New evidence for the unlearnability of non-conservative quantifiers*

Tyler Knowlton[1], John Trueswell[2], and Anna Papafragou[3]

[1] MindCORE, University of Pennsylvania, Philadelphia, PA, U.S.A.
tzknowlt@upenn.edu
[2] Department of Psychology, University of Pennsylvania, Philadelphia, PA, U.S.A.
trueswel@psych.upenn.edu
[3] Department of Linguistics, University of Pennsylvania, Philadelphia, PA, U.S.A.
anna4@sas.upenn.edu

1 Introduction

The 'conservativity' of quantificational determiners is perhaps the most robust and renowned semantic universal [1, 7, 9, 21]. The typological generalization is this: duplicating a quantifier's internal (NP) argument in its external (predicative) argument is logically insignificant. Put another way, abstracting from variations in word order and morphology, the entailment pattern in (1) holds for any quantifier Q (e.g., *every fish swims* iff *every fish is a fish that swims*).

(1) [[Q NP] PRED] ↔ [[Q NP] [be NP that PRED]]

And where potential exceptions to this pattern have been proposed (e.g., *only*), there are independent reasons for thinking the purported counter-examples are not quantificational determiners after all [6, 5, 2, 23, 16, 17].

That all quantifiers are conservative is significant, in part because it is unexpected given the standard view that quantifiers express relations between two sets. If we grant that languages have conservative quantifiers like *every* that express the subset relation in (2a), then what explains why languages lack a hypothetical non-conservative quantifier like *yreve*, which expresses the superset relation in (2b), or like *ident*, which expresses the identity relation in (2c)?

(2) a. ⟦Every fish swims⟧ ≈ FISH ⊆ SWIMMERS
b. ⟦Yreve fish swims⟧ ≈ FISH ⊇ SWIMMERS
c. ⟦Ident fish swims⟧ ≈ FISH = SWIMMERS

There is nothing conceptually more complicated about the latter two relations [20]. And it seems unlikely that an explanation can be found in terms of communicative need. After all, languages have expressions like *only* (*yreve* is *only* if it were a determiner instead of a focus operator) and *are identical to*. The fact to be explained is that no language seems to have a quantificational determiner with those sorts of meanings.

A reductionist (though logically possible) explanation would be that determiner conservativity is a historical accident. Maybe non-conservative meanings could be lexicalized as determiners, but they just haven't been in any natural language. Alternatively (and more interestingly), it has been argued that conservativity has an architectural explanation. It might be that only conservative relations and conservativity-preserving operations are part of the primitives out of

*For helpful discussion, we thank Jeffrey Lidz, Paul Pietroski, Alexander Williams, and audiences at the UMass Amherst Psycholinguistics Workshop, BUCLD 47, and the University of Pennsylvania.

which determiner meanings can be constructed [9], or that details of how syntactic movement is interpreted serve to filter out or otherwise disguise would-be non-conservative determiners [3, 13, 18], or that determiners never express non-conservative relations because they don't express relations in the first place [15, 11, 12, 14]. What these explanations have in common is that they all maintain that determiner conservativity reflects a fundamental fact about the language faculty, a constraint stemming from the architecture of the grammar, not one due to historical coincidence, communicative pressures, or domain general cognitive considerations. This claim predicts that children (and adults) should be unable to learn non-conservative quantifiers. But evidence bearing out this bold prediction has proven elusive.

2 Prior work on conservativity and learnability

Two studies have previously looked for evidence of a learnability advantage for novel conservative over novel non-conservative determiners. In one [8], 20 5-year-old children were introduced to a picky puppet who 'likes' certain scenes and 'dislikes' others. The child's task was to help the experimenter sort scenes into piles according to whether the puppet liked them. Children were told that the puppet likes it when "gleeb girls are on the beach", and were left to discern what *gleeb* means. In the Conservative condition, *gleeb* meant *not all*, so "gleeb girls are on the beach" was true just in case there was at least one girl not on the beach. In the Non-Conservative condition, *gleeb* meant *not only*, so "gleeb girls are on the beach" was true just in case there was at least one boy on the beach. After watching the experimenter sort five cards, children were presented with five new cards and asked to sort them. They showed the predicted pattern: only participants in the Conservative condition performed significantly above chance.

In an attempt to replicate this result, though, another study [19] found that children showed no evidence of learning in either condition. Children also failed to show the expected result when the task was modified to allow the puppet to 'correct' the situation to his liking (e.g., by moving a character). Moreover, the effect failed to replicate in 18 English-speaking adults.

The experiments reported below aim to improve on these previous designs in the hopes of demonstrating that non-conservative meanings cannot be learned as quantificational determiners even when (i) their conservative counterparts can be learned as quantifiers in the same experimental setting and (ii) they can be learned as instances of other syntactic categories.

3 Current experiments

The current experiments differ from prior work in a number of ways. First, instead of using *not all* and *not only* as the novel conservative and non-conservative meanings, the current experiments compare the conservative *all but one* against the non-conservative *outnumbers by one* (Experiments 1-4) and the conservative *every* against the non-conservative *equi* (Experiment 5). The rationale behind the change is that *not all* and *not only* are both negations of existing quantifier meanings, and learners may be slower to learn meanings that involve negation [4].

Second, instead of being asked to sort scenes into piles based on whether a puppet liked them (which may have highlighted the puppet's preferences more than the novel word's meaning), participants here were explicitly told that their job is to learn a novel word. And their knowledge was tested in ways that served to highlight the word-learning aspect of the task (e.g., by being asked whether a sentence with *gleeb* is true relative to a picture).

Third, the experiments presented here use sentences in the partitive frame (e.g., *gleeb of the circles are blue*), which provide an unambiguous signal of the novel word being a determiner.

In contrast, past work avoided the partitive, opting instead for sentences like *gleeb girls are on the beach*. This is a sensible choice: the partitive frame might encourage attention to the quantifier's internal argument as such, leading participants to focus on the girls or the circles independent of their hypothesized meaning for *gleeb*. But the lack of partitive may have led some participants to posit a different syntactic category for *gleeb* (e.g., *gleeb* could be a focus operator like *only* or an adjective like *friendly*). And while phrases like *of the circle* do likely encourage attending to the circles, quantificational phrases that lack the partitive, like *every circle*, have since been shown to drive attention to the circles as well [10, 11]. This fact and conservativity likely share a common cause, making it difficult and perhaps futile to control for.

Finally, the current study departs from previous work in that it focuses only on adult participants. This is in part out of convenience, but also because adults have the full complement of cognitive resources at their disposal. If they are able to learn novel conservative quantifiers but unable to learn novel non-conservative ones, it is hard to imagine children faring any better.

Each experiment is discussed in its own subsection below. But for ease of display, all data are plotted together in Figure 1, along with example training and test trials for each experiment. The prediction throughout – on the hypothesis that conservativity has a grammatical explanation – is that only participants in the Conservative condition will successfully learn the meaning of a novel quantifier *gleeb*, but that any such learnability advantage will disappear if *gleeb* is instead taught as a novel verb.

3.1 Experiment 1: Learning by example

In Experiment 1, participants (60 English-speaking adults recruited on `prolific.co`) were trained with a series of example images and corresponding sentences using the novel quantifier *gleeb*. Participants in the Conservative condition were given training trials consistent with the sentence in (3) having the meaning in (3a), whereas those in the Non-conservative condition were given training trials consistent with (3) having the meaning in (3b).

(3) Gleeb of the circles are blue.

a. $|\textsc{circles}| - 1 = |\textsc{blue-circles}|$ *all but one of the circles are blue*

b. $|\textsc{circles}| - 1 = |\textsc{blue-things}|$ *the circles outnumber the blue things by one*

Training consisted of 16 images of blue and orange circles and squares. Each picture was either described with the sentence "gleeb of the circles or blue" or the sentence "it's not the case that gleeb of the circles are blue", depending on the condition. Participants were also explicitly told the number of circles and blue shapes present in each display (this information was intended to tilt the odds in favor of success in the Non-conservative condition). All text presented on screen was also read aloud to participants. The particular training trials used were designed to rule out easily-hypothesized meanings for *gleeb* (e.g., training trials were included where *more than half of the circles are blue* was false but *gleeb of the circles are blue* was true).

Immediately after training, participants were given six new images and asked "is it true that gleeb of the circles are blue?". For three images, the correct answer was yes; for the other three images, the correct answer was no. Both the training and the test images were held constant across conditions, but the correct answer flipped (a true/'yes' image in the Conservative condition would be a false/'no' image in the Non-conservative condition).

Results. Participants in the Conservative condition were more accurate than those in the Non-conservative condition (comparing a model with condition as a fixed effect to an intercept-only model with the same random effects structure: $\chi^2(1) = 23.28, p < .001$; main effect of

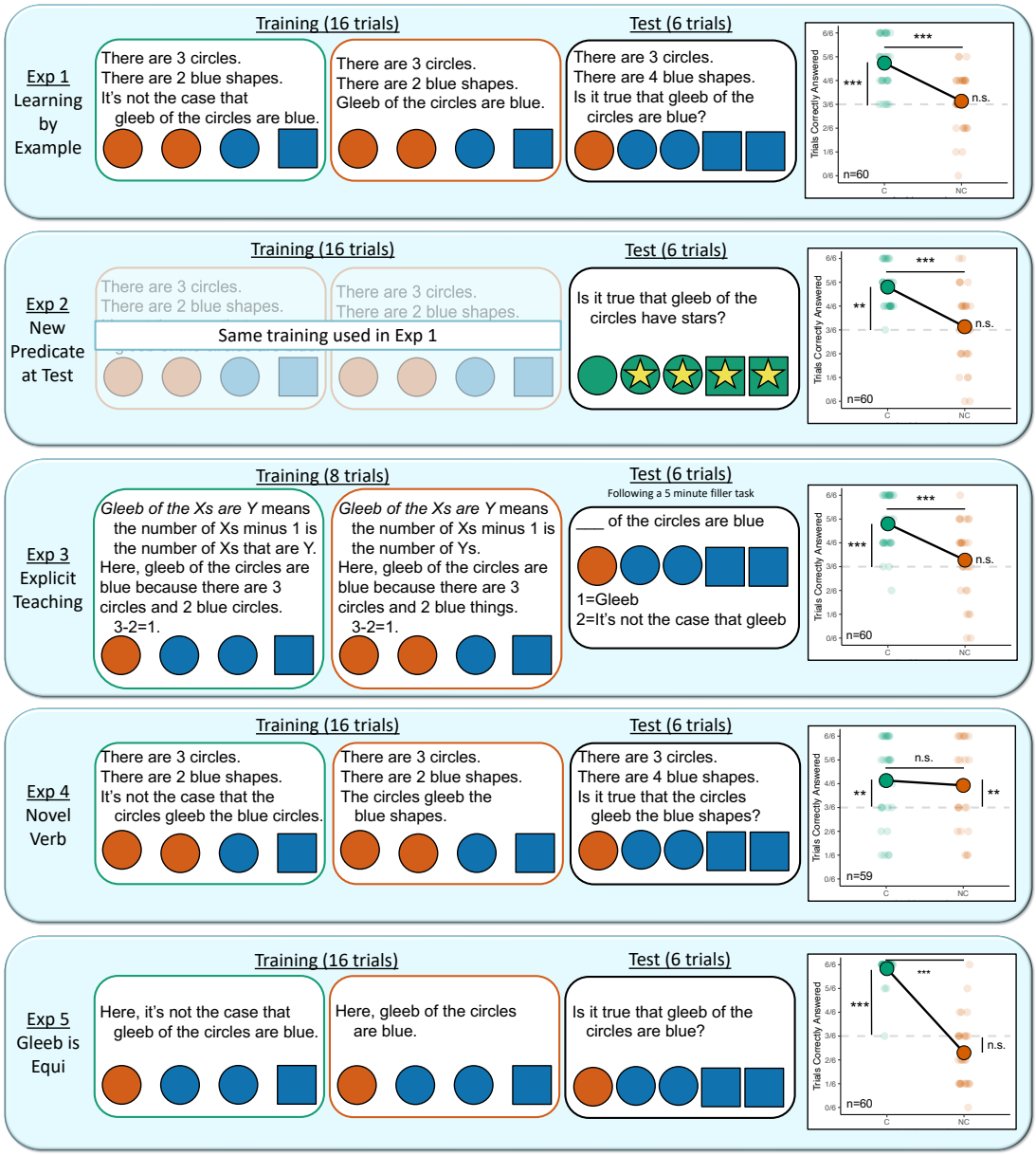

Figure 1: Example stimuli and results from all five experiments. Example training trials from the Conservative (green outline) and Non-conservative (orange outline) conditions are shown for each experiment. One test trial (black outline) is also shown. Results for each experiment are shown on the far right. Large points represent average performance, translucent points represent the number of test trials each individual participant correctly answered. Chance performance, 3/6, is represented by the grey dashed line. Experiments were designed in PCIbex [22].

condition: $\beta = .66$ [95% CI .52 to .79], $z = 4.93, p < .001$). And while those in the Conservative condition performed significantly above chance (intercept: $\beta = 1.72$ [95% CI 1.22 to 2.22], $z = 3.43, p < .001$), those in the Non-conservative condition did not (intercept: $\beta = .09$ [95% CI $-.06$ to .24], $z = 0.59, p = .553$). That is, only participants in the Conservative condition showed evidence of learning the novel meaning, as predicted.

3.2 Experiment 2: New predicate at test

In Experiment 2, participants (60 English-speaking adults) were trained exactly as in Experiment 1 (with 16 pictures of blue and orange circles and squares). The only difference came in the test phase: instead of differing in color, the shapes were all green, and they either had or lacked yellow stars. Everything else about the distribution of shapes was the same as in Experiment 1: the shapes that were blue in Experiment 1 had stars in Experiment 2, and those that were orange in Experiment 1 lacked stars in Experiment 2. Importantly, the test sentences used a different predicate ("is it true that gleeb of the circles have stars?"). This change helps ensure that participants who succeed at the task actually do so by learning the meaning of *gleeb*, as opposed to relying on visual similarities between the training and test images.

Results. The results from Experiment 1 were replicated. Participants in the Conservative condition were more accurate than those in the Non-conservative condition ($\chi^2(1) = 21.03, p < .001$; main effect of condition: $\beta = .73$ [95% CI .58 to .88], $z = 4.75, p < .001$). Moreover, while those in the Conservative condition performed significantly above chance (intercept: $\beta = 2.04$ [95% CI 1.33 to 2.75], $z = 2.87, p < .01$), those in the Non-conservative condition did not (intercept: $\beta = .1$ [95% CI $-.12$ to .32], $z = 0.46, p = .647$). As in Experiment 1, then, only participants in the Conservative condition showed evidence of learning the novel meaning.

3.3 Experiment 3: Explicit teaching

Given the difficulties experienced in the Non-conservative condition in Experiments 1 and 2, Experiment 3 sought to make the task easier through an 'explicit teaching' paradigm. In this version, participants (60 English-speaking adults) were explicitly told the meaning of *gleeb* before being guided through the training trials. Additionally, an explanation was given as to why each image made the sentence true or false (e.g., "Here, gleeb of the circles are blue because there are 3 circles and 2 blue shapes..."). Participants were then given a 5-minute break before moving on to the test phase, which had them fill in a blank in a sentence like "___ of the circles are blue" (with either "Gleeb" or "It's not the case that gleeb").

Results. Even with the 'brute force' nature of the training, Experiment 3 again replicates the effect found in Experiments 1 and 2. Participants in the Conservative condition were more accurate than those in the Non-conservative condition ($\chi^2(1) = 16.34, p < .001$; main effect of condition: $\beta = .69$ [95% CI .52 to .85], $z = 4.05, p < .001$). And while those in the Conservative condition performed significantly above chance (intercept: $\beta = 1.74$ [95% CI 1.27 to 2.21], $z = 3.72, p < .001$), those in the Non-conservative condition did not (intercept: $\beta = .21$ [95% CI $-.01$ to .43], $z = 0.94, p = .346$). So despite being explicitly taught, only those in the Conservative condition were able to successfully encode and remember *gleeb*'s meaning.

3.4 Experiment 4: Novel verb instead of novel determiner

Since conservativity is specific to quantificational determiners, any learnability asymmetry that exists should disappear if the novel word is of a different syntactic category. Experiment 4 aimed to test this prediction; to make the effect observed in Experiments 1-3 go away. To do so, participants (59 English-speaking adults) were taught the same two novel meanings, but *gleeb* was introduced as a verb instead of as a determiner. In the Conservative condition, images were described with sentences like "The circles gleeb the blue circles". In the Non-conservative condition, images were described with sentences like "The circles gleeb the blue shapes." Otherwise, the same scene-sentence pairs were used as in the above three experiments.

Results. As predicted, the learnability advantage for the conservative *gleeb* disappeared when both novel words were taught as verbs. Participants in the Conservative condition were not significantly more accurate than those in the Non-conservative condition ($\chi^2(1) = 0.27, p = .601$; no effect of condition: $\beta = .11$ [95% CI $-.1$ to $.33$], $z = 0.53, p = .598$). And both groups performed significantly better than chance (Conservative intercept: $\beta = 1.13$ [95% CI 0.78 to 1.48], $z = 3.22, p < .01$; Non-conservative intercept: $\beta = 0.8$ [95% CI 0.53 to 1.08], $z = 2.93, p < .01$). This suggests the results from Experiments 1-3 were driven by leaning the non-conservative meaning as a quantifier per se, not any difficulties with the meaning itself.

3.5 Experiment 5: Testing a different non-conservative meaning

The results of Experiments 1-4 suggest that pairing a non-conservative meaning with a novel determiner is at least exceedingly unnatural, if not outright impossible. But one might worry that the observed learnability asymmetry is a quirk of the particular meaning tested. To address that concern, Experiment 5 used the same paradigm to test another hypothetical non-conservative quantifier. In the Non-conservative condition of Experiment 5, *gleeb of the circles are blue* meant *the circles and the blue things are equinumerous* (i.e., |CIRCLES| = |BLUE-THINGS|, also known as *equi*). This non-conservative meaning was compared against its conservative counterpart: |CIRCLES| = |BLUE-CIRCLES| (which is truth-conditionally equivalent to *every*). Different images were used, but the structure of the task was the same as Experiment 1.

Results. Participants in the Conservative condition were again more accurate than those in the Non-conservative condition ($\chi^2(1) = 77.34, p < .001$; main effect of condition: $\beta = 2.58$ [95% CI 2.21 to 2.96], $z = 6.93, p < .001$). And while those in the Conservative condition performed significantly above chance (intercept: $\beta = 54.51$ [95% CI 47.03 to 62], $z = 7.28, p < .001$), those in the Non-conservative condition did not (intercept: $\beta = -.69$ [95% CI -1.28 to -0.09], $z = -1.15, p = .249$). That is, the effect replicates with a different pair of quantifiers.

4 Conclusion

Cross-linguistically, all quantifiers are conservative. The robustness of this semantic universal invites a learnability claim: if conservativity reflects a deep fact about the language faculty, then non-conservative quantifiers should be impossible to learn. The five experiments presented above bear out this prediction. Adults were able to pair conservative meanings, but not non-conservative meanings, with novel quantifiers, and, as predicted, this effect disappeared when the novel word was a verb. As such, the experiments reported above lend support to views on which conservativity is a cornerstone of the semantics of determiners (e.g., [11, 15, 14]).

References

[1] Jon Barwise and Robin Cooper. Generalized quantifiers and natural language. In *Philosophy, Language, and Artificial Intelligence*, pages 241–301. Springer, 1981. https://www.jstor.org/stable/25001052.

[2] Ariel Cohen. Relative readings of many, often, and generics. *Natural Language Semantics*, 9:41–67, 2001. http://dx.doi.org/10.1023/A:1017913406219.

[3] Danny Fox. Antecedentcontained deletion and the copy theory of movement. *Linguistic Inquiry*, 33(1):63–96, 2002. https://doi.org/10.1162/002438902317382189.

[4] Angela Xiaoxue He and Alexis Wellwood. "Most" is easy but "least" is hard: Novel determiner learning in 4-year-olds. In Jennifer Culbertson, Andrew Perfors, Hugh Rabagliati, and Veronica Ramenzoni, editors, *Proceedings of the 44th Annual Conference of the Cognitive Science Society*, 2022. https://escholarship.org/uc/item/5hh4m526.

[5] Elena Herburger. Focus and weak noun phrases. *Natural Language Semantics*, 5(53):53–78, 1997. https://doi.org/10.1023/A:1008222204053.

[6] Elena Herburger. *What counts: Focus and quantification*. MIT Press, 2000.

[7] James Higginbotham and Robert May. Questions, quantifiers and crossing. *The Linguistic Review*, 1(1):41–80, 1981. http://dx.doi.org/10.1515/tlir.1981.1.1.41.

[8] Tim Hunter and Jeffrey Lidz. Conservativity and learnability of determiners. *Journal of Semantics*, 30(3):315–334, 2013. doi.org/10.1093/jos/ffs014.

[9] Edward L Keenan and Jonathan Stavi. A semantic characterization of natural language determiners. *Linguistics and Philosophy*, 9(3):253–326, 1986. https://www.jstor.org/stable/25001246.

[10] Tyler Knowlton, Paul Pietroski, Justin Halberda, and Jeffrey Lidz. The mental representation of universal quantifiers. *Linguistics and Philosophy*, 45:911–941, 2022. https://doi.org/10.1007/s10988-021-09337-8.

[11] Tyler Knowlton, Paul Pietroski, Alexander Williams, Justin Halberda, and Jeffrey Lidz. Determiners are "conservative" because their meanings are not relations: Evidence from verification. *Semantics and Linguistic Theory*, 30:206–226, 2021. https://journals.linguisticsociety.org/proceedings/index.php/SALT/article/view/30.206.

[12] Peter Lasersohn. Common nouns as modally non-rigid restricted variables. *Linguistics and Philosophy*, 44(2):363–424, 2021. https://doi.org/10.1007/s10988-019-09293-4.

[13] Peter Ludlow. Lf and natural logic. In Gerhard Preyer and Georg Peter, editors, *Logical form and language*, pages 132–168. Oxford University Press, 2002.

[14] Peter Ludlow and Sašo Zivanović. *Language, Form, and Logic: In Pursuit of Natural Logic's Holy Grail*. Oxford University Press, 2022. https://doi.org/10.1093/oso/9780199591534.001.0001.

[15] Paul Pietroski. *Conjoining Meanings: Semantics Without Truth Values*. Oxford University Press, 2018. https://doi.org/10.1093/oso/9780198812722.001.0001.

[16] Maribel Romero. The conservativity of *many*. In *20th Amsterdam Colloquium*, pages 20–29, 2015. https://ling.sprachwiss.uni-konstanz.de/pages/home/romero/papers/romero_v8-AC20.pdf.

[17] Maribel Romero. The conservativity of *many*: Split scope and *most*. *Topoi*, 37:393–404, 2018. https://doi.org/10.1007/s11245-017-9477-5.

[18] Jacopo Romoli. A structural account of conservativity. *Semantics-Syntax Interface*, 2(1):28–57, 2015. https://pure.ulster.ac.uk/en/publications/a-structural-account-of-conservativity-3.

[19] Jennifer Spenader and Jill de Villiers. Are conservative quantifiers easier to learn? evidence from novel quantifier experiments. In *22nd Amsterdam Colloquium*, 2019. http://hdl.handle.net/11370/79051821-bdba-4168-8214-5cf1033b8451.

[20] Shane Steinert-Threlkeld and Jakub Szymanik. Learnability and semantic universals. *Semantics*

and Pragmatics, 12(4), 2019. https://doi.org/10.3765/sp.12.4.

[21] Kai von Fintel and Lisa Matthewson. Universals in semantics. *The Linguistic Review*, 25(1-2):139–201, 2008. https://doi.org/10.1515/TLIR.2008.004.

[22] Jeremy Zehr and Florian Schwarz. PennController for internet based experiments (IBEX). 2018. https://doi.org/10.17605/OSF.IO/MD832.

[23] Richard Zuber. A class of non-conservative determiners in polish. *Lingvisticæ Investigationes*, 27(1):147–165, 2004. https://doi.org/10.1075/li.27.1.07zub.