# In Search of True Contextuality in Natural Language

Quantum Contextuality in Quantum Mechanics and Beyond (QCQMB)

Daphne Wang (UCL) joint work with Mehrnoosh Sadrzadeh (UCL), Samson Abramsky (University of Oxford) and Víctor H. Cervantes (University of Illinois at Urbana-Champaign) 1. Introduction

Contextuality

Ambiguity in Natural Language

2. Contextuality in Natural language

Logical contextuality

Contextuality-by-Default contextuality

3. Beyond contextuality: analysis of signalling

4. Conclusion

Introduction

#### Sheaf-theoretic Contextuality

*Contextuality* can be seen as the impossibility to find a global section of the presheaf:

$$\mathcal{D}_{R}\mathcal{E}:\mathcal{P}(X)^{op}\to\mathsf{Set}$$
 (1)

which is compatible with observations.

#### Sheaf-theoretic Contextuality

*Contextuality* can be seen as the impossibility to find a global section of the presheaf:

$$\mathcal{D}_{R}\mathcal{E}:\mathcal{P}(X)^{op}\to\mathsf{Set}$$
 (1)

which is compatible with observations.

Logical/Possibilistic Contextuality

In some systems, only the support of the distributions is needed to show contextuality

#### Sheaf-theoretic Contextuality

*Contextuality* can be seen as the impossibility to find a global section of the presheaf:

$$\mathcal{D}_R \mathcal{E} : \mathcal{P}(X)^{op} \to \mathsf{Set}$$
 (1)

which is compatible with observations.

Logical/Possibilistic Contextuality

In some systems, only the support of the distributions is needed to show contextuality

Framework only works for no-signalling systems

А	В	(0,0)	(0,1)	(1,0)	(1,1)
а			1	1	1
а	b'	0	1	1	1
a'	b	0	1	1	1
a'	b'	1	1	1	0

• 0 • 1

•

b

• b'

• a

Α	В	(0,0)	(0,1)	(1,0)	(1,1)
а	b	1	1	1	1
а	b'	0	1	1	1
a'	b	0	1	1	1
a'	b'	1	1	1	0



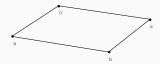
Α	В	(0,0)	(0,1)	(1,0)	(1,1)
а	b	1	1	1	1
а	b′	0	1	1	1
a'	b	0	1	1	1
a'	b'	1	1	1	0



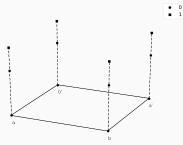
	4	В	(0,0)	(0,1)	(1,0)	(1,1)
	а	b	1	1	1	1
	а	b'	0	1	1	1
(	2'	b	0	1	1	1
(	יג	b'	1	1	1	0



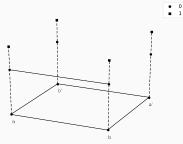
Α	В	(0,0)	(0,1)	(1,0)	(1,1)
а	b	1	1	1	1
а	b'	0	1	1	1
a'	b	0	1	1	1
a'	b'	1	1	1	0



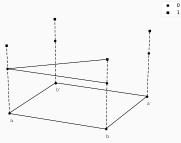
А	В	(0,0)	(0,1)	(1,0)	(1,1)
а			1	1	1
	b'	0	1	1	1
a'	b	0	1	1	1
a'	b'	1	1	1	0



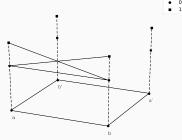
А	В	(0,0)	(0,1)	(1,0)	(1,1)
а	b	1	1	1	1
		0	1	1	1
a'	b	0	1	1	1
a'	b'	1	1	1	0



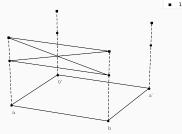
А	В	(0,0)	(0,1)	(1,0)	(1,1)
а	b	1	1	1	1
а	b'	0	1	1	1
a'	b	0	1	1	1
a'	b'	1	1	1	0



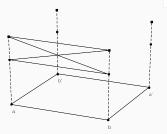
Α	В	(0,0)	(0,1)	(1,0)	(1,1)
а	b	1	1	1	1
а	b'	0	1	1	1
a'	b	0	1	1	1
a'	b'	1	1	1	0



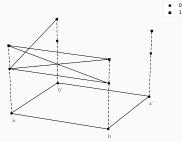
Α	В	(0,0)	(0,1)	(1,0)	(1,1)
а	b	1	1	1	1
а	b'	0	1	1	1
a'	b	0	1	1	1
a'	b'	1	1	1	0



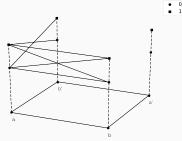
Α	В	(0,0)	(0,1)	(1,0)	(1,1)
а	b	1	1	1	1
а	b'	0	1	1	1
a'	b	0	1	1	1
a'	b'	1	1	1	0



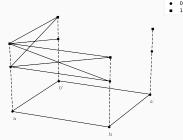
Α	В	(0,0)	(0,1)	(1,0)	(1,1)
а	b	1	1	1	1
а	b'	0	1	1	1
a'	b	0	1	1	1
a'	b'	1	1	1	0



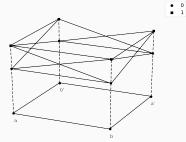
А	В	(0,0)	(0,1)	(1,0)	(1,1)
а	b	1	1	1	1
а	b'	0	1	1	1
a'	b	0	1	1	1
a'	b'	1	1	1	0



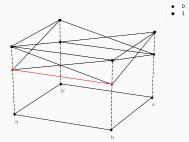
Α	В	(0,0)	(0,1)	(1,0)	(1,1)
а	b	1	1	1	1
а	b'	0	1	1	1
a'	b	0	1	1	1
a'	b'	1	1	1	0



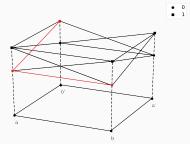
Α	В	(0,0)	(0,1)	(1,0)	(1,1)
а	b	1	1	1	1
а	b'	0	1	1	1
a'	b	0	1	1	1
a'	b'	1	1	1	0



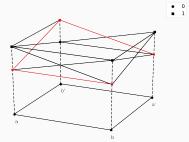
Α	В	(0,0)	(0,1)	(1,0)	(1,1)
а	b	1	1	1	1
а	b'	0	1	1	1
a'	b	0	1	1	1
a'	b'	1	1	1	0



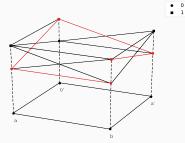
Α	В	(0,0)	(0,1)	(1,0)	(1,1)
а	b	1	1	1	1
а	b'	0	1	1	1
a'	b	0	1	1	1
a'	b'	1	1	1	0



Α	В	(0,0)	(0,1)	(1,0)	(1,1)
а	b	1	1	1	1
а	b'	0	1	1	1
a'	b	0	1	1	1
a'	b'	1	1	1	0



Α	В	(0,0)	(0,1)	(1,0)	(1,1)
а	b	1	1	1	1
а	b'	0	1	1	1
a'	b	0	1	1	1
a'	b'	1	1	1	0



• Measurements between contexts are fundamentally unrelated

- Measurements between contexts are fundamentally unrelated
  - Maximise the probability of agreement between contents ("local measurement")

- Measurements between contexts are fundamentally unrelated
  - Maximise the probability of agreement between contents ("local measurement")
  - Allows systems to be signalling

- Measurements between contexts are fundamentally unrelated
  - Maximise the probability of agreement between contents ("local measurement")
  - Allows systems to be signalling
- More general definition of *contexts*

#### Cyclic systems

#### Definition

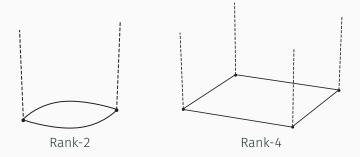
A system is a *cyclic system* if every content if binary, if every context contains exactly 2 contents, and every content is contained in exactly 2 contexts.

#### Contextuality-by-Default

## Cyclic systems

#### Definition

A system is a *cyclic system* if every content if binary, if every context contains exactly 2 contents, and every content is contained in exactly 2 contexts.



#### Cyclic systems

#### Definition

A system is a *cyclic system* if every content if binary, if every context contains exactly 2 contents, and every content is contained in exactly 2 contexts.

#### (Non)Contextuality criterion for cyclic systems

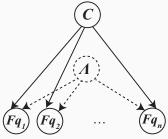
A cyclic system is non-contextual within the CbD framework iff:

$$s_{odd}\left(\left\langle R_{i}^{i}R_{i\oplus1}^{i}\right\rangle \right)_{i=1,...,n} \leq n-2+\Delta$$
 (1)

where:

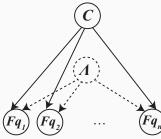
$$\Delta = \sum_{i=1}^{n} \left| \left\langle R_{i}^{i} \right\rangle - \left\langle R_{i}^{i \ominus 1} \right\rangle \right| \tag{2}$$

## Direct influences via M-Contextuality



Canonical model

## Direct influences via M-Contextuality



Canonical model

#### **Direct Influence**

Given a pair of contexts c, c' in a given canonical model, the *direct influence* on content variable  $F_q$  is quantified as:

$$\Delta_{c,c'}(F_q) = \Pr\left[\lambda|F_q(\lambda,c) \neq F_q(\lambda,c')\right]$$
<sup>(1)</sup>

### Direct influences via M-Contextuality

#### **M-Contextuality**

A system is contextual is there is no canonical model which simulteneously minimises all direct influences.

# Direct influences via M-Contextuality

#### **M-Contextuality**

A system is contextual is there is no canonical model which simulteneously minimises all direct influences.

#### M-Contextuality and CbD

- M-Contextuality is equivalent to CbD-contextuality
- If  $\Delta_{c_q,c_q'}^*(F_q)$  is the minimal direct influence across all canonical models:

$$\Delta = 2\sum_{q} \Delta_{c_q, c'_q}^* (F_q) \tag{1}$$

#### Lexical Ambiguity

- The beach was mined
- I saw a log
- They can fish

#### Lexical Ambiguity

- The beach was mined
- I <mark>saw</mark> a log
- They can fish

#### Syntactic Ambiguity

- I shot an elephant in my pyjamas (Attachment ambiguity)
- He bought blue socks and hats (Scope ambiguity)

#### Lexical Ambiguity

- The beach was mined
- I <mark>saw</mark> a log
- They can fish

#### Syntactic Ambiguity

- I shot an elephant in my pyjamas (Attachment ambiguity)
- He bought blue socks and hats (Scope ambiguity)

#### **Co-reference Ambiguity**

- John put the bowl on the plate and it broke
- The man couldn't lift his son because he was so weak. (Winograd challenge)

#### Lexical Ambiguity

- The beach was mined
- I <mark>saw</mark> a log
- They can fish

#### Syntactic Ambiguity

- I shot an elephant in my pyjamas (Attachment ambiguity)
- *He bought blue socks and hats* (Scope ambiguity)

#### **Co-reference Ambiguity**

- John put the bowl on the plate and it broke
- The man couldn't lift his son because he was so weak. (Winograd challenge)

#### Homonymy Unrelated interpretations

Unrelated interpretations e.g. *pitcher* 

Unrelated interpretations e.g. *pitcher* 



Unrelated interpretations e.g. *pitcher* 



Unrelated interpretations e.g. *pitcher* 





### **Polysemy** Related interpretations e.g. *paper*

Unrelated interpretations e.g. *pitcher* 





# **Polysemy** Related interpretations

#### e.g. paper



Unrelated interpretations e.g. *pitcher* 





# **Polysemy** Related interpretations

#### e.g. paper



• Homonymous words VS polysemous words:

- Homonymous words VS polysemous words:
  - **Frequency** effect in homonymous words: Most common activated first

- Homonymous words VS polysemous words:
  - Frequency effect in homonymous words: Most common activated first
  - Context effect: Context can change the activation weights

- Homonymous words VS polysemous words:
  - Frequency effect in homonymous words: Most common activated first
  - Context effect: Context can change the activation weights
- Verbs VS Nouns:

- Homonymous words VS polysemous words:
  - Frequency effect in homonymous words: Most common activated first
  - Context effect: Context can change the activation weights
- Verbs VS Nouns:
  - Nouns: disambiguation happens when disambiguating context is given;

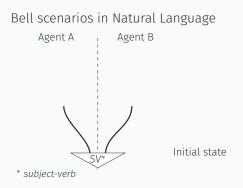
- Homonymous words VS polysemous words:
  - Frequency effect in homonymous words: Most common activated first
  - Context effect: Context can change the activation weights
- Verbs VS Nouns:
  - Nouns: disambiguation happens when disambiguating context is given;
  - Verbs: wait to know its arguments first (for homonymous verbs) or end of clause/sentence (for polysemous verbs).

# Contextuality in Natural language

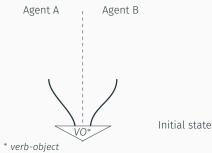
#### Bell scenarios in Natural Language

#### Bell scenarios in Natural Language

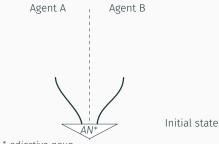
Agent A , Agent B





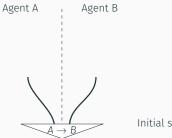




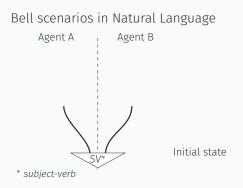


\* adjective-noun

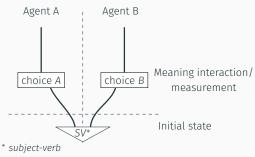
#### Bell scenarios in Natural Language

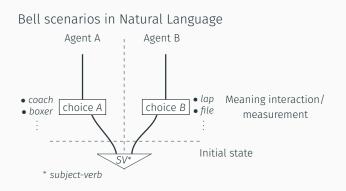


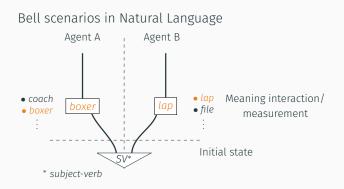
Initial state



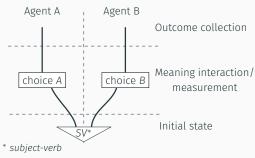






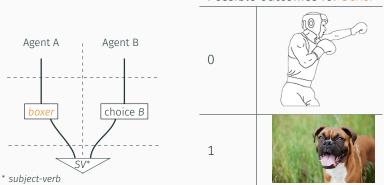


#### Bell scenarios in Natural Language



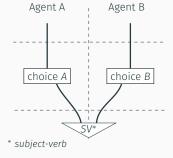
# Logical contextuality

#### Bell scenarios in Natural Language



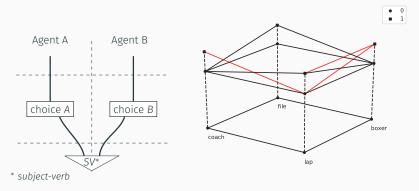
#### Possible outcomes for *boxer*

#### Bell scenarios in Natural Language : Possibilistic models



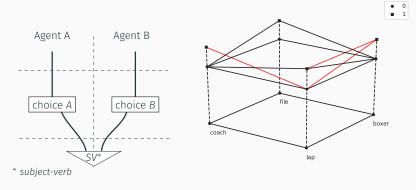
subject	verb	(0,0)	(0,1)	(1,0)	(1,1)	
coach	lap	1	1	1	0	
coach	file	1	1	0	0	
boxer	lap	1	1	1	1	
boxer	file	1	1	0	0	

#### Bell scenarios in Natural Language : Possibilistic models



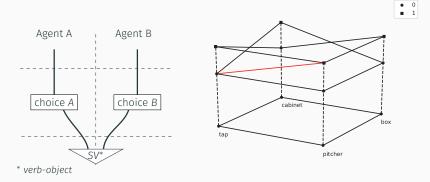
### Logical contextuality





Signalling

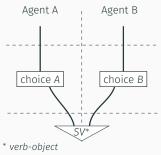
Bell scenarios in Natural Language : Possibilistic models



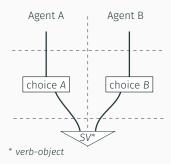
### (Weakly) logically contextual

٠

### Bell scenarios in Natural Language : Probabilistic models

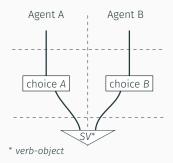


Bell scenarios in Natural Language : Probabilistic models Probabilities from corpora:



Bell scenarios in Natural Language : Probabilistic models Probabilities from corpora:

• BNC

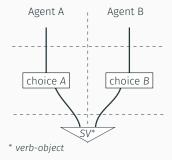


Bell scenarios in Natural Language : Probabilistic models

Probabilities from corpora:

• BNC

100 million words, spread across press articles, fiction, transcription of spoken language, and academic publications ...



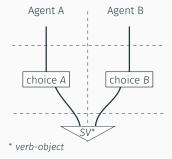
Bell scenarios in Natural Language : Probabilistic models

Probabilities from corpora:

• BNC

100 million words, spread across press articles, fiction, transcription of spoken language, and academic publications ...

• ukWaC



Bell scenarios in Natural Language : Probabilistic models

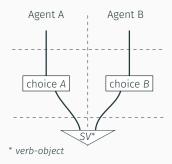
Probabilities from corpora:

• BNC

100 million words, spread across press articles, fiction, transcription of spoken language, and academic publications ...

• ukWaC

2 billion words, constructed from the Web (*.uk* domains)



Bell scenarios in Natural Language : Probabilistic models

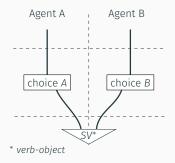
Probabilities from corpora:

• BNC

100 million words, spread across press articles, fiction, transcription of spoken language, and academic publications ...

• ukWaC

2 billion words, constructed from the Web (.*uk* domains)



Bell Scenarios: Cyclic systems of rank 4

<sup>1</sup>Out of 1682 models

### Bell Scenarios: Cyclic systems of rank 4

 $\cdot$  No CbD-contextual instances (yet)<sup>1</sup>

### Bell Scenarios: Cyclic systems of rank 4

- No CbD-contextual instances (yet)<sup>1</sup>
  - Corpus data: incomplete?

• 2 words, 2 different contexts

- 2 words, 2 different contexts
  - 1 verb + 1 noun: verb-object, subject-verb contexts

- 2 words, 2 different contexts
  - 1 verb + 1 noun: verb-object, subject-verb contexts

Verb: eat	Noui	n: chicken
Con	text	Example
Verb-obj	ect:	Humans <mark>eat chicken</mark> .
Subject-v	erb:	The <mark>chicken eats</mark> grains.

	Adopt	Boxer	
	They are thinking	The heavyweight	
0	about adopting a	boxer won the	
	child	fight	
	They are thinking	The boxer dog	
1	about adopting a	harked	
	new policy	burkeu.	

	Adopt	Boxer	
	They are thinking	The heavyweight	
0	about adopting a	boxer won the	
	child	fight	
	They are thinking	The boxer dog	
1	about adopting a	harked	
	new policy	Durkeu.	

(adopt, boxer)	(0,0)	(0,1)	(1,0)	(1,1)
adopt boxer	0	29/30	1/30	0
boxer adopts	1/4	0	0	3/4

	Adopt	Boxer	
	They are thinking	The heavyweight	• Contextual!
0	about adopting a	boxer won the	
	child	fight	
	They are thinking	The boxer dog	
1	about adopting a	harked.	
	new policy	Durkeu.	

(adopt, boxer)	(0,0)	(0,1)	(1,0)	(1,1)
adopt boxer	0	29/30	1/30	0
boxer adopts	1/4	0	0	3/4

	Adopt	Boxer
	They are thinking	The heavyweight
0	about adopting a	boxer won the
	child	fight
	They are thinking	The boxer dog
1	about adopting a	harked.
	new policy	Durkeu.

(adopt, boxer)	(0,0)	(0,1)	(1,0)	(1,1)
adopt boxer	0	29/30	1/30	0
boxer adopts	1/4	0	0	3/4

• Contextual!

2 > 28/15

	Adopt	Boxer	
	They are thinking	The heavyweight	•
0	about adopting a	boxer won the	
	child	fight	
	They are thinking	The boxer dog	
1	about adopting a	harked	
	new policy	Durkeu.	

Contextual!

2 > 28/15

• Contextuality measure: 1/30

(adopt, boxer)	(0,0)	(0,1)	(1,0)	(1,1)
adopt boxer	0	29/30	1/30	0
boxer adopts	1/4	0	0	3/4

	Adopt	Boxer
	They are thinking	The heavyweight
0	about adopting a	boxer won the
	child	fight
	They are thinking	The boxer dog
1	about adopting a	harked
	new policy	burkeu.

(adopt, boxer)	(0,0)	(0,1)	(1,0)	(1,1)	
adopt boxer	0	29/30	1/30	0	
boxer adopts	1/4	0	0	3/4	

• Contextual!

2 > 28/15

- Contextuality measure: 1/30
- Non-contextuality probability: > 0.56

### Throwing pitchers and Pitchers throwing

	Throw	Pitcher	
0	She threw the ball across the field.	The pitcher was filled with water.	
1	They are throw- ing a party.	In baseball, pitchers usually bats as well.	

(throw, pitcher)	(0,0)	(0,1)	(1,0)	(1,1)
throw pitcher	2/5	0	1/10	1/2
pitcher throws	0	2/3	1/3	0

• Also contextual!

9/5 > 13/15

- Contextuality measure: 7/30
- Non-contextuality probability: > 0.08

# Beyond contextuality: analysis of signalling

• Models considered are signalling

- Models considered are signalling
- Contextuality-by-Default and M-contextuality offers a way to quantify the degree of signalling

- Models considered are signalling
- Contextuality-by-Default and M-contextuality offers a way to quantify the degree of signalling

### Motivation

Can we use the "measures" of signalling to detect context effects in natural language?

- Models considered are signalling
- Contextuality-by-Default and M-contextuality offers a way to quantify the degree of signalling

#### Motivation

Can we use the "measures" of signalling to detect context effects in natural language?

• Focus on cyclic models of rank 2

- Models considered are signalling
- Contextuality-by-Default and M-contextuality offers a way to quantify the degree of signalling

### Motivation

Can we use the "measures" of signalling to detect context effects in natural language?

- Focus on cyclic models of rank 2
- Compare the data for models in the following categories: Verbs with multiple meanings - Noun with multiple meanings
   Verbs with multiple senses - Noun with multiple meanings
   Verbs with multiple senses - Noun with multiple senses

• Average essentially equal across different classes

• Average essentially equal across different classes

Verbs Nouns	Meanings	Senses	Overall
Meanings	$1.24\pm0.29$	$1.36\pm0.15$	$1.33\pm0.14$
Senses	$1.50\pm0.43$	$1.38\pm0.36$	$1.41 \pm 0.29$
Overall	$1.30\pm0.25$	$1.36\pm0.14$	$1.35\pm0.12$

- Average essentially equal across different classes
- Small disparities in standard deviations

- Average essentially equal across different classes
- Small disparities in standard deviations

Classes of models	SD
Verbs with multi. meanings	1.04
Verbs with multi senses	1.20
Nouns with multi meanings	1.18
Nouns with multi senses	1.11

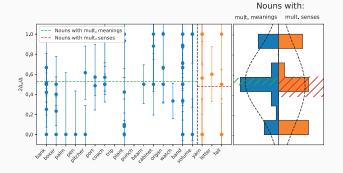
# Proportion of direct influence from the verb (2 $\Delta^{*}_{c_{1},c_{2}}$ (F<sub>v</sub>) / $\Delta$ )

# Proportion of direct influence from the verb (2 $\Delta^{*}_{c_{1},c_{2}}$ (F<sub>v</sub>) / $\Delta$ )

• Combinations involving nouns with multiple meanings VS multiple senses: basically unchanged

# Proportion of direct influence from the verb (2 $\Delta_{c_1,c_2}^*$ ( $F_v$ ) / $\Delta$ )

• Combinations involving nouns with multiple meanings VS multiple senses: basically unchanged



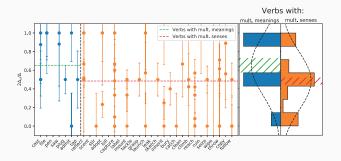
# Proportion of direct influence from the verb (2 $\Delta^{*}_{c_{1},c_{2}}$ (F<sub>v</sub>) / $\Delta$ )

- Combinations involving nouns with multiple meanings VS multiple senses: basically unchanged
- Combinations involving verbs with multiple meanings VS multiple senses: larger variation when combination involves verbs with multiple meanings

#### Analysis of signalling

## Proportion of direct influence from the verb (2 $\Delta^{*}_{c_{1},c_{2}}$ (F<sub>v</sub>) / $\Delta$ )

- Combinations involving nouns with multiple meanings VS multiple senses: basically unchanged
- Combinations involving verbs with multiple meanings VS multiple senses: larger variation when combination involves verbs with multiple meanings



Conclusion

• Natural Language exhibits contextuality in a way that is similar to contexutality arising from Quantum Mechanics

- Natural Language exhibits contextuality in a way that is similar to contexutality arising from Quantum Mechanics
  - Logical contextuality in Bell-like scenarios

- Natural Language exhibits contextuality in a way that is similar to contexutality arising from Quantum Mechanics
  - Logical contextuality in Bell-like scenarios
  - Contextuality-by-Default contextuality in cyclic models of rank 2

- Natural Language exhibits contextuality in a way that is similar to contexutality arising from Quantum Mechanics
  - Logical contextuality in Bell-like scenarios
  - Contextuality-by-Default contextuality in cyclic models of rank 2
- Framework of contextuality can be used for formalising the analysis of context effects in natural language

- Natural Language exhibits contextuality in a way that is similar to contexutality arising from Quantum Mechanics
  - Logical contextuality in Bell-like scenarios
  - Contextuality-by-Default contextuality in cyclic models of rank 2
- Framework of contextuality can be used for formalising the analysis of context effects in natural language
  - Bounds on context representation for computational applications

- Natural Language exhibits contextuality in a way that is similar to contexutality arising from Quantum Mechanics
  - Logical contextuality in Bell-like scenarios
  - Contextuality-by-Default contextuality in cyclic models of rank 2
- Framework of contextuality can be used for formalising the analysis of context effects in natural language
  - · Bounds on context representation for computational applications
  - Quantification of influences of context

• Collect human judgements

- Collect human judgements
- Analysis of signalling: add causality in the equation

- Collect human judgements
- Analysis of signalling: add causality in the equation
  - Is our data compatible with certain causal structure?

- Collect human judgements
- Analysis of signalling: add causality in the equation
  - Is our data compatible with certain causal structure?
  - Which ones and why?

- Collect human judgements
- Analysis of signalling: add causality in the equation
  - Is our data compatible with certain causal structure?
  - Which ones and why?
- Use contextuality analysis to further the analogy with the work from psycholinguistics

- Collect human judgements
- Analysis of signalling: add causality in the equation
  - Is our data compatible with certain causal structure?
  - Which ones and why?
- Use contextuality analysis to further the analogy with the work from psycholinguistics
  - Frequency effect (dominant VS subordinate interpretations)

- Collect human judgements
- Analysis of signalling: add causality in the equation
  - Is our data compatible with certain causal structure?
  - Which ones and why?
- Use contextuality analysis to further the analogy with the work from psycholinguistics
  - Frequency effect (dominant VS subordinate interpretations)
  - Priming context VS no priming context

- Collect human judgements
- Analysis of signalling: add causality in the equation
  - Is our data compatible with certain causal structure?
  - Which ones and why?
- Use contextuality analysis to further the analogy with the work from psycholinguistics
  - Frequency effect (dominant VS subordinate interpretations)
  - Priming context VS no priming context
  - Garden path sentences e.g. *The old man the boat*

- Collect human judgements
- Analysis of signalling: add causality in the equation
  - Is our data compatible with certain causal structure?
  - Which ones and why?
- Use contextuality analysis to further the analogy with the work from psycholinguistics
  - Frequency effect (dominant VS subordinate interpretations)
  - Priming context VS no priming context
  - Garden path sentences e.g. *The old man the boat*
  - Senses following rules
     e.g. chicken (animal) → chicken(meat)

#### Background References i

- S. Abramsky, R. Soares Barbosa, K. Kishida, R. Lal, and S. Mansfield.

### Contextuality, Cohomology and Paradox.

arXiv e-prints, page arXiv:1502.03097, Feb. 2015.

E. N. Dzhafarov and J. V. Kujala.

### Context–content systems of random variables: The contextuality-by-default theory.

*Journal of Mathematical Psychology*, 74:11–33, 2016. Foundations of Probability Theory in Psychology and Beyond.

#### M. Jones.

**Relating causal and probabilistic approaches to contextuality.** *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 377(2157):20190133, 2019.



M. Pickering and S. Frisson. **Processing ambiguous verbs: Evidence from eye movements.**  *Journal of experimental psychology. Learning, memory, and cognition, 27:556–73, 03 2001.*