Plurals under quantification: new experimental perspectives

NihiL seminar, ILLC.

Claire Rong
clairerongfr@gmail.com

Supervised by Benjamin Spector

Department of Cognitive Studies, École Normale Supérieure, Paris

January 16, 2026

Outline

Theories on the interpretation of plurals

Previous experimental studies and gradient effects

Comprehension study in English: bare plurals

Follow-ups: *several NPs*; cumulativity

Comprehension study in English: *Several NPs*

Comprehension study in English: cumulativity

Comprehension study in English: *Some NPs* (continuous)

Comprehension study in English: *Some NPs* (binary)

Comprehension study in Mandarin: *xie*

Outline

Theories on the interpretation of plurals

Previous experimental studies and gradient effects

Comprehension study in English: bare plurals

Follow-ups: *several NPs*; cumulativity

Comprehension study in English: *Several NPs*

Comprehension study in English: cumulativity

Comprehension study in English: *Some NPs* (continuous)

Comprehension study in English: *Some NPs* (binary)

Comprehension study in Mandarin: *xie*

A logical gap

Bare plurals in English most often give rise to a **multiplicity inference**:

(1) The box contains books.

In some environments (e.g. negative sentence), the meaning of bare plurals is **not** the negation of their meaning in simple affirmative sentences:

(2) The box doesn't contain books.

Logical gap: situations where the box contains **exactly one book**.
Is **exactly one book** included in the denotation of "books"?

How do we account for the logical gap?

Accounting for the logical gap: (main) existing theories

Two families of theoretical approaches:

- **Bivalent approaches.**
  Bare plurals have an *at least one* denotation
  → pragmatically strengthened to *at least two*.
  - Higher-Order implicature (Spector 2007)
  - Zweig(+Ivlieva)'s approach (Zweig 2009, Ivlieva 2020)

- **Trivalent approaches.**
  Bare plurals have truth conditions (*at least two*), falsity conditions (*zero*) and are undefined for *exactly one*.
  - Presuppositional Exhaustification approach (Ahn et al. 2020)
  - Homogeneity-based approach (Križ 2017)

Accounting for the logical gap: (main) existing theories

Overview of bivalent approaches.

- Bare plurals have **weak semantics.**
  Literal meaning: *at least one*
- Multiplicity inference arises via **scalar implicature:** $\langle \mathrm{PL}, \mathrm{SG} \rangle$.
- Two main variants considered:
  - Higher-Order Implicatures (HOI, Spector 2007)
  - Dependent plural–based implicatures (Zweig 2008; Zweig 2009; Ivlieva 2020)
- Predict **optional** strengthening, sensitive to alternatives.

Accounting for the logical gap: (main) existing theories

Overview of trivalent approaches.

- Based on the trivalent semantics of the Strong Kleene framework.
- Two main variants considered:
  - based on homogeneity (Križ 2017)
  - based on Presuppositional Exhaustification (Ahn et al. 2020; Bassi et al. 2021)
- For our data, both make the **same empirical predictions.**

Bare plurals under universal quantification

What readings do these theories predict?

"Each box contains books."

1. **Literal** reading: each box contains at least one book.
2. **Weak** reading: each box contains at least one book and it is not the case that each box contains exactly one book.
3. **Strong** reading: each box contains at least two books.

Logical strengths: strong > weak > literal

Consequence: we **cannot** test **any combination** of readings.

Higher-order implicature	{literal, weak, strong}
Zweig(+Ivlieva)'s approach	{literal, strong}
Presuppositional Exhaustification	{literal, strong}
Homogeneity-based approach	{literal, strong}

Bare plurals under universal quantification

What readings do these theories predict?

"Each box contains books."

1. **Literal** reading: each box contains at least one book.
2. **Weak** reading: each box contains at least one book and it is not the case that each box contains exactly one book.
3. **Strong** reading: each box contains at least two books.

Logical strengths: strong > weak > literal

uniformly-singular

Bare plurals under universal quantification

What readings do these theories predict?

"Each box contains books."

1. **Literal** reading: each box contains at least one book.
2. **Weak** reading: each box contains at least one book and it is not the case that each box contains exactly one book.
3. **Strong** reading: each box contains at least two books.

Logical strengths: strong $>$ $\underbrace{\textbf{weak} > \textbf{literal}}_{\textbf{mixed}}$

Bare plurals under universal quantification

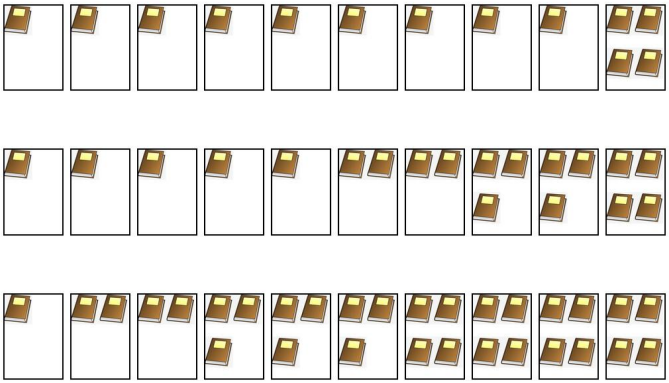


Figure: Examples of mixed scenarios for "each box contains books".

Bare plurals under universal quantification

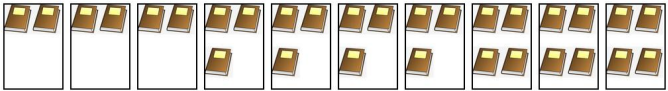
What readings do these theories predict?

"Each box contains books."

1. **Literal** reading: each box contains at least one book.
2. **Weak** reading: each box contains at least one book and it is not the case that each box contains exactly one book.
3. **Strong** reading: each box contains at least two books.

Logical strengths: strong > weak > literal

uniformly-plural



Goals

Theoretical question 1

What are the available readings?

Goals

Theoretical question 1

What are the available readings?

Theoretical question 2

How universal are the mechanisms of plural interpretation?
More specifically, as a case study, what are the available readings in Mandarin, a language with optional number marking?

Outline

Theories on the interpretation of plurals

Previous experimental studies and gradient effects

Comprehension study in English: bare plurals

Follow-ups: *several NPs*; cumulativity

Comprehension study in English: *Several NPs*

Comprehension study in English: cumulativity

Comprehension study in English: *Some NPs* (continuous)

Comprehension study in English: *Some NPs* (binary)

Comprehension study in Mandarin: *xie*

Previous experimental work

Two experimental studies looked at plurals under universal quantification:

- Stateva et al. 2016
- Jiang and Sudo 2023

Collected **graded semantic judgments:**

- Likert scales
- scenarios: uniformly-singular, mixed, uniformly-plural

Previous experimental work

Two experimental studies looked at plurals under universal quantification:

- Stateva et al. 2016 ← **also investigated readings**
- Jiang and Sudo 2023

Collected **graded semantic judgments:**

- Likert scales
- scenarios: uniformly-singular, mixed, uniformly-plural

How to detect readings

Recall the three readings: literal, weak, strong.

- **uniformly-singular scenarios:** only literal reading true
- **mixed scenarios:** literal + weak readings true
- **uniformly-plural scenarios:** literal + weak + strong readings true

Prediction about ratings: **uniformly-plural** > **mixed** > **uniformly-singular**.

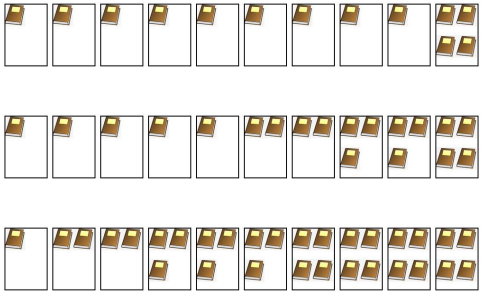
All theories predict: **uniformly-plural** > **uniformly-singular.**

Underlying idea: *the more readings are satisfied, the higher the ratings.* (see e.g. Chemla and Spector 2011)

A crucial confound: gradient effects

uniformly-plural > **mixed** > **uniformly-singular**

⚠ **Spurious inference:**
intermediate ratings for mixed scenarios ⇒ evidence for a weak reading



Gradient effects: intuition

"Each box contains books."

Highest ratings: uniformly-plural scenarios.
Lowest ratings: uniformly-singular scenarios.

Mixed scenarios intuitively not all equally good:

- intermediate ratings can arise without a weak reading,
- ratings may increase with 'similarity to uniformly-plural scenarios'.

Gradient effects: when different instantiations of the same reading might systematically receive different truth-value ratings.

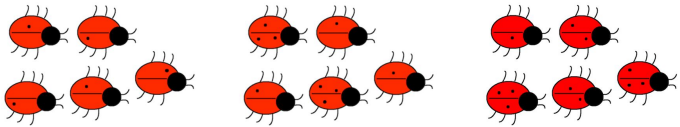
Stateva et al. 2016: design

(In Italian) **"Every ladybug has dots."** / **"Every ladybug has some dots."**

Same experimental logic as ours.

Factor: number of single-dotted ladybugs out of 5.

$$\underbrace{5}_{\text{uniformly-singular}}, \underbrace{1,3}_{\text{mixed}}, \underbrace{0}_{\text{uniformly-plural}}$$

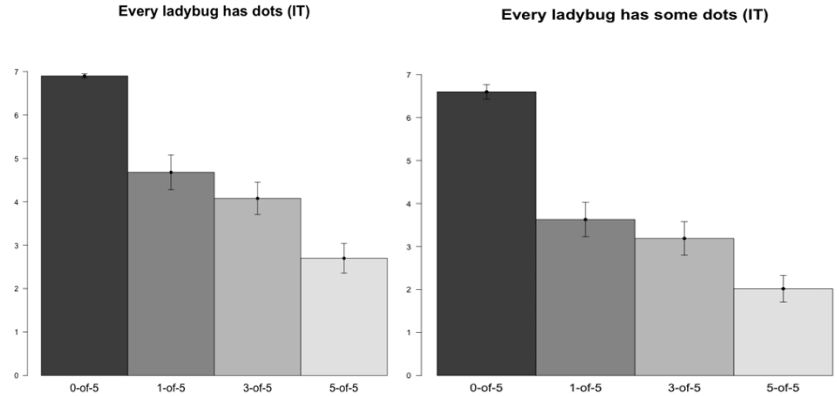


Stateva et al. 2016: original interpretation

Significant differences found:

- between uniformly-singular and mixed
- between mixed and uniformly-plural.

Original interpretation: evidence for both weak and strong readings.



Stateva et al. 2016: alternative interpretation

Suppose only two readings exist: literal + strong.

Mixed scenarios:

- do not satisfy the strong reading
- but vary in similarity to uniformly-plural scenarios.

Intermediate ratings might not reflect a distinct weak reading, but only **gradient effects** from the literal to the strong reading.

Original interpretation misses a key comparison:

- difference between uniformly-singular and closest mixed
- vs. difference within mixed conditions.

Nothing can be concluded from 'one contrast is significant, the other not'. (Gelman and Stern 2006, "The difference between 'significant' and 'not significant' is not itself statistically significant".)

Confound from previous studies

Previous studies:

- did not explicitly control for gradience
- even though hints were already present

Two competing accounts:

- literal + strong readings → intermediate ratings due to gradience
- literal + weak + strong readings → possibly with additional gradience

→ Our experiments control for **gradient effects**, a confounding factor overlooked in previous work.

Methodological question

Experimentally, how can we disentangle readings from gradient effects?

Outline

Theories on the interpretation of plurals

Previous experimental studies and gradient effects

Comprehension study in English: bare plurals

Follow-ups: *several NPs*; cumulativity

Comprehension study in English: *Several NPs*

Comprehension study in English: cumulativity

Comprehension study in English: *Some NPs* (continuous)

Comprehension study in English: *Some NPs* (binary)

Comprehension study in Mandarin: *xie*

Bare plurals (continuous judgments): design

We conducted several experiments of language comprehension, asking for judgments (on a continuous scale) of a sentence against a picture.
Structure of the sentences from all experiments:

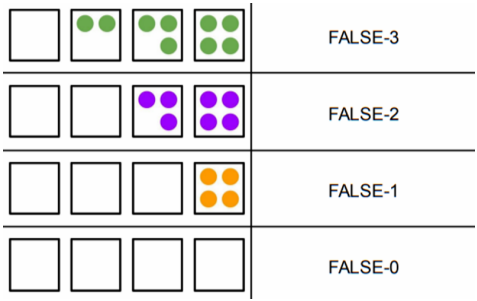
Each box contains [plural expression].

A different plural expression in each experiment:

- bare plurals
- *several NPs*
- *some NPs*

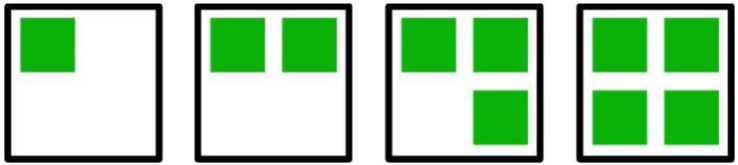
Bare plurals (continuous judgments): design

A box containing several shapes is called a **strong verifier**.



Bare plurals (continuous judgments): design

Example of a trial:



Bare plurals (continuous judgments): results

200 participants (after exclusions) recruited through Prolific.
Each participant saw each condition 3 times.

Bare plurals (continuous judgments): results

Visually:

- Qualitative shifts from **FALSE** to **LITERAL**, and from **WEAK** to **STRONG**.
- Quantitative shifts within **FALSE** and within **LITERAL**+**WEAK**.

Bare plurals (continuous judgments): results

Four predictors:

- c_{vrf} (number of strong verifiers)
- c_{lit} (binary variable indicating whether the condition is literally true)
- c_{weak} (indicating whether the condition supports a weak reading)
- c_{str} (indicating whether the condition supports a strong reading)

On literally true conditions: we fit a linear mixed-effects model predicting responses as a function of c_{vrf}, with random intercepts and slopes by participant.

$$\text{response} \sim c_{\text{vrf}} + (1 + c_{\text{vrf}} \mid \text{participant})$$

Result (as expected): a positive slope in the linear model and a significant LRT p-value (comparison with a null model). $\chi^2(1) = 395.0$, $p < 10^{-15}$.
LRT on **WEAK** conditions alone: $\chi^2(1) = 18.8$, $p < 10^{-4}$.
$\rightarrow$ **Gradience is indeed present within the same reading.**

Bare plurals (continuous judgments): results

Four predictors:

- c_{vrf} (number of strong verifiers)
- c_{lit} (binary variable indicating whether the condition is literally true)
- c_{weak} (indicating whether the condition supports a weak reading)
- c_{str} (indicating whether the condition supports a strong reading)

Full model:

$$\text{response} \sim c_{\text{vrf}} + c_{\text{lit}} + c_{\text{weak}} + c_{\text{str}} + (1 \mid \text{participant})$$

Which of these factors actually matter for explaining the data?
$\rightarrow$ **Model comparisons** using the Bayesian information criterion (BIC), comparing all $2^4 = 16$ possible sub-models.

Bare plurals (continuous judgments): results

Best-fitting model across all 9 conditions:

$$\text{response} \sim c_{\text{vrf}} + c_{\text{lit}} + c_{\text{str}} + (1 \mid \text{participant})$$

$\rightarrow$ Seems to favor approaches that **do not predict a weak reading.**

Second best-fitting model ($\Delta BIC = 9$):

$$\text{response} \sim c_{\text{vrf}} + c_{\text{lit}} + c_{\text{weak}} + c_{\text{str}} + (1 \mid \text{participant})$$

Bare plurals (continuous judgments): results

In both BIC and AIC rankings, this model

$$\text{response} \sim c_{\text{lit}} + c_{\text{weak}} + c_{\text{str}} + (1 \mid \text{participant})$$

is judged far better than this model

$$\text{response} \sim c_{\text{lit}} + c_{\text{str}} + (1 \mid \text{participant})$$

Compare these 2 models through LRT, the p-value is significant.

Bare plurals (continuous judgments): results

In both BIC and AIC rankings, this model

$$\text{response} \sim c_{\text{lit}} + c_{\text{weak}} + c_{\text{str}} + (1 \mid \text{participant})$$

is judged far better than this model

$$\text{response} \sim c_{\text{lit}} + c_{\text{str}} + (1 \mid \text{participant})$$

Compare these 2 models through LRT, the p-value is significant.

→ **Spurious inference: "the weak reading exists".**
→ Precisely because **we didn't control for gradience.**

Bare plurals: interim summary of findings

Systematic gradience within readings.
→ **Gradience must be an independent factor.**

From **LITERAL-0** to **WEAK-1**:

- only a quantitative increase
- driven by c_{vrf}.

From **WEAK-3** to **STRONG-4**:

- a qualitative increase
- driven by c_{str}.

Model comparisons: weak reading does not improve the model.
→ Favors theories that do not predict a weak reading.

Outline

1 Theories on the interpretation of plurals

2 Previous experimental studies and gradient effects

3 Comprehension study in English: bare plurals

4 Follow-ups: *several NPs*; cumulativity

- Comprehension study in English: *Several NPs*
- Comprehension study in English: cumulativity

5 Comprehension study in English: *Some NPs* (continuous)

6 Comprehension study in English: *Some NPs* (binary)

7 Comprehension study in Mandarin: *xie*

8 Conclusion

Follow-up experiments needed

Two observations from the bare plurals experiment:

- **Gradience might reflect proximity to a uniformly-plural scenario.**
  Prediction: gradience should persist even with no ambiguity.
  → **1st follow-up, *several NPs*, an unambiguous plural.**

- **Unexpectedly high ratings for uniformly-singular scenarios.**
  Possible explanation: marginal cumulative interpretation.
  → **2nd follow-up, cumulativity experiment with different plural expressions.**

Outline

Theories on the interpretation of plurals

Previous experimental studies and gradient effects

Comprehension study in English: bare plurals

Follow-ups: *several NPs*; cumulativity

Comprehension study in English: *Several NPs*

Comprehension study in English: cumulativity

Comprehension study in English: *Some NPs* (continuous)

Comprehension study in English: *Some NPs* (binary)

Comprehension study in Mandarin: *xie*

Several NPs (continuous judgments): design

"Each box contains [several NPs]."

Same basic design as bare plurals.

Only one predictor of interest: c_{vrf} (number of strong verifiers).

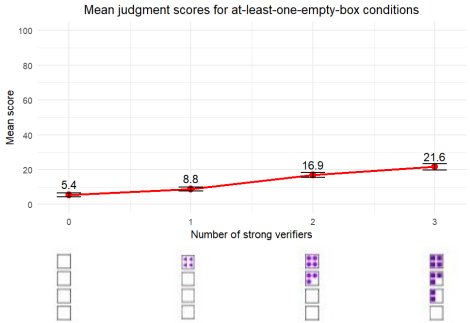
Gradient effects expected:

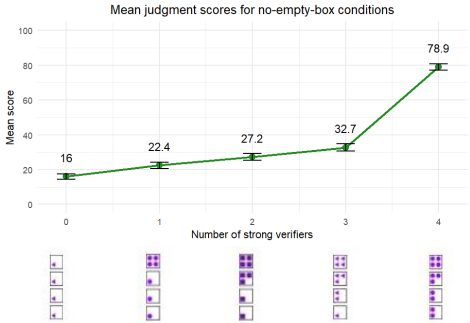
- within no-empty-box conditions
- within at-least-one-empty-box conditions

Several NPs (continuous judgments): results

Visually:

- Gradience persists even with a single reading.
- Mean ratings do not follow the visual ordering of conditions.

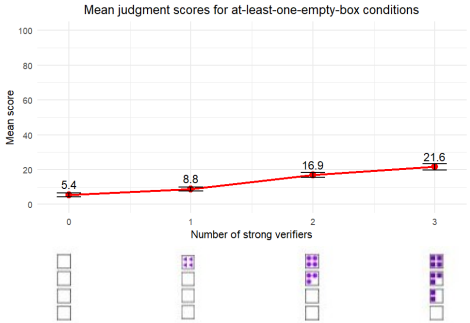


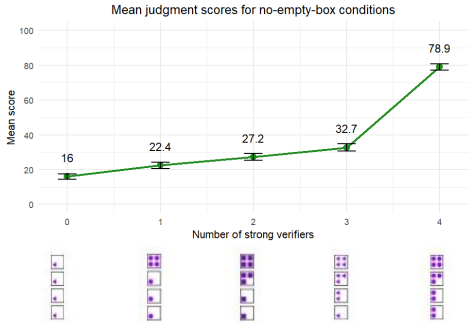


Several NPs (continuous judgments): results

LRT on no-empty-box conditions: $\chi^2(1) = 232.31$, $p < 10^{-49}$.
LRT on at-least-one-empty-box conditions: $\chi^2(1) = 380.01$ $p < 10^{-81}$.

→ In both cases, c_{vrf} significantly improves model fit.
Supports gradience as an independent factor.





Outline

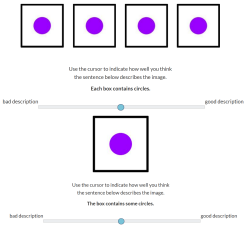
Cumulativity experiment (continuous judgments): design

Marginal cumulative interpretation: sentence judged acceptable if there is a plurality in total.

Does the magnitude of this effect depend on the plural expression?

Factors:

- Plural expression:
  - bare plurals
  - *some NPs*
  - *several NPs*
- Picture type:
  - 1 box (sentences like "The box contains circles");
  - 4 boxes (sentences like "Each box contains circles").



600 participants total, each completing exactly one trial.

Cumulativity (continuous judgments): results

Across all sentence types: **higher ratings with 4 boxes than with 1 box.**

	Bare plural	***Some NPs***	***Several NPs***
1-box	32.8	18.9	3.4
4-boxes	75.8	46.8	14.5

Table: Mean ratings (scale 1-100) by sentence type and image type.

Size of the effect varies:

- largest for bare plurals
- intermediate for *some NPs*
- smallest for *several NPs*

Cumulativity (continuous judgments): results

Open question: existing theories do not predict differences in cumulative availability, especially between bare plurals and *some NPs*.

But otherwise, regarding multiplicity inference, *some NPs* have predicted behavior identical to bare plurals, in all theories.

Motivation for experiment with *some NPs*:
Lower 'baseline' ratings in uniformly-singular scenarios.
→ More room for increase within literally true scenarios.

Design: identical to bare plurals experiments.

Outline

Theories on the interpretation of plurals

Previous experimental studies and gradient effects

Comprehension study in English: bare plurals

Follow-ups: *several NPs*; cumulativity

Comprehension study in English: *Several NPs*

Comprehension study in English: cumulativity

Comprehension study in English: *Some NPs* (continuous)

Comprehension study in English: *Some NPs* (binary)

Comprehension study in Mandarin: *xie*

Some NPs (continuous judgments): results

200 participants (after exclusions) recruited through Prolific.
Each participant saw each condition 3 times.

Some NPs (continuous judgments): results

On literally true conditions:

$$\text{response} \sim c_{\text{vrf}} + (1 + c_{\text{vrf}} \mid \text{participant})$$

LRT: $\chi^2(1) = 1052.9$, $p < 10^{-15}$.
LRT on **WEAK** conditions alone: $\chi^2(1) = 65.19$, $p < 10^{-15}$.

$\rightarrow$ **Gradience is again present within the same reading.**

Some NPs (continuous judgments): results

Best-fitting model across all 9 conditions:

$$\text{response} \sim c_{\text{vrf}} + c_{\text{lit}} + c_{\text{str}} + (1 \mid \text{participant})$$

$\rightarrow$ Again, seems to favor approaches that **do not predict a weak reading.**

Second best-fitting model ($\Delta BIC = 9$):

$$\text{response} \sim c_{\text{vrf}} + c_{\text{lit}} + c_{\text{weak}} + c_{\text{str}} + (1 \mid \text{participant})$$

Outline

Theories on the interpretation of plurals

Previous experimental studies and gradient effects

Comprehension study in English: bare plurals

Follow-ups: *several NPs*; cumulativity

Comprehension study in English: *Several NPs*

Comprehension study in English: cumulativity

Comprehension study in English: *Some NPs* (continuous)

Comprehension study in English: *Some NPs* (binary)

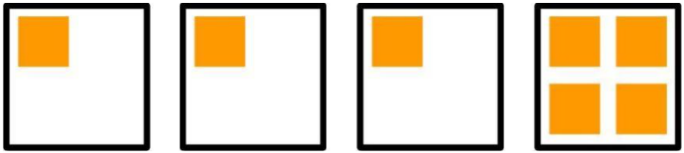
Comprehension study in Mandarin: *xie*

Some NPs (binary judgments): design

Goal: are gradient effects still present with binary judgments?

Same sentences and pictures as in the version with continuous judgments.

Example of a trial:



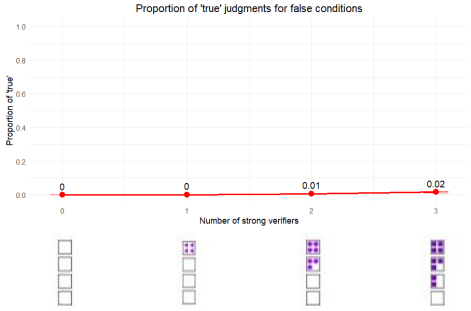
Do you think the sentence below is true or false?

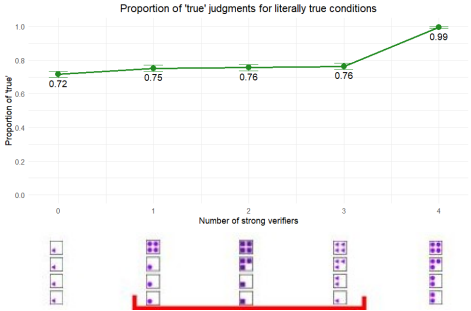
Each box contains some squares.

☐ false ☐ true

Some NPs (binary judgments): results

200 participants (after exclusions) recruited through Prolific.
Each participant saw each condition 3 times.

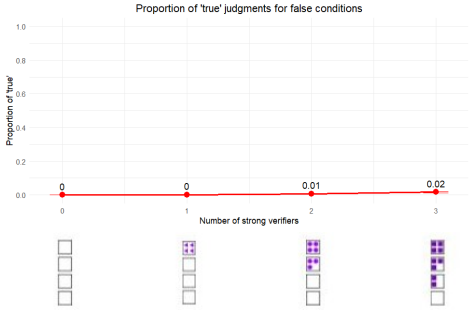


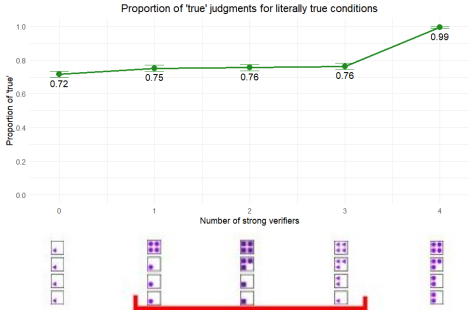


Some NPs (binary judgments): results

Visually:

- Hardly any gradience within **FALSE** cases, or within **WEAK+LITERAL** cases.
- The only noticeable qualitative shifts:
  from **FALSE** to **LITERAL** and from **WEAK** to **STRONG**.





Some NPs (binary judgments): results

Same four predictors:

- c_{vrf} (number of strong verifiers)
- c_{lit} (binary variable indicating whether the condition is literally true)
- c_{weak} (indicating whether the condition supports a weak reading)
- c_{str} (indicating whether the condition supports a strong reading)

Best-fitting model across all 9 conditions:

$$\text{response} \sim c_{\text{vrf}} + c_{\text{lit}} + c_{\text{str}} + (1 \mid \text{participant})$$

$\rightarrow$ Seems to favor approaches that **do not predict a weak reading.**

BUT this could well be due to **limitations of the logistic model**: exaggerated effects of tiny (insignificant) gradience within **FALSE**.

Some NPs (binary judgments): results

Follow-up analysis: (post-hoc)
model comparison without c_{lit} on the subset of literally true conditions.

Best-fitting model across the subset of 5 conditions:

$$\text{response} \sim c_{\text{weak}} + c_{\text{str}} + (1 \mid \text{participant})$$

This is the **only time** in all our English experiments that c_{weak} **was present in the best-fitting model.**

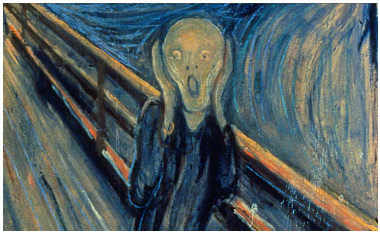
Some NPs (binary judgments): results

Follow-up analysis: (post-hoc)
model comparison without c_{lit} on the subset of literally true conditions.

Best-fitting model across the subset of 5 conditions:

$$\text{response} \sim c_{\text{weak}} + c_{\text{str}} + (1 \mid \text{participant})$$

This is the **only time** in all our English experiments that c_{weak} **was present in the best-fitting model.**



Some NPs (binary judgments): results

Follow-up analysis: (post-hoc)
model comparison without c_{lit} on the subset of literally true conditions.

Best-fitting model across the subset of 5 conditions:

$$\text{response} \sim c_{\text{weak}} + c_{\text{str}} + (1 \mid \text{participant})$$

This is the **only time** in all our English experiments that c_{weak} **was present in the best-fitting model.**

... still not a strong argument in favor of the weak reading:

- visually not convincing
- analysis not preregistered on the subset of 5 conditions
- properties of the logistic model
- if a continuous response type 'concealed' the weak reading, it is conversely possible that a binary response type is 'concealing' cognitive gradience...

Outline

Theories on the interpretation of plurals

Previous experimental studies and gradient effects

Comprehension study in English: bare plurals

Follow-ups: *several NPs*; cumulativity

Comprehension study in English: *Several NPs*

Comprehension study in English: cumulativity

Comprehension study in English: *Some NPs* (continuous)

Comprehension study in English: *Some NPs* (binary)

Comprehension study in Mandarin: *xie*

A language with optional number marking

A threefold number marking in Mandarin:

1. Bare noun: **number-neutral** (Zhang 2014; Cheng and Sybesma 1999...).

2. [one + CL] where CL is the '**singular**' classifier ≈ *a NP*
   In Mandarin: *yige* **一个**、*yiben* **一本**⋯⋯
   Triggers a **uniqueness** inference.

3. [one + *xie*] where *xie* **些** is the '**plural**' classifier ≈ *(some) NPs*
   In Mandarin: *yixie* **一些**.
   Triggers a **multiplicity** inference.

It can be shown that all three forms have the **same truth conditions**.

Comprehension study in Mandarin

Picture stimuli identical to those of the English comprehension experiment, with translated instructions and the following stimulus sentence:

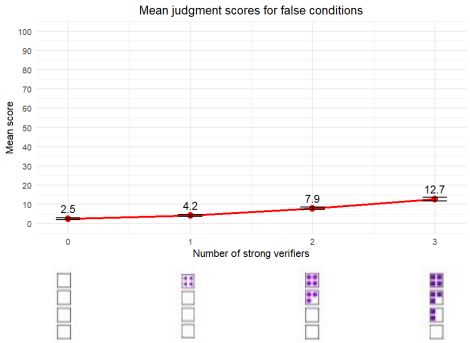
(3) 每 个 盒子 里 都 有 **一** **些** *[NP]*
měi gè hé-zi lǐ dōu yǒu yī xiē
each CL box in DOU EXIST **one** **xie** [NP]

'Each box contains [one + *xie* + NP].'

xie experiment: results

Visually:

- Qualitative shifts from **FALSE** to **LITERAL**, from **WEAK** to **STRONG**, but also (it seems!) from **LITERAL** to **WEAK**.
- Quantitative shifts within **FALSE** and within **LITERAL**+**WEAK**.



xie experiment: results

Same four predictors:

- c_{vrf} (number of strong verifiers)
- c_{lit} (binary variable indicating whether the condition is literally true)
- c_{weak} (indicating whether the condition supports a weak reading)
- c_{str} (indicating whether the condition supports a strong reading)

On literally true conditions, we fitted a linear mixed-effects model:

$$\text{response} \sim c_{\text{vrf}} + (1 + c_{\text{vrf}} \mid \text{participant})$$

Result (as expected): a positive slope in the linear model and a significant LRT p-value (comparison with a null model). $\chi^2(1) = 858.13$, $p < 10^{-15}$.

LRT on **WEAK** conditions alone, with the same conclusions:
$\chi^2(1) = 38.34$, $p < 10^{-9}$.

xie experiment: results

Best-fitting model across all 9 conditions:

$$\text{response} \sim c_{\text{vrf}} + c_{\text{lit}} + c_{\text{weak}} + c_{\text{str}} + (1 \mid \text{participant})$$

Second best-fitting model ($\Delta BIC = 62$):

$$\text{response} \sim c_{\text{vrf}} + c_{\text{lit}} + c_{\text{str}} + (1 \mid \text{participant})$$

$\rightarrow$ Contrary to continuous *some NPs*,
c_{weak} is present in the best model.

$\rightarrow$ Contrary to binary *some NPs*,
presence of c_{weak} is at least visually clear.

Possibly because baseline (= uniformly-singular) judgments are lower?
(+ Caveat: differences in participants socio-demographic background.)

Outline

Theories on the interpretation of plurals

Previous experimental studies and gradient effects

Comprehension study in English: bare plurals

Follow-ups: *several NPs*; cumulativity

Comprehension study in English: *Several NPs*

Comprehension study in English: cumulativity

Comprehension study in English: *Some NPs* (continuous)

Comprehension study in English: *Some NPs* (binary)

Comprehension study in Mandarin: *xie*

Answers to our questions

Theoretical question 1

What are the available readings?

- In every experiment, the **literal** and **strong** readings are present in the best model.
- At first glance, **with continuous judgments**, weak reading not supported in English comprehension (no improvement with c_{weak}).
- Best model: response $\sim c_{\text{vrf}} + c_{\text{lit}} + c_{\text{str}} + (1 \mid \text{participant})$
- Weak reading detected in binary *some NPs* task $\rightarrow$ methodological challenge.

Higher-order implicature	{literal, weak, strong}
Zweig(+Ivlieva)'s approach	{literal, strong}
Presuppositional Exhaustification	{literal, strong}
Homogeneity-based approach	{literal, strong}

Answers to our questions

Theoretical question 2

How universal are the mechanisms of plural interpretation?
More specifically, as a case study, what are the available readings in Mandarin, a language with optional number marking?

- Mandarin shows gradient effects *and* supports the weak reading (c_{weak} improves model fit).
- Further theoretical work needed on link between optional number marking systems and possible availability of weak reading.

Answers to our questions

Methodological question

Experimentally, how can we disentangle readings from gradient effects?

- Gradience modeled via strong verifier count (c_{vrf}) + binary factors for readings (c_{lit}, c_{weak}, c_{str}).
- Alternative: weights on readings + for each reading, distance to closest situation that satisfies the reading (e.g., Chemla and Spector 2014).

Remaining puzzles

- **Cumulative readings:** acceptability varies across plural expressions.
- **Binary vs. continuous responses:** modeling binary responses as a function of continuous responses.
- **Production vs. comprehension:** underexplored link.

Future directions

- **Mandarin:** investigate plural interpretation when number marking is optional.
- Investigate **other lexical scales** with scalar implicatures (revisit Chemla and Spector 2011).
- Extend to **other quantifiers** and modals.
- Explore **context-sensitivity** by controlling for the QUD.
- Refine ambiguity resolution using **probabilistic models** (e.g. RSA).

Thank you!

References I

Ahn, Dorothy, Ankana Saha, and Uli Sauerland (2020). "Positively polar plurals: Theory and predictions". In: Semantics and linguistic theory, pp. 450–463.

Bassi, Itai, Guillermo Del Pinal, and Uli Sauerland (2021). "Presuppositional exhaustification". Unpublished manuscript.

Chemla, Emmanuel and Benjamin Spector (2011). "Experimental evidence for embedded scalar implicatures". In: Journal of semantics 28.3, pp. 359–400.

— (2014). "Distinguishing typicality and ambiguities, the case of scalar implicatures". Unpublished manuscript.

Cheng, Lisa Lai-Shen and Rint Sybesma (1999). "Bare and not-so-bare nouns and the structure of NP". In: Linguistic inquiry 30.4, pp. 509–542.

References II

Gelman, Andrew and Hal Stern (2006). "The Difference Between "Significant" and "Not Significant" is not Itself Statistically Significant". In: The American Statistician 60.4, pp. 328–331. DOI: 10.1198/000313006X152649. eprint: https://doi.org/10.1198/000313006X152649. URL: https://doi.org/10.1198/000313006X152649.

Ivlieva, Natalia (2020). "Dependent plurality and the theory of scalar implicatures: Remarks on Zweig 2009". In: Journal of Semantics 37.3, pp. 425–454.

Jiang, Yizhen and Yasutada Sudo (2023). "Putting bare plurals into context". In: Hnm2—gaps and imprecision in natural language semantics: homogeneity effects

Križ, Manuel (2017). "Bare plurals, multiplicity, and homogeneity". Unpublished manuscript.

References III

Spector, Benjamin (2007). "Aspects of the pragmatics of plural morphology: On higher-order implicatures". In: Presupposition and implicature in compositional semantics. Springer, pp. 243–281.

Stateva, Penka, Sara Andreetta, and Arthur Stepanov (2016). "On the nature of the plurality inference: Ladybugs for Anne". In: Papers dedicated to Anne Reboul. Lyon: CNRS.

Zhang, Niina Ning (2014). "Expressing number productively in Mandarin Chinese". In: Linguistics.

Zweig, Eytan (2008). "Dependent plurals and plural meaning". PhD thesis. New York University.

— (2009). "Number-neutral bare plurals and the multiplicity implicature". In: Linguistics and Philosophy 32.