Evidence of accurate logical reasoning in online sentence comprehension

34th Annual CUNY Conference on Human Sentence Processing

Maksymilian Dąbkowski[1], Roman Feiman[2]

March 4, 2021

[1]University of California, Berkeley, [2]Brown University

Introduction

logic in thought

- what is the status of logic in thought?
- logic studies relations among propositions

Dictum de omni

All rats love to eat.
∴ All spotted rats love to eat.

- do such schemata capture the nature of thought?

logic in psychology

- psychology has focused on difficulties in logical reasoning
  - Wason's (1968) selection tasks easier when ecologically valid (Cheng and Holyoak, 1985, 1989; Cheng, Holyoak, et al., 1986)
  - dual-process theories (Evans and Stanovich, 2013; Kahneman, 2011)

A	$\gg$	B
Birds have an ulnar artery.		Birds have an ulnar artery.
$\therefore$ Robins have an ulnar artery.		$\therefore$ Penguins have an ulnar artery.

Sloman (1993)

system 1 … has little understanding of logic and statistics

Kahneman (2011)

- **formal semantics** presupposes logical ability
  the logical notions are embedded in our deepest nature, in the very form of our language and thought
  Chomsky (1988, p. 99)
- linguists predict some logical thought as effortless as language
- can we find evidence for **spontaneous logical computation**?
- **entailment**: if p is true, then q is also true

Dictum de omni

All rats love to eat.
∴ All spotted rats love to eat.

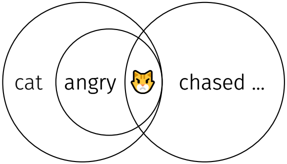
Entailment directions

some: the first argument

$\textit{some of}\ \underbrace{\textit{the cats}}_{\text{first argument}}\ \underbrace{\textit{chased a mouse}}_{\text{second argument}}$

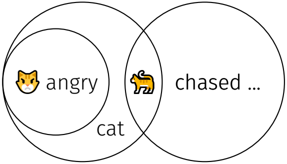
$\textit{some of the}\ \overbrace{\textit{angry cats}}^{\text{subset}}\ \textit{chased a mouse}$

$\models$ *some of the cats chased a mouse*



$\textit{some of}\ \overbrace{\textit{the cats}}^{\text{superset}}\ \textit{chased a mouse}$

$\not\models$ *some of the angry cats chased a mouse*



some, the first argument:

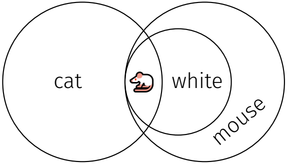
you can go from a subset to a larger set (*angry cat* $\rightsquigarrow$ *cat*)

some: the second argument

$\textit{some of } \underbrace{\textit{the cats}}_{\text{first argument}} \; \underbrace{\textit{chased a mouse}}_{\text{second argument}}$

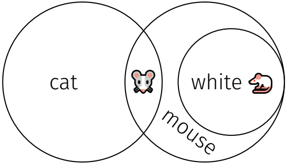
$\textit{some of the cats chased } \overbrace{\textit{a white mouse}}^{\text{subset}}$

$\models$ *some of the cats chased a mouse*



$\textit{some of the cats chased } \overbrace{\textit{a mouse}}^{\text{superset}}$

$\not\models$ *some of the cats chased a white mouse*



some, the second argument:

you can go from a subset to a larger set (*white mouse* $\rightsquigarrow$ *mouse*)

entailment direction

- *upward entailment*: can go from a subset to a larger set
- 1st arg of *some*: upward-entailing (*angry cat* ⇝ *cat*)
- 2nd arg of *some*: upward-entailing (*white mouse* ⇝ *mouse*)

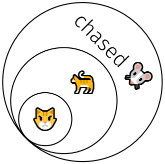
- *downward entailment*: can go from a superset to a smaller set

all: both arguments

$$\textit{all of}\ \underbrace{\textit{the cats}}_{\text{first argument}}\ \underbrace{\textit{chased a mouse}}_{\text{second argument}}$$

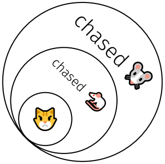
all of $\overbrace{\textit{the cats}}^{\text{superset}}$ *chased a mouse*

$\models$ *all of the angry cats chased a mouse*



all of the cats chased $\overbrace{\textit{a white mouse}}^{\text{subset}}$

$\models$ *all of the cats chased a mouse*



all: downward-entailing on the 1st argument (*cat* $\rightsquigarrow$ *angry cat*),

upward-entailing on the 2nd argument (*white mouse* $\rightsquigarrow$ *mouse*)

entailment direction by quantifier and argument

	SOME	*NOT ALL*	*ALL*	*NONE*
FIRST ARG	upward	upward	downward	downward
SECOND ARG	upward	downward	upward	downward

evidence for entailment computation

- presupposed by accounts of:
  - Gricean implicature computation
  - distribution of negative polarity items (NPIs) (e.g. Ladusaw, 1983)
- yet, little evidence for online logical computation outside of acceptability judgements
- can be challenged on empirical grounds
  - the distribution of NPIs is more complex (cf. Hoeksema, 2012)

- previous studies:
  - Deschamps et al. (2015): signature of quantifier's direction of entailment
  - Agmon et al. (2019): signatures of both negative polarity and downward entailment
- limitation: inferences tested indirectly

Hoeksema's (2012) 12 classes of polarity items

1. negation
2. *yes*/*no*-questions
3. WH-questions
4. comparatives of inequality
5. conditional clauses
6. restriction of universals
7. restriction of *the only*
8. restriction of superlatives
9. scope of *only*

	1	2	3	4	5	6	7	8	9
Any	+	+	+	+	+	+	+	+	+
Ever	+	+	+	+	+	+	+	+	+
Ook maar	+	+	+	+	+	+	+	+	+
Minimizer	+	+	+	+	+	+	+	-	-
Remotely	+	+	+	+	+	+	+	-	+
At all	+	+	+	+	+	+	+	-	+
Adv. Any	+	+	+	+	+	+	+	-	+
Yet	+	+	-	+	+	-/+	+	+	+
Either	+	+	-	+	-	-	-	-	-
In X	+	-	-	+	-	-	+	+	-
Can help	+	+	+	+	+	-/+	-	-	-
Can blame	+	+	+	-	-	-	+	-	-
Kwaad kunnen	+	+	+	-	-	-	+	-	+
Need, etc.	+	+	+	+	-	-/+	+	-	+
Anymore (US)	+	-	-	-	-	-	+	-	-
Squat	+	-	-	-	-	-	+	-	-
Exactly	+	-	-	-	-	-	-	-	-
Meer/mehr	+	-	-	-	-	-	-	-	-

Methods

methods

- three novel self-paced reading experiments
- tested for signatures of accurate inferences between quantified sentences
- experiment 1 involved detecting logical contradictions
- experiments 2 and 3 leveraged variable entailments of the first and second arguments of quantifiers to detect incorrect inferences
- preregistered design and analyses on OSF

Experiment 1

experiment 1

- tested whether speakers detect logical contradictions
- 400 participants on Amazon Mechanical Turk
- 12 target items displayed line by line
- 6 conditions differing in quantifiers

Test item

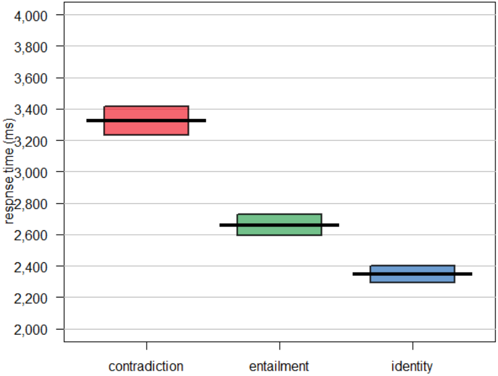
(1) *A group of scientists wanted to know whether spotted rats,*
(2) *who are pickier eaters than other rats, liked a new kind of food.*
(3) *They tested white, black, and spotted rats of both sexes.*
(4) *The scientists discovered that QUANT1 of the rats loved the food.*
(5) *Now that they knew that QUANT2 of the rats loved the food,*
(6) *they decided to issue a recommendation based on their findings.*

- measured variable: RT of the conclusion line (5)
- participants were asked unrelated comprehension questions
  - *The researchers studied rodents.* TRUE FALSE

experiment 1 conditions

	QUANT1	QUANT2
IDENTITY	***some*** *of the rats loved . . . .*	*they knew that* ***some*** *of the rats . . .*
IDENTITY	***not all*** *of the rats loved . . .*	*they knew that* ***not all*** *of the rats . . .*
ENTAILMENT	***all*** *of the rats loved . . . .*	*they knew that* ***some*** *of the rats . . .*
ENTAILMENT	***none*** *of the rats loved . . . . .*	*they knew that* ***not all*** *of the rats . . .*
CONTRADICTION	***none*** *of the rats loved . . . . .*	*they knew that* ***some*** *of the rats . . .*
CONTRADICTION	***all*** *of the rats loved . . . .*	*they knew that* ***not all*** *of the rats . . .*

experiment 1 results



- identity
- entailment
- contradiction

LMER effect between contradiction and entailment: $\chi^2 = 161.31$

$p < 0.001$

Experiment 2

experiment 2

- same paradigm to detect subtler unlicensed inferences (n = 400)
- manipulated quantifiers and premise quantifier's 1st arg

Test item

(1) *A group of scientists wanted to know whether spotted rats,*
(2) *who are pickier eaters than other rats, liked a new kind of food.*
(3) *They tested white, black, and spotted rats of both sexes.*
(4) *The scientists discovered that QUANT of the ((male) spotted) rats loved the food.*
(5) *Now that they knew that QUANT of the spotted rats loved the food,*
(6) *they decided to issue a recommendation based on their findings.*

- 4 quantifiers × 3 containment relations = 12 conditions
  - 4 conditions: premise identical to (trivally entails) conclusion
  - 4 conditions: premise entails conclusion
  - 4 conditions: premise does not entail conclusion
- within quantifier, critical lines have identical lexical content

experiment 2 conditions, full

	SOME	NOT ALL	ALL	NONE
SUBSET → of spotted rats →	... some of the male spotted rats loved the food. Now that they knew that some of the spotted rats ...	... not all of the male spotted rats loved the food. Now that they knew that not all of the spotted rats ...	... all of the male spotted rats loved the food. Now that they knew that all of the spotted rats ...	... none of the male spotted rats loved the food. Now that they knew that none of the spotted rats ...
IDENTICAL → to spotted rats →	... some of the spotted rats loved the food. Now that they knew that some of the spotted rats ...	... not all of the spotted rats loved the food. Now that they knew that not all of the spotted rats ...	... all of the spotted rats loved the food. Now that they knew that all of the spotted rats ...	... none of the spotted rats loved the food. Now that they knew that none of the spotted rats ...
SUPERSET → of spotted rats →	... some of the rats loved the food. Now that they knew that some of the spotted rats ...	... not all of the rats loved the food. Now that they knew that not all of the spotted rats ...	... all of the rats loved the food. Now that they knew that all of the spotted rats ...	... none of the rats loved the food. Now that they knew that none of the spotted rats ...

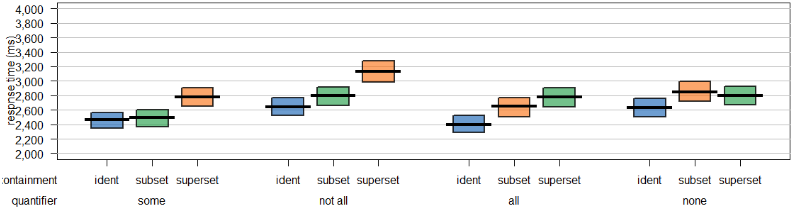
- trivially entailed
- entailed
- not entailed

experiment 2 conditions, abridged

	SOME	*NOT ALL*	*ALL*	*NONE*
SUBSET	entl’d	entl’d	¬entl’d	¬entl’d
IDENT	triv’l	triv’l	triv’l	triv’l
SUPERSET	¬entl’d	¬entl’d	entl’d	entl’d

- trivially entailed
- entailed
- not entailed

experiment 2 results



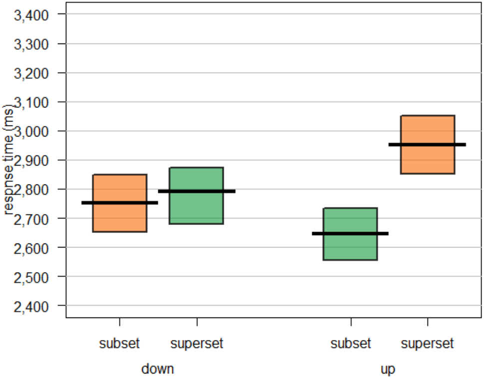
- trivial
- entailed
- not entailed

subset: *male spotted rats ≺ spotted rats*

ident: *spotted rats ≺ spotted rats*

superset: *rats ≺ spotted rats*

experiment 2 results, quantifiers grouped by entailment



- entailed
- not entailed

subset: *male spotted rats* $\prec$ *spotted rats*

superset: *rats* $\prec$ *spotted rats*

containment (subset vs. superset) × entailment (up vs. down): $\chi^2 = 10.9$, $p < 0.001$

Experiment 3

experiment 3

- manipulated quantifiers and premise quantifier's 2nd arg

Test item

(1) *A group of scientists wanted to know what rats liked to eat.*
(2) *They gave rats a choice of different meats,*
(3) *as well as leafy and root vegetables, both fresh and frozen.*
(4) *They discovered that QUANT of the rats ate ((frozen) leafy) vegetables.*
(5) $\boxed{\textit{Now that they knew that QUANT of the rats ate leafy vegetables,}}$
(6) *they decided to issue a recommendation based on their findings.*

- 12 conditions, with different interactions of quantifier × containment relation

experiment 3 conditions, full

	SOME	NOT ALL	ALL	NONE
SUBSET → of leafy veg. →	… **some** of the rats ate **frozen leafy vegetables**. Now that they knew that **some** of the rats ate **leafy vegetables** …	… **not all** of the rats ate **frozen leafy vegetables**. Now that they knew that **not all** of the rats ate **leafy vegetables** …	… **all** of the rats ate **frozen leafy vegetables**. Now that they knew that **all** of the rats ate **leafy vegetables** …	… **none** of the rats ate **frozen leafy vegetables**. Now that they knew that **none** of the rats ate **leafy vegetables** …
IDENTICAL → to leafy veg. →	… **some** of the rats ate **leafy vegetables**. Now that they knew that **some** of the rats ate **leafy vegetables** …	… **not all** of the rats ate **leafy vegetables**. Now that they knew that **not all** of the rats ate **leafy vegetables** …	… **all** of the rats ate **leafy vegetables**. Now that they knew that **all** of the rats ate **leafy vegetables** …	… **none** of the rats ate **leafy vegetables**. Now that they knew that **none** of the rats ate **leafy vegetables** …
SUPERSET → of leafy veg. →	… **some** of the rats ate **vegetables**. Now that they knew that **some** of the rats ate **leafy vegetables** …	… **not all** of the rats ate **vegetables**. Now that they knew that **not all** of the rats ate **leafy vegetables** …	… **all** of the rats ate **vegetables**. Now that they knew that **all** of the rats ate **leafy vegetables** …	… **none** of the rats ate **vegetables**. Now that they knew that **none** of the rats ate **leafy vegetables** …

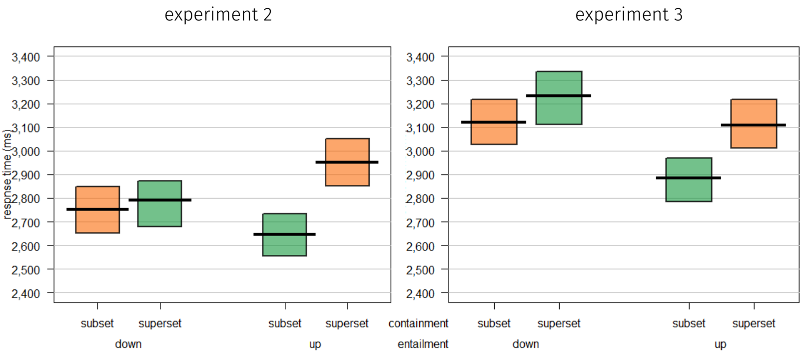
- trivially entailed
- entailed
- not entailed

experiment 2 and 3 conditions, compared

exp. 2: 1st arg of	*SOME*	*NOT ALL*	*ALL*	*NONE*
SUBSET	entl'd	entl'd	¬entl'd	¬entl'd
IDENT	triv'l	triv'l	triv'l	triv'l
SUPERSET	¬entl'd	¬entl'd	entl'd	entl'd

exp. 3: 2nd arg of	*SOME*	*NOT ALL*	*ALL*	*NONE*
SUBSET	entl'd	¬entl'd	entl'd	¬entl'd
IDENT	triv'l	triv'l	triv'l	triv'l
SUPERSET	¬entl'd	entl'd	¬entl'd	entl'd

experiment 2 and 3 results



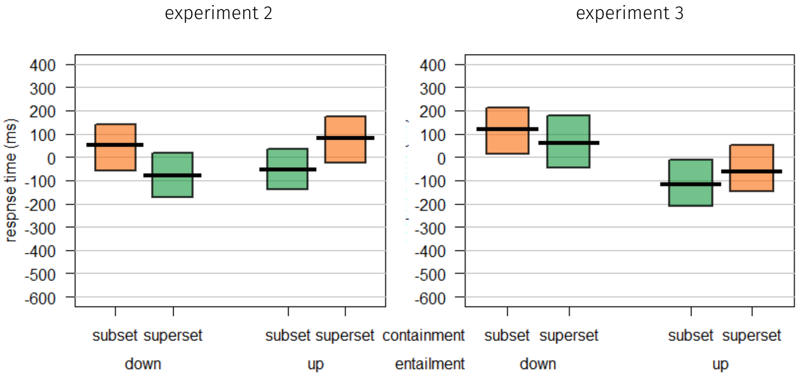
■ entailed

■ not entailed

experiment 3

containment (subset vs. superset)
× entailment (up vs. down):
$\chi^2 = 6.21$
$p = 0.013$

experiments 2 and 3, partial residual graphs



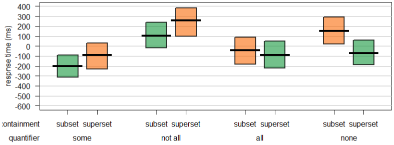
■ entailed

■ not entailed

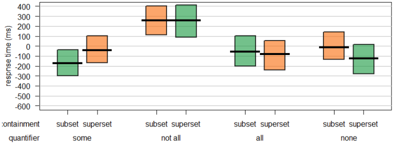
between experiments

containment × experiment × entailment flip (yes vs. no):
$\chi^2 = 0.98$
$p = 0.32$

experiment 2



experiment 3



- entailed
- not entailed

Discussion

discussion

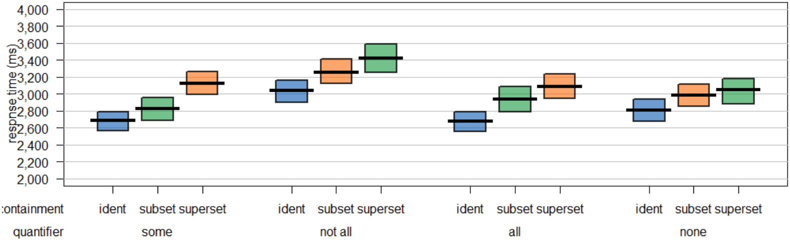
- language involves accurate and spontaneous logical computations
- differs from dual-process theories of cognition

  > *it is assumed that people's intuitive logical knowledge emerges from a learning process in which key principles have been practiced to automaticity*
  >
  > De Neys and Pennycook (2019)

- consistent with some logic being naturally intuitive
  - natural logic in reasoning (e. g. Braine and O'Brien, 1998)
  - logic (L-analyticity) in grammar (e. g. Gajewski, 2002)
- inference derives from compositionality?
- some logical competence revealed more easily in natural language comprehension than in puzzles and tests
- new empirical terrain: which inferences follow from structure of language?

thank you!

experiment 3 results



- trivial
- entailed
- not entailed

subset: *frozen leafy vegetables* $\prec$ *leafy vegetables*

ident: *leafy vegetables* $\prec$ *leafy vegetables*

superset: *vegetables* $\prec$ *leafy vegetables*

Agmon, Galit, Yonatan Loewenstein, and Yosef Grodzinsky (2019). "Measuring the cognitive cost of downward monotonicity by controlling for negative polarity". In: *Glossa: A Journal of General Linguistics* 4.1.

Braine, Martin D. S. and David P. O'Brien (1998). *Mental logic*. Psychology Press.

Cheng, Patricia W. and Keith J. Holyoak (1985). "Pragmatic Reasoning Schemas". In: *Cognitive Psychology* 17.4, pp. 391–416.

Cheng, Patricia W. and Keith J. Holyoak (1989). "On the natural selection of reasoning theories". In: *Cognition*.

Cheng, Patricia W., Keith J. Holyoak, Richard E. Nisbett, and Lindsay M. Oliver (1986). "Pragmatic versus syntactic approaches to training deductive reasoning". In: *Cognitive Psychology* 18.3, pp. 293–328.

De Neys, Wim and Gordon Pennycook (2019). "Logic, fast and slow: Advances in dual-process theorizing". In: *Current Directions in Psychological Science* 28.5, pp. 503–509.

Deschamps, Isabelle, Galit Agmon, Yonatan Loewenstein, and Yosef Grodzinsky (2015). "The processing of polar quantifiers, and mumerosity perception". In: *Cognition* 143, pp. 115–128. ISSN: 0010-0277. DOI: https://doi.org/10.1016/j.cognition.2015.06.006. URL: https://www.sciencedirect.com/science/article/pii/S0010027715300160.

Evans, Jonathan St. B. T. and Keith E. Stanovich (2013). "Dual-process theories of higher cognition: Advancing the debate". In: *Perspectives on Psychological Science* 8.3, pp. 223–241.

Gajewski, Jon (2002). "L-analyticity and natural language". Manuscript. Cambridge, MA: MIT.

Hoeksema, Jack (2012). "On the natural history of negative polarity items". In: *Linguistic Analysis* 38.1/2, pp. 3–33.

Kahneman, Daniel (2011). *Thinking, Fast and Slow*. Farrar, Straus and Giroux.

Ladusaw, William A. (1983). "Logical form and conditions on grammaticality". In: *Linguistics and Philosophy* 6.3, pp. 373–392.

Sloman, Steven A. (1993). "Feature-based induction". In: *Cognitive Psychology* 25.2, pp. 231–280.

Wason, Peter C. (1968). "Reasoning about a rule". In: *Quarterly Journal of Experimental Psychology* 20.3, pp. 273–281.