). Individuation of objects and events: A developmental study. *Cognition*, 90, 163–191.
- Waxman, S. R., & Markow, D. B. (1995). Words as invitations to form categories: Evidence from 12- to 13-month-old infants. *Cognitive Psychology*, 29, 257–302.
- Wiggins, D. (1967). *Identity and spatio temporal continuity*. Oxford: Blackwell.
- Wittgenstein, L. (1958). *Philosophical investigations*. London: Blackwell.
- Wynn, K. (1990). Children's understanding of counting. *Cognition*, 36, 155–193.
- Wynn, K. (1992). Children's acquisition of number words and the counting system. *Cognitive Psychology*, 24, 220–251.
- Xu, F. (2007). Sortal concepts, object individuation, and language. *Trends in Cognitive Sciences*, 11, 400–406.