

Analysis of Conversations

Lecture 04 (Analysis I)

Alessandro Vinciarelli



University
of Glasgow



Social AI



Engineering and
Physical Sciences
Research Council

Outline

- Why Conversations
- Example: Disagreement Analysis
- Conclusions

Outline

- Why Conversations
- Example: Disagreement Analysis
- Conclusions

Why Conversation

“The **most important determinants** of successful development, mental health, and our individual traits and preferences **arise from interacting** with other individuals. Social interaction underpins **who we are, how we think, and how we behave.**”

Why Conversation

“[...] the fundamental or primordial scene of social life is that of direct interaction between members of a social species [...] **talking in interaction** appears to be a distinctive form of this **primary constituent of social life** [...]”

Outline

- Why Conversations
- Example: Disagreement Analysis
- Conclusions

The Six Steps

- **Definition of the Construct** (what are we analysing?)
- **Identification of the Layers** (physical vs inferential)
- **Data Collection** (where is the phenomenon?)
- **Task Definition** (shaping the AI approach)
- **Performance Analysis** (finding confounding factors)
- **The Interdisciplinary Loop** (psychological insight)

The Six Steps

- **Definition of the Construct** (what are we analysing?)
- **Identification of the Layers** (physical vs inferential)
- **Data Collection** (where is the phenomenon?)
- **Task Definition** (shaping the AI approach)
- **Performance Analysis** (finding confounding factors)
- **The Interdisciplinary Loop** (psychological insight)

Disagreement

“[...] an **oppositional stance** (verbal or non-verbal) to an antecedent verbal (or non-verbal) **action**.”

Agreement

“[...] a relation of identity, similarity or congruence between the mental states of two or more persons, where these mental states are in any case opinions.”

The Six Steps

- Definition of the Construct (what are we analysing?)
- **Identification of the Layers** (physical vs inferential)
- Data Collection (where is the phenomenon?)
- Task Definition (shaping the AI approach)
- Performance Analysis (finding confounding factors)
- The Interdisciplinary Loop (psychological insight)

The Layers

- Both definitions include two **layers**;
- The **physical** layer corresponds to **observable, machine detectable** phenomena (action, identity, etc.);
- The **inferential** layer corresponds to social, cognitive and psychological phenomena **not accessible to senses** (opinions, stance, etc.);
- **Analysis** approaches **bridge the gap** between physical and inferential layer.

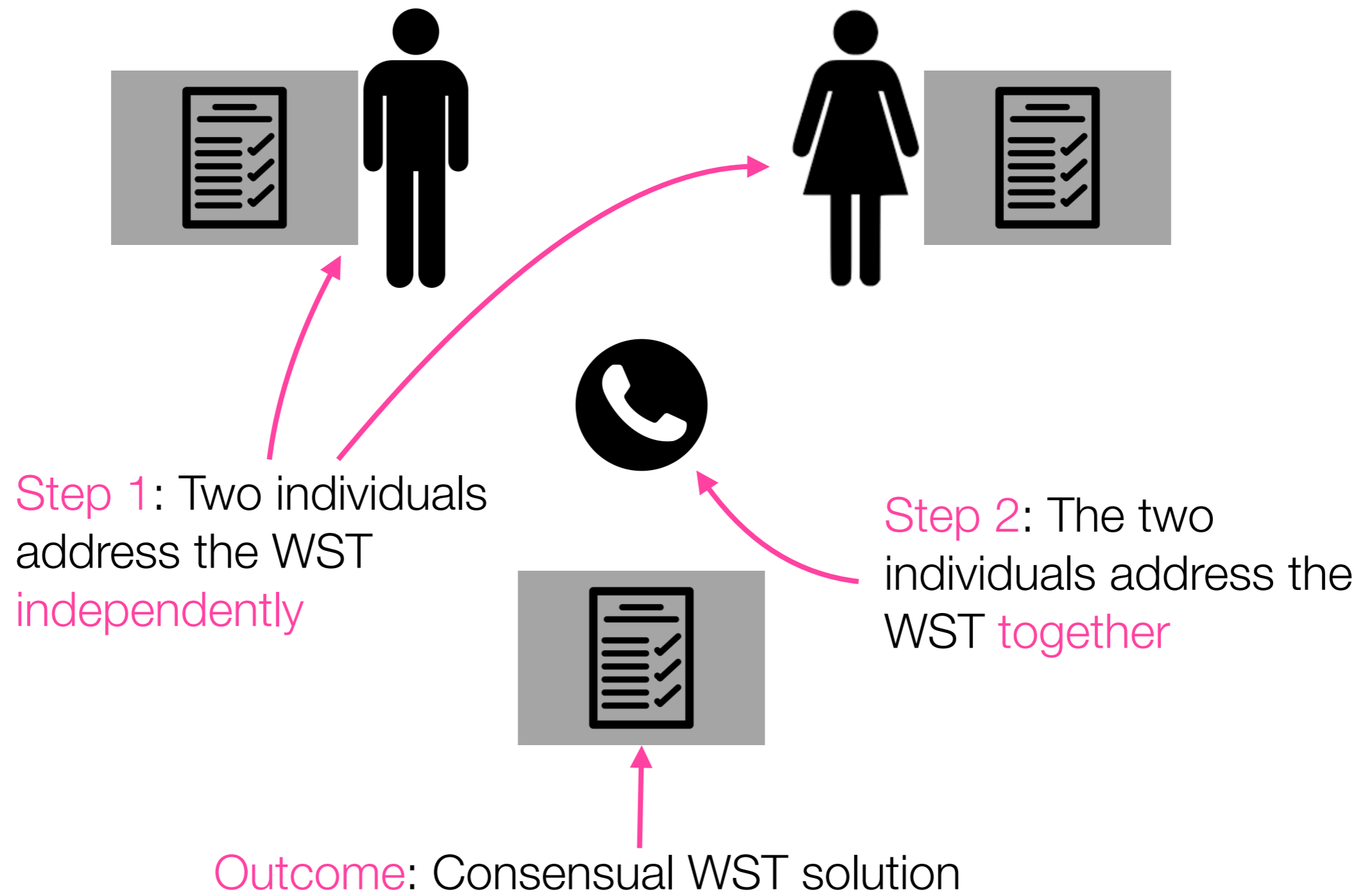
The Six Steps

- **Definition of the Construct** (what are we analysing?)
- **Identification of the Layers** (physical vs inferential)
- **Data Collection** (where is the phenomenon?)
- **Task Definition** (shaping the AI approach)
- **Performance Analysis** (finding confounding factors)
- **The Interdisciplinary Loop** (psychological insight)

The Winter Survival Task

- You must **help the survivors** of a plane crash in a polar area during the winter;
- You must suggest the **items most likely to increase the chances of survival** (as a few as possible);
- It is **often used in psychological experiments** because people have no technical knowledge and **social dynamics dominate**;
- The 12 items are clothing, compass, chocolate, canvas, steel wool, whisky, newspaper, butter, airmap, pistol, lighter, and axe.

Eliciting Disagreement



Why Decision Making

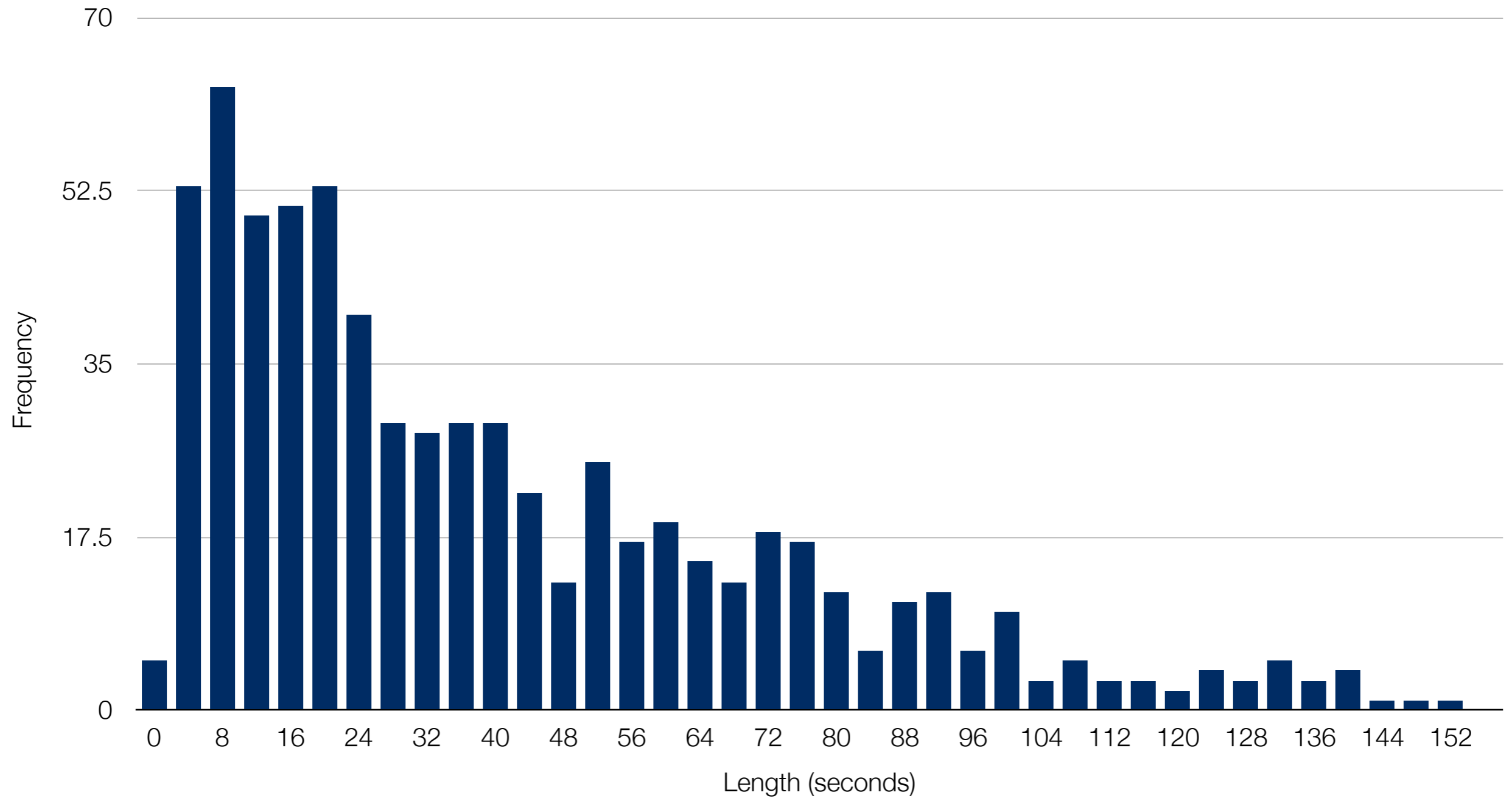
“[...] certain practices are prone to contain disagreement [...] for example [...] decision making and problem solving talk in either every day or professional contexts [...]”

The Glasgow Disagreement Corpus

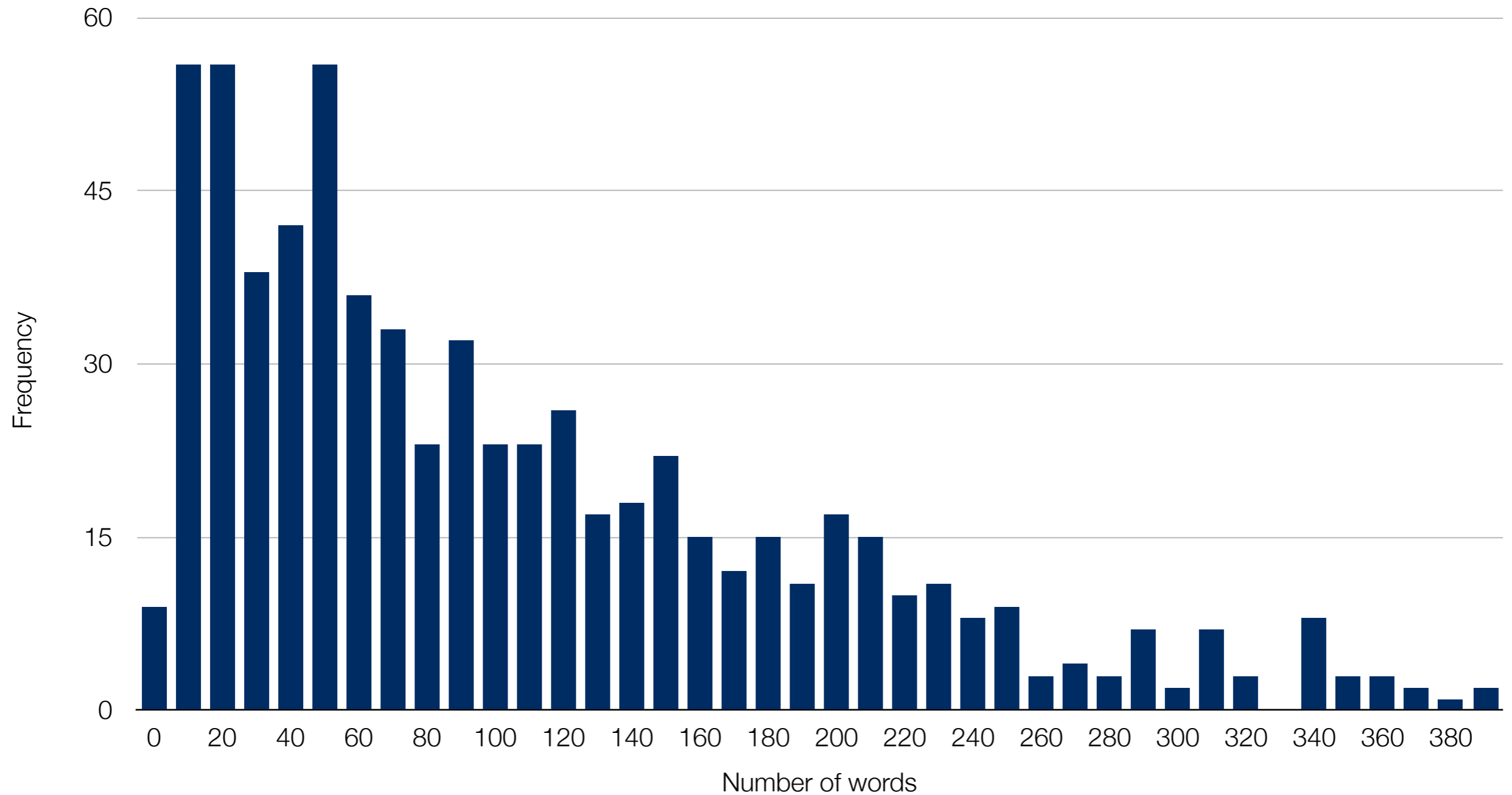
Composition	FF	MM	FM
Number	196 (28.7%)	157 (22.9%)	331 (48.4%)

- 120 participants (63 female and 57 male);
- 684 clips of average length 41.9 ± 32.9 sec.;
- The average number of words is 112.3 ± 90.5 .
- <https://github.com/disagreementdataset/disagreement-dataset>

Length Distribution



Word Number Distribution



The Annotation

- The key-problem is that there is **no rigorous definition** of disagreement **accessible to the “naive” observer**;
- This makes it **difficult** to address the problem of disagreement **directly**;
- Every clip was rated by 10 different annotators in terms of **percentage of time spent in disagreement** (0% to 100%);
- The **10 ratings** obtained for every clip were **averaged** to obtain the percentage of time associated to a clip;
- **Publicly available**: <https://www.prolific.com/>

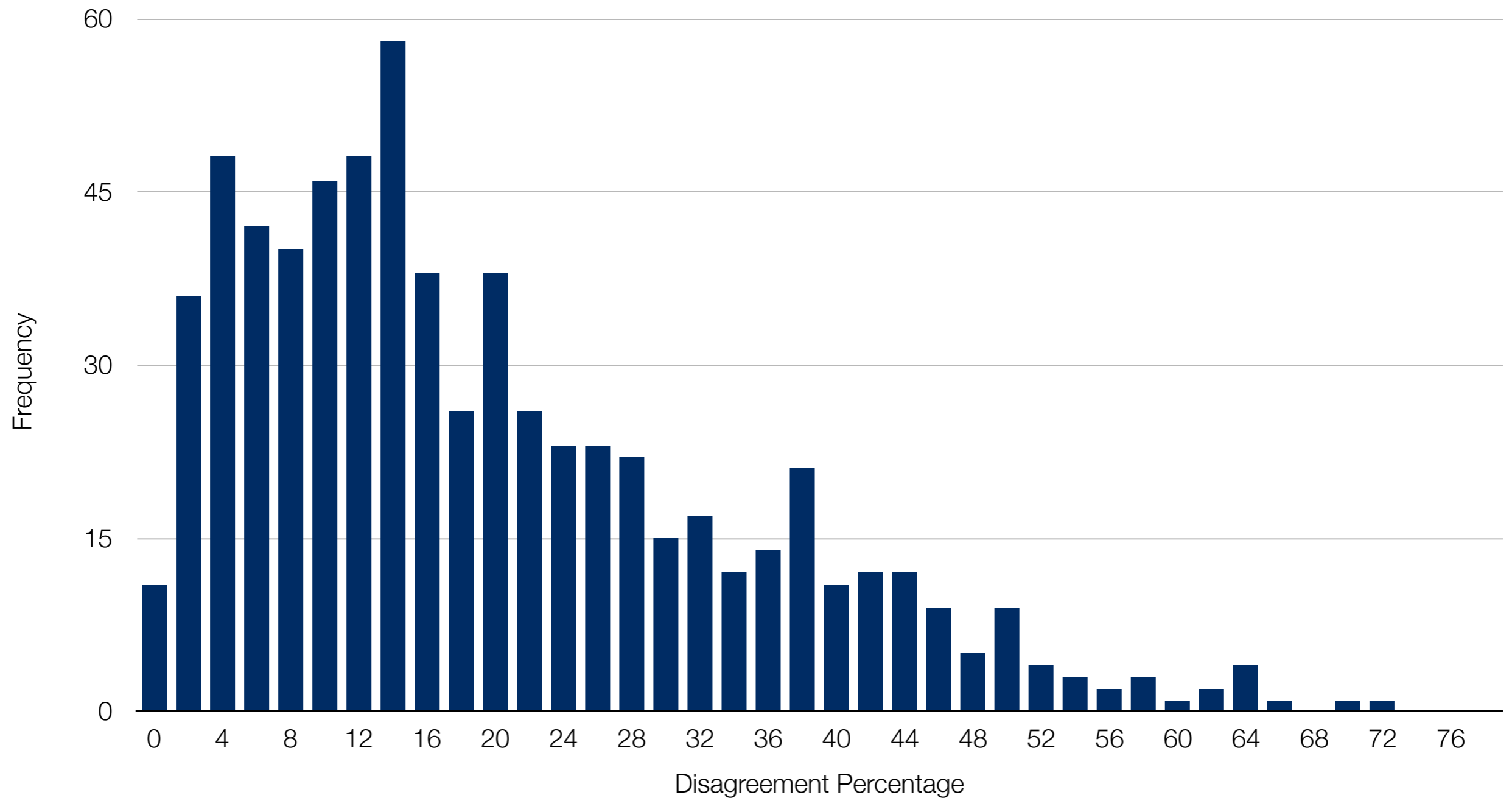
Disagreement Perception

“[...] human judgment of **perceived** agreement or disagreement.”

Disagreement Duration

“[...] evidence suggests that disagreement is beneficial, because it exposes problems and it instigates change. After those benefits are realised, however, additional fighting starts to corrode relationships.”

Disagreement Percentage Distribution



The Annotation

- Roughly **half of the clips** show **low percentages** of disagreement ($<15\%$);
- Roughly **5% of the clips** show **high percentages** of disagreement ($>50\%$);
- The **distribution is continuous** and not bi-modal, there is no clear distinction between agreement and disagreement;
- The distribution does not reach 100% because **participants are asked to achieve consensus.**

Disagreement Perception

“[...] people are supposed to **not like to disagree** and prefer to **pursue consensus** where possible.”

Reliability (Spearman Brown)

$$\rho_{SB} = \frac{N\rho}{1 + (N - 1)\rho}$$

- N is the **number of raters** who assessed the data;
- ρ is the **average correlation** between two raters that assessed the same data;
- ρ_{SB} is bound between 0 and 1 and **it measures the reliability**, the typical requirement is $\rho_{SB} \geq 0.7$ ($\rho_{SB} = 0.87$ for the data).

Reliability and Subjectivity

- The minimum number of necessary raters depends on their correlation, i.e., on the **consistency between stimulus** (physical layer) **and perception** (inferential layer);
- **High reliability** suggests that **different raters interpret the data in the same way**;
- While no definition of (dis)agreement was provided, **raters show consensus** on what it is;
- The annotation has **worked**.

The Glasgow Disagreement Corpus



4%



26%



57%



71%

- <https://github.com/disagreementdataset/disagreement-dataset>

The Six Steps

- **Definition of the Construct** (what are we analysing?)
- **Identification of the Layers** (physical vs inferential)
- **Data Collection** (where is the phenomenon?)
- **Task Definition** (shaping the AI approach)
- **Performance Analysis** (finding confounding factors)
- **The Interdisciplinary Loop** (psychological insight)

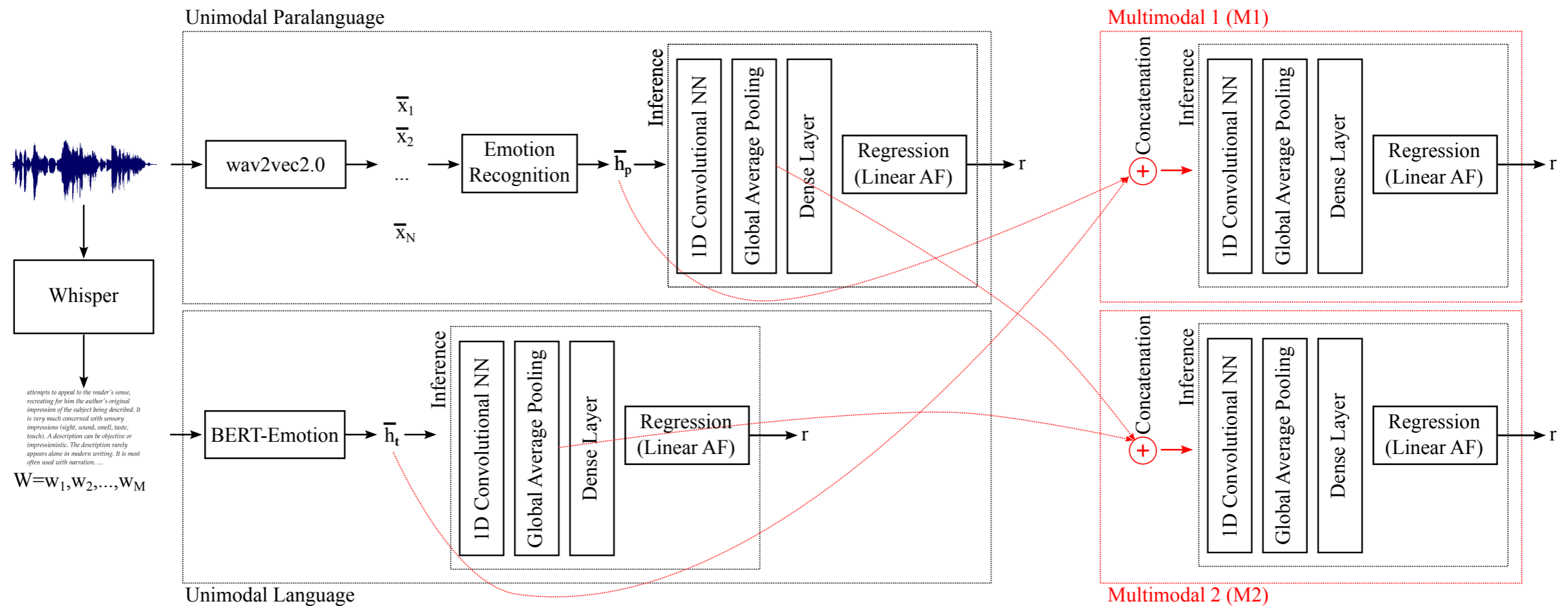
The Role of Emotions

“[...] disagreement [...] may or may not be verbalized, and its form may incorporate a variety of **affective markers** to signal an **emotional association**”

The Approach (I)

- The key-assumption is that people experience and manifest **different emotions when they agree and when they disagree**;
- Emotions leave traces and, therefore, **different emotions leave different machine detectable traces**;
- The expectation is that the “**oppositional stance**” at the core of the disagreement definition **leaves traces in terms of detectable emotions**;
- The analysis of disagreement takes place through the analysis of emotions.

The Approach (II)



<https://github.com/m-bain/whisperX>

<https://huggingface.co/jonatasgrosman/wav2vec2-large-xlsr-53-english>

<https://huggingface.co/AreejB/wav2vec2-xlsr-english-speech-emotion-recognition>

<https://huggingface.co/bhadresh-savani/bert-base-uncased-emotion>

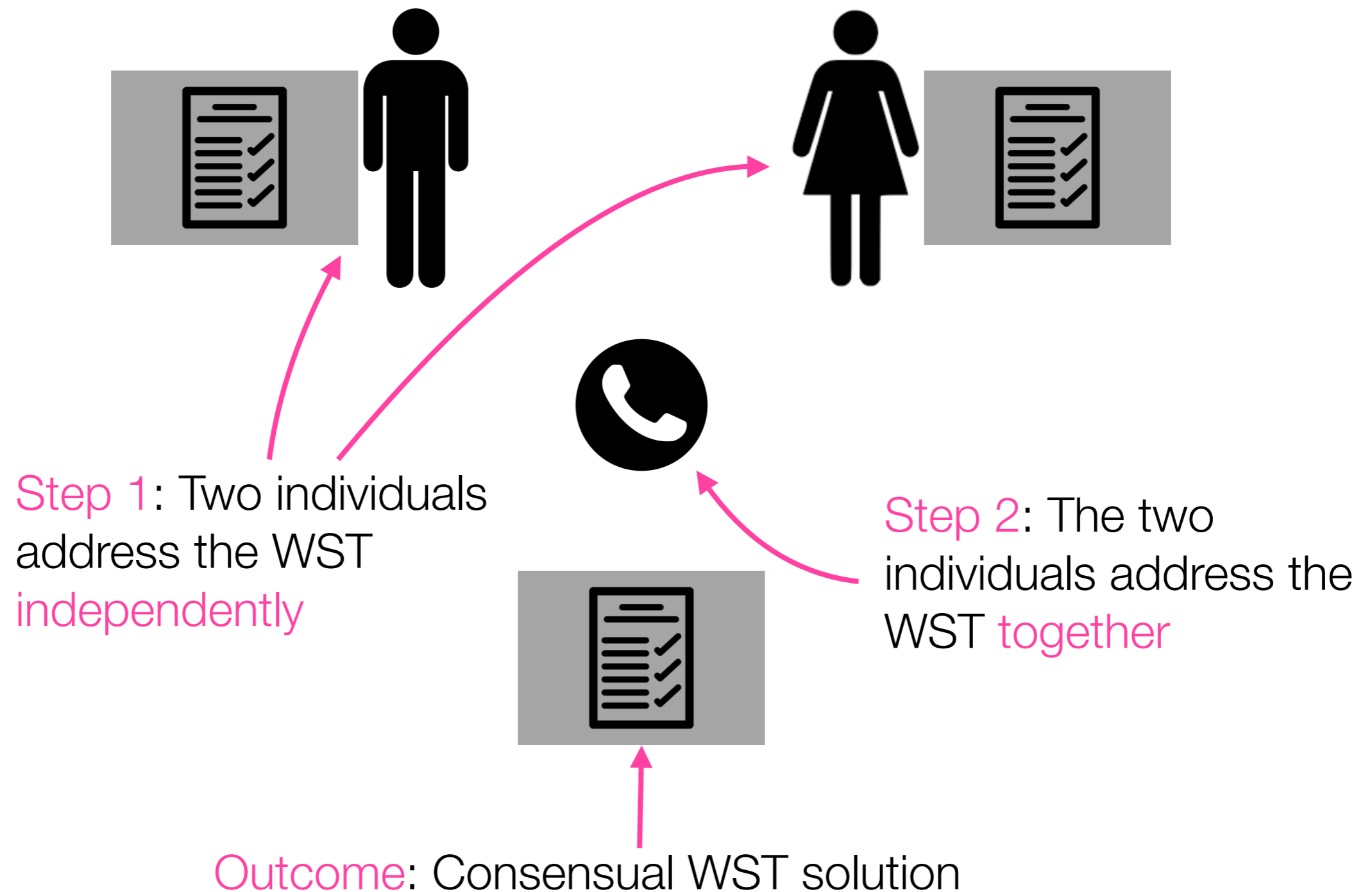
Results

Approach	MAE	RMSE	Correlation
Language	10.2±0.1	13.4±0.1	0.43±0.02
Paralanguage	10.4±0.0	13.3±0.1	0.41±0.02
Multimodal 1	9.7±0.1	12.6±0.1	0.52±0.01
Multimodal 2	10.1±1.0	13.1±1.7	0.50±0.06
Baseline	15.8	20.7	0.0

The Six Steps

- **Definition of the Construct** (what are we analysing?)
- **Identification of the Layers** (physical vs inferential)
- **Data Collection** (where is the phenomenon?)
- **Task Definition** (shaping the AI approach)
- **Performance Analysis** (finding confounding factors)
- **The Interdisciplinary Loop** (psychological insight)

Eliciting Disagreement



Possible Item Biases

- There is a **risk** that the **approach does not learn about disagreement**, but about the association between items and tendency to agree or disagree;
- If this is the case, the approach is not predicting the percentage of time spent in disagreement (as planned), but the **tendency of an item to consistently elicit short or long disagreement discussions**;
- The risk is **especially important when using language-based methodologies** (all dyads talk about the same items).

Relative Entropy (I)

$$\sum_{k=1}^N p_k = 1$$

- Consider a **probability distribution** over N mutually exclusive events;
- The value of p_k is the probability of event k taking place.

Relative Entropy (II)

$$H = - \frac{\sum_{k=1}^N p_k \log(p_k)}{\log(N)}$$

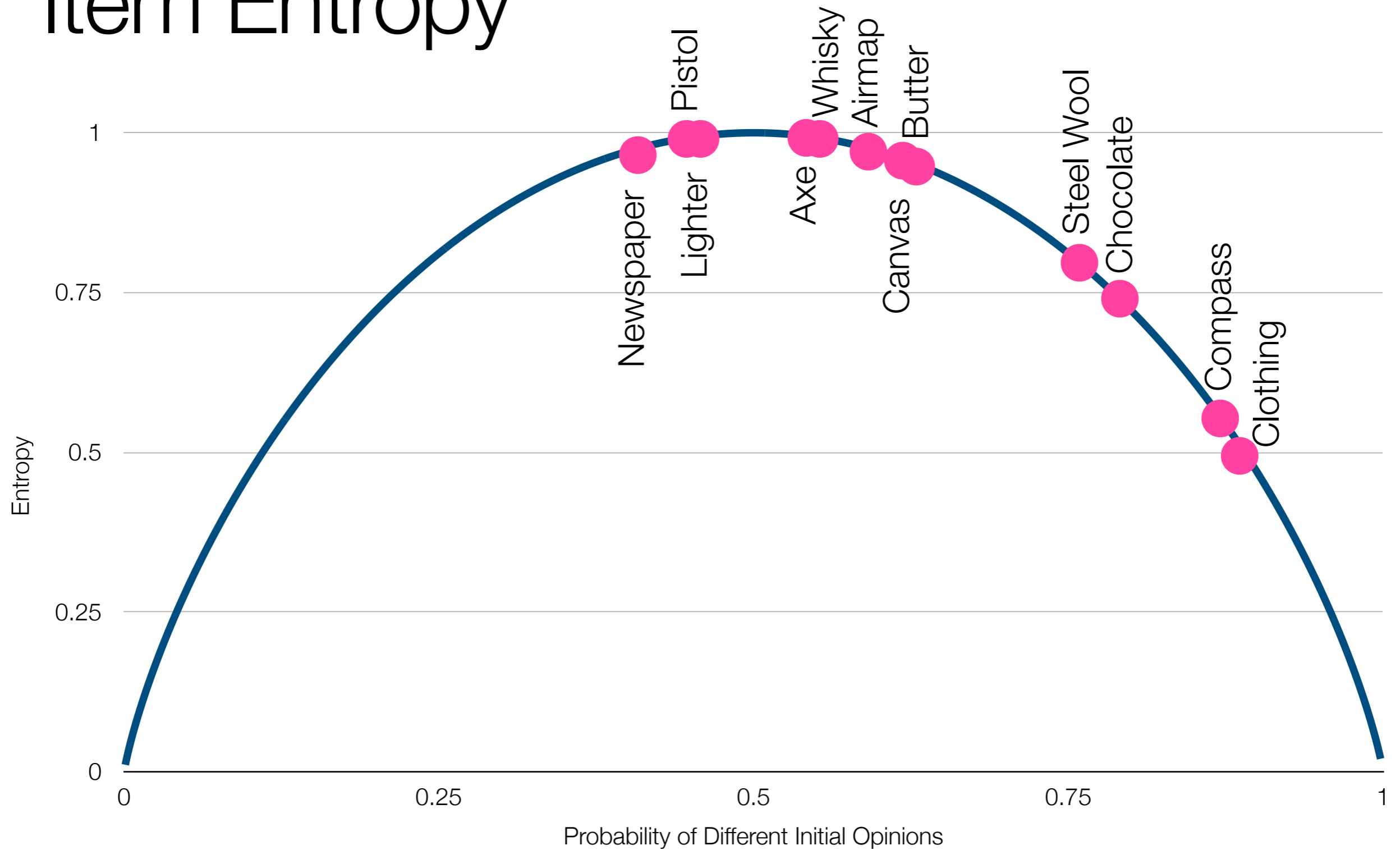
- H is the **relative entropy** of the distribution and its value is bound between 0 (one of the p_k is equal to one) and 1 ($p_k = 1/N \forall k \in \{1, \dots, N\}$);
- $H = 0$ means **maximum certainty**, $H = 1$ means **maximum uncertainty**.

Relative Entropy for $N = 2$

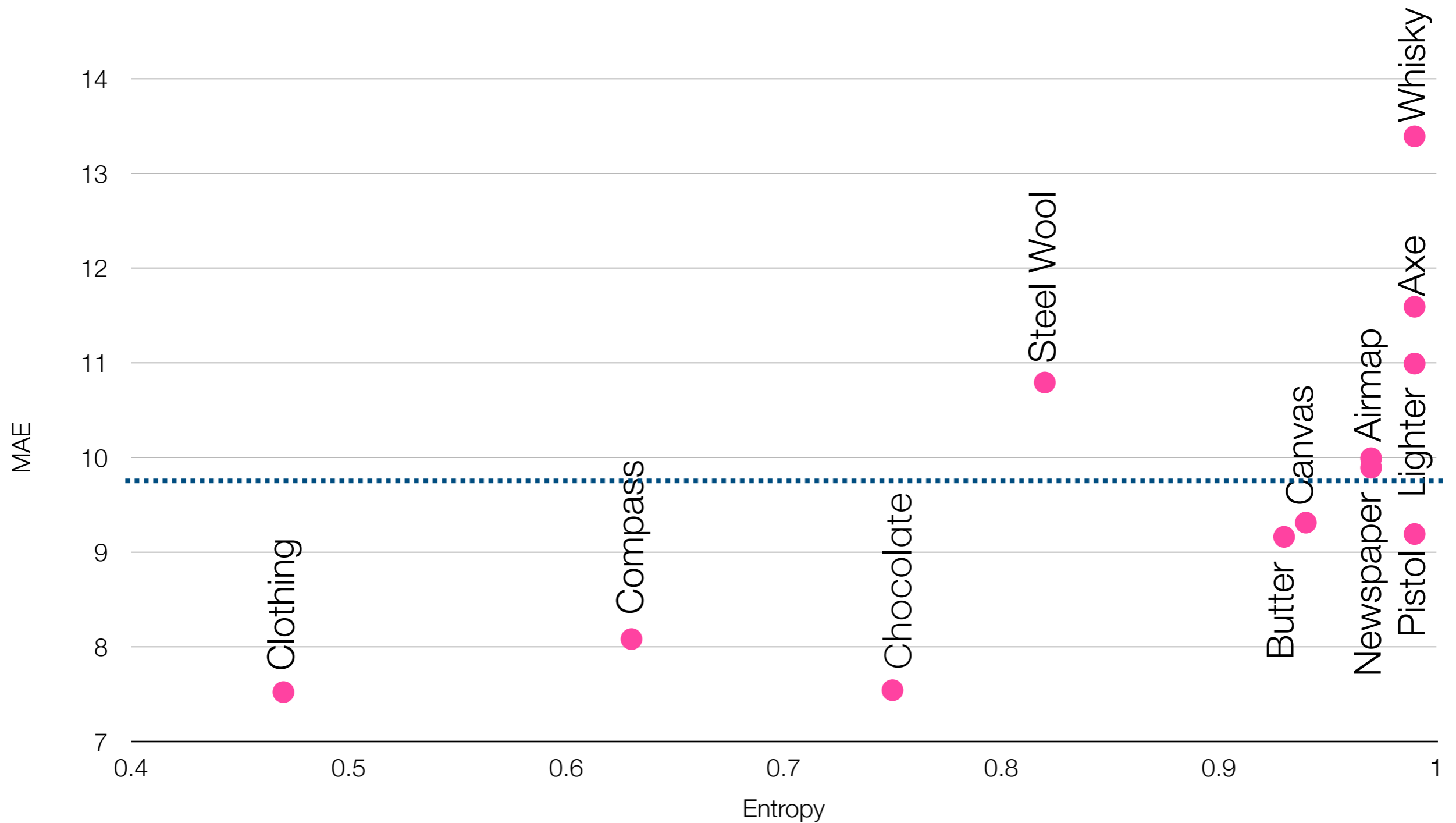
$$H = -\frac{p \log(p) + (1 - p) \log(1 - p)}{\log(2)}$$

- For a given item of the WST, p is the **fraction of times** the dyad members have different opinion;
- If H is **too low**, dyad members tend to have either the same or a different opinion, thus establishing a **relationship between item and disagreement**.

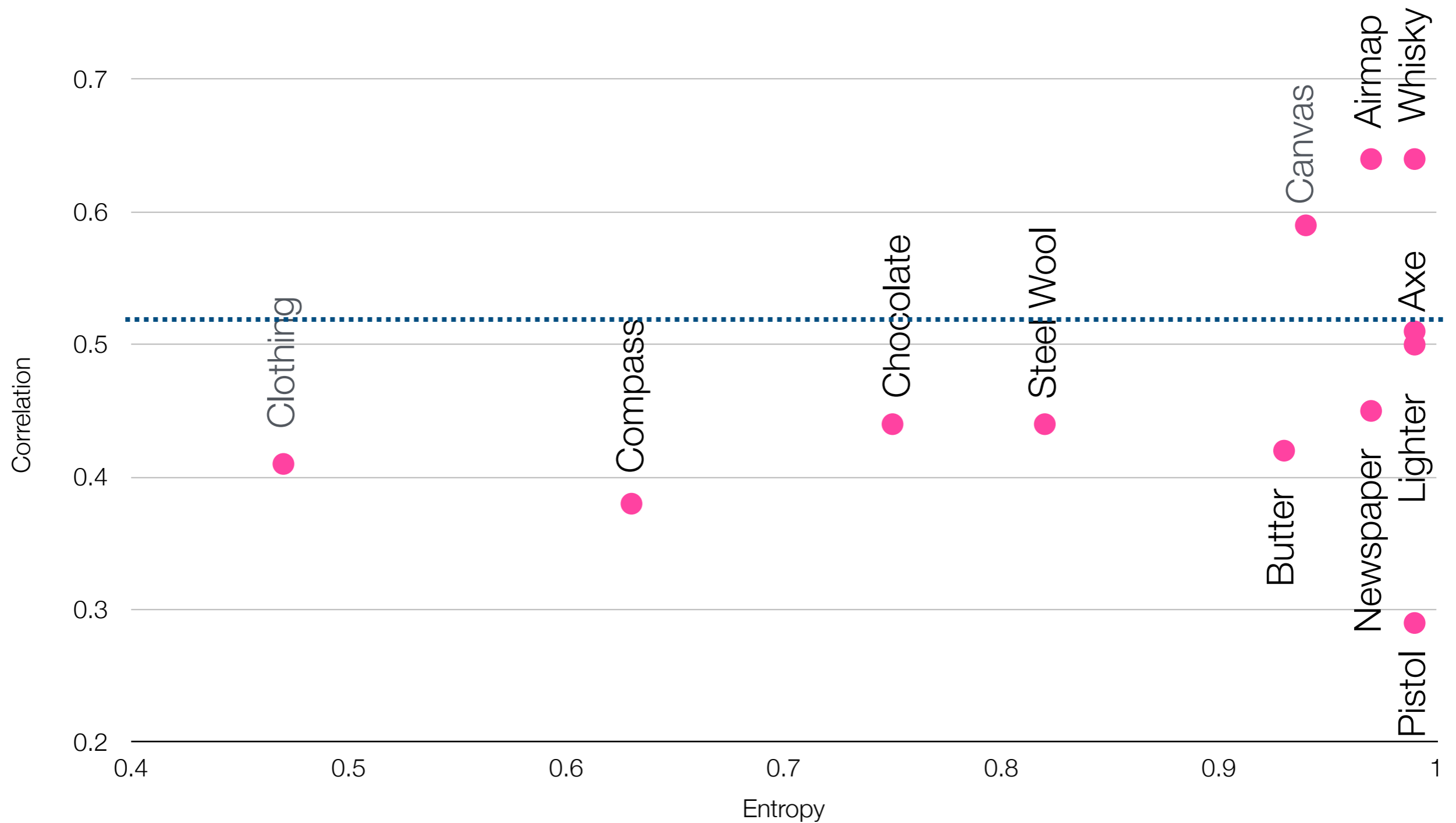
Item Entropy



MAE vs Entropy (MC1)



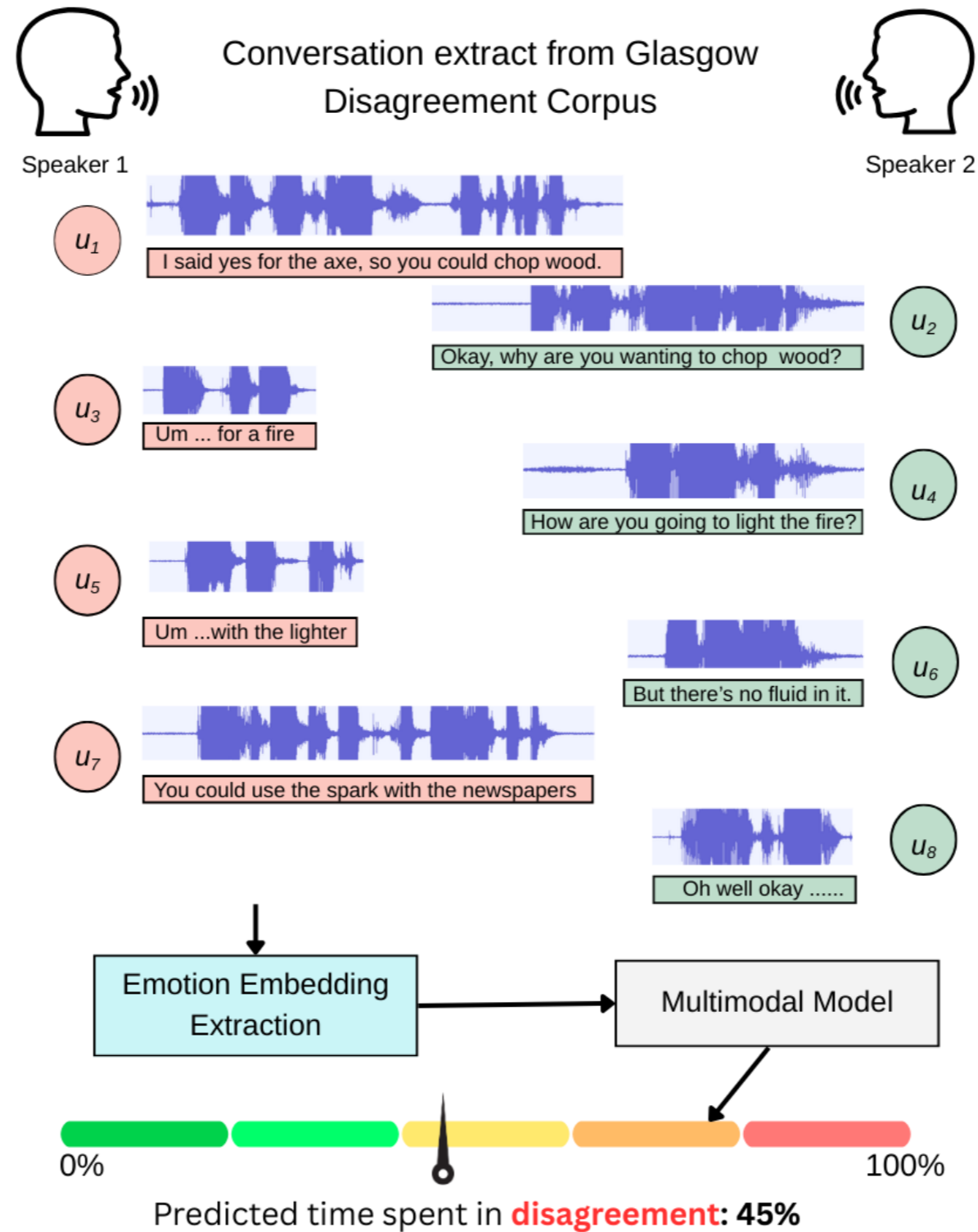
Correlation vs Entropy (MC1)



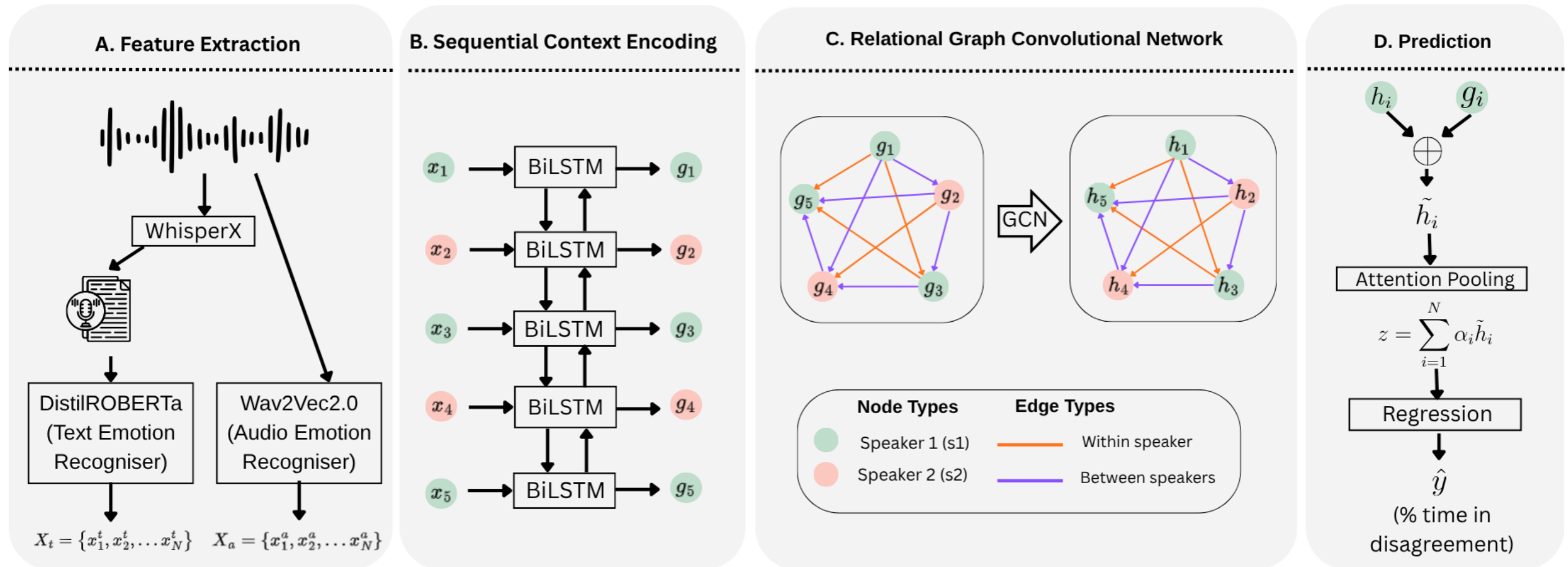
The Six Steps

- **Definition of the Construct** (what are we analysing?)
- **Identification of the Layers** (physical vs inferential)
- **Data Collection** (where is the phenomenon?)
- **Task Definition** (shaping the AI approach)
- **Performance Analysis** (finding confounding factors)
- **The Interdisciplinary Loop** (psychological insight)

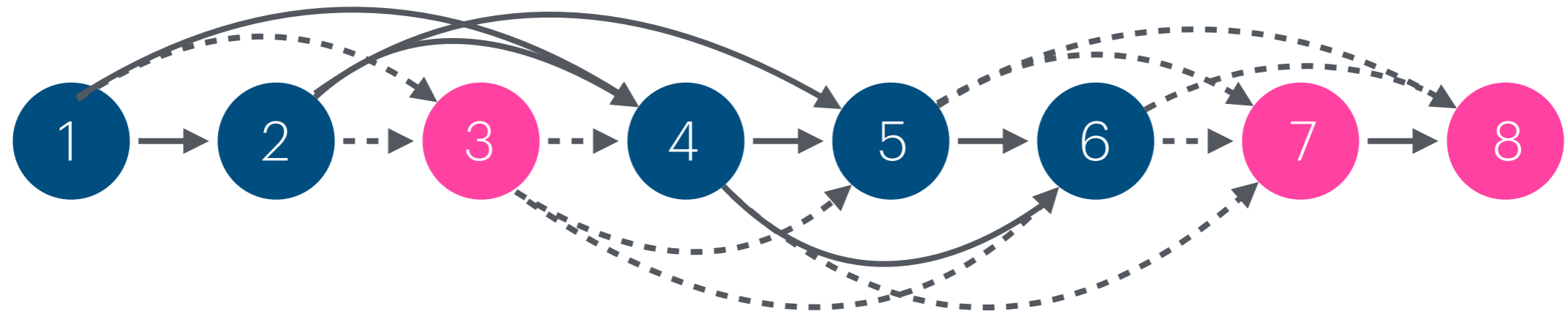
What is a Conversation?



Refining the Approach



Why a Graph?



- The nodes (utterances) are **locally** represented as a weighted sum of the connected nodes (**local weights**);
- The same nodes (utterances) are **globally** represented as a weighted sum of all nodes (**global weights**).

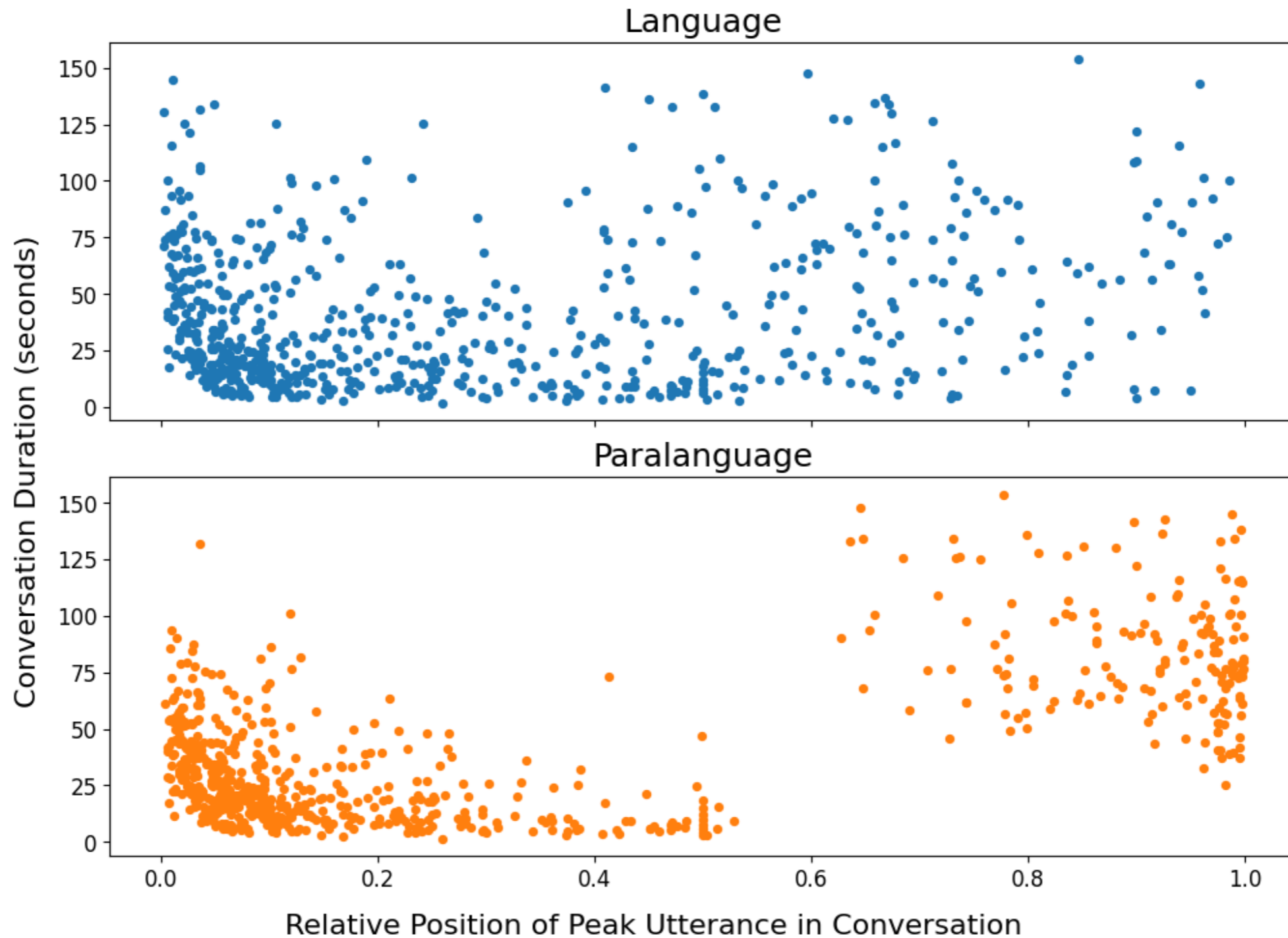
Graph Representation

- The local weights sum to 1 and they allow one to analyse the **interplay between consecutive utterances**;
- The global weights sum to 1 and they allow one to identify the “**peak**” utterances;
- The analysis of the weights provides **insight** about the **dynamics of disagreement**.

Results

Approach	MAE	RMSE	Correlation
Baseline	15.8	20.7	0.0
Previous Approach	9.7±0.1	12.6±0.1	0.52±0.01
Language	85±0.3	11.0±0.1	0.63±0.01
ParaLanguage	8.5±0.3	11.0±0.1	0.63±0.01
Multimodal	8.1±0.3	10.5±0.2	0.67±0.01

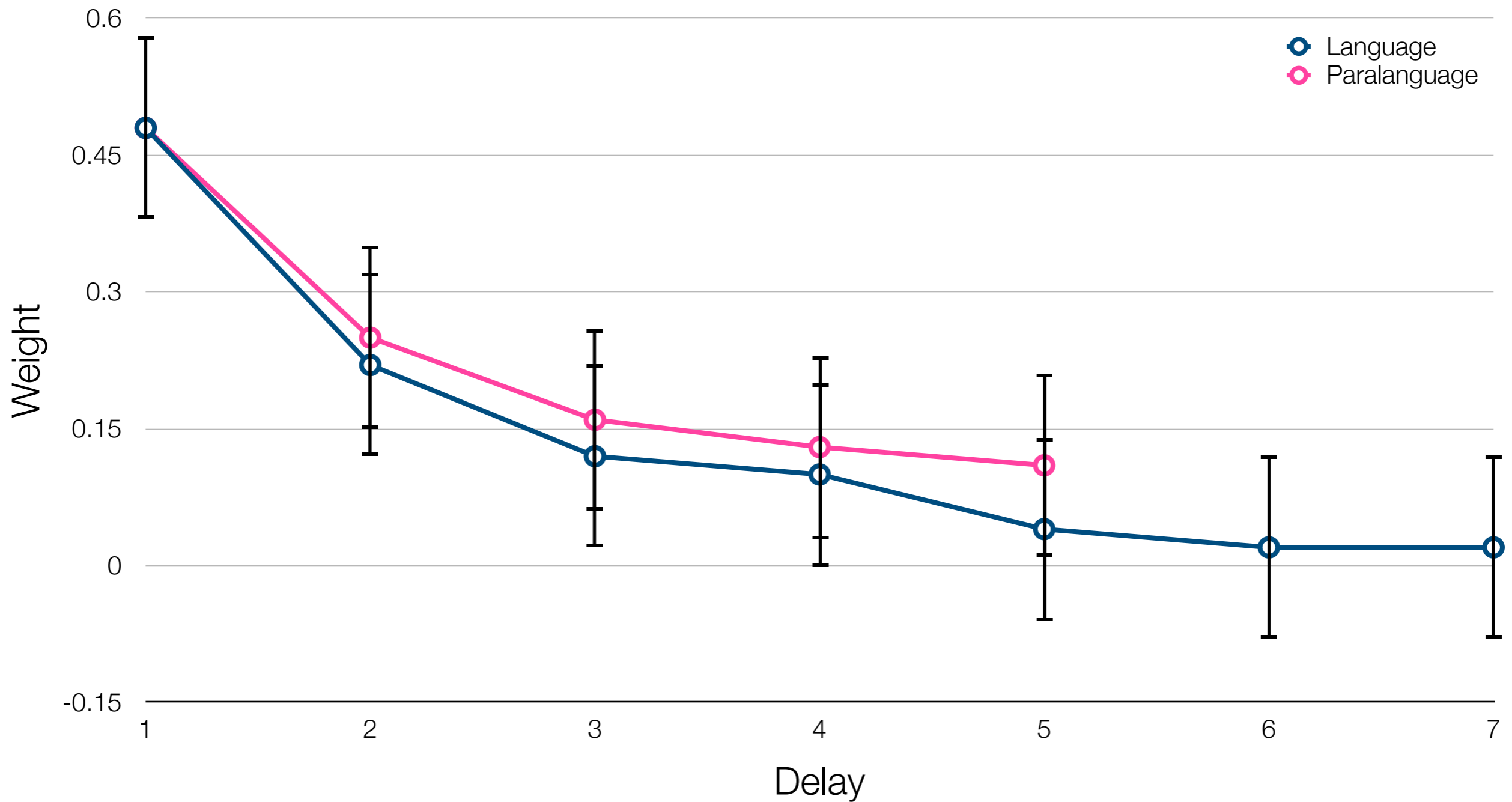
Global Weight Analysis (I)



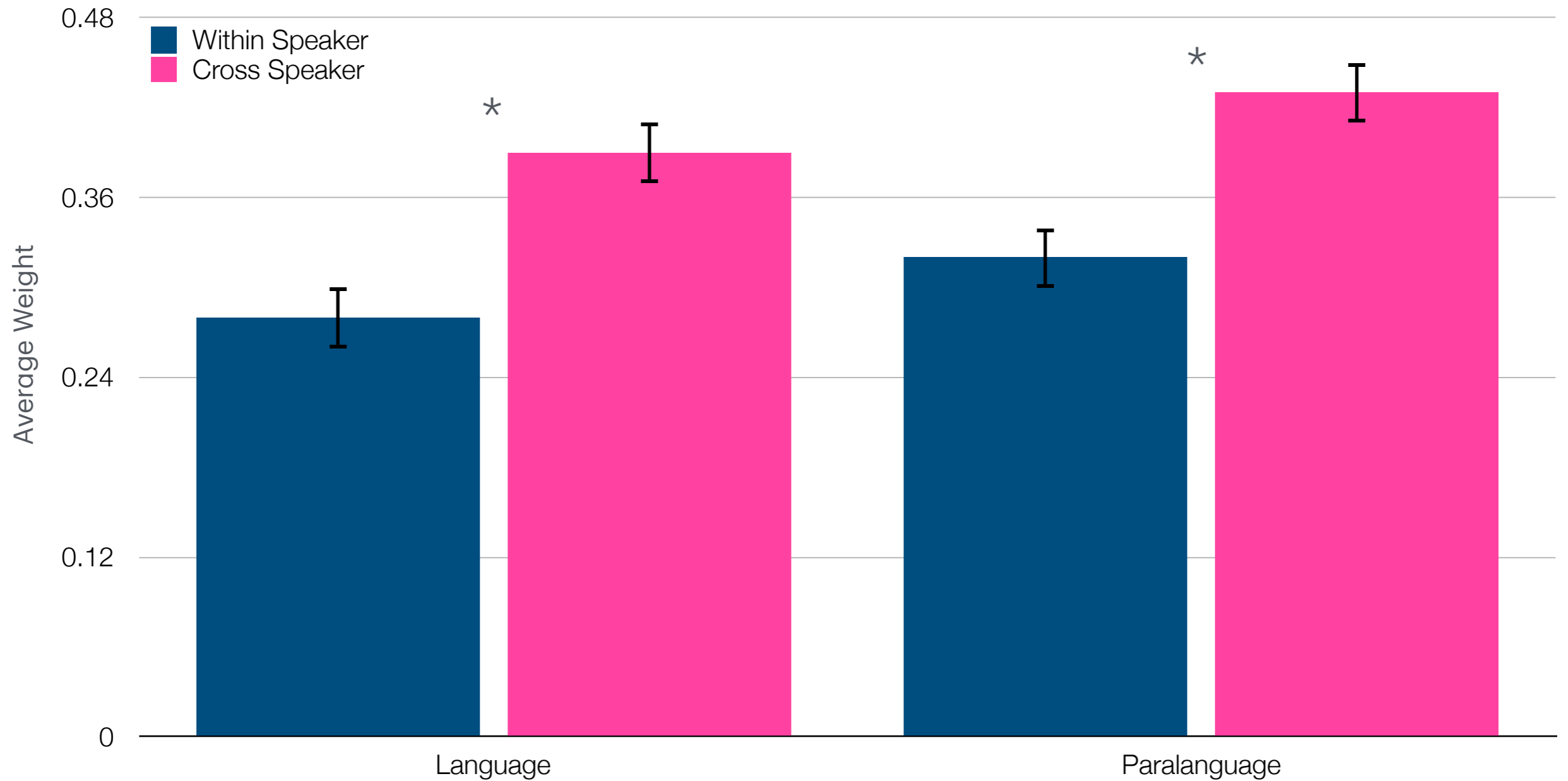
Global Weight Analysis (II)

- The utterances with the highest global weights (“**peak**” utterances) tend to concentrate **early when there is agreement**;
- Vice versa, **peak utterances** tend to concentrate **late** when there is **high disagreement percentage**;
- In the latter case, the model needs to wait the end of the conversation.

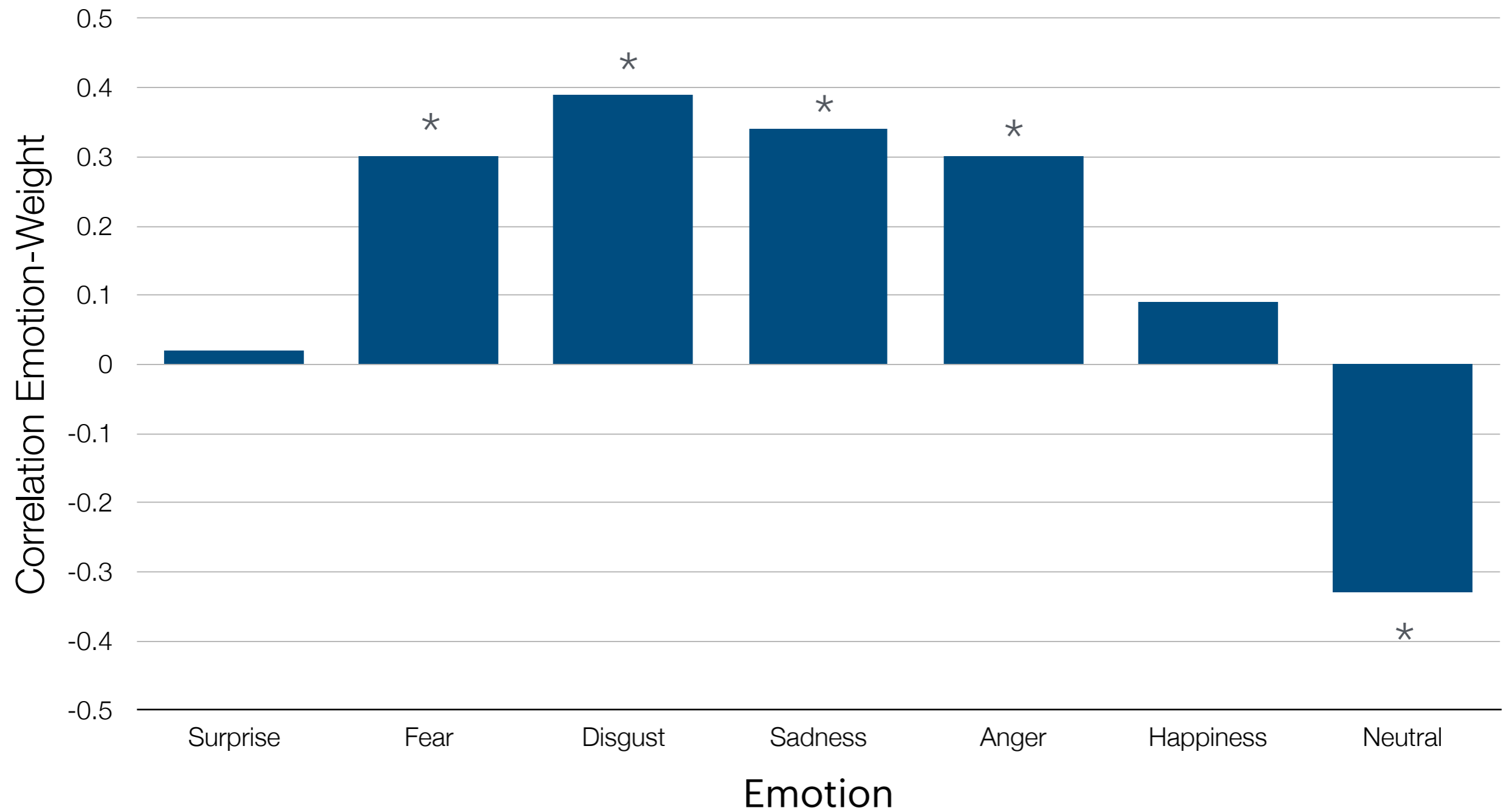
Local Weight Analysis (I)



Local Weight Analysis (II)



Local Weight Analysis (III)



Local Weight Analysis (IV)

- The **closest previous utterance weights**, on average, **half of the total** in the local representations;
- On average, **Cross Speaker weights are greater**, to a statistically significant extent, **than Within Speaker ones**;
- What a speaker says (and how s/he says it) **depends on** what **the other** says (and on how s/he says it);
- Local weights **correlate significantly with the emotional content** (especially negative) of the utterances.

Disagreement

“[...] an **oppositional stance** (verbal or non-verbal) to an antecedent verbal (or non-verbal) **action**.”

Outline

- Why Conversations
- Example: Disagreement Analysis
- Conclusions

Conclusions

- The experiments would not be possible without **knowledge about the psychological phenomenon** (disagreement in this case);
- Psychology shapes the **collection of the data**, a stage that is **part of the research effort**;
- The approach bridges the gap between **observable behaviour** (physical layer) and **its interpretation** (inferential layer).

Thank You!