First DETEC was great. Really looking forward to this, also to seeing what progress we've made. I'll be telling you about the progress I've made since 2013.
- Best evidence for radical pragmatics. Also best evidence against.
- Successfully communicates. But would have been better with right words.
- [In story, was prank phrasebook. Only works because words have meaning.]
- Like many of us, I’m interested in that interplay of words, syntax, and discourse context.
Rich, nuanced effects of context.
rich, nuanced effects of context

‘he’ = ?

A  A frightened B. He carried a knife.
B  A feared B. He carried a knife.
B  A frightened B, so he carried a knife.
A  A feared B, so he carried a knife.
A  A frightened B, so he carried a knife in order to heighten the effect.
B  A feared B, so he carried a knife in order to heighten the effect.
Pronouns
(he)

definite noun phrases
(the dog)

homophones
(is your refrigerator running?)

proper names
(George Washington)
(David Morse)

- Pronouns important in own right, but also serve as lesson.
Established Phenomena

Verb Bias (implicit causality, etc.)
Verb Bias (implicit causality, etc.)

A frightened B because she...
A feared B because she...

Empirical Log-Odds of Choosing Subject

(Brown & Fish, 1983; Ferstl, Garnham, & Manouilidou, 2011; Garvey & Caramazza, 1974; Hartshorne, O’Donnell, & Tenenbaum, 2015; Hartshorne & Shedeke, 2012; Rudolph & Forsterling, 1997)
Established Phenomena

Verb Bias (implicit causality, etc.)
Connectives

‘he’ = ?

A  A frightened B. He carried a knife.
B  A feared B. He carried a knife.
B  A frightened B, so he carried a knife.
A  A feared B, so he carried a knife.

A number of theories try to account for people’s expectations about pronoun reference.

- Centering theory: pronouns preferentially refer to the backward-looking center, a sort of “topic”.
- Learned heuristics: people learn from the input that certain features of the context — verbs, connectives, etc. — predict certain patterns of pronoun reference.
- Expectancy: People learn features of the context that predict what will be mentioned next. They assume that if a pronoun is used, it will refer to that thing.
- to some extent centering theory & expectancy — but definitely for learned heuristics — the patterns of verb bias and connective bias are unmotivated.
- BDM: Andy Kehler’s Bayesian Discourse Model tries to derive them from first principles. He draws on earlier work by Jerry Hobbs, which posts that two statements X & Y in a discourse can be related to one another by specific types of relations.
<table>
<thead>
<tr>
<th>Example</th>
<th>Relations between X &amp; Y</th>
</tr>
</thead>
<tbody>
<tr>
<td>A frightened B because she…</td>
<td>Explanation (“X because Y”)</td>
</tr>
<tr>
<td>A feared B because she…</td>
<td>Event Y provides explanation for event X</td>
</tr>
<tr>
<td></td>
<td>Subj(Y) &amp; cause(X) co-refer</td>
</tr>
<tr>
<td>Because A frightened B, she…</td>
<td>Result (“Because X, Y”)</td>
</tr>
<tr>
<td>Because A feared B, she…</td>
<td>Event X has consequence Y</td>
</tr>
<tr>
<td></td>
<td>Patient(X) &amp; Subj(Y) co-refer</td>
</tr>
<tr>
<td>A frightened B, and then she…</td>
<td>Sequence (“X then Y”)</td>
</tr>
<tr>
<td>A feared B, and then she…</td>
<td>Event Y happens after event X</td>
</tr>
<tr>
<td></td>
<td>Topic(X) &amp; Topic(Y) co-refer</td>
</tr>
</tbody>
</table>

(Kehler, 2002; Kehler et al., 2008; Kehler & Rohde, 2013)

Take implicit causality.
A frightened B because she...
A feared B because she...
Because A frightened B, she...
Because A feared B, she...

(Hartshorne, O’Donnell, & Tenenbaum, 2015)
**Established Phenomena**

- Verb Bias (implicit causality, etc.)
- Connectives

**Theories**

- **Centering Theory** (Grosz, Joshi, & Weinstein, 1995)
- **Learned Heuristics** (van Berkum, various)
- **Expectancy** (Arnold, various)
- **Bayesian Discourse Model** (Kehler, 2002; Kehler & Rohde, 2013)

**Limitation:** Only account for expectations
Theories

- Centering Theory (Grosz, Joshi, & Weinstein, 1995)
- Learned Heuristics (van Berkum, various)
- Expectancy (Arnold, various)
- Bayesian Discourse Model (Kehler, 2002; Kehler & Rohde, 2013)

Limitation: Only account for expectations

- Minimally, big hole in theory
- Worse: Ultimately, these theories only account for expectations. Actual interpretation will require something else. Seems like a lot of work to go to, for something that is ultimately superfluous. And that probably make us wonder if we're on the right track.
Theories

Centering Theory (Grosz, Joshi, & Weinstein, 1995)
Learned Heuristics (van Berkum, various)
Expectancy (Arnold, various)
Bayesian Discourse Model (Kehler, 2002; Kehler & Rohde, 2013)

Limitation: Only account for expectations

- A frightened B because she...
- A frightened B because she was easily frightened.

- Minimally, big hole in theory
- Worse: Ultimately, these theories only account for expectations. Actual interpretation will require something else. Seems like a lot of work to go to, for something that is ultimately superfluous. And that probably make us wonder if we’re on the right track.
Outline

- Basics of “new” theory
- Account for
  - Ultimate interpretation
  - Discourse expectations
- Open Qs / Future directions

Not Included
- Choice of referential form

- “new” - because lots of influences: Kehler & Rohde, Arnold, Brown & Fish, Grice, Relevance, RSA
- Will try to shout out as I go, but for sake of time, may miss connections to prior work you feel are important. For which I apologize.
- Open Qs: Won’t solve everything. Sorry.
- Choice of ref: Have some thoughts, but no data so not in talk.
How do we know the speaker’s knowledge or goals? Well, sometimes, they’ve told us.
Can also use our prior expectations.
- Cooperative principle. Didn’t observe the fall of the Roman Empire personally.
- Also, you have some direct observations of world.
There are no doubt other things that should be added, but this is all I need. Also, some of these boxes are probably more complex. So maybe linguistic system
Also, the speaker’s conversational goals are probably complex. Particularly if they are engaging in recursive Gricean reasoning a la RSA. Really hard to draw. But, again, these details aren’t going to matter much for present purposes. So suppressed.
Crucial: Most of this structure is latent.
- Observed: The utterance, prior testimony, your own experience. Rest must be inferred — or, in case of priors, possibly innate.
utterance: A beat B at tug-of-war because he is strong.

message: cause(strong(A), beat(A,B))
cause(strong(B), beat(A,B))
cause(strong(C), beat(A,B))
cause(strong(D), beat(A,B))
...

\[
P(\text{message}|\text{utterance}) \propto P(\text{utterance}|\text{message}) \cdot P(\text{message})
\]

- Except, we actually have more than 2 variables!
- [PAUSE] So let's call all those other variables “V” for the moment.
utterance: A beat B at tug-of-war because he is strong.

message: cause(strong(A), beat(A,B))
cause(strong(B), beat(A,B))
cause(strong(C), beat(A,B))
cause(strong(D), beat(A,B))
...

$P(message|utterance, \mathbf{V}) = \prod P(message|utterance, \mathbf{V})$
- Don't know anything about this speaker. No prior testimony to worry about.
Likewise, don’t know anything about specific goals. So reduces to our priors about speakers in general.
Likewise, we know nothing specific about speaker’s knowledge or experience, so that falls away.
listener isn’t observing anything new about the world — other than this conversation — so we can ignore that part for now, too.
prior on speaker’s goals

Listener’s theory of speaker
- Why show all that other stuff?
- Again, it’s actually the same model. Beautiful aspect of inferential models like this. When you have no specific info about a variable, you can ignore it. But it’s actually still there to be used in other situations. That’s part of what makes inference over a generative model so powerful.
A beat B at tug-of-war because he is strong.

P(message|utterance, V) \propto P(utterance|message, V) \cdot P(message|V)

cause(strong(A), beat(A, B))
cause(strong(B), beat(A, B))
cause(strong(C), beat(A, B))
cause(strong(D), beat(A, B))
...

So we need to figure out the probability of the utterance, given the message, the linguistic system, and the prior on speaker’s goals. As far as syntax and semantics, our utterance is a good choice for any of these messages, right? It’s syntactically well-formed, and it’s semantically compatible. But speaker’s goals matter. If the speaker wants to be understood, it’s a bad idea to use a pronoun to refer to somebody new out the blue. Children are terrible at this. They’ll come home from daycare and tell you, “He is in trouble,” and you have absolutely no idea who is being talked about. It could be someone in a cartoon they watched. Let’s assume a competent speaker.
A beat B at tug-of-war because he is strong.

\[ P(\text{message}|\text{utterance}, V) \propto P(\text{utterance}|\text{message}, V) \cdot P(\text{message}|V) \]

| message                  | \( P(\text{utterance}|\text{message}, V) \) | \( P(\text{message}|V) \) |
|--------------------------|---------------------------------|------------------------|
| cause(strong(A), beat(A,B)) | likely                          | likely                |
| cause(strong(B), beat(A,B)) | likely                          | likely                |
| cause(strong(C), beat(A,B)) | unlikely                         | unlikely              |
| cause(strong(D), beat(A,B)) | unlikely                         | unlikely              |
| …                        |                                 |                       |

- You just wouldn’t say this sentence if you meant C being strong caused A to beat B, because your listener wouldn’t understand you. This is something that probably could be derived from a recursive pragmatics model like RSA, but to keep the math simple, I just hard-coded it.
- Probability of the message)
- depends on: prior on speaker’s goals & world. Let’s assume a cooperative speaker who tells the truth. Then all we need to know is what’s likely to be true about the world. So we built a model for that.
“Because” introduces an explanation. Explanations refer to causes. So we paraphrase “A beat B because A is strong” as
A beat B because he is strong.

“A being strong caused A to beat B.” Now, it’s an open question as to how exactly people reason about causation. Luckily, it has been shown that for the relatively simple scenarios we’re considering, a counter-factual analysis works pretty well. So we can paraphrase again.
A beat B because he is strong.

A beat B because A is strong
⇒ A strong caused A beat B.
⇒ If A had not been strong, A would not have beaten B.

[READ] So we need to decide how likely is it that this would be true of the world.
Here’s how we set up the model. First, we assume that people’s strengths are normally distributed, which is more or less right. Now we need definitions for strong and weak. While there are a number of proposals about how to define scalar adjectives, work by Dave Barner, Josh Tenenbaum, and others shows that for many purposes they make roughly the same predictions. So for convenience, we just used a very simple prototype semantics. There is some strength such that they closer someone is to that strength, the more likely you are to label them as strong. Similarly for weak and average. Implicitly, there are other labels here, too, like “very strong” and “very weak”, though we don’t explicitly model them. Computationally, this is equivalent to a soft threshold semantics. As I said before, for first-pass approximations, the exact semantics you choose won’t make a big difference.
Now, we make one further simplifying assumption, which is that the strongest player wins. We could have made this probabilistic, but it would only make the math harder without really changing the results.
Now, how do we determine the truth of “A being strong caused A to beat B”? 
First, we draw a strength for A conditioned on A being the kind of person you would call “strong”. Now, we know that A beat B, so we draw a B conditioned on B being weaker than A,
following our assumption that the stronger person always wins. Now we need to check the counterfactual: If A had not be strong, what would have happened? We draw a new strength for A — A’ — conditioned on A not being the sort of person you would call strong.
A being strong caused A to beat B.

Some of the time, A’ will still be stronger than B and A’ wins anyway.
A being strong caused A to beat B.

But more often than not, A’ will be weaker than B. We repeat this whole process a large number of times in order to calculate the probability of the counterfactual. What we’ll find, again, is that if A wasn’t strong, A would probably have not beaten B.
A beat B because he is strong.

likely

Which I hope you agree seems right.
A beat B because he is strong.

<table>
<thead>
<tr>
<th>A beat B because A is strong</th>
<th>A beat B because B is strong</th>
</tr>
</thead>
<tbody>
<tr>
<td>A strong caused A beat B.</td>
<td>B strong caused A beat B.</td>
</tr>
<tr>
<td>If A had not been strong, A would not have beaten B.</td>
<td>If B had not been strong, A would not have beaten B.</td>
</tr>
</tbody>
</table>

In contrast, “A beat B because B is strong” means “If B had not been strong, A would not have beaten B.” Hopefully you agree that this seems unlikely. And that’s something we can derive from the model.
This time, we draw B conditioned on B being called ‘strong’. A beat B, so A has to be stronger.
Now we do our counterfactual. What are the chances that if B wasn’t strong, A wouldn’t have beaten B?
This time, when we redraw B’ such that B’ isn’t strong, B’ almost always looses. So all else equal, B being strong would not cause A to beat B.
A beat B because he is strong.

A beat B because A is strong

= A strong caused A beat B.
= If A had not been strong, A would not have beaten B.

A beat B because B is strong

= B strong caused A beat B.
= If B had not been strong, A would not have beaten B.

likely unlikely

I don’t think I need to show you that the model predicts that whether or not C and D are strong has absolutely no effect on A beating B. So now we have the probabilities we need.
A beat B at tug-of-war because he is strong.

\[
P(\text{message|utterance, } \mathbf{V}) \propto P(\text{utterance|message, } \mathbf{V}) \cdot P(\text{message|V})
\]

| message | \(P(\text{utterance|message, } \mathbf{V})\) | \(P(\text{message|V})\) |
|---------|-----------------|-----------------|
| cause(strong(A), beat(A,B)) | likely | likely |
| cause(strong(B), beat(A,B)) | likely | unlikely |
| cause(strong(C), beat(A,B)) | unlikely | very unlikely |
| cause(strong(D), beat(A,B)) | unlikely | very unlikely |

For instance, because stronger people are more likely to win at tug-of-war, [CLICK] it is much more likely that A beat B because A is strong than that A beat B because B is strong. [CLICK] It’s just about impossible to think up a scenario in which A beat B because of the strength of someone who isn’t A or B.

[PAUSE] Now, we’ll combine those probabilities using Bayes’ Rule, which you may have noticed I’ve already sneakily set up.
Bayes’ Rule is the normatively correct way of combining probabilities. So the implication of using Bayes’ Rule is that we’re assuming that when the listener misinterprets the speaker, it’s because the listener’s beliefs about the speaker are incorrect or his beliefs about language are incorrect, and not that he’s just bad at probabilities.

So the most likely message is the first one, which hopefully matches your intuition.
A beat B because he is strong.
A beat B because he is weak.
A lost to B because he is strong.
A lost to B because he is weak.
A beat B although he is strong.
A beat B although he is weak.
A lost to B although he is strong.
A lost to B although he is weak.

So in addition to this sentence, we had its pair where “he is weak”. We made two more sentences by replacing “beat” with “lost to” [CLICK]. And we created four more by changing “because to “although”. [CLICK]
You can ask me later if you want to know how we defined ‘although’.
A beat B because he is strong.
A beat B although he is strong.
A lost to B because he is strong.
A lost to B although he is strong.
A beat B because he is weak.
A beat B although he is weak.
A lost to B because he is weak.
A lost to B although he is weak.

In addition to these eight sentence, we made eight more by adding the word “almost”
Ask me later if you are interested in how we formalized “almost”. It’s pretty straightforward. [PAUSE] So we asked both the model and a group of Turkers to decide who the pronoun referred to in each of our sentences.
And you can see that we got a very tight fit, with a correlation of 0.92. Concern: Really good model of talking about tug-of-war?
Listener’s theory of speaker

Problem: World model. Used next best thing
- AMT. Ask people about probability of different types of events.
40 pairs of sentences:

*Agnes frightened Beatrice because she is reckless.*
\[
P(A \text{ frightened } B \mid A \text{ reckless}) > P(A \text{ frightened } B \mid B \text{ reckless})
\]

*Agnes frightened Beatrice because she is timid.*
\[
P(A \text{ frightened } B \mid A \text{ timid}) < P(A \text{ frightened } B \mid B \text{ timid})
\]

- log-odds choosing subject
- Note: No free parameters here!
- not all sentences involve explanations

\[ r = .88, p < .001 \]
40 pairs of sentences:

Because Agnes frightened Beatrice she got in trouble.
P(A trouble | A frightened B) > P(B trouble | A frightened B)

Because Agnes frightened Beatrice she ran away.
P(A ran | A frightened B) < P(B ran | A frightened B)

\[ r = 0.73, p < 0.001 \]
So far, the characters in our stories are random people. There was no a priori reason to suppose that A is any stronger than B. But what if that wasn’t the case.
What if it was “Wimpy beat Popeye at tug-of-war because he strong”. We know that Popeye is at least sometimes strong but Wimpy never is. So if we’re trying to identify the interpretation that results in a statement that’s like to be true
Wimpy beat Popeye at tug-of-war because he is strong.

\[ P(\text{Wimpy beat Popeye}|\text{Wimpy strong}) \]

Maybe we want to weight, for instance, the probability…
Wimpy beat Popeye at tug-of-war because he is strong.

\[ P(\text{Wimpy beat Popeye}|\text{Wimpy strong}) \times P(\text{Wimpy strong}) \]

- As judged by Turkers
- Identified 61 scenarios where this prior knowledge made big difference to model, and we asked whether it made a difference to people
Indeed, people thought that ‘he’ was more likely to refer to the subject of the first sentence when Popeye — the person who is likely to be strong — was the subject of that sentence than when he was the object. -Wait, there’s more!
Manipulating prior knowledge: God of the Weaklings

Ensures the weaker participant wins.

Al beat Bart at tug-of-war because he is strong.

<table>
<thead>
<tr>
<th>god of weaklings?</th>
<th>he = Al?</th>
</tr>
</thead>
<tbody>
<tr>
<td>no</td>
<td>0.978</td>
</tr>
<tr>
<td>yes</td>
<td>0.190</td>
</tr>
</tbody>
</table>

Adynamo

![Graph showing the relationship between probability of prior knowledge and participant's decision.](image)

normal condition

prob pro = subj

god condition

prob pro = subj
Manipulating prior knowledge: God of the Weaklings

Ensures the weaker participant wins.

<table>
<thead>
<tr>
<th>god of weaklings?</th>
<th>he = Al?</th>
</tr>
</thead>
<tbody>
<tr>
<td>no</td>
<td>0.978</td>
</tr>
<tr>
<td>yes</td>
<td>0.190</td>
</tr>
</tbody>
</table>

Al beat Bart at tug-of-war because he is strong.

Adynamo

![Graph showing the relationship between probability of perception and condition.](image)
Manipulating prior knowledge:
God of the Weaklings

Ensures the weaker participant wins.

God of weaklings?  
he = Al?  

Al beat Bart at tug-of-war  
because he is strong  
no  
yes  
0.978  
0.190

Adynamo
- “new” - because lots of influences: Kehler & Rohde, Arnold, Brown & Fish, Grice, Relevance, RSA-
  - Will try to shout out as I go, but for sake of time, may miss connections to prior work you feel are important. For which I apologize.
  - Open Qs: Won’t solve everything. Sorry.
  - Choice of ref: Have some thoughts, but no data so not in talk.
- “new” - because lots of influences: Kehler & Rohde, Arnold, Brown & Fish, Grice, Relevance, RSA-
  - Will try to shout out as I go, but for sake of time, may miss connections to prior work you feel are important. For which I apologize.
  - Open Qs: Won’t solve everything. Sorry.
  - Choice of ref: Have some thoughts, but no data so not in talk.
Ultimate Interpretation:

A frightened B because she carried a knife.
A feared B because she carried a knife.

Discourse expectations:

A frightened B because she...
A feared B because she...
Ultimate Interpretation:

A frightened B because she carried a knife.
A feared B because she carried a knife.

\[ P(m=cause(carry(Agnes,knife), frighten(Agnes, Beatrice)),
    u="Agnes frightened Beatrice because she carried a knife") \]

Discourse expectations:

A frightened B because she...
A feared B because she...

\[ P(m=cause(X, frighten(Agnes, Beatrice)),
    u="Agnes frightened Beatrice because she...") \]
Ultimate Interpretation:

A frightened B because she carried a knife.
A feared B because she carried a knife.

\[ P(m=\text{cause(carry(Agnes, knife), frighten(Agnes, Beatrice))}, \]
\[ u=\text{"Agnes frightened Beatrice because she carried a knife"} \]

Discourse expectations:

A frightened B because she…
A feared B because she…

\[ \sum_{X} P(m=\text{cause}(X, \text{frighten(Agnes, Beatrice)}), \]
\[ u=\text{"Agnes frightened Beatrice because she…"} ) \]
Summing over Continuations
(implicit causality)

Agnes scared Beatrice because…
Agnes scared Beatrice because…

Agnes had an intimidating personality.
Agnes hid and jumped out.
Agnes had just killed someone.
… (27 more)

Beatrice didn’t see Agnes.
Beatrice went into the haunted house.
Beatrice was prejudiced.
… (27 more)
Summing over Continuations
(implicit causality)

Agnes scared Beatrice because…

Agnes had an intimidating personality.
Agnes hid and jumped out.
Agnes had just killed someone.
… (27 more)

Beatrice didn’t see Agnes.
Beatrice went into the haunted house.
Beatrice was prejudiced.
… (27 more)

60 continuations X 40 verbs
-prob of causing
-prior prob
What's role of verb class?
Epiphenomenon? Important, because related to syntax.
Summing over Continuations
(implicit causality)

Agnes scared Beatrice because...

Agnes had an intimidating personality.
Agnes hid and jumped out.
Agnes had just killed someone.
... (27 more)

Beatrice didn’t see Agnes.
Beatrice went into the haunted house.
Beatrice was prejudiced.
... (27 more)

This looks continuous. But…
Agnes scared Beatrice because…

Agnes had an intimidating personality.
Agnes hid and jumped out.
Agnes had just killed someone.
… (27 more)

Beatrice didn’t see Agnes.
Beatrice went into the haunted house.
Beatrice was prejudiced.
… (27 more)

- No evidence of model predicting judgments within class, though only 10 verbs/class.
- Also, tested youngest kids (5yo) who know a lot of fear & frighten-type verbs. Same results as adults. Children have less experience with world.
Evidence for Role of Verb Class

1. Little evidence of within-class variation
2. Young children show adult-like biases
3. Verb class effects with novel verbs

(Hartshorne, Skorb, & Snedeker, in prep)
- “new” - because lots of influences: Kehler & Rohde, Arnold, Brown & Fish, Grice, Relevance, RSA-
  - Will try to shout out as I go, but for sake of time, may miss connections to prior work you feel are important. For which I apologize.
  - Open Qs: Won’t solve everything. Sorry.
  - Choice of ref: Have some thoughts, but no data so not in talk.
Outline

- Basics of “new” theory
- Account for
  - Ultimate interpretation
  - Discourse expectations
- Open Qs / Future directions

- That’s my take-home
- Remind you of the evidence
- ultimate interpretation
- Depends on nuanced beliefs about world, can update quickly
- ultimate interpretation
- Depends on nuanced beliefs about world, can update quickly
Agnes scared Beatrice because...

- Agnes had an intimidating personality.
- Agnes hid and jumped out.
- Agnes had just killed someone.
  ... (27 more)

Beatrice didn't see Agnes.

Beatrice went into the haunted house.

Beatrice was prejudiced.
  ... (27 more)
Open Qs / Future directions

- Discourse relations
- Implement rest of model
  - linguistic system, speaker's goals, speaker's history
- Recursive pragmatic reasoning?
- Production biases?
- Computability

LAST SLIDE!