Where linguistic meaning meets non-linguistic cognition

Tim Hunter and Paul Pietroski

NASSLLI 2016
Friday: Putting things together (perhaps)
Outline

5 Experiments with kids on ‘most’

6 So: How should we think about meanings?

7 Multiple non-equivalent interpretations

8 Application to ‘most’
## Outline

5. **Experiments with kids on ‘most’**

6. **So: How should we think about meanings?**

7. **Multiple non-equivalent interpretations**

8. **Application to ‘most’**
Kids and ‘most’

“Are most of the animals frogs, or rabbits?”

(Halberda et al. 2008, Odic et al. submitted)
Kids and ‘most’

“Are most of the animals frogs, or rabbits?”

Non-full-counter stage:

- some kids at chance
- some kids with above-chance but noisy accuracy (ANS)

(Halberda et al. 2008, Odic et al. submitted)
Kids and ‘most’

“Are most of the animals frogs, or rabbits?”

Non-full-counter stage:
- some kids at chance
- some kids with above-chance but noisy accuracy (ANS)

Full-counter stage:
- some kids at chance
- some kids with above-chance and adult-like accuracy
- some kids with above-chance but noisy accuracy (ANS)

(Halberda et al. 2008, Odic et al. submitted)
Summary

- Experiment 1: adults neglect one-to-one information and persist with ANS
- Experiment 2: adults neglect direct number-of-non-yellows information and persist with subtraction
- Experiment 3: children neglect precise cardinality information and persist with ANS
Summary

Experiments with kids on ‘most’

Summary

‘Most of the dots are yellow’

\[
\text{OneToOnePlus}(\text{DOT} \cap \text{YELLOW}, \text{DOT} \setminus \text{YELLOW})
\]

\[
\begin{align*}
\text{ANS representation } & G(y, w^2 y^2) \text{ for } |\text{YELLOW} \cap \text{DOT}| \\
\text{ANS representation } & G(b, w^2 b^2) \text{ for } |\text{BLUE} \cap \text{DOT}| \\
\text{ANS representation } & G(d, w^2 d^2) \text{ for } |\text{DOT}|
\end{align*}
\]

Judgements with accuracy poorer than one-to-one control task

Judgements with accuracy poorer than simple ‘more’ task

Judgements with accuracy dependent on ratio

Precise number label ‘\( y \)’ for \(|\text{YELLOW} \cap \text{DOT}|\)

Precise number label ‘\( b \)’ for \(|\text{BLUE} \cap \text{DOT}|\)
Experiments with kids on ‘most’

So: How should we think about meanings?

Multiple non-equivalent interpretations

Application to ‘most’
What are meanings?

Experiments with kids on 'most'

So: How should we think about meanings?

Multiple non-equivalent interpretations

Application to 'most'

'Most of the dots are yellow'?

true/false

48/62
What are meanings?

 DOT ∩ YELLOW > DOT ∩ \overline{YELLOW} 

ˈMost of the dots are yellowˈ \rightarrow \text{linguistic processing} \rightarrow ? \rightarrow \text{non-linguistic processing} \rightarrow \text{true/false}
Experiments with kids on ‘most’

So: How should we think about meanings?

Multiple non-equivalent interpretations

Application to ‘most’

What are meanings?

\[ |\text{DOT} \cap \text{YELLOW}| > |\text{DOT}| - |\text{DOT} \cap \text{YELLOW}| \]

‘Most of the dots are yellow’

linguistic processing

\?

non-linguistic processing

true/false
What are meanings?

\[ |\text{DOT} \cap \text{YELLOW}| > |\text{DOT} \cap \overline{\text{YELLOW}}| \]

'Most of the dots are yellow'

linguistic processing

? 

non-linguistic processing

true/false

Experiments with kids on 'most'

So: How should we think about meanings?

Multiple non-equivalent interpretations

Application to 'most'
So: What is a meaning?

We’ve been guided by the idea of “choosing between” things like the following:

\[
\begin{align*}
\text{DOT} \cap \text{YELLOW} &\succ \text{DOT} \cap \overline{\text{YELLOW}} \\
|\text{DOT} \cap \text{YELLOW}| &> |\text{DOT} \cap \overline{\text{YELLOW}}| \\
|\text{DOT} \cap \text{YELLOW}| &> |\text{DOT} - |\text{DOT} \cap \text{YELLOW}| |
\end{align*}
\]
So: What is a meaning?

We’ve been guided by the idea of “choosing between” things like the following:

\[
\text{DOT} \cap \text{YELLOW} \succ \text{DOT} \cap \overline{\text{YELLOW}} \\
|\text{DOT} \cap \text{YELLOW}| > |\text{DOT} \cap \overline{\text{YELLOW}}| \\
|\text{DOT} \cap \text{YELLOW}| > |\text{DOT}| - |\text{DOT} \cap \text{YELLOW}|
\]

But what exactly is the relation between

- the things we have discovered about the ways people go about making true/false judgements, and
- the forms of the mathematical expressions above?

What linking hypotheses connect these mental representations to observations? (cf. “When the state to transition to is not determined by the input, things take longer.”)
So: What is a meaning?

We’ve been guided by the idea of “choosing between” things like the following:

\[
\text{DOT} \cap \text{YELLOW} \succ \text{DOT} \cap \overline{\text{YELLOW}} \\
|\text{DOT} \cap \text{YELLOW}| > |\text{DOT} \cap \overline{\text{YELLOW}}| \\
|\text{DOT} \cap \text{YELLOW}| > |\text{DOT}| - |\text{DOT} \cap \text{YELLOW}|
\]

But what exactly is the relation between

- the things we have discovered about the ways people go about making true/false judgements, and
- the forms of the mathematical expressions above?

What linking hypotheses connect these mental representations to observations? (cf. “When the state to transition to is not determined by the input, things take longer.”)

Idea: say something about how the symbols are interpreted by non-linguistic cognitive systems (cf. Chomsky on “instructions for performance systems”)

I will not be claiming to have *derived* the connection between ‘most’ and subtraction-ish, non-one-to-one-ish verification procedures, from other independently justified assumptions. This remains a stipulation (for now).
Expectation-lowering disclaimer

I will not be claiming to have derived the connection between ‘most’ and subtraction-ish, non-one-to-one-ish verification procedures, from other independently justified assumptions. This remains a stipulation (for now).

I’m just trying to find a way to at least encode these stipulations into our theories. (Something like “observational adequacy”.)

Hopefully, in the process we might get ideas about what to look for in order to get rid of some of these stipulations.
I will not be claiming to have derived the connection between ‘most’ and subtraction-ish, non-one-to-one-ish verification procedures, from other independently justified assumptions. This remains a stipulation (for now).

I’m just trying to find a way to at least encode these stipulations into our theories. (Something like “observational adequacy”.)

Hopefully, in the process we might get ideas about what to look for in order to get rid of some of these stipulations.

(But I don’t really know whether we’re getting anywhere yet . . .)
Outline

5 Experiments with kids on ‘most’

6 So: How should we think about meanings?

7 Multiple non-equivalent interpretations

8 Application to ‘most’
Notation for semantic rules

\[
[[S \ [NP \ \alpha] \ [VP \ \beta]]] = [[VP \ \beta]]([[NP \ \alpha]])
\]

\[
\frac{[[NP \ \alpha]] = x \quad [[VP \ \beta]] = f}{[[S \ [NP \ \alpha] \ [VP \ \beta]]] = f(x)}
\]
Analogy from programming languages

Program $P_1$

```plaintext
let x = 3;
if ((x > 0) || (f(x) > 10)) then
  return "yes";
else
  return "no";
```

Program $P_2$

```plaintext
let x = 3;
if ((f(x) > 10) || (x > 0)) then
  return "yes";
else
  return "no";
```

Evaluating ‘$A || B$’ (as in many programming languages) means:

- evaluate $A$
- if the result is true, stop; the whole expression evaluates to true
- otherwise, evaluate $B$, ...

\[
\begin{align*}
\lbrack A \rbrack = \text{true} & \quad \Rightarrow \quad \lbrack (A \ || \ B) \rbrack = \text{true} \\
\lbrack A \rbrack = \text{false} & \quad \text{and} \quad \lbrack B \rbrack = v \\
\Rightarrow \quad \lbrack (A \ || \ B) \rbrack = v
\end{align*}
\]

Therefore:

- The two programs differ in that $P_1$ will not evaluate $(f(x) > 10)$, but $P_2$ will
- Nonetheless, the two programs will both return the same result
Experiments with kids on ‘most’

So: How should we think about meanings?

Multiple non-equivalent interpretations

Application to ‘most’

Analogy from programming languages

Program $P_1$

```plaintext
let x = 3;
if ((x > 0) || (f(x) > 10)) then
    return "yes";
else
    return "no";
```

Program $P_2$

```plaintext
let x = 3;
if ((f(x) > 10) || (x > 0)) then
    return "yes";
else
    return "no";
```

What is the meaning of $P_1$? Of $P_2$? Are they the same?

- If we are interested only in the result returned, we can define a suitable interpretation function $\llbracket \cdot \rrbracket$ which maps programs to $\{ \text{true}, \text{false} \}$ such that $\llbracket P_1 \rrbracket = \llbracket P_2 \rrbracket$

\[
\begin{align*}
\llbracket A \rrbracket &= \text{true} & \llbracket A \rrbracket &= \text{false} \\
\llbracket B \rrbracket &= v & \llbracket (A \; \text{||} \; B) \rrbracket &= \text{true} \\
\llbracket (A \; \text{||} \; B) \rrbracket &= v
\end{align*}
\]

- But there are things that might make us prefer a different interpretation function
Analogy from programming languages

Program $P_1$

```
let x = 3;
if ((x > 0) || (f(x) > 10)) then
  return "yes";
else
  return "no";
```

Program $P_2$

```
let x = 3;
if ((f(x) > 10) || (x > 0)) then
  return "yes";
else
  return "no";
```

Perhaps $f(x)$ is very complicated and slow to compute.

We could define an interpretation function which maps programs to ordered pairs, a boolean result and a number of steps.

\[
\begin{align*}
\lceil A \rceil &= \langle \text{true}, n_1 \rangle \\
\lceil (A \lor B) \rceil &= \langle \text{true}, n_1 + 1 \rangle \\
\lceil A \rceil &= \langle \text{false}, n_1 \rangle \\
\lceil B \rceil &= \langle v, n_2 \rangle \\
\lceil (A \lor B) \rceil &= \langle v, n_1 + n_2 + 1 \rangle
\end{align*}
\]

Here $\lceil P_1 \rceil \neq \lceil P_2 \rceil$.

Program $P_1$

\[
\lceil (x > 0) \rceil = \langle \text{true}, 1 \rangle
\]

\[
\lceil (x > 0) \lor (f(x) > 10) \rceil = \langle \text{true}, 2 \rangle
\]

Program $P_2$

\[
\lceil (f(x) > 10) \rceil = \langle \text{true}, 100 \rangle
\]

\[
\lceil (f(x) > 10) \lor (x > 0) \rceil = \langle \text{true}, 101 \rangle
\]
Analogy from programming languages

Program $P_1$

\[
\begin{align*}
&\text{let } x = 3; \\
&\text{if } ((x > 0) \lor (f(x) > 10)) \text{ then} \\
&\quad \text{return } "yes"; \\
&\text{else} \\
&\quad \text{return } "no";
\end{align*}
\]

Program $P_2$

\[
\begin{align*}
&\text{let } x = 3; \\
&\text{if } ((f(x) > 10) \lor (x > 0)) \text{ then} \\
&\quad \text{return } "yes"; \\
&\text{else} \\
&\quad \text{return } "no";
\end{align*}
\]

Perhaps $f(x)$ changes the value of a global variable (say, sets $y$ to 5)

We could define $[\cdot]^s$ which maps programs to ordered pairs, a boolean result and a new state

\[
\begin{align*}
[A]^s &= \langle \text{true}, s' \rangle \\
[(A \lor B)]^s &= \langle \text{true}, s' \rangle \\
[A]^s &= \langle \text{false}, s' \rangle \\
[B]^s &= \langle v, s'' \rangle \\
[(A \lor B)]^s &= \langle v, s'' \rangle
\end{align*}
\]

Here $[P_1]^s \neq [P_2]^s$.
Multiple Interpretations

For many different questions we might want to ask about the interpretation of a program/expression, we can define a compositionally-determined function from programs/expressions to the appropriate answer

- returned value
- returned value and number of steps
- returned value and updated state
Multiple Interpretations

\[ |\text{DOT} \cap \text{YELLOW}| > |\text{DOT}| - |\text{DOT} \cap \text{YELLOW}| \]
\[ |\text{DOT} \cap \text{YELLOW}| > |\text{DOT} \cap \overline{\text{YELLOW}}| \]

- We can give this language a variety of interpretations, just as we gave ‘\(A \mid\mid B\)’ multiple different interpretations, each answering a different question about its semantic properties.
Multiple Interpretations

\[ |\text{DOT} \cap \text{YELLOW}| > |\text{DOT}| - |\text{DOT} \cap \text{YELLOW}| \]
\[ |\text{DOT} \cap \text{YELLOW}| > |\text{DOT} \cap \overline{\text{YELLOW}}| \]

- We can give this language a variety of interpretations, just as we gave ‘\(A \ | | B\)’ multiple different interpretations, each answering a different question about its semantic properties.

- A “normal” interpretation can answer the verification-agnostic questions about truth values.

- A new interpretation can answer other questions: e.g. How much noise accompanies the result of a particular computation?
Outline

5 Experiments with kids on ‘most’

6 So: How should we think about meanings?

7 Multiple non-equivalent interpretations

8 Application to ‘most’
Composing a sentence meaning

\[[\text{most of the dots are yellow}]^{SEM} = |D \& Y| > |D| - |D \& Y|\]
Composing a sentence meaning

\[
[\text{most of the dots are yellow}]^{\text{SEM}} = |D \& Y| > |D| - |D \& Y|
\]

\[
[\text{most}]^{\text{SEM}} = \lambda a \lambda b. |a \& b| > |a| - |a \& b|
\]

\[
[\text{dot}]^{\text{SEM}} = D
\]

\[
[\text{yellow}]^{\text{SEM}} = Y
\]

\[
[\text{Det of the Ns are Adj}]^{\text{SEM}} = [\text{Det}]^{\text{SEM}} ([\text{N}]^{\text{SEM}})([\text{Adj}]^{\text{SEM}})
\]
Composing a sentence meaning

\[
\text{[[most of the dots are yellow]]}^{\text{SEM}} = |D \& Y| > |D| - |D \& Y|
\]

\[
\text{[[most]]}^{\text{SEM}} = \lambda a \lambda b. |a \& b| > |a| - |a \& b|
\]

\[
\text{[[dot]]}^{\text{SEM}} = D
\]

\[
\text{[[yellow]]}^{\text{SEM}} = Y
\]

\[
\text{[[Det of the Ns are Adj]]}^{\text{SEM}} = \text{[[Det]]}^{\text{SEM}} (\text{[[N]]}^{\text{SEM}}) (\text{[[Adj]]}^{\text{SEM}})
\]

\[
\text{[[N]]}^{\text{SEM}} = a \quad \text{[[Adj]]}^{\text{SEM}} = b
\]

\[
\text{[[most of the Ns are Adj]]}^{\text{SEM}} = |a \& b| > |a| - |a \& b|
\]
Definition of ANS interpretation

\[
\begin{align*}
\begin{bmatrix} e_1 \end{bmatrix}^{\text{ANS}} &= G(n_1, \sigma_1^2) & \begin{bmatrix} e_2 \end{bmatrix}^{\text{ANS}} &= G(n_2, \sigma_2^2) \\
\begin{bmatrix} e_1 - e_2 \end{bmatrix}^{\text{ANS}} &= G(n_1 - n_2, \sigma_1^2 + \sigma_2^2) \\
\begin{bmatrix} e_1 \end{bmatrix}^{\text{ANS}} &= G(n_1, \sigma_1^2) & \begin{bmatrix} e_2 \end{bmatrix}^{\text{ANS}} &= G(n_2, \sigma_2^2) \\
\begin{bmatrix} e_1 > e_2 \end{bmatrix}^{\text{ANS}} &= \frac{1}{2} \text{erfc} \left( \frac{n_2 - n_1}{\sqrt{2} \sqrt{\sigma_1^2 + \sigma_2^2}} \right)
\end{align*}
\]
Examples

Our ‘most’ experiments with adults:

\[\text{Most of the dots are yellow}^{\text{SEM}} = |D \& Y| > |D| - |D \& Y|\]

\[
\begin{align*}
|D \& Y|^{\text{ANS}} &= G(y, w^2 y^2) \\
|D|^{\text{ANS}} &= G(y + b, w^2(y + b)^2) \\
|D \& Y|^{\text{ANS}} &= G(y, w^2 y^2) \\
|D| - |D \& Y|^{\text{ANS}} &= G(b, w^2(y + b)^2 + w^2 y^2) \\
|D \& Y| > |D| - |D \& Y|^{\text{ANS}} &= \frac{1}{2} \text{erfc} \left( \frac{b - y}{\sqrt{2 \sqrt{w^2((y+b)^2+2y^2)}}} \right)
\end{align*}
\]
Examples

Our ‘most’ experiments with adults:

\[ \text{Most of the dots are yellow}^{\text{SEM}} = |D \& Y| > |D| - |D \& Y| \]

\[
\begin{align*}
|D \& Y|^{\text{ANS}} &= G(y, w^2 y^2) \\
|D|^{\text{ANS}} &= G(y + b, w^2 (y + b)^2) \\
|D \& Y|^{\text{ANS}} &= G(b, w^2 (y + b)^2 + w^2 y^2)) \\
|D| - |D \& Y|^{\text{ANS}} &= \frac{1}{2} \text{erfc} \left( \frac{b-y}{\sqrt{2} \sqrt{w^2 ((y+b)^2 + 2y^2)}} \right)
\end{align*}
\]

Classic ANS experiment:

\[ \text{There are more yellow dots than blue dots}^{\text{SEM}} = |D \& Y| > |D \& B| \]

\[
\begin{align*}
|D \& Y|^{\text{ANS}} &= G(y, w^2 y^2) \\
|D \& B|^{\text{ANS}} &= G(b, w^2 b^2) \\
|D \& Y| > |D \& B|^{\text{ANS}} &= \frac{1}{2} \text{erfc} \left( \frac{b-y}{\sqrt{2} \sqrt{w^2 y^2 + w^2 b^2}} \right)
\end{align*}
\]
Examples

Our ‘most’ experiments with adults:

\[
\left[ \text{Most of the dots are yellow} \right]^{\text{SEM}} = |D \& Y| > |D| - |D \& Y|
\]

\[
\begin{align*}
\left[ |D \& Y| \right]^{\text{ANS}} &= G(y, w^2y^2) \\
\left[ |D| \right]^{\text{ANS}} &= G(y + b, w^2(y + b)^2) \\
\left[ |D \& Y| \right]^{\text{ANS}} &= G(y, w^2y^2) \\
\left[ |D| - |D \& Y| \right]^{\text{ANS}} &= G(b, w^2(y + b)^2 + w^2y^2) \\
\left[ |D \& Y| > |D| - |D \& Y| \right]^{\text{ANS}} &= \frac{1}{2} \text{erfc} \left( \frac{b - y}{\sqrt{2 \sqrt{w^2((y+b)^2 + 2y^2)}}} \right)
\end{align*}
\]

Classic ANS experiment:

\[
\left[ \text{There are more yellow dots than blue dots} \right]^{\text{SEM}} = |D \& Y| > |D \& B|
\]

\[
\begin{align*}
\left[ |D \& Y| \right]^{\text{ANS}} &= G(y, w^2y^2) \\
\left[ |D \& B| \right]^{\text{ANS}} &= G(b, w^2b^2) \\
\left[ |D \& Y| > |D \& B| \right]^{\text{ANS}} &= \frac{1}{2} \text{erfc} \left( \frac{b - y}{\sqrt{2 \sqrt{w^2y^2 + w^2b^2}}} \right)
\end{align*}
\]

- In other ‘most’ situations, some “reaching” happens that draws a direct comparison of count-based cardinalities into answering the ‘most’ question.
- In other ‘most’ situations, some “reaching” happens that draws the one-to-one info into answering the ‘most’ question.
- In other ‘most’ situations, ...
Experiments with kids on ‘most’

Subtraction vs. selection

16 yellow dots, 10 blue dots

With only $y$ yellow dots and $b$ blue dots present:

- **Subtraction Procedure:** non-yellow numerosity $\sim \mathcal{N}(b, w^2((y + b)^2 + y^2))$
- **Selection Procedure:** non-yellow numerosity $\sim \mathcal{N}(b, w^2 b^2)$
Definition of “precise cardinality” interpretation

\[
\begin{align*}
[e]^{\text{CARD}} &= S \\
[|e|]^{\text{CARD}} &= |S| \\
[e_1]^{\text{CARD}} &= n_1 \\
[e_2]^{\text{CARD}} &= n_2 \\
[e_1 - e_2]^{\text{CARD}} &= n_1 - n_2 \\
\end{align*}
\]

etc.
Examples

Kids who can count, but use the ANS for ‘most’:

\[ \text{[Most of the dots are yellow]}^{\text{SEM}} = |D \& Y| > |D| - |D \& Y| \]

These kids haven’t yet learned to apply \[ \text{[ ]}^{\text{CARD}} \] to this expression, only \[ \text{[ ]}^{\text{ANS}} \].

So they can use

\[ \text{[ |D \& Y|}^{\text{ANS}} = G(y, w^2 y^2) \quad \text{[ |D \& B|}^{\text{ANS}} = G(b, w^2 b^2) \]

but they can’t use

\[ \text{[ |D \& Y|}]^{\text{CARD}} = y \quad \text{[ |D \& B|}]^{\text{CARD}} = b \]
References


