Conversations Gone Awry:

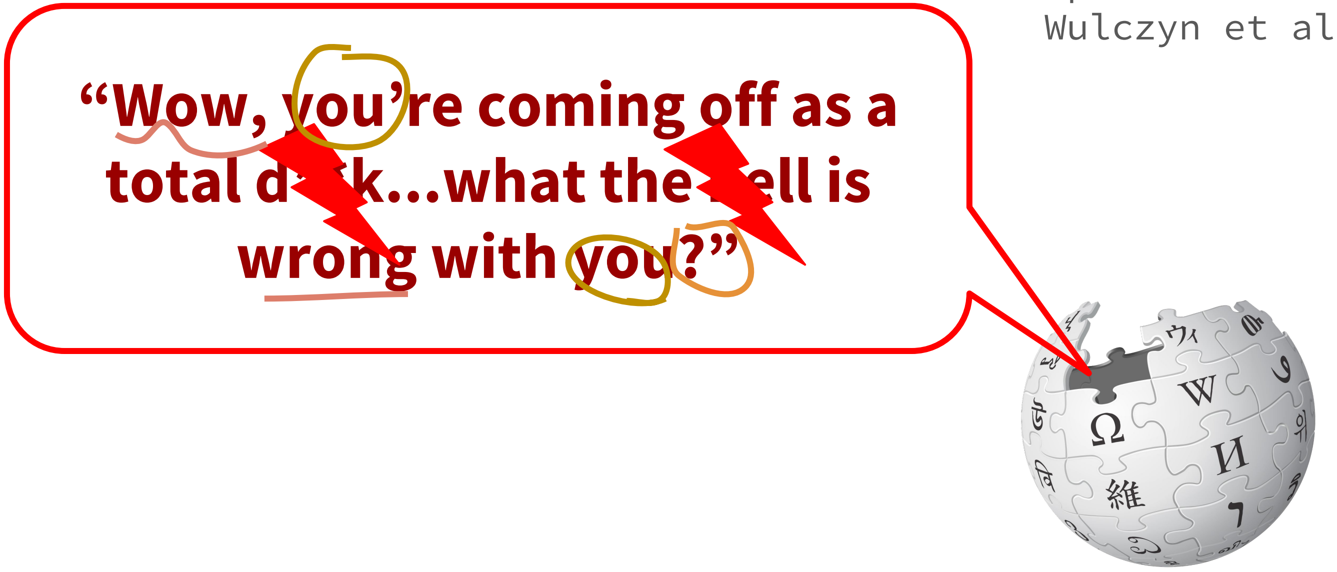
Detecting Early Signs of Conversational Failure

Justine Zhang, Jonathan P. Chang, Cristian Danescu-Niculescu-Mizil,
Lucas Dixon, Yiqing Hua, Dario Taraborelli, Nithum Thain

Cornell University, Jigsaw, Wikimedia Foundation

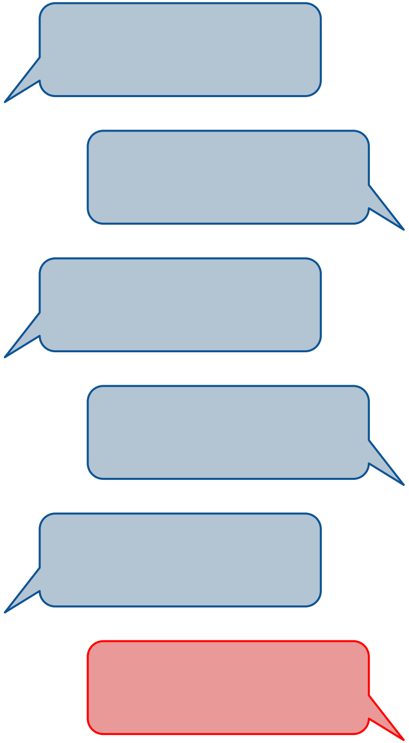
Motivating problem: toxic behaviour

Yin et al., 2009
Sood et al., 2012
Nobata et al., 2016
Cheng et al., 2017
Gambäck & Sikdar, 2017
Pavlopoulos et al., 2017
Wulczyn et al., 2017



"Wow, you're coming off as a total d[illegible]k...what the [illegible]ell is wrong with you?"

Conversations can start civil...



... but later **derail**

Our goal: detecting early signals of eventual derailment

Testing intuitions of early signals: a guessing game

“Wow, you’re coming off as a total d**k...what the hell is wrong with you?”

A

Is the St. Petersberg Times considered a reliable source by Wikipedia? I'm going to maybe do a rewrite of the article. I don't think the bulk of this article should rely on this one so-so source, which seems to speculate about UFOs.

Why's there no mention of it here? Namely, an altercation with a foreign intelligence group? True, by *some* standards the source has some weak points, but that doesn't mean it shouldn't exist.

B

I would assume that it's as reliable as any other mainstream news source.

So what you're saying is we should put a bad source in the article because it exists?

[...]

[...]

"Wow, you're coming off as a total d**k...what the hell is wrong with you?"

A

Is the St. Petersberg Times considered a reliable source by Wikipedia? I'm going to maybe do a rewrite of the article. I don't think the bulk of this article should rely on this one so-so source, which seems to speculate abo

I would assume that it's as r as any other mainstream ne source.

[...]

B

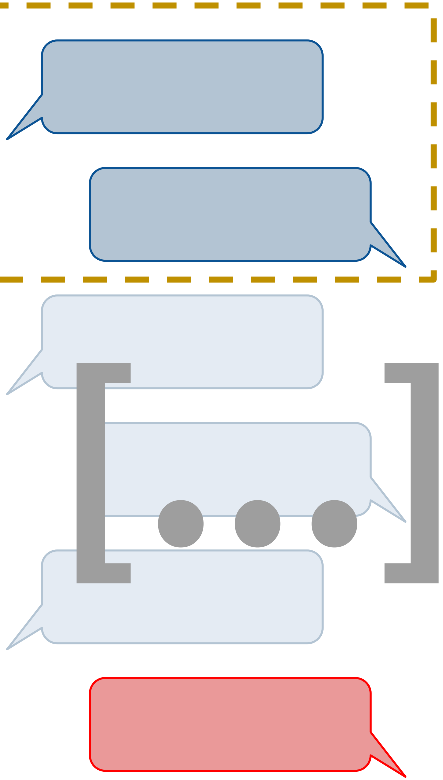
Why's there no mention of it here? Namely, an altercation with a foreign intelligence group? True, by *some* standards the source has some weak points, but that doesn't mean it ist.

at you're saying is we should put a bad source in the article because it exists?

[...]

Try more examples at
awry.infosci.cornell.edu

"Wow, you're coming off as a total d**k...what the hell is wrong with you?"



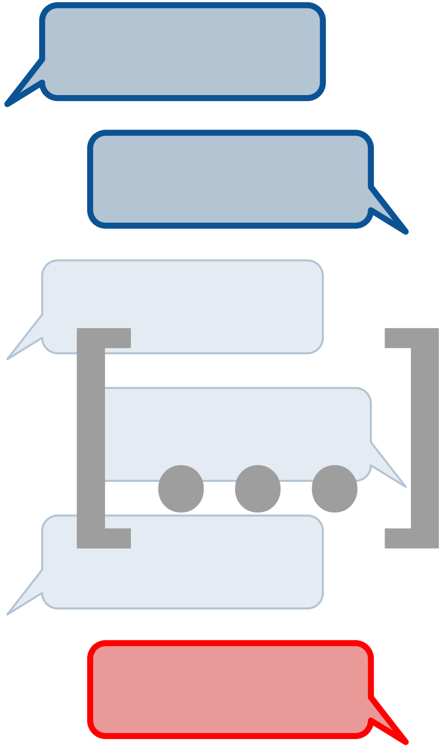
Other humans:

72% Accuracy

Can we reconstruct some of this intuition?

Identifying awry-turning conversations

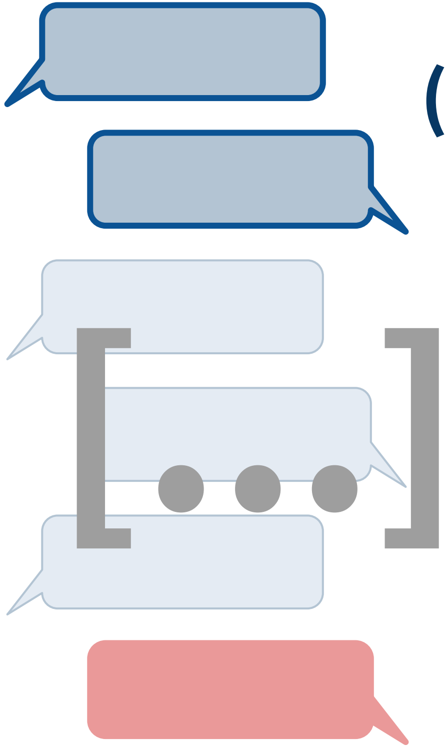
Civil Start:



← *Toxic Comment*

Identifying awry-turning conversations

Civil Start:



(no *toxic* behaviour)

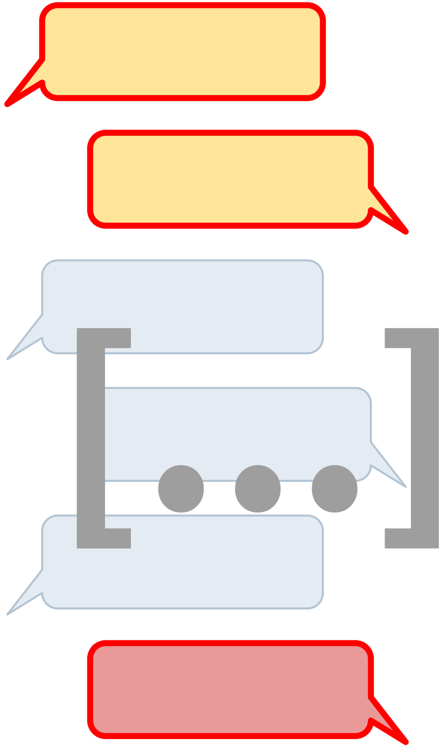
rude, insulting, or disrespectful *towards a person/group or that person/group's actions, comments, or work*

Wulczyn et al., 2017

← ***Toxic Comment***

Identifying awry-turning conversations

Civil Start:

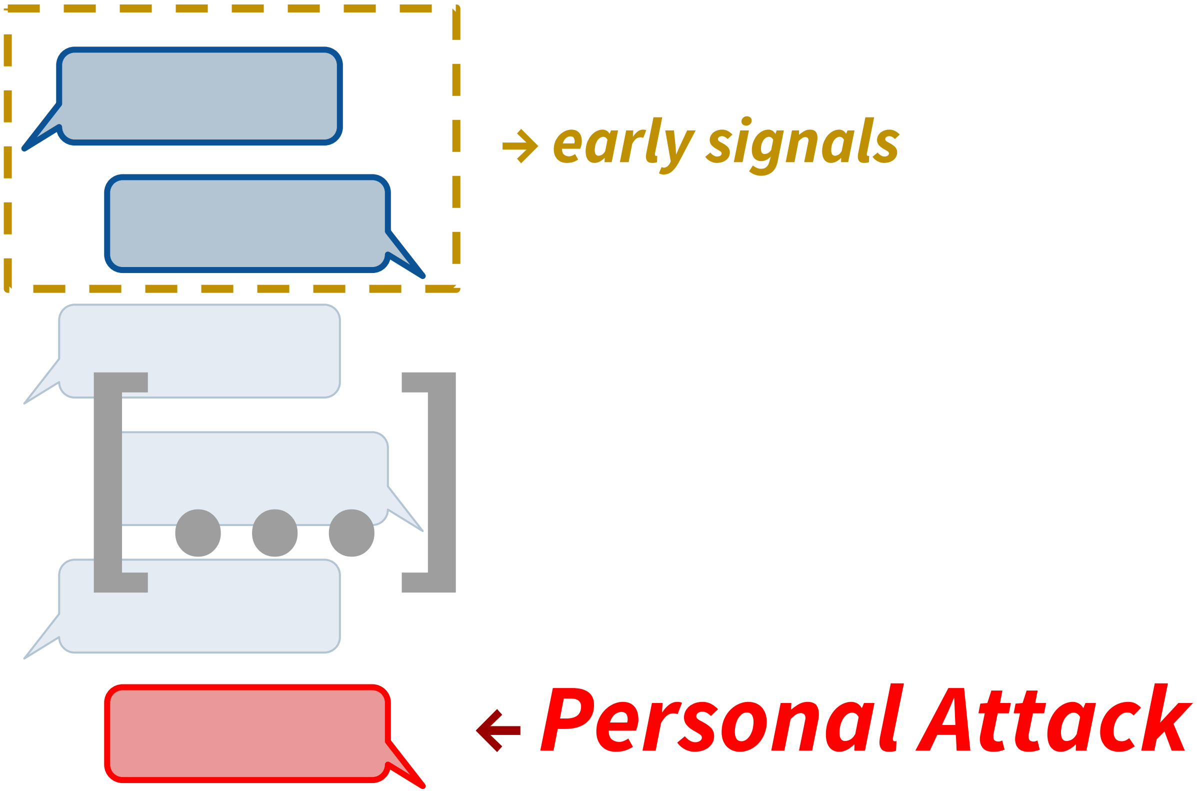


← ***Personal Attack***

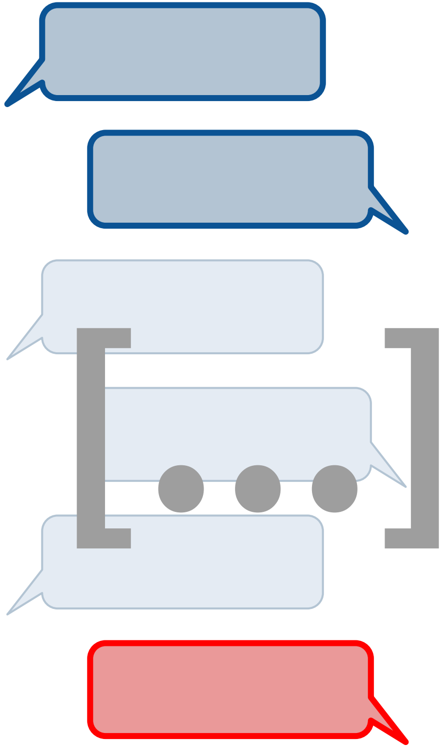
(one of the initial commenters eventually attacks the other)

Arazy et al., 2013

Civil Start:



Civil Start:

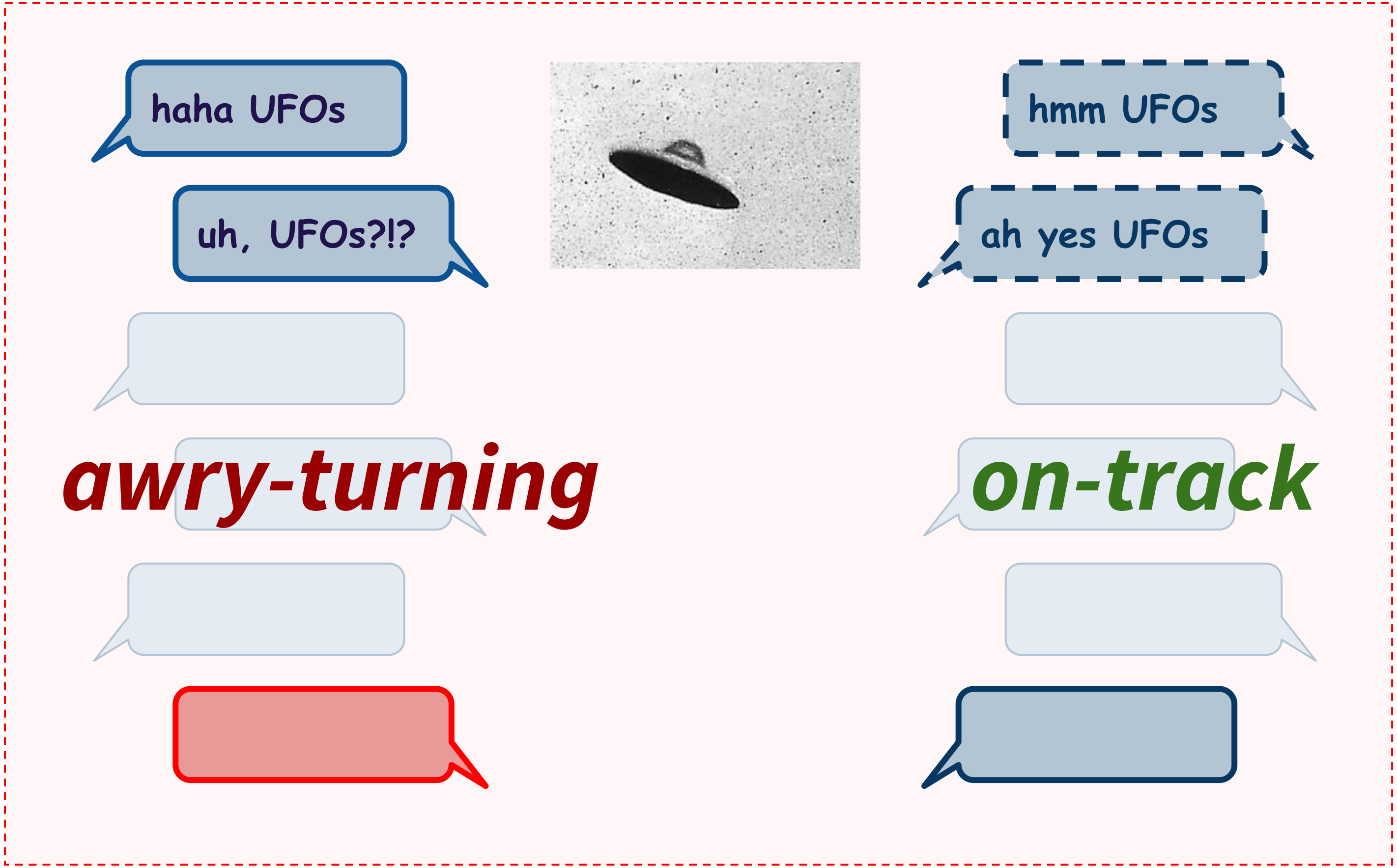


← ***Personal Attack***

Coser, 1956
De Dreu & Weingart, 2003
Galley et al., 2004
Andreas et al., 2012
Hillard et al., 2012
Allen et al., 2014
Wang & Cardie, 2014
Rosenthal & McKeown, 2015
inter alia

(different from disagreement)

controlling for context:



haha UFOs

uh, UFOs?!?

hmm UFOs

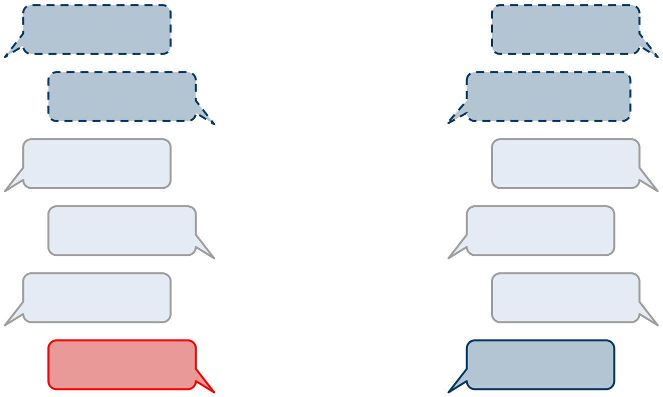
ah yes UFOs

awry-turning

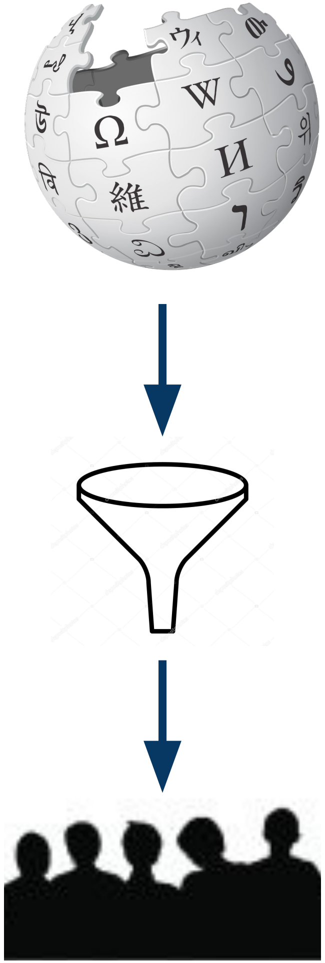
on-track

Data: Wikipedia Talk Pages

*(same **talk page**)*



635 pairs



(details in paper; data available!)

Detecting early signals: *(im)politeness*

A

Is the St. Petersberg Times considered a reliable source by Wikipedia? I'm going to maybe do a rewrite of the article. I don't think the bulk of this article should rely on this one so-so source, which seems to speculate about UFOs.

I would assume that it's as reliable as any other mainstream news source.

B

Why's there no mention of it here? Namely, an altercation with a foreign intelligence group? True, by *some* standards the source has some weak points, but that doesn't mean it shouldn't exist.

So what you're saying is we should put a bad source in the article because it exists?

direct questioning

hedging

Detecting early signals: *(im)politeness*

A

Is the St. Petersberg Times considered a reliable source by Wikipedia? I'm going to maybe do a rewrite of the article. I don't think the bulk of this article should rely on this one so-so source, which seems to speculate about UFOs.

I would assume that it's as reliable as any other mainstream news source.

B

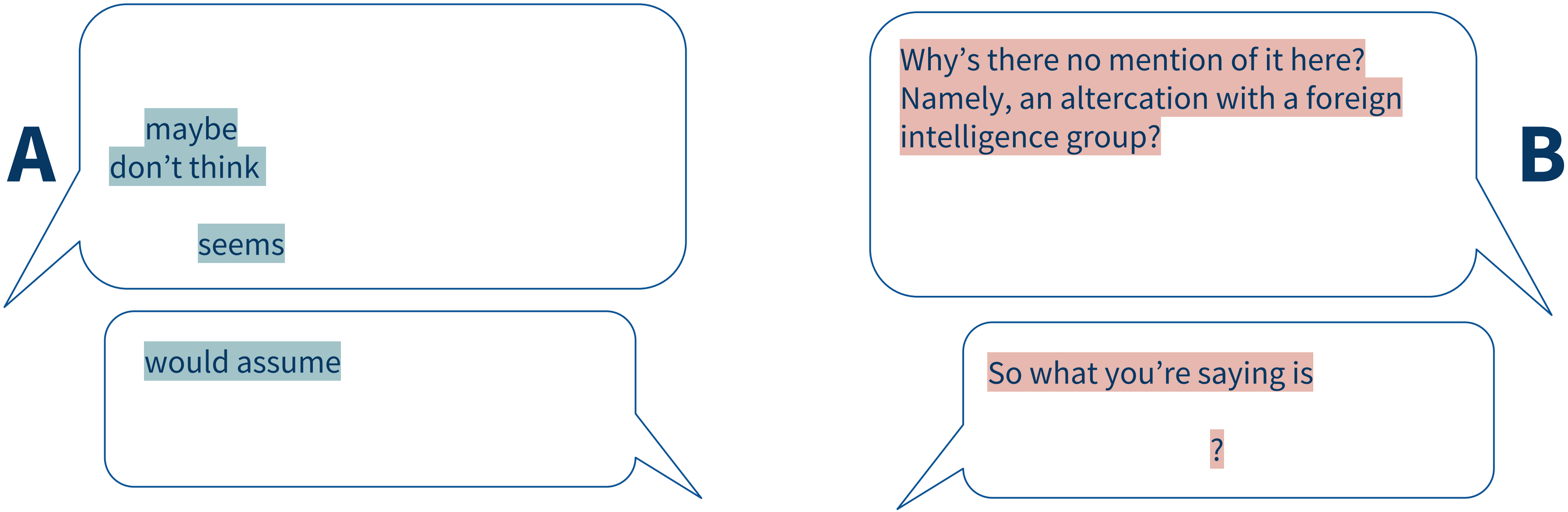
Why's there no mention of it here? Namely, an altercation with a foreign intelligence group? True, by *some* standards the source has some weak points, but that doesn't mean it shouldn't exist.

So what you're saying is we should put a bad source in the article because it exists?

"building rapport, softening potential conflicts"

Goffman, 1955
Fraser, 1980
Brown & Levinson, 1987

Detecting early signals: *(im)politeness*



A

maybe

don’t think

seems

would assume

B

Why’s there no mention of it here? Namely, an altercation with a foreign intelligence group?

So what you’re saying is

?

38 politeness strategies features

Goffman, 1955
Fraser, 1980
Brown & Levinson, 1987

D-N-M et al., 2013
politeness.cornell.edu

Detecting early signals

Is the St. Petersberg Times considered a reliable source by Wikipedia? I'm going to maybe do a rewrite of the article. I don't think the bulk of this article should rely on this one so-so source, which seems to speculate about UFOs.

I have been working on creating a new section…

Let me know if you agree with this…

Domain-specific intuition:

"let's coordinate"

Kittur & Kraut, 2008

Detecting early signals: *rhetorical intentions*

Is the St. Petersberg Times considered a reliable source by Wikipedia? I'm going to maybe do a rewrite of the article. I don't think the bulk of this article should rely on this one so-so source, which seems to speculate about UFOs.

I have been working on creating a new section…

Let me know if you agree with this…

Automatically inferring

"let's coordinate"

idea:

similar rhetorical intentions prompt similar responses

Zhang, Spirling & D-N-M, 2017

Detecting early signals: *rhetorical intentions*

Is the St. Petersberg Times considered a reliable source by Wikipedia? I'm going to maybe do a rewrite of the article. I don't think the bulk of this article should rely on this one so-so source, which seems to speculate about UFOs.

I have been working on creating a new section…

Let me know if you agree with this…

typical responses

Nice work, thanks for helping.

If you can do it I would appreciate it.

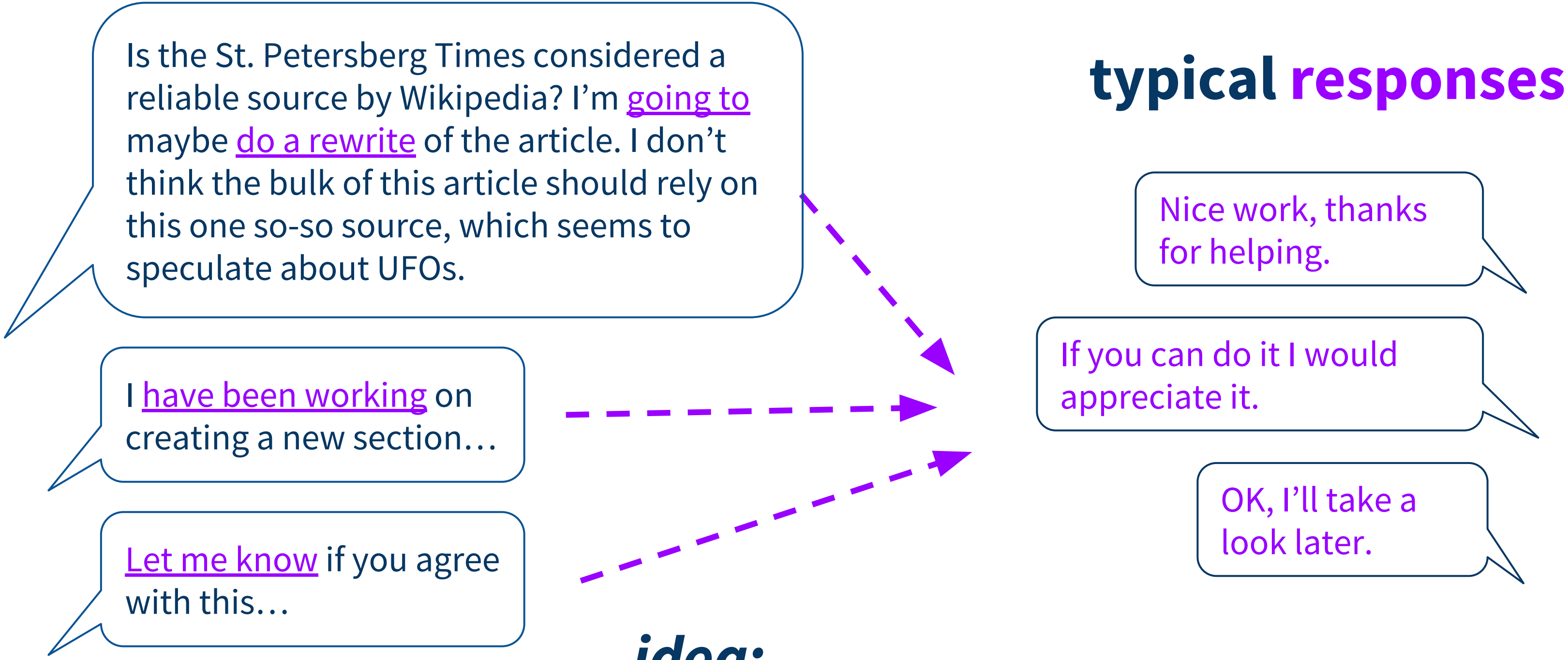
OK, I'll take a look later.

idea:

similar rhetorical intentions prompt similar responses

Zhang, Spirling & D-N-M, 2017

Detecting early signals: *rhetorical intentions*



Is the St. Petersberg Times considered a reliable source by Wikipedia? I'm going to maybe do a rewrite of the article. I don't think the bulk of this article should rely on this one so-so source, which seems to speculate about UFOs.

I have been working on creating a new section…

Let me know if you agree with this…

typical responses

Nice work, thanks for helping.

If you can do it I would appreciate it.

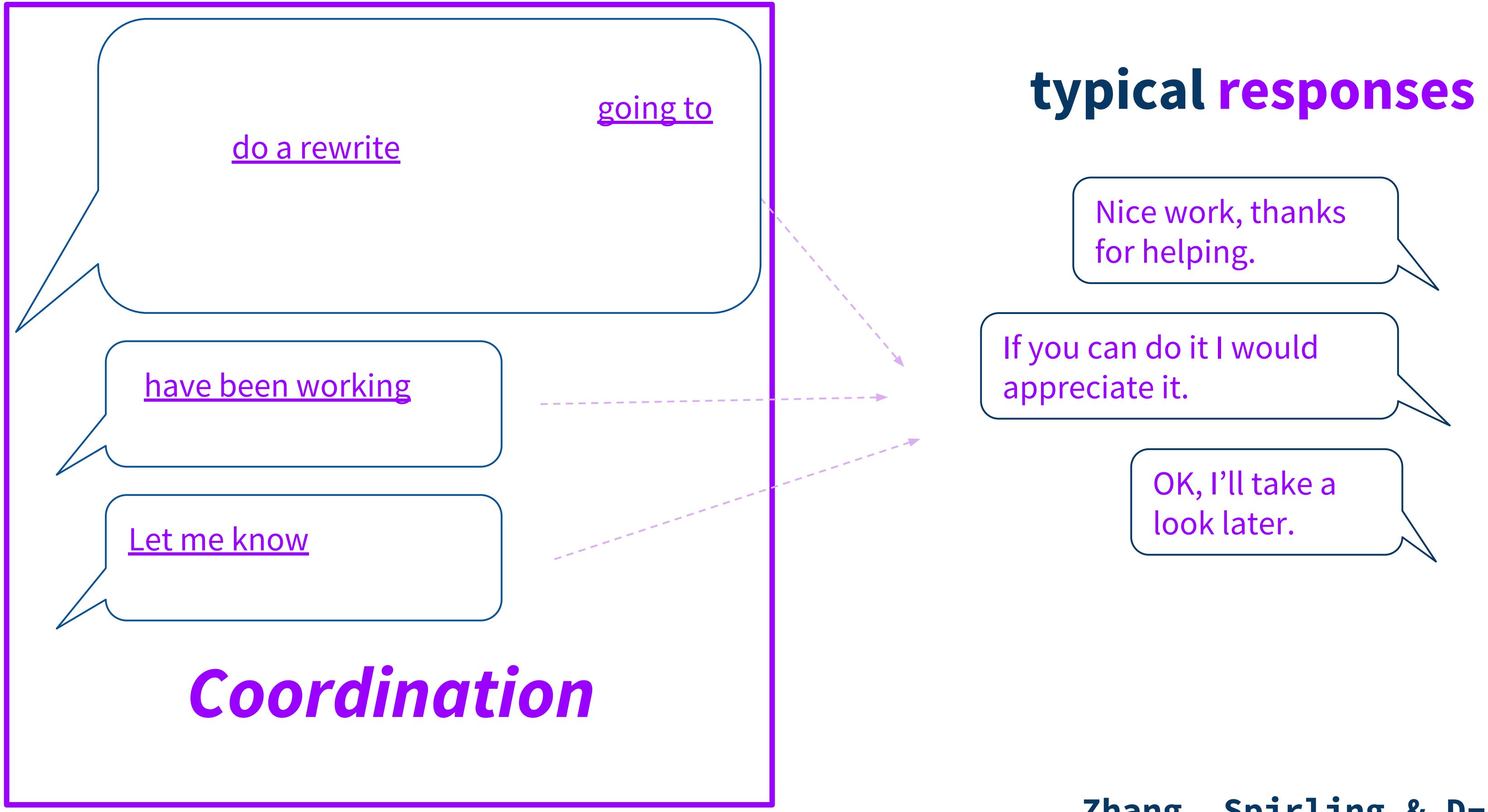
OK, I'll take a look later.

idea:

similar rhetorical intentions prompt
similar responses

Zhang, Spirling & D-N-M, 2017

Detecting early signals: *rhetorical intentions*



going to

do a rewrite

have been working

Let me know

Coordination

typical responses

Nice work, thanks for helping.

If you can do it I would appreciate it.

OK, I'll take a look later.

Zhang, Spirling & D-N-M, 2017

Detecting early signals: *rhetorical intentions*

Coordination

Moderation

Factual Check

Casual remark

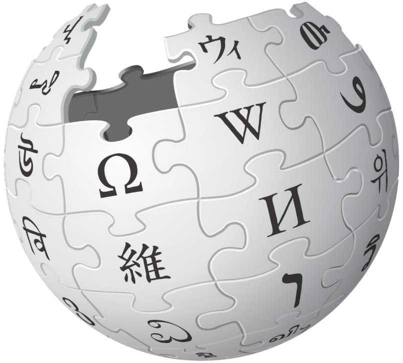
Action statement

Opinion

I'm going to maybe do a rewrite of the article...

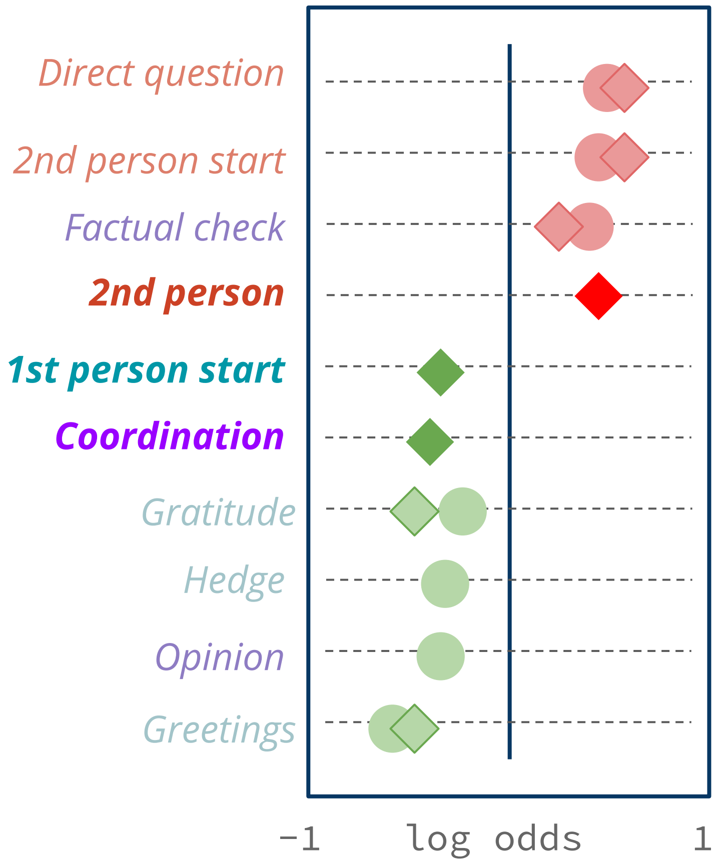
The census is not talking about families here.

It is hard to address both of these issues.



12 rhetorical intention features

Comparing awry-turning and on-track conversations



occurrence in

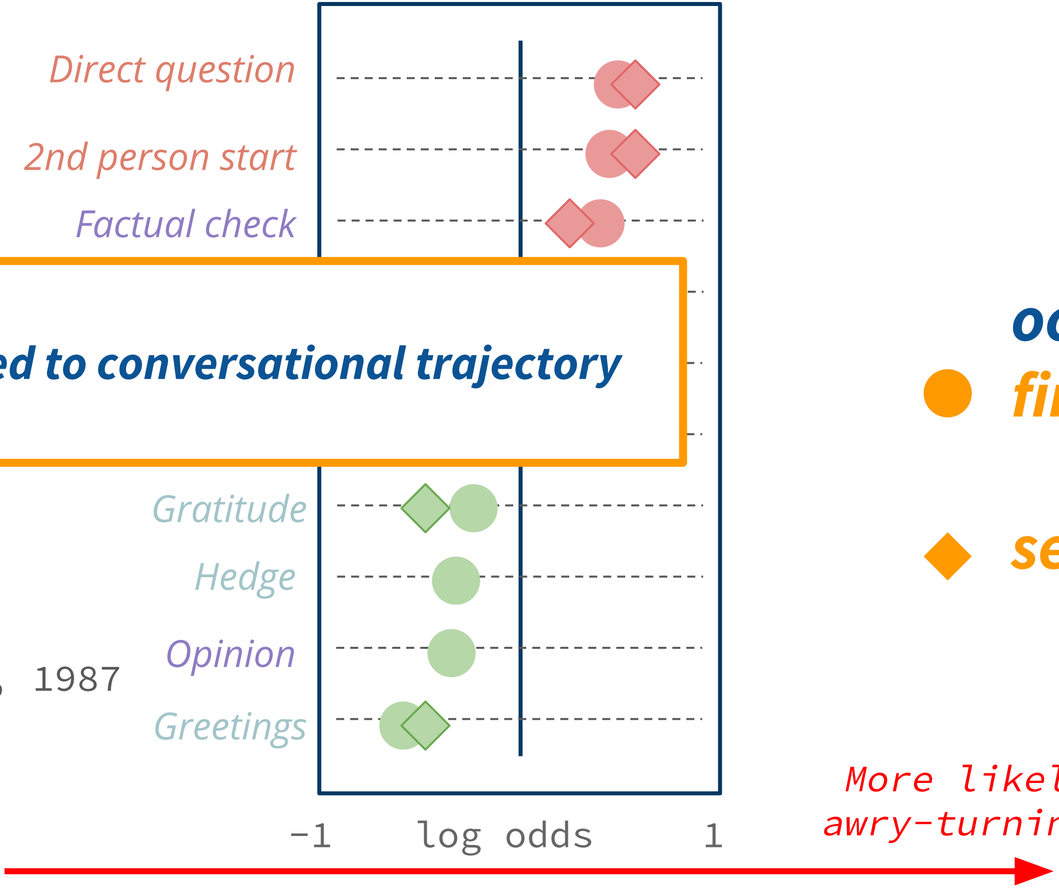
● ***first* comment**

◆ ***second* comment**

binomial $p < 0.05$

More likely to occur in awry-turning conversation

Comparing awry-turning and on-track conversations



politeness is tied to conversational trajectory

Goffman, 1955
Fraser, 1980
Brown & Levinson, 1987

occurrence in

● **first comment**

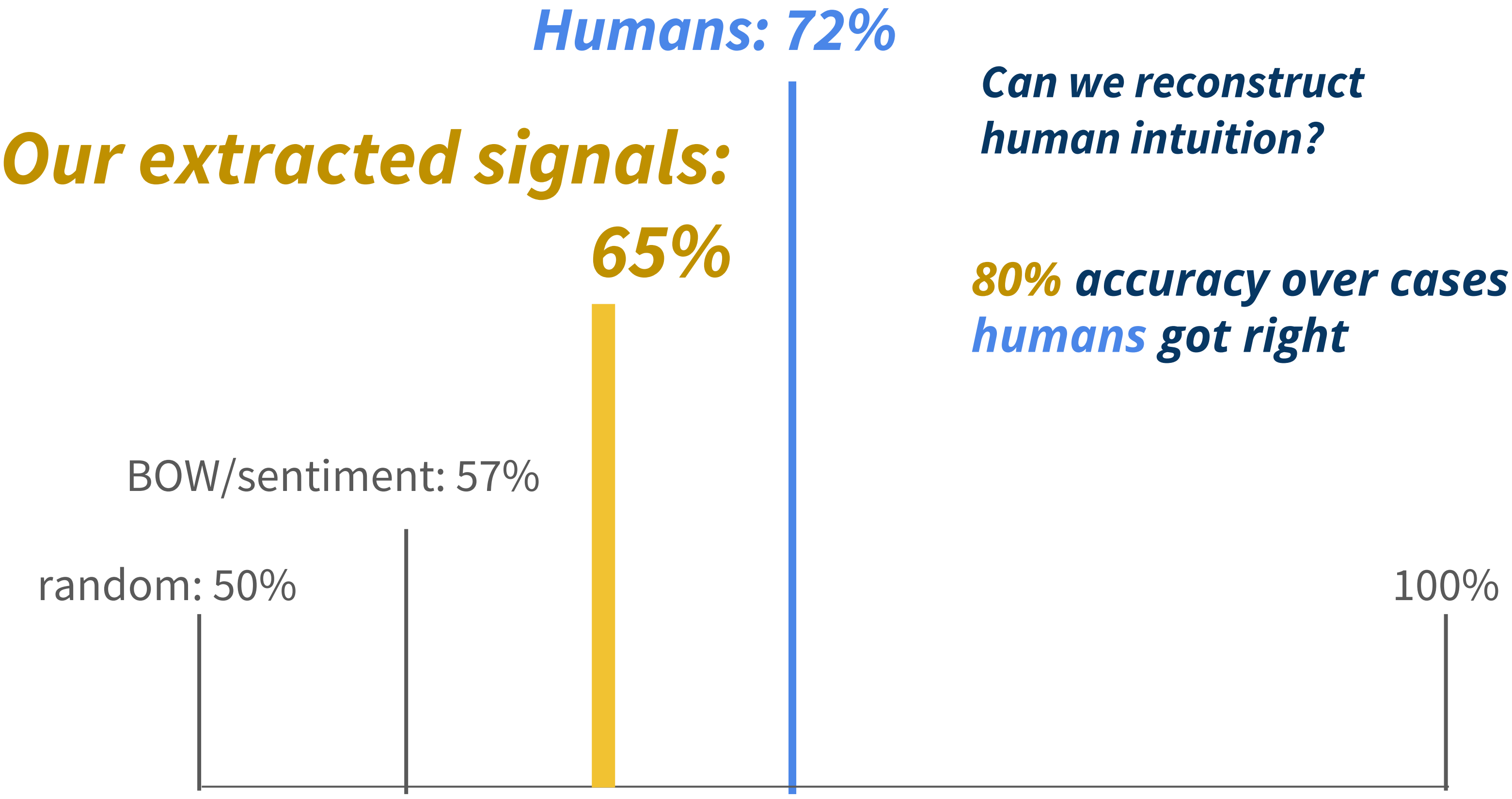
◆ **second comment**

More likely to occur in awry-turning conversation

Measuring the strength of our early signals

How well do our signals (politeness strategies, rhetorical intentions) perform at the guessing game?

Measuring the strength of our early signals

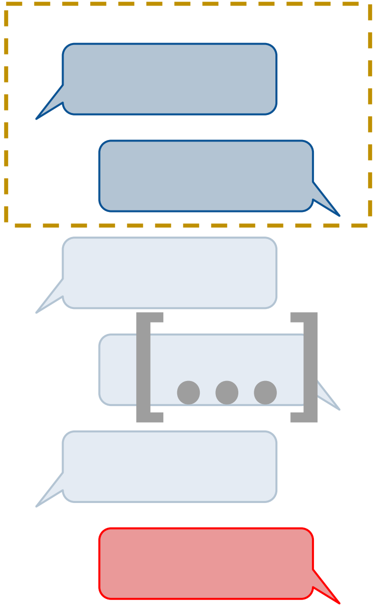


Can we reconstruct human intuition?

80% accuracy over cases that humans got right

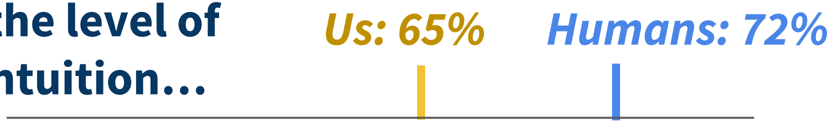
Conclusion

Conversations are dynamic:
They can start civil…
but later turn awry.

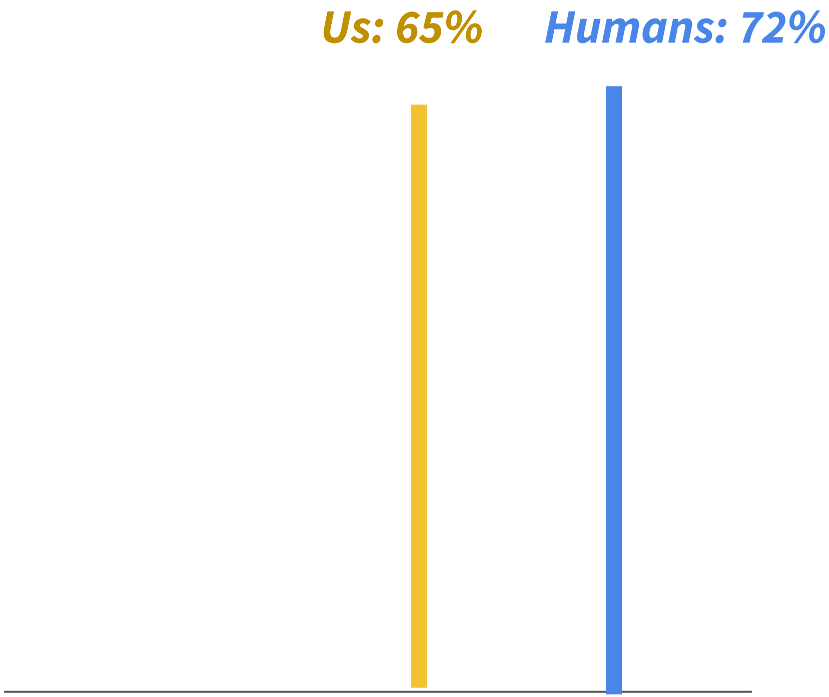


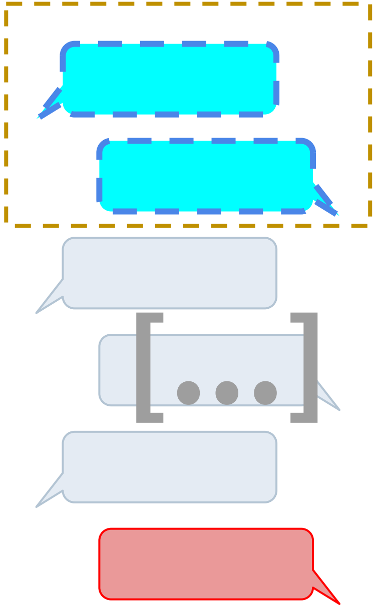
We detect early signals
(politeness strategies,
rhetorical intentions)
of eventual derailment,

***towards* the level of human intuition…**

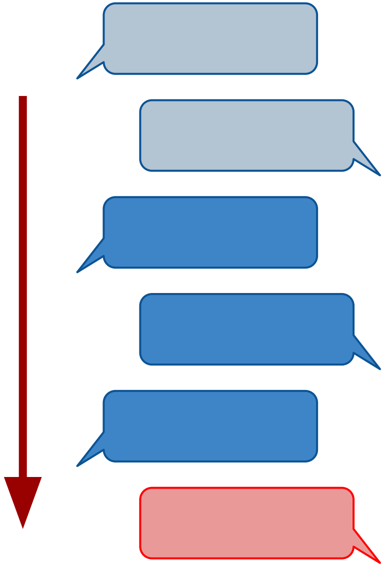


Future work: early signals we might have missed

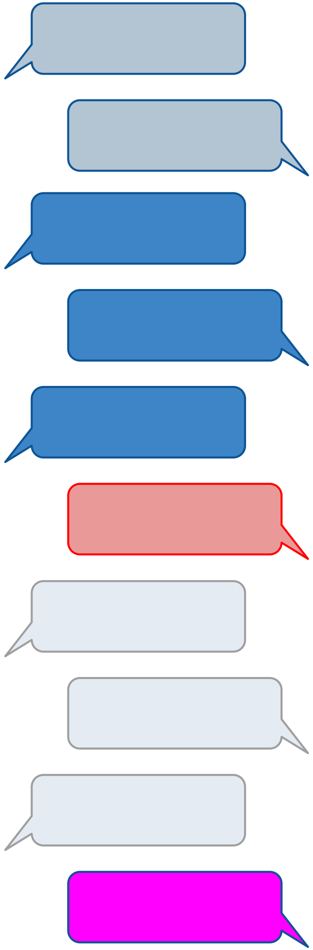




Future work: understanding the derailment *process*

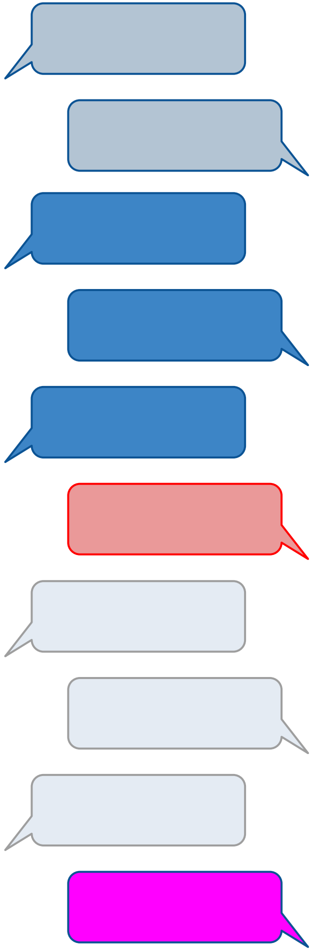


Future work: other trajectories



how might derailed conversations recover?

Questions?



Data and code:
Cornell Conversational Analysis Toolkit
convokit.infosci.cornell.edu

Online guessing game:
awry.infosci.cornell.edu