Truth conditions at scale, and beyond

Guy Emerson

What I'll Cover...

- Truth conditions at scale
  - Learning from text, images, ontologies
  - Generalisation

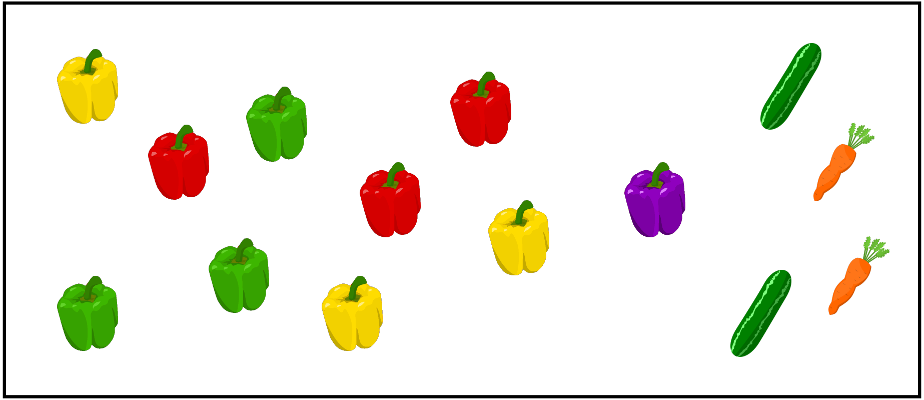
What I'll Cover...

- Truth conditions at scale
  - Learning from text, images, ontologies
  - Generalisation
- Beyond truth conditions
  - (In)tractability of inference
  - A new kind of probabilistic model

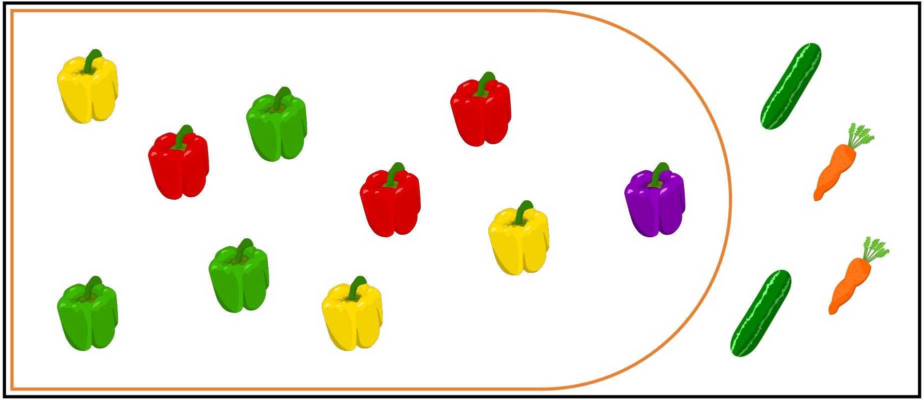
Truth-Conditional Semantics

"the meaning of a sentence is the method of its verification… there is some uniform means of deriving all the other features of the use of any sentence from this one feature" — Dummett (1976)

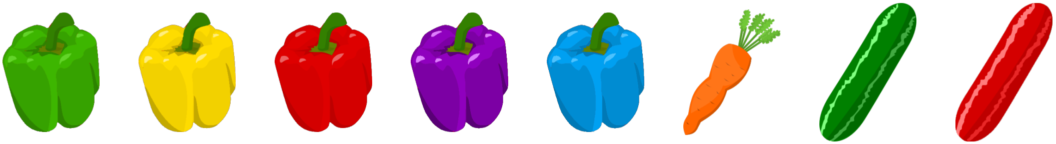
Truth-Conditional Semantics



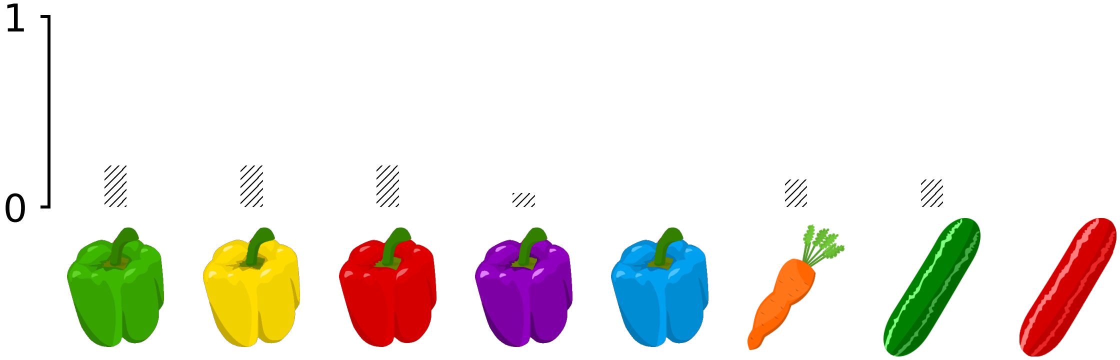
Truth-Conditional Semantics



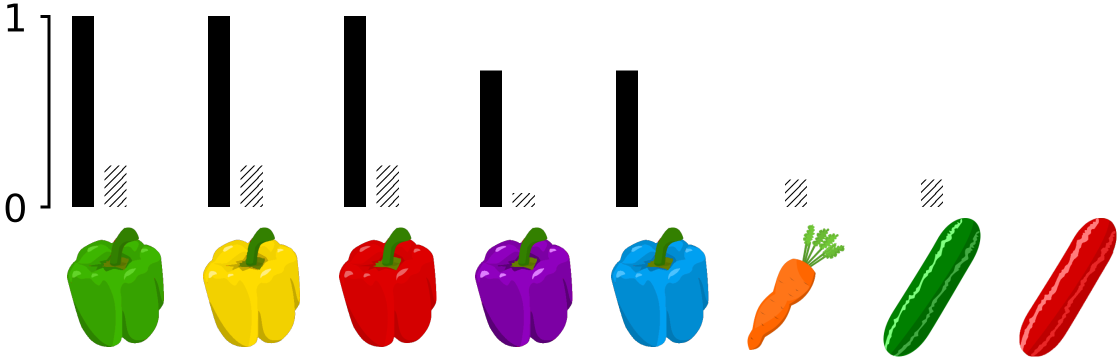
Truth-Conditional Functions



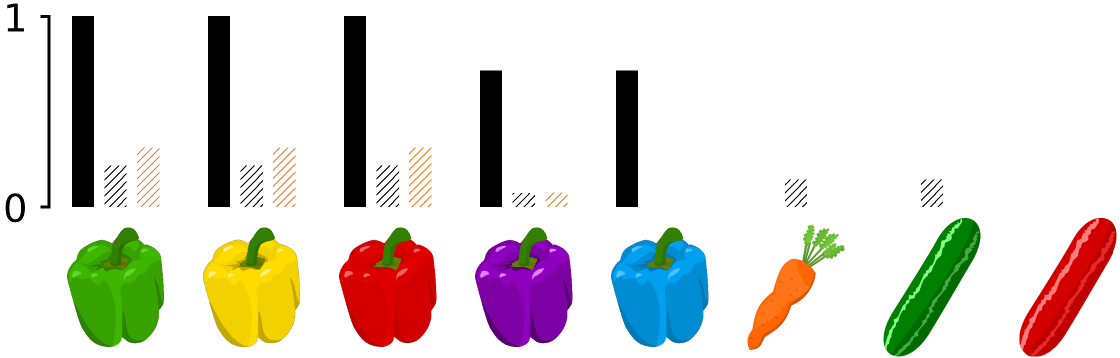
Truth-Conditional Functions



Truth-Conditional Functions



Truth-Conditional Functions



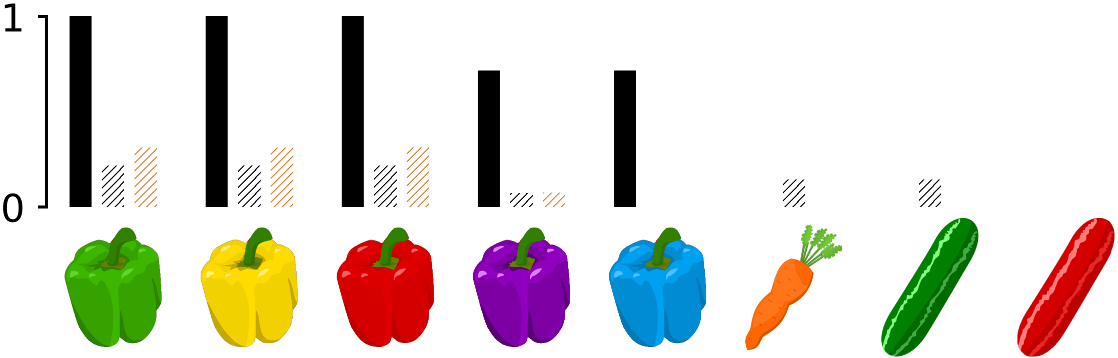
Interim Summary

- Predicates as functions:
  entity representation $\mapsto$ probability of truth

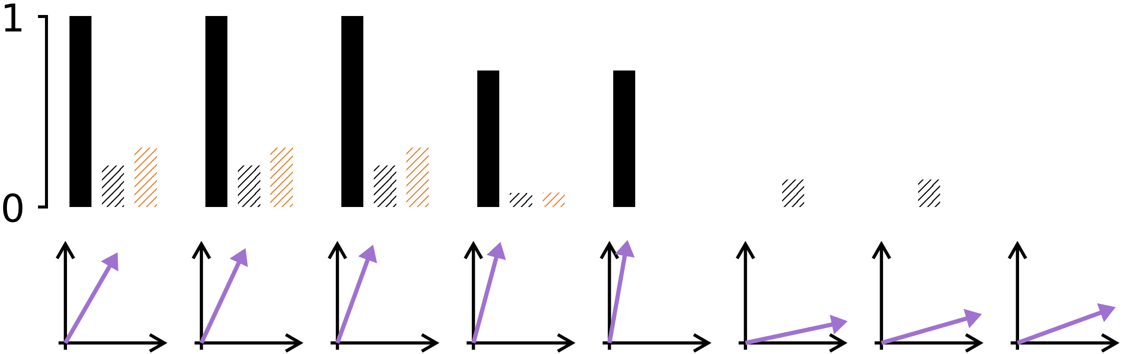
Interim Summary

- Predicates as functions:
  pixie $\mapsto$ probability of truth
- Pixie: entity representation

Truth-Conditional Functions



Truth-Conditional Functions



Situation Semantics

x

pepper(x)

Situation Semantics

x

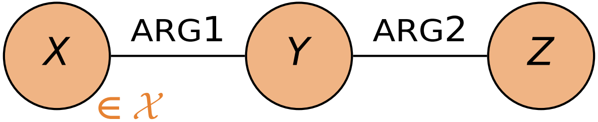
pepper(x)
vegetable(x)
animal(x)
dog(x)
cat(x)

Situation Semantics

$$x \xleftarrow{\text{ARG1}} y \xrightarrow{\text{ARG2}} z$$

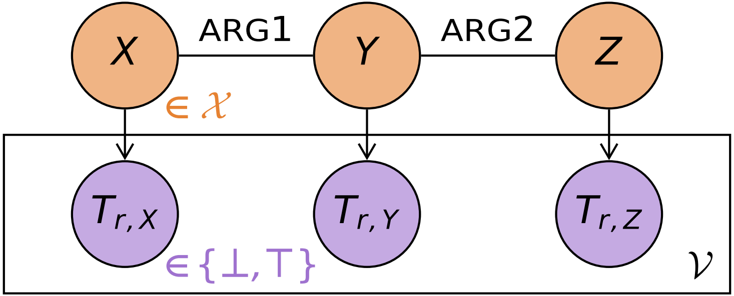
dog(x)	chase(y)	cat(z)
animal(x)	pursue(y)	animal(z)
chase(x)	dog(y)	chase(z)
pursue(x)	cat(y)	pursue(z)
cat(x)	animal(y)	dog(z)

Probabilistic Situation Semantics



dog(X)	chase(Y)	cat(Z)
animal(X)	pursue(Y)	animal(Z)
chase(X)	dog(Y)	chase(Z)
pursue(X)	cat(Y)	pursue(Z)
cat(X)	animal(Y)	dog(Z)

Probabilistic Situation Semantics



Probabilistic Situation Semantics

- World model: $\mathbb{P}(x, y, z)$
- Lexical truth-conditional model: $\mathbb{P}(t_{r,X} \mid x)$

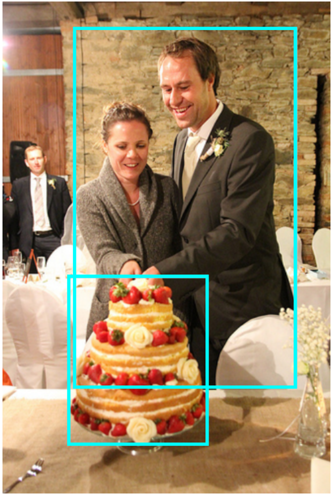
Probabilistic Situation Semantics

- World model: $\mathbb{P}(x, y, z)$
- Lexical truth-conditional model: $\mathbb{P}(t_{r,X} \mid x)$
- Aim: learn these at scale!

Probabilistic Truth Conditions at Scale

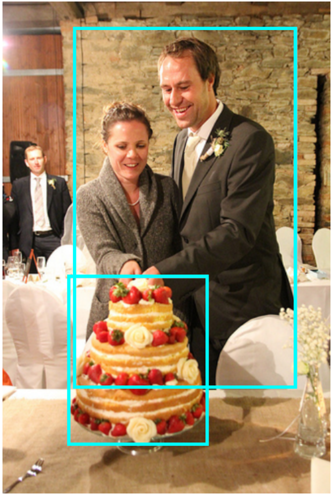
- Learn from:
  - Labelled images (Liu & Emerson, 2022)
  - Parsed text (Lo et al., 2023)
  - WordNet (Cheng et al., 2023)

Visual Genome (Krishna et al., 2017)

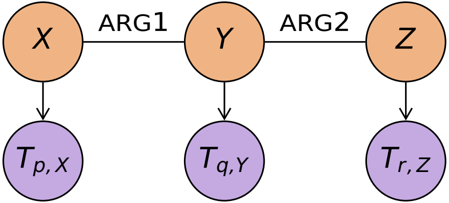


"couple cutting cake"

Visual Genome (Krishna et al., 2017)



“couple cutting cake”



Learning from Visual Genome

1. Data
2. Objective
3. Model
4. Training

Learning from Visual Genome

1. Data: 2.3m of form (, , , couple, cut, cake)

2. Objective

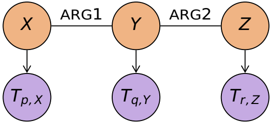
3. Model

4. Training

Learning from Visual Genome

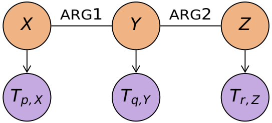
1. Data: 2.3m of form (, , , couple, cut, cake)
2. Objective: $\mathbb{P}$(, , , couple, cut, cake)
3. Model
4. Training

Learning from Visual Genome

1. Data: 2.3m of form (, , , couple, cut, cake)
2. Objective: $\mathbb{P}$ (, , , couple, cut, cake)
3. Model: 

4. Training

Learning from Visual Genome

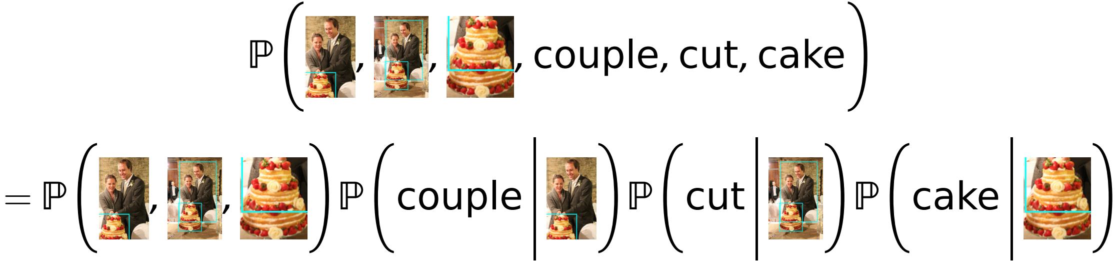
1. Data: 2.3m of form $\Big($ , , , couple, cut, cake $\Big)$
2. Objective: $\mathbb{P}\Big($ , , , couple, cut, cake $\Big)$
3. Model: 

4. Training: gradient descent

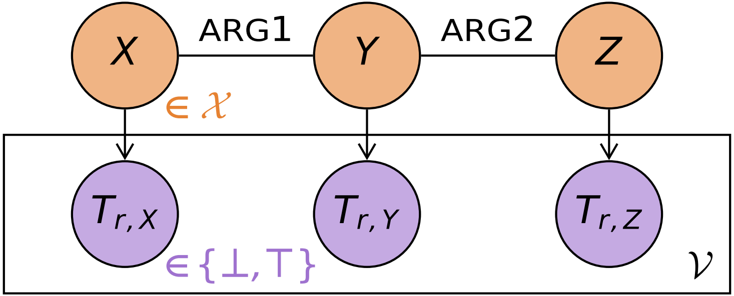
Probabilistic Situation Semantics

$\mathbb{P}\Big($ , , , couple, cut, cake $\Big)$

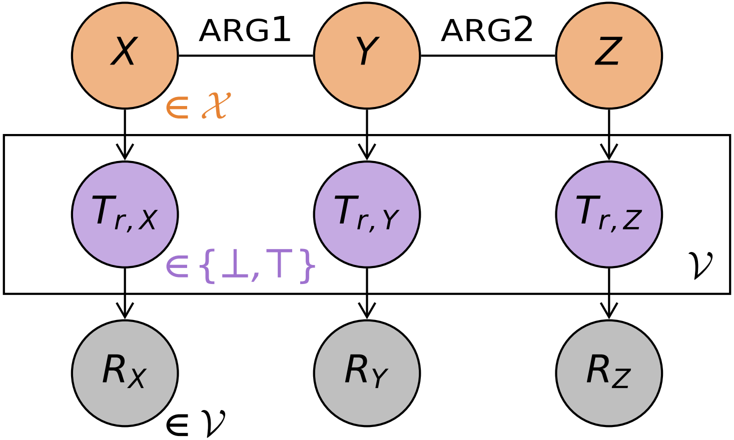
Probabilistic Situation Semantics



Probabilistic Situation Semantics



Functional Distributional Semantics



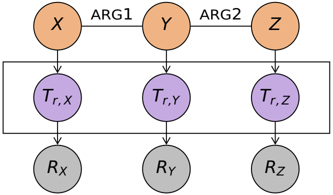
Functional Distributional Semantics

- World model $\mathbb{P}(x, y, z)$
- Lexical model $\mathbb{P}(t_{r,X} | x)$
- Extended lexical model $\mathbb{P}(r_X | x)$

Functional Distributional Semantics

- World model $\mathbb{P}(x, y, z)$
- Lexical model $\mathbb{P}(t_{r,X} | x)$
- Extended lexical model $\mathbb{P}(r_X | x) \propto \mathbb{P}(t_{r,X} | x)$

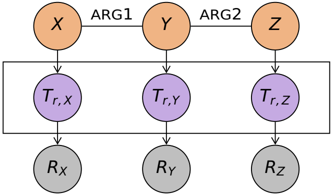
Truth Conditions from Images

1. Data: 2.3m of form (, , , couple, cut, cake)
2. Objective: $\mathbb{P}$ (, , , couple, cut, cake)
3. Model: 

4. Training: gradient descent

Truth Conditions from Text

1. Data: 36m of form $\left(\text{couple} \xleftarrow{\text{ARG1}} \text{cut} \xrightarrow{\text{ARG2}} \text{cake}\right)$
2. Objective: $\mathbb{P}\left(\text{couple} \xleftarrow{\text{ARG1}} \text{cut} \xrightarrow{\text{ARG2}} \text{cake}\right)$
3. Model:



4. Training: gradient descent

Truth Conditions from Text

- Only observe an utterance, not a situation...

Truth Conditions from Text

- Only observe an utterance, not a situation...
- Training objective:
  - $\mathbb{P}(u) = \sum_{s} \mathbb{P}(u \mid s)\,\mathbb{P}(s)$

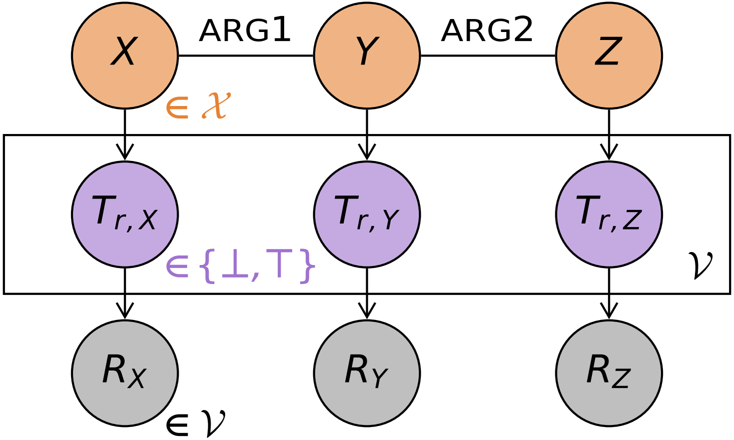
Truth Conditions from Text

- Only observe an utterance, not a situation...
- Training objective:
  - $\mathbb{P}(u) = \sum_{s} \mathbb{P}(u \mid s)\,\mathbb{P}(s)$
- Summing over all s is intractable!

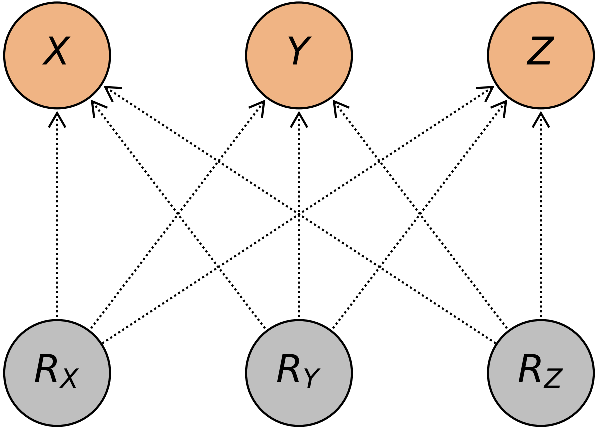
Truth Conditions from Text

- Only observe an utterance, not a situation...
- Training objective:
  - $\mathbb{P}(u) = \sum_{s} \mathbb{P}(u \mid s)\,\mathbb{P}(s)$
- Summing over all s is intractable!
- Approximation: only consider likely s

Functional Distributional Semantics



Amortised Variational Inference



Amortised Variational Inference

- Variational Inference: use a simple distribution to approximate $\mathbb{P}(s|u)$
- Amortised Variational Inference: train a neural net to approximately optimise the simple distribution

Amortised Variational Inference

- Variational Inference: use a simple distribution to approximate $\mathbb{P}(s \mid u)$

- Amortised Variational Inference: train a neural net to approximately optimise the simple distribution

- When applied to a latent-variable model, called a "Variational Autoencoder" (VAE)

Scaling Up

- Emerson (2020) “Autoencoding Pixies”
  - Semantic graphs with two or three predicates

Scaling Up

- Emerson (2020) “Autoencoding Pixies”
  - Semantic graphs with two or three predicates
- Lo et al. (2023) “Functional Distributional Semantics at Scale”
  - Arbitrary semantic graphs

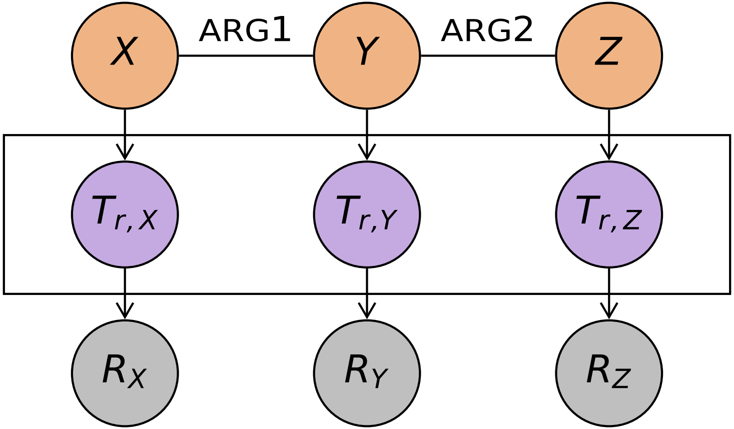
Scaling Up

- Emerson (2020) “Autoencoding Pixies”
  - Semantic graphs with two or three predicates
- Lo et al. (2023) “Functional Distributional Semantics at Scale”
  - Arbitrary semantic graphs
  - Inference network scales well
  - World model scales badly...

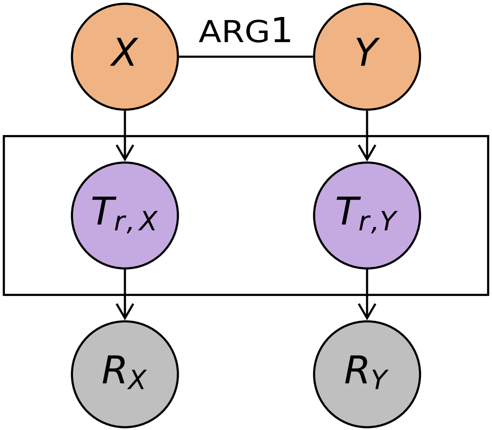
Scaling Up

- Emerson (2020) “Autoencoding Pixies”
  - Semantic graphs with two or three predicates
- Lo et al. (2023) “Functional Distributional Semantics at Scale”
  - Arbitrary semantic graphs
  - Inference network scales well
  - World model scales badly… remove it (!!!)

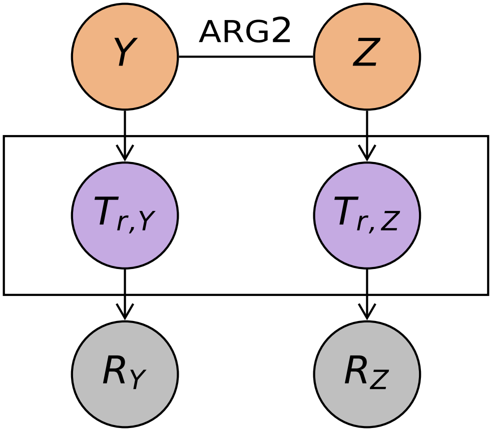
Family of Distributions



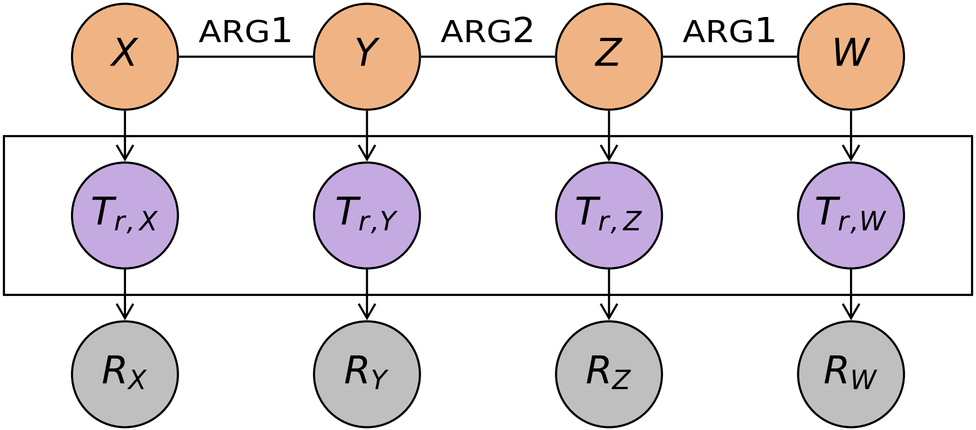
Family of Distributions



Family of Distributions



Family of Distributions



World Models at Scale

- World model is a family of distributions
- $P(s)$ must sum to 1
- Need to scale to many entities

World Models at Scale

- Emerson (2020): discrete vectors (RBM), normalisation constant intractable

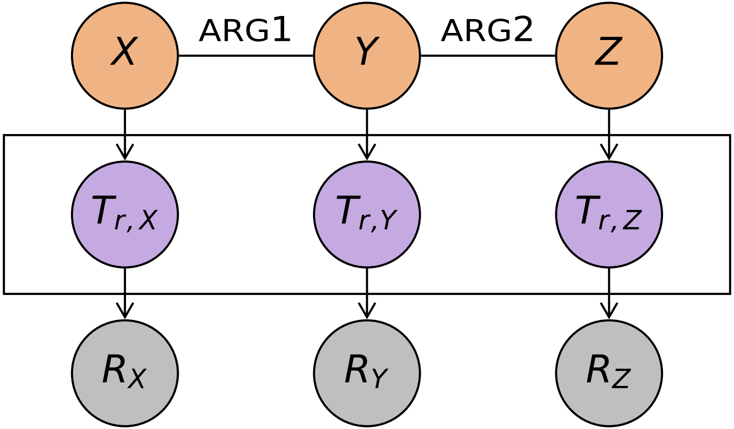
World Models at Scale

- Emerson (2020): discrete vectors (RBM), normalisation constant intractable
- Fabiani (2022), Liu & Emerson (2022): real vectors (Gaussian), normalisation constant scales as $O(n^3)$

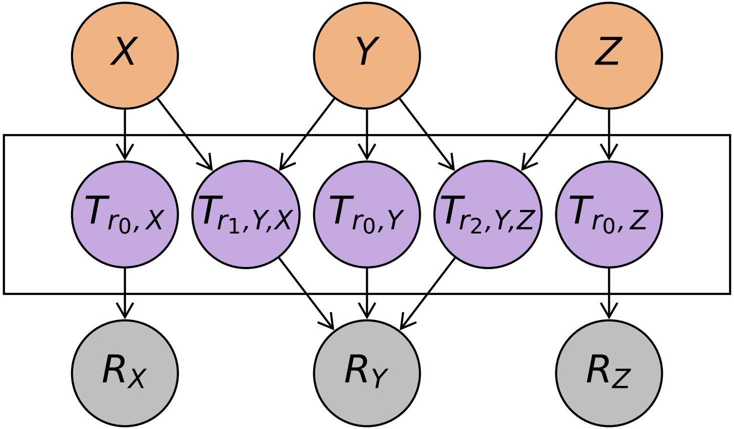
World Models at Scale

- Emerson (2020): discrete vectors (RBM), normalisation constant intractable
- Fabiani (2022), Liu & Emerson (2022): real vectors (Gaussian), normalisation constant scales as $O(n^3)$
- Lo et al. (2023): trivial world model, interactions moved to lexical model

Family of Distributions



Family of Distributions



Evaluating a Model

- Has the model learnt something useful?

Evaluating a Model

- Has the model learnt something useful?
- Can it generalise?

Evaluating a Model

- Has the model learnt something useful?
- Can it generalise?
  - Logical inference

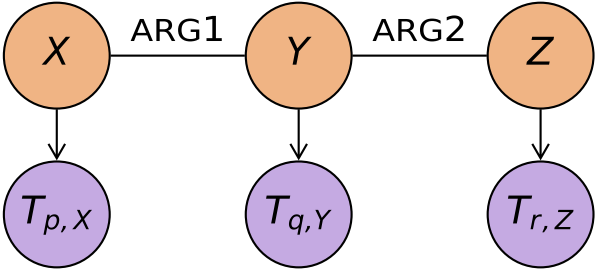
Logical Inference

- Is an animal that has a tail a cat?

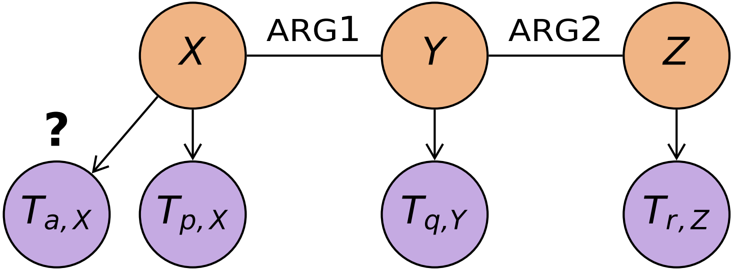
Logical Inference

- Is an animal that has a tail a cat?
- Is an animal that has a tail a computer?

Logical Inference with Latent Entities

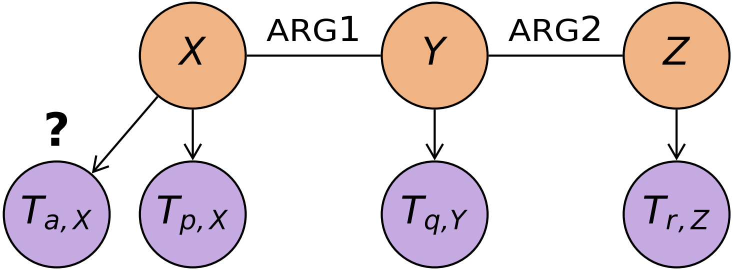


Logical Inference with Latent Entities



$$\mathbb{P}\left(t_{a,X} \,\middle|\, t_{p,X}, t_{q,Y}, t_{r,Z}\right)$$

Logical Inference with Latent Entities



$$\mathbb{P}\left(t_{a,X} \,\middle|\, t_{p,X}, t_{q,Y}, t_{r,Z}\right)$$

$$\mathbb{P}\left(t_{cat,X} \,\middle|\, t_{animal,X}, t_{have,Y}, t_{tail,Z}\right)$$

Variational Inference for Logical Inference

$$\mathbb{P}\left(t_{cat,X} \mid t_{animal,X}, t_{have,Y}, t_{tail,Z}\right)$$

Variational Inference for Logical Inference

$$\mathbb{P}\left(t_{cat,X} \,|\, t_{animal,X}, t_{have,Y}, t_{tail,Z}\right)$$

$$= \sum_{x,y,z} \mathbb{P}\left(t_{cat,X} \,|\, x\right) \mathbb{P}\left(x, y, z \,|\, t_{animal,X}, t_{have,Y}, t_{tail,Z}\right)$$

Variational Inference for Logical Inference

$$\mathbb{P}\left(t_{cat,X} \,|\, t_{animal,X}, t_{have,Y}, t_{tail,Z}\right)$$

$$= \sum_{x,y,z} \mathbb{P}\left(t_{cat,X} \,|\, x\right) \mathbb{P}\left(x, y, z \,|\, t_{animal,X}, t_{have,Y}, t_{tail,Z}\right)$$

- Exact inference is computationally intractable

Variational Inference for Logical Inference

$$\mathbb{P}\left(t_{cat,X} \,\middle|\, t_{animal,X}, t_{have,Y}, t_{tail,Z}\right)$$

$$= \sum_{x,y,z} \mathbb{P}\left(t_{cat,X} \,\middle|\, x\right) \mathbb{P}\left(x, y, z \,\middle|\, t_{animal,X}, t_{have,Y}, t_{tail,Z}\right)$$

- Exact inference is computationally intractable
- Use (amortised) variational inference

RELPRON Dataset (Rimell et al., 2016)

telescope	*device that astronomers use*
telescope	*device that detects planets*
saw	*device that cuts wood*
philosopher	*person that defends rationalism*
survivor	*person that helicopter saves*
farming	*activity that soil supports*
...	*...*

RELPRON Dataset (Rimell et al., 2016)

telescope	*device that astronomers use*
	device that detects planets
	device that cuts wood
	person that defends rationalism
	person that helicopter saves
	activity that soil supports
	...

RELPRON Dataset (Rimell et al., 2016)

saw	*device that astronomers use*
	device that detects planets
	device that cuts wood
	person that defends rationalism
	person that helicopter saves
	activity that soil supports
	...

RELPRON Dataset (Rimell et al., 2016)

philosopher	*device that astronomers use*
	device that detects planets
	device that cuts wood
	person that defends rationalism
	person that helicopter saves
	activity that soil supports
	...

Similarity in Context (GS2011)

student	write	name
student	spell	name
scholar	write	book
scholar	spell	book

Evaluation Dataset Summary

- Evaluation datasets for visual model
  - RELPRON: inference with relative clauses
  - GS2011: similarity in context
  - MEN, SL999: similarity (no context)

Evaluation Dataset Summary

- Evaluation datasets for visual model
  - RELPRON: inference with relative clauses
  - GS2011: similarity in context
  - MEN, SL999: similarity (no context)
  - (All filtered for Visual Genome vocabulary)

Results (Visual Models)

Model	MEN	SL999	GS2011	RELPRON
VG-count (Herbelot, 2020)	.336	.224	.063	.038
VG-retrieval	.420	.190	.072	.045
EVA (Herbelot, 2020)	.543	.390	.068	.032
FDS (Liu & Emerson, 2022)	.639	.431	.171	.117

Results (Visual Models)

Model	MEN	SL999	GS2011	RELPRON
VG-count (Herbelot, 2020)	.336	.224	.063	.038
VG-retrieval	.420	.190	.072	.045
EVA (Herbelot, 2020)	.543	.390	.068	.032
FDS (Liu & Emerson, 2022)	.639	.431	.171	.117

- Truth-conditional structure helps generalisation

Evaluation Dataset Summary

- Evaluation datasets for textual model
  - RELPRON: inference with relative clauses
  - GS2011, GS2012: similarity in context
  - GS2013: similarity in context (plus adjectives)

Results (Textual Models)

Model	RELPRON	GS2011	GS2012	GS2013
BERT	.667	.519	.608	.562
FDSAS	.580	.552	.660	.601

Results (Textual Models)

Model	RELPRON	GS2011	GS2012	GS2013
BERT	.667	.519	.608	.562
FDSAS	.580	.552	.660	.601

- Competitive with BERT, but with 10% data

Results (Textual Models)

Model	RELPRON	GS2011	GS2012	GS2013
BERT	.667	.519	.608	.562
FDSAS	.580	.552	.660	.601

- Competitive with BERT, but with 10% data
- BERT requires template tuning: consistency better than grammaticality, punctuation crucial!

Truth Conditions from WordNet

- Hyponymy is straightforwardly truth-conditional:
  - f is a hyponym of g iff $\forall x\, f(x) \to g(x)$

Truth Conditions from WordNet

- Hyponymy is straightforwardly truth-conditional:
  - f is a hyponym of g iff $\forall x\, f(x) \rightarrow g(x)$
- With probabilistic truth conditions:
  - f is a hyponym of g iff $\forall x\, f(x) \leq g(x)$

Truth Conditions from WordNet

- Assume:
  - x on the unit sphere, $|x| = 1$
  - f and g logistic regression classifiers,
    $f(x) = \sigma(a_f.x + b_f)$
    $g(x) = \sigma(a_g.x + b_g)$

Truth Conditions from WordNet

- Assume:
  - x on the unit sphere, $|x| = 1$
  - f and g logistic regression classifiers,
    $f(x) = \sigma(a_f.x + b_f)$
    $g(x) = \sigma(a_g.x + b_g)$
- Then the following are equivalent:
  - $\forall x\, f(x) \leq g(x)$
  - $b_g - b_f - |a_g - a_f| \geq 0$

Results (WordNet Models)

Model	Link Pr.
TransE (Bordes et al., 2013)	.345
DistMult (Yang et al., 2015)	.425
rGAT (Chen et al., 2021)	.500

Results (WordNet Models)

Model	Link Pr.	Sim.	Ana.	POS	NER
TransE (Bordes et al., 2013)	.345	.486	.320	.765	.492
DistMult (Yang et al., 2015)	.425	.288	.116	.672	.484
rGAT (Chen et al., 2021)	.500	.289	.132	.716	.307

Results (WordNet Models)

Model	Link Pr.	Sim.	Ana.	POS	NER
TransE (Bordes et al., 2013)	.345	.486	.320	.765	.492
DistMult (Yang et al., 2015)	.425	.288	.116	.672	.484
rGAT (Chen et al., 2021)	.500	.289	.132	.716	.307
FuncE (Cheng et al., 2023)	.259	.512	.353	.772	.545

Summary

- Truth conditions feasible at scale
- Approximations required
- Improves generalisation

Crucial Approximations & Simplifications

- Images (Liu and Emerson): variational inference, no latent variables
- Text (Lo et al.): amortised variational inference, simple world model
- Ontology (Cheng et al.): simple truth-conditional model, simple world model

Crucial Approximations & Simplifications

- Images (Liu and Emerson): variational inference, no latent variables
- Text (Lo et al.): amortised variational inference, simple world model
- Ontology (Cheng et al.): simple truth-conditional model, simple world model
- Pragmatics (RSA) needs further approximation...

Bitter Lesson

- At scale, truth-conditional semantics is intractable

Bitter Lesson

- At scale, truth-conditional semantics is intractable
- An intractable model is cognitively implausible

Bitter Lesson

- At scale, truth-conditional semantics is intractable
- An intractable model is cognitively implausible
- Unavoidable “approximations” must be seen as part of the theory...

Truth-Conditional Semantics

"the meaning of a sentence is the method of its verification... there is some uniform means of deriving all the other features of the use of any sentence from this one feature" — Dummett (1976)

Beyond Truth Conditions

- Goal: a theory of language understanding that is tractable at scale

Beyond Truth Conditions

- Goal: a theory of language understanding that is tractable at scale
- Idea: some processes of language understanding are not *reducible* to truth conditions, but instead *mutual* to truth conditions

Beyond Truth Conditions

- Truth-conditional model:
  - $\mathbb{P}(t_u \mid s)$

Beyond Truth Conditions

- Truth-conditional model:
  - $\mathbb{P}(t_u | s)$
- Inference model:
  - $\mathbb{P}(s | t_u)$

Beyond Truth Conditions

- Truth-conditional model:
  - $\mathbb{P}(t_u \mid s)$
- Inference model:
  - $\mathbb{P}(s \mid t_u)$
- Bayesian inference is intractable:
  - $\mathbb{P}(s \mid t_u) = \dfrac{\mathbb{P}(t_u \mid s)\,\mathbb{P}(s)}{\sum_{s'} \mathbb{P}(t_u \mid s')\,\mathbb{P}(s')}$

Amortised Variational Inference Revisited

- VAE objective: inference network approximates Bayesian inference for generative model

Amortised Variational Inference Revisited

- VAE objective: inference network approximates Bayesian inference for generative model
- Zhao et al. (2019) alternative view:
  - VAE objective minimises KL-divergence between
  - generative model $\mathbb{P}_\theta(z)\mathbb{P}_\theta(x|z)$
  - inference model $\mathbb{P}_\phi(x)\mathbb{P}_\phi(z|x)$

Amortised Variational Inference Revisited

- Truth-conditional model $\mathbb{P}(t_u | s)$
- World-inferential model $\mathbb{P}(s | t_u)$

Amortised Variational Inference Revisited

- Truth-conditional model $\mathbb{P}(t_u | s)$
- World-inferential model $\mathbb{P}(s | t_u)$
- Treat them as *mutual*:
  - Neither is primary
  - Each approximates the other
  - No coherent joint $\mathbb{P}(s, t_u)$

Masked Language Modelling Revisited

- Masked language model predictions:
  - $\mathbb{P}\left(w_i \,\middle|\, w_1, \ldots, w_{i-1}, w_{i+1}, \ldots, w_n\right)$

Masked Language Modelling Revisited

- Masked language model predictions:
  - $\mathbb{P}\left(w_i \,\middle|\, w_1, \ldots, w_{i-1}, w_{i+1}, \ldots, w_n\right)$
- Can be seen as *mutual*:
  - No w_i is primary
  - No coherent joint $\mathbb{P}(w_1, \ldots, w_n)$
  - *Approximately* coherent

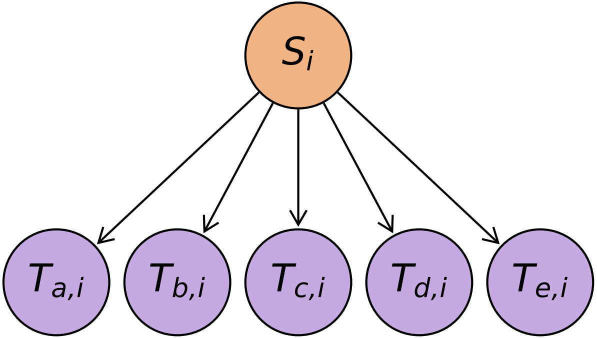
Mutual Models

- Bundle of component models
- Each component makes some conditional inference
- Jointly trained with an objective for which a coherent (but intractable) model would be optimal

Mutual Models

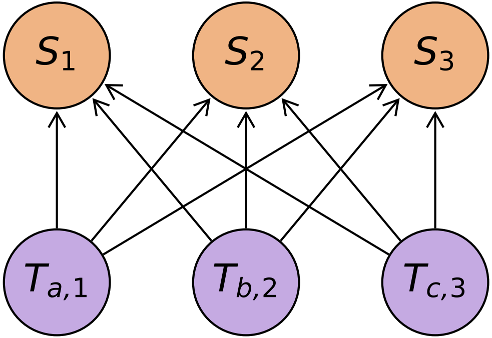
- Bundle of component models
- Each component makes some conditional inference
- Jointly trained with an objective for which a coherent (but intractable) model would be optimal
- Rigorous framework for modelling “incoherence”: systematic divergence between components

Lexical Truth-Conditional Model



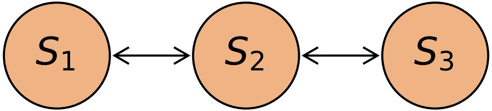
$$\mathbb{P}\left(t_{r,i} \mid s_i\right)$$

World-Inferential Model



$$\mathbb{P}(s \mid t_u)$$

Conditional World Model



$$\mathbb{P}\left(s_i \,\middle|\, s_{\neq i}\right)$$

Mutual Models

- Component models:
  - Truth-conditional model $\mathbb{P}(t_u \mid s)$
  - World-inferential model $\mathbb{P}(s \mid t_u)$
  - Conditional world model $\mathbb{P}(s_i \mid s_{\neq i})$
- Trained jointly (without a coherent joint distribution!)

Mutual Models for Cognitive Science

- Rigorous framework for *bounded rationality*
  (vs. Icard, 2018; Chater et al., 2020; Lieder & Griffiths, 2020)

Mutual Models for Cognitive Science

- Rigorous framework for *bounded rationality* (vs. Icard, 2018; Chater et al., 2020; Lieder & Griffiths, 2020)
- Cognitive processes in different directions will systematically diverge

Mutual Models for Cognitive Science

- Rigorous framework for *bounded rationality*
  (vs. Icard, 2018; Chater et al., 2020; Lieder & Griffiths, 2020)
- Cognitive processes in different directions will systematically diverge
  - Classification and production
  - Classification and imagination

Mutual Models for Cognitive Science

- Rigorous framework for *bounded rationality*
  (vs. Icard, 2018; Chater et al., 2020; Lieder & Griffiths, 2020)

- Cognitive processes in different directions will systematically diverge
  - Classification and production
  - Classification and imagination
  - Classification and generation

Classification and Production

- Two mutual processes:
  - Classifying instances of a concept
  - Producing instances of a concept

Example: Looptail g

Example: Looptail g

- Classified without effort
- Produced with difficulty (if at all)

Classification and Production

- Aim to quantify:
  - How does divergence depend on learning?
  - How do mutual processes support each other during learning?

Example: Ge’ez Script for Amharic

Classification and Production

- Hand-drawn characters:
  - Visually complex, hard to describe
  - Physically simple to reproduce

Classification and Production

- Hand-drawn characters:
  - Visually complex, hard to describe
  - Physically simple to reproduce
- Plan:
  - Observe classification and production behaviour, under different learning conditions
  - Compare with mutual model predictions

Classification and Imagination

- Two mutual processes:
  - Classifying instances of a concept
  - Imagining instances of a concept

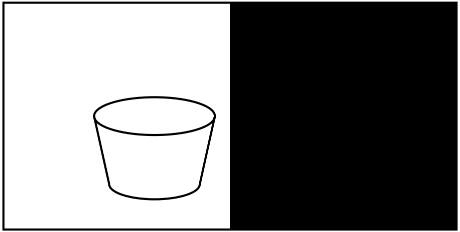
Classification and Imagination

- Two mutual processes:
  - Classifying instances of a concept
  - Imagining instances of a concept
- Methodological challenge: can’t observe imagination

Classification and Imagination

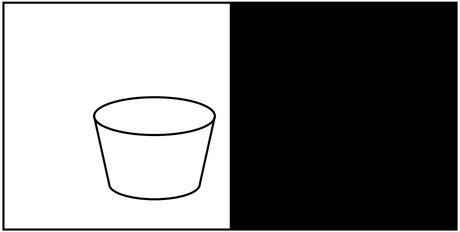
- Two mutual processes:
  - Classifying instances of a concept
  - Imagining instances of a concept
- Methodological challenge: can’t observe imagination
- Idea: provide some features, probe others

Classification and Imagination



“a cup and a bowl”

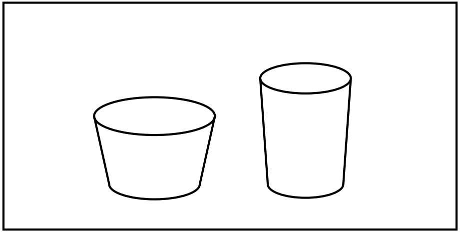
Classification and Imagination



“a cup and a bowl”

Can you see the bowl?

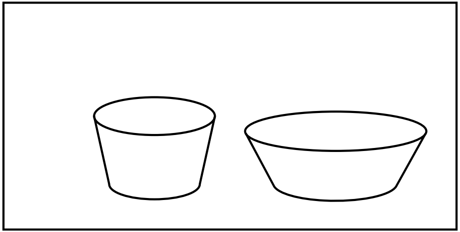
Classification and Imagination



“a cup and a bowl”

Can you see the bowl?

Classification and Imagination



"a cup and a bowl"

Can you see the bowl?

Classification and Imagination

- Divergence between:
  - Classification-based Bayesian inference
  - Imagination-based inference

Classification and Imagination

- Divergence between:
  - Classification-based Bayesian inference
  - Imagination-based inference
- “Mode collapse” in machine learning

Summary of Plans

- Mutual Models
  - New framework for probabilistic modelling
  - New tools for studying human behaviour

Summary of Plans

- Mutual Models
  - New framework for probabilistic modelling
  - New tools for studying human behaviour
- Next steps:
  - Mutual models at scale
  - Experiments with human participants

Conclusion

- Truth conditions at scale
  - Feasible (with approximations…)
  - Truth helps generalisation

Conclusion

- Truth conditions at scale
  - Feasible (with approximations…)
  - Truth helps generalisation
- Beyond truth conditions
  - Reducing to truth conditions is intractable
  - Instead: *mutual models*