Syntax and the world agree on mass/count distinctions

Charlotte Moore & Elika Bergelson
Duke University

Introduction

Syntactically:
- Many languages distinguish mass vs. count nouns[1]:
  - Mass nouns generally can't be pluralized[2, 5]
  - 2-year-olds generalize mass/count nouns based on syntactic cues[3]

Perceptually:
- Discrete objects and substances have different properties[2]
- Mass nouns denote objects without a characteristic shape[1]
- 8-month-olds expect objects to behave differently from substances[3]
- Perceptual categories are available to infants well before syntactic ones

Research questions:

1. What kinds of syntactic frames do infants hear with mass nouns?
2. What is the relationship between mass noun syntax and perceptual appearance?

Methods

- Used the **28** nouns from MCDI[4] that can take mass syntax[2]
- Analyzed videos from the SEEDLingS corpus (528 hours of footage):
  - Yearlong longitudinal corpus from 44 families
  - 1-hour videos collected monthly from 6-17 months
- Tagged all mass noun tokens where mentioned noun was present (N=2494)
- Annotated whether noun looked like substance or object
  - Reliability: Cohen's $\varkappa$ = 0.67, 85% agreement
- Annotated determiner phrase of each noun as count, mass, or ambiguous
  - See "Syntax Types" below

Included mass words (frequency)

water (531)	cheese (150)	juice (71)	sky (40)	pizza (18)	glass (9)	medicine (5)
hair (272)	chicken (125)	carrot (67)	coffee (39)	toast (17)	butter (8)	snow (4)
fish (270)	bread (85)	orange (66)	soap (28)	meat (15)	trash (8)	candy (1)
milk (267)	paper (82)	cake (44)	cereal (22)	money (12)	spaghetti (6)	television (1)

Syntax types

Mass only syntax:
- Do you want some milk?
- *Do you want some cat?

Count only syntax[5]:
- Do you want another cat?
- #Do you want another milk?

Ambiguous syntax:
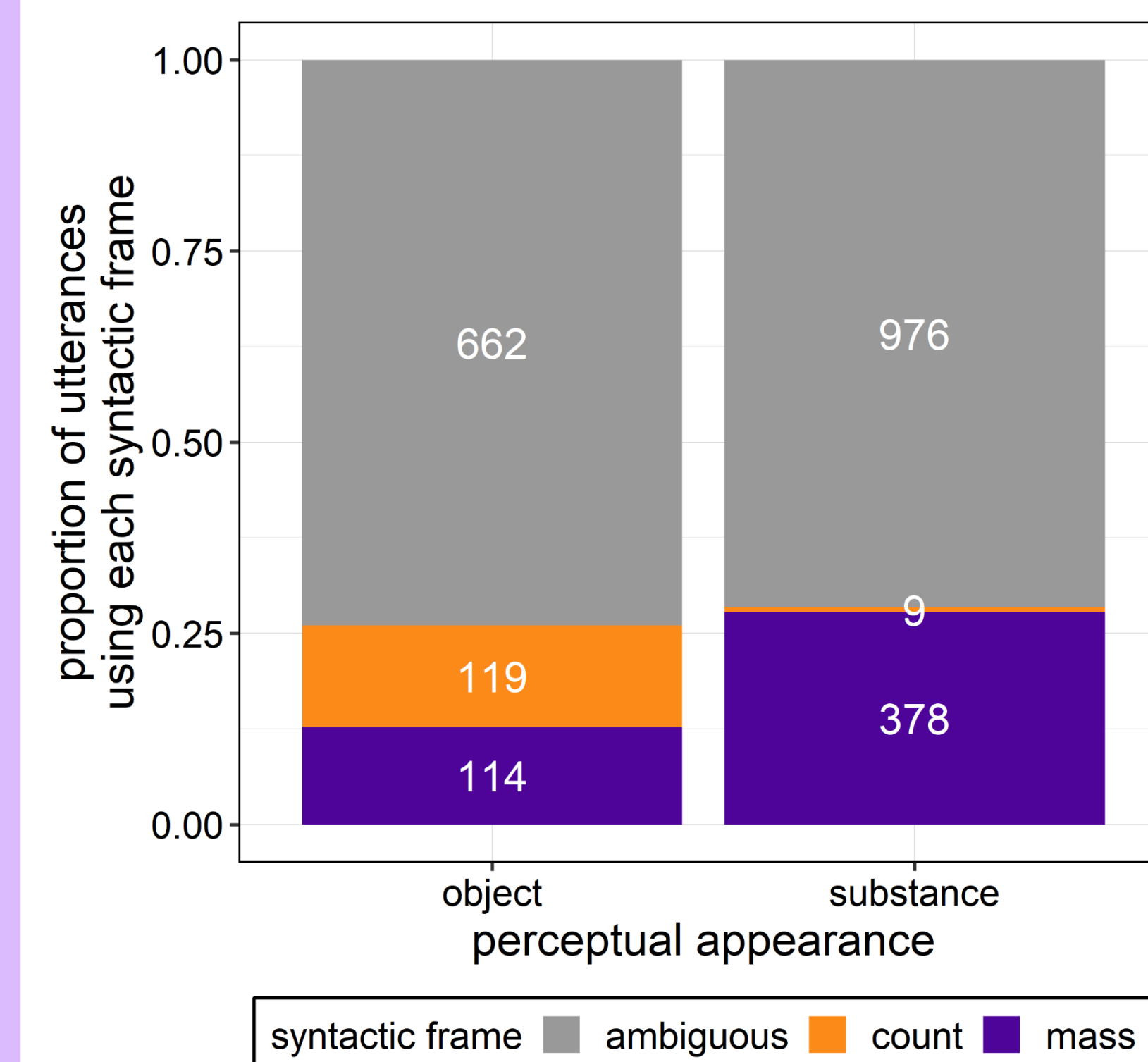- Do you want the milk?
- Do you want the cat?

Top Frames

- some, more, [unit] of, a lot of, any

- a, an, another, two, the other

- [none], your, the, my, her

Most mass syntax was ambiguous


- Most syntactic frames were ambiguous.
- Syntax was *more* reliably informative when mass nouns looked like substances.
- When words looked like substances **98%** informative syntax indicated a mass noun.
- When words looked like discrete objects, **51%** of informative syntax indicates a count noun.

Perception - Object vs. Substance
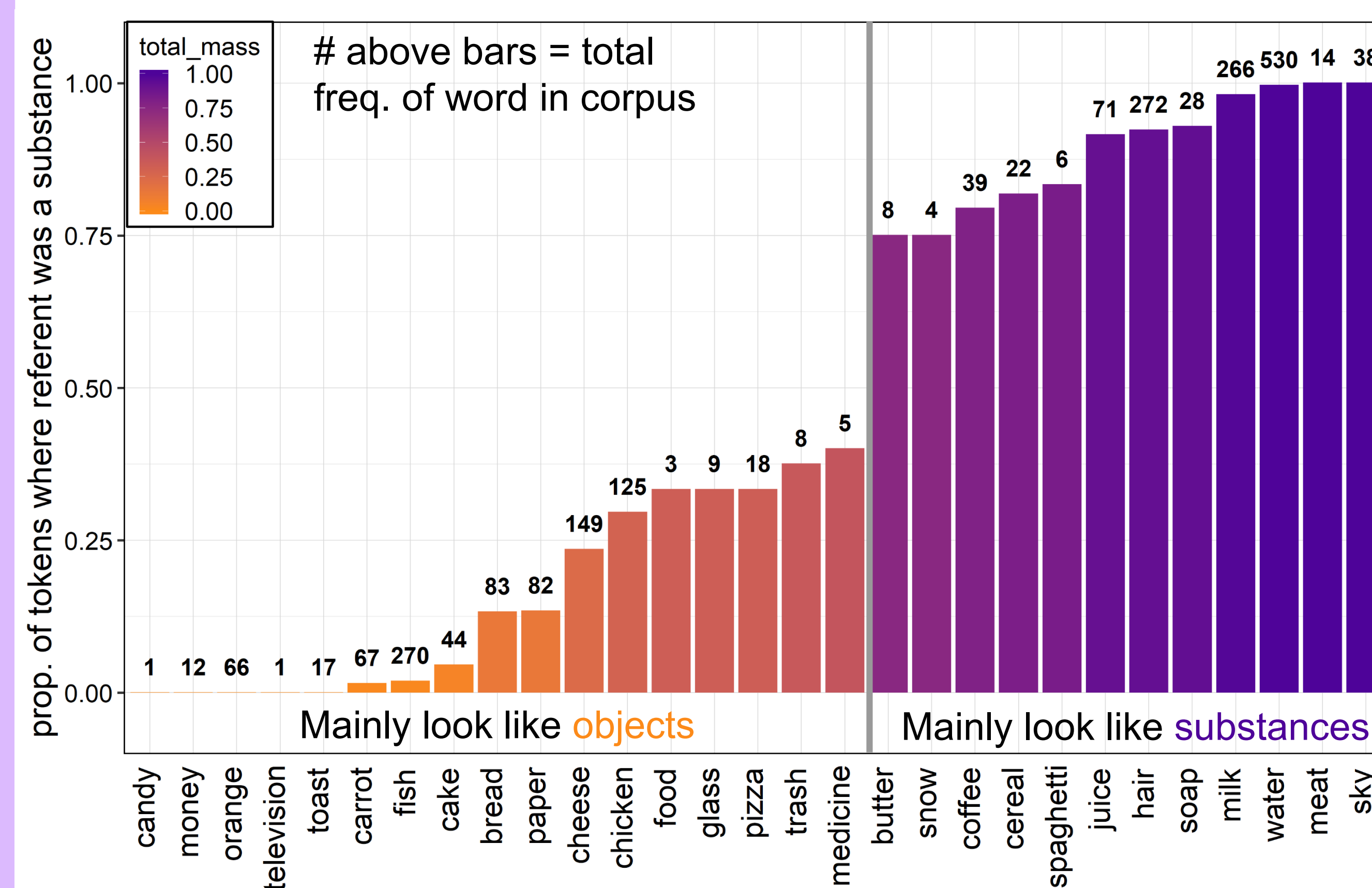
- Real world falls into separable categories:
  - all included words take mass syntax, but only 12 of them are substances most of the time.
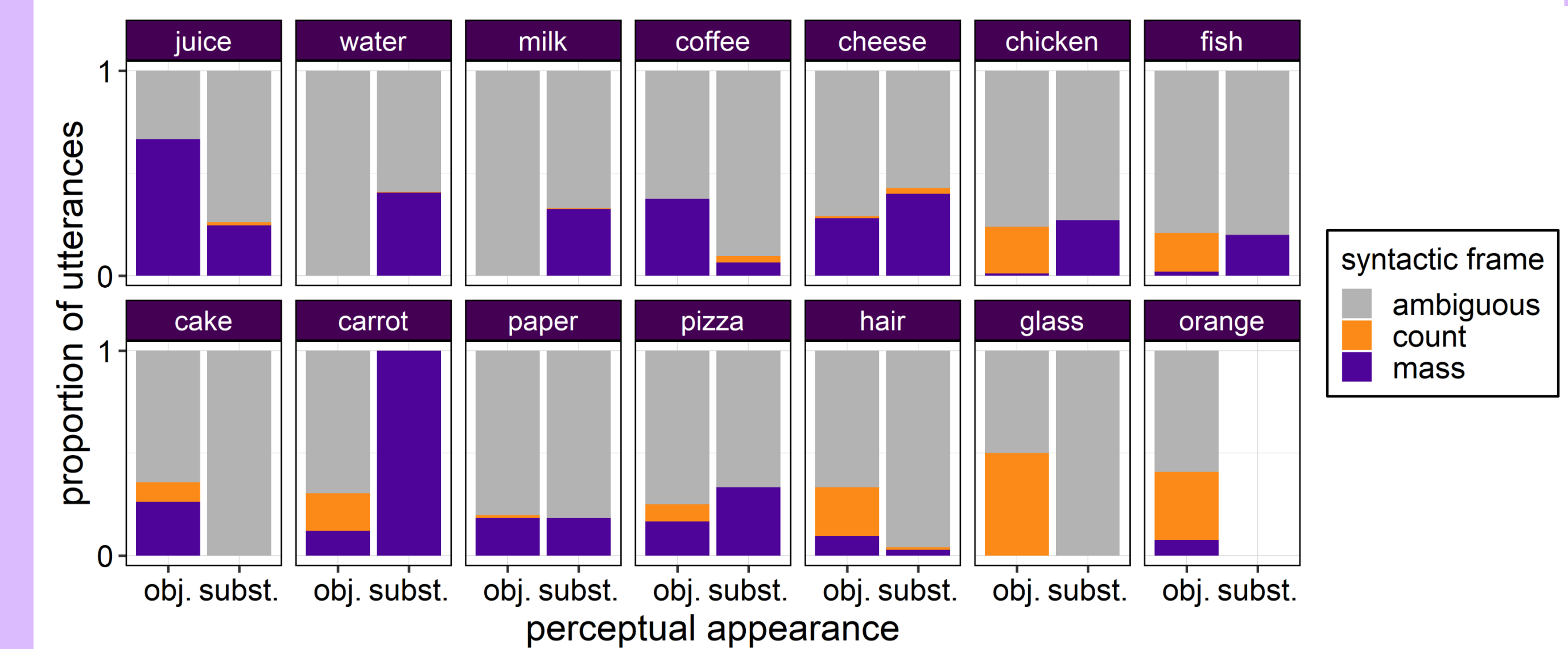- Token freq. **not** correlated with appearance (Spearman's ρ=0.17, p=0.38)

Patterns stable from 6-17mo
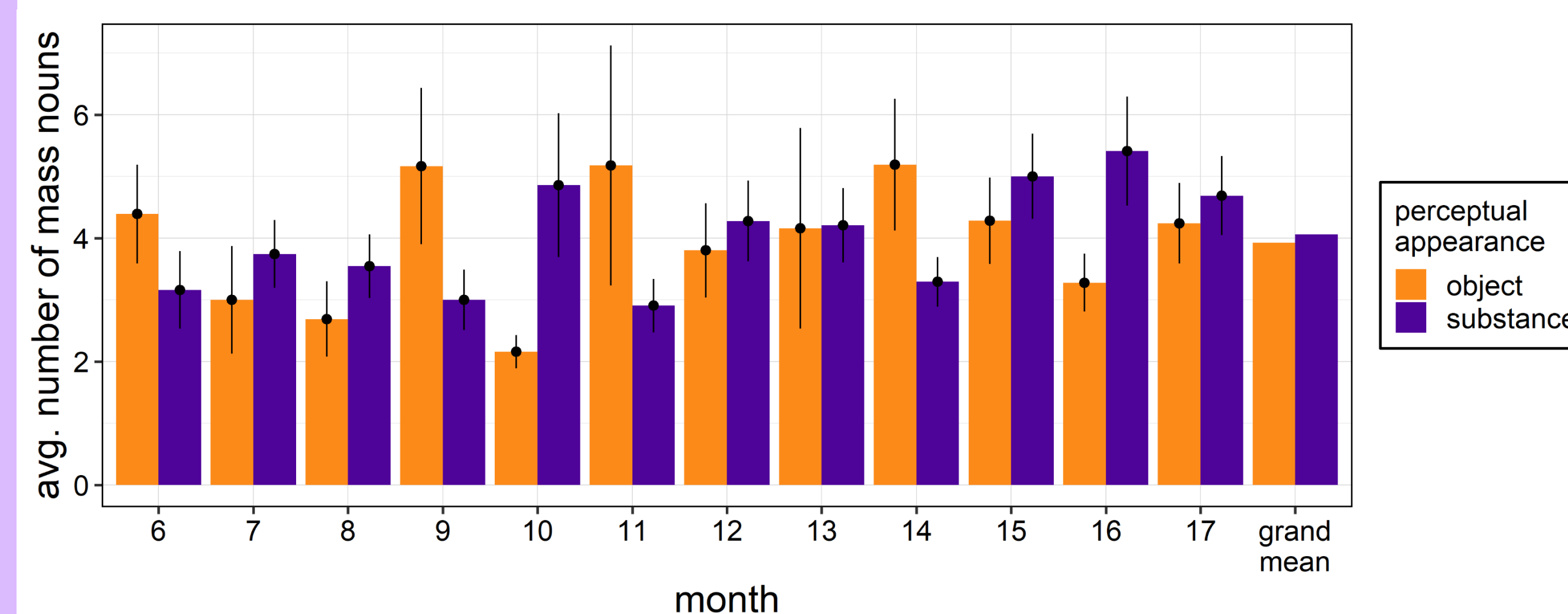
- No systematic increase or decrease in either perceptual experience over time

A Closer Look

Mass nouns that appeared with count syntax

- Most count syntax occurred when the words looked like objects.
  - All nouns appeared with ambiguous syntax at least some of the time.
  - 15 nouns **never** appeared with count syntax
  - 5 (potentially mass) nouns **never** appeared with mass syntax

Discussion & Open Questions

- Ongoing examination of count noun syntax and perceptual properties
  - Will address symmetry across count/mass
  - How often do count nouns appear with mass syntax?
  - Will tackle related questions about how plurals appear in input

- Can mass-count syntax be learned from the bottom up, or do infants need to expect these categories in order to learn them?
  - To be investigated by training a model on this data and comparing model's predictions to real infants' behaviour.

- These findings predict that it should be easier to learn often-substance mass nouns before often-count nouns; to be tested in ongoing work

Conclusions

1. Infants mostly hear about mass nouns with ambiguous syntax, resulting in a lot of uninformative data that they must sift through
2. Syntax that *is* informative:
   a) For substances: aligns well with visual percepts for substances, creating **coherent statistics** to learn from
   b) For objects: only points in the right direction half the time, which could make it harder to learn these nouns

Citations

1. Bale, A., & Barner, D. (2018). Quantity judgment and the mass-count distinction across languages: Advances, problems, and future directions for research. *Glossa: A Journal of General Linguistics, 3*(1), 63.
2. Samuelson, L. K., & Smith, L. B. (1999). Early noun vocabularies: Do ontology, category structure and syntax correspond? *Cognition, 73*(1), 1–33.
3. Huntley-Fenner, G., Carey, S., & Solimando, A. (2002). Objects are individuals but stuff doesn't count: Perceived rigidity and cohesiveness influence infants' representations of small groups of discrete entities. *Cognition, 85*(3), 203–221.
4. Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., & Pethick, S. J. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development, 59*(5), 1–173; discussion 174-185.
5. Barner, D., & Snedeker, J. (2005). Quantity judgments and individuation: Evidence that mass nouns count. *Cognition, 97*(1), 41–66.