

# Towards a framework for empirical measurement of conceptualization qualities

**Sotirios Liaskos**

York University

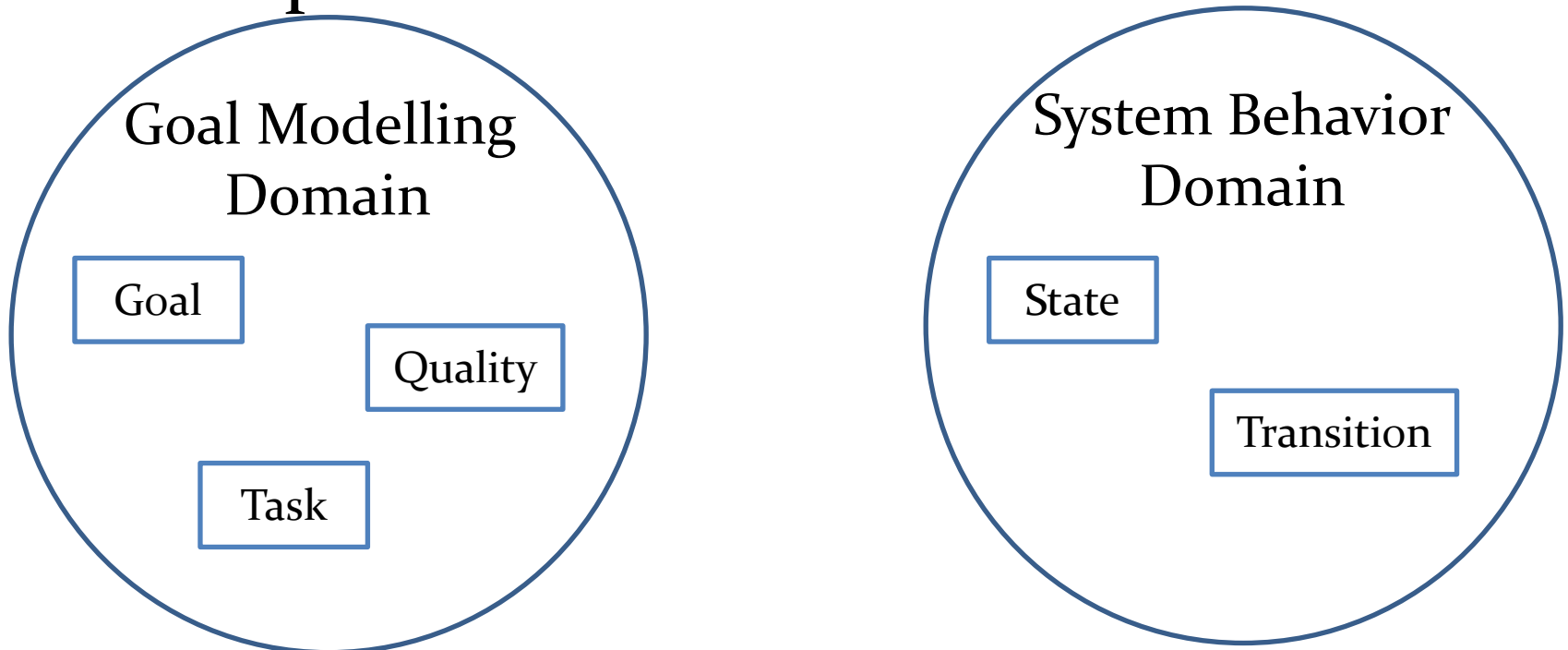
**Ibrahim Jaouhar**

York University

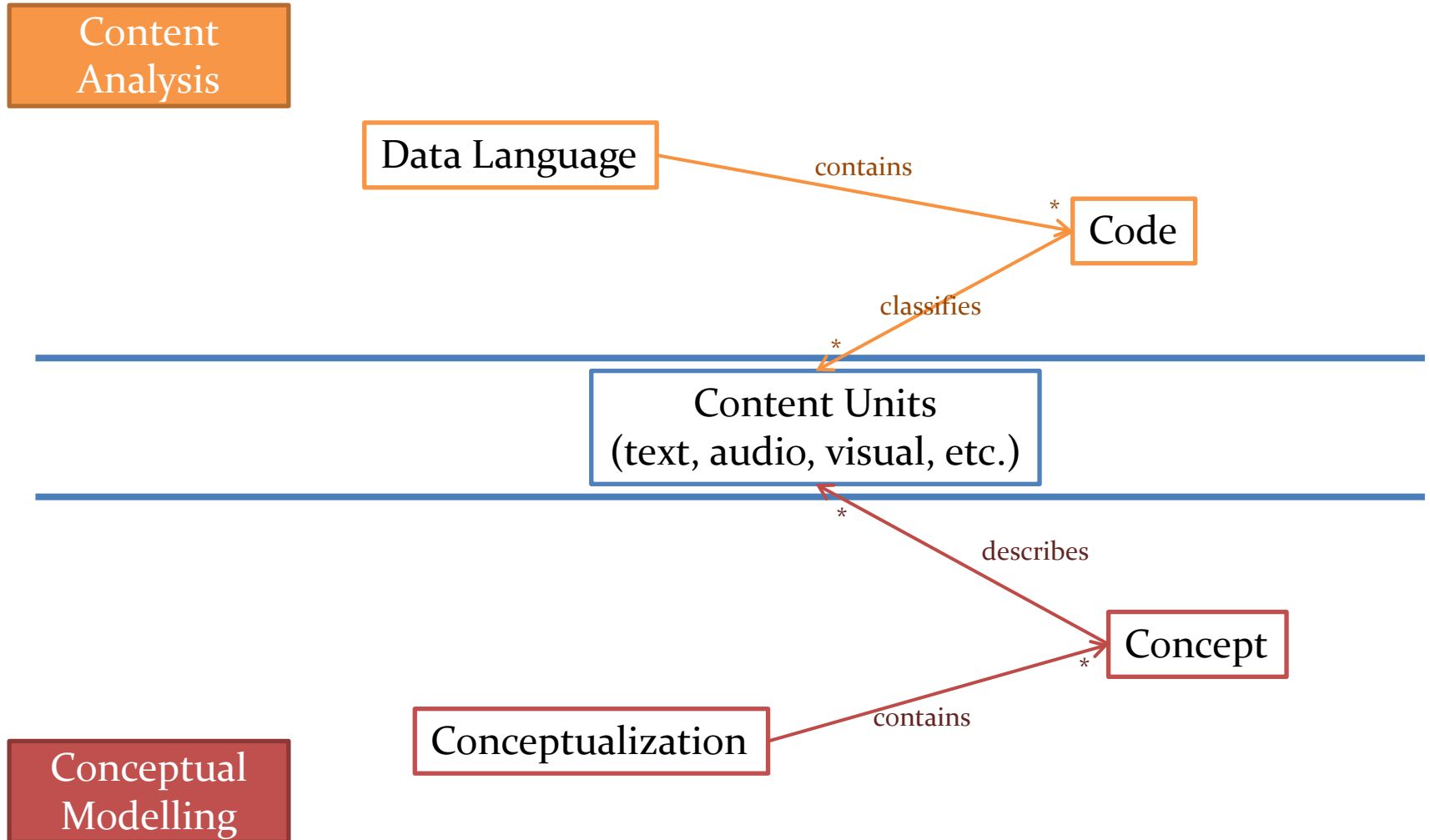


# Conceptualizations

- Used to articulate abstraction of states of affairs in a given domain.
- Examples:



# Content Analysis vs. Conceptual Modelling



# Main Argument

- The act of modelling can be seen as an act of qualitatively analyzing content taken from the domain.

*thus*

- Measures of quality of content analysis should be applicable to measuring quality of a modelling process (and the language used in that process).

# Main Argument

- *Reliability is an* indicator of good quality content analysis.
  - Different people under different circumstances would still code (classify) content units to the same codes.
  - OR: something wrong with data language, training, process etc.
- For conceptual modelling:
  - Different **modellers** under different circumstances would still **describe in the model** content units using the same **concepts**.
  - OR: : something wrong with **conceptualization**, training, process etc.

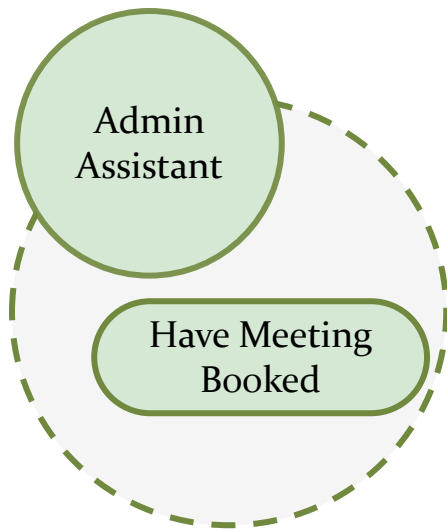
# Example

- Content Unit:
  - “The admin assistant wants to have a meeting booked.”
- Conceptualization: “actor”, “goal”, “belief”

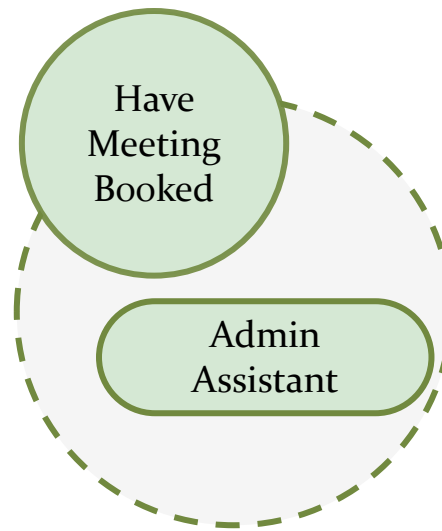


# Example

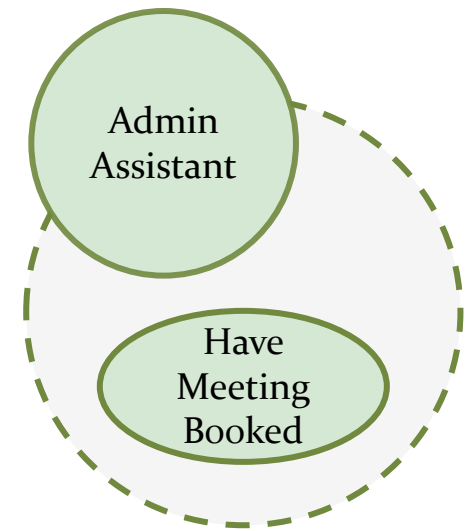
- Content Unit:
  - “The admin assistant wants to have a meeting booked.”
- Conceptualization: “actor”, “goal”, “belief”



**Language Designer:**  
[Admin Assistant]<sub>actor</sub>  
[Have Meeting Booked]<sub>goal</sub>



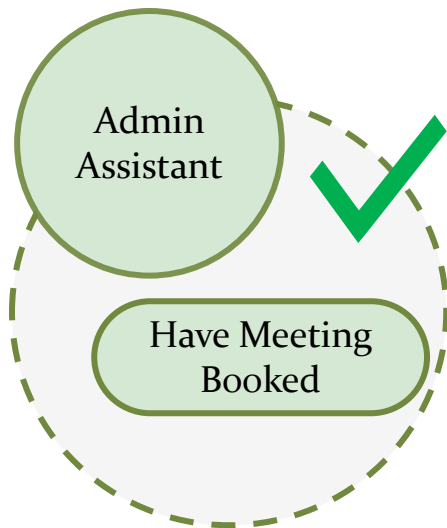
**Modeller 1:**  
[Admin Assistant]<sub>goal</sub>  
[Have Meeting Booked]<sub>actor</sub>



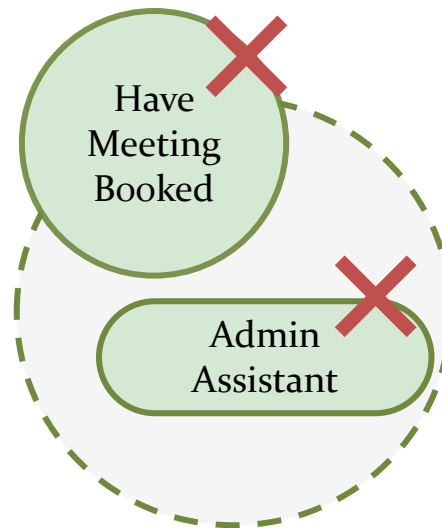
**Modeller 2:**  
[Admin Assistant]<sub>actor</sub>  
[Have Meeting Booked]<sub>belief</sub>

# Example

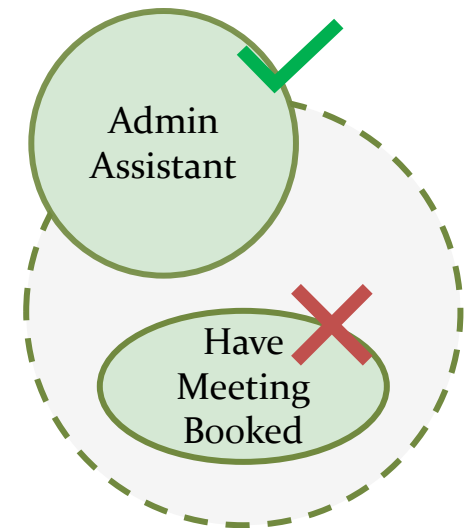
- Content Unit:
  - “The admin assistant wants to have a meeting booked.”
- Conceptualization: “actor”, “goal”, “belief”



**Language Designer:**  
[Admin Assistant]<sub>actor</sub>  
[Have Meeting Booked]<sub>goal</sub>



**Modeller 1:**  
[Admin Assistant]<sub>goal</sub>  
[Have Meeting Booked]<sub>actor</sub>

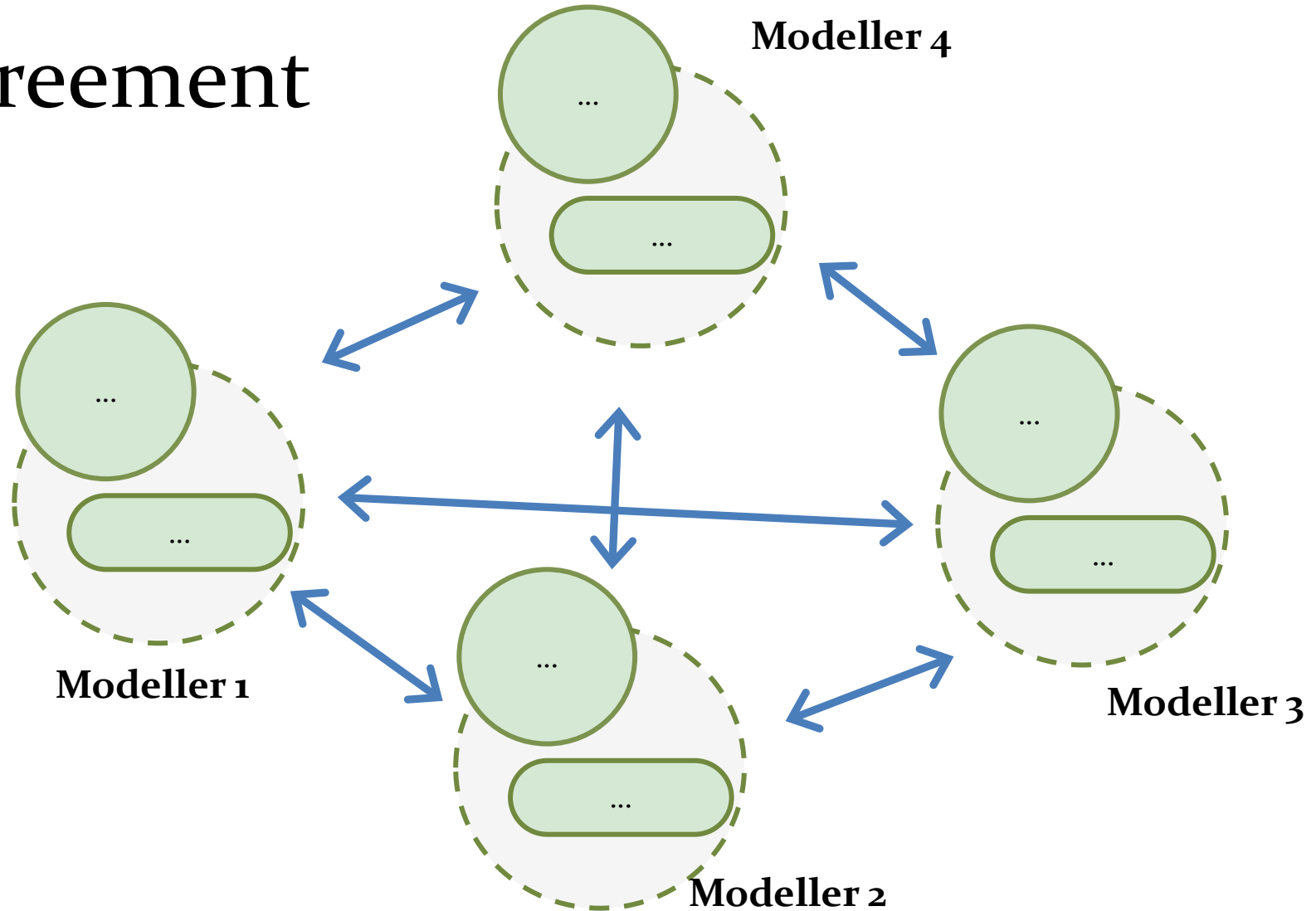


**Modeller 2:**  
[Admin Assistant]<sub>actor</sub>  
[Have Meeting Booked]<sub>belief</sub>



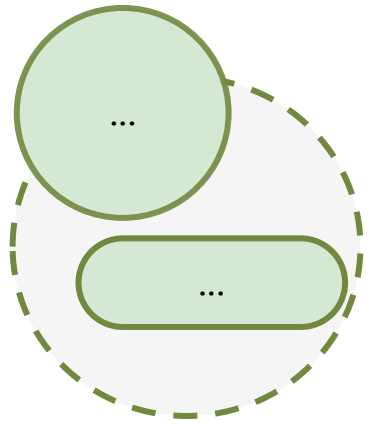
# Reliability Measures

## Agreement

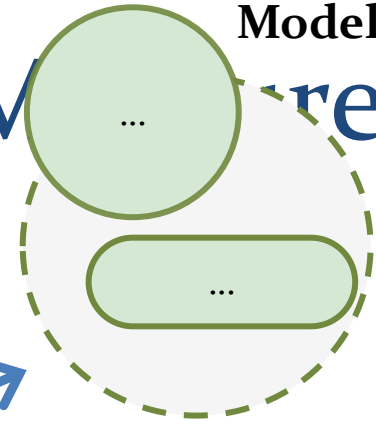


# Reliability Measures

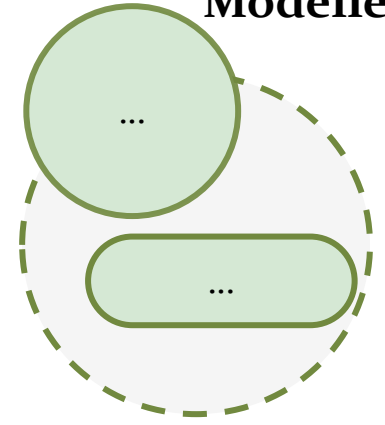
Accuracy



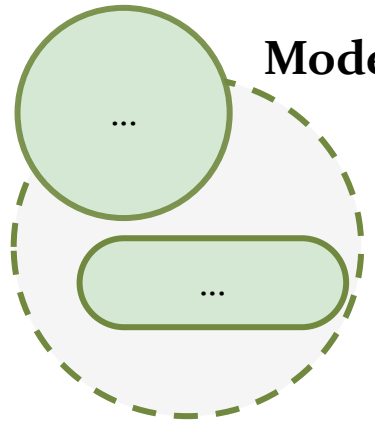
Language Designer



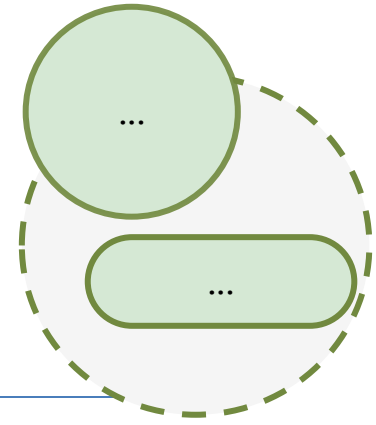
Modeller



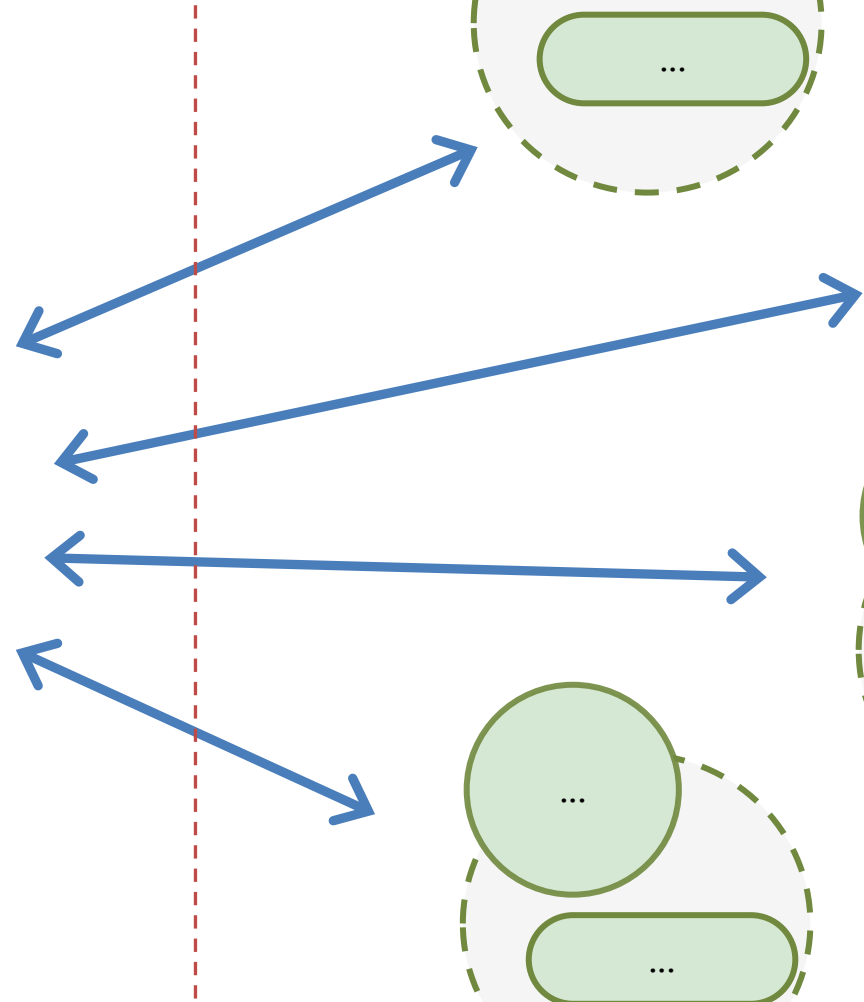
Modeller



Modeller



Modeller



# Measuring Reliability

- Assume a candidate conceptualization.
  - A set of concepts  $o \in O$
- Collect a sample of domain content units.
  - A set of *expressions*  $l \in L$
- Gather a sample of participants from the modeler (= language user) population.
  - A set of participants  $s \in S$
- Ask *participants* to classify *expressions* to *concepts*.
- Measure:
  - Agreement
  - Accuracy
  - Overlap

# Operationalizations

$f(l, o)$	: proportion of respondents that classified content unit $l$ into concept $o$ .
$a(l)$	: proportion of respondents that classified content unit $l$ in agreement with designers.
$O$	: set of concepts in conceptualization.
$L$	: set of all sampled content units (“expressions”).

## Measures per expression $l \in L$

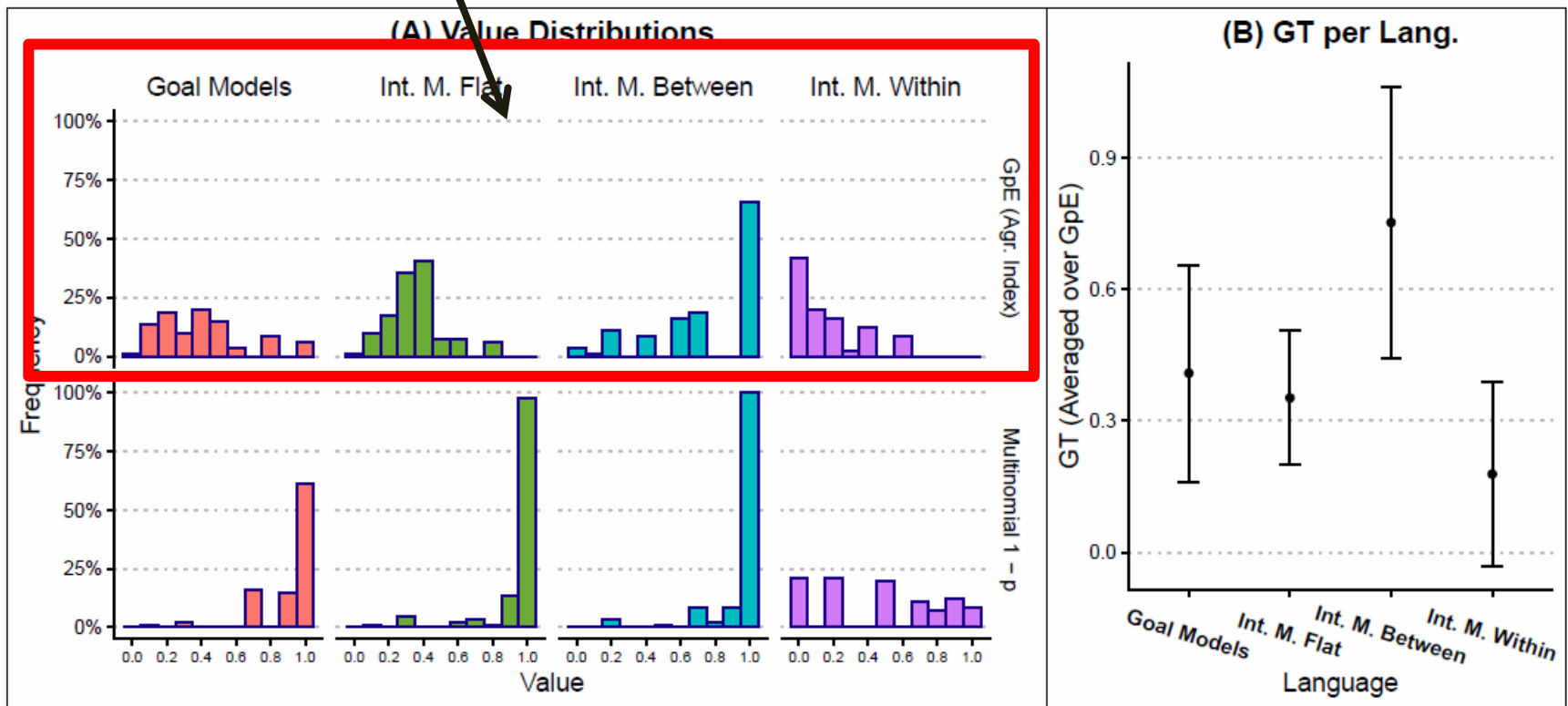
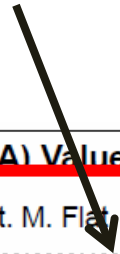
Agreement	Accuracy	Overlap between $o_1$ and $o_2$
$\frac{\sum_{o_i \in O} f(l, o_i) - 1/ O }{1/ O }$	$\frac{a(l) - 1/ O }{1/ O }$	$\frac{\min\{f(l, o_1), f(l, o_2)\}}{\max\{f(l, o_1), f(l, o_2)\}} [f(l, o_1) + f(l, o_2)]$

# Case Study

- Two conceptualizations
- Goal Models: *goal, quality, task, belief*
- “Intention Models”: *goal, objective, claim, assertion*
  - Manufactured so that concepts are synonyms in pairs.
- Sample content units from:
  - iStar 2.0 Guide
  - Archimate<sup>®</sup> 3.1 Specification
- 41 participants to *classify* to content units.
  - Between subjects wrt. conceptualization.

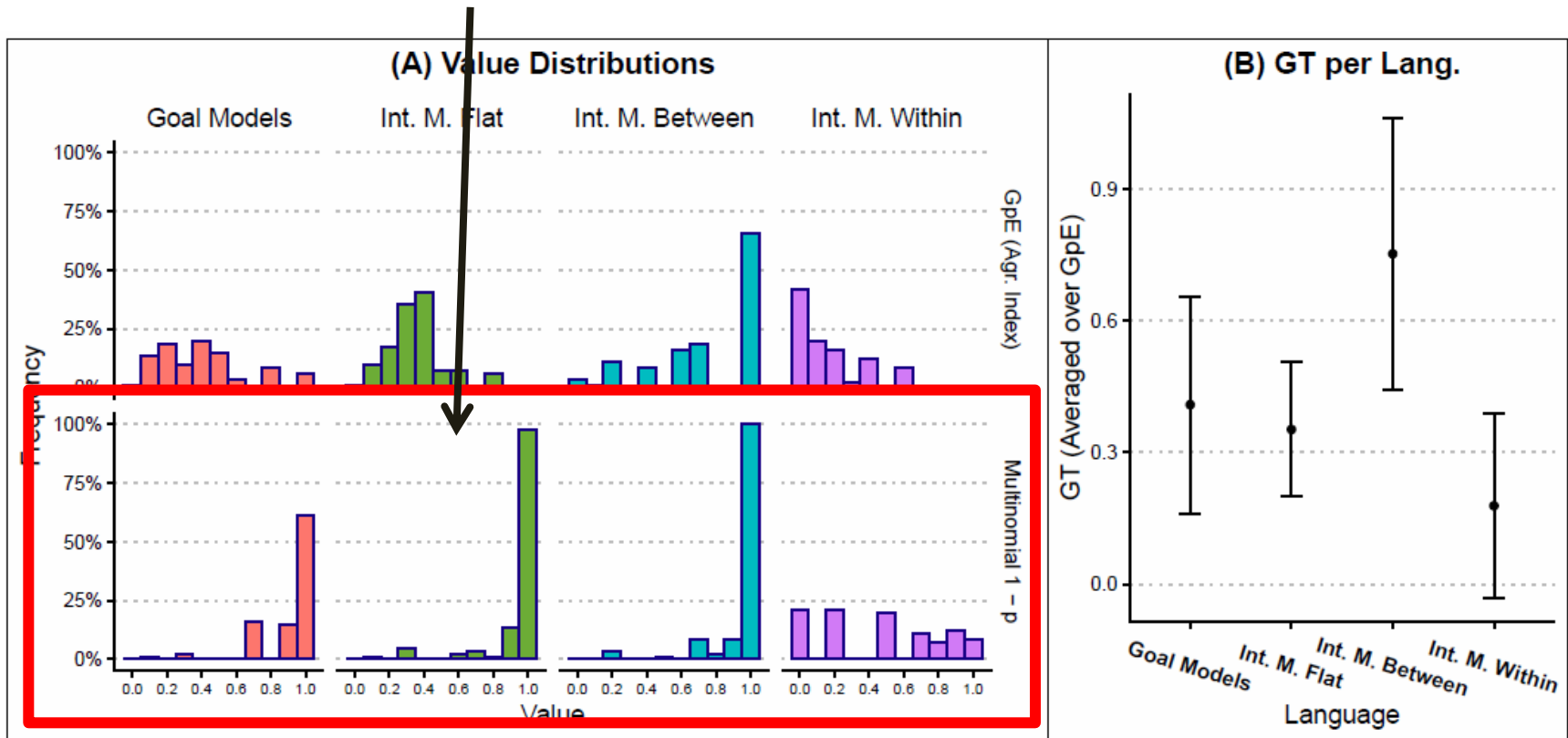
# Case Study

## Agreement histograms



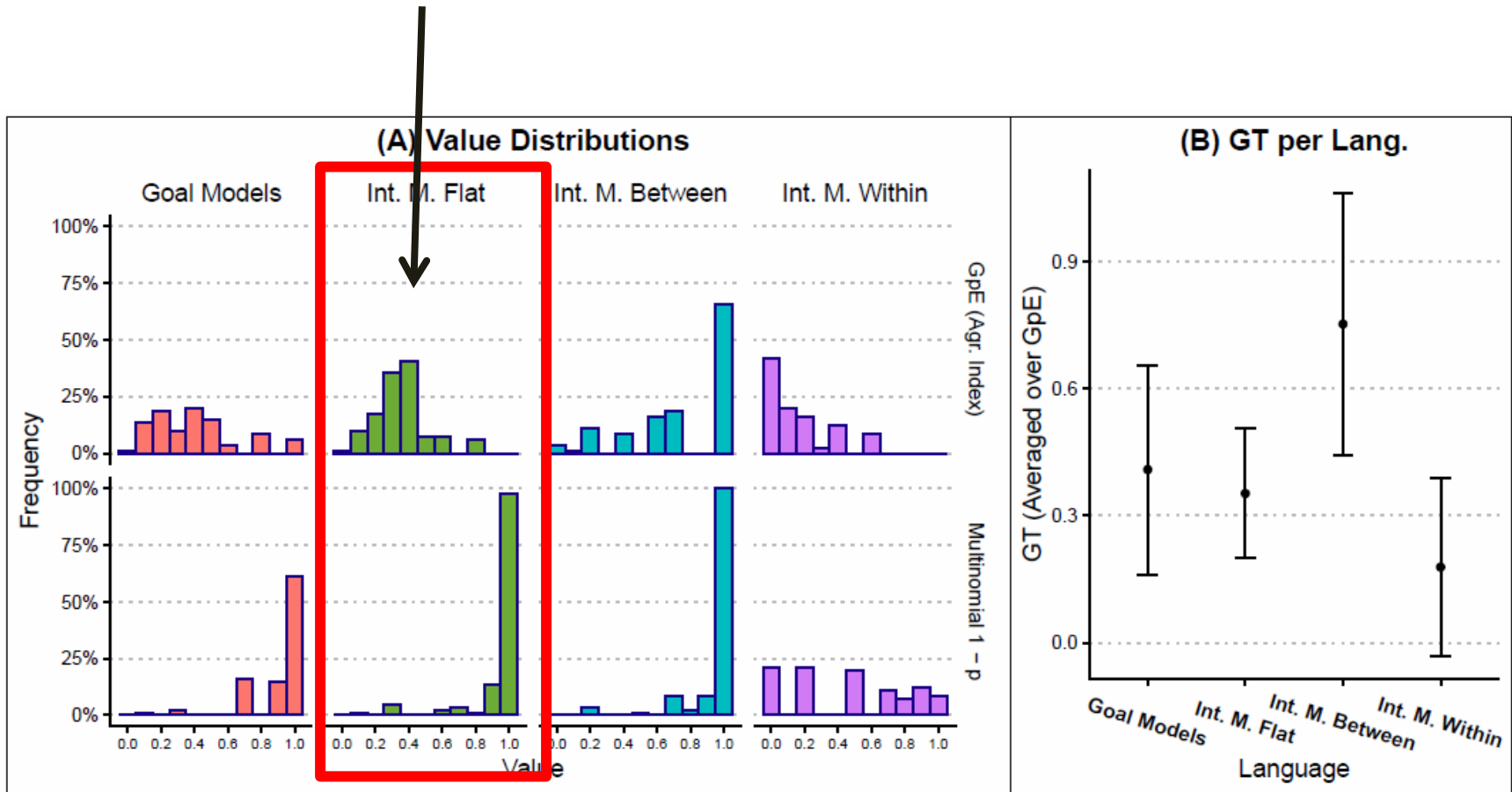
# Case Study

Likelihood that responses are not random.



# Case Study

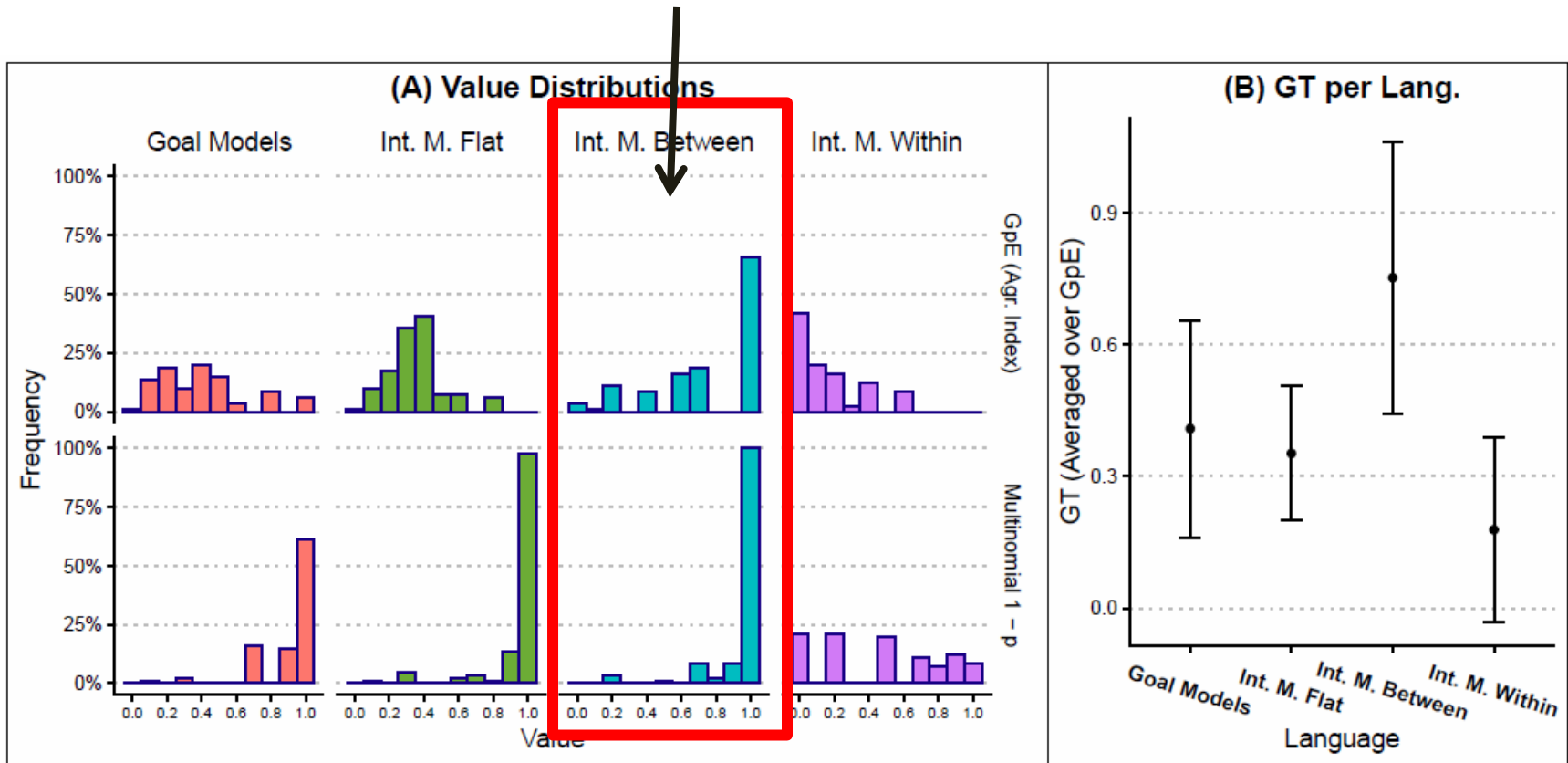
## Intention models “as-is”





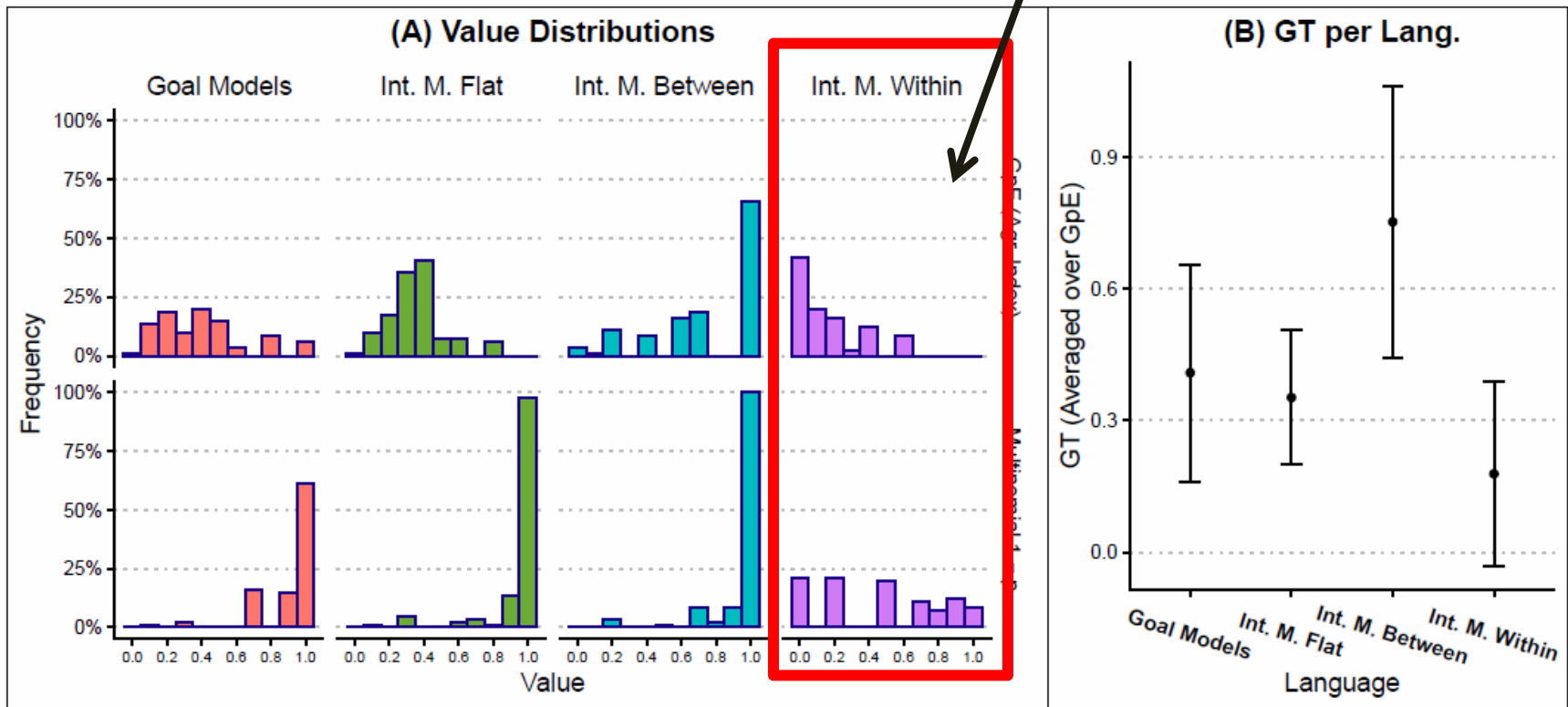
# Case Study

Intention models after merging {goal, objective} into {goal} and {claim, assertion} into {claim}.



# Case Study

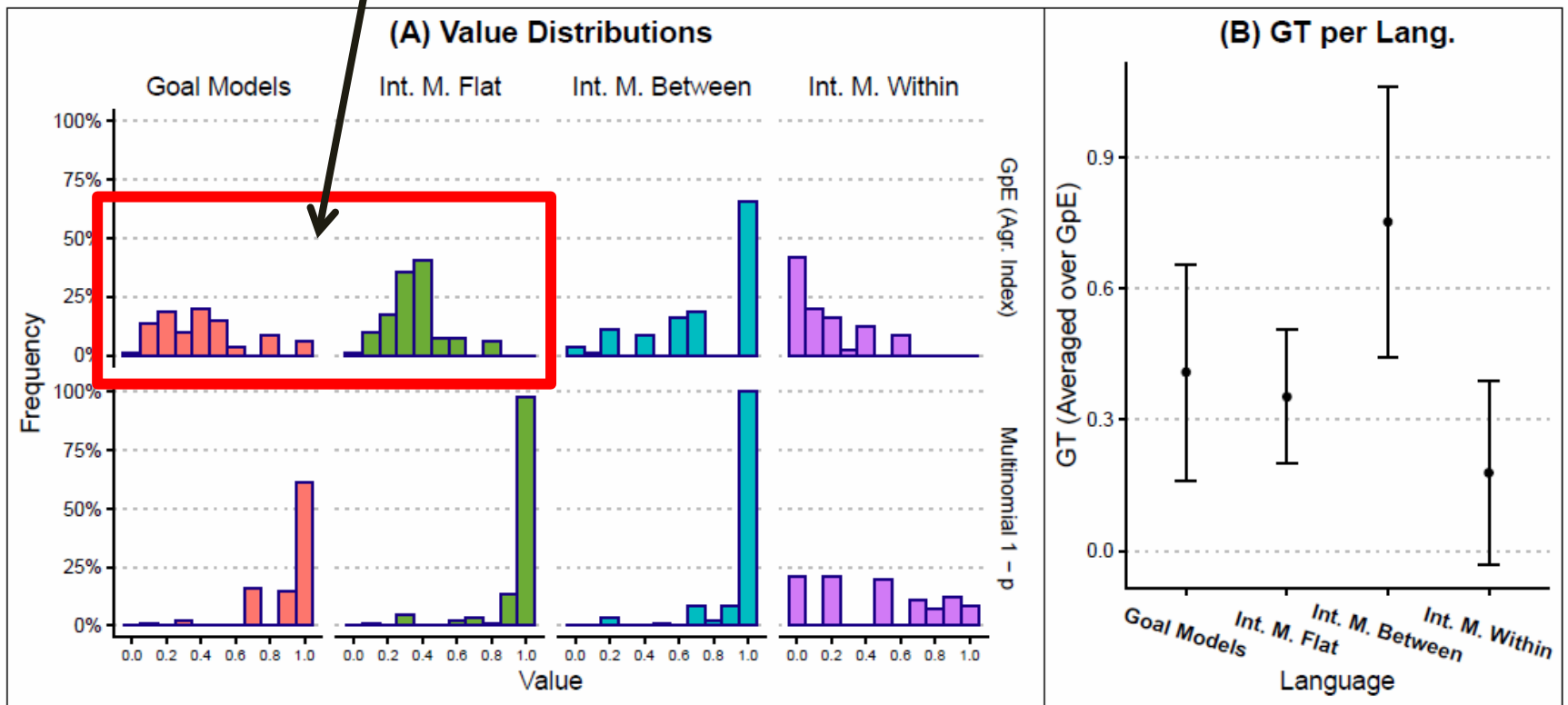
Pairs {goal, objective} and {claim, assertion} as separate languages.



# Case Study

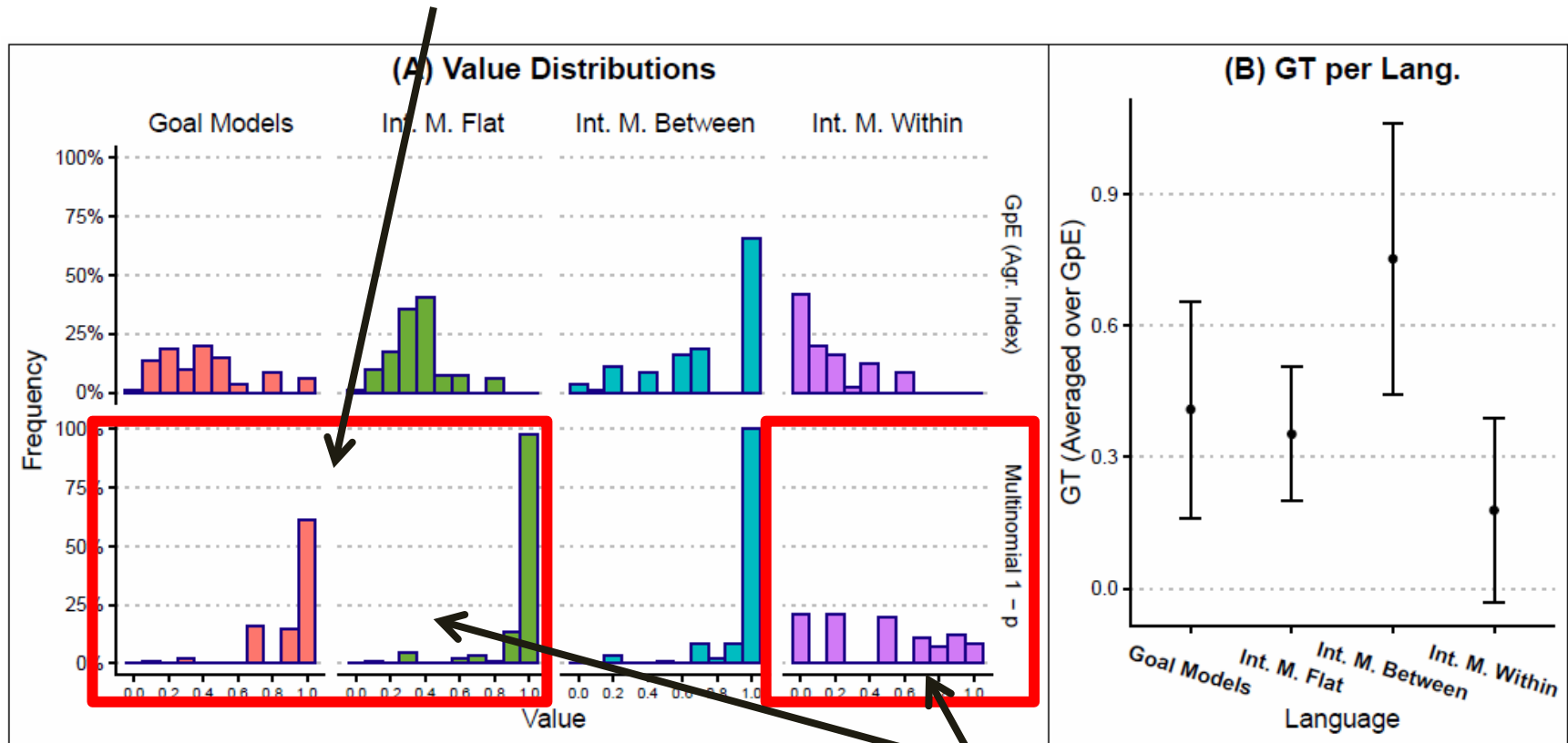
Agreement low due to overlaps

BUT...



# Case Study

... disagreement is systematic  
(responses concentrate on subset)



.. compare with this..

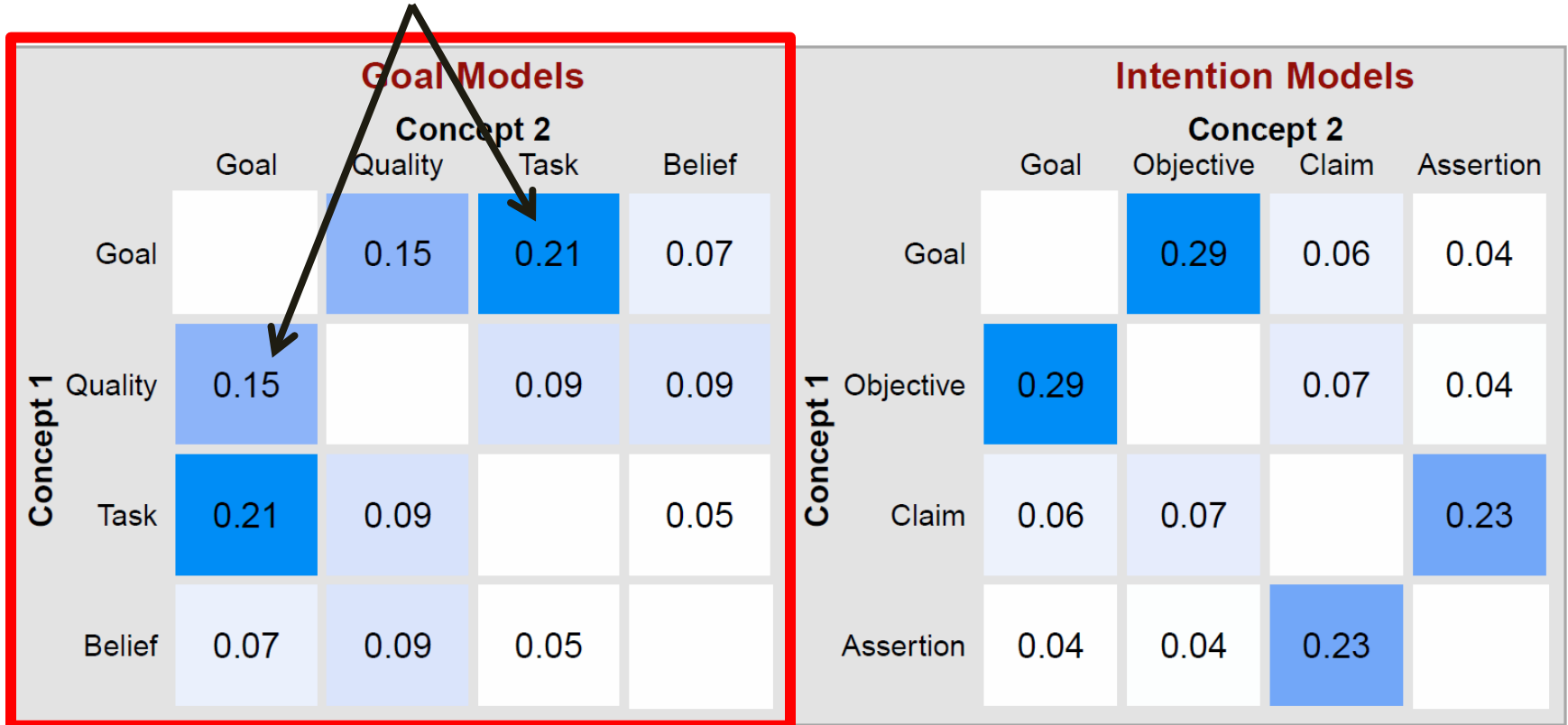
# Case Study

## “Concept Overlap Maps”

		Goal Models				Intention Models				
		Concept 2				Concept 2				
		Goal	Quality	Task	Belief	Goal	Objective	Claim	Assertion	
Concept 1	Goal		0.15	0.21	0.07		0.29	0.06	0.04	
	Quality	0.15		0.09	0.09	0.29		0.07	0.04	
	Task	0.21	0.09		0.05	0.06	0.07		0.23	
	Belief	0.07	0.09	0.05		0.04	0.04	0.23		

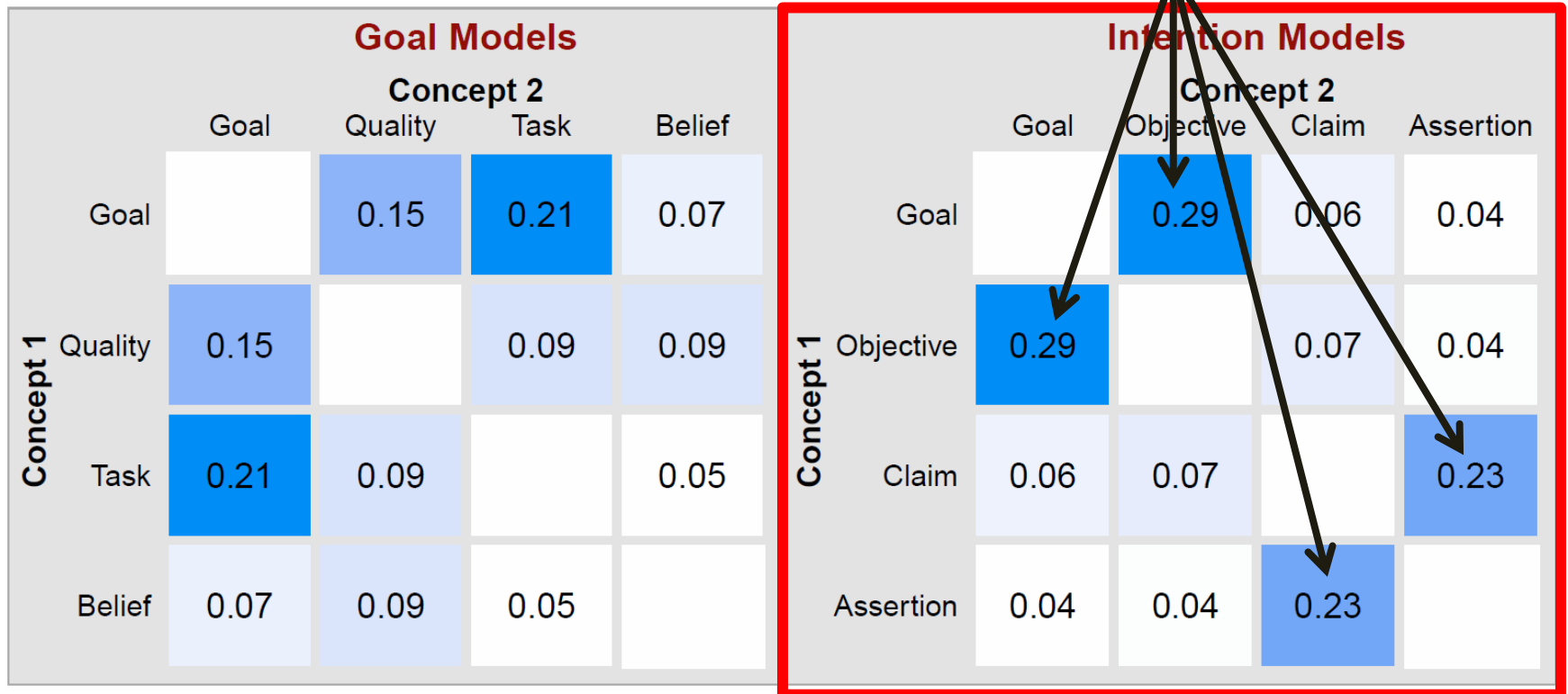
# Case Study

Overlaps between Goals and Qualities and (primarily) Goals and Tasks. Quite Expected.



# Case Study

Contrived overlaps captured  
by metric.



# Future Work

- Compare alternative operationalizations.
- Connect with established comprehensibility appropriateness measures:
  - *lucidity, laconicity, soundness and completeness*
- Place metrics within established quality frameworks.



# Thank you!