



NEW YORK UNIVERSITY

Question asking as program induction

Anselm Rothe



NEW YORK UNIVERSITY

WITH

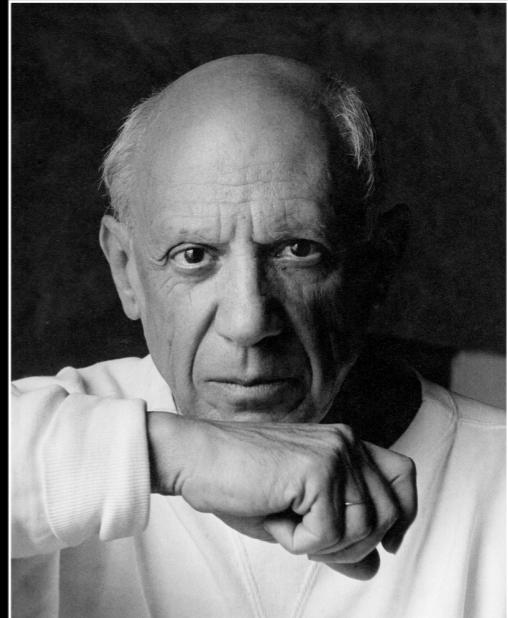


Todd Gureckis



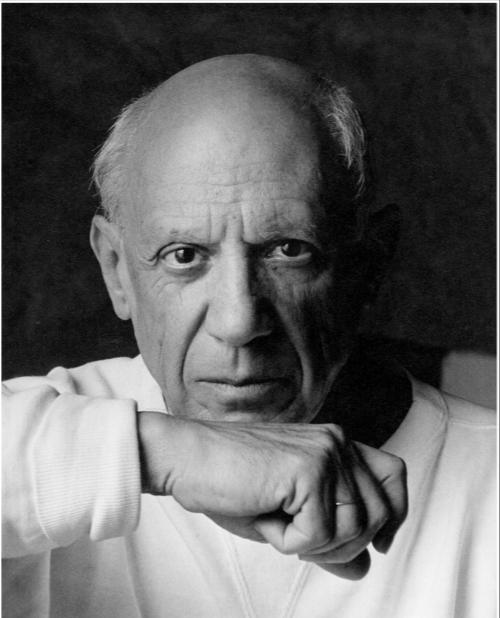
Brenden Lake

**Computers are
useless. They can only
give you answers.**



(attributed to) **Pablo Picasso**

**Computers are
useless. They can only
give you answers.**



(attributed to) **Pablo Picasso**

**What does it take
to build a machine
that asks good
questions?**



What does it take to build a machine that asks good questions?

- Representing questions as **programs** that, when executed on the state of the world, output an **answer**

Key ingredients

- Generativity
- Compositionality
- Informativeness
- Simplicity

**What does it take
to build a machine
that asks good
questions?**

- Representing questions as **programs** that, when executed on the state of the world, output an **answer**

HUMAN QUESTIONS

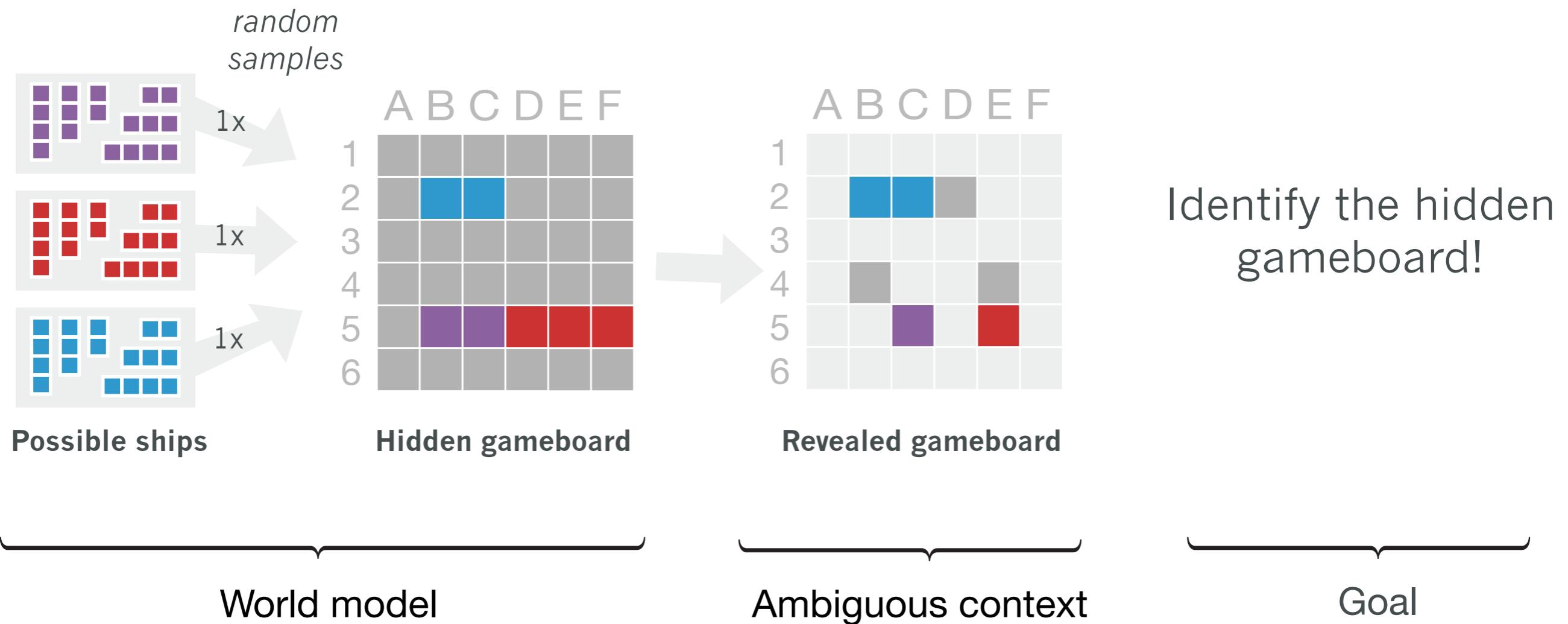
We need a task that allows people to intuitively ask **interesting questions** and is still amenable to **formal modeling**

World model

Ambiguous context

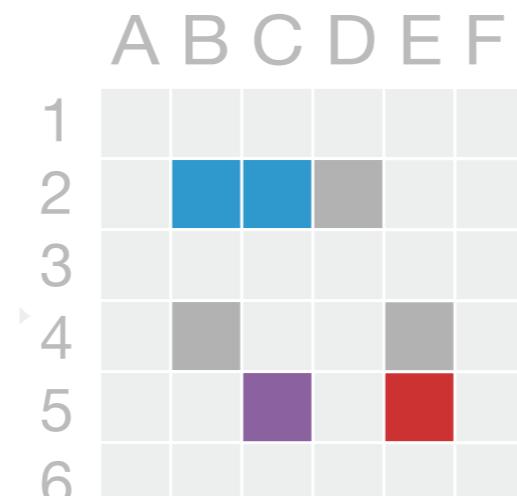
Goal

HUMAN QUESTIONS



HUMAN QUESTIONS

People were dropped
into the middle of a
game and were given
the 'magic' opportunity
to **ask whatever
they want***



Revealed gameboard

Ambiguous context

Identify the hidden
gameboard!

Goal

* only one-word-answer questions,
no combination of questions

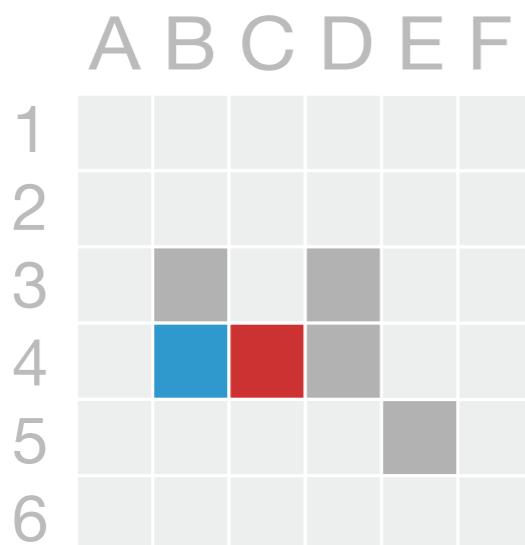
type in your question



Is the red ship horizontal?

HUMAN QUESTIONS

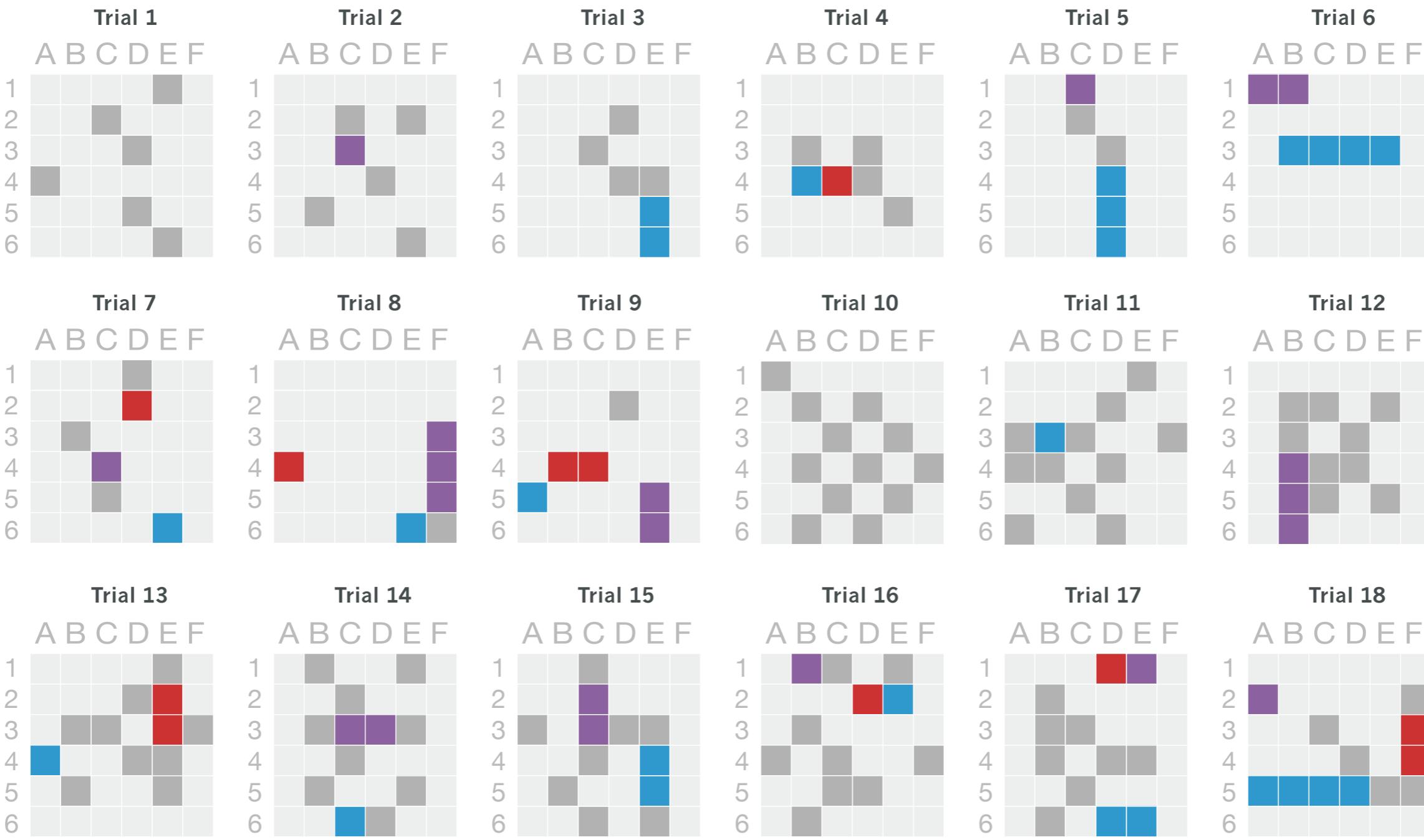
Context



Example questions from people

- At what location is the top left part of the purple ship?
- What is the location of one purple tile?
- Is the blue ship horizontal?
- Is the red ship 2 tiles long?
- Is the purple ship horizontal?
- Is the red ship horizontal?

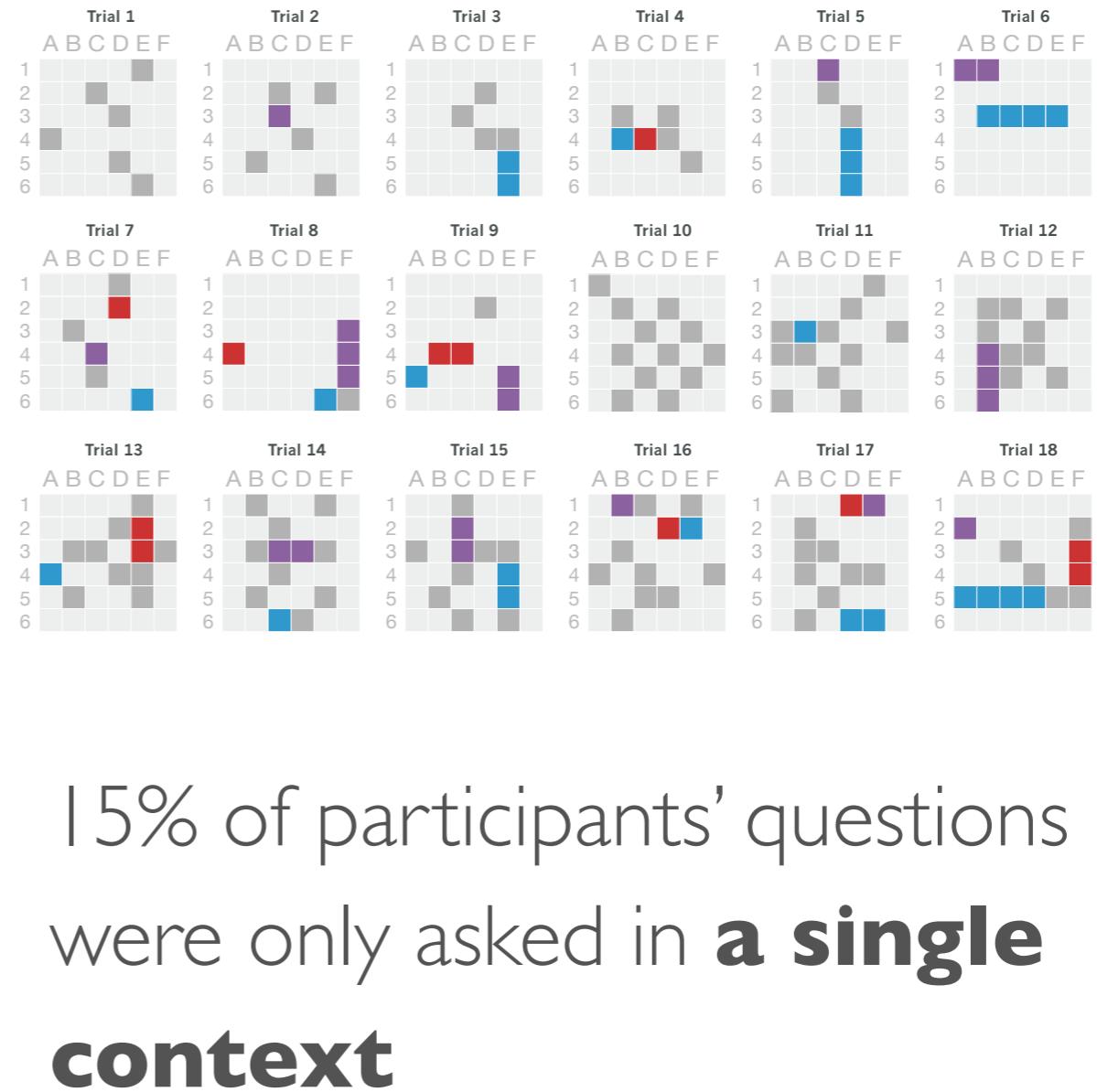
...



- 40 MTurk participants
- 605 human questions

Key ingredients

- Generativity
- Compositionality
- Informativeness
- Simplicity



- 15% of participants' questions were only asked in **a single context**
- Our model needs the ability to generate **novel questions**

Key ingredients

- Generativity
- **Compositionality**
- Informativeness
- Simplicity

“How long is the blue ship?”

“Is the blue ship less than 4 tiles?”

“Are there any ships with 4 tiles?”

“Does the blue ship have 3 tiles?”

“Does the red ship have more tiles than the blue ship?”

“Are all 3 ships the same size?”

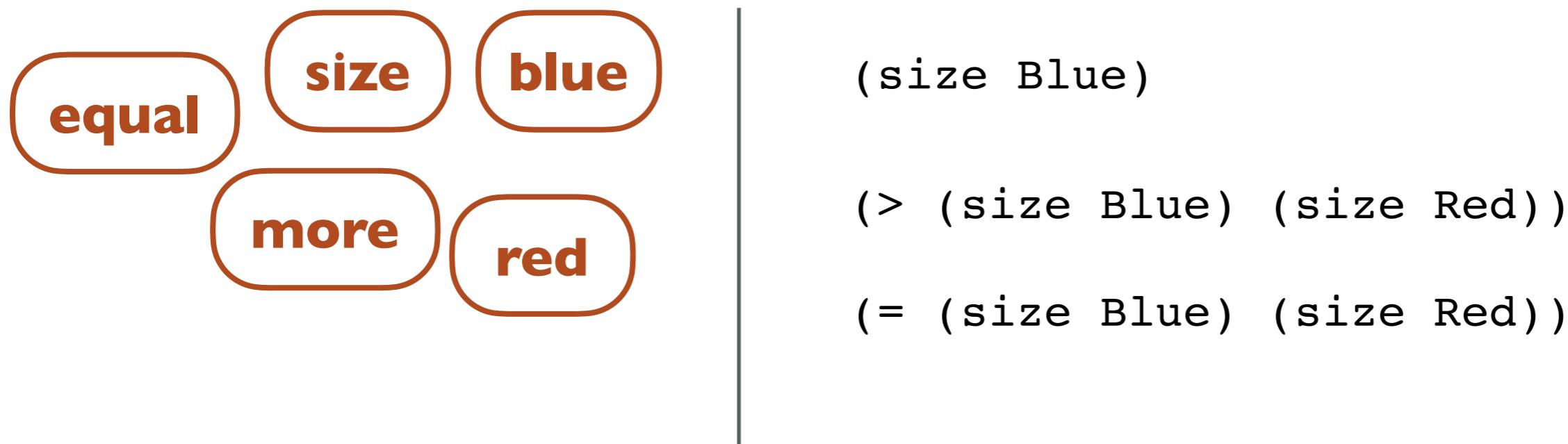


COMPOSITIONALITY IN QUESTION STRUCTURE

size **blue** | (size Blue)

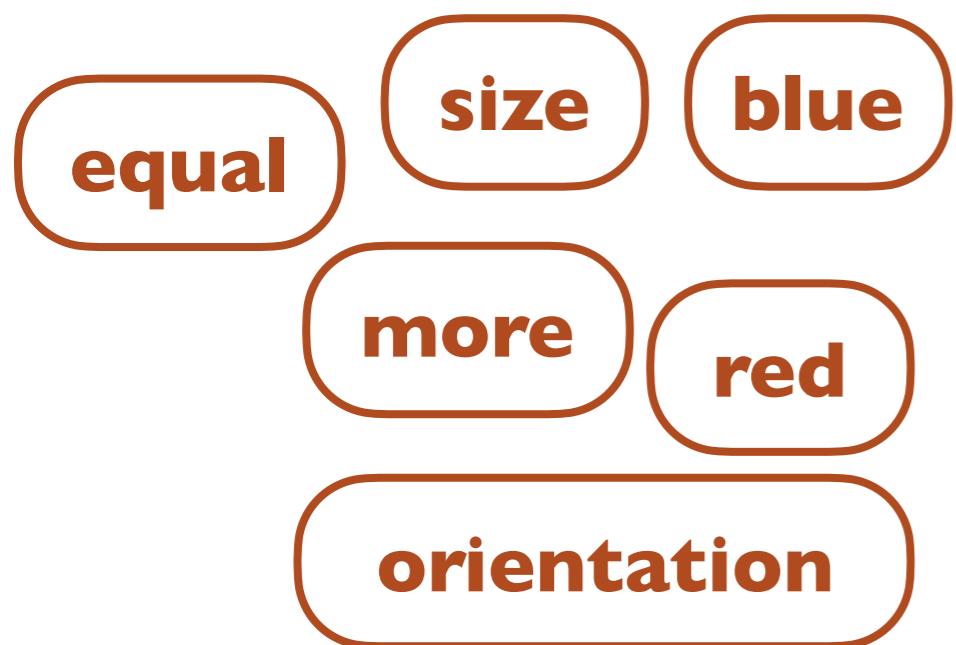
- Questions are represented as programs that, when executed on the state of the world, output an answer

COMPOSITIONALITY IN QUESTION STRUCTURE



- Questions are represented as programs that, when executed on the state of the world, output an answer

COMPOSITIONALITY IN QUESTION STRUCTURE



(size Blue)

(> (size Blue) (size Red))

(= (size Blue) (size Red))

(= (orientation Blue) (orientation Red))

“Are the blue ship and the red ship parallel?”

- Questions are represented as programs that, when executed on the state of the world, output an answer

COMPOSITIONALITY IN QUESTION STRUCTURE

How many ships are three tiles long?

```
( +  
  ( map  
    ( lambda  
      x  
      ( =  
        ( size x )  
        3  
      )  
    )  
    ( set Blue Red Purple )  
)
```

Are any ships 3 tiles long?

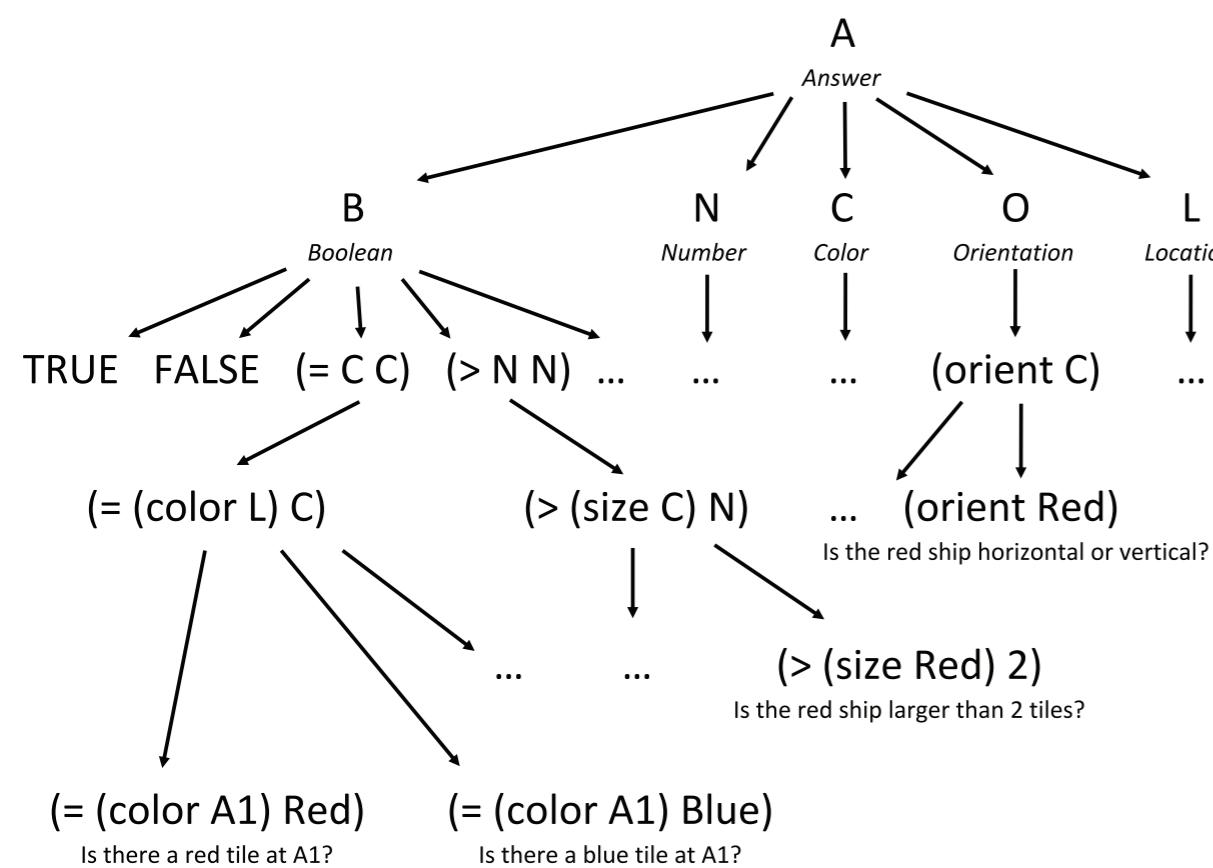
```
( >  
  ( +  
    ( map  
      ( lambda  
        x  
        ( =  
          ( size x )  
          3  
        )  
      )  
      ( set Blue Red Purple )  
)  
  )  
0  
)
```

Are all ships three tiles long?

```
( =  
  ( +  
    ( map  
      ( lambda  
        x  
        ( =  
          ( size x )  
          3  
        )  
      )  
      ( set Blue Red Purple )  
)  
  )  
3  
)
```

- Questions are represented as programs that, when executed on the state of the world, output an answer

A GRAMMAR OF QUESTIONS



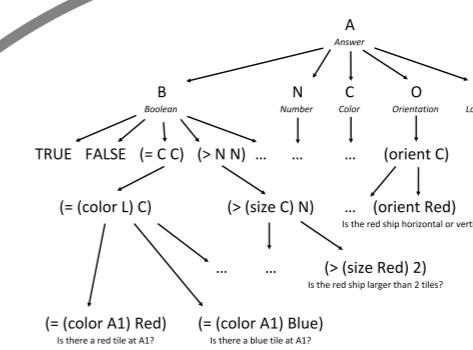
$A \rightarrow B \text{ (boolean)}$	$N \rightarrow 0$	$O \rightarrow H$
$A \rightarrow N \text{ (number)}$	\dots	$O \rightarrow V$
$A \rightarrow C \text{ (color)}$	$N \rightarrow 10$	$O \rightarrow (\text{orient } S)^b$
$A \rightarrow O \text{ (orientation)}$	$N \rightarrow (+ N N)$	$L \rightarrow A1$
$A \rightarrow L \text{ (location)}$	$N \rightarrow (+ B B)$	\dots
	$N \rightarrow (+ \text{set} N)$	$L \rightarrow F6$
	$N \rightarrow (+ \text{set} B)$	$L \rightarrow (\text{topleft } S)^b$
	$N \rightarrow (- N N)$	$L \rightarrow (\text{bottomright } S)^b$
	$N \rightarrow (\text{size } S)^b$	$L \rightarrow (\text{draw set} L)^*$
	$N \rightarrow (\text{row } L)$	
	$N \rightarrow (\text{col } L)$	
$B \rightarrow \text{TRUE}$		
$B \rightarrow \text{FALSE}$		
$B \rightarrow (\text{not } B)$		
$B \rightarrow (\text{and } B B)$		
$B \rightarrow (\text{or } B B)$		
$B \rightarrow (= B B)$		
$B \rightarrow (= N N)$		
$B \rightarrow (= O O)$		
$B \rightarrow (= \text{set} N)$		
$B \rightarrow (> N N)$		
$B \rightarrow (\text{touch } S S)^b$		
$C \rightarrow S \text{ (ship color)}$		
$C \rightarrow \text{Water}$		
$C \rightarrow (\text{color } L)^b$		
$S \rightarrow \text{Blue}$		
$S \rightarrow \text{Red}$		
$S \rightarrow \text{Purple}$		
$S \rightarrow x^\lambda$		
$\text{set} B \rightarrow (\text{map } fxB \text{ set} S)$		
$fxB \rightarrow (\lambda x B)$		
$\text{set} N \rightarrow (\text{map } fxN \text{ set} S)$		
$fxN \rightarrow (\lambda x N)$		
$\text{set} S \rightarrow (\text{set Blue Red Purple})$		
$\text{set} L \rightarrow (\text{set } A1 \dots F6)$		
$\text{set} L \rightarrow (\text{shipTiles } S)^b *$		
$\text{set} L \rightarrow (\text{map } fxL \text{ set} S)$		
$fxL \rightarrow (\lambda x L)$		

Rothe, Lake, & Gureckis 2017, NIPS

Key ingredients

- Generativity ✓
- Compositionality ✓
- Informativeness
- Simplicity

cost / fitness function



Question space
as defined by grammar

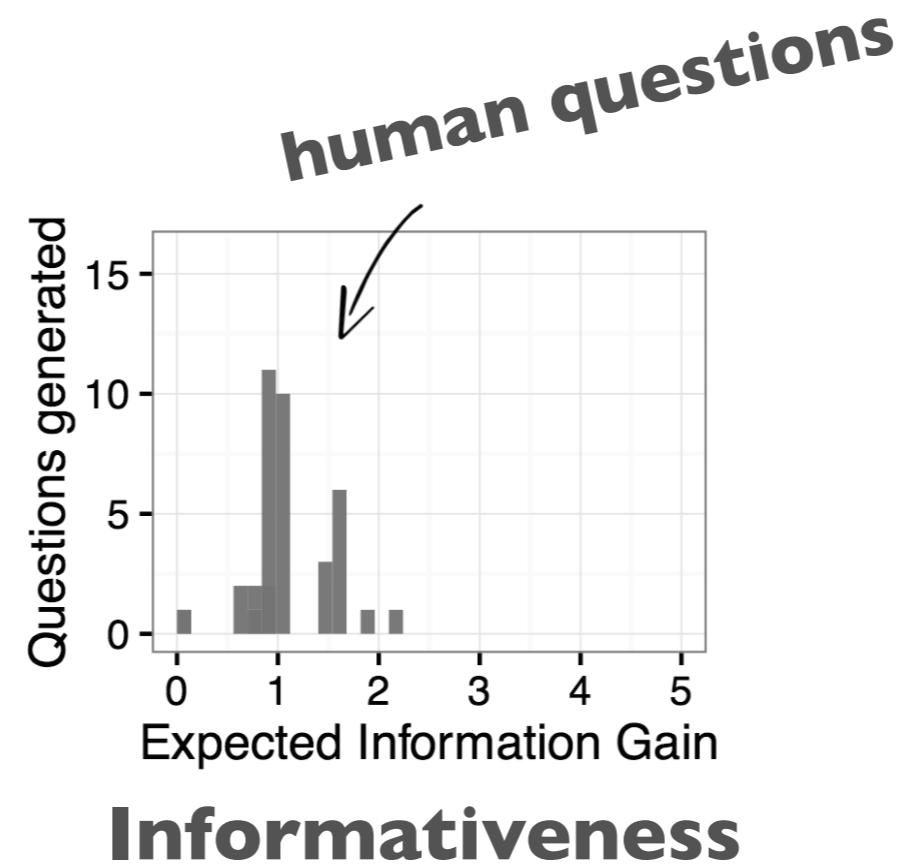
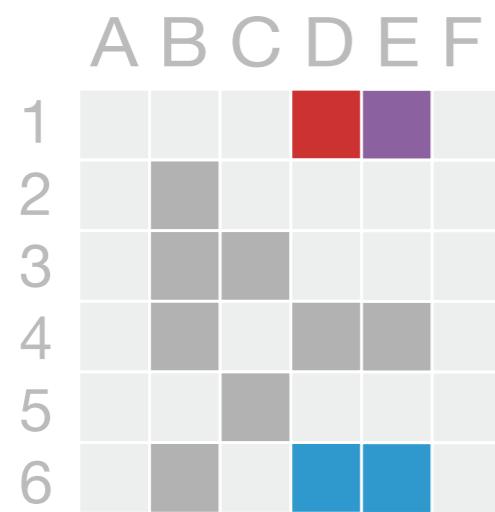


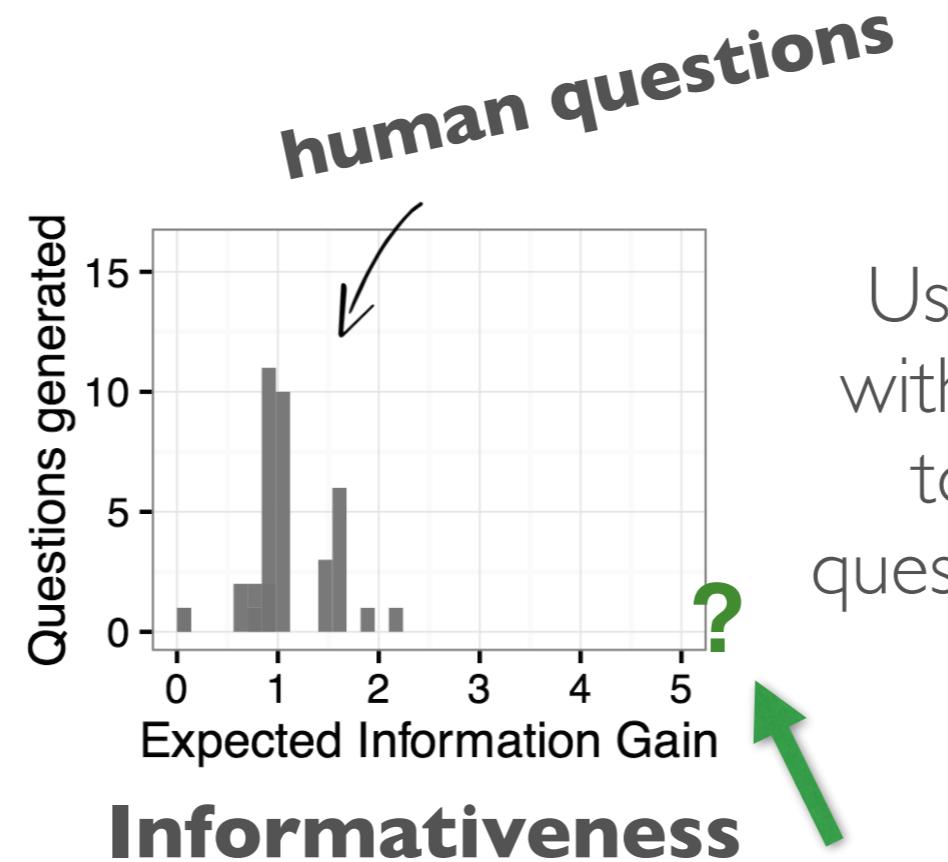
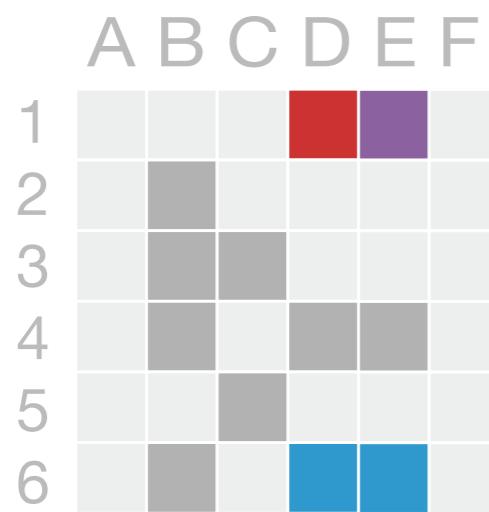
Generating questions

- Drawing samples from grammar
- Evolutionary search

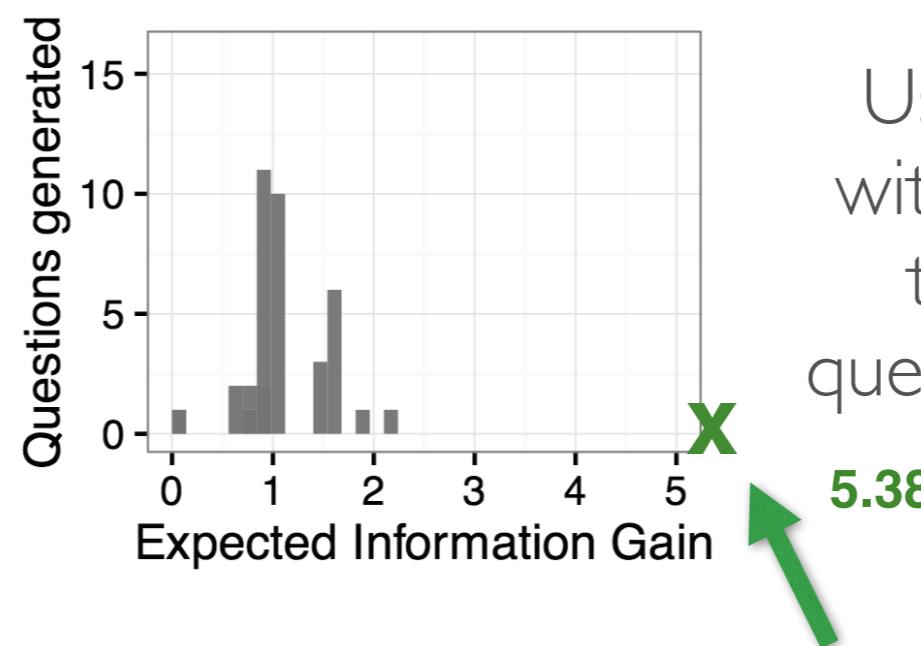
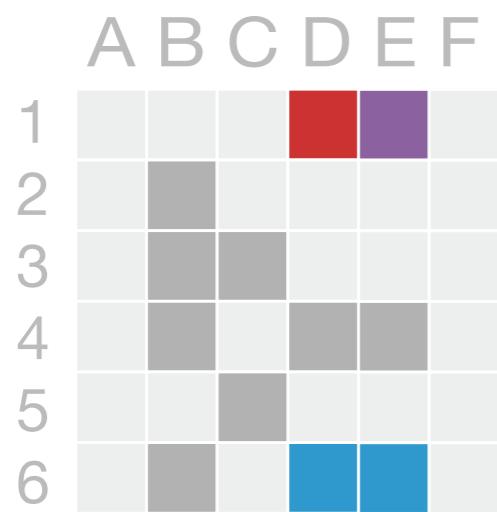
Key ingredients

- Generativity
- Compositionality
- **Informativeness**
- Simplicity





Using a genetic algorithm
with EIG as fitness function
to search for the “best
question” for a given context



Using a genetic algorithm with EIG as fitness function to search for the “best question” for a given context

```
(- (- (+ (- (- (+ (size Purple) (coll (topleft Red)))
(size Blue)) (- (+ (size Blue) (size Red)) (coll (topleft
Red)))) (coll (bottomright Purple))) (+ (+ (coll (topleft
Red)) (+ (- (- (+ (size Purple) (coll (topleft Red)))
(size Blue)) (- (+ (size Blue) (size Red)) (coll (topleft
Blue)))) (coll (topleft Red)))) (+ (- (- (+ (size Purple)
(coll (topleft Red))) (size Blue)) (- (+ (size Blue) (size
Red)) (coll (topleft Red)))) (coll (topleft Red)))) (size
Red)) (- (+ (size Blue) (size Blue)) (coll (topleft
Red))))
```

Key ingredients

- Generativity
- Compositionality
- **Informativeness**
- **Simplicity**

What **features** are relevant for people to ask a question?

- f_1 **Informativeness**
Informative questions
- f_2 **Complexity**
Short questions

- Combine features of question x via weighted sum

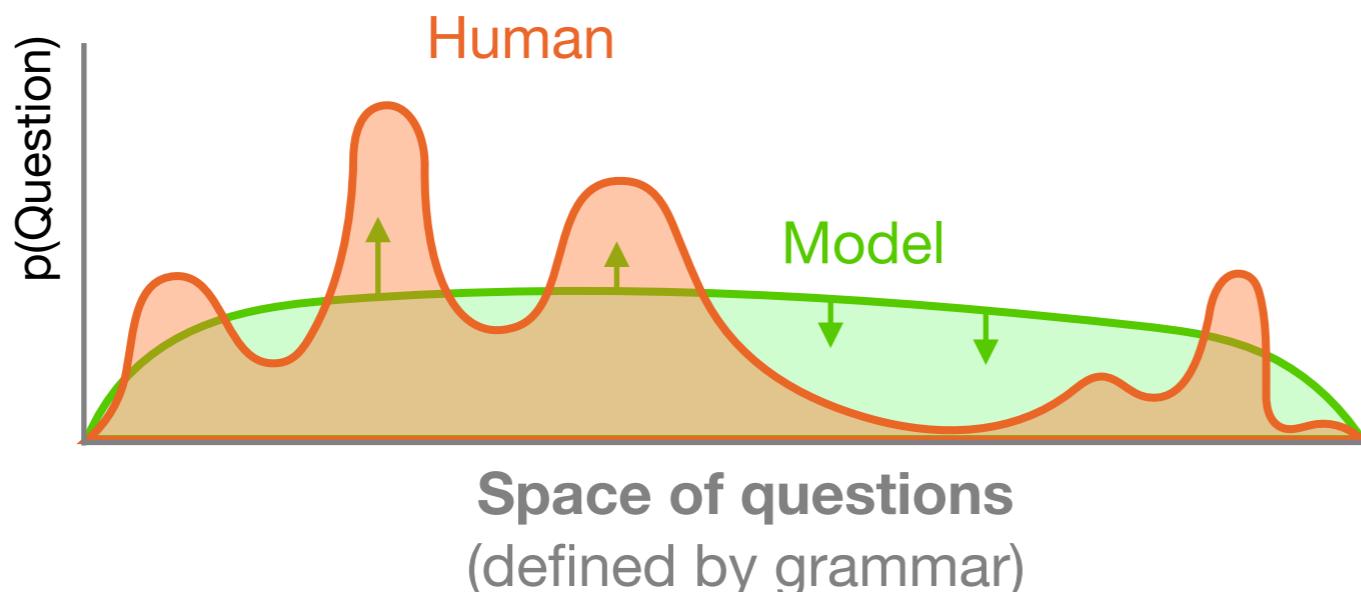
$$\mathcal{E}(x) = \theta_1 f_1(x) + \theta_2 f_2(x) + \dots + \theta_K f_K(x)$$

- Predict probability of question x being asked

$$p(x) = \frac{\exp(-\mathcal{E}(x))}{\sum_{x \in X} \exp(-\mathcal{E}(x))}$$

What **features** are relevant for people to ask a question?

- f_1 **Informativeness**
Informative questions
- f_2 **Complexity**
Short questions

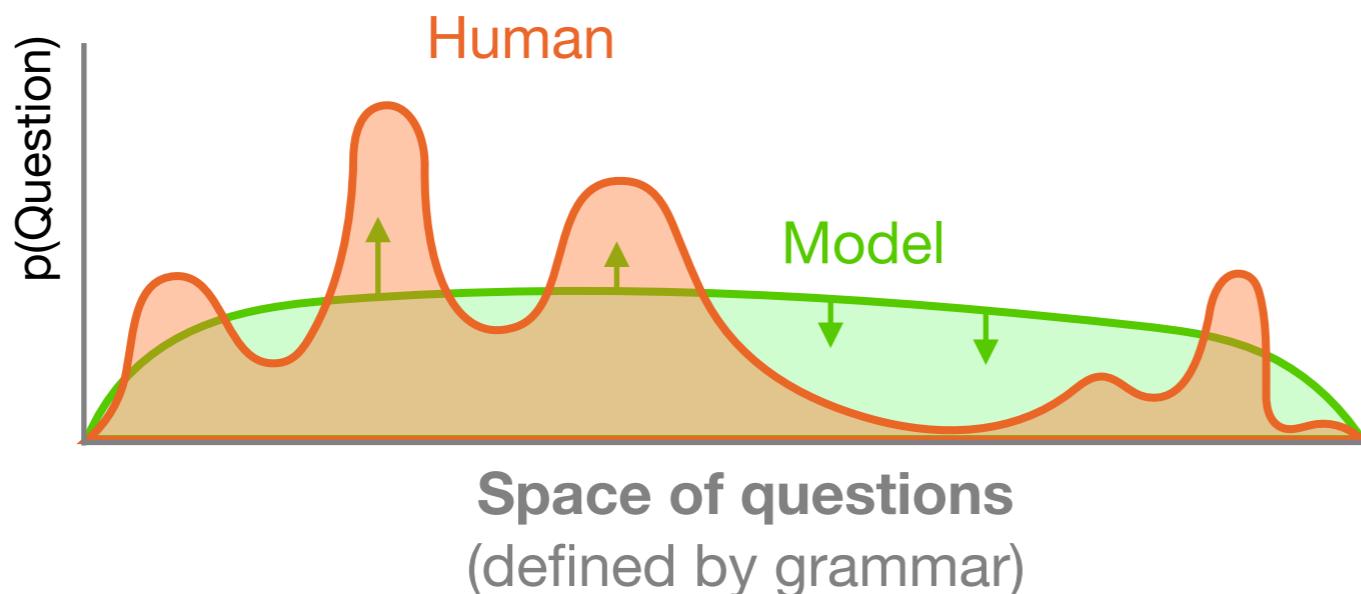


Rothe, Lake, & Gureckis 2017, NIPS

What **features** are relevant for people to ask a question?

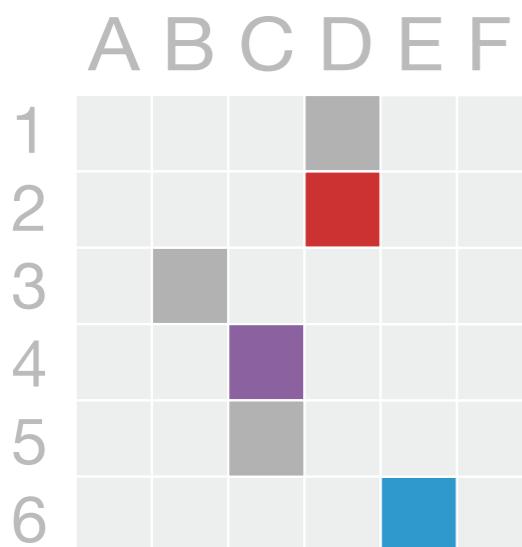
Model	Features	Log-likelihood (out-of-sample predictions)
Full	all	-1400.06
Information-agnostic	not f_1	-1464.65
Complexity-agnostic	not f_2	-22993.38

- f_1 **Informativeness**
Informative questions
- f_2 **Complexity**
Short questions



Rothe, Lake, & Gureckis 2017, NIPS

MODEL OR HUMAN?



Are all the ships horizontal?

```
(all (map (lambda x (= H (orient x))) (set Blue Red Purple)))
```

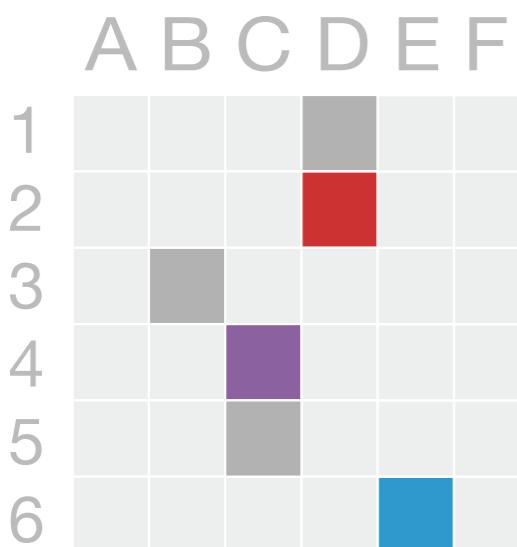
Are any of the ship sizes greater than 2?

```
(any (map (lambda x (> (size x) 2)) (set Blue Red Purple)))
```

How many ships are 4 tiles long?

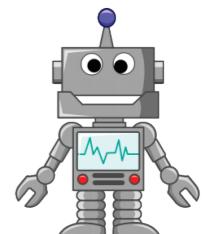
```
(++ (map (lambda x (= (size x) 4)) (set Blue Red Purple)))
```

MODEL OR HUMAN?



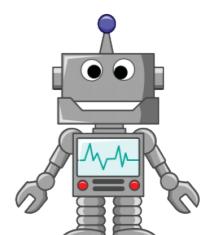
Are all the ships horizontal?

```
(all (map (lambda x (= H (orient x))) (set Blue Red Purple)))
```



Are any of the ship sizes greater than 2?

```
(any (map (lambda x (> (size x) 2)) (set Blue Red Purple)))
```

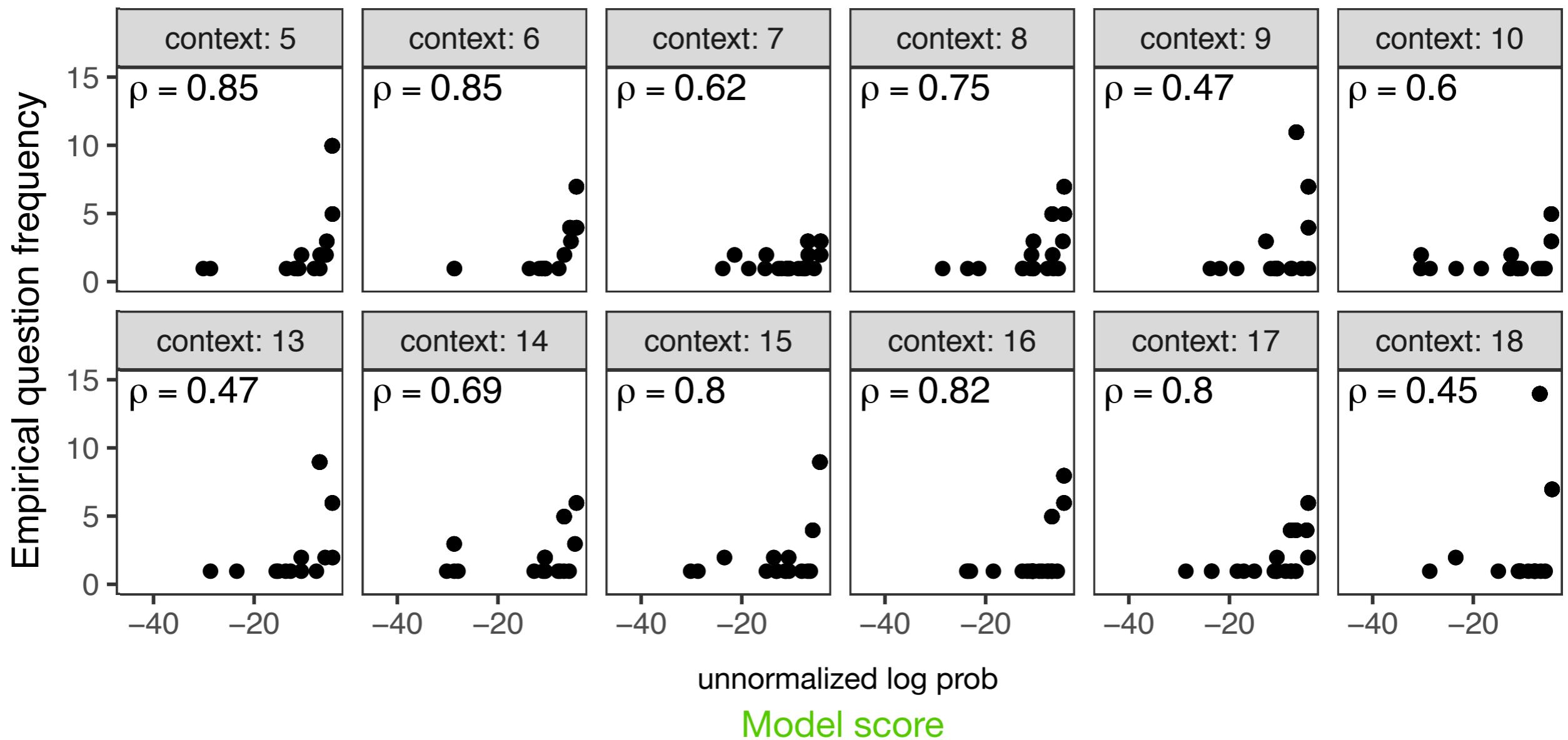


How many ships are 4 tiles long?

```
(++ (map (lambda x (= (size x) 4)) (set Blue Red Purple)))
```



Human



Average rank correlation $\rho = .64$

Rothe, Lake, & Gureckis 2017, NIPS

Key ingredients

- Generativity
- Compositionality
- Informativeness
- Simplicity

What does it take to build a machine that asks good questions?

We represent **questions** as **programs** that, when executed on the state of the world, **output an answer**.

We achieve **generativity** through **compositionality**.

Good, human-like questions are **informative** but **simple**.

You may now generate **your questions**

