Computing implicatures under QUDs

Eszter Ronai & Ming Xiang

The University of Chicago

49th Annual Meeting of the North East Linguistic Society
Proposal and roadmap

Main claim

Questions Under Discussion, rather than e.g. the complexity of alternatives, determine the (reaction time) cost of implicature calculation.

1 Background
   • Processing implicatures
   • Lexical Access
   • The role of context

2 Experiment 1: QUD elicitation

3 Experiment 2: QUD manipulation

4 Discussion
Messages that hearers infer do not always equal literal messages, e.g. implicatures.

(1) Mary ate some of the cookies.
   a. Literal: Mary ate some and possibly all of the cookies.
   b. Inference: Mary ate some but not all of the cookies.

(2) It is a cookie that Mary ate.
   a. Literal: Mary ate a cookie and possibly other things too.
   b. Inference: Mary only ate a cookie.
Background

Processing implicatures

Implicatures: window into the **integration of semantic and pragmatic knowledge**.

**Time delay cost** for the inference-enriched as compared to the literal reading.

- Reaction time (Bott & Noveck, 2004).
- Eye-tracking (Huang & Snedeker, 2009).
- ERP (Noveck & Posada, 2003).

What makes implicature calculation a costly process?

Idea from the theoretical literature: how do we construct the alternatives the speaker could have? Use complexity of alternatives to characterise them (Katzir, 2007; Fox & Katzir, 2011).

- deletion
- contraction
- substitution, from e.g. the lexicon

No processing claims though - how would this manifest?

Van Tiel & Schaeken (2017, following Chemla & Bott, 2014) present a particular implementation:

Lexical Access hypothesis

Retrieving items from the lexicon to construct the relevant alternatives is what triggers cost.
Background
Lexical Access

Van Tiel & Schaeken (2017):

Lexical Access hypothesis

Retrieving items from the lexicon to construct the relevant alternatives is what triggers cost.

They compare: scalar inference, *it*-cleft exhaustivity, free choice inference, conditional perfection.

- Only scalar inference shows a reaction time cost.
- Support for Lexical Access: retrieving all to construct relevant alternative (some but not all) is what triggers cost.
- The other inference types: no (lexical) alternatives, or alternative construction via deletion.

Key predictions and findings

Scalar inference incurs a reaction time cost, but *it*-cleft exhaustivity does not.
The role of context

A potential problem: sentences were presented in isolation, and previous research has highlighted the importance of context.

QUDs modulate how likely a scalar inference is to arise.

- explicit questions (Zondervan et al., 2008)
- background story (Degen, 2013)
- focus prosody (Cummins & Rohde, 2015)

We might predict this effect to extend to a) other types of inferences, b) cost of computation.
Our study

Hypothesis

Context is what determines the cost of implicature calculation.

We compare scalar inference (SI) and *it*-cleft exhaustivity (EXH), embedded under QUDs.

In contrast to earlier studies, QUDs are elicited empirically.
→ more systematic comparison

Known problem: how can we track down the QUDs relevant for a given context, other than relying on our own intuition? This is a first attempt at using elicitation.
General experimental design

**Background story:** Anne is asking questions from Bob, about pictures that only Bob can see.
- Control: Bob’s answers unambiguously good/bad descriptions of the picture.
- Target: descriptions either good (on literal reading) or bad (inference-enriched).

**Bob:** *Some of the shapes are blue.*

- Control: Good
- Control: Bad
- Target: Underinformative

**Bob:** *It is the square that is blue.*

- Control: Good
- Control: Bad
- Target: Underinformative
Experiment 1: QUD elicitation
Participants, procedure and task

• 40 native monolingual speakers of American English.
• Participants saw SI and EXH target sentences paired with pictures, and were told that the sentences were Bob’s answers to Anne’s questions.

(3) Anne: __________________________?
    Bob: *Some of the..., It is the...*
    Picture (Good Control or Target, between-participants)

• **Task:** guess what Anne’s question was.
Experiment 1: QUD elicitation

Results

Dominant SI questions:
- what: What color are the shapes?
- any: Are any (of the) shapes black? Are there (any) red shapes?
- all: Are all of the shapes yellow?
- some: Are some of the shapes yellow?

Dominant EXH questions:
- which: Which/what shape is black? Which one (of them) is blue?
- any: Are any of the shapes yellow? Are there any black shapes?
- what: What color are the shapes? What color is the square?

<table>
<thead>
<tr>
<th></th>
<th>SI</th>
<th>EXH</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>what</td>
<td>any</td>
</tr>
<tr>
<td>Target</td>
<td>42%</td>
<td>25%</td>
</tr>
<tr>
<td>Good Control</td>
<td>32%</td>
<td>33%</td>
</tr>
</tbody>
</table>
Experiment 2: QUD manipulation
Participants, procedure and task

- 85 native monolingual speakers of American English (different from Exp. 1).
  - 25-30 in each of the QUD conditions.
- **Sentence-picture verification task**: participants saw a dialogue between Anne and Bob, together with a picture.
- Task: make a **binary judgment** about whether Bob gave a good answer to Anne’s question, given the picture he saw.
- We are interested in their response (Good/Not Good) and reaction time.
Experiment 2: QUD manipulation
Participants, procedure and task

3 × 3 design: Picture (within-participants) × QUD (between-participants)

Anne’s questions: most frequent questions elicited from Exp. 1.:

(4) QUD manipulation in SI
   wh-word: **What** color are the shapes?
   indefinite: Are there **any** blue shapes?/Are any shapes blue?
   quantifier: Are **all** shapes blue?

(5) QUD manipulation in EXH
   wh-word: **Which*/What shape is blue?
   indefinite: Are there **any** blue shapes?
   quantifier: Are **both** shapes blue?
Experiment 2: QUD manipulation

SI example trial

QUD condition: Any, Picture condition: Target

(6) Anne: Are any shapes blue?
    Bob: Some of the shapes are blue.

Task: choose “Good” or “Not Good”.

△ □ ○ ● □ ○
Experiment 2: QUD manipulation

Predictions

“Good” responses to Target: higher % indicate a lower rate of implicature calculation.

Some of the shapes are blue.

Literal-biasing QUD → higher %
Inference-biasing QUD → lower %

Eszter Ronai & Ming Xiang (UChicago)
Experiment 2: QUD manipulation

Calculation rate results

“Good” responses to Target: higher % indicate a lower rate of implicature calculation.

significant Target differences:
- any vs. all (p<0.001)
- any vs. what (p<0.05)

replication of earlier findings
Experiment 2: QUD manipulation

Calculation rate results

“Good” responses to Target: higher % indicate a lower rate of implicature calculation.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Percent</th>
<th>Question</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bad</td>
<td>0%</td>
<td>which</td>
</tr>
<tr>
<td>Good</td>
<td>25%</td>
<td>any</td>
</tr>
<tr>
<td>Target</td>
<td>50%</td>
<td>both</td>
</tr>
</tbody>
</table>

EXH: % of Good responses

significant Target differences:
- both vs. any (p<0.001)
- both vs. which (p<0.001)

extends earlier findings to EXH
Experiment 2: QUD manipulation

Calculation rate results

Interim finding

- SI: any is a **Literal-biasing**, while what, all are **Inference-biasing** QUDs
- EXH: any and which are **Literal-biasing**, while both is an **Inference-biasing** QUD

Hypothesis: this is reflected in reaction time cost.
Experiment 2: QUD manipulation

Reaction time results

**Cost of implicature calculation:** longer reaction time when responding **Not Good (NG)** to **Target**, relative to the reaction time when responding NG to **Control**.

- Bad Control: target sentence is unambiguously a bad description.
- Target: responding “Not Good” implies the participant has gone through the inference calculation process.

---

**Some of the shapes are blue.**

Control: Bad

Target: Underinformative

---

**It is the square that is blue.**

Control: Bad

Target: Underinformative
**Experiment 2: QUD manipulation**

Reaction time results

Cost: difference in NG to Target vs. NG to Control.

significant interaction of QUD-Response (p < 0.01):

- **any**: cost
- **what, all**: no cost

→ **SI computation is only costly** when preceded by non-supportive QUDs.
Experiment 2: QUD manipulation

Reaction time results

![Graphs showing mean reaction times by question type (which, any, both) for G and NG responses.]

Cost: difference in NG to Target vs. NG to Control.

- **any**: cost (RT for NG, Target vs. Control, \(p<0.05\))
- **which**: similar but not exactly the same pattern
- **both**: unexpected cost for responding NG to Control
Discussion

Main findings

For both SI and EXH, calculation rates and \textbf{processing cost} are \textit{strongly modulated by QUD}.

- QUDs that bias towards deriving the implicature make it a cost-free process.
- QUDs that bias against it make it incur a processing cost.

Differences signal cost of inference calculation, not just naturalness of dialogues:

- QUDs were empirically elicited.
- In the binary task, acceptance rates (“Good” response) are at ceiling in the Good Control picture condition, no matter the QUD.
Discussion

Cost predicted?

<table>
<thead>
<tr>
<th></th>
<th>SI</th>
<th>EXH</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lexical Access</td>
<td>✓</td>
<td>×</td>
</tr>
<tr>
<td>QUD</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Literal-biasing</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Inference-biasing</td>
<td>×</td>
<td>×</td>
</tr>
</tbody>
</table>

Our findings: processing cost of implicature calculation is not directly/uniquely accounted for by alternative construction and the complexity of alternatives, rather it is context-dependent.

Conclusion
A QUD-based account better explains the current findings than a Lexical Access-based account.
Thank you!

We would like to thank Chris Kennedy for his support throughout the whole project, Bob van Tiel for the experimental materials and Zsolt Veraszto for technical help.

Eszter Ronai - ronai@uchicago.edu
Ming Xiang - mxiang@uchicago.edu


Experiment 2: QUD manipulation

Reaction time results

Why are the EXH results less clear?

Unexpected cost for responding NG to Control may be a “side-effect” of the picture stimuli.

- Two-step verification process for Bad Control - something is indeed blue, but not the correct thing. In SI, nothing is blue.
- Evidence: rate of “Good” responses to Bad Control is higher in EXH than SI.

Problem with “fixing” this: existential presupposition - there *is* something that is blue.