Unless: an experimental approach

Prerna Nadathur — Stanford University
Daniel Lassiter — Stanford University

Abstract. The semantics of *unless* has posed a challenge for compositional theories of semantics, and its role as a restrictor of quantifiers is a source of empirical as well as theoretical controversy. We report on an experiment which tests the predictions of the prominent "exceptive" account of *unless*, in particular with respect to how it differs from *if not*. Our results reveal categorical as well as graded patterns of difference between the two conditionals. These patterns falsify previous accounts and motivate a new theoretical picture in which the compositionality issue does not arise, and presupposition and implicature play a central role.

Keywords: conditionals, presupposition, conditional perfection, *unless*, compositionality

1. Introduction

The connective *unless* is often cited as a potential counterexample to semantic compositionality, on the grounds that it contributes a different meaning when embedded under positive quantifiers than it does under negative ones (Higginbotham 1986; Janssen 1997; Szabó 2008). In its most up-to-date form, the problem is that *unless* seems to contribute a biconditional meaning in positive contexts, but only a unidirectional conditional meaning in negative ones (see (1); Leslie 2008). An account of *unless* is consequently of some interest with respect to the status of compositionality: the challenge is to develop a semantics which reflects the perceived strength of *unless* statements (versus *if not*), but also captures the contextual split in (1).

(1) a. Every student will succeed unless he goofs off.
All students who don't goof off succeed, and all students who goof off don't succeed.
b. No student will succeed unless he works hard.
No student who doesn't work hard succeeds (but hard work doesn't guarantee success).

The best previous approach to this problem treats *unless* as an exceptive operator on quantifier domains (von Fintel 1992). In its most current form (due to Leslie 2008), the exceptive account handles the compositionality question by exploiting formal differences between positive and negative quantifiers to build the biconditionality/unidirectionality split directly into the semantic form of *unless*. This account makes a number of as-yet unexplored predictions about interpretive differences between *unless* and *if not*. We report here on an experimental test of these predictions using sentences with *every* and *no*. Our results reveal a number of empirical issues for existing exceptive theories: semantic biconditionality is too strong a requirement with *every*, but unidirectionality is too weak to account for certain contexts in which *if not* is acceptable and *unless* is not.

Our results argue for three main conclusions. First, while the perception of a positive/negative split in (1) reflects an empirical reality, this is not a difference in asserted content, and *unless* is not noncompositional. Second, *unless* encodes a prohibition on use in "Across-the-Board" contexts: *q unless p* is infelicitous when *q* holds unconditionally. We argue that this prohibition is presuppositional. Third, the biconditional interpretation is a *conditional perfection* phenomenon (Geis and Zwicky 1971), which affects both *if not* and *unless* and leads to reduced (but non-zero) acceptability in contexts intermediate between biconditional (*q* iff *not p*) and Across-the-Board (*q*). This leads to a new empirical puzzle: the pragmatic inference to conditional perfection is stronger with *unless* than *if not*, but only under the positive quantifier *every* – no difference arises under *no*. We conclude by describing this puzzling phenomenon and suggesting some directions for future work.

2. *Unless* and biconditionality

2.1. Exceptionality: previous accounts

Classically, *unless* is equated with the negative material conditional *if not* (e.g. Quine 1959). This produces an incorrect interpretation when embedded under negative quantifiers (Higginbotham 1986).[1] Many alternatives have been suggested: for example, Clark and Clark (1977) claim that *q unless p* is *q only if not p*, while Fillenbaum (1986) proposes *p only if q*. These proposals reflect an intuition that *unless* is stronger than *if not*, as Dancygier (1985) argues explicitly: on her account, *q unless p* asserts *q* while acknowledging *p* as an exception to the rule (see also Geis's (1973) comparison of *unless* to *except if*). The core idea is that *q unless p* not only reports that *q* is true when *p* is false, but also draws attention to potential uncertainty regarding *q* when *p* holds.

Von Fintel (1992) implements this analysis by treating *unless* as an exceptive operator on quantifier domains (cf. *except for* in "Everyone except for John left"). On this view, an *unless*-statement is a conjunction of two assertions: one making a (quantified) generalization over a domain from which the *unless*-complement is subtracted, and a second stating that this complement represents the unique smallest set on which the generalization fails. Letting *Q* represent the interpretation of the quantifier, *C* its restriction, *M* its nuclear scope, and *R* the *unless*-complement or excepted set:

(2) **Analysis based on von Fintel 1992:**[2]
$$Q[C]M \textit{ unless } R := Q[C-R]M \wedge \forall S \subseteq C : Q[C-S]M \rightarrow R \subseteq S$$

(3) Every student will succeed unless he goofs off.
$$\text{ALL}\big[\text{STUDENT} - \text{GOOF}\big]\text{SUCC} \wedge \forall S \subseteq \text{STUDENT}: \ \text{ALL}\big[\text{STUDENT} - S\big]\text{SUCC} \rightarrow \text{GOOF} \subseteq S$$

[1] This is a familiar problem with material implication in conditional semantics; see Higginbotham (1986) for details.

[2] Von Fintel's original proposal is stated for those cases where *Q* is a modal quantifier (adverbial or otherwise). He does not provide an explicit formula interpreting *unless*-statements with a nominal quantifier, so there is some uncertainty as to how to adapt the proposal for these cases. Leslie (2008) shows that allowing the quantifier to take wide scope over a (covert) universal modal quantifier results in the same problem as the classical account. Given these considerations, (2) seems to us to be the most plausible extension of von Fintel's proposal for the cases at hand.

This gives us the interpretation in (3) for (1a), where STUDENT represents the students, GOOF the individuals who goof off, and SUCC those who succeed. The first conjunct asserts that all students who do not goof off are successful. The second clause (*uniqueness*) produces the effect of the reverse conditional (all Ms are not Rs), by stipulating that students who goof off are necessarily excluded from any arbitrary set containing only successful students. This entails that no students who goof off succeed. Von Fintel's exceptive account thus gives us biconditionality for (3).

Von Fintel also predicts biconditionality with *no*. For (4) we have: no student who does not work hard will succeed, and no student who works hard is contained in any set of unsuccessful students. Consequently, working hard is both necessary and sufficient for success.

(4) No student will succeed unless he works hard.
$\textsc{no}[\textsc{student} - \textsc{work}]\textsc{succ} \ \wedge\ \forall S \subseteq \textsc{student} :\ \textsc{no}[\textsc{student} - S]\textsc{succ} \rightarrow \textsc{work} \subseteq S$

Leslie (2008) claims that this is too strong, and provides a number of supporting examples. To take a parallel case, suppose we are discussing a university course that is notoriously difficult. We know that students taking this class must work very hard to pass, but in some cases even this may not suffice. In this context, (4) seems to be neither invalid nor infelicitous. This suggests that, in (4), working hard should only be necessary for success, and crucially not sufficient. Leslie captures non-sufficiency under the negative quantifier by modifying the second (uniqueness) clause as in (5). Since *no* is symmetric — *No As are Bs* $\equiv$ *No Bs are As* — this has the desired result of preserving biconditionality under *every* but eliminating it under *no*, as shown in (6).

(5) **Leslie's proposal:**[3]
$Q[C]M \text{ unless } R := Q[C-R]M \wedge \mathbf{Q}[\mathbf{C} \cap \mathbf{M}](-\mathbf{R})$

(6) a. Every student will succeed unless he goofs off.
$= \textsc{all}[\textsc{student} - \textsc{goof}]\textsc{succ} \wedge \textsc{all}[\textsc{student} \cap \textsc{succ}](-\textsc{goof})$
All students who don't goof off succeed, and all students who succeed don't goof off.

b. No student will succeed unless he works hard.
$= \textsc{no}[\textsc{student} - \textsc{work}]\textsc{succ} \wedge \textsc{no}[\textsc{student} \cap \textsc{succ}](-\textsc{work})$
No student who doesn't work hard succeeds, and no student who succeeds doesn't work hard.
Equivalently: *No student who doesn't work hard succeeds.*

[3] Due to space limitations, some of the details of Leslie's account have been glossed over here. Crucially, Leslie holds that *unless* can restrict nominal quantifiers as well as quantificational adverbs, and this is reflected in (6).

2.2. Biconditionality as a pragmatic inference

Although Leslie's modification certainly improves matters for negatively-quantified *unless*-statements, it does not seem to go far enough. Pushing farther on the idea that a biconditional interpretation is not always correct for *unless*, we find naturally-occurring examples such as the following:

(7) Mantou is always late unless she's already out before we meet, but she's often just less late then.

On both exceptive proposals, the pre-comma clause of (7) requires that all relevant situations are (a) such that Mantou is late if she's not already out, and (b) she is not late if she is out. However, the *but*-clause specifies that at least some of the situations where Mantou is out are ones in which she is still late, albeit less so. On these theories, (7) ought to appear as contradictory as (8).

(8) #Roses are always red and violets are always blue, but sometimes violets are not blue.

This difference shows that uniqueness is *defeasible*, and thus is not entailed by *unless*. Crucially, though, if the *but*-clause is suppressed there is a strong tendency to interpret (7) biconditionally. This suggests that uniqueness may be a pragmatic inference.

Nadathur (2014a) provides empirical arguments for viewing uniqueness as a conversational implicature. (9a) shows it can be reinforced without redundancy and (9b) shows that it can be questioned without contradiction. Along with defeasibility, these properties are normally associated with implicatures, and are incompatible with any theory treating uniqueness as an entailment.

(9) a. "Always be yourself, unless you are Fernando Torres. Then, always be someone else."
(*vs:* "Always be yourself, unless you are Fernando Torres. #Otherwise, always be yourself.")
b. "The answer is no unless you ask. If you do ask the answer might be no."
(*vs:* "The answer is no unless you ask. #If you don't ask the answer might be yes.")

Nadathur (2014a) also shows that uniqueness — specifically, the *not* ... *if* direction of biconditionality — can be *backgrounded* without redundancy (ex 10) and does not cause infelicity when it is suspended prior to an *unless*-statement (ex 11). These properties rule out a conventional implicature classification (cf. Potts 2005) and a presuppositional treatment of uniqueness, respectively.

(10) John won't fail if he studies. He will fail unless he studies.
(*vs:* John is a student. John, ?the student, will fail unless he studies.)

(11) The student might not fail if he studies, but he'll fail unless he studies.
(*vs:* ?There might not be a student, but the student will fail unless he studies.)

Descriptively, uniqueness is best classified as a generalized conversational implicature (GCI). As well as being defeasible, reinforceable, and questionable, it seems to arise by *default* when it is not directly blocked, as is characteristic of GCIs á la Levinson 2000). It also bears a striking resemblance to *conditional perfection* (the biconditional interpretation given to certain *if*-conditionals; Geis and Zwicky 1971), which is usually regarded as a GCI (see e.g. van der Auwera 1997).

3. Marbles and dots: an experimental investigation

We are faced with a puzzling collection of intuitions about *unless*. There is some empirical difference between *unless* and *if not*, and it seems to reside in the stronger tendency of the former toward biconditionality. At the same time, the *not if* direction is apparently cancelable with *unless*, and thus not an entailment; but if it is merely an implicature, it is unclear why it should be stronger with *unless* than *if not*. The puzzle is rendered more complex by the need to account for the divergence between *unless* under *every* and under *no*. Having rejected a semantic uniqueness clause in favour of a pragmatic account, we can no longer rely on Leslie's elegant explanation of the asymmetry.

This section reports on an experiment with two main motivations. First, as section 2 demonstrates, any analysis based only on a small set of intuitions (even from corpus-drawn data) is empirically limited: in particular, such intuitions do not distinguish easily between semantic and pragmatic aspects of interpretation. Second, as the examples in (12) show (*Q*=*every*), intuitions about appropriate and inappropriate uses of *unless* are robust only in extreme cases, where either *Q*-many of the individuals in the excepted set do not have the property picked out by the nuclear scope (biconditional), or *Q*-many of them do (Across-the-Board).

(12) [Context: *Half of the students goofed off.*] Every student passed unless he goofed off.
  a. **Clearly appropriate** if all of the students who did not goof off passed, and all of the students who goofed off did not pass. (Biconditional context; uniqueness satisfied)
  b. **Clearly inappropriate** if all of the students passed, including all of those who goofed off. (Across-the-Board context; uniqueness not satisfied)
  c. **Unclear/??** if all of the students who did not goof off passed, and {a few/half/most/...} of the students who goofed off passed. (Intermediate context; uniqueness not satisfied)

The intermediate cases are equally unclear when *Q*=*no*. Collecting naïve judgements on a large (experimental) scale seems to be the only way to clarify these cases. Furthermore, as discussed below, the quantitative details of these cases turn out to be highly informative about the status of exceptionality with *unless* and the nature of the difference between *unless* and *if not*. These details ultimately place significant constraints on the parameters of a revised theoretical account of *unless*.

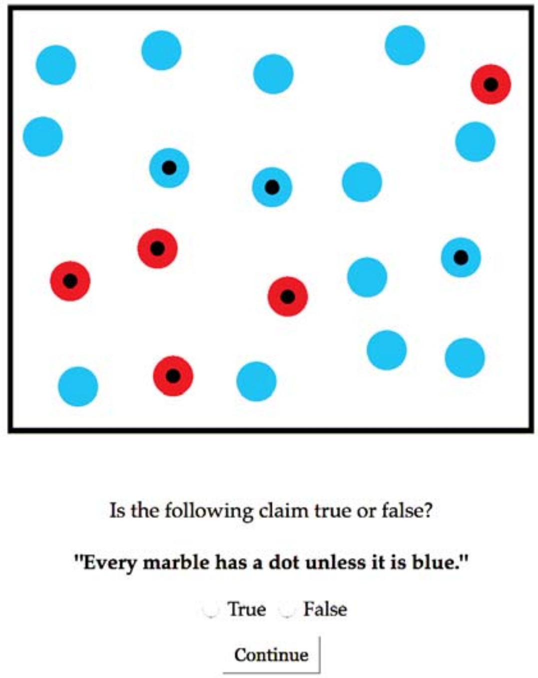

Figure 1: Sample test trial

3.1. Design

Participants were shown a display of 20 red and blue marbles, and were asked to decide whether a given stimulus statement about the display was true or false (forced-choice; Figure (3.1) is a sample trial). Test stimuli (in (13)-(14)) contained either *unless* or *if not*, embedded under either *every* or *no*. We also varied the proportion of target-colour marbles with dots from among 0, 0.2, 0.4, 0.6, 0.8, and 1. This gave us a total of 24 test conditions.

(13) a. Every marble has a dot unless it is [target colour].
b. Every marble has a dot if it is not [target colour].

(14) a. No marble has a dot unless it is [target colour].
b. No marble has a dot if it is not [target colour].

To increase display variety, we randomly varied the target colour between red and blue, and the ratio of target:non-target marbles from 5:15, 10:10, and 15:5 (each test condition thus had 6 display variants). To avoid overwhelming participants with false sentences, we set the proportion of dotted non-target marbles in each test condition to satisfy the minimal truth conditions of both *unless* and *if not*, as given by $Q[C-R]M$ (the first conjunct in (2) and (5)).

We also included a number of fillers. A “sampling” condition asked participants to imagine that a marble was chosen at random, and then judge a sentence of the form “The selected marble has a dot {if it is not/unless it is} [colour].” The display varied as described above.[4] Additional filler statements varied along three parameters: quantifier (*every, no*, none), whether they mentioned “red,” “blue,” or no colour, and construction type (see (15)-(17)). For the last parameter, we used positive *if*-sentences, single-clause quantified statements, and *there*-existentials. In these filler displays, we also varied the red:blue ratio from amongst 5:15, 10:10, and 15:5, and selected both red and blue dot proportions randomly from amongst 0, 0.2, 0.4, 0.6, 0.8, and 1. Consequently, any given filler statement could occur with any one of 108 possible displays.

(15) [bare, red, *if*]: The selected marble has a dot if it is red.

(16) [every, dot, single-clause]: Every marble has a dot.

(17) [no, blue, *there*]: There are dots on no blue marbles.

3.2. Method

Using Amazon’s Mechanical Turk platform, we recruited 160 participants, all of whom were financially compensated. They viewed the experiment in a frame through the Mechanical Turk website. Participants were given detailed instructions together with a sample display and a non-test stimulus, and told to judge whether the stimulus was true or false of the display. They were unable to move from one trial to the next without selecting an answer. At the end of the experiment, participants were asked to enter their native language; this did not affect payment, and no indication was given that it would. The full experiment is available on the second author’s website (http://web.stanford.edu/∼danlass/experiment/marbles/marbles.html).

Each participant saw 48 trials, in a random order. 24 trials were randomly selected from a set containing the 24 test conditions and 12 “sampling” conditions; the displays for these were selected randomly from the 6 variants for the given condition. 24 trials were randomly selected from the remaining 29 filler types, with one of the 108 display variants randomly generated for each filler. On average, each of our 24 test items was seen by 105 participants (min:80, max:124, median:109).

3.3. Results

We excluded data from 5 participants who reported being native speakers of languages other than English. The analysis below includes data from the remaining 155 participants. Figure 2 shows

[4] We included this filler condition out of curiosity about “sampling”-based judgments. The results were statistically indistinguishable from those in the *every* test condition, consistent with theories in which bare conditionals contain a covert *must* (Kratzer 1986). We do not analyze these results further here, however.

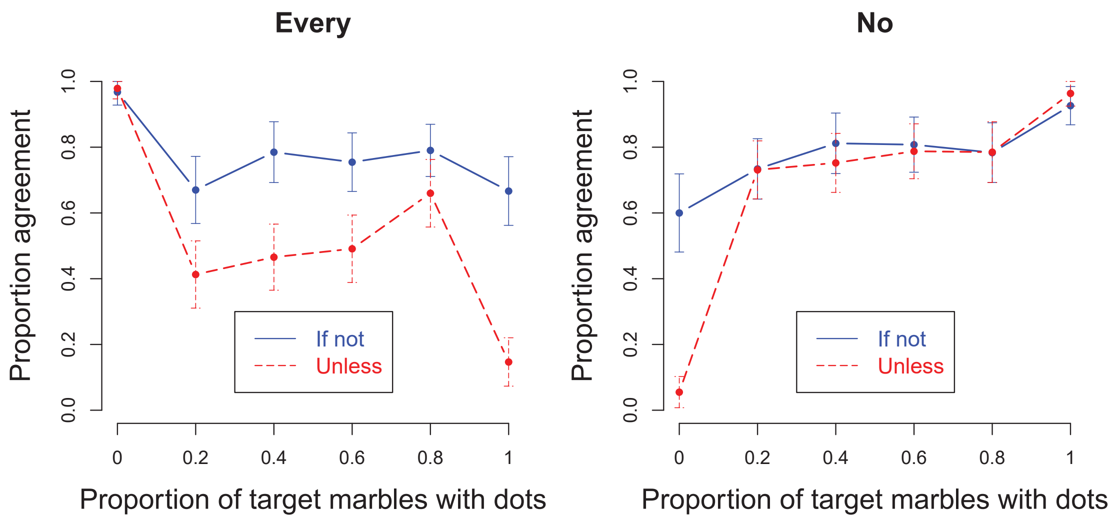

Figure 2: Experimental results, with 97% binomial confidence intervals (see main text).

our results in both quantifier conditions (left=*every*; right=*no*). In each graph, the x-axis is the proportion of target marbles with dots, and the y-axis represents the fraction of participants who judged the sentence true. Results from the *if not* condition are in blue with a solid line interpolated between points. Results from the *unless* condition are in red with a dashed line interpolated. Error bars represent 97% binomial confidence intervals, corresponding to the corrected significance criterion ($\alpha = .03$) selected by Monte Carlo simulation to ensure a Type 1 error rate less than 5% despite multiple comparisons (cf. Edwards and Berry 1987). Non-overlapping error bars at a single proportion thus represent statistically significant differences between *if not* and *unless* in a two-sample t-test, $p < .05$. Table 1 gives the corresponding numerical data.

We analyzed the data using a separate linear mixed-effects models for each quantifier, using the lme4 package (Bates et al. 2014) in R (R Core Team 2014). Proportion was coded as a categorical variable, and we included random effects of participant, target colour, and red/blue distribution (5:15, 10:10, 15:5). We included random intercepts only because the maximal models with random slopes did not converge. For each quantifier, we tested for a main effect of conditional type (*if not* vs. *unless*) following the procedure outlined by Levy (2014). Specifically, we converted the categorical proportion variable to a sum-coded numeric variable, and then calculated the likelihood ratio of two models differing only in the inclusion of a fixed main effect of conditional type. Both models included a fixed main effect of proportion and an interaction between proportion and context, in addition to random intercepts as noted above. These tests revealed a highly significant main effect of conditional type for both quantifiers: $p < .001$, df=1 in both cases.

Table 4 summarizes responses to filler stimuli paired with displays which clearly falsified them. Participants very rarely responded “true” in such scenarios (37 of 1549 judgments; 2.4%).

	if not				***unless***			
	every		*no*		*every*		*no*	
Proportion	N	% agree	N	% agree	N	% agree	N	% agree
0.0	93	96.8 ± 4.0	80	60.0 ± 11.9	95	97.9 ± 3.2	109	5.5 ± 4.7
0.2	100	67.0 ± 10.2	109	73.4 ± 9.2	109	41.3 ± 10.2	119	73.1 ± 8.8
0.4	93	78.5 ± 9.3	85	81.2 ± 9.2	116	46.6 ± 10.0	109	75.2 ± 9.0
0.6	110	75.5 ± 8.9	104	80.8 ± 8.4	112	49.1 ± 10.3	113	78.8 ± 8.4
0.8	124	79.0 ± 7.9	97	78.4 ± 9.1	100	66.0 ± 10.3	93	78.5 ± 9.3
1.0	96	66.7 ± 10.4	95	92.6 ± 5.8	109	14.7 ± 7.4	110	96.4 ± 3.9

Table 1: Endorsement rates in test conditions

every/no marble is [colour]		existential variant	
N	% agree	N	% agree
538	1.8 ± 1.2	456	3.5 ± 1.9

every/no marble has a dot		existential variant	
N	% agree	N	% agree
276	2.2 ± 1.9	269	1.8 ± 1.8

Table 2: Endorsement rates for false filler items

3.4. Discussion of qualitative patterns and implications for previous theories

This section discusses the predictions of von Fintel's and Leslie's theories and compares them to the experimental results. The results are inconsistent with the predictions of both previous theories.

Both formulations of the exceptive account discussed above make unambiguous predictions about the truth-values of the relevant *unless* sentences (examples 18-19). Table 3 shows the predicted distribution of truth-values by theory in each experimental condition.

(18) Every marble has a dot unless it is blue.
von Fintel/Leslie: TRUE iff *all red marbles have dots and no blue marbles have dots.*

(19) No marble has a dot unless it is blue.
von Fintel: TRUE iff *no red marbles have dots and all blue marbles have dots.*
Leslie: TRUE iff *no red marbles have dots.*

	von Fintel						**Leslie**					
	Target dot proportion						Target dot proportion					
	0.0	*0.2*	*0.4*	*0.6*	*0.8*	*1.0*	*0.0*	*0.2*	*0.4*	*0.6*	*0.8*	*1.0*
every	T	F	F	F	F	F	T	F	F	F	F	F
no	F	F	F	F	F	T	T	T	T	T	T	T

Table 3: Predictions for *unless* by condition and theory

In general, we assume that sentences in which *if not* restricts a nominal quantifier have the interpretation in (20). The corresponding predictions are given in (21). These predictions are assumed to hold across all theories considered here.

(20) $Q[C]M$ *if not* $R := Q[C - R]M$

(21) Every/no marble has a dot if it is not blue.
TRUE iff *all/no red marbles have dots.*

Since the experiment was designed so that the truth-conditions for *if not* sentences were satisfied in all test conditions, this account predicts a response of TRUE in all cases (see Table 4). Results from the *if not* conditions are thus expected to provide a baseline for interpreting *unless* results.

	Target dot proportion					
	0.0	*0.2*	*0.4*	*0.6*	*0.8*	*1.0*
every	T	T	T	T	T	T
no	T	T	T	T	T	T

Table 4: Predictions for *if not* by condition

We discuss results for *no* first. Under *no*, participants almost unanimously endorsed stimuli with both *if not* and *unless* at target proportion 1. Recall that the non-target dot proportion in the *no* condition was set at 0: target proportion 1 therefore represents the biconditional context in which no non-target marbles have dots and all target marbles do (satisfying uniqueness). This result is consistent with both von Fintel's and Leslie's accounts.

Endorsement rates dropped off slightly at proportions less than 1 but greater than 0, and appear to have done so identically for *unless* and *if not*. If we allow that the lowered acceptability is due to some pragmatic pressure on both connectives, this result is consistent with Leslie's account, under which *unless* and *if not* are semantically identical under *no* (see Tables 3 and 4). However, it is not consistent with the elaboration of von Fintel's theory, which predicts that the *unless* sentences are always biconditional, and thus false under *no* when the proportion of target marbles with dots is less than 1 (Table 3). Since our participants categorically rejected false items (only .024 acceptance, cf. Table 2), strong endorsement in these conditions falsifies von Fintel's theory.

The remaining data point under *no* is problematic for Leslie's account: *unless* diverges from *if not* sharply at target proportion 0. While a majority of participants (60%) found *No marble has a dot if it is not blue* acceptable in the "Across-the-Board" scenario where no marbles of either colour had dots, endorsement of the corresponding *unless*-sentence approached zero (5.5%) in the same scenario. Indeed, the divergence at 0 was so sharp that it appears to have been entirely responsible for the significant main effect of connective under *no* reported above: endorsement rates for the two connectives were indistinguishable at all other proportions. Confirming this impression, a model comparison procedure identical to the one outlined above, but with proportion 0 removed from the data in all conditions, found no main effect of connective choice under *no* ($p > .6$, df=1). Since Leslie predicts identical truth-values for *unless* and *if not* across all target proportions under *no* (see Tables 3 and 4), this divergence is inconsistent with her theory.

Results under *every* were in some respects the mirror image of those under *no*, but there were some significant differences. Here again we found near-unanimous endorsement in the biconditional context (target proportion 0), just as with *no*. That is, virtually all participants endorsed *Every marble has a dot COND it is blue* for both conditionals when uniqueness/biconditionality was satisfied: all red marbles have dots, and no blue ones do. This result is consistent with both previous semantic accounts. In addition, as under *no*, *if not* and *unless* diverged sharply in the "Across-the-Board" context (target proportion 1; all marbles have dots regardless of colour). A majority of participants (66.7%) accepted *Every marble has a dot if it is not blue* when every marble had a dot regardless of colour, but few (14.7%) accepted *Every marble has a dot unless it is blue* in the same scenario. These data points are consistent with both exceptive accounts, since both predict that *unless* under *every* is true only in biconditional scenarios.

Results in the intermediate range of target proportions (0.2-0.8) were problematic for both previous theories. As in the *no* condition, endorsement rates under *every* were non-maximal at these proportions, but remained much higher than rates for false filler items. However, with *every* there was a reliable difference between *unless* and *if not* also in the intermediate range: *Every marble has a dot if it is not blue* was more likely to be accepted than *Every marble has a dot unless it is blue* when all red marbles have dots and the proportion of blue marbles with dots was .2, .4, .6, or .8. However, it appears that neither is false or otherwise totally unacceptable: endorsement of *every-unless* was reliably above floor in this range (between .41 and .66, all lower CIs above .3), while false filler items were very rarely endorsed (.03 on average). While von Fintel and Leslie are both correct that there is a difference between *every-if not* and *every-unless*, the specific diagnosis of a truth-conditional difference appears to be incorrect: *every-unless* is not simply false in all non-biconditional contexts. In other words, the difference between *unless* and *if not* under *every* appears to involve graded factors affecting felicity, rather than categorical factors involving truth.

In sum, our results argue against a semantically biconditional account of *unless* in either quantifier context. Such an account would incorrectly predict an overwhelming preference for "false" responses in non-biconditional contexts (0-0.8 under *no*, 0.2-1 under *every*). The results also tell against Leslie's one-directional account of *unless* under *no*, given the divergence between *unless* and *if not* in Across-the-Board contexts. A one-directional theory cannot explain our participants' near-unanimous rejection of *unless* in these contexts. Finally, the results suggest that there is a non-categorical difference between the two conditionals in intermediate scenarios under *every*. Both previous theories wrongly predict a categorical, truth-conditional difference here.

4. Three puzzles and two proposals

The experimental investigation reported in section 3 suggests that a new account of *unless* is needed. This account should ideally explain each of the following three puzzles:

(A) The categorical divergence of *if not* and *unless* in Across-the-Board contexts (at target proportion 1 under *every*, and at 0 under *no*).

(B) The degraded but non-zero acceptability of both types of conditionals in the middle range of target proportions (and, for *if not*, in the across-the-board contexts as well).

(C) The reliable but non-categorical divergence between *if not* and *unless* under *every*, and the fact that no such divergence appears under *no*.

In this section we propose solutions to (A) and (B) and sketch some directions for explaining (C).

4.1. Puzzle (A): Categorical divergence in Across-the-Board (AtB) scenarios

When the quantifier (Q) is either *every* or *no*, AtB scenarios for statements of the form $Q[C]M$ COND R can be characterized very simply: they are those in which $Q[C]M$ holds, i.e., the sentence would be true if the conditional clause were omitted. For instance, an AtB scenario for *Every marble has a dot unless it is blue* is one in which every marble has a dot, regardless of colour.

If it were not for Puzzle (A), it would be tempting to suppose that *if not* and *unless* have exactly the same semantic content, and that all differences between them have a pragmatic origin. While there is at least one difference that could in principle be implicated in pragmatic reasoning — *if not* consists of two syntactically separable items, rather than a single word — we do not know of any well-motivated pragmatic mechanisms that would apply only in AtB scenarios and could be expected to create the effect of a categorical divergence. Rather, it seems that *unless* lexically encodes a prohibition against being used in these contexts which is not present in the meaning of *if not*. We will call this requirement the **AtB prohibition**. Two ways of formulating this prohibition are given by the boldfaced portions of (22) and (23) (bracketing for the moment the question of the theoretical status of this conjunct as an entailment, presuppostion, CI, etc.).

(22) $Q[C]M \textit{ unless } R \Rightarrow Q[C-R]M \wedge \neg\mathbf{Q}[\mathbf{C}]\mathbf{M}$ **(Option 1)**

(23) $Q[C]M \textit{ unless } R \Rightarrow Q[C-R]M \wedge \neg\mathbf{Q}[\mathbf{C}\cap\mathbf{R}]\mathbf{M}$ **(Option 2)**

Both (22) and (23) accurately predict the results in Puzzle (A), and in fact Options 1 and 2 are logically equivalent when Q is *every* or *no*. Taking non-universal quantifiers into account, however, Option 2 is clearly preferable: consider example (24) when Q is *some*. As (24a) shows, Option 1 is a contradiction, but Option 2 is satisfiable when some red marbles have dots and none of the blue ones do. The choice of Option 2 is consistent with results from unpublished experimental work: when at least one red marble had a dot, participants rejected (24) if one or more blue marbles had dots, but reliably endorsed it when no blue marbles did.[5]

(24) Some marbles have a dot unless they are blue.

[5] Data and preliminary analysis from the unpublished experiment, which employed similar methods but included quantifiers *most, few*, and *some*, are reported in Nadathur (2014b). The discussion and proposals in this section are in close alignment with the theoretical account of *unless* developed in Nadathur (2014a,b).

a. $\textsc{some}[\textsc{marble} - \textsc{blue}]\textsc{dot} \wedge \neg\mathbf{SOME}[\mathbf{MARBLE}]\mathbf{DOT}$ **(Option 1)**
b. $\textsc{some}[\textsc{marble} - \textsc{blue}]\textsc{dot} \wedge \neg\mathbf{SOME}[\mathbf{MARBLE} \cap \mathbf{BLUE}]\mathbf{DOT}$ **(Option 2)**

The AtB prohibition is much weaker than biconditionality. When Q is *every* or *no*, its only effect is to rule out AtB scenarios: intermediate proportions are predicted to be acceptable.

(25) Every marble has a dot unless it is blue.
$\textsc{all}[\textsc{marble} - \textsc{blue}]\textsc{dot} \wedge \neg\textsc{all}[\textsc{marble} \cap \textsc{blue}]\textsc{dot}$
Every red marble has a dot, and some blue marble does not.

(26) No marble has a dot unless it is blue.
$\textsc{no}[\textsc{marble} - \textsc{blue}]\textsc{dot} \wedge \neg\textsc{no}[\textsc{marble} \cap \textsc{blue}]\textsc{dot}$
No red marble has a dot, and some blue marble does.

Having concluded that the AtB prohibition is lexically associated with *unless*, we must decide whether it is an entailment, a presupposition, or something else. The empirical endorsement rate of *unless* in AtB contexts was comparable to the endorsement rate of false fillers, but since our experiment did not include control conditions with false presuppositions or conventional implicatures (CIs), we cannot be sure of the source of this pattern. However, independent arguments suggest that the bold clause in (23) is neither an entailment nor a CI: it can be reinforced without redundancy, which argues against an entailment treatment (example 27), and it can be backgrounded, which argues against a CI treatment (ex. 28, cf. Potts 2005).

(27) Every marble has a dot unless it is blue, and some blue marbles do not have dots.
(*vs*: #Every marble has a dot unless it is blue, and every red marble has a dot.)

(28) No blue marbles have dots. However, every marble has a dot unless it is blue.
(*vs*: No blue marbles have dots. #The blue marbles, none of which have dots, are my favourites.)

Moreover, efforts to suspend the AtB prohibition explicitly as in (29a) seem to create infelicity, unless they are framed as corrections as in (29b). Like the possibility of backgrounding in (28), this is reminiscent of the behaviour of presuppositions (compare examples 29a-29b with 30a-30b).

(29) a. #Every blue marble has a dot, and every marble has a dot unless it is blue.
b. Every marble has a dot unless it is blue. In fact, every blue marble has a dot, too.

(30) a. #It's not raining, and Mary doesn't realize that it is raining.
b. Mary doesn't realize that it's raining. In fact, it isn't raining.

While we remain somewhat uncertain about this diagnosis — due, in part, to the difficulty of applying standard projection tests to quantified *unless* sentences — we suggest that these data favour a presuppositional treatment of the AtB prohibition. *Modulo* the presupposition, *unless* is semantically equivalent to *if not* on this proposal:

(31) $Q[C]M$ *unless* $R \begin{cases} \text{is a presupposition failure if } Q[C \cap R]M\text{; otherwise,} \\ \text{is true if and only if } Q[C - R]M. \end{cases}$

Regardless of the precise status of the AtB prohibition, however, it seems evident that this condition must be part of any account of *unless*: it provides a solution to Puzzle (A), and uniquely picks out those points at which *unless* (but not *if not*) empirically receives near-zero agreement ratings.

4.2. Puzzle (B): Biconditionality inferences

Puzzle (B) involves the less-than-unanimous but clearly nonzero acceptance rates for both *unless* and *if not* across the middle range of target dot proportions, and in AtB scenarios for *if not*. We suggest that this pattern can be given a natural pragmatic explanation in terms of *conditional perfection* (Geis and Zwicky 1971). A conditional sentence like (32a) is in many contexts strengthened to a biconditional — that is, to the conjunction of (32a) and (32b). The inference to biconditionality is typically regarded as a GCI (van der Auwera 1997; Horn 2000), where the default inference adds the content of (32b) to the truth-conditional content provided in (32a).

(32) a. The marble Bill selected has a dot if it is not blue.
b. $\rightsquigarrow$ The marble Bill selected does not have a dot if it is blue.
c. (32a) & (32b) $\equiv$ The marble Bill selected has a dot if and only if it is blue.

We suggest that quantified conditionals with *every* and *no* can also be pragmatically perfected.

(33) a. Every marble has a dot if it is not blue.
b. $\rightsquigarrow$ Every marble does not have a dot if it is blue.
c. (33a) & (33b) $\equiv$ All and only non-blue marbles have dots.

(34) a. No marble has a dot if it is not blue.
b. $\rightsquigarrow$ No marble does not have a dot if it is blue.
c. (34a) & (34b) $\equiv$ All and only blue marbles have dots.

Previous work has revealed a tendency for experimental participants in a truth-value judgement task to reject true sentences which are associated with false implicatures. For instance, Doran et al. (2012) investigated truth-value judgments involving true sentences that were associated with false GCIs. Participants in their baseline condition received instructions most similar to those given to our participants; rejection rates in this condition ranged between 15% and 63% over a large variety of GCI triggers. In terms of our experiment, then, the effect of a default pragmatic inference to biconditionality should be to render *if not* sentences systematically less acceptable in the intermediate target proportion range, as well as in AtB scenarios.

Specifically, suppose that our participants were inclined, with some small probability p, to reject sentences associated with false conditional perfection inferences. *Modulo* experimental noise, we would then expect to find an endorsement rate of $1 - p$ in the *every* condition for those stimuli where (33a) is true but (33b) is false. Since the truth-conditions of (33a) were always satisfied in this condition (i.e., all red marbles had dots), we expect high but non-maximal endorsement in the *every*-*if not* conditions at any target dot proportion greater than 0 (i.e. for any item where some blue marbles have dots). For the *no* condition, we expect an endorsement rate of roughly $1 - p$ when (34a) is true and (34b) is false: since all stimuli in this condition verified (34a) (no red marbles have dots), this occurs whenever the target dot proportion is strictly below 1 (whenever there is a blue marble without a dot). As far as *if not*-conditionals are concerned, these predictions describe Puzzle (B) precisely: a small but robust tendency for participants to reject *if not* statements in the non-biconditional scenarios.

It is a relatively small step to extend the explanation to Puzzle (B) as it pertains to *unless*. We proposed in (31) above that the semantic meaning of *unless* is identical to that of *if not* when the presupposed AtB prohibition is satisfied. It is thus reasonable to expect that *unless* will be associated with a conditional perfection inference in the same way that *if not* is. This is also supported by the arguments in section 2.2 for treating uniqueness/biconditionality as a pragmatic inference. This derives the observed pattern just as described above for *if not*: outside of AtB scenarios, *unless* is dispreferred under *no* in non-biconditional contexts to a degree which is almost perfectly matched with the dispreference for *if not* under *no*. The AtB scenario is the only clear point of deviation, and this is explained by the additional effect of presupposition failure for *unless*.

Unless is also dispreferred under *every* in non-biconditional contexts, excluding again the AtB scenario (where presupposition failure leads to reduced acceptability with *unless*). This effect is consistent with the account from conditional perfection; however, our account so far does not predict the higher rejection rate for *unless* under *every*. This is the subject of Puzzle (C).

4.3. Puzzle (C): Grades of biconditionality?

Puzzle (C) is the most perplexing. It involves the reliable but non-categorical divergence between *if not* and *unless* over the target dot proportions 0.2-0.8 — and, in particular, the fact that this divergence is only observed under *every*, while the connectives are indistinguishable in the intermediate range under *no*. In light of the conclusions above, we see two strategies for explaining Puzzle (C).

First, we might attribute the divergence to a second pragmatic pressure which (i) is triggered only in sentences with *every*, and (ii) interacts additively with the downward pressure exerted by the falsity of the conditional perfection inference in intermediate cases. While this possibility is intriguing, we do not at this time have any speculations about what the nature of such a pressure might be.

A second possibility is that the biconditionality implicature might be somehow "stronger" for *unless* in positive contexts than it is in negative ones. Interestingly, this diagnosis is not too different from the intuition, reported by Leslie (2008) and discussed above, that *unless* feels "more" biconditional under *every* than under *no*. While we have argued that it is a misdiagnosis to treat this intuition as reflecting a truth-conditional difference, a pragmatic approach which (somehow) allows for different strengths of conditional perfection could honour Leslie's empirical insight.

For the purposes of the current paper, we remain noncommittal about the source of this divergence.[6] We hope that we have nevertheless made important steps toward the resolution of Puzzle (C) in this paper by (i) uncovering it empirically, (ii) showing that it cannot be treated as a straightforward truth-conditional difference, (iii) framing it with respect to the AtB prohibition and conditional perfection, and (iv) suggesting some routes for addressing it in future work. For future work, point (ii) is especially important: given our experimental results, a semantic account is not viable because participants do not respond to intermediate *every-unless* stimuli as they do to false control stimuli. Consequently, a pragmatic approach to Puzzle (C) seems unavoidable.

5. Conclusions

Based on the results of our experimental investigation, we have argued for three main points and posed a new empirical question for future work. First, neither von Fintel (1992) nor Leslie (2008) provide the correct semantic account for *unless*: *unless* is not semantically biconditional under either *every* or *no*. Second, *unless* and *if not* diverge categorically in acceptability under both quantifiers. To explain this pattern, we suggested that the only lexically encoded difference between *if not*- and *unless*-sentences is that the latter presupposes an "Across-the-Board prohibition" which proscribes the use of an *unless*-conditional *q unless p* in contexts where the unqualified statement *q* could be used truthfully. Third, we argued that biconditionality generated by conditional perfection implicatures plays a role in the interpretation of both types of conditionals, producing degraded acceptability under both quantifiers in non-biconditional scenarios. Finally, we demonstrated that *if not* and *unless* sentences diverge in intermediate scenarios under *every*, but pattern identically under *no*. While we did not resolve this puzzle fully, we identified several strategies for future investigation, and argued that the most profitable direction for its explication resides in the pragmatics of conditionals rather than the semantics of *unless*.

Our primary aim was to investigate the exceptive account of *unless* experimentally. In addition to showing that neither von Fintel's nor Leslie's version makes the right predictions, our data lead

[6] However, Nadathur 2014b provides a detailed pragmatic proposal involving the interaction of downward-entailing contexts and scalar reasoning.

naturally to an alternative set of theoretical proposals, as outlined above. These proposals converge to a large extent with the alternative theory of *unless* offered in Nadathur (2014a). That proposal comprises two main points: first, a biconditional interpretation is pragmatically (not semantically) associated with *unless*, and second, *unless* (and not *if not*) invokes a presupposition which would produce the effect of the AtB prohibition in the empirical cases examined here. Since the claims in this paper were made on the basis of experimental data, we see this convergence as lending support to Nadathur (2014a) – or any alternative proposal in which non-truth-conditional mechanisms are used to accurately predict the patterns described in Puzzles (A)-(C).

These conclusions highlight the importance of utilizing large-scale judgement studies when investigating complex issues in semantics and pragmatics. The difficulty in giving a theoretical account of *unless* has in no small part been driven by the complex and graded cluster of acceptability intuitions associated with it. This work would not have been possible if we had relied exclusively on introspective judgments. However, by collecting a large number of judgments in controlled conditions and analyzing them quantitatively, we were able to distinguish intuitions that are categorical and truth-conditional in nature from those that are graded and contextual/pragmatic in nature (see also Wasow and Arnold 2005; Gibson and Fedorenko 2010, 2013; Scontras and Gibson 2011).

More broadly, our investigation supports the conclusion that the puzzling interaction between *unless* and various quantifiers is explained primarily by pragmatics and presuppositional content, rather than the truth-conditional content. We thus concur strongly with Leslie (2008)'s conclusion, albeit for different reasons: *unless*, while intriguing, poses no threat to compositionality.

References

van der Auwera, J. (1997). Pragmatics in the last quarter century: the case of conditional perfection. *Journal of Pragmatics 27*, 261–274.

Bates, D., M. Maechler, B. Bolker, and S. Walker (2014). *lme4: Linear mixed-effects models using Eigen and S4*. R package version 1.1-7.

Clark, H. and E. Clark (1977). *Psychology and language: an introduction to psycholinguistics*. London: Thomson Learning.

Dancygier, B. (1985). If, unless, and their Polish equivalents. *Papers and Studies in Contrastive Linguistics 20*, 64–72.

Doran, R., G. Ward, M. Larson, Y. McNabb, and R. E. Baker (2012). A novel experimental paradigm for distinguishing between what is said and what is implicated. *Language 88*, 124–154.

Edwards, D. and J. J. Berry (1987). The efficiency of simulation-based multiple comparisons. *Biometrics*, 913–928.

Fillenbaum, S. (1986). The use of conditionals in inducements and deterrents. In E. Traugott, A. ter Meulen, J. Reilly, and C. Ferguson (Eds.), *On Conditionals*, pp. 179–195. Cambridge: Cambridge University Press.

von Fintel, K. (1992). Exceptive conditionals: the meaning of "unless". In *Proceedings of the North East Linguistics Society*, Volume 22, pp. 135–151.

Geis, M. (1973). "If" and "unless". In B. B. Kachru, R. B. Lees, Y. Malkiel, A. Pietrangeli, and S. Saporta (Eds.), *Issues in linguistics: papers in honor of Henry and Renee Kahane*, pp. 231–253. Urbana, IL: University of Illinois Press.

Geis, M. and A. Zwicky (1971). On invited inferences. *Linguistic Inquiry 2*, 561–566.

Gibson, E. and E. Fedorenko (2010). Weak quantitative standards in linguistics research. *Trends in cognitive sciences 14*, 233–234.

Gibson, E. and E. Fedorenko (2013). The need for quantitative methods in syntax and semantics research. *Language and Cognitive Processes 28*, 88–124.

Higginbotham, J. (1986). Linguistic theory and Davidson's program in semantics. In E. Lepore (Ed.), *Truth and interpretation: perspectives on the philosophy of Donald Davidson*, pp. 29–48. Oxford: Blackwell.

Horn, L. (2000). From if to iff: conditional perfection as pragmatic strengthening. *Journal of Pragmatics 32*, 289–326.

Janssen, T. (1997). Compositionality. In J. van Benthem and A. ter Meulen (Eds.), *Handbook of logic and linguistics*, pp. 417–475. Amsterdam: Elsevier.

Kratzer, A. (1986). Conditionals. In A. M. Farley, P. Farley, and K. E. McCollough (Eds.), *Proceedings of Chicago Linguistic Society*, Volume 22, pp. 1–15.

Leslie, S.-J. (2008). "If," "unless," and quantification. In R. Stainton and C. Viger (Eds.), *Compositionality, context, and semantic value: essays in honor of Ernie Lepore*, pp. 3–30. Amsterdam: Springer.

Levinson, S. (2000). *Presumptive meanings: the theory of generalized conversational implicature*. Cambridge MA: MIT Press.

Levy, R. (2014). Using R formulae to test for main effects in the presence of higher-order interactions. *ArXiv e-prints*.

Nadathur, P. (2014a). *Unless*, exceptionality, and conditional strengthening. In R. de Haan (Ed.), *Proceedings of the ESSLLI 2014 Student Session*, pp. 109–120.

Nadathur, P. (2014b). *Unless*, exceptionality, and the pragmatics of conditional statements. Ms, Stanford University, http://web.stanford.edu/∼pnadath/pdfs/Nadathur_QP1.pdf.

Potts, C. (2005). *The logic of conventional implicatures*. Oxford: Oxford University Press.

Quine, W. (1959). *Methods of logic*. New York: Holt, Rinehart & Winston.

R Core Team (2014). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Scontras, G. and E. Gibson (2011). A quantitative investigation of the imperative-and-declarative construction in English. *Language 87*, 817–829.

Szabó, Z. (2008). Compositionality. In *The Stanford Encyclopedia of Philosophy*. Winter 2008 edition.

Wasow, T. and J. Arnold (2005). Intuitions in linguistic argumentation. *Lingua 115*, 1481–1496.