Memory for cardinality supports a one-set account of conservativity

All quantifiers have conservative meanings. Using a sentence verification task to probe which set(s) participants represent during evaluation, we test the predictions of three explanations for this semantic universal: on the "lexical" view, quantifiers express relations between two sets, but only certain relations make good determiner meanings; on the "interface" view, the interpretation of quantifier raising ensures that non-conservative quantifiers either couldn't exist or couldn't be detected; and on the "logical" view, quantifiers express first-order relations relativized to a single set, so conservativity is a natural consequence. Our results support the logical view.

Conservativity states that for any quantificational relation Q, its first argument (its *restriction*) can be intersected with its second argument (its *scope*) without a resulting change in truth conditions, as in (1) [1,2]. For example, (2a) is truth-conditionally equivalent to (2b).

(1) $Q(A, B) \leftrightarrow Q(A, A \cap B)$

(2) a. Every big circle is green
b. Every big circle is a big circle that's green

Put another way, the truth of a conservative quantifier depends on nothing outside of the extension of its restriction. Languages thus have no determiners that express relations like (4) – the inverse of *every* in (3) – or like (5), which is true when its two arguments are coextensional. And children do not seem to consider such non-conservative meanings when learning a novel quantifier [3], suggesting that conservativity reflects a deep fact about the language faculty.

(3) $\text{Every}(A, B) \equiv A \subseteq B$ (4) $\text{Yreve}(A, B) \equiv A \supseteq B$ (5) $\text{Equi}(A, B) \equiv A = B$

Given the standard view that quantifier meanings express relations between sets [1], the lexical approach deals with the apparent ban on non-conservative relations by stipulating a restriction on which relations can be lexicalized (e.g., subset but not superset or identity) [2].

The interface approach [e.g., 4] retains the idea that quantifier meanings express set relations and derives conservativity as a consequence of the syntax-semantics interface. On this view, sentences like (2a) have logical forms like (2b), due to quantifier raising and independently plausible assumptions about the semantic interpretation of traces. Conservative meanings like *every*'s are unaffected, as in (3'). But some non-conservative relations would lead to trivial meanings, as in (4'), and are ruled out by a filter that renders trivial meanings ungrammatical. Other non-conservative relations might be lexicalized as quantifiers but remain undetectable. For example, *every* might express the identity relation in (5), but any sentence with it would end up with the conservative meaning in (5'), which is truth-conditionally equivalent to (3').

(3') $\text{Every}(A,B) \equiv A \subseteq A \cap B$ (4') $\text{Yreve}(A,B) \equiv A \supseteq A \cap B$ (5') $\text{Equi}(A,B) \equiv A = A \cap B$

The logical approach – versions of which are pursued in [5-6] – rules out non-conservative meanings by rejecting the standard view that they express relations between two sets. Instead, quantifiers express one-place first-order relations relativized to the set expressed by their restriction, as in (6) ($x \restriction A$ represents the restriction of x to A): (6) $\text{Every}/A(B) \equiv \forall x[Bx] \restriction A$

Unlike (3), which treats A and B as independent sets related in a particular way (e.g., via subset), the meaning specification in (6) implicates only the set denoted by the restriction (A). The predicate (B) then attributes a property to the members of this set in a first-order way (e.g., *relativized to the set of big circles, each thing is green*). All attested quantifiers are stateable in these terms but non-conservative relations are not [6]. Intuitively, non-conservative relations implicate the full extension of their scope, independently of that of their restriction. Stating such a relation is easy if both A and B are on a par. But if quantification is restricted to the elements of A, there is no way to state a relation that involves the elements of B – A.

To pit these three views against each other, we asked whether sentences like (2a) are represented by speakers in terms of a relation between two independent sets, as in (3) and (3'), or in terms of attributing properties to the members of a single set, as in (6). Participants saw circles of different sizes and colors and were asked to judge the truth of statements (Fig. A).

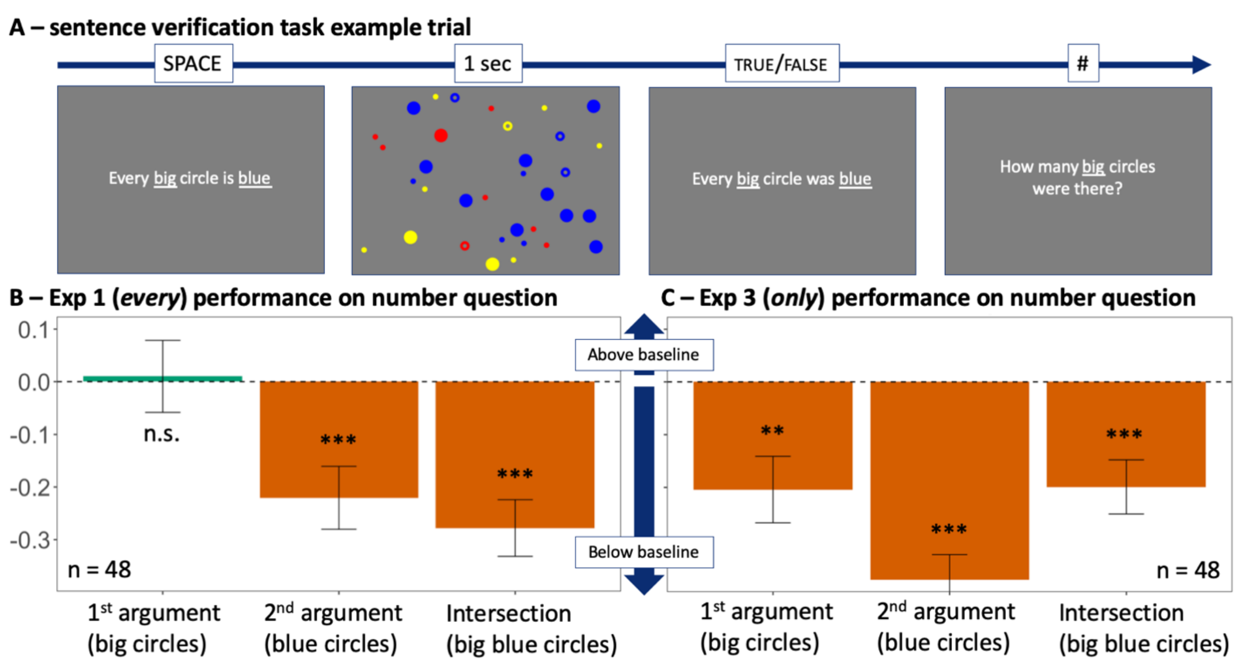
After responding, they were asked how many circles of a particular type there were (e.g., *big*, *green*, or *big green*). To determine the accuracy of their estimates, their responses were fit with the standard psychophysical model of cardinality estimation [7]. The resulting parameter was compared against a baseline task using the same images alongside questions like *how many big circles are there?* This baseline offers a measure of the best possible performance the visual system will afford for sets of each type (size, color, and intersection). We expect participants to have a reliable estimate of cardinality only if they represented the relevant set during evaluation. And we expect participants to represent and use sets implicated by the meaning [e.g., 8].

Consistent with the logical view, we find that when asked about the cardinality of the set denoted by the restriction (e.g., big circles), participants (n=48) performed as well as baseline (χ^2=0.02, p=.88), meaning they knew the cardinality as well as their visual systems would allow. But when asked about the set denoted by the scope (e.g., blue things) or by the intersection of both arguments (e.g., big blue circles), they performed significantly worse than baseline (scope: χ^2=13.61, p<.001; intersection: χ^2=26.61, p<.001; Fig. B).

In a second experiment, participants (n=54) had the option to opt out of the "how many?" questions with an "I don't know" button. We observe two patterns: participants are more likely than baseline to opt out when asked about the second argument (t_{49}=2.94, p<.005) or about the intersection (t_{49}=3.09, p<.005), but not when asked about the first argument (t_{49}=0.19, p=.85); and when performance is measured on trials they did not opt out of, the pattern matches exp1 (restriction: χ^2=2.59, p=.11; scope: χ^2=40.02, p<.001; intersection: χ^2=31.46, p<.001).

In a third experiment, participants (n=48) were shown similar images but asked to verify statements with the focus-operator *only* (e.g., *only big circles are blue*). Participants performed worse than baseline on all questions (size: χ^2=10.51, p<.005; color: χ^2=62.34, p<.001; intersection: χ^2=15.05, p<.005; Fig. C). This excludes the possibility that participants represent the set described by the first NP regardless of the sentence's meaning.

Taken together, these results suggest that *every*'s meaning implicates only the set described by its restriction, as in (6). To the extent that this result generalizes beyond *every*, it supports an explanation of conservativity rooted in the logical structure of quantifiers.



[1] Barwise & Cooper (1981) [2] Keenan & Stavi (1986) [3] Hunter & Lidz (2012) [4] Romoli (2015) [5] Pietroski (2018) [6] Westerståhl (2019) [7] Odic et al. (2016) [8] Lidz et al. (2011)