Non-conservative quantifiers are unlearnable

Tyler Knowlton, John Trueswell, Anna Papafragou

[last updated April 24, 2024]

Abstract

Linguistic universals have long been a cornerstone of linguistic theories. Perhaps the most well-known universal in the semantic domain is the observation that all quantificational determiners (e.g., “every”, “some”, “no”) have ‘conservative’ meanings: only the noun phrase with which the quantifier combines matters for the truth of the sentence. If it’s true that “every fish swims” then it’s true that “every fish is a fish that swims” (*cf.* “only fish swim”, which is not true in all the same situations as “only fish are fish that swim”). Accordingly, no language has a ‘non-conservative’ quantifier like “equi”, where “equi fish swims” means “the fish and the swimmers are numerically equivalent” (this quantifier fails to be conservative because the swimmers also matter). This robust cross-linguistic generalization has been argued to reflect a fundamental property of quantifier semantics, a linguistically-specific constraint. If conservativity is a genuine semantic universal, as opposed to the result of a historical accident, then non-conservative quantifiers should be unlearnable. Across seven experiments, we show that this prediction is borne out. Adult participants fail to learn three novel non-conservative meanings, even when explicitly taught, but succeed at learning their conservative counterparts. And since conservativity is a property of *quantifier* semantics, this effect disappears when an intended non-conservative meaning is instead paired with verbal syntax. These results suggest that the conservativity universal is tied to learnability, and support semantic theories on which conservativity reflects a deep fact about the human language faculty.

1. Introduction

Linguistic typology is often informative about learning biases (e.g., Chomsky 1965; Tesar & Smolensky 1998; Gentner & Bowerman 2010). Grammatical patterns that are easier to learn are frequently reflected in the typology of the world's languages, and those that are harder to learn are more likely to 'die out' (e.g., Tily, Frank, & Jaeger 2011; Culbertson, Smolensky, & Legendre 2012). To be sure, cross-linguistic universals (features present in all languages) and cross-linguistic gaps (features present in no languages) can sometimes result from mere historical accident. But they can also reflect general cognitive biases, communicative pressures (e.g., pragmatic considerations about what things are useful to express), or architectural constraints on grammar (i.e., restrictions stemming specifically from language). Typology can thus offer a useful lens from which to study cognition, language acquisition, and the language faculty.

Compared to syntax or phonology, robust universals in the domain of linguistic meaning are relatively hard to come by (see von Fintel & Matthewson 2008 for review). Where they have been most fruitfully pursued is in the domain of quantification. Here, we focus on perhaps the most well-known semantic universal, 'conservativity' (described in detail below). We show that this universal is related to learnability, suggesting that (as many in the literature have suspected) it is no historical accident, but an informative typological generalization. In particular, we argue that this particular learnability-typology relationship resists explanation in terms of general cognitive constraints or communicative pressures. Instead, the fact that all quantifiers are 'conservative' is a consequence of grammatical architecture.

1.1 The conservativity universal

The conservativity of quantificational determiners (henceforth, quantifiers) is perhaps the most robust and renowned semantic universal (Barwise & Cooper 1981; Higginbotham & May 1981; Keenan 1981; Keenan & Stavi 1986). A first-pass, intuitive characterization of the observation is this: for any quantifier, the noun (phrase) with which it combines 'sets the scene' such that one need not look beyond the things described by that noun (phrase) to figure out whether the sentence is true.

Take the quantifier "every" as an example: A sentence like (1a) is about the fish, not the swimmers or anything else. In this sense, "every" can be thought of as restricting its domain to just the things described by the expression that it directly combines with ("fish"), or as Barwise and Cooper (1981) put it, it can be said to 'live on' its internal argument. An often-deployed test for conservativity is to show that biconditionals hold between sentences like (1a) and (1b), which duplicates "fish" in the predicate.

(1) a. Every fish swims.

b. Every fish is a fish that swims.

The idea is that since the domain is already restricted to the fish, adding it as an extra predicate is informationally insignificant. And so (1a) is true if and only if (1b) is true. This much is obvious in the case of "every": every fish is a fish, so if every fish swims, then it must be that every fish is a fish that swims; likewise, if it's not the case that every fish swims, then it can't be that every fish is a swimming fish. The same biconditional holds if "every" is replaced by "most", "some", "no", or any other quantifier. Indeed, this pattern seems to hold for all quantifiers cross-linguistically (and in the few cases where potential counterexamples have been proposed, their status as genuine quantificational determiners has been debated; see e.g., von Fintel 1997; Herburger 1997; 2000; Cohen 2001; Zuber 2004; Romero 2015; 2018; Ahn & Sauerland 2017; Ahn & Ko 2022; Pasternak & Sauerland 2022 and below).

It almost seems trivial to point out that only fish matter for the truth of quantificational sentences like “every/some/no fish swim”, or that biconditionals like the one between (1a) and (1b) hold. But it is easy to imagine communicatively useful and conceptually simple ‘non-conservative’ quantifiers that would be counterexamples to this generalization. For instance, consider the hypothetical quantifier “equi”, meaning “the two arguments are equinumerous”, as in (2a). Because one can imagine a situation in which (2a) is true but (2b) is false – imagine two fish, one that can swim and one that is injured and cannot, and one human, who can swim – we know that “equi” is not conservative. And indeed, no language has a determiner with the meaning of “equi”.

(2) a. Equi fish swims.

≈ *the fish and the swimming things are numerically equivalent*

b. Equi fish are fish that swim.

≈ *the fish and the swimming fish are numerically equivalent*

To take another example, consider the hypothetical “everynon”, which has a meaning that combines the meaning of “every” with negation, as in (3a). Again, we can imagine a case where the truth of (3a) and (3b) come apart. If the domain consists of two humans who both swim, (3a) is true, whereas (3b) is trivially false, since the non-fish will never be a subset of the swimming fish, no matter how many there are. So “everynon” fails to be conservative and, as with “equi”, no such quantifier exists cross-linguistically.

(3) a. Everynon fish swims.[1]

≈ the non-fish are a subset of the swimming things

b. Everynon fish are fish that swim.

≈ the non-fish are a subset of the swimming fish

The hypothetical “equi” and the hypothetical “everynon” fail to be conservative in slightly different ways. As noted above, what it means for quantifiers like “every” to be conservative is that they don’t require looking beyond the fish; they ‘live on’ their internal arguments. “Equi” in (2) does require looking beyond the fish; in particular it requires comparing the fish to another group, the swimmers. So “equi” doesn’t ‘live on’ either argument. “Everynon” in (3), on the other hand, requires looking to things named by the complement of its internal argument, namely, the non-fish. So while “everynon” also doesn’t ‘live on’ either argument, it could be said to ‘live on’ the complement of its internal argument.

In light of these two ways of being non-conservative – ‘living on’ neither argument and ‘living on’ the complement of the internal argument – consider a sentence like “only fish swim”. Here, things that aren’t fish matter. The sentence is made false by humans who swim, for example. But “only” doesn’t require looking to the non-fish like “everynon” does. “Only” requires looking to the swimmers, the things named by its external argument. So “only” fails to

[1] Compare the non-conservative “everynon” to the conservative “every”, which can be used in a sentence like “every non-fish swims”, where the negation is part of the internal argument instead of the determiner. Whereas (3a) is not logically equivalent to (3b), “every non-fish swims” is logically equivalent to “every non-fish is a non-fish that swims”. As such, those two sentences follow the pattern in (1).

be conservative in a third way: it does 'live on' an argument, just the wrong one; it 'lives on' its external argument. And as expected, "only" fails the biconditional test: it's not the case that "only fish swim" is true if and only if "only fish are fish that swim" is true. The latter sentence is still true in a case where one human swims (since a human that swims isn't a fish that swims), whereas the former sentence is false, as noted above).

This might initially seem to be a counterexample to the conservativity generalization. But there are independent reasons for thinking that "only" is not a genuine quantificational determiner, and thus not subject to conservativity. For one, it is syntactically more promiscuous than a standard quantifier: "only" can occur between any two words in a sentence like "the cat thought that the dog found the bone", whereas "every", "some", "no" and the like can only appear in place of a "the". Moreover, "only" is focus-sensitive in a way that quantifiers are not. Compare "students only ordered coffee$_F$" (they didn't also order tea or juice) with "students only ordered$_F$ coffee" (they didn't also brew it or pay for it). The same cannot be said for a sentence like "every student ordered coffee", where shifting focus can serve to highlight different alternatives, but does not affect the sentence's truth-conditions: compare "every student ordered coffee$_F$" with "every student ordered$_F$ coffee". Both sentences' truth-conditions are unaffected by whether some students ordered tea or juice in addition and by whether students also paid for their drink. If "only" is not a quantifier, then it is not a counterexample to conservativity.

For those unconceived by these differences, Zuber and Keenan (2019) propose a weakening of conservativity that brings "only" under the generalization as well. The gist of the proposal is this: instead of maintaining that quantifiers 'live on' their internal argument, maybe the right generalization is that quantifiers 'live on' either their internal or their external argument. This weakening would permit meanings like "only" (which 'lives on' its external argument) along with all classically conservative meanings, like “every” (which ‘lives on’ its internal argument). But it would forbid meanings like “equi” and “everynon” for reasons discussed above. For our purposes, we will adopt the classical understanding of conservativity (and the corollary that “only” isn’t a counterexample). We return to the issue of how to define conservativity in Experiment 1b.

But however one states it, the generalization that all quantificational determiners are conservative is surprising given standard views about what types of entities quantifier meanings are. On the standard treatment, quantifiers are said to express relations between the sets named by the quantifier’s two arguments (e.g., Barwise & Cooper 1981; Larson & Segal 1995; Heim & Kratzer 1998). So “every”, for example, is thought to express the subset relation. The sentence “every fish swims” then roughly means “the fish are a subset of the swimming things”. But if the subset relation can be lexicalized as a quantifier, what stops the equi-numerosity relation (“the fish are equal in number to the swimming things”) – the meaning of the hypothetical “equi” – from being lexicalized as a quantifier? And what stops the superset relation (“the fish are a superset of the swimming things”) – essentially the meaning of “only”, if it were a quantifier – from being lexicalized as a quantifier? The hypothetical “everynon” is likewise trivial to specify in relational terms. In short, standard approaches to quantification make conservativity a particularly puzzling generalization because the idea that quantifiers express relations between sets, on its own, is not constraining enough.

What, then, is the source of the conservativity generalization? It seems unlikely that the cross-linguistic lack of non-conservative quantifiers has an explanation in terms of communicative need or general cognitive/computational complexity for the simple reason that languages across the world have expressions like “only” and “equal in number”. That is, languages have expressions with non-conservative meanings, sometimes in the form of single lexical items. So it can't be that the underlying thoughts are in general too complex to express or too hard to lexicalize.[2] The fact to be explained is that languages don't have *quantifiers* with these sorts of meanings, so any explanation not in terms of historical accident will need to make reference to syntactic category (see Section 8.2). Of course, it is possible that conservativity is merely a historical accident. On such a view, there would be no general learnability constraint against non-conservative determiner meanings: non-conservative meanings could in principle be lexicalized as quantifiers, they just haven't been.

An alternative to the deflationary proposals above is that conservativity has a linguistic explanation. We return to what such an explanation might look like in Section 8.2, but briefly, it might be (i) that only conservative relations and conservativity-preserving operations are available for quantifier lexicalization, (ii) that details of how syntactic movement is interpreted serve to filter out or otherwise disguise would-be non-conservative quantifier meanings, or (iii) that quantifiers never express non-conservative relations because quantifiers do not express relations in the first place (see Knowlton, Pietroski, Williams, Halberda, & Lidz 2021 for a review of these differing grammatical approaches). These three explanations all posit that

[2] In support of this idea, Steinert-Threlkeld and Szymanik (2019) show that a neural network had no difficulty learning the non-conservative relations 'not only' and 'greater than', despite being slower to learn certain relations that ran afoul of other typological generalizations. This offers another reason for thinking that non-conservative relations (both those of the sort that 'live on' the wrong argument and those of the sort that fail to 'live on' either argument) are no more computationally complex than conservative ones, at least with respect to the assumptions inherent in the model tested. Relatedly, Steinert-Threlkeld (2021) discusses meanings in terms of optimizing a tradeoff between simplicity and informativeness. Assuming a particular language of thought and a particular measure of complexity, it is argued that meanings that are conservative and monotonic are more efficient. This might be (assuming the logical primitives and measure of complexity used turn out to have psychological reality), but it's not clear to us what would prevent the same reasoning from applying to other syntactic categories, which are not subject to a restriction against non-conservative meanings. See also van de Pol et al. (2023), which found that, depending on the measure of complexity used and the nature of the experiment, conservative meanings can appear to be more, less, or equally as complex as their non-conservative counterparts. They explicitly note that their measures of complexity do not take into account syntactic category. If one did want to explain the conservativity generalization in terms of complexity, one would need to also explain why a given measure (the one for which non-conservative meanings are more complex) applies to a particular syntactic category (determiners) but not others.

conservativity is a fundamental fact about the language faculty. It is a constraint that stems from the very architecture of the grammar, not a generalization that arose because of historical coincidence, communicative pressures, or domain general cognitive considerations. This claim makes a bold prediction: children (as well as adults) should be unable to learn non-conservative determiners despite being able to learn to express the same meanings with non-determiner syntax.

1.2 Conservativity and learnability

Only a few studies have previously looked for evidence of a learnability advantage for novel conservative over novel non-conservative determiners. Hunter and Lidz (2013) tested 5-year-old children with a 'picky puppet task' (Waxman & Gelman 1986). Participants in their experiment were introduced to a puppet who 'likes' certain scenes and 'dislikes' others. The child's task was to help the experimenter sort scenes according to whether the puppet likes them. They were told that the puppet said he likes it when "gleeb girls are on the beach" (but did not tell the experimenter what "gleeb" means). The training phase consisted of the experimenter showing participants five cards, each depicting a scene of girls and boys either on the beach or on the grass. The experimenter sorted the card into the puppet's 'like' pile, saying (4a), or into the puppet's 'doesn't like' pile, saying (4b).

(4) a. The puppet told me that he likes this card because gleeb girls are on the beach.

b. The puppet told me that he doesn't like this card because it's not true that gleeb girls are on the beach.

In the Conservative condition, “gleeb girls are on the beach” was true when not all of the girls were on the beach (i.e., when there was at least one girl on the grass), whereas in the Non-conservative condition, “gleeb girls are on the beach” was true when not only girls were on the beach (i.e., when there was also at least one boy on the beach). This non-conservative “gleeb” fails to be conservative in the same way that “only” fails to be conservative: it ‘lives on’ the wrong argument (i.e., it can be thought of as restricting the domain to the things described by the quantifier’s external argument, the beachgoers, instead of its internal argument, the girls). After being trained in one of those two conditions, children were then shown five new cards and asked to sort them into one of the two piles (the contents of which remained visible throughout). The 20 participants tested showed the predicted pattern: children in the Conservative condition were more accurate than those in the Non-conservative condition (an average of 4.1/5 versus 3.1/5 cards correctly sorted), and only children in the Conservative condition performed significantly above chance (2.5/5 cards correctly sorted).

Spenader and de Villiers (2019) attempted to replicate this result, but found no evidence of learning in either condition: children in the Conservative “not all” condition (n=10) and Non-conservative “not only” condition (n=10) failed to perform above chance. And when they adapted the paradigm to allow the puppet to ‘correct’ the situation to his liking (e.g., by moving one character), children still failed to learn either novel meaning. Moreover, they failed to replicate the effect on 18 English-speaking adults, again finding no evidence of learning in either condition. Simplifying the design, Ramotowska (2022) also attempted to teach adult participants (n=58) “not all” and “not only” by exposing them to 96 trials of sentences such as “gleeb triangles are red” in conjunction with pictures of different colored triangles and circles. On each trial, participants were asked to render a true/false judgment and were given feedback. By the last quarter of trials, participants in both conditions were achieving around 80% accuracy. However, given the large number of trials and the nature of the feedback, we find it unlikely that participants were trying to pair a meaning with a novel quantifier instead of approaching the task as one of pattern-matching. (The conservativity constraint does not preclude adults from recognizing and responding to non-conservative relations, just from taking those sorts of thoughts to be the meaning of a quantificational determiner.) In sum, even though the lack of non-conservative quantifiers in natural language leads to semantic theories that make concrete learnability predictions, evidence bearing out these predictions remains elusive.

1.3 Current study

In a series of seven experiments, we revisit the connection between conservativity and learnability to probe whether the absence of non-conservative determiners from natural language is a design feature of the language faculty. Our study departs from prior experimental work in several ways. A first important difference is the particular novel meanings probed. All three studies cited in the preceding section used "not all" and "not only" as their conservative and non-conservative meanings for "gleeb". These are sensible selections, as on standard theories, both "all" and "only" express different kinds of inclusion relations. If "all" is thought of as expressing the subset relation, then "only" could be said to express the superset relation (notwithstanding differences like its focus sensitivity). So "not all" and "not only" can be regimented as negations of relations that standard theories maintain can get lexicalized. This preempts recourse to alternative explanations like the relative cognitive complexity of one relation over the other. But it may also result in difficult sorts of things to lexicalize, as many studies find that learners are slower to acquire negative meanings or are otherwise biased against them (e.g., Donaldson &

Balfour 1968; Clark 1971; Ehri 1976; Townsend 1976; Ryalls 2000; He & Wellwood 2022). In our study, we compare the conservative "all but one" with the non-conservative "outnumbers by one" (Experiments 1a; 1c; 2-4) and the non-conservative "all but one" with inverted arguments (Experiment 1b). We also compare the conservative "every" with the non-conservative "equi" (Experiment 5). These novel meanings are not paraphrasable by negating an existing lexical item.

A second major difference is that our task is more explicitly a novel word learning task. Participants were told at the outset that their goal was to learn the meaning of a new word, "gleeb", and that they would do so by seeing a series of examples of it used in a sentence. During the test phase, they were either asked whether it was true that "gleeb of the Xs are Y" or they were asked to fill in a blank like "___ of the circles are blue". All of these methodological choices were made to highlight to our participants that the task was about learning a novel quantifier. In contrast, past work either highlighted a puppet's preferences or trained participants to associate pictures with particular responses. These setups may have taken the focus away from the word-learning aspect of the task and shifted it to discerning the pattern present in the displays.

A third, more subtle difference between our task and past work is that, in most of our experiments, our training sentences used the partitive frame – "gleeb of the circles are blue" – whereas past studies avoided the partitive in favor of constructions like "gleeb girls are on the beach" or "gleeb triangles are red". We chose to embrace the partitive in the majority of our experiments because it provides an unambiguous signal to participants that the novel words in question were determiners. Without the partitive "of the", participants might not have been sure of the novel word's syntactic category (e.g., recall that "only" is not a determiner but can be used pre-nominally, as in “only girls are on the beach”). That said, Hunter and Lidz (2013) note that they made the conscious choice to avoid the partitive because phrases like “gleeb of the girls” might encourage representation of the girls *per se*, independent of conservativity. That is, they worried that participants might have been driven to attend to the girls (and not the boys or the beachgoers or the grassgoers) because of the partitive frame and not because of the novel quantifier’s conservative meaning. As we were also concerned about this, Experiment 1c eschews the partitive in favor of sentences like “gleeb circles are blue”.

Lastly, we depart from previous work in that we focus solely on adult participants, as they have a full arsenal of cognitive resources at their disposal. If cognitively sophisticated adults are able to learn novel conservative quantifiers but unable to learn non-conservative quantifiers in the same experimental setup, it is unlikely that children would fare any better.

With these methodological changes in place, we find striking evidence that non-conservative quantifiers are unlearnable (i.e., that non-conservative quantificational meanings cannot be paired with determiner syntax in the natural course of language acquisition). In particular, we find that adult participants are able to learn novel conservative quantifiers, but fail to learn novel non-conservative quantifiers. In the same context though, adults are able to learn a non-conservative meaning when given verbal syntax. These results lend support to the idea that conservativity is a semantic universal; they bolster the case for theories that explain the lack of non-conservative determiners in the world’s languages in terms of the inability of the human language faculty to consider such form-meaning pairs.

2. Experiment 1: Learning by example

The seven experiments reported below follow the same general structure. Participants were told they were going to learn a new word, "gleeb", and were given a series of training trials pairing affirmative and negative "gleeb"-sentences with images of blue and orange circles and squares. After some number of training trials, participants were tested by being asked whether the same kinds of "gleeb"-sentences were true of novel images. Finally, they were asked explicitly to paraphrase a meaning for the novel word.

In Experiment 1, participants were trained 'by example'. They saw images either described as a situation in which "gleeb (of the) circles are blue" or a situation in which "it's not the case that gleeb (of the) circles are blue". Participants in the Conservative condition were given both affirmative and negative training trials consistent with the sentence in (5) having the conservative meaning in (5a). Participants in the Non-conservative condition of Experiment 1a were given both affirmative and negative training trials consistent with the sentence in (5) having the non-conservative meaning in (5b).

(5) Gleeb of the circles are blue.

a. ≈ *the circles outnumber the blue circles by 1*	conservative
b. ≈ *the circles outnumber the blue things by 1*	non-conservative
c. ≈ *the blue things outnumber the blue circles by 1*	non-conservative

Intuitively, (5a) is conservative as it does not require 'looking beyond' the circles, which is the internal argument of "gleeb" (it 'lives on' the circles). In contrast, (5b) is not conservative because by making reference to the blue things, one must 'look beyond' the circles.[3] Put another

[3] We can also apply the test given in Section 1 to demonstrate the conservativity of (5a) and the non-conservativity of (5b). "Gleeb" is conservative so long as "gleeb of the N are PRED" is true just in case "gleeb of the N are N that

way, this quantifier fails to 'live on' either argument (i.e., (5b) fails to be conservative in the same way that "equi" in (2) fails to be conservative). Experiment 1b probes performance on the non-conservative (5c), which is the same as (5a), but with the arguments behaving as if they were inverted. That is, (5c) can be thought of as expressing the same relation as (5a) does, but it 'lives on' its external argument instead of its internal argument. For this reason, (5c) fails to be conservative in the same what that "only" fails to be conservative ((5c) is to (5a) what "only" is to "every"). This makes (5c) non-conservative on the classical understanding, but conservative given Zuber and Keenan (2019)'s weakened notion of conservativity.

In general, if conservativity is a genuine semantic universal, we should observe superior learning in the Conservative condition (suggesting that conservative quantifiers are easier to learn than non-conservative ones) and above chance performance during the test phase only in the Conservative condition (suggesting that novel quantifiers are learnable in our experimental context, but non-conservative quantifiers are not). And, depending on which notion of conservativity turns out to be right, (5c) might either count as another Conservative condition (in which case it should pattern like (5a)) or another Non-conservative condition (in which case it should pattern like (5b)). Stimuli, results, and analyses for all experiments are available here: https://osf.io/qd4xp/?view_only=3eb005bc009c41cba69dcb2365e33eeb

2.1. Method

2.1.1. Subjects

are PRED" is true. Given the meaning for "gleeb" in (5a), we can convince ourselves that this equivalence holds. For example, *the circles outnumber the blue circles by 1* if and only if *the circles outnumber the circles that are blue circles by 1*. Since every circle that is blue is a circle, the second mention of the N "circle" is logically inert. On the other hand, the meaning for "gleeb" in (5b) fails this test. Consider a simple counter-example: there are two blue circles, one orange circle, and one blue square. Here, the circles do not outnumber the blue things by 1 (3 circles, 3 blue things), but the circles do outnumber the circles that are blue things by 1 (3 circles, 2 circles that are blue things). The issue is that the set of blue things and the set of circles that are blue things need not be identical.

In Experiment 1a, sixty English-speaking adults participated (half in the Conservative condition described in (5a), half in the Non-conservative condition described in (5b)). Another thirty English-speaking adults participated in Experiment 1b (all in the altered Non-conservative condition described in (5c)). All participants (here and in the experiments that follow) were native English speakers living in the United States. They were recruited on Prolific (www.prolific.co) and gave informed consent prior to participating. No participants were removed from analysis.

2.1.2. Procedure and materials

The experiment was designed using PCIbex (Zehr & Schwarz 2018), as were all subsequent experiments reported in this paper. Training consisted of 16 images of blue and orange circles and squares alongside text either describing the picture affirmatively as "gleeb of the circles are blue" or negatively as "it's not the case that gleeb of the circles are blue" (see Figure 1). Past artificial language learning studies have drawn conclusions about the (un)learnability of meanings on the basis of similar numbers of training trials (e.g., 10 in Saratsli & Papafragou 2023; 12 in Maldonado & Culbertson 2022), including in novel quantifier learning experiments (e.g., 5 in Hunter & Lidz 2013; 6 in He & Wellwood 2022).

To increase the salience of the non-conservative meaning, participants (in both conditions) were also reminded of the number of circles and blue shapes present in each display. Since this information is only helpful in the Non-conservative condition, it served to stack the deck against the Conservative condition without explicitly giving away the intended non-conservative meaning. All text presented on-screen was automatically read aloud to participants by the female voice "Joanna" from the text-to-speech synthesizer VoiceMaker (voicemaker.in).

Eight distinct scene-sentence pairs were used (4 affirmative, 4 negative), each displayed twice over the course of training, in a semi-random order (two randomly ordered blocks each containing all eight trial types). The particular training trials used were designed to rule out easily-hypothesized meanings such as "more than half". A full table of training trials is given in the Appendix. Training was immediately followed by six test trials in which participants were shown new displays, reminded of the number of circles and blue shapes, and asked "is it true that gleeb of the circles are blue?" (3 yes, 3 no). None of the training images were reused in the test portion (see the Appendix for details).

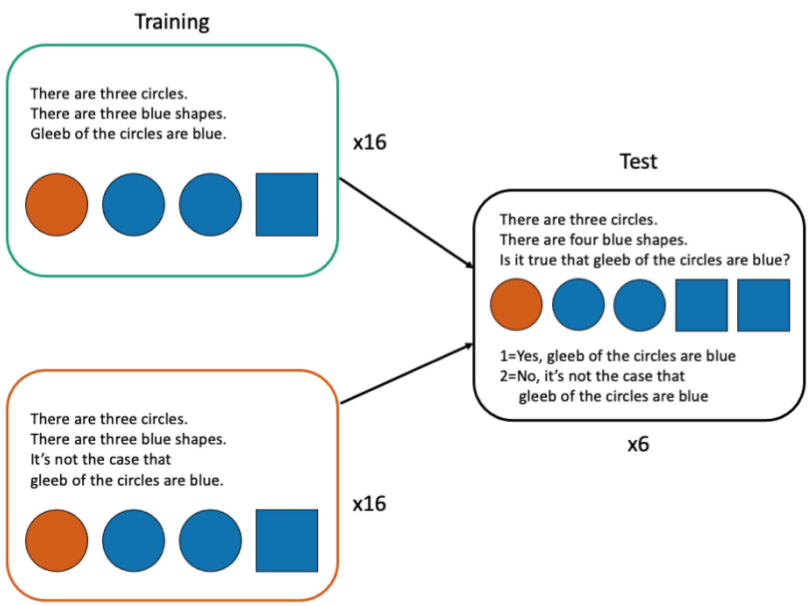

Figure 1: Example Experiment 1a training and test items from the Conservative (top; green outline) and Non-conservative (bottom; orange outline) conditions. Both conditions used affirmative and negative examples (an image that was described as "Gleeb of the circles are blue" in the Conservative condition might be described as "It's not the case that gleeb of the

circles are blue" in the Non-conservative condition, as in this particular example, but the inverse pairing also occurred). One particular trial is shown here for illustrative purposes, but the number of affirmative and negative training trials was matched in the two conditions (see complete stimuli details in the Appendix).

2.2. Results

In this and all subsequent experiments, we ask three main questions about participants' performance accuracy at test. First, did participants perform better in one condition than the other? We expect participants in the Conservative condition to perform better during the test portion than those in the Non-conservative condition (since conservative quantifiers should be easier to learn than non-conservative ones). Second, did participants in either condition perform better than chance (which, in this case, is answering 3 out of the 6 test trials correctly)? If non-conservative quantifiers are truly unlearnable, and not merely harder to learn, we expect above chance performance in only the Conservative condition. And third, how many participants in either condition achieved perfect scores? (We include this measure for purposes of comparison with past work, which reported it; see Section 1.2.) As a secondary, qualitative measure, we also examine the responses explicitly offered by participants.

As seen in Figure 2, participants in the Conservative condition of Experiment 1a were more accurate than those in the Non-conservative condition (78.9% vs. 52.2%). To verify this statistically, we fit a mixed effects model with condition as a fixed effect and random intercepts for participants and item. Model comparisons revealed that this model was a significantly better fit than an intercept-only model with the same random effects structure (according to a chi-square test for goodness of fit: $\chi^2(1) = 23.28$, $p<.001$; the model including condition versus the null model also had a lower AIC (436.4 vs. 457.7) and a lower BIC (451.9 vs. 469.3)). The model also showed a main effect of condition ($\beta = .66$ [95% CI .52 to .79], $z = 4.93$, $p<.001$). Moreover, while those in the Conservative condition achieved significantly higher accuracy than chance performance of 50% (significant intercept of an intercept-only model for the Conservative condition: $\beta = 1.72$ [95% CI 1.22 to 2.22], $z = 3.43$, $p<.001$), those in the Non-conservative condition did not (non-significant intercept of an intercept-only model for the Non-conservative condition: $\beta = .09$ [95% CI -.06 to .24], $z = 0.59$, $p=.553$). Lastly, those in the Conservative condition were more likely to obtain perfect accuracy than those in the Non-conservative condition (10 out of 30 vs. 0 out of 30). So while participants reliably did learn the novel conservative meaning from only 16 training trials, we find no evidence of learning in the Non-conservative condition of Experiment 1a.

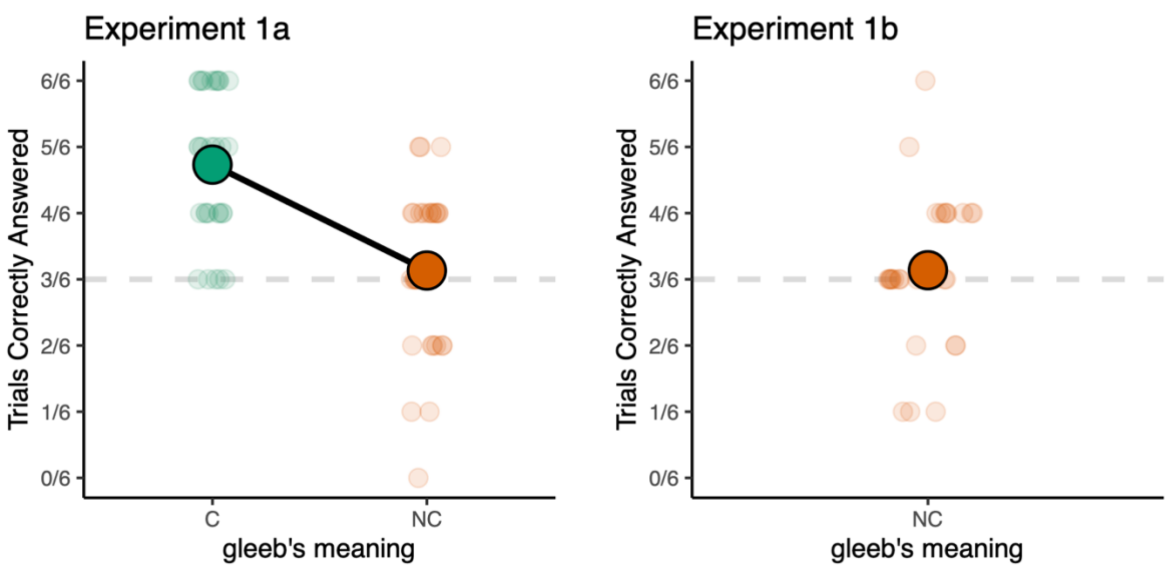

Figure 2: Number of test trials correctly answered in Experiment 1a (Conservative condition in green on the left; Non-conservative condition in orange on the right) and Experiment 1b (in which participants were exclusively trained with a non-conservative meaning). Chance

performance is 3/6 trials correctly answered. Large dots represent average performance; small translucent dots represent individual participant performance.

We likewise find no evidence of learning in the Non-conservative condition of Experiment 1b (non-significant intercept: $\beta = .09$ [95% CI -.08 to .26], $z = 0.55$, $p=.582$), where participants' performance was at chance level (52.2%) and only a single participant answered all six test trials correctly. This suggests that participants' failure to learn the non-conservative "gleeb" in Experiment 1a was not due to its being non-conservative in a particular way ('living on' neither argument and thus requiring a comparison between two sets). Instead, both kinds of non-conservativity ('living on' neither argument versus 'living on' the wrong argument) seem equally difficult to learn, especially when compared to a minimally-different conservative meaning (which 'lives on' its first argument).

Turning to the open-ended measure – asking participants what they think "gleeb" means – we find that participants, on the whole, were unable to paraphrase the novel meaning following the test phase. Even those who achieved perfect accuracy struggled to explicitly report the meaning they had learned for "gleeb". That said, participants tended to guess conservative meanings: 21 out of 30 participants in the Conservative condition offered responses that were conservative (e.g., "more than half, but not all"), despite not being entirely consistent with the intended "all but one" meaning in (5a), and 17 out of 30 participants in the Non-conservative condition of Experiment 1a offered conservative guesses, as did 16 out of 30 participants in the Non-conservative condition of Experiment 1b. This suggests that participants were treating the task as one of learning the meaning of a novel determiner, not merely one of learning a pattern. In the Conservative condition, only 6 out of 30 participants expressed some degree of uncertainty when asked to paraphrase “gleeb” (e.g., “I have no idea what gleeb means!”), whereas in the Non-conservative condition of Experiment 1a, nearly half of the participants (14 out of 30) expressed uncertainty (e.g., “No idea – everything seemed contradictory”).[4] In the Non-conservative condition of Experiment 1b, 8 out of 30 participants expressed some degree of uncertainty, and 3 others speculated that the meaning changes “at random/based on context”.

3. Experiment 1c: Removing the partitive frame

The results of Experiments 1a-b convincingly demonstrate the predicted learnability asymmetry between conservative and non-conservative meanings. But as noted in Section 1.3, one might wonder whether the results could be explained by the use of the partitive frame alone. On our view, “of the circles” signals to participants that they are learning a quantifier, and having decided on this syntactic category, they are unwilling to consider a non-conservative meaning. But an alternative view might hold that “of the circles” encourages attending to the circles more than anything else, and that such added attention will encourage participants to consider only conservative meanings.[5] To address this concern, Experiment 1c replicates Experiment 1a but with sentences that don’t use the partitive frame.

[4] A single participant in the Non-conservative condition of Experiment 1a expressed uncertainty but also offered a response consistent with the intended meaning in (5b). However, this dawned on them only after completing the test phase: “I think it means there are more circles than blue shapes, but I would need to go through the entire thing again to test that- I just realized that was probably it. But I'm not sure”. Their performance during the test portion was at chance.

[5] We are not particularly moved by this concern because while the frame “of the circles” does likely drive attention to the circles *per se*, the same can be said of quantificational phrases like “every circle” and “all circles”, which have been shown to drive attention to the circles even without the partitive frame (Knowlton 2021; Knowlton et al. 2021; Knowlton, Pietroski, Halberda, & Lidz 2022). We suspect this finding and the fact that quantifiers are conservative both have a common source (see Section 8.2). Perhaps relatedly, Matthewson (2001) argues that the partitive form is more ‘basic’ than the non-partitive, and is thus more revealing of the true nature of quantification.

3.1. Method

3.1.1. Subjects

Sixty English-speaking adults living in the United States participated (half in the Conservative condition, half in the Non-conservative condition). In this an all other experiments reported in this paper, participants were not allowed to participate if they had already participated in one of the other experiments. No participants were removed from analysis.

3.1.2. Procedure and materials

As in Experiment 1a-b, training consisted of 16 images of blue and orange circles and squares alongside text describing the picture. Unlike Experiment 1a-b though, the partitive frame was not used. Instead, pictures were described as “gleeb circles are blue” or “it’s not the case that circles are blue” . The same eight scene-sentence pairs from Experiment 1a-b were used, each displayed twice over the course of training, in a semi-random order (two randomly ordered blocks each containing all eight trial types). Training was immediately followed by six test trials, which differed from those used in Experiment 1a-b only in that the partitive frame was not used.

3.2. Results

As seen in Figure 3, participants in the Conservative condition were more accurate than those in the Non-conservative condition (68.3% vs. 47.2%). Mixed effects modeling confirmed that this difference was significant (condition model better fit than null model: $\chi^2(1) = 16.77$, $p<.001$, with lower AIC (480.5 vs. 495.3) and lower BIC (496.1 vs. 507.0); main effect of condition: $\beta = .45$ [95% CI .34 to .56], $z = 4.04$, $p<.001$). Participants were also more likely to obtain perfect accuracy (3 out of 30 vs. 0 out of 30). And while those in the Conservative condition achieved significantly higher accuracy than chance performance of 50% (significant intercept: β = .86 [95% CI .52 to 1.20], z = 2.51, p<.05), those in the Non-conservative condition did not (non-significant intercept: β = -.11 [95% CI -.26 to .04], z = 0.75, p=.456). As in Experiment 1a-b then, participants reliably learned the novel conservative meaning, but not the novel non-conservative meaning, in spite of the fact that here, the partitive frame was removed.

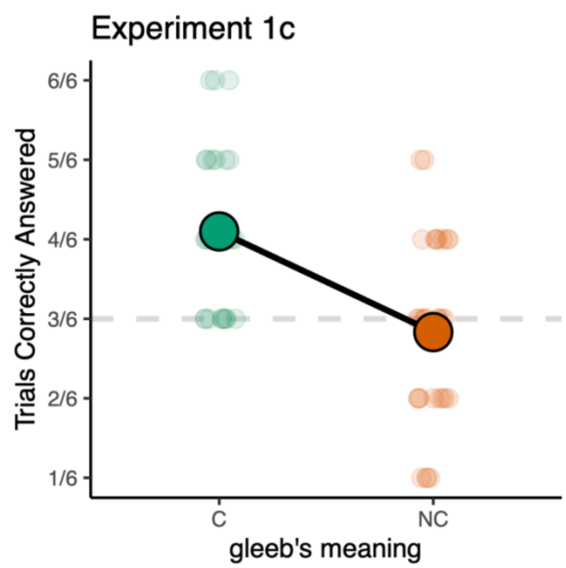

Figure 3: Number of test trials correctly answered in Experiment 1c (Conservative condition in green on the left; Non-conservative condition in orange on the right). Chance performance is 3/6 trials correctly answered. Large dots represent average performance; small translucent dots represent individual participant performance.

When asked what they thought "gleeb" meant at the end of the experiment, 11/30 participants in the Conservative condition and 5/30 participants in the Non-conservative offered conservative responses (none gave the conservative meaning in (5a)). In the Conservative condition, 3/30 participants expressed some uncertainty about the meaning of "gleeb"; in the

Non-conservative condition, 6/30 expressed uncertainty. Most illuminating, we think, is that across both conditions, only 28/60 participants offered meanings that referred to quantity. Non-quantity meanings for "gleeb" included things like "blue", "in between two objects of a different nature", "round", and "good and big". This confirmed our suspicion that the partitive frame is helpful as a signal to syntactic category, and that without it, some participants assumed that "gleeb" was an adjective or adverb. This potentially explains the slightly degraded performance in the Conservative condition compared to Experiment 1a, as well as the differences in qualitative responses.

4. Experiment 2: Generalizing to new predicates

Experiment 2 served as another replication of Experiment 1a, but with a more stringent test phase. In particular, participants were trained exactly as in Experiment 1a, with images of blue or orange shapes (and sentences like "gleeb of the circles are blue"), but were tested on images of shapes that either had stars or not (with sentences like "gleeb of the circles have stars"). In this way, Experiment 2 ensures that participants in the Conservative condition actually learned the meaning of the novel quantifier "gleeb", and were not merely relying on any visual similarities between the training and test images.

4.1. Method

4.1.1. Subjects

Sixty English-speaking adults participated over the internet (half in the Conservative condition, half in the Non-conservative condition). No participants were removed from analysis.

4.1.2. Procedure and materials

As in Experiment 1a, training consisted of 16 images of blue and orange circles and squares alongside text either describing the picture as “gleeb of the circles are blue” or “it’s not the case that gleeb of the circles are blue” (see Figure 4). To increase the salience of the non-conservative meaning, participants (in both conditions) were also reminded of the number of circles and blue shapes present in each display. All text presented on-screen was automatically read aloud to participants by a text-to-speech synthesizer.

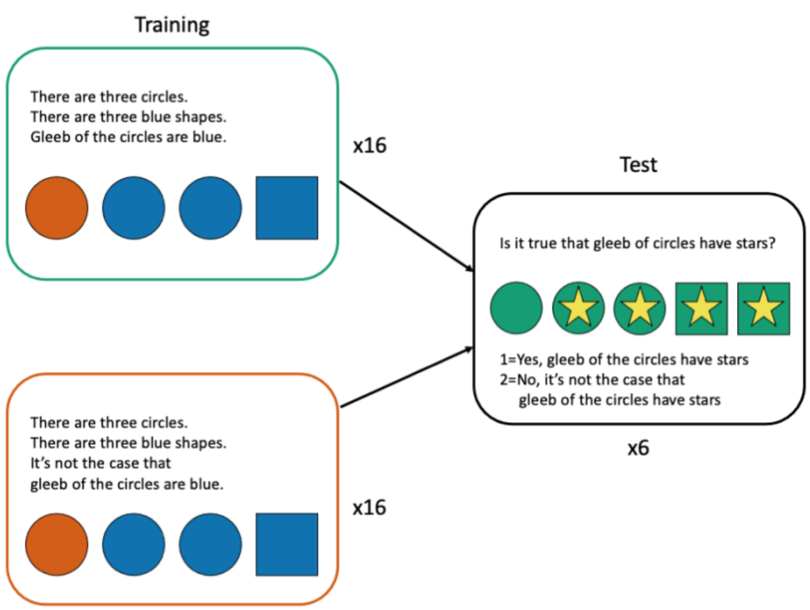

Figure 4: Example Experiment 2 training and test items from the Conservative (green) and Non-conservative (orange) conditions.

The same eight scene-sentence pairs from Experiment 1a were used in Experiment 2, each displayed twice over the course of training, in a semi-random order (two randomly ordered blocks each containing all eight trial types). Training was immediately followed by six test trials, which differed from those used in Experiment 1a in the following ways. Instead of differing in color, circles and squares in Experiment 2 were all green. Some were adorned with yellow stars whereas others were not. Unlike in Experiment 1a, participants were not given any extra information (i.e., the number of circles and the number of shapes with stars was not highlighted like the number of circles and the number of blue shapes was in Experiment 1a). Participants were asked "is it true that gleeb of the circles have stars?" (3 yes, 3 no). The training images were the same in all other respects (i.e., the number of circles and squares was not altered; any shapes that satisfied the relevant predicate in Experiment 1a – in that they were blue – satisfied the predicate in Experiment 2 in that they had a star).

4.2. Results

As seen in Figure 5, participants in the Conservative condition were more accurate than those in the Non-conservative condition (80% vs. 52.2%). Mixed effects modeling confirmed that this difference was significant (condition model better fit than null model: $\chi^2(1) = 21.03$, $p<.001$, with lower AIC (426.0 vs. 445.0) and lower BIC (441.6 vs. 456.7); main effect of condition: $\beta = .73$ [95% CI .58 to .88], $z = 4.75$, $p<.001$). Participants were also more likely to obtain perfect accuracy (8 out of 30 vs. 2 out of 30). And moreover, while those in the Conservative condition achieved significantly higher accuracy than chance performance of 50% (significant intercept: $\beta = 2.04$ [95% CI 1.33 to 2.75], $z = 2.87$, $p<.01$), those in the Non-conservative condition did not (non-significant intercept: $\beta = .1$ [95% CI -.12 to .32], $z = 0.46$, $p=.647$). As in Experiment 1 then, participants reliably learned the novel conservative meaning, but not the novel non-conservative meaning.

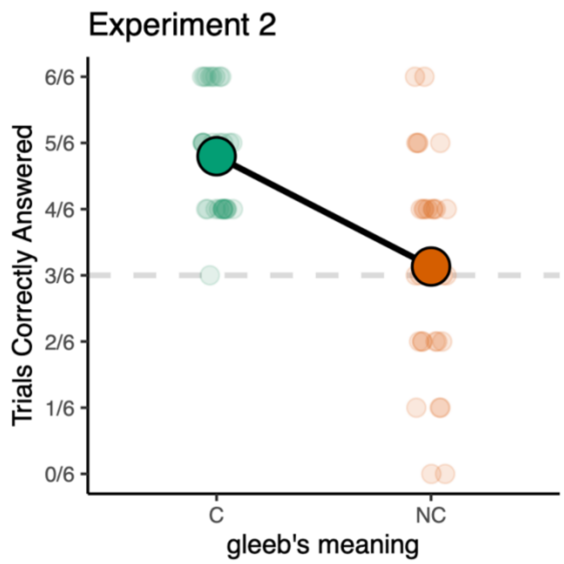

Figure 5: Number of test trials correctly answered in Experiment 2 (Conservative condition in green on the left; Non-conservative condition in orange on the right). Chance performance is 3/6 trials correctly answered. Large dots represent average performance; small translucent dots represent individual participant performance.

When asked what they thought "gleeb" meant at the end of the experiment, 24/30 participants in the Conservative condition and 24/30 participants in the Non-conservative offered conservative responses, though none gave the conservative meaning in (5a). In the Conservative condition, 8/30 participants expressed some uncertainty about the meaning of "gleeb", whereas in the Non-conservative condition, 13/30 expressed uncertainty.

5. Experiment 3: Explicit teaching

Whereas Experiment 2 aimed to make the task slightly more difficult than Experiment 1 by asking participants to generalize to a new predicate at test, Experiment 3 offered participants the best chance to learn the two novel meanings by implementing an 'explicit teaching' paradigm. Participants were explicitly told the meaning of "gleeb" before being walked through the same eight training trials from Experiments 1a and 2, with an explanation of why each sentence was a truthful description of each image.[6] After a five-minute break, participants were then asked to complete the test portion from Experiment 1a, save for one difference: instead of being asked "is it true that gleeb of the circles are blue?", participants were asked to fill in the blank in the sentence "___ of the circles are blue". Their choices were "Gleeb" and "It's not the case that gleeb".

Both of these additions – the five minute break and the fill-in-the-blank test – were included because we wanted to ensure that participants would succeed at this task because they paired the novel meaning with the determiner "gleeb", not merely by learning an algorithm by which to decide if a picture should get the answer "true" or "false". They delay served to permit forgetting any such algorithmic procedure. And the modified test trials served to remind participants that the task was about learning the meaning of the "gleeb", not sorting images into "yes" and "no" categories. Given these measures to defend against participants learning a picture-truth-value matching algorithm instead of a quantifier meaning, our expectation is that conservative meanings will still be learned but non-conservative ones will not. But, given the explicit teaching, if it is possible to pair the non-conservative meaning in (5b) with a quantificational determiner, participants in Experiment 3 should have an even better chance at doing so. And moreover, any worries about the length of training in Experiments 1-2 being too

[6] There was one difference between the training trials used in Experiment 3 and Experiments 1a and 2, noted in the Appendix: an extra blue circle was added to one of the false training trials to further help rule out the possible meaning "more than half but not all".

short should be assuaged by Experiment 3: no amount of training can replace giving the target of learning away at the beginning of the experiment.

5.1. Method

5.1.1. Subjects

Sixty English-speaking adults participated over the internet (half in the Conservative condition, half in the Non-conservative condition). No participants were removed from analysis.

5.1.2. Procedure and materials

After being given the general instructions, participants were told what “gleeb” means, as in (6). Those in the Conservative condition saw (and heard) (6a), whereas those in the Non-conservative condition received (6b).

(6) ‘Gleeb of the Xs are Y’ means that:

a. The number of Xs minus one is the number of Xs that are Y.

b. The number of Xs minus one is the number of Ys.

Training subsequently commenced. It consisted of 8 images of blue and orange circles and squares alongside text either describing the picture with “here, gleeb of the circles are blue because…” or “here, it’s not the case that gleeb of the circles are blue because…” (see Figure 6). The explanations referenced the number of circles and blue circles in the Conservative condition, and the number of circles and blue things in the Non-conservative condition. As in Experiments 1-2, all text presented on-screen was automatically read aloud to participants by a text-to-speech synthesizer. The same eight scene-sentence pairs from Experiments 1a and 2 were used, each displayed once in a random order.

Training was followed by an unrelated filler experiment, which took approximately five minutes. This filler experiment consisted of participants rating a series of images as “more like stuff” or “more like a thing”. After completing 41 trials of ‘stuff-thing’ rating, participants were tested on their knowledge of “gleeb”. They were given six test trials in which they were shown new displays and asked to fill in the blank: “___ of the circles are blue” (3 “gleeb”, 3 “it’s not the case that gleeb”). As noted above, the fill-in-the-blank question was used to remind participants that the task was about learning a novel quantifier, not merely sorting pictures according to some pattern. As in Experiments 1-2, none of the training images were reused in the test portion.

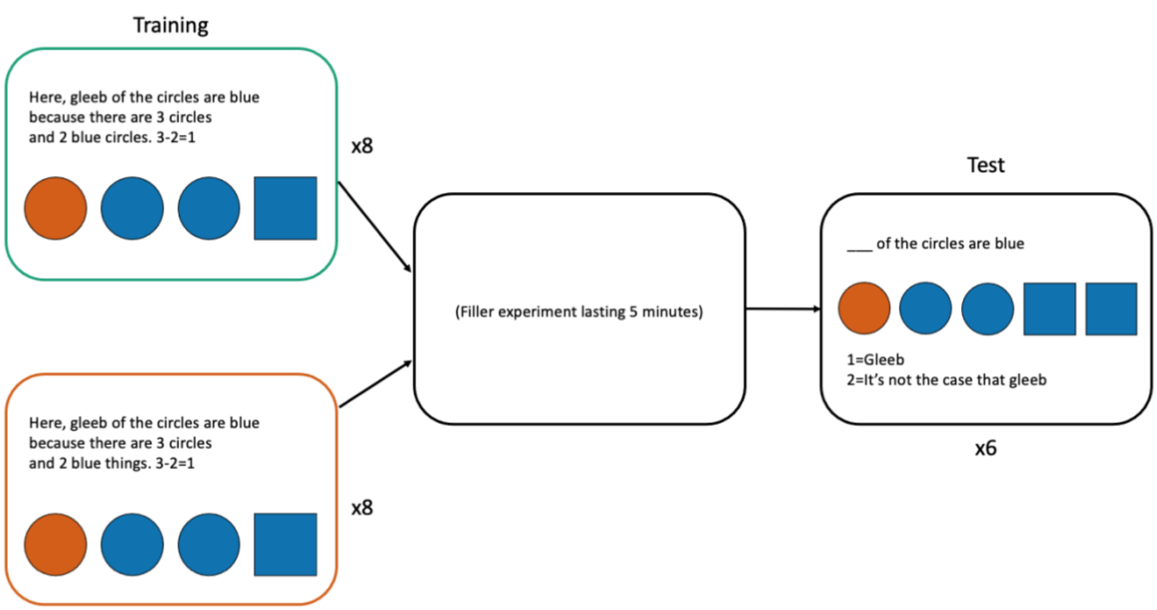

Figure 6: Example training and test items from the Conservative (green) and Non-conservative (orange) conditions of Experiment 3.

5.2. Results

As seen in Figure 7, participants in the Conservative condition were more accurate than those in the Non-conservative condition (80% vs. 54.6%). Mixed effects modeling confirmed that this difference was significant (condition model better fit than null model: $\chi^2(1) = 16.34$, $p<.001$, with lower AIC (419.7 vs. 434.0) and lower BIC (435.1 vs. 445.6); main effect of condition: $\beta = .69$ [95% CI .52 to .85], $z = 4.05$, $p<.001$). Participants were also more likely to obtain perfect accuracy (12 out of 30 vs. 1 out of 30). Moreover, while those in the Conservative condition achieved significantly higher accuracy than chance performance of 50% (significant intercept: $\beta = 1.74$ [95% CI 1.27 to 2.21], $z = 3.72$, $p<.001$), those in the Non-conservative condition did not (non-significant intercept: $\beta = .21$ [95% CI -.01 to .43], $z = 0.94$, $p=.346$). This bolsters the finding from Experiments 1-2: even when explicitly told the meanings, we only find evidence of learning in the Conservative condition.

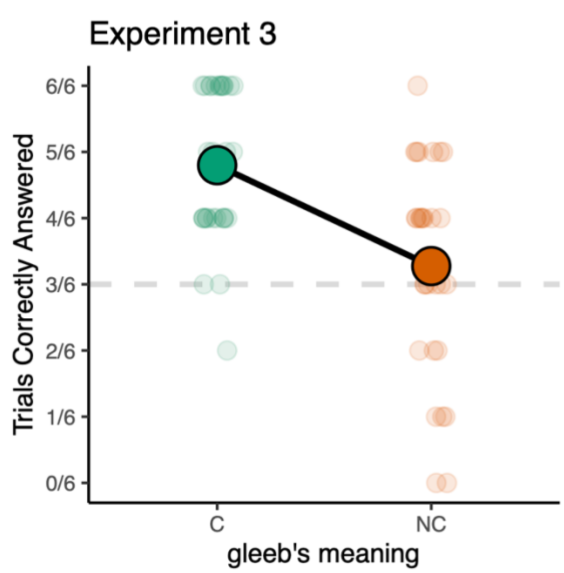

Figure 7: Number of test trials correctly answered in Experiment 3 (Conservative condition in green on the left; Non-conservative condition in orange on the right). Chance performance is 3/6 trials correctly answered. Large dots represent average performance; small translucent dots represent individual participant performance.

When asked what they thought "gleeb" meant, 10/30 participants in the Conservative condition correctly paraphrased the conservative (6a) (e.g., "yes, it means X-1 or all but one"; "Essentially, all but one"), 8 of whom achieved perfect accuracy at test. Only 5/30 participants in the Non-conservative condition correctly paraphrased (6b) (e.g., "x minus one equals y"; "the number of something minus one is the same as the other thing"), and none of these participants achieved perfect accuracy. Rates of reported uncertainty were similar to those in Experiments 1 and 2 (Conservative condition: 7/30; Non-conservative condition: 9/30).

6. Experiment 4: Verbal syntax

As noted in Section 1, conservativity is specific to quantifiers (in particular, the generalization has most often been discussed in the context of quantificational determiners). As such, all views that explain conservativity as a fact about grammar predict that the conservativity generalization is a constraint on what sorts of meanings can map to items from this particular syntactic category. This constraint prevents quantificational determiners from having non-conservative meanings, but it does not constitute a ban on those meanings being expressed in other ways. Therefore, any learnability constraint present is predicted to disappear if the meanings in question are taught without determiner syntax. In Experiment 4, we put this prediction to the test by teaching the same meanings from Experiments 1-3, but with verbal syntax (e.g., "the circles gleeb the blue shapes").

6.1. Method

6.1.1. Subjects

Sixty English-speaking adults participated over the internet (half in the Conservative condition, half in the Non-conservative condition). One participant (from the Non-conservative condition) was removed from analysis for failing to complete the task.

6.1.2. Procedure and materials

Training and test trials were nearly identical to those used in Experiment 1a. Training consisted of 16 images of blue and orange circles and squares paired with text which was read aloud by a text-to-speech synthesizer. The major difference in Experiment 4 was that the text paired with the images used "gleeb" as a verb. The text either described the images as a case where "(it's not the case that) the circles gleeb the blue circles" (in the Conservative condition) or "(it's not the case that) the circles gleeb the blue things" (in the Non-conservative condition) (see Figure 8). In both conditions, then, the verb itself expressed the relation 'outnumbers', but the whole sentence had a meaning that was either in line with the conservative (5a) or with the non-conservative (5b).

The same eight scene-sentence pairs from Experiment 1a were used, each displayed twice over the course of training, in a semi-random order (two randomly ordered blocks each containing all eight trial types). And as in Experiment 1a, participants in both conditions were reminded of the number of circles and blue shapes present in each display, alongside the critical training sentence. The training portion was immediately followed by six test trials in which participants were shown new displays, reminded of the number of circles and blue shapes, and either asked “is it true that the circles gleeb the blue circles?” (in the Conservative condition) or “is it true that the circles gleeb the blue things?” (in the Non-conservative condition) (3 yes, 3 no). None of the training images were reused in the test portion.

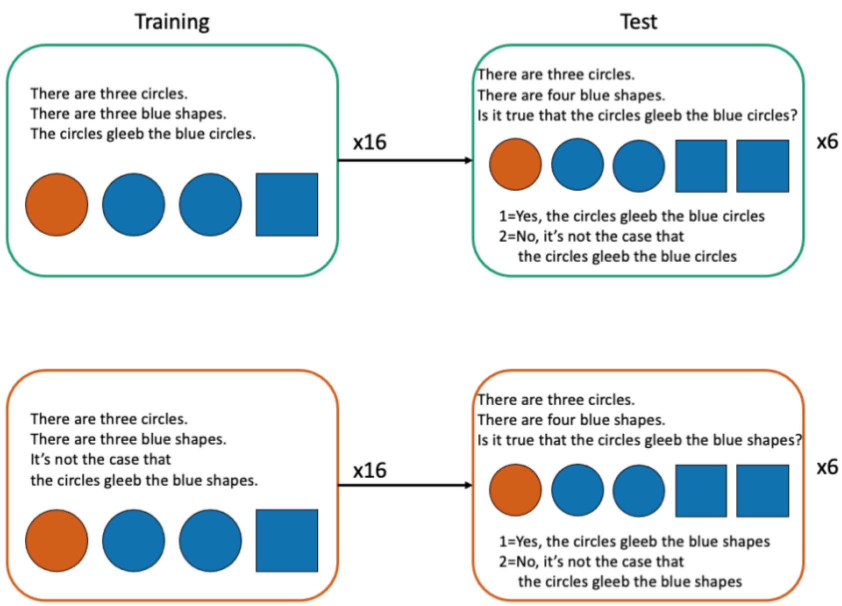

Figure 8: Example Experiment 4 training and test items from the Conservative (green) and Non-conservative (orange) conditions.

6.2. Results

As seen in Figure 9, participants in the Conservative condition were no more accurate than those in the Non-conservative condition (68.9% vs. 65.5%). Mixed effects modeling confirmed that this difference was not significant (condition model not a better fit than null model: $\chi^2(1) = 0.27$, p=.601, with higher AIC (425.1 vs. 423.4) and higher BIC (440.6 vs. 435.0);

no significant effect of condition: $\beta = .11$ [95% CI -.1 to .33], $z = 0.53$, $p=.598$). The gap between participants obtaining perfect accuracy likewise narrowed (9 out of 30 vs. 7 out of 29). And both groups of participants performed significantly better than chance performance of 50% accuracy (Conservative condition significant intercept: $\beta = 1.13$ [95% CI 0.78 to 1.48], $z = 3.22$, $p<.01$; Non-conservative condition significant intercept: $\beta = 0.8$ [95% CI 0.53 to 1.08], $z = 2.93$, $p<.01$).

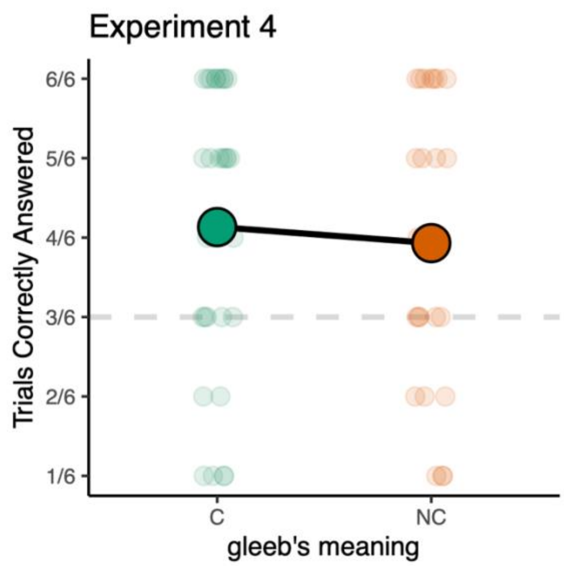

Figure 9: Number of test trials correctly answered in Experiment 4 (Conservative condition in green on the left; Non-conservative condition in orange on the right). Chance performance is 3/6 trials correctly answered. Large dots represent average performance; small translucent dots represent individual participant performance.

Finally, a similar proportion of participants in both conditions expressed some degree of uncertainty when asked to explicitly paraphrase “gleeb” (Conservative: 6/30; Non-conservative: 4/29). So as predicted, the advantage in the Conservative condition disappears when the same two meanings are taught with verbal syntax instead of as novel quantifiers. This further bolsters the case for thinking the differences observed here result from pairing particular kinds of meanings with determiners, not from any general conceptual differences between the two novel meanings. Here, as in Experiment 1c, the slightly degraded performance in the Conservative condition compared to earlier experiments likely stems from the fact that participants cannot use knowledge of syntactic category to constrain their hypotheses about possible meanings (i.e., quantificational determiners have more restrictive meanings than verbs).

7. Experiment 5: Testing another non-conservative meaning

Experiments 1-4 make the case that even cognitively sophisticated adults cannot pair the non-conservative meaning in (5b) or the non-conservative meaning in (5c) with a quantificational determiner. But the possibility remains that this failure to learn is a quirk of the particular non-conservative meanings used.[7] To address this potential concern, Experiment 5 used the same paradigm to test a third non-conservative meaning: (7b), often called "equi" in the literature.

---

[7] It is unclear what this quirk would be. The non-conservative meaning in (5b) doesn't strike us as communicatively useless, and there is no evidence that it is more conceptually complex than its conservative counterpart. It does run afoul of the often-discussed generalization that morphologically simple determiners are monotonic (Barwise & Cooper 1981; Icard & Moss 2014). That is, they either license 'subset to superset' inferences (e.g., if "every frog eats flies" it follows that "every frog eats bugs") or 'superset to subset' inferences (e.g., if "no frog eats flies" it follows that "no frog eats fruit flies"). Our non-conservative "gleeb" fails to license either kind of inference. If "gleeb of the frogs eat flies" means "the frogs outnumber the fly-eaters by one" then from "gleeb of the frogs eat flies" it does not follow that "gleeb frogs eat bugs" (imagine one frog that eats flies and one frog that eats ants) and it similarly does not follow that "gleeb frogs eat fruit flies" (imagine one frog that eats fruit flies and one dragonfly that eats fruit flies). Of course, our conservative "gleeb" – which essentially means "all but one" – runs afoul of this generalization too. If all but one frog eats flies it follows neither that all but one frog eats bugs (imagine one frog that eats flies and one frog that eats ants) nor that all but one frog eats fruit flies (imagine one frog that eats horse flies and one frog that eats ants). So a constraint on morphologically simple determiners being monotonic cannot explain the present results, even if we assume that participants are treating "gleeb" as morphologically simple.

(7) Gleeb of the circles are blue.

a. ≈ *the circles and the blue circles are equal in number* conservative

b. ≈ *the circles and the blue things are equal in number* non-conservative

The contrasting, conservative, meaning is given in (7a).[8] We predict a similar contrast will exist in Experiment 5 between the learnability of (7a) and (7b).

7.1. Method

7.1.1. Subjects

Sixty English-speaking adults participated over the internet (half in the Conservative condition, half in the Non-conservative condition). No participants were removed from analysis.

7.1.2. Procedure and materials

The training and test procedure was nearly identical to that of Experiment 1. Training consisted of 16 images of blue and orange circles and squares paired with text which was read aloud by a text-to-speech synthesizer. The text either described the images with "here, gleeb of the circles are blue" or "here, it's not the case that gleeb of the circles are blue" (see Figure 10).

The major difference in Experiment 5 was the actual images used. Eight scene-sentence pairs were designed and each displayed twice over the course of training, in a semi-random order (two randomly ordered blocks each containing all eight trial types). A full table of training trials

[8] It is worth noting that this conservative "gleeb" can be paraphrased with one word in English: "all". That is, whenever (7a) is true of some scene, "all of the circles are blue" will likewise be true. As such, it would be relatively unsurprising if participants had an easy time learning the meaning of "gleeb" in (7a). Nonetheless, participants' failure to learn the non-conservative (7b) would point to our earlier results generalizing beyond the particular non-conservative meaning used in Experiments 1-4.

is given in the Appendix. Training was immediately followed by six test trials in which participants were shown new displays, and asked “is it true that the gleeb of the circles are blue?” (3 yes, 3 no). As in prior experiments, none of the training images were reused in the test portion.

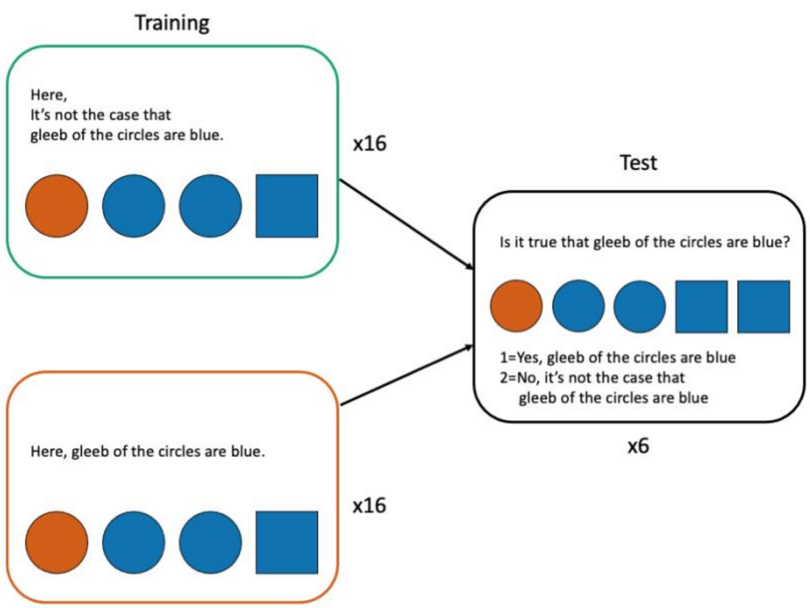

Figure 10: Experiment 5 training and test items from the Conservative (green) and Non-conservative (orange) conditions.

7.2. Results

As seen in Figure 11, participants in the Conservative condition were more accurate than those in the Non-conservative condition (97.2% vs. 38.3%). Mixed effects modeling confirmed that this difference was significant (condition model better fit than null model: $\chi^2(1) = 77.34$, $p<.001$, with lower AIC (269.3 vs. 344.7) and lower BIC (284.9 vs. 356.3); main effect of condition: $\beta = 2.58$ [95% CI 2.21 to 2.96], $z = 6.93$, $p<.001$). Participants were also more likely to obtain perfect accuracy (27 vs. 1). And moreover, while those in the Conservative condition achieved significantly higher accuracy than chance performance of 50% (significant intercept: β = 54.51 [95% CI 47.03 to 62], z = 7.28, p<.001), those in the Non-conservative condition did not (non-significant intercept: β = -.69 [95% CI -1.28 to -0.09], z = -1.15, p=.249). As in Experiments 1-3 then, participants reliably learned the novel conservative meaning, but, crucially, failed to show any evidence of learning the novel non-conservative meaning.

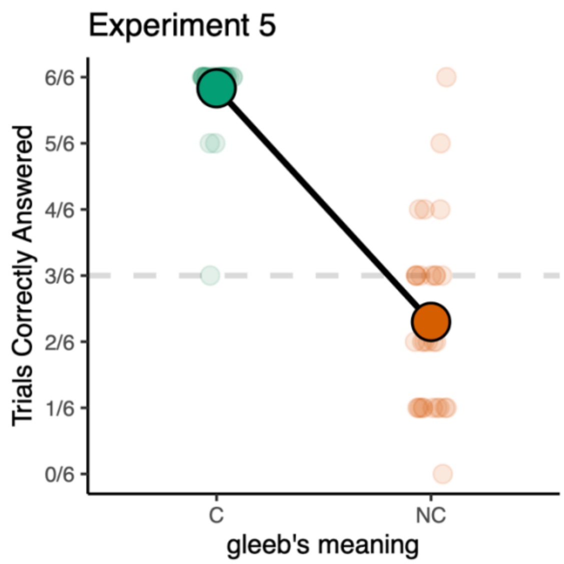

Figure 11: Number of test trials correctly answered in Experiment 5 (Conservative condition in green on the left; Non-conservative condition in orange on the right). Chance performance is 3/6 trials correctly answered. Large dots represent average performance; small translucent dots represent individual participant performance.

When asked what they thought “gleeb” meant, 24/30 participants in the Conservative condition offered responses consistent with the conservative meaning in (7a) (e.g., “a word for all of the”), whereas 0/30 participants in the Non-conservative condition offered responses with the intended meaning in (7b). Instead, participants mostly either guessed a conservative meaning (18) or expressed uncertainty (7).

8. Discussion

Quantificational determiners across the world's languages obey the conservativity generalization: they 'live on' their internal argument. This generalization, which is tied to a particular syntactic category, seems unrelated to general conceptual complexity or communicative usefulness and instead seems a prime target for an internal design feature of human language. If so, the (non-)conservativity of quantifiers should be expected to have implications for their learnability. The seven experiments presented here (see Table 1 for summary of findings) support this prediction. We find that adults show a considerable advantage for pairing a novel conservative meaning with a quantifier but fail to show any evidence of learning to pair any of the three novel non-conservative meanings tested with a quantifier. This is true even when the deck is stacked in favor of the non-conservative meaning, and, surprisingly, even when the non-conservative meaning is explicitly taught. Importantly, adults in the same experimental context are able to learn to express the same novel relations with verbal syntax, confirming that the learnability difference is about making this particular form-meaning pairing (pairing a non-conservative meaning with a quantificational determiner).

Exp	Conservative mean % correct	Non-conservative mean % correct
1a	78.9 (3.4) *	52.2 (3.8)
1b	-	52.3 (3.4)
1c	68.3 (2.9) *	47.2 (3.5)
2	80.0 (2.7) *	52.2 (5.0)
3	80.0 (3.5) *	54.6 (4.9)
4	68.9 (5.4) *	65.5 (5.0) *
5	97.2 (1.8) *	38.3 (4.2)

Table 1: Summary of mean accuracy by condition in each experiment (standard error in parenthesis). Participants in the Conservative condition outperformed those in the Non-conservative condition in every experiment except Experiment 4 (verbal syntax). Stars indicate that accuracy is significantly different from chance (50%). Performance in the Non-conservative condition only differed from chance in Experiment 4 (verbal syntax).

As discussed in Section 1.2, previous experiments probing the link between conservativity and learnability had not led to as decisive results. We suspect there were two main factors that differentiated our experiments from past work. First, that our experiments used completely novel meanings might have made learning easier in general. As noted “not all”, which has been used in the past, might have been difficult to learn for other (possibly pragmatic) reasons. Second, the fact that our experiments were explicitly about word learning and (mostly) used novel words in the partitive frame may have served to highlight that the point of the task was to learn a novel quantifier. It is only in the context of quantifier learning that we should expect to see the predicted learnability asymmetry. This prediction, which was borne out in our results, had not been tested in prior work.[9] In any case, the present set of findings confirms the long-held expectation among linguists that the typological generalization of conservativity is unlikely to be a historical accident or epiphenomenon (*cf.* Steinert-Threlkeld 2021), but rather a genuine semantic universal, fundamentally revealing about the nature of the language faculty.

[9] An interesting direction for future work that eschews the partitive in favor of sentences like “gleeb circles are blue” might be to ask what syntactic category participants think “gleeb” belongs to. That is, participants might be able to learn a non-conservative meaning for “gleeb” if they think the term is a focus operator like “only”. This might be diagnosed by asking participants to use the novel word in a production experiment, and seeing if they produce it in floating positions (e.g., “the circles gleeb are green”). It also might be informative to teach “gleeb” with explicitly adverbial morphology (e.g., “the circles are gleebly green”), though this could introduce unnecessary ambiguity (e.g., it might be taken to mean “often” or “apparently”).

8.1 Alternative accounts

Could our learnability results merely reflect a case of transfer from participants' native language? After all, since all quantifiers are conservative, all of the quantifiers that speakers in our experiments know are conservative. Maybe this knowledge – however it got there – is responsible for how they behave when they're asked to learn a new quantifier. In response to this worry, we think an important lesson from the large literature using artificial language learning experiments is that one cannot assume that participants will only (or will more easily) acquire things that are present in their native language. To take a recent example, Saratsli and Papafragou (2023) show that participants can quite easily learn an evidentiality marker that is not present in their native language (and they can do so with only 10 training trials, which is fewer than we used here). Non-conservative quantifiers could have turned out the same way, but didn't, even when given the best possible chance. Namely, in our "explicit teaching" experiment (Experiment 3), we told participants the intended conservative and non-conservative meanings and still found that they can only learn the conservative one. Any native-language bias that pushes against a non-conservative meaning should have been eliminated in that experiment.

One might argue that the above experiments only show that non-conservative quantifiers are hard to learn, not unlearnable (especially in light of results like those reported in Ramotowska 2022, discussed in Section 1.2). This point is well taken. We cannot provide dispositive evidence that something is unlearnable; no experiment can. This concern is not unique to the present study, though. Artificial language learning experiments often show that unattested but licit patterns are learnable under certain conditions whereas other patterns are not learned under the same conditions. These results are used to draw non-demonstrative inferences about (un)learnability. There are, of course, always ways to strengthen the inferential chain, including: more training, explicit teaching, and incentivization. Our explicit teaching experiment, longer training sessions, and simplified displays go some way toward strengthening the inference in the present case. The same can be said for the fact that in the Non-conservative conditions we find performance no different from chance except in the case where the meaning was taught as a novel verb.

Moreover, the unlearnability claim has an additional point in its favor: theoretical proposals on which non-conservative quantifiers are not able to be generated by the language faculty. If they can't be generated, then they can't be hypotheses available for learning. The subsequent section discusses three such proposals, all of which predict non-conservative quantifiers to be unlearnable, rather than merely hard to learn.

8.2 The source of conservativity

As noted in Section 1, there are at least three views about the source of the generalization that fall under the header of 'grammatical explanations'. Our results do not directly rule out any of these three possibilities, though the possibilities considered do differ in terms of how central they take conservativity to be, their explanatory power, and their supplementary commitments.

The most straightforward grammatical view is that non-conservative determiner meanings are 'filtered out' at the lexical level. That is, only conservative relations and conservativity-preserving operations are available as building blocks for lexicalization. This is not to say that non-conservative relations – like identity or numerical equivalence – are absent from cognition more generally, just that these relations are not in learners' hypothesis space when they are tasked with acquiring determiner meanings. Keenan and Stavi (1986) show how such a view might work by defining a set of basic conservative relations (including the relations expressed by “every” and “at least four”) and a set of basic operations, which preserve conservativity when combined with any of the basic relations (e.g., “at least four *and* at most ten”). We could stipulate that only these relations and operations are available for lexicalization in the quantificational domain. But while such a claim would situate conservativity as a central fact about quantifier semantics – it would be a fact about language *per se* – the explanation of its existence would essentially amount to a redescription of the facts.

An explanatorily richer approach would make use of the fact that determiners are a distinct syntactic category that combine with internal (NP) and external (tensed VP) arguments. That is, we might attempt to explain conservativity as a consequence of the syntax-semantics interface. Romoli (2015) offers a spelled-out version of one such view. Building on earlier suggestions (Fox 2002; Ludlow 2002; Sportiche 2005), Romoli argues that non-conservative relations might well get lexicalized, but are ‘filtered out’ at the syntax-semantics interface in one of two ways. Either they result in sentence meanings that are truth-conditionally equivalent to sentences with conservative determiners or they lead to ungrammatical sentences.

Intuitively, the proposal is that a sentence like “every frog is green” literally means “every frog is a frog that is green”. Formally, this is achieved thanks to an ancillary hypothesis about how traces of quantifier raising are interpreted. But the consequence is that any quantificational sentence has an instance of its internal argument (the NP with which the quantifier combines) inside its external argument. So if the non-conservative “equi” – meaning “the internal and external arguments are equal in number” – were to exist in English, a sentence like “equi frog is green” would in fact mean “the frogs and the green frogs are equal in number”. As noted above, this would make “equi” truth-conditionally equivalent to “every”.[10] And if

[10] A potentially surprising consequence of this view is that “every” might not express the (conservative) subset relation, as is often assumed by standard accounts, but the (non-conservative) numerical equivalence relation.

“only” were to be lexicalized as a determiner, perhaps pronounced “donly”, then a sentence like “donly frogs are green” would in fact mean “only frogs are green frogs”. But this meaning is, in some sense, trivial: it will always be true since nothing that is a green frog can fail to be a frog (i.e., it is just as trivial as “only frogs are frogs”). If one adopts the auxiliary assumption that trivial sentence meanings are ungrammatical (e.g., Fox & Hackl 2006; Gajewski 2002), then “donly” would thus be ungrammatical in any sentence in which it was used.

An alternative syntax-semantics interface explanation of conservativity advocates jettisoning the relational conception of quantifier meanings altogether, and replacing it with the older notion of restricted quantification (Pietroski 2004, 2018; Westerståhl 2019; Lasersohn 2021; Knowlton et al. 2021; Ludlow & Živanović 2022). The leading idea behind this ‘non-relational’ view is that quantifiers are not tools for relating two sets supplied by both arguments (“the fish are a subset of the swimming things”), but are instead tools for restricting attention to the internal argument and subsequently specifying how many things meet the condition supplied by the external argument (“the fish are such that the predicate *swim* applies universally”). Given this architecture, stating non-conservative meanings is impossible. For example, one cannot paraphrase “equi” from (2) or “everynon” from (3) or “donly” from the example above if the domain is initially restricted to the fish.

What makes this proposal a syntax-semantics interface explanation is that it respects the syntactic difference between the quantifier’s two arguments. Instead of treating both arguments on a semantic par by supposing that they both denote sets (which obscures any syntactic differences between them), the non-relational view maintains that the internal (NP) argument is treated as a group of things whereas the external (tensed VP) argument is treated as an open sentence which can apply or fail to apply to some things. So instead of resulting from a conspiracy of other factors, conservativity, on this 'non-relational' view, is a logical consequence of what kinds of things determiner meanings and the meanings of their arguments are. This view thus offers a simple explanation of conservativity, respects the syntactic distinction between a determiner's two arguments, and avoids taking on any ancillary hypotheses about the interpretation of quantifier raising or the (un)grammaticality of certain trivial sentences.

Moreover, Knowlton, Pietroski, Williams, Halberda, and Lidz (2023) provide psycholinguistic evidence in favor of this proposal. They show that when asked to evaluate sentences like "every big circle is blue", participants mentally represent the big circles as such, but avoid mentally representing the blue things or even the big blue circles as such. For example, in a case where participants are shown 9 big circles, all of which are blue, and asked whether "every big circle is blue", they correctly answer in the affirmative but can only accurately recall the number of big circles; not the number of blue things or the number of big blue circles (which is identical in this example). This asymmetry between how the two arguments are treated is exactly the predicted result given the 'non-relational' view. In contrast, it is potentially a surprising result given the other two views, which maintain that a sentence like "every big circle is blue" expresses a relation between the big circles and a second set (either the blue things or the big blue circles).

In any case, the results presented here are compatible with all three of these views about the source of the conservativity generalization. Each view predicts participants to be unable to pair our non-conservative meanings with a determiner, though for different reasons. The case of "equi", used in Experiment 5, was discussed above. But consider our non-conservative "gleeb" from Experiments 1-3, meaning "the internal argument outnumbers the external argument by one". On the 'lexical filtering' view, this non-conservative "gleeb" is stipulated to be absent from the hypothesis space of possible determiner meanings that learners can consider. On the ‘interface filtering’ view, “gleeb of the circles are blue” in fact means “the circles outnumber the circles that are blue by one”, given the posited interpretation of quantifier raising. But this sentence meaning is truth-conditionally equivalent to a sentence with a conservative determiner; in fact, it’s identical to our conservative “gleeb”.[11] Finally, on the ‘non-relational’ view, the conservative “gleeb” meaning is stateable (“the circles are such that the predicate *be blue* applies to everything but one”), whereas the non-conservative “gleeb” meaning is not (“the circles are such that the predicate *be blue* applies to ???”).

But while our results do not lend support to one particular hypothesis about the source of the generalization, they do point to the notion that it deserves a deep grammatical explanation. That is, conservativity seems to be an important consequence of how the language faculty is organized. It is more like the fact that learners expect determiners to be about quantity (Syrett, Musolino, & Gelman 2012; Wellwood, Gagliardi, & Lidz 2016) than it is like the fact that they have difficulty learning negative meanings for determiners, which is likely to have a more general conceptual basis (He & Wellwood 2022). As such, the learnability results presented above lend support to semantic theories that account for this generalization as a core phenomenon indicative of the nature of the grammar as opposed to a peripheral one deserving an ad hoc justification. And, relating this discussion to the previous section, the fact that these language-specific views all predict the unlearnability of non-conservative quantifiers supports the idea that our experimental results are best interpreted as suggesting that non-conservative quantifiers are unlearnable, not just that they are hard to learn.

[11] On the ‘interface filtering’ view, our conservative “gleeb” also has a different meaning, but it is truth-conditionally equivalent to the one we intended. Roughly, our conservative “gleeb” means “all but one”. So, given the proposed interpretation of quantifier raising, a sentence like “gleeb of the circles are blue” does not mean “all but one of the circles are blue”, but instead “all but one of the circles are blue circles”.

8.3 Conclusion

Natural language quantifiers are conservative: their internal argument ‘sets the scene’. A sentence like “every frog is green” is about the frogs (and consequently, it is logically equivalent to a sentence like “every frog is a frog that’s green”). Since this generalization is perhaps the most robust semantic universal to ever be discovered, it invites a learnability claim: no non-conservative quantifiers exist because such meanings are absent from the hypothesis space available to learners. It might be that human minds are not built to entertain non-conservative determiners. And since the conservativity generalization seems unfriendly to explanations in terms of general cognitive or communicative constraints, conservativity is a good candidate for being a typological universal that reflects a fundamental property of the language faculty. This position predicts that non-conservative determiners should be unlearnable. In the results discussed above, this prediction is borne out. Thus the present study supports semantic theories on which conservativity is a cornerstone of the human semantic potential.

References

Barwise, J. & Cooper, R. (1981). Generalized quantifiers and natural language. In *Philosophy, language, and artificial intelligence*, 241-301. http://dx.doi.org/10.1007/BF00350139

Chomsky, N. (1965). *Aspects of the theory of syntax*. MIT Press.

Clark, E. V. (1971). On the acquisition of the meaning of before and after. *Journal of Verbal Learning and Verbal Behavior*, *10*, 266-275. https://doi.org/10.1016/S0022-5371(71)80054-3

Cohen, A. (2001). Relative readings of many, often and generics. *Natural Language Semantics*, *9*, 41-67. http://dx.doi.org/10.1023/A:1017913406219

Culbertson, J., Smolensky, P., & Legendre, G. (2012). Learning biases predict a word order universal. *Cognition*, *122*(3), 306-329. http://dx.doi.org/10.1016/j.cognition.2011.10.017

Donaldson, M., & Balfour, G. (1968). Less is more: A study of language comprehension in children. *British Journal of Psychology*, *59*(4), 461-471. http://dx.doi.org/10.1111/j.2044-8295.1968.tb01163.x

Ehri, L. C. (1976). Comprehension and production of adjectives and seriation. *Journal of Child Language*, *3*(3), 369-384. http://dx.doi.org/10.1017/S0305000900007248

von Fintel, K. & Matthewson, L. (2008). Universals in semantics. *The Linguistic Review*, *25*, 139-201. https://doi.org/10.1515/TLIR.2008.004

Fintel, K. V. (1997). Bare plurals, bare conditionals, and only. *Journal of Semantics*, *14*(1), 1-56. https://doi.org/10.1093/jos/14.1.1

Ahn, D., & Sauerland, U. (2017). Measure constructions with relative measures: Towards a syntax of non-conservative construals. *The Linguistic Review*, *34*(2), 215-248. https://doi.org/10.1515/tlr-2017-0001

Ahn, D., & Ko, H. (2022). On non-conservativity of Korean floating quantifiers. *Glossa: a journal of general linguistics*, *7*(1). https://doi.org/10.16995/glossa.5776

Fox, D. (2002). Antecedent-contained deletion and the copy theory of movement. *Linguistic Inquiry 33*(1). 63–96. https://doi.org/10.1162/002438902317382189

Fox, D. & Hackl, M. (2006). The universal density of measurement. *Linguistics and Philosophy* 29(5). 537–586. https://doi.org/10.1007/s10988-006-9004-4

Gajewski, J. (2002). L-analyticity and natural language. *Manuscript, MIT* http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.488.3195

Gentner, D., & Bowerman, M. (2010). Why some spatial semantic categories are harder to learn than others: The typological prevalence hypothesis. In Lieven, E., Guo, J., Nakamura, K., Budwig, N., Ozcaliskan, S., & Ervin-Trupp, S. (Eds.) *Crosslinguistic approaches to the psychology of language: Research in the tradition of Dan Isaac Slobin*, 465-480. Taylor & Francis.

He, A. X. & Wellwood, A. C. (2022). "Most" is easy but "least" is hard: Novel determiner learning in 4-year-olds. In Culbertson, J., Perfors, A., Rabagliati, H., & Ramenzoni, V. (Eds.) *Proceedings of the 44th Annual Conference of the Cognitive Science Society*, 1056-1063. https://escholarship.org/uc/item/5hh4m526

Heim, I. & Kratzer, A. (1998). *Semantics in generative grammar*. Oxford: Blackwell.

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

Herburger, E. (2000). *What counts: Focus and quantification*. MIT Press.

Higginbotham, J. & May, R. (1981). Questions, quantifiers and crossing. http://dx.doi.org/10.1515/tlir.1981.1.1.41

Hunter, T. & Lidz, J. (2013). Conservativity and learnability of determiners. *Journal of Semantics*, *30*(3), 315-334. http://dx.doi.org/10.1093/jos/ffs014

Icard III, T. F., & Moss, L. S. (2014). Recent progress on monotonicity. *Linguistic Issues in Language Technology*, *9*, 167-194. https://doi.org/10.33011/lilt.v9i.1325

Keenan, E. L. (1981). A boolean approach to semantics. In Groenendijk, J. (Ed.). *Formal Methods in the Study of Language*, 343–79. Mathematisch Centrum.

Keenan, E. L. & Stavi, J. (1986). A semantic characterization of natural language determiners. *Linguistics and Philosophy*, *9*(3), 253–326. https://doi.org/10.1007/BF00630273

Knowlton, T. (2021). *The psycho-logic of universal quantifiers*. University of Maryland dissertation. https://doi.org/10.13016/fdr8-3qqh

Knowlton, T., Pietroski, P., Williams, A., Halberda, J., & Lidz, J. (2021). Determiners are "conservative" because their meanings are not relations: Evidence from verification. *Semantics and Linguistic Theory*, *30*, 206-226.

Knowlton, T., Pietroski, P., Halberda, J., & Lidz, J. (2022). The mental representation of universal quantifiers. *Linguistics and Philosophy*, *45*, 911-941. https://doi.org/10.1007/s10988-021-09337-8

Knowlton, T., Pietroski, P., Williams, A., Halberda, J., & Lidz, J. (2023). Psycholinguistic evidence for restricted quantification. *Natural Language Semantics*, *31*(2), 219-251. https://doi.org/10.1007/s11050-023-09209-w

Larson, R. K. & Segal, G. (1995). *Knowledge of meaning: An introduction to semantic theory*. MIT Press.

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

Ludlow, P. (2002). LF and natural logic. In Preyer, G. & Peter, G. (Eds.) *Logical form and language*, 132-168. Oxford University Press.

Ludlow, P., & Zivanović, S. (2022). *Language, Form, and Logic: In Pursuit of Natural Logic's Holy Grail*. Oxford University Press.

Matthewson, L. (2001). Quantification and the nature of crosslinguistic variation. *Natural Language Semantics*, *9*(2), 145-189. https://doi.org/10.1023/A:1012492911285

Maldonado, M., & Culbertson, J. (2022). Person of interest: Experimental investigations into the learnability of person systems. *Linguistic Inquiry*, *53*(2), 295-336. https://doi.org/10.1162/ling_a_00406

Pasternak, R., & Sauerland, U. (2022). German measurement structures: Case-marking and non-conservativity. *The Journal of Comparative Germanic Linguistics*, *25*(2), 221-272. https://doi.org/10.1007/s10828-022-09134-y

Pietroski, P. M. (2004). *Events and Semantic Architecture*. Oxford University Press. http://dx.doi.org/10.1093/acprof:oso/9780199244300.001.0001

Pietroski, P. M. (2018). *Conjoining Meanings: Semantics Without Truth Values*. Oxford University Press.

van de Pol, I., Lodder, P., van Maanen, L., Steinert-Threlkeld, S., & Szymanik, J. (2023). Quantifiers satisfying semantic universals have shorter minimal description length. *Cognition*, *232*, 105150. https://doi.org/10.1016/j.cognition.2022.105150

Ramotowska, S. (2022). *Quantifying quantifier representations: Experimental studies, computational modeling, and individual differences* (Doctoral dissertation, University of Amsterdam). https://eprints.illc.uva.nl/id/eprint/2202/

Romero, M. (2015). The conservativity of many. In *Proceedings of the 20th Amsterdam Colloquium*, 20-29.

Romero, M. The Conservativity of *Many*: Split Scope and *Most*. *Topoi*, 37, 393–404 (2018). https://doi.org/10.1007/s11245-017-9477-5

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

Ryalls, B. O. (2000). Dimensional adjectives: Factors affecting children's ability to compare objects using novel words. *Journal of Experimental Child Psychology*, *76*(1), 26-49. https://doi.org/10.1006/jecp.1999.2537

Saratsli, D., & Papafragou, A. (2023). Pragmatic effects on semantic learnability: Insights from evidentiality. *Journal of Memory and Language*, *131*, 104426. https://doi.org/10.1016/j.jml.2023.104426

Spenader, J. & De Villiers, J. (2019). Are conservative quantifiers easier to learn? Evidence from novel quantifier experiments. In *Proceedings of the 22nd Amsterdam Colloquium*, 504-512.

Steinert-Threlkeld, S. & Szymanik, J. (2019). Learnability and semantic universals. *Semantics and Pragmatics*, *12*(4). https://doi.org/10.3765/sp.12.4

Steinert-Threlkeld, S. (2021). Quantifiers in natural language: Efficient communication and degrees of semantic universals. *Entropy*, *23*(10), 1335. https://doi.org/10.3390/e23101335

Syrett, K., Musolino, J., & Gelman, R. (2012). How can syntax support number word acquisition?. *Language Learning and Development*, *8*(2), 146-176. https://doi.org/10.1080/15475441.2011.583900

Tesar, B. & Smolensky, P. (1998). Learnability in optimality theory. *Linguistic Inquiry*, *29*(2), 229-268. https://doi.org/10.1162/002438998553734

Tily, H., Frank, M., & Jaeger, F. (2011). The learnability of constructed languages reflects typological patterns. In *Proceedings of the 33rd Annual Meeting of the Cognitive Science Society*, 1364-1369. https://escholarship.org/uc/item/17q7h67f

Townsend, D. J. (2018). Do children interpret 'marked' comparative adjectives as their opposites? *Journal of Child Language*, *3*, 385-396. https://doi.org/10.1017/S030500090000725X

Waxman, S. & Gelman, R. (1986). Preschoolers' use of superordinate relations in classification and language. *Cognitive Development*, *1*(2), 139-156. https://doi.org/10.1016/S0885-2014(86)80016-8

Wellwood, A., Gagliardi, A., & Lidz, J. (2016). Syntactic and lexical inference in the acquisition of novel superlatives. *Language Learning and Development*, *12*(3), 262-279. https://doi.org/10.1080/15475441.2015.1052878

Westerståhl, Dag. 2019. Generalized quantifiers. In Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*, Metaphysics Research Lab, Stanford University winter 2019 edn. https://plato.stanford.edu/archives/win2019/entries/generalized-quantifiers

Zehr, J., & Schwarz, F. (2018). PennController for Internet Based Experiments (IBEX). https://doi.org/10.17605/OSF.IO/MD832

Zuber, R. (2004). A class of non-conservative determiners in Polish. *Lingvisticae Investigationes*, *27*(1), 147-165. https://doi.org/10.1075/li.27.1.07zub

Zuber, R. & Keenan, E. L. (2019). A Note on Conservativity. *Journal of Semantics*, *36*(4), 573-582. https://doi.org/10.1093/jos/ffz007

Appendix

Table A1: Training and test items used in Experiments 1a, 1c, and 4. The training items used in Experiment 2 were identical, and the test items were conceptually identical but differed in that all shapes were green and either had or lacked yellow stars. The test items Experiment 3 were identical, though the training items differed in that item 7 had one extra blue circle (it was still false in both conditions).

		Conservative #(circles) – 1 = #(blue circles)	**Non-conservative** #(circles) – 1 = #(blue things)
Training Phase			
1	●●●■	True	True
2	●●●■	False	False
3	●●■	True	False
4	●●■	False	True
5	●●●■	True	False
6	●●●■	False	True
7	●●●●■■	False	False
8	●●●●■■	True	True
Test Phase			
9	●●●■■	True	False
10	●●●■■	False	True
11	●●●●■■	True	False
12	●●●●■■	False	True
13	●●●●■	True	False
14	●●●●■	False	True

Table A2: Training and test items used in Experiment 1b (only items 5 and 9 differ from the items in Table A1; this was done to maintain a 50/50 balance between true and false trials).

		Non-conservative #(blue) – 1 = #(blue circles)
Training Phase		
1		False
2		False
3		True
4		True
5		True
6		True
7		False
8		False
Test Phase		
9		True
10		False
11		False
12		False
13		True
14		True

Table A3: Training and test items used in Experiment 5.

		Conservative #(circles) = #(blue circles)	**Non-conservative** #(circles) = #(blue things)
Training Phase			
1		True	True
2		False	False
3		False	True
4		True	False
5		False	True
6		True	False
7		False	False
8		True	True
Test Phase			
9		True	False
10		False	True
11		True	False
12		False	True
13		True	False
14		False	True