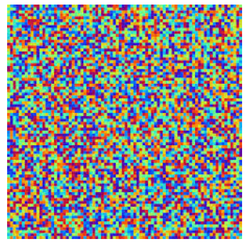
Neuro-Symbolic Commonsense Knowledge and Reasoning

Yejin Choi

Paul G. Allen School of Computer Science & Engineering
University of Washington &
Allen Institute for Artificial Intelligence

Despite (super-) human-level performances on leaderboards…

- SOTA neural models are brittle if given adversarial or out-of-domain samples


+

=


Giant panda

Gibbon

Object Recognition

Szegedy et al, 2014….

…. Nikola Tesla moved to Prague in 1880. … Tadakatsu moved to Chicago in 1881.

Where did Tesla move in 1880? Chicago

QA

Jia et al, 2017

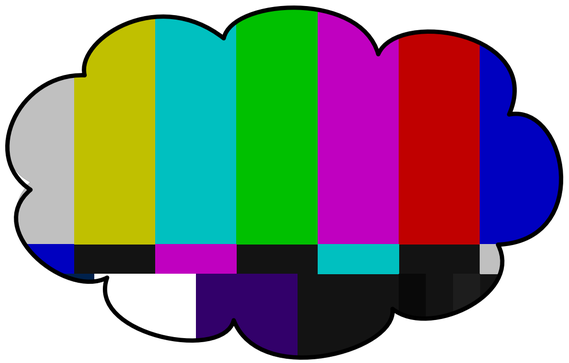
A horse standing in the grass.

Captioning

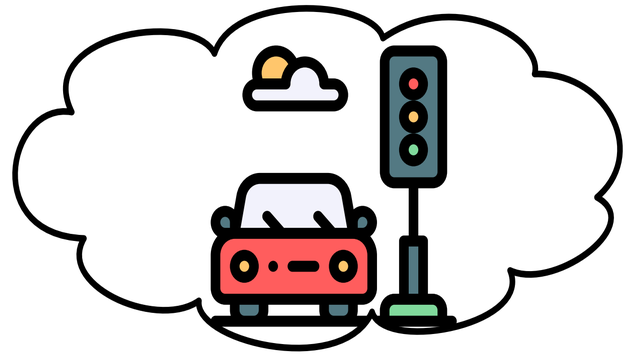
MacLeod et al, 2017

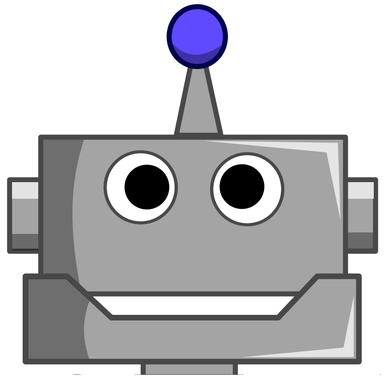
Solving only a "dataset"
without solving the underlying "task"!

(lacking "systematic generalization" (lake & baron 2017, Bahdanauet al & Courville ICLR 2019, Bengio's keynote at NeurIPS 2019))

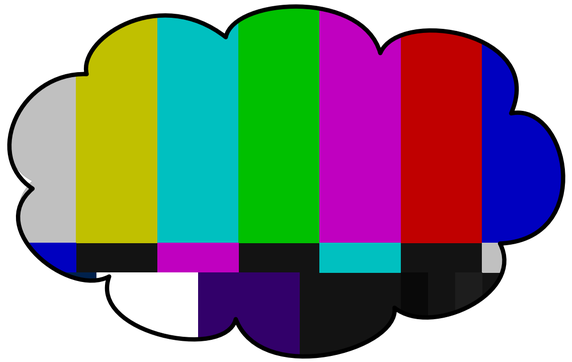


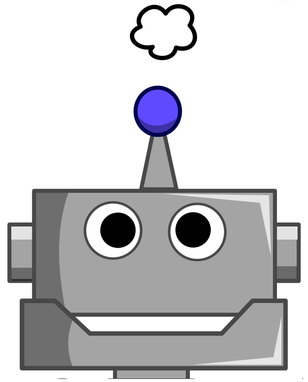
Let's bridge this gap!



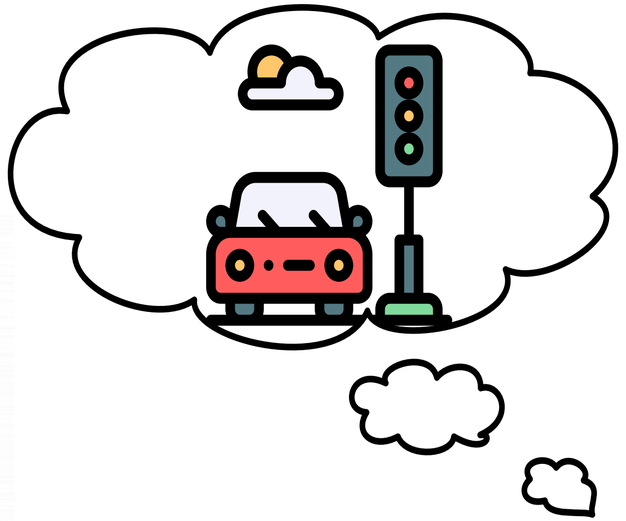


Peters et al., 2018;
Devlin et al., 2018





Let's bridge this gap!



SYSTEM 1
Intuition & instinct

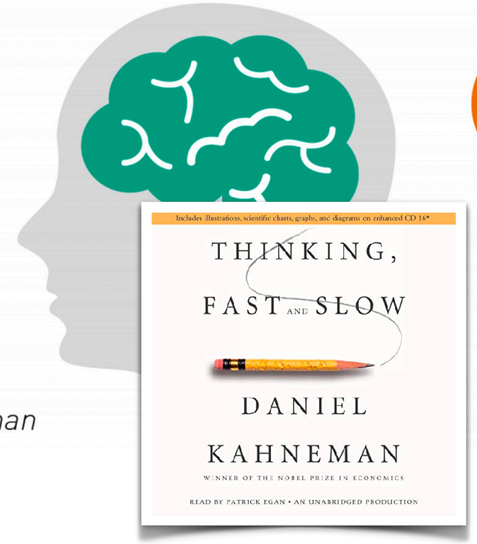
SYSTEM 2
Rational thinking

95%

Unconscious
Fast
Associative
Automatic pilot

5%

Takes effort
Slow
Logical
Lazy
Indecisive

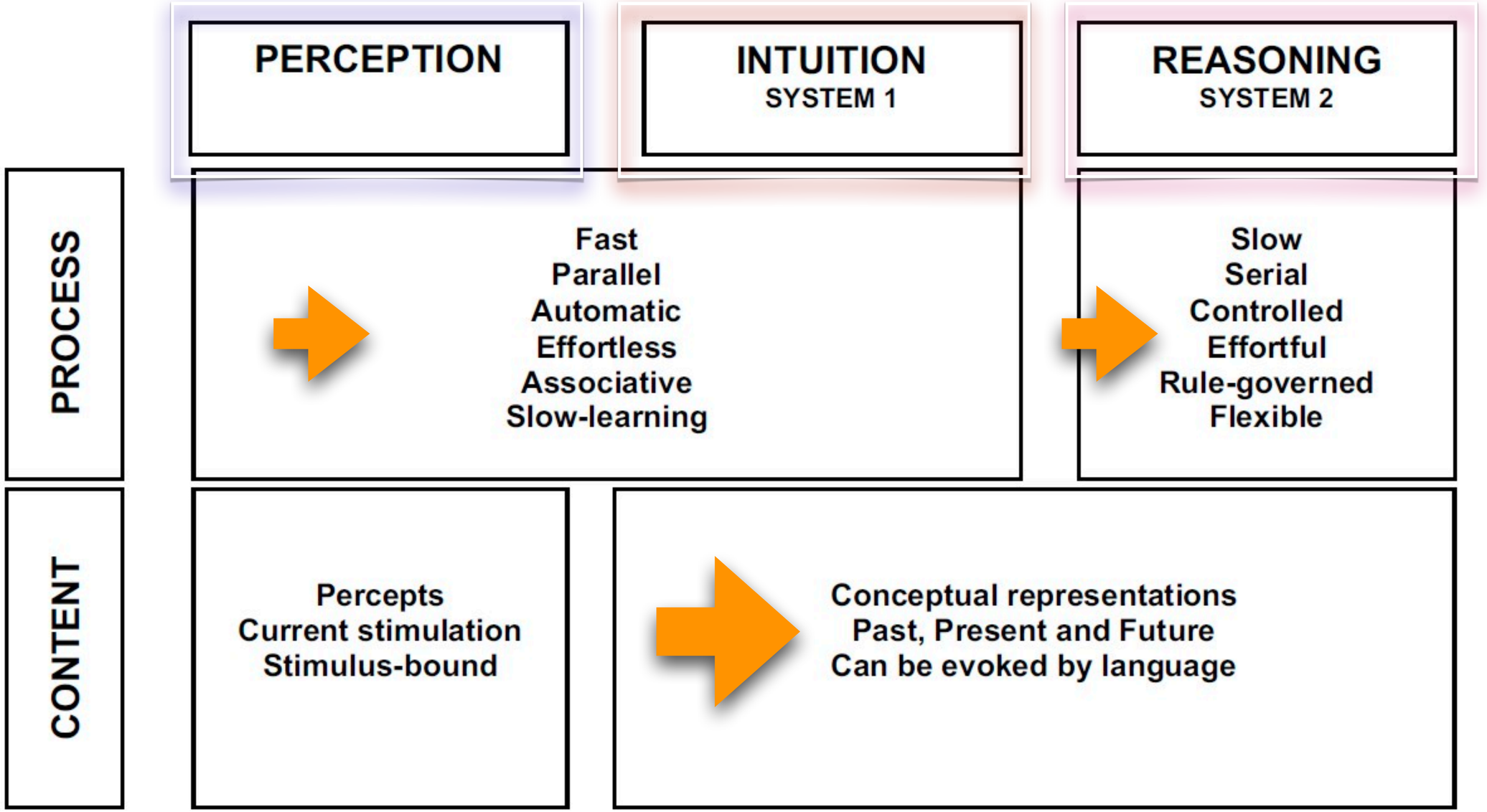


Source: Daniel Kahneman

- **Myth**: we know how to do **[system-1 reasoning]** with deep learning so we only need to figure out **[system-2 reasoning]**?

Kahneman's "three cognitive systems"

— "*Maps of Bounded Rationality: …*" (Kahneman 2003)



Kahneman's "three cognitive systems"

— "*Maps of Bounded Rationality: …*" (Kahneman 2003)

PERCEPTION

- object recognition
- image segmentation

INTUITION
SYSTEM 1

- Intuitive inferences on
  - pre-conditions and post-conditions
  - what happens before and after?
  - motivations and intents
  - mental and emotional states

=> This is what humans do every waking minute

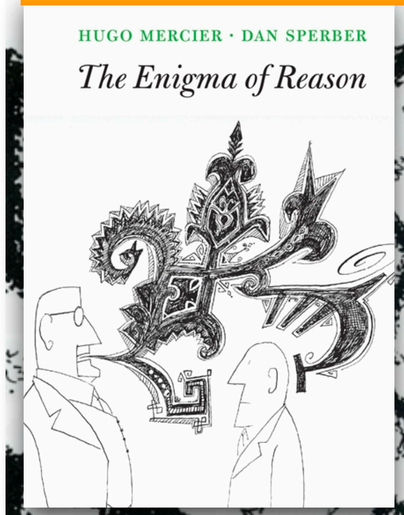
REASONING
SYSTEM 2

- solving puzzles
- writing programs
- proving logic theorems

- reviewing CVPR papers
- crafting CVPR rebuttals
- giving an invited talk
- writing an op-ed

=> Humans often spend hours (or days) ***not*** doing this sort of reasoning at all…

Examples from p58 - p60 of



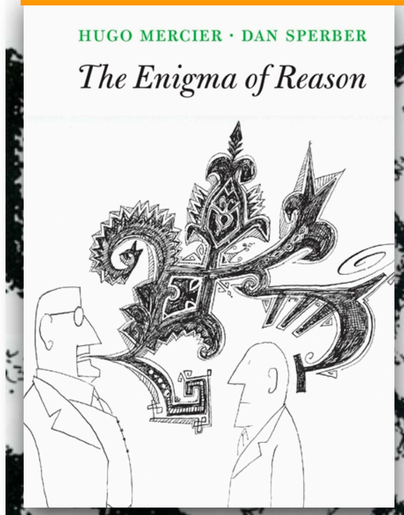
Roger Shepard's "monsters in a tunnel"

- **Two monsters are running** (rather than standing still on one foot)
- **One is chasing another** (rather than trying to copy his movements)
- **The chaser has hostile intentions and the chased is afraid** (even though two faces are identical)

INTUITION
SYSTEM 1

- Intuitive inferences on
  - pre-conditions and post-conditions
  - what happens before and after?
  - motivations and intents
  - mental and emotional states

Examples from p58 - p60 of



Roger Shepard's "monsters in a tunnel"

- **Two monsters are running** (rather than standing still on one foot)
- **One is chasing another** (rather than trying to copy his movements)
- **The chaser has hostile intentions and the chased is afraid** (even though two faces are identical)

Important Observations:

- None of these inferences is absolutely true. The inferences are stochastic in nature. Everything is defeasible with additional context.
- A great deal of intuitive inferences are commonsense inferences, a great deal of which can be best described in natural language — **full scope of language**, not **just words**, or even **graphs of words**

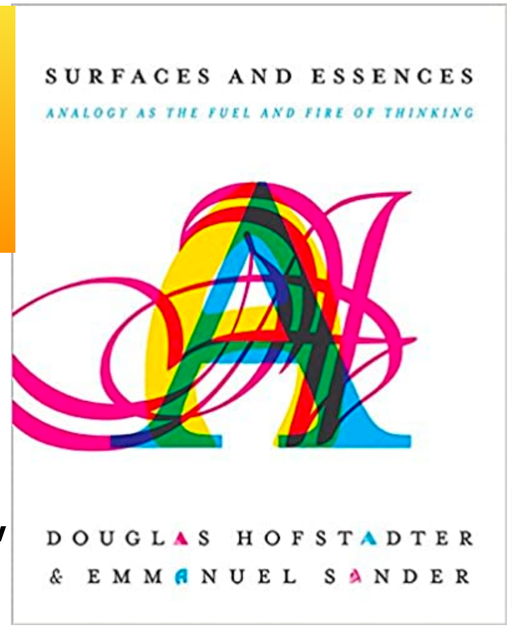
INTUITION
SYSTEM 1

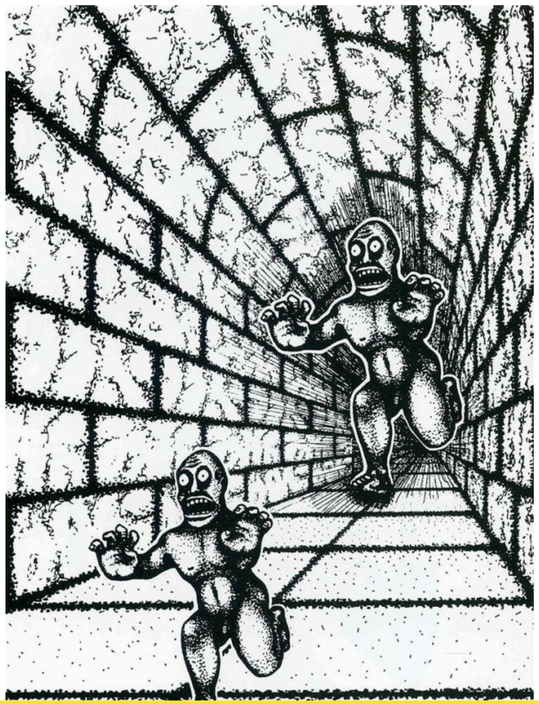
- Intuitive inferences on
  - pre-conditions and post-conditions
  - what happens before and after?
  - motivations and intents
  - mental and emotional states

Tldr; 🔥 Language and Symbols 🔥 🔥 Reasoning as Generation🔥

1. ***Language*** as the ***symbols***
   - all of it —
     - not just ***words*** (ImageNet labels, …)
     - not just small ***sets of words*** (scene graphs,
2. Reasoning as ***generative tasks***
   - As opposed to ***discriminative tasks*** (i.e., categorization)
   - Because the space of reasoning in language is ***infinite***

"Categories (concepts) vastly outnumber words, and require free-form open text descriptions"



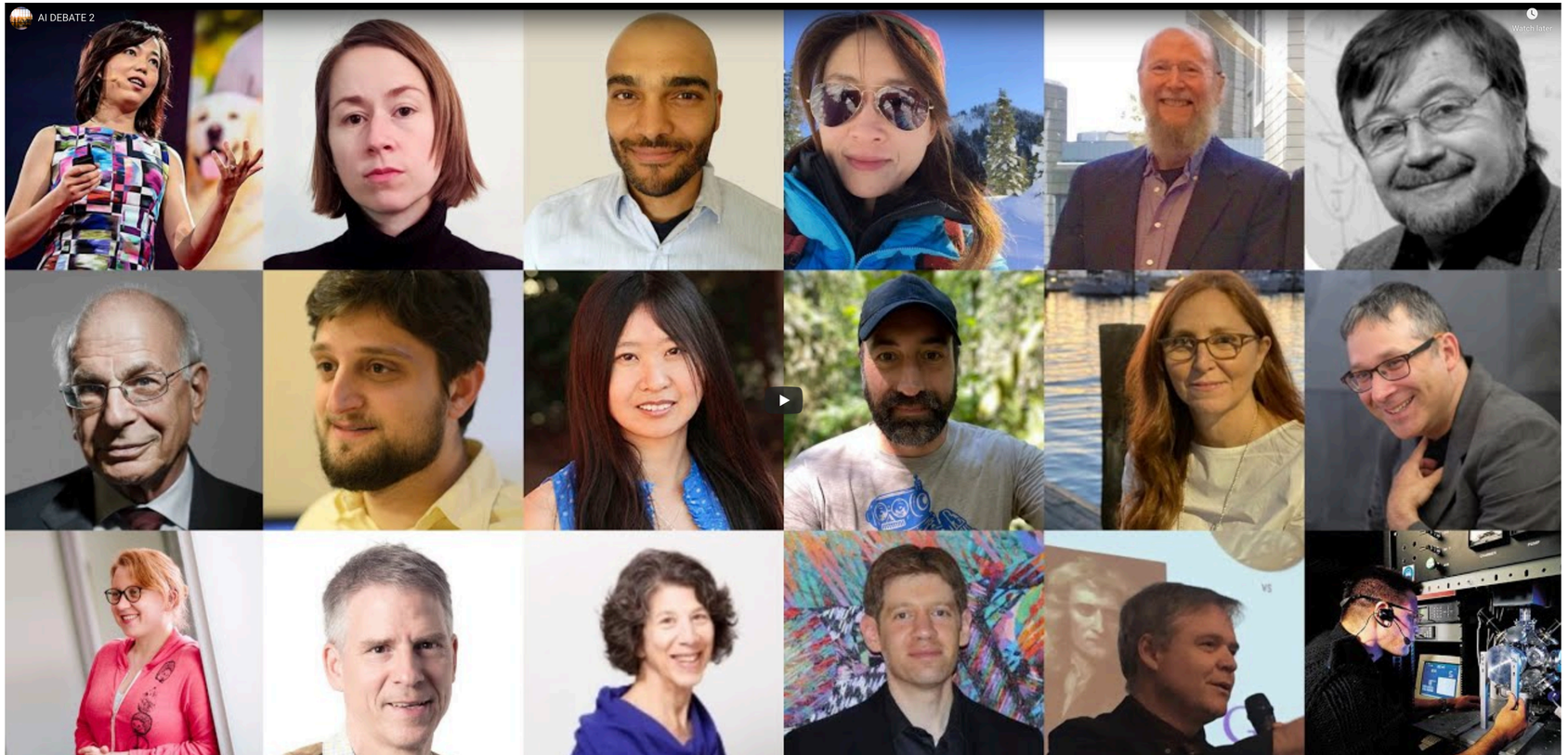


"thinking out loud"

We often think as we speak, on the fly, word-by-word without enumerating all possible alternative sentences

AI Debate 2 at montreal.ai

- https://montrealartificialintelligence.com/aidebate2.html
  - **Daniel Kahneman: https://youtu.be/2zNd69ZGZ8o**
- https://venturebeat.com/2021/01/02/leading-computer-scientists-debate-the-next-steps-for-ai-in-2021/
- Stay tuned for AI Dabate 3 (Dec 23 2021!)



In this talk: Reasoning as Generation

- **Part 1:** unsupervised inference-time algorithms

Reasoning thru Neural Backpropagation	Reasoning thru Search with Logical Constraints	Reasoning thru Distributional Neural Imagination
DeLorean	NeuroLogic	Reflective Decoding

=> How to make 

from (off-the-shelf) neural language models

- **Part 2:** supervised, but with declarative knowledge

?	?	?

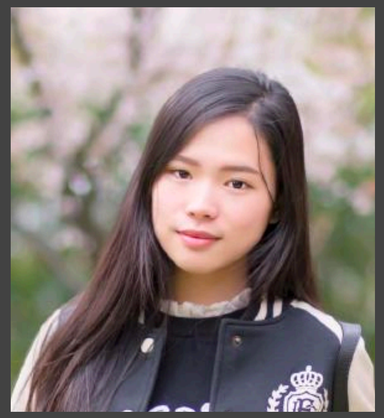
- **Part 3:** benchmarks and algorithmic bias reduction

Back to the Future:

Unsupervised Backprop-based Decoding for Counterfactual and Abductive Commonsense Reasoning

EMNLP 2020

Lianhui Qin



Vered Shwartz

Peter West

Chandra Bhagavatula

Jena Hwang

Ronan LeBras

Antoine Bosselut

Me

Abductive Reasoning

(Bhagavatula et al., 2019)

Past Observation

Ray hung a tire on a rope to make his daughter a swing.

What happened in between?

Future Observation

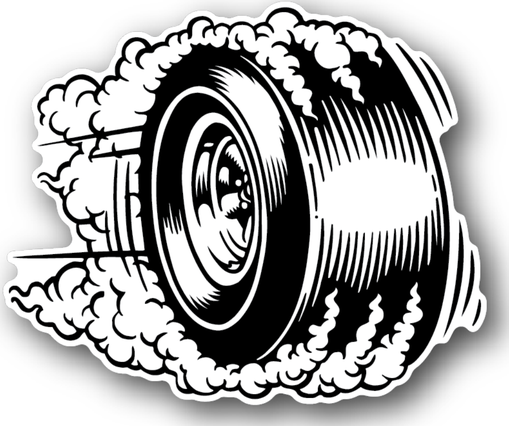
Ray ran to his daughter to make sure she was okay.

Abductive Reasoning

(Bhagavatula et al., 2019)

Past Observation

Ray hung a tire on a rope to make his daughter a swing.



Hypothesis

She hit the rope and the tire fell on top of her.

Future Observation

Ray ran to his daughter to make sure she was okay.

Abductive Reasoning = Inference to the best explanation to partial observation (Peirce 1960)

Abduction != Induction or Deduction

Abductive Reasoning

(Bhagavatula et al., 2019)

Past Observation

Ray hung a tire on a rope to make his daughter a swing.

Hypothesis

She hit the rope and the tire fell on top of her.

Future Observation

Ray ran to his daughter to make sure she was okay.

Counterfactual Reasoning

(Qin et al., 2019)

An example from the "TimeTravel" dataset (Qin et al., EMNLP 2019)

Zeke was throwing a party.

All his friends were dressing up for this Halloween party. Story context changes…

Zeke thought about being a vampire or a wizard.

Then he decided on a scarier costume.

Zeke dressed up like a skeleton.

What if this is a Game of Thrones themed party instead of a Halloween party? 🤔

An example from the "TimeTravel" dataset (Qin et al., EMNLP 2019)

Zeke was throwing a party.

All his friends were dressing up for this Halloween party. Story context changes…

Zeke thought about being a vampire or a wizard.

Then he decided on a scarier costume.

Zeke dressed up like a skeleton.

Story ending doesn't make sense now…

What if this is a Game of Thrones themed party instead of a Halloween party? 🤔

An example from the "TimeTravel" dataset (Qin et al., EMNLP 2019)

Zeke was throwing a party.

All his friends were dressing up for this Halloween party.

Story context changes…

Zeke thought about being a vampire or a wizard.

Then he decided on a scarier costume.

Zeke dressed up like a skeleton.

Story ending doesn't make sense now…

What if this is a Game of Thrones themed party instead of a Halloween party?

Counterfactual Reasoning

(Qin et al., 2019)

Reasoning about the alternative ***future*** based on counterfactual ***past***.

An example from the "TimeTravel" dataset (Qin et al., EMNLP 2019)

Zeke was throwing a party.

All his friends were dressing up for this Halloween party. Story context changes…

Zeke thought about ~~being a vampire or a wizard.~~

~~Then he decided on a scarier costume.~~

Zeke dressed up like a ~~skeleton.~~

Story ending doesn't make sense now…

What if this is a Game of Thrones themed party instead of a Halloween party?

Zeke was throwing a party.

All his friends were dressing up for this Halloween party.

Only do minimal edit!

Zeke thought about *Lannister, but he didn't want to look like a Lannister.*

He wanted to look like a Stark.

Zeke dressed up like a *Stark*.

Consistent now!!

Abductive Reasoning

(Bhagavatula et al., 2019)

Past Observation

Ray hung a tire on a rope to make his daughter a swing.

Hypothesis

She hit the rope and the tire fell on top of her.

Future Observation

Ray ran to his daughter to make sure she was okay.

Story Context

Zeke was throwing a party.

All his friends were dressing up for this Halloween party.

All his friends were dressing up for this Game of Thrones themed party.

Rewritten Ending

Zeke thought about Lannister, but he didn't want to look like a Lannister.

He wanted to look like a Stark.

Zeke dressed up like a Stark.

Original Ending

Zeke thought about being a vampire or a wizard.

Then he decided on a scarier costume.

Zeke dressed up like a skeleton.

Counterfactual Reasoning

(Qin et al., 2019)

Abductive Reasoning

(Bhagavatula et al., 2019)

Both involve ***nonmonotonic reasoning*** with past context X and future constraint Z

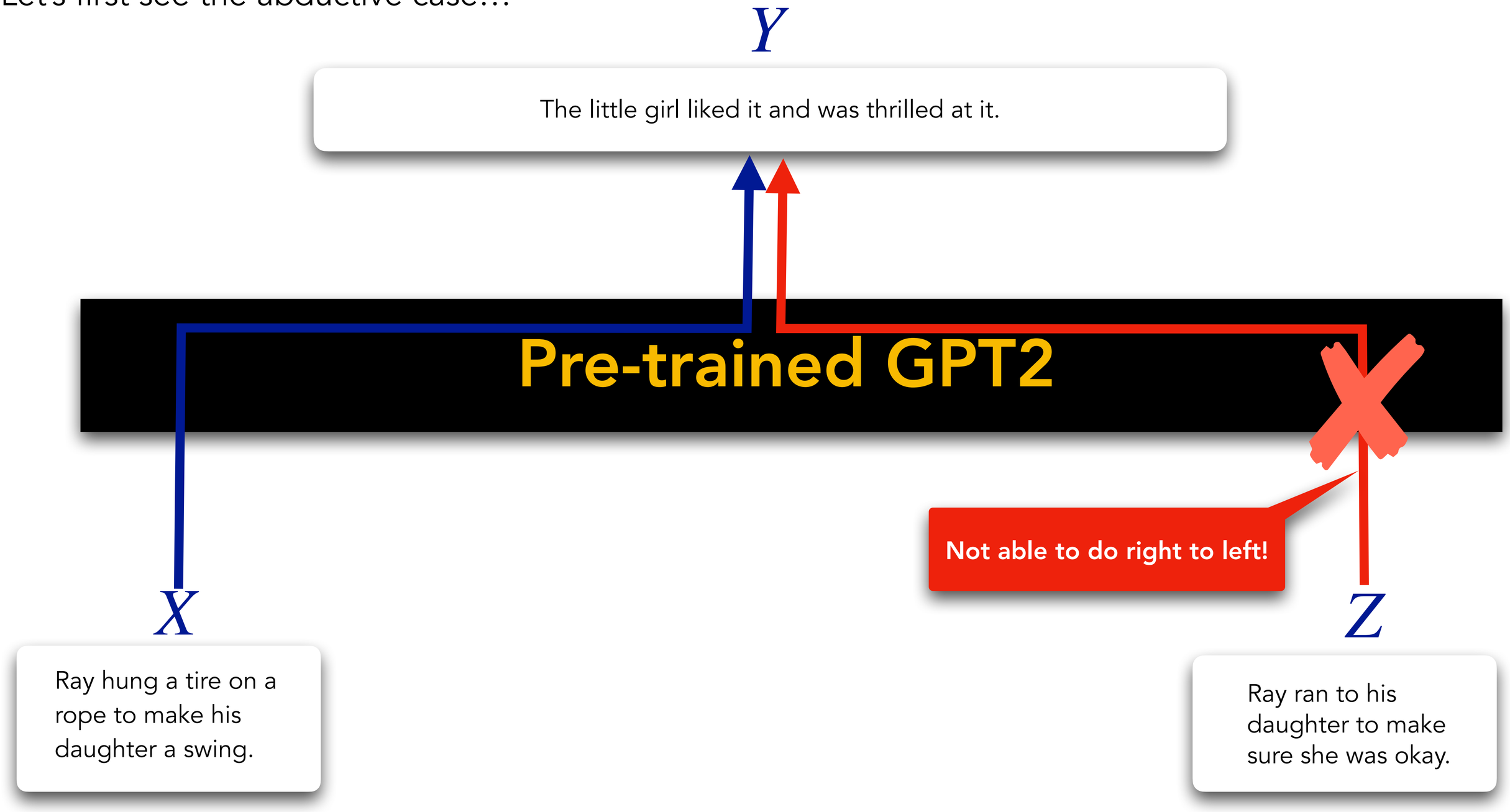
Counterfactual Reasoning

(Qin et al., 2019)

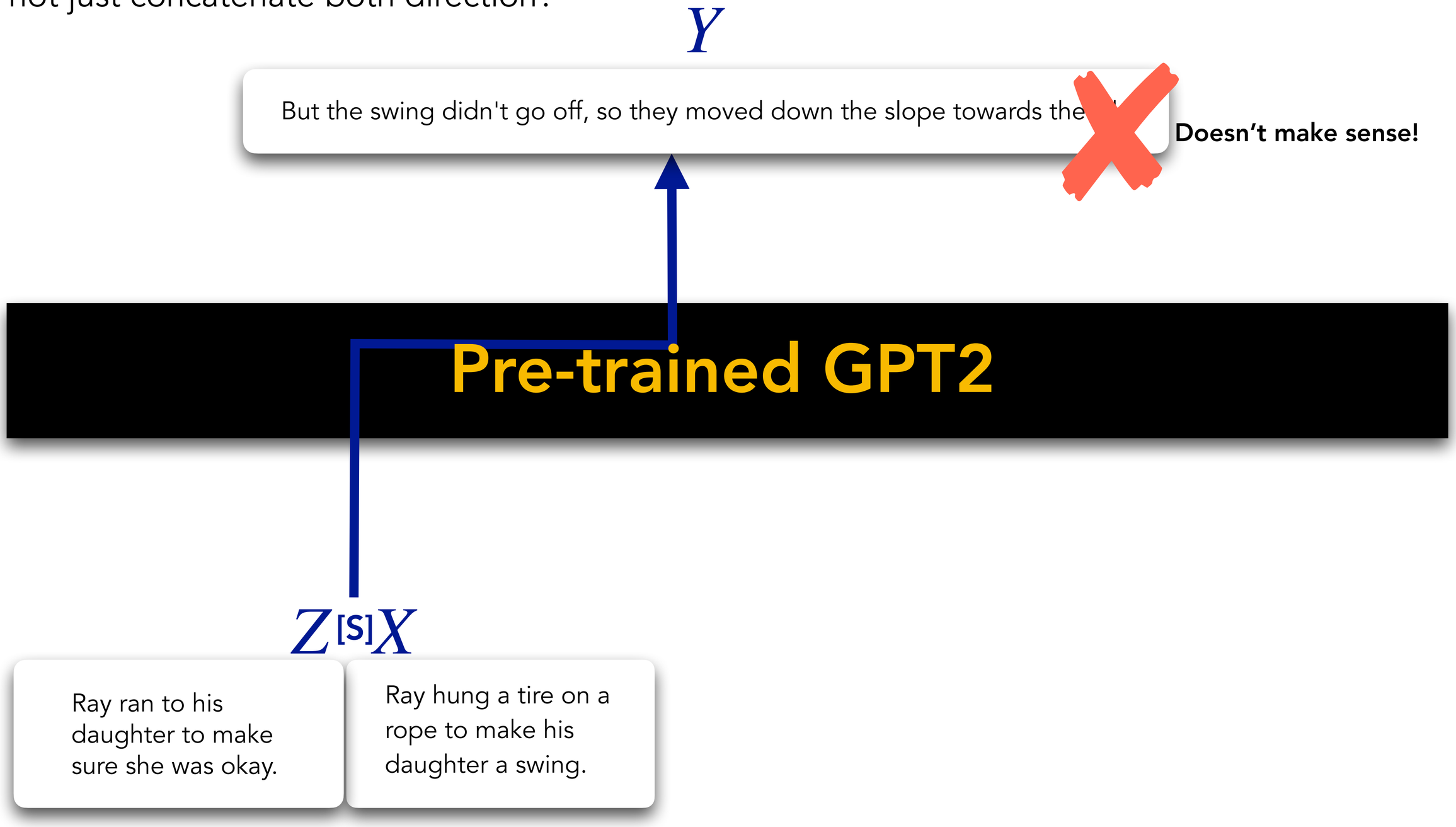
Pretrained Language Models are successful on many tasks…

How are Pretrained LMs on the Nonmonotonic Reasoning?

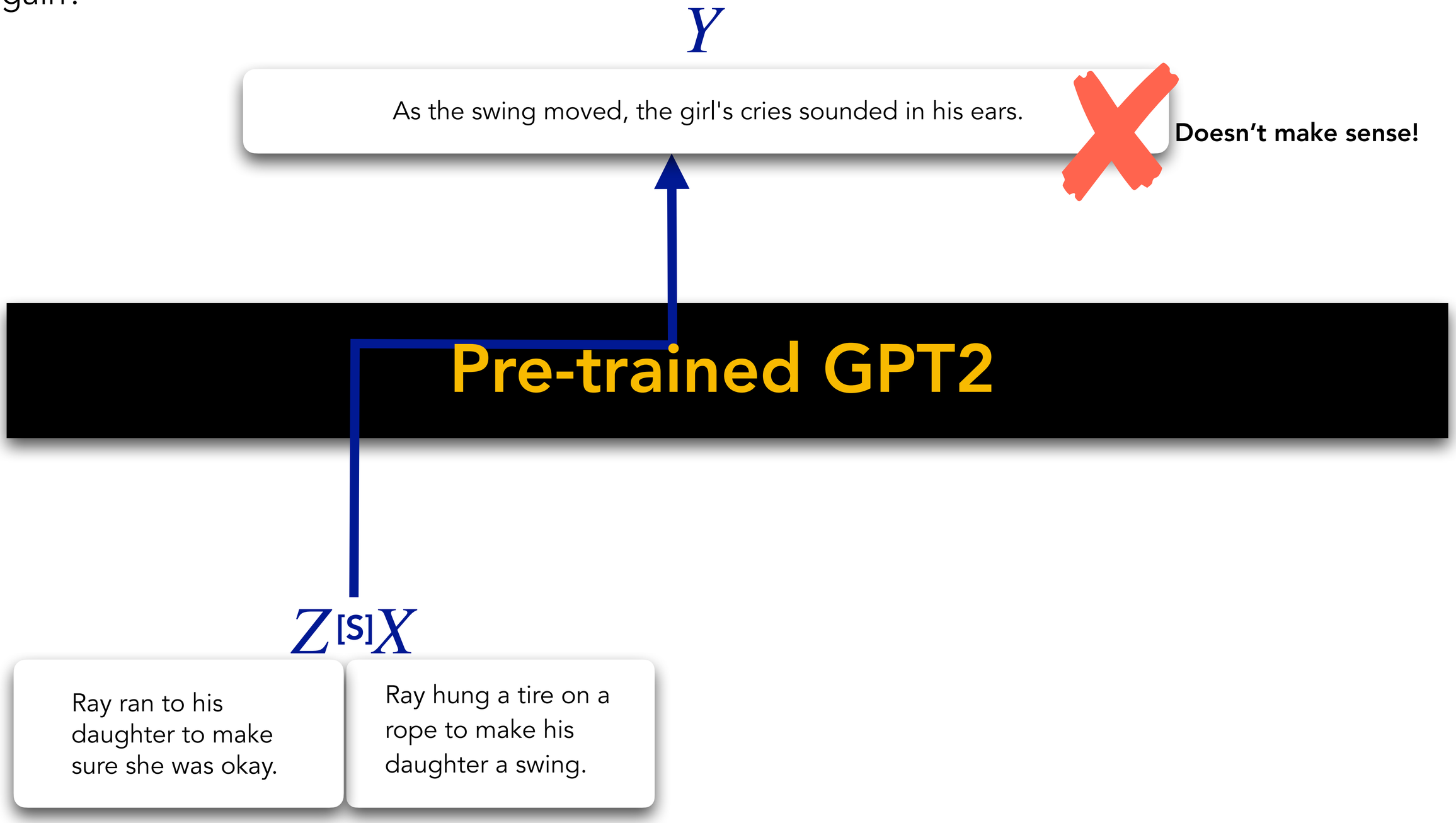
Let’s first see the abductive case…



Why not just concatenate both direction?

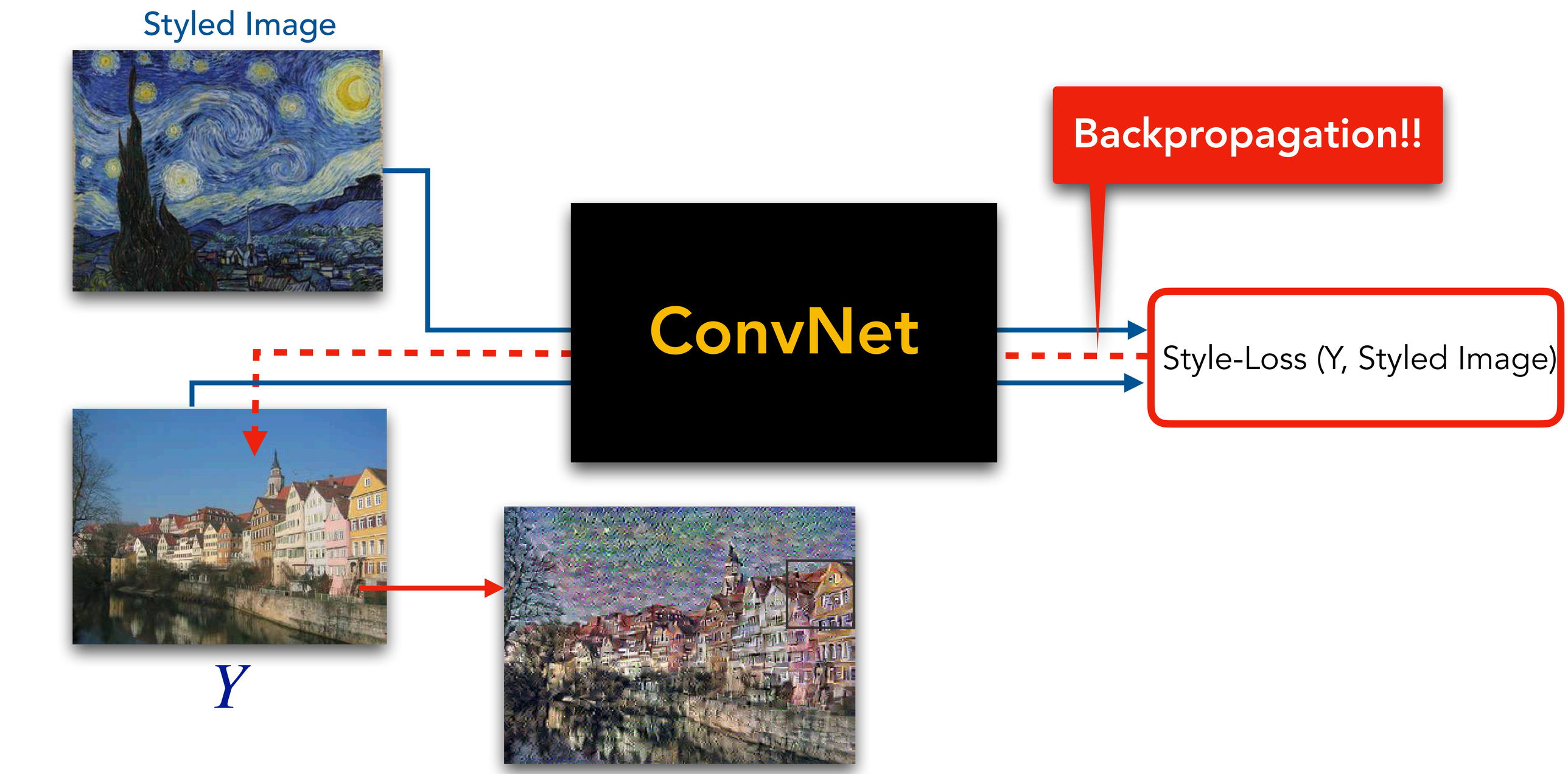


Try again?

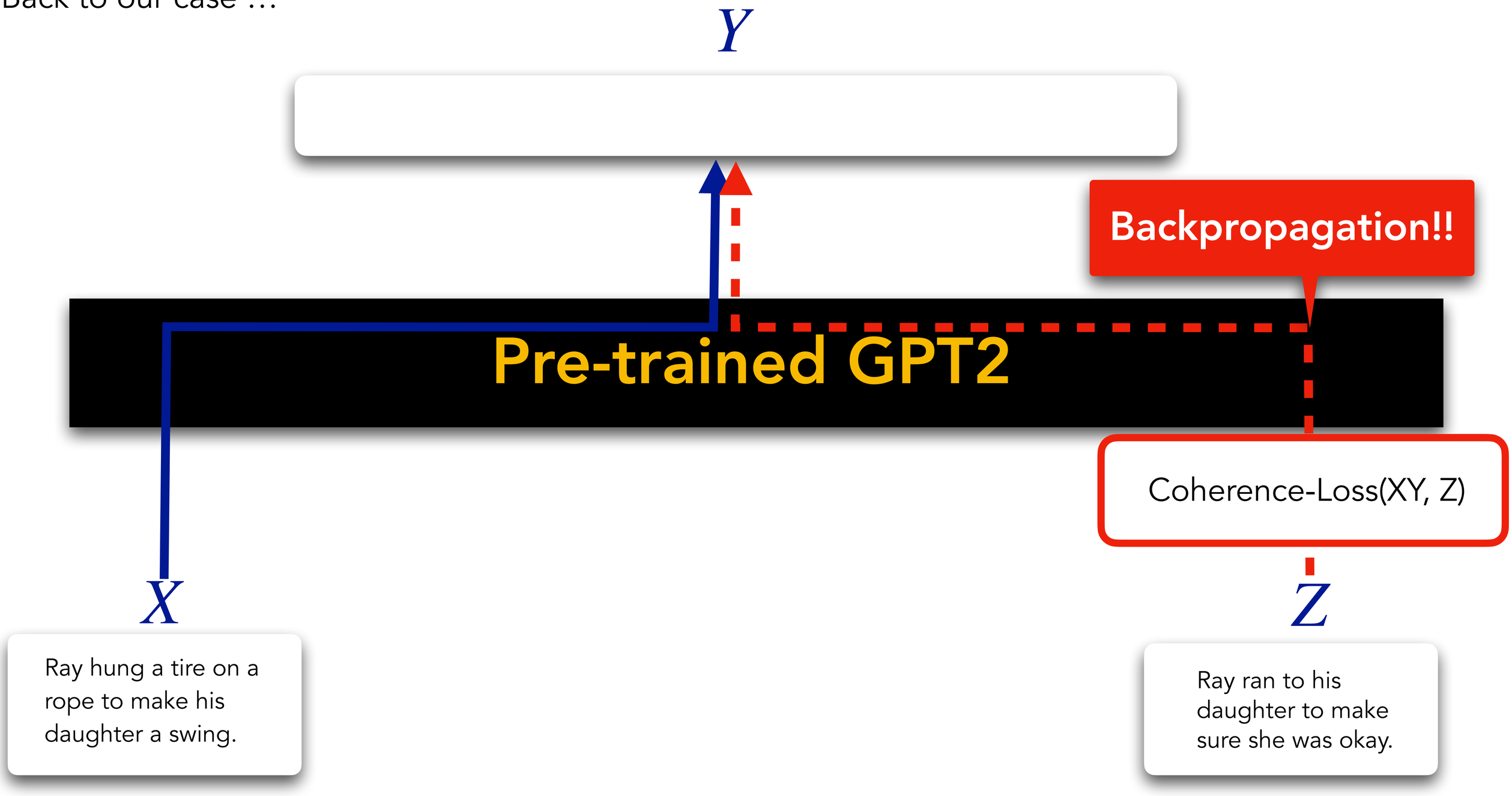

Y
As the swing moved, the girl's cries sounded in his ears.
Doesn't make sense!
Pre-trained GPT2
$Z^{[S]}X$
Ray ran to his daughter to make sure she was okay.
Ray hung a tire on a rope to make his daughter a swing.

Something might have been missing here…

Inspired by “Image Style Transfer” (Gatys et al, 2016)…



Back to our case …



DELOREAN

(DEcoding for nonmonotonic LOgical REAsoNing)

Initialization

Just as how you do regular decoding

Y

$\tilde{y}_1$ $\tilde{y}_2$ ... $\tilde{y}_N$

LM

x_1 x_2 ... x_{N_X}

Ray hung a tire on a rope to make his daughter a swing.

Ray ran to his daughter to make sure she was okay.

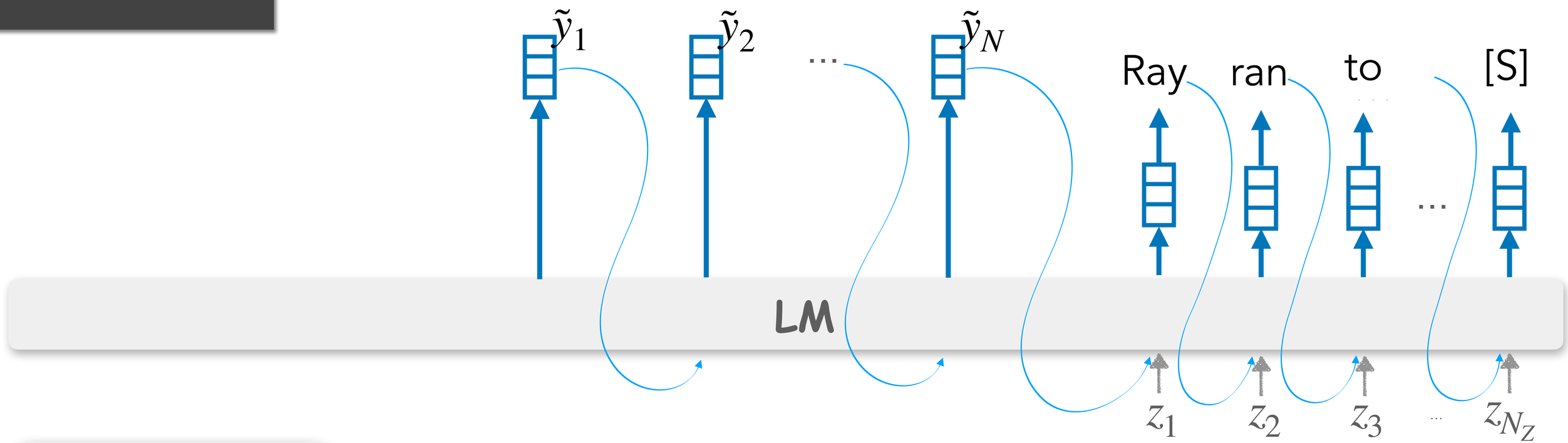
X

Z

Backward Pass

Backpropagate future information

Y



Ray hung a tire on a rope to make his daughter a swing.

Ray ran to his daughter to make sure she was okay.

X

Z

Backward Pass

Backpropagate future information

Loss(Z | X, Y)

Y

$\tilde{y}^b_1$ $\tilde{y}^b_2$ ... $\tilde{y}^b_N$

Backpropagation

$\mathcal{L}(X, \tilde{Y}, Z) := -\sum_{n=1}^{N_Z} \log P_{\mathrm{LM}}(z_n | X, \tilde{Y}, Z_{1:n-1})$

$\tilde{y}_1$ $\tilde{y}_2$ ... $\tilde{y}_N$

Ray ran to ... [S]

LM

z_1 z_2 z_3 ... z_{N_Z}

Ray hung a tire on a rope to make his daughter a swing.

Ray ran to his daughter to make sure she was okay.

X

Z

Forward Pass

Mix both past and future information

Y

$\tilde{y}_1^b$ $\tilde{y}_2^b$ ... $\tilde{y}_N^b$

$\tilde{y}_1$ $\tilde{y}_2$ ... $\tilde{y}_N$

Backpropagation

$\mathcal{L}(X, \tilde{Y}, Z) := -\sum_{n=1}^{N_Z} \log P_{\mathrm{LM}}(z_n | X, \tilde{Y}, Z_{1:n-1})$

Ray ran to ... [S]

LM

z_1 z_2 z_3 ... z_{N_Z}

Ray hung a tire on a rope to make his daughter a swing.

X

Ray ran to his daughter to make sure she was okay.

Z

Forward Pass

Mix both past and future information

Y

$\tilde{y}_1^b$ $\tilde{y}_2^b$ ... $\tilde{y}_N^b$

$\tilde{y}_1$ $\tilde{y}_2$ ... $\tilde{y}_N$

$\tilde{y}_1^f$ $\tilde{y}_2^f$ ... $\tilde{y}_N^f$

Backpropagation

$\mathcal{L}(X, \tilde{Y}, Z) := -\sum_{n=1}^{N_Z} \log P_{\mathrm{LM}}(z_n | X, \tilde{Y}, Z_{1:n-1})$

Ray ran to ... [S]

LM

x_1 x_2 ... x_{N_X}

z_1 z_2 z_3 ... z_{N_Z}

Ray hung a tire on a rope to make his daughter a swing.

Ray ran to his daughter to make sure she was okay.

X

Z

Repeat *T* times

Y

Backpropagation

$\mathcal{L}(X, \tilde{Y}, Z) := -\sum_{n=1}^{N_Z} \log P_{\mathrm{LM}}(z_n | X, \tilde{Y}, Z_{1:n-1})$

$\tilde{y}_1^b$ $\tilde{y}_2^b$ … $\tilde{y}_N^b$

$\tilde{y}_1$ $\tilde{y}_2$ … $\tilde{y}_N$

$\tilde{y}_1^f$ $\tilde{y}_2^f$ … $\tilde{y}_N^f$

Ray ran to … [S]

LM

x_1 x_2 … x_{N_X}

z_1 z_2 z_3 … z_{N_Z}

Ray hung a tire on a rope to make his daughter a swing.

Ray ran to his daughter to make sure she was okay.

X

Z

Sampling

Y

Output: She hit the rope and the tire fell on top of her.

Backpropagation

$\mathcal{L}(X, \tilde{Y}, Z) := -\sum_{n=1}^{N_Z} \log P_{\mathrm{LM}}(z_n | X, \tilde{Y}, Z_{1:n-1})$

$\tilde{y}_1^b$ $\tilde{y}_2^b$ ... $\tilde{y}_N^b$

$\tilde{y}_1$ $\tilde{y}_2$ ... $\tilde{y}_N$

$\tilde{y}_1^f$ $\tilde{y}_2^f$ ... $\tilde{y}_N^f$

Ray ran to ... [S]

LM

x_1 x_2 ... x_{N_X}

z_1 z_2 z_3 ... z_{N_Z}

Ray hung a tire on a rope to make his daughter a swing.

Ray ran to his daughter to make sure she was okay.

X

Z

Counterfactual Reasoning?

Y

Backpropagation

?

$\tilde{y}^b_1$ $\tilde{y}^b_2$ … $\tilde{y}^b_N$

$\tilde{y}_1$ $\tilde{y}_2$ … $\tilde{y}_N$

$\tilde{y}^f_1$ $\tilde{y}^f_2$ … $\tilde{y}^f_N$

LM

x_1 x_2 … x_{N_X}

z_1 z_2 z_3 … z_{N_Z}

X

Z

Counterfactual Reasoning?

Distance-Loss: (Y, Z)

Y

$\tilde{y}^b_1$ $\tilde{y}^b_2$ ... $\tilde{y}^b_N$

Backpropagation

$\mathcal{L}(X, \tilde{Y}, Z) := \mathrm{KL}(Z \| \mathrm{softmax}(\tilde{Y}/\tau))$

?

$\tilde{y}_1$ $\tilde{y}_2$ ... $\tilde{y}_N$

$\tilde{y}^f_1$ $\tilde{y}^f_2$ ... $\tilde{y}^f_N$

LM

x_1 x_2 ... x_{N_X}

z_1 z_2 z_3 ... z_{N_Z}

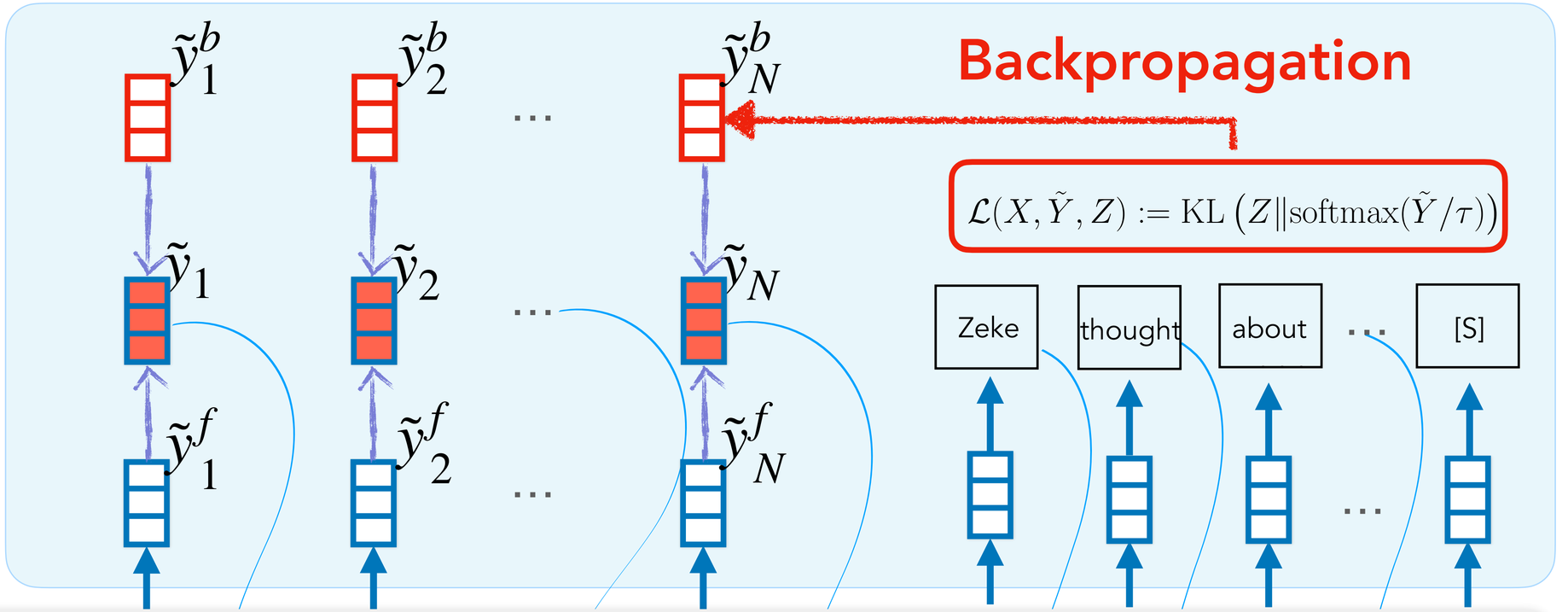
X

Z

Counterfactual Reasoning?

Distance-Loss: (Y, Z)

Y



LM

x_1 x_2 ... x_{N_X}

z_1 z_2 z_3 ... z_{N_Z}

Zeke was throwing a party. [Counterfactual] All his friends were dressing up for this Game of Thrones themed party.

Zeke thought about being a vampire or a wizard. Then he decided on a scarier costume. Zeke dressed up like a skeleton.

X

Z

Counterfactual Reasoning?

Distance-Loss: (Y, Z)

Y

Zeke thought about Lannister, but he didn't want to look like a Lannister. He wanted to look like a Stark.Zeke dressed up like a *Stark*.

Backpropagation

$$\mathcal{L}(X, \tilde{Y}, Z) := \mathrm{KL}\left(Z \| \mathrm{softmax}(\tilde{Y}/\tau)\right)$$

$\tilde{y}^b_1$ $\tilde{y}^b_2$ … $\tilde{y}^b_N$

$\tilde{y}_1$ $\tilde{y}_2$ … $\tilde{y}_N$

$\tilde{y}^f_1$ $\tilde{y}^f_2$ … $\tilde{y}^f_N$

Zeke thought about … [S]

LM

x_1 x_2 … x_{N_X}

z_1 z_2 z_3 … z_{N_Z}

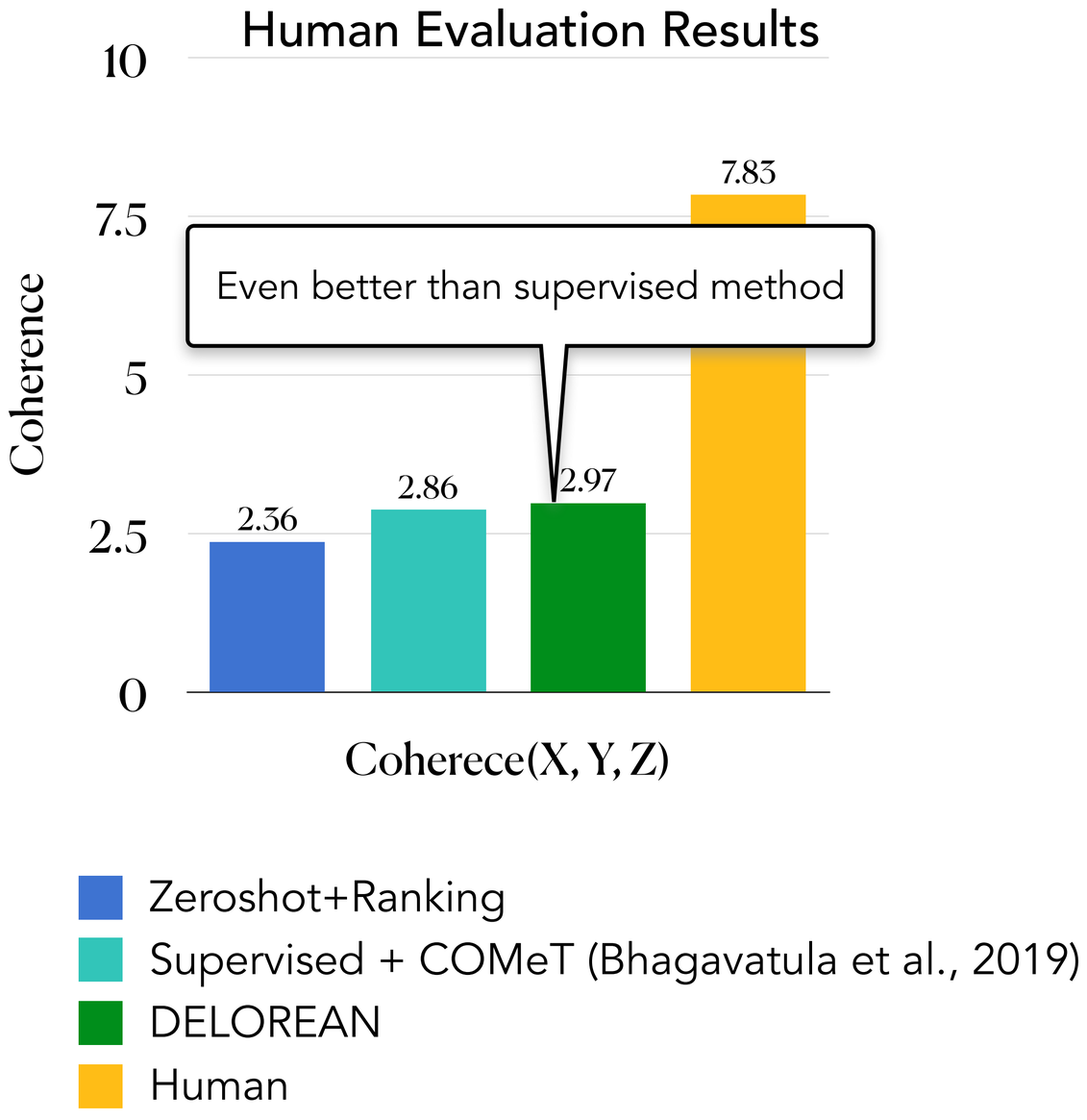
Zeke was throwing a party. [Counterfactual] All his friends were dressing up for this Game of Thrones themed party.

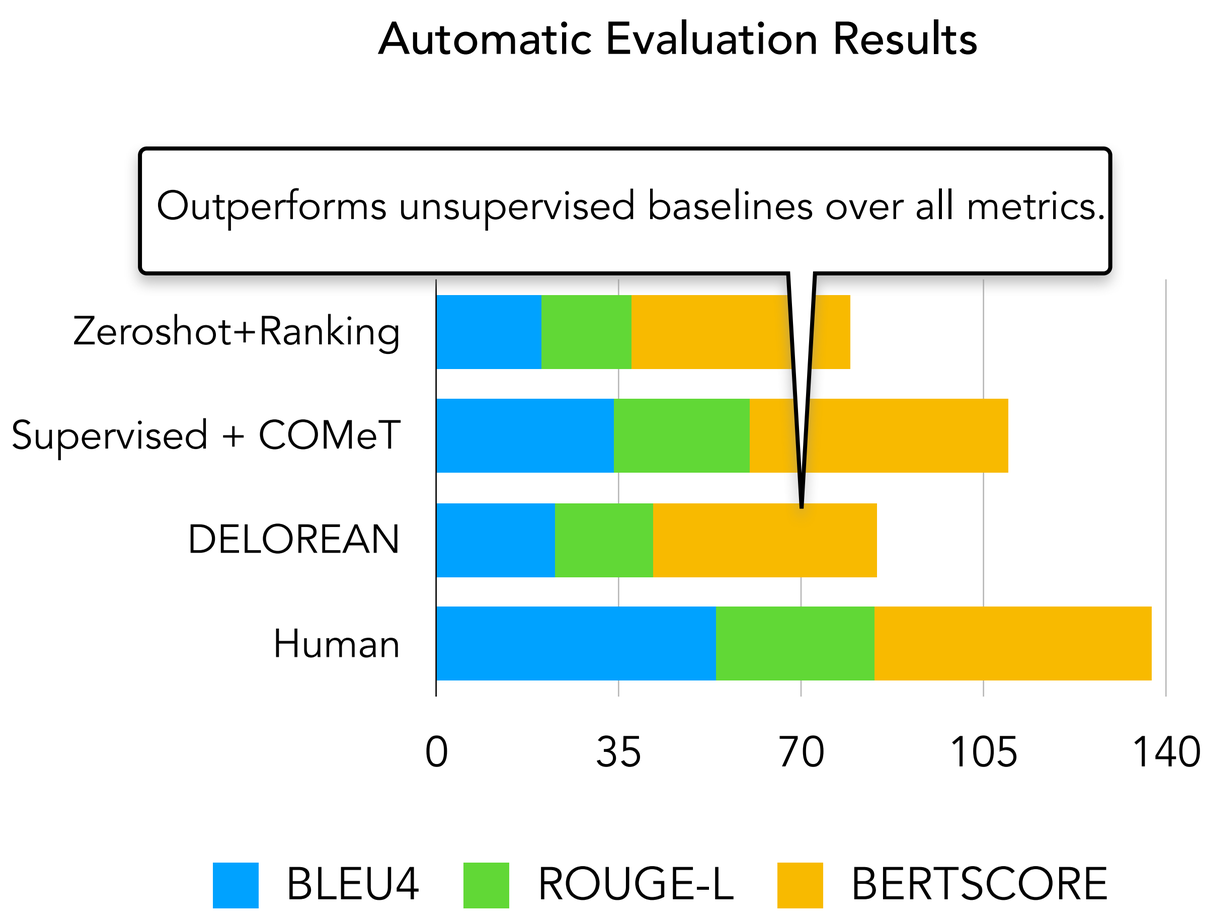
Zeke thought about being a vampire or a wizard. Then he decided on a scarier costume. Zeke dressed up like a skeleton.

X

Z

Abductive Reasoning



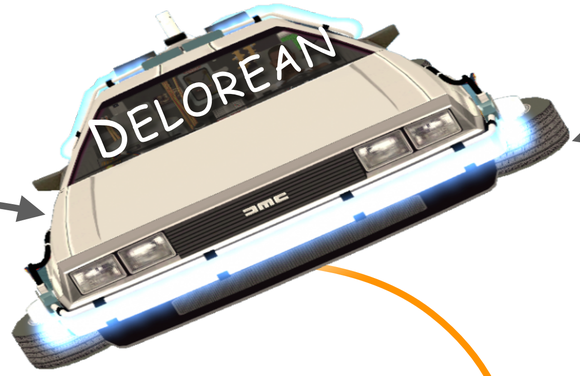


Please check the paper for more baselines …

Abductive Reasoning

Past Observation

Ray drove his car on a steep mountain road.



Future Observation

Ray was fine but his car was totaled.

Hypothesis

As he drove the car to the top of the mountain, his car is hit by a car.

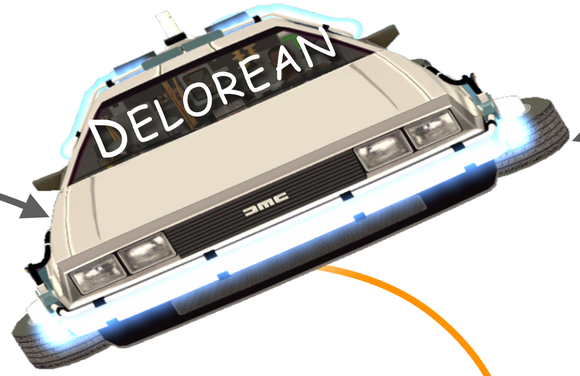
Counterfactual Reasoning

Story Context

Tara wanted to buy a new shirt for her upcoming school formal.

[Original] She went to the mall with her mom.

[Counterfactual] She knew of a cool place online that did custom fits really cheaply, and ordered from there.



Original Ending

They browsed shirts from a variety of stores.

Tara picked out a floral patterned shirt that she liked best.

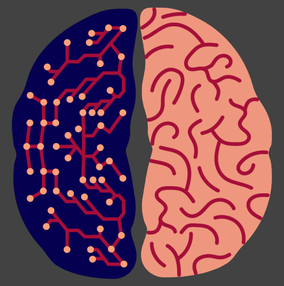
Tara looked forward to wearing it.

Rewritten Ending

They sent her a shirt that fit her perfectly.

Tara was so excited to wear it.

She looked forward to wearing it.



NeuroLogic Decoding

Controlling Neural Language Generation with Logic Constraints

Ximing Lu

Peter West

Ronan LeBras

Rowan Zellers

Chandra Bhagavatula

Yejin Choi

Seq2Seq

Machine Translation

X

The physician told the baker that she had cancer.

Y

Der Arzt sagte dem Bäckerin, dass er Krebs habe.

Dialogue Response

X

type	hotel
count	182
dogs allowed	don't care

Y

There are 182 hotels if you do not care whether dogs are allowed .

Y

Language Model

X

COMMONGEN

X

{ food, table, sit, front }

Y

The man sat with his food at the front of the table.

Image Captioning

X

Y

Man in blue wetsuit is surfing on wave.

CommonGen (Lin et al. EMNLP 2020)

missing keyword { lose, ride }

A man is trying to keep his balance as he falls off a board .

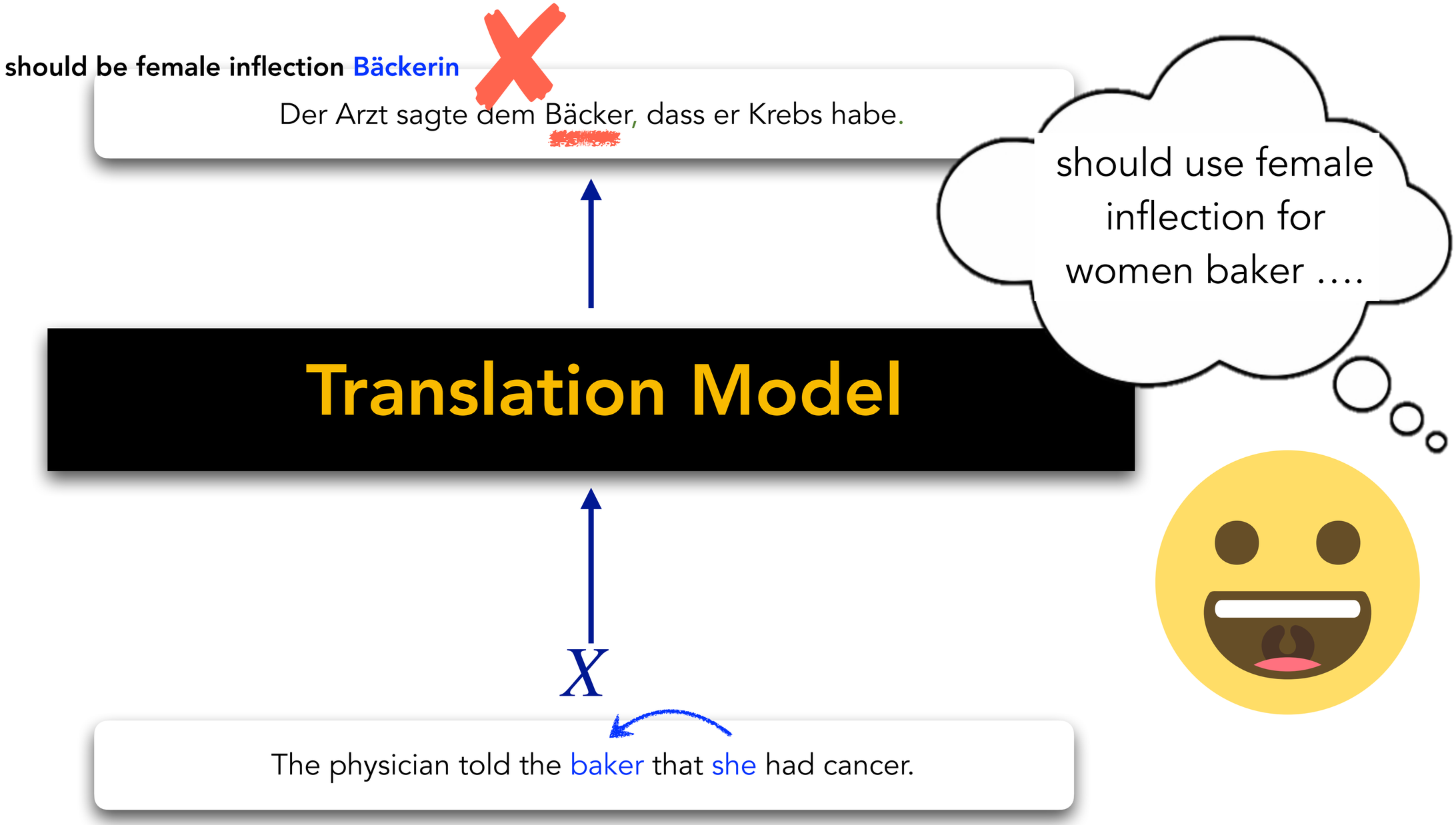
should use **all** given keywords ....

Fine-tuned Language Model

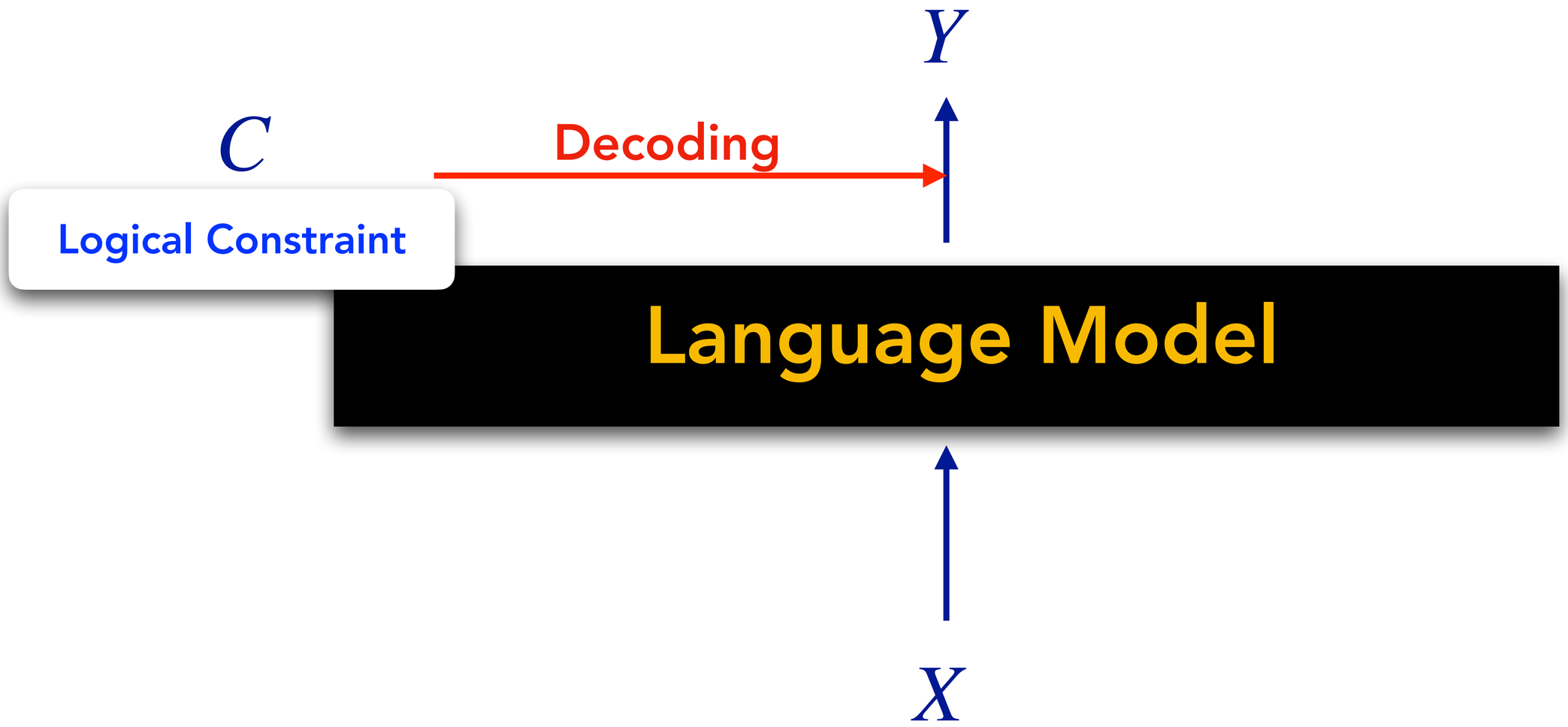
X

{ board, lose, ride, fall, balance }

Machine Translation (Stanovsky et al., 2019)



NEUROLOGIC DECODING



Logical Constraints

Generate response to user's query that must contain retrieved information

$$\mathcal{D}(\text{Hilton Hotel}) \wedge \mathcal{D}(\text{address}) \wedge \mathcal{D}(\text{phone number})$$

Generate a oil-free chow mien recipe that uses
at least one type of proteins and one type of vegetables

$$\mathcal{D}(\text{noodle}) \wedge \neg\mathcal{D}(\text{oil}) \wedge \big(\mathcal{D}(\text{beef}) \vee \mathcal{D}(\text{chicken}) \vee \dots\big) \wedge \big(\mathcal{D}(\text{broccoli}) \vee \mathcal{D}(\text{lettuce}) \vee \dots\big)$$

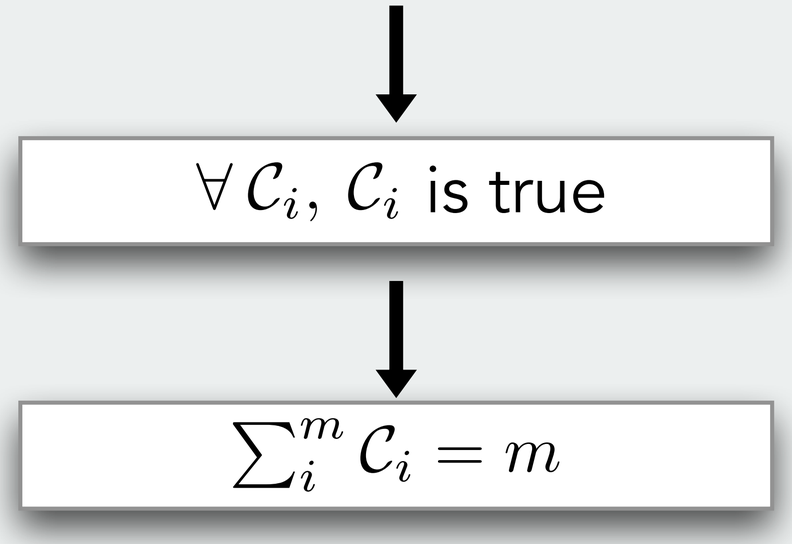
Logical Constraints

CNF Form	a product of sums or an AND of ORs

$$\underbrace{(\mathcal{D}_1 \vee \mathcal{D}_2 \cdots \vee \mathcal{D}_i)}_{\mathcal{C}_1} \wedge \cdots \wedge \underbrace{(\mathcal{D}_k \vee \mathcal{D}_{k+1} \cdots \vee \mathcal{D}_l)}_{\mathcal{C}_m}$$

$\mathcal{D}_i$: literal

$\mathcal{C}_i$: clauses



Objective

$$\hat{\mathbf{y}} = \arg\max_{\mathbf{y}\in\mathcal{Y}} P_\theta(\mathbf{y}|\mathbf{x})$$

subject to

some logical constraints

CNF

$$\hat{\mathbf{y}} = \arg\max_{\mathbf{y}\in\mathcal{Y}} P_\theta(\mathbf{y}|\mathbf{x})$$

subject to

$$\sum_{i=1}^{m} \mathcal{C}_i = m$$

penalty method (Fiacco, 1976)

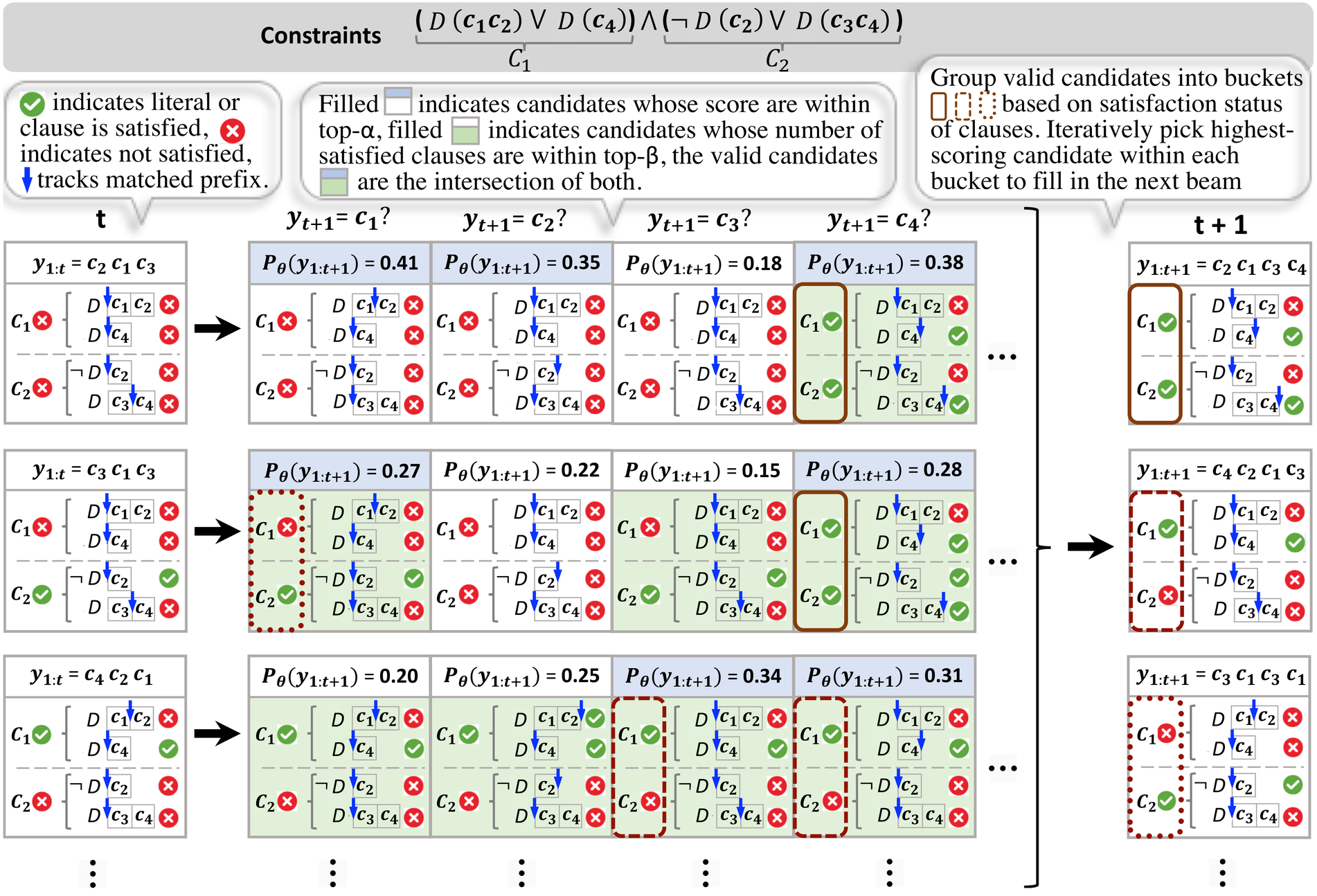
penalty term

$$\hat{\mathbf{y}} = \arg\max_{\mathbf{y}\in\mathcal{Y}} P_\theta(\mathbf{y}|\mathbf{x}) - \alpha\Big(\sum_{i=1}^{m} \mathcal{C}_i - m\Big)^2$$

monotonicity

$$\hat{\mathbf{y}} = \arg\max_{\mathbf{y}\in\mathcal{Y}} P_\theta(\mathbf{y}|\mathbf{x}) + \alpha' \sum_{i=1}^{m} \mathcal{C}_i$$

Perturbed Beam Search



CommonGen

X

{ board, lose, ride, fall, balance }

Finetuned Language Model (i.e., Supervised)

C

$(\textbf{board} \vee \textbf{boards}) \wedge (\textbf{lose} \vee \textbf{loses} \vee \textbf{lost}) \wedge (\textbf{ride} \vee \textbf{rides} \vee \textbf{rode} \vee \textbf{ridden}) \wedge (\textbf{fall} \vee \textbf{falls} \vee \textbf{fell} \vee \textbf{fallen}) \wedge (\textbf{balance} \vee \textbf{balances})$

beam search

Someone loses his balance and falls off his bike .

missing keyword { board, ride }

nucleus sampling

A man is trying to keep his balance as he falls off a board .

missing keyword { lose, ride }

NueroLogic decoding

Someone loses balance on the ride and falls off the board .

CommonGen

X

{ board, lose, ride, fall, balance }

Pretrained Language Model (Unsupervised!!!)

C

(**board** ∨ **boards**) ∧ (**lose** ∨ **loses** ∨ **lost**) ∧ (**ride** ∨ **rides** ∨ **rode** ∨ **ridden**) ∧ (**fall** ∨ **falls** ∨ **fell** ∨ **fallen**) ∧ (**balance** ∨ **balances**)

beam search

The woman, whose name has not been released, was taken to a local hospital, where she was listed in stable condition, according to the sheriff's office.

completely irrelevant

nucleus sampling

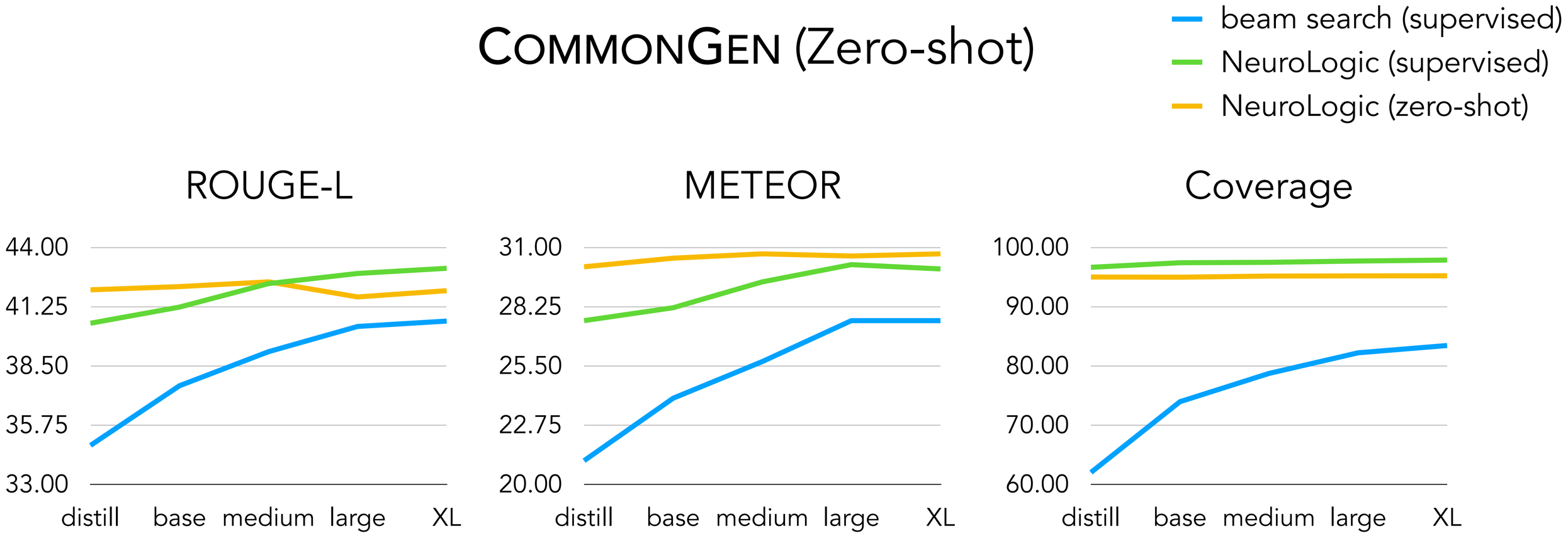
A woman and a man have been arrested in connection with the shooting death of an unarmed black man in Ferguson, Missouri, on Aug. 12, 2014.

completely irrelevant

NueroLogic decoding

A woman lost her balance riding a horse, falling off the horse, and hitting her head on a board .

CommonGen (Zero-shot)



Unsupervised NeuroLogic outperforms **supervised approaches**

Unsupervised NeuroLogic on smaller networks outperforms **supervised approaches on larger networks!**

Recipe Generation

X

Garlic-Butter Steak

{ 2 tsp butter, 1 beef steak, 1/4 tsp soy sauce, 1 tsp parsley, 1/8 tsp salt , 1/2 tsp garlic}

Finetuned Language Model

missing ingredient { soy sauce }
unprovided ingredient { pepper }

C

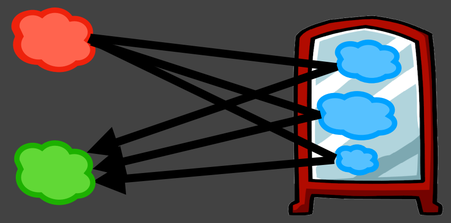
butter $\wedge$ (**beef** $\vee$ **steak** $\vee$ **meat**) $\wedge$ **soy sauce** $\wedge$ (**parsley** $\vee$ **herb**) $\wedge$ **salt** $\wedge$ (**garlic** $\vee$ **vegetable**) $\wedge$ ($\neg$**pork** $\wedge$ $\neg$**bean** $\wedge$ $\neg$…)

beam search

Preheat a grill for high heat. In a bowl, mix together the butter, garlic, parsley, salt and pepper. Place the steak on the hot grill and cook for 3 to 4 minute per side. Transfer the steak to a cutting board and cut it into 1 inch wide strips.

NueroLogic decoding

Mix 1 tablespoon butter, parsley, garlic and soy sauce. Sprinkle steak with salt. In a large skillet, heat remaining butter over medium heat. Add steak; cook until meat reaches desired doneness, 4-7 minutes per side. Serve with garlic butter.



Reflective Decoding: Unsupervised Paraphrasing and Abductive Reasoning

Peter West

Ximing Lu

Ari Holtzman

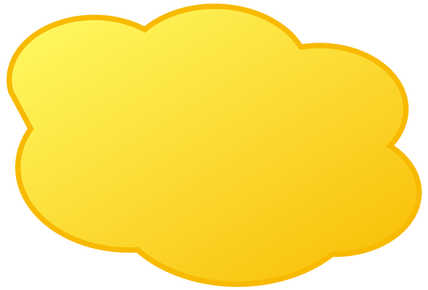
Chandra Bhagavatula

Jena Hwang

Yejin Choi

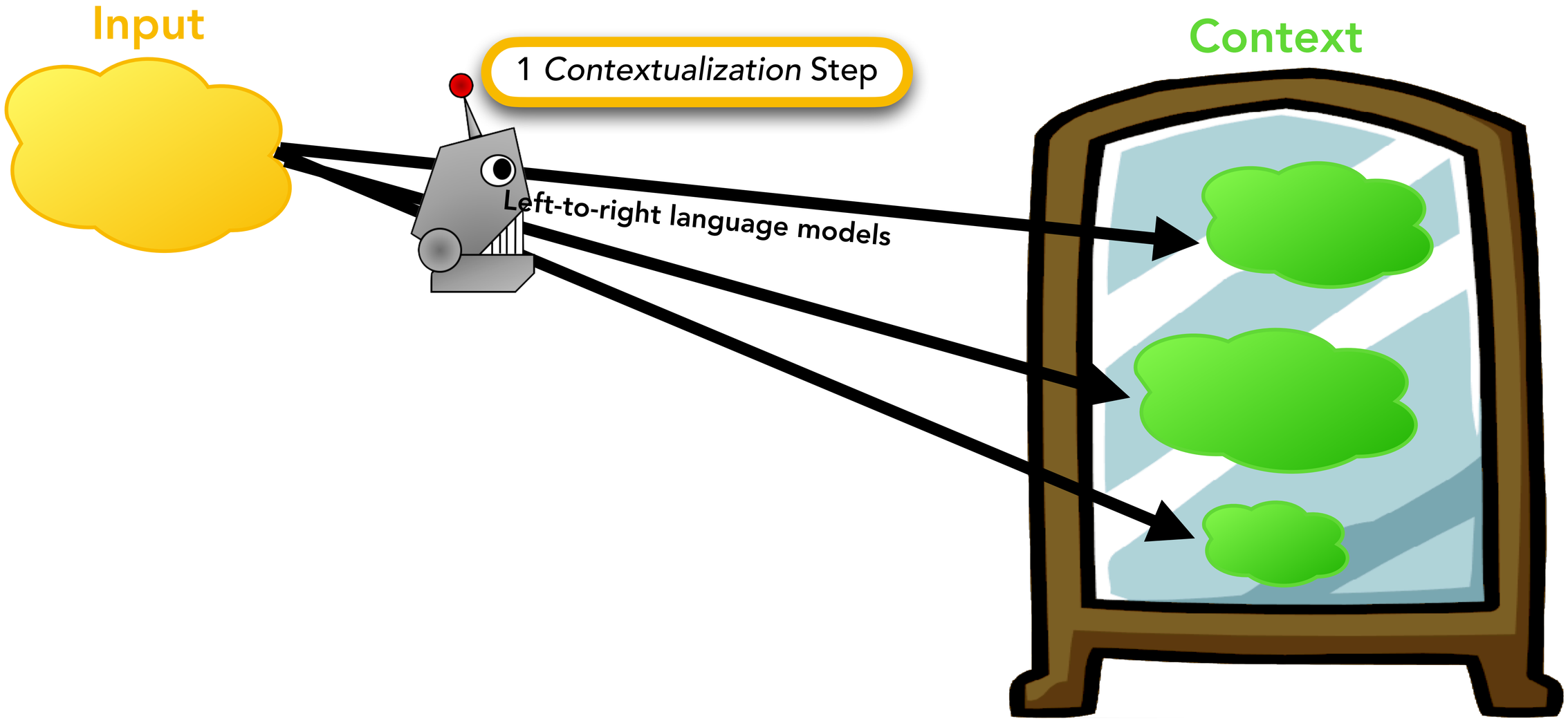
Paraphrasing with Reflective Decoding

Input

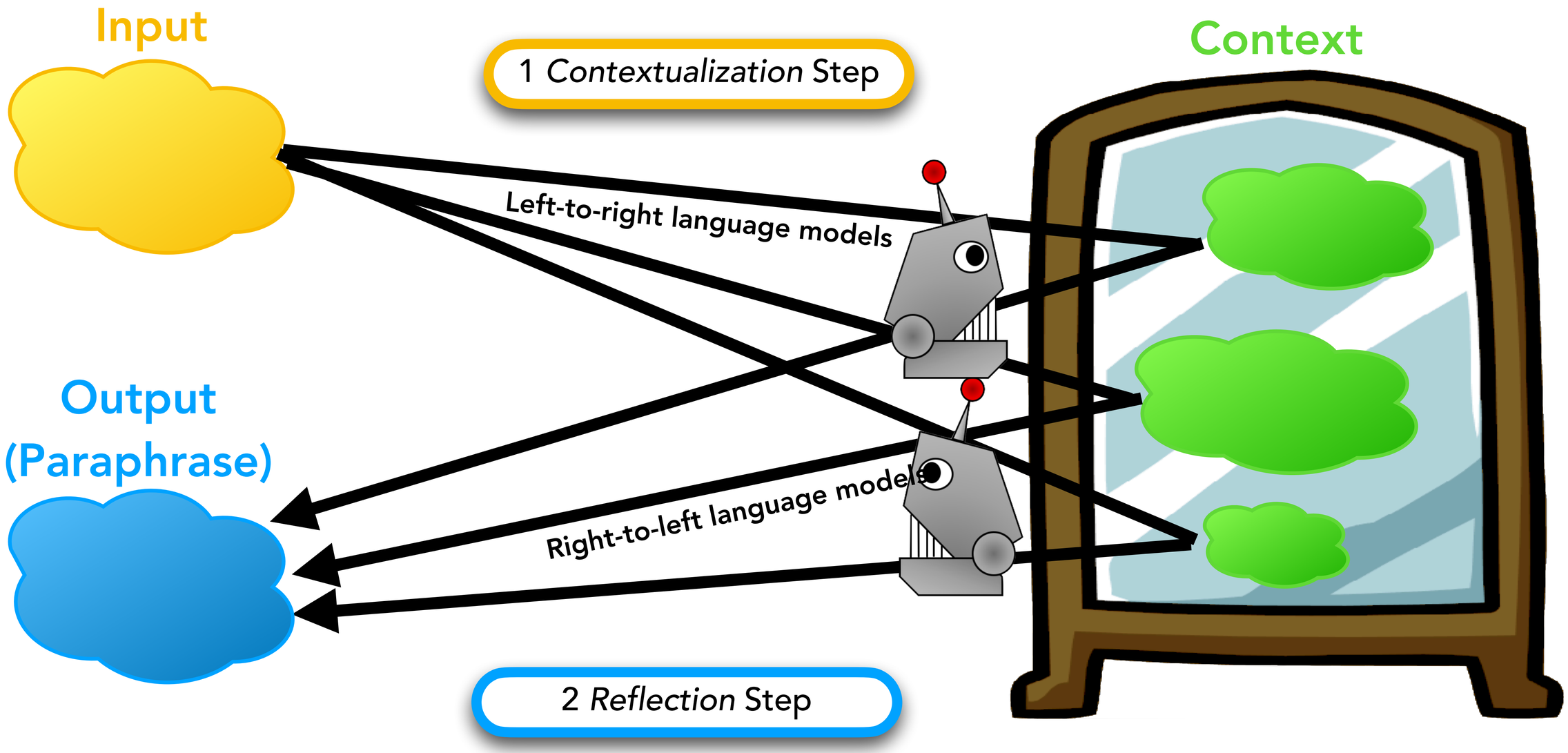


Context

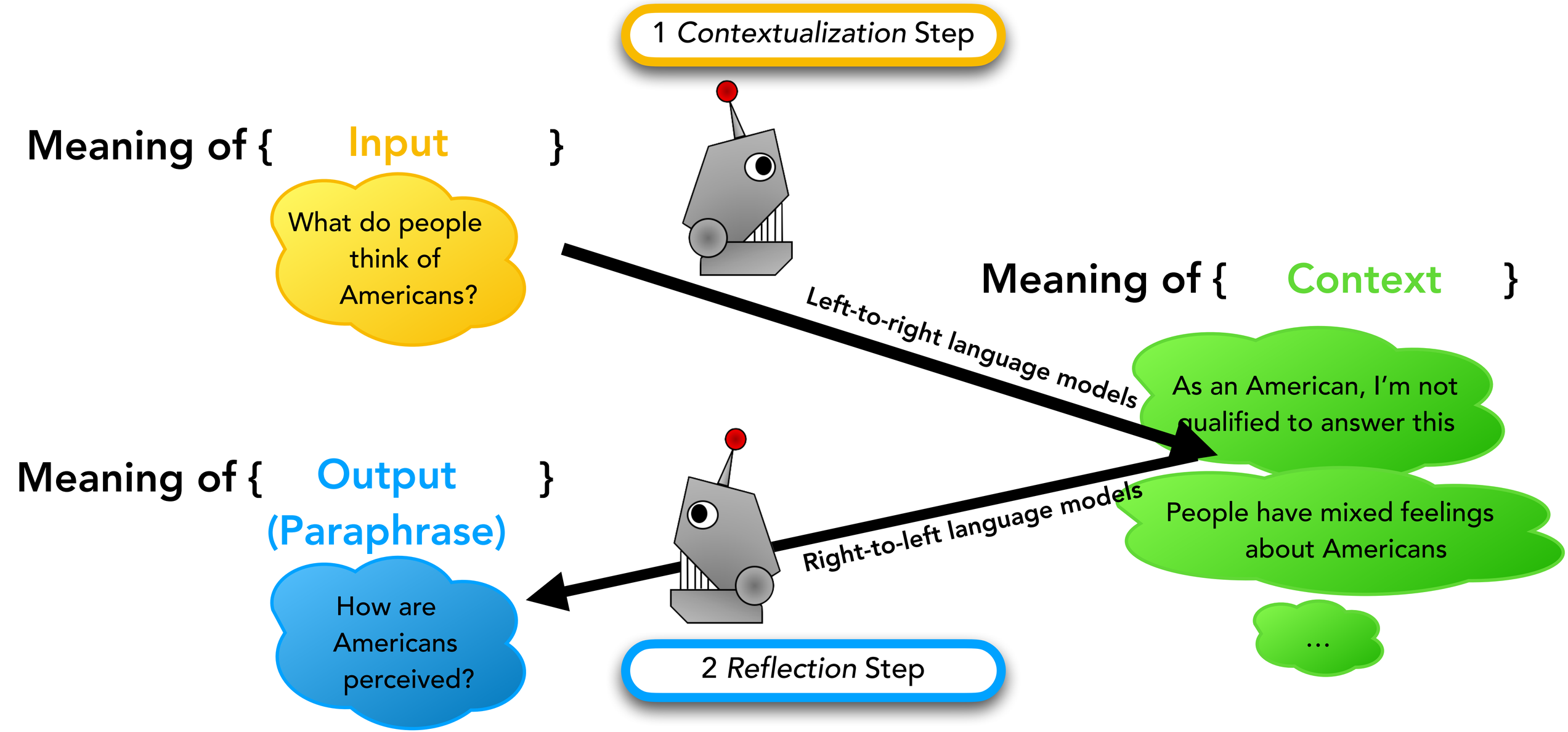
Paraphrasing with Reflective Decoding



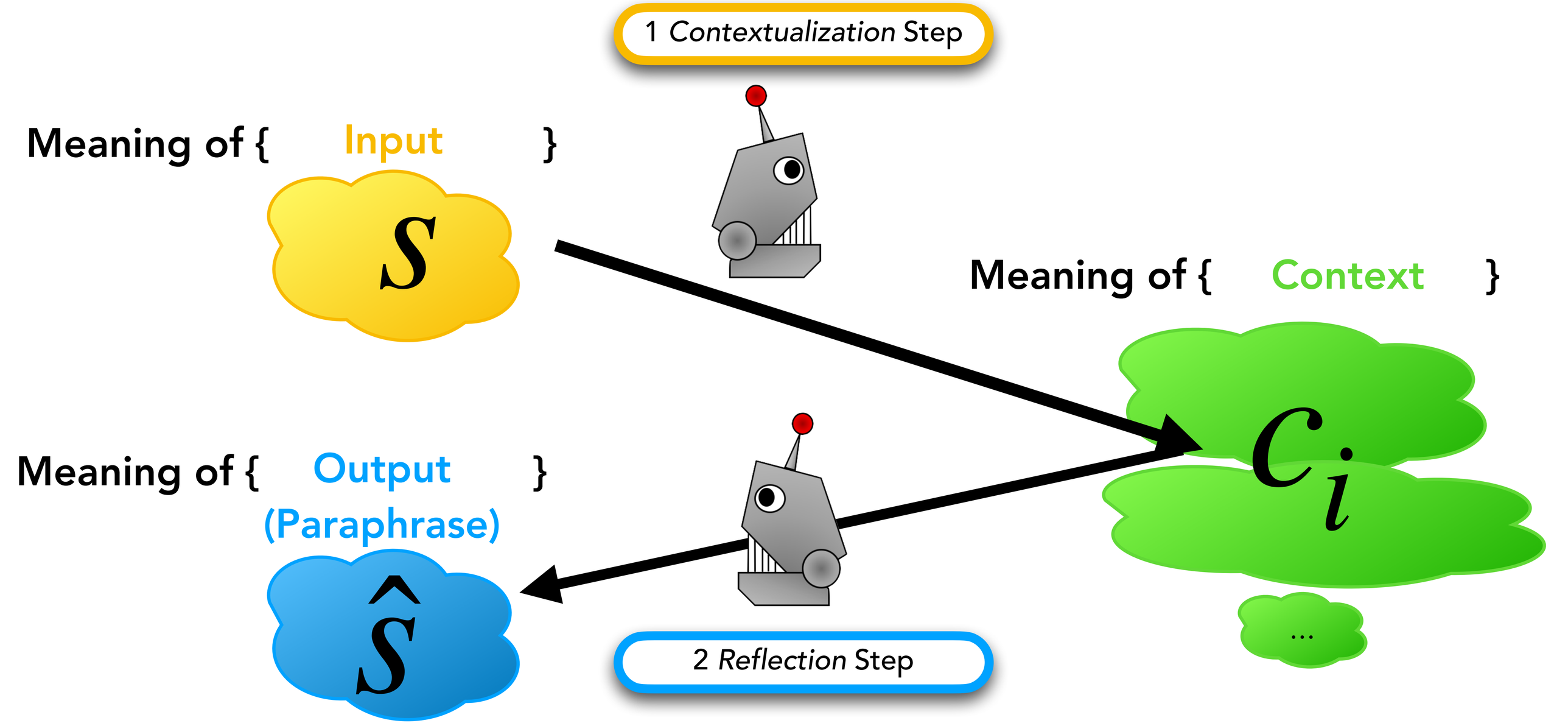
Paraphrasing with Reflective Decoding



Paraphrasing with Reflective Decoding

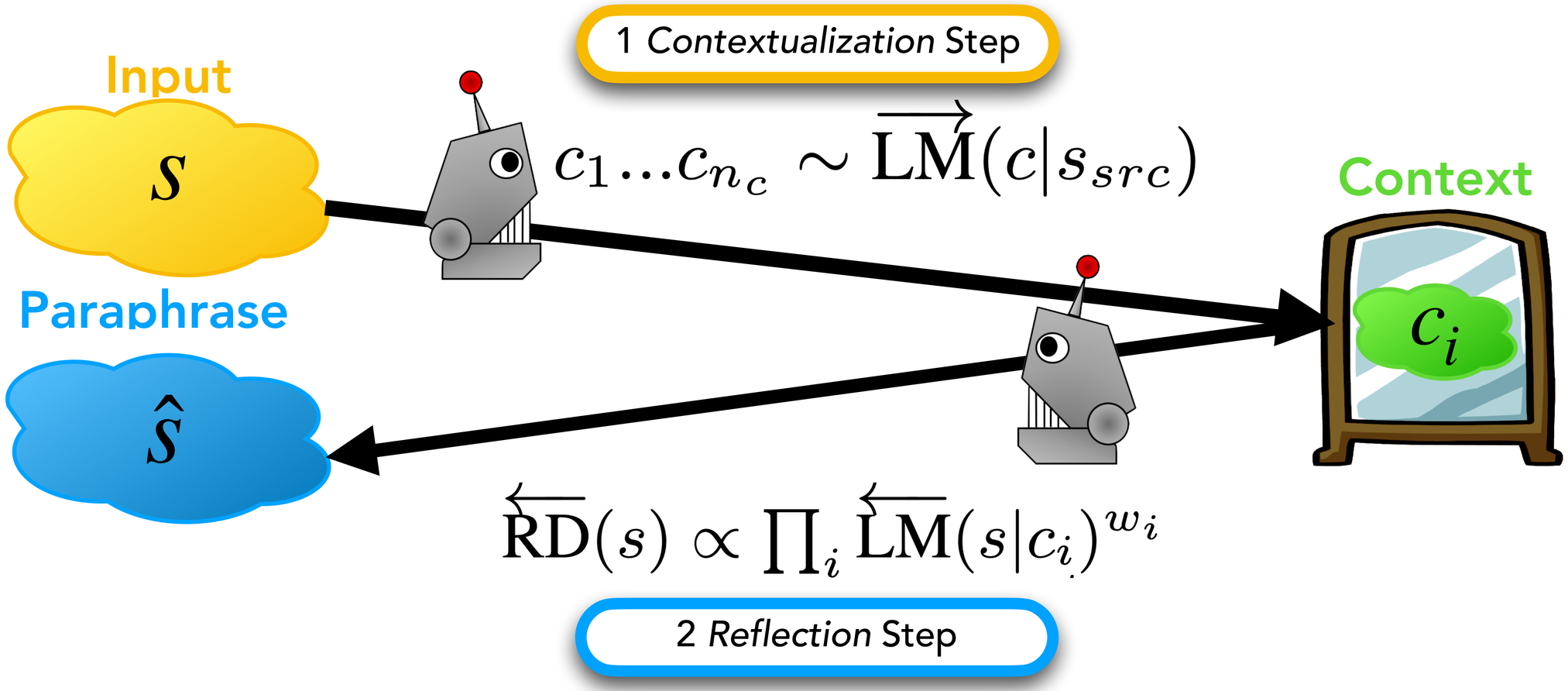


Paraphrasing with Reflective Decoding

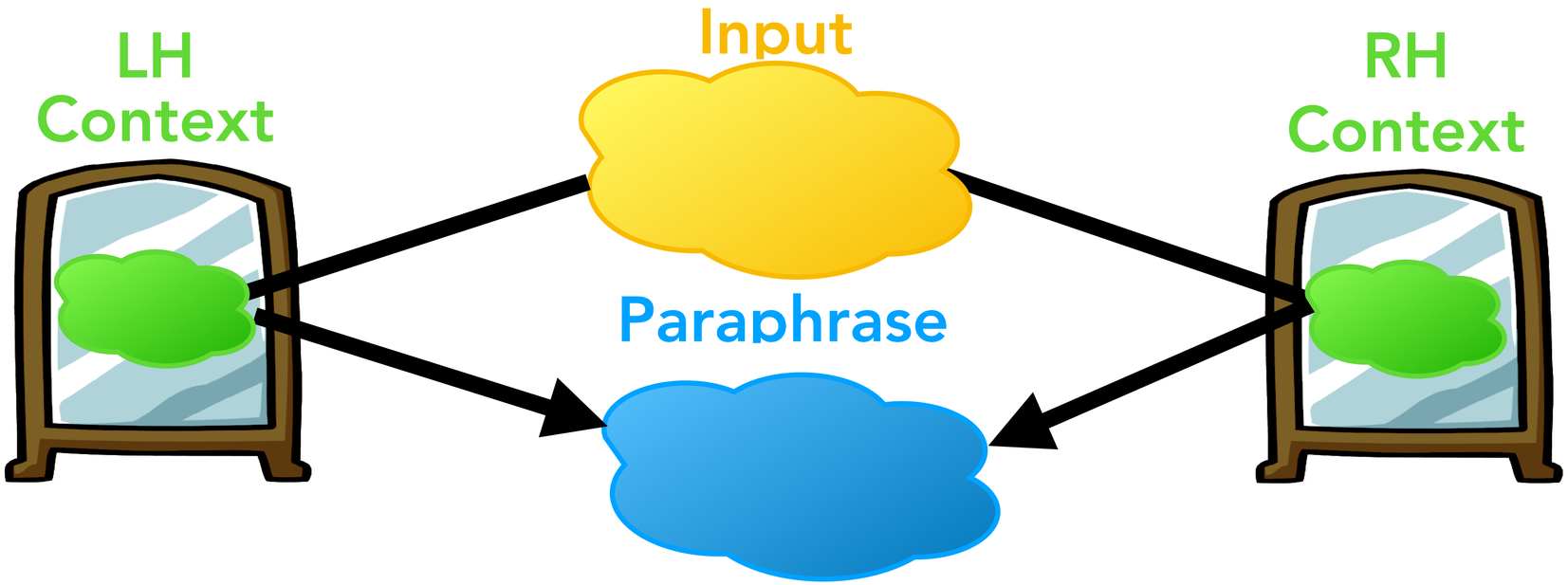


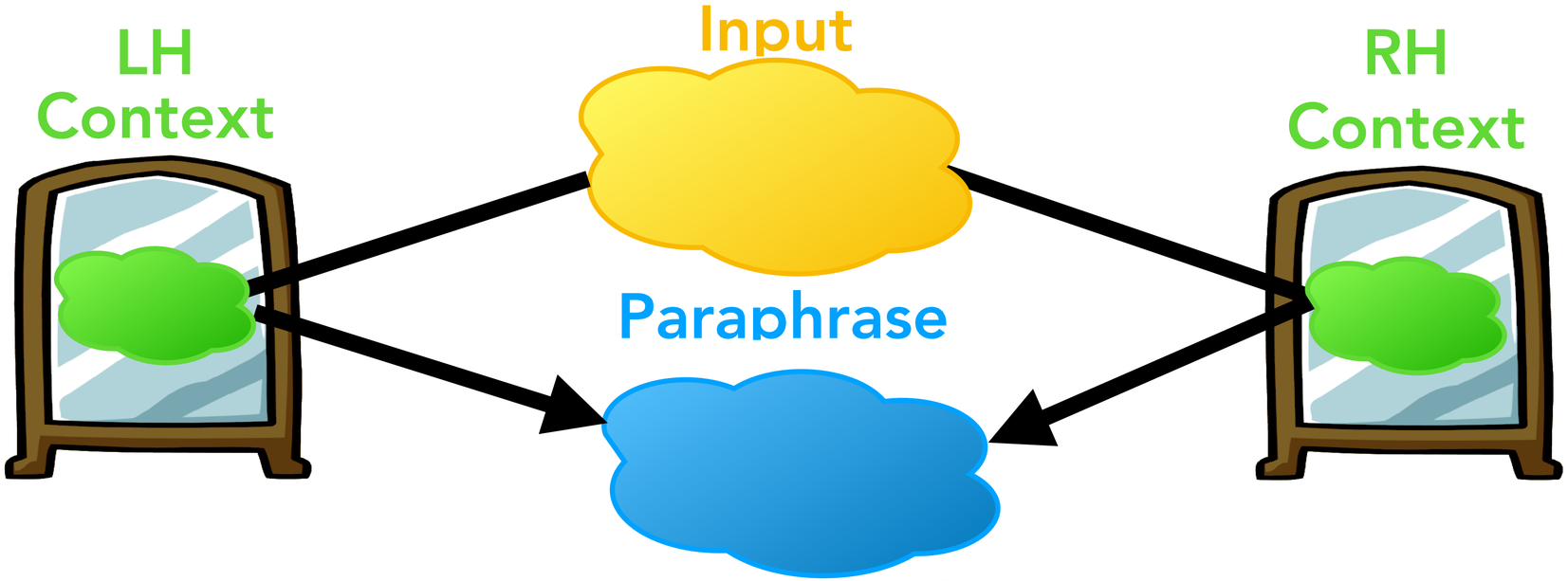
Paraphrasing with Reflective Decoding

$$D_{KL}(\overrightarrow{\text{LM}}(c|s_{src}), \overrightarrow{\text{LM}}(c|\hat{s}))$$

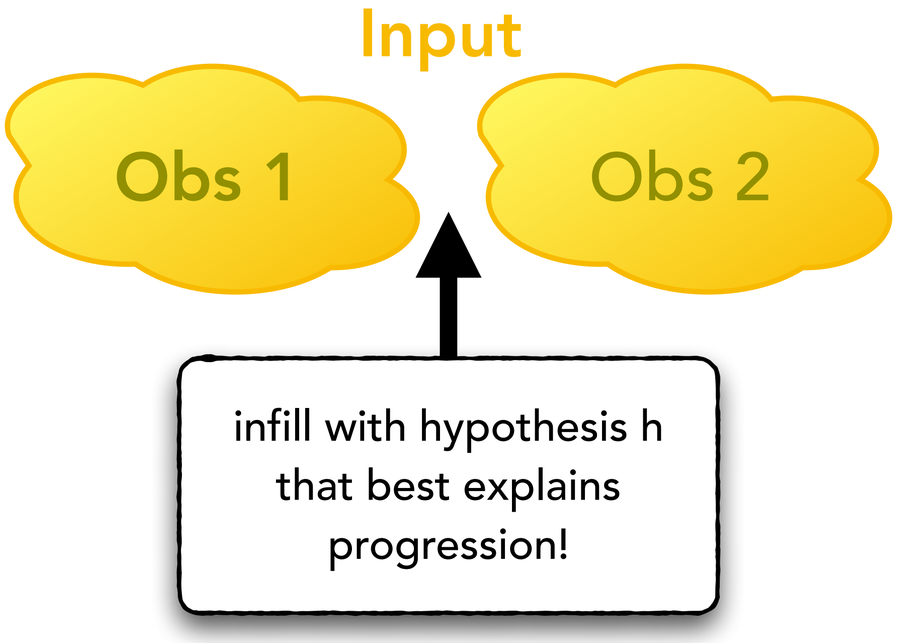


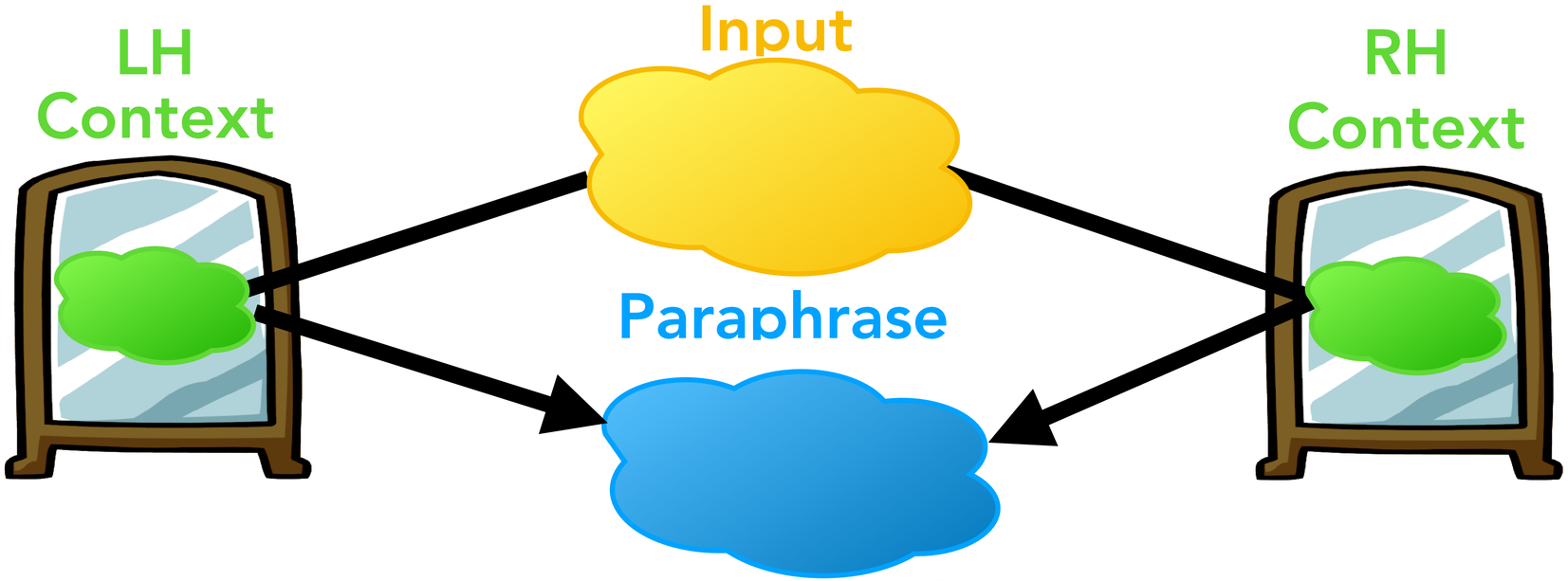
Use case 1: paraphrasing



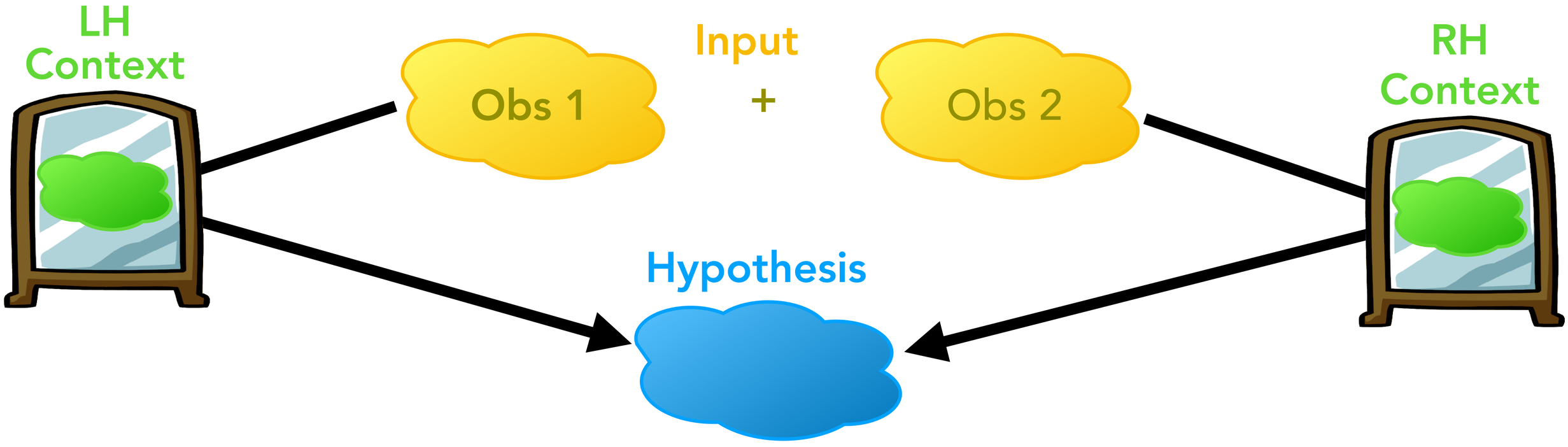


Use case 2: Abductive NLG

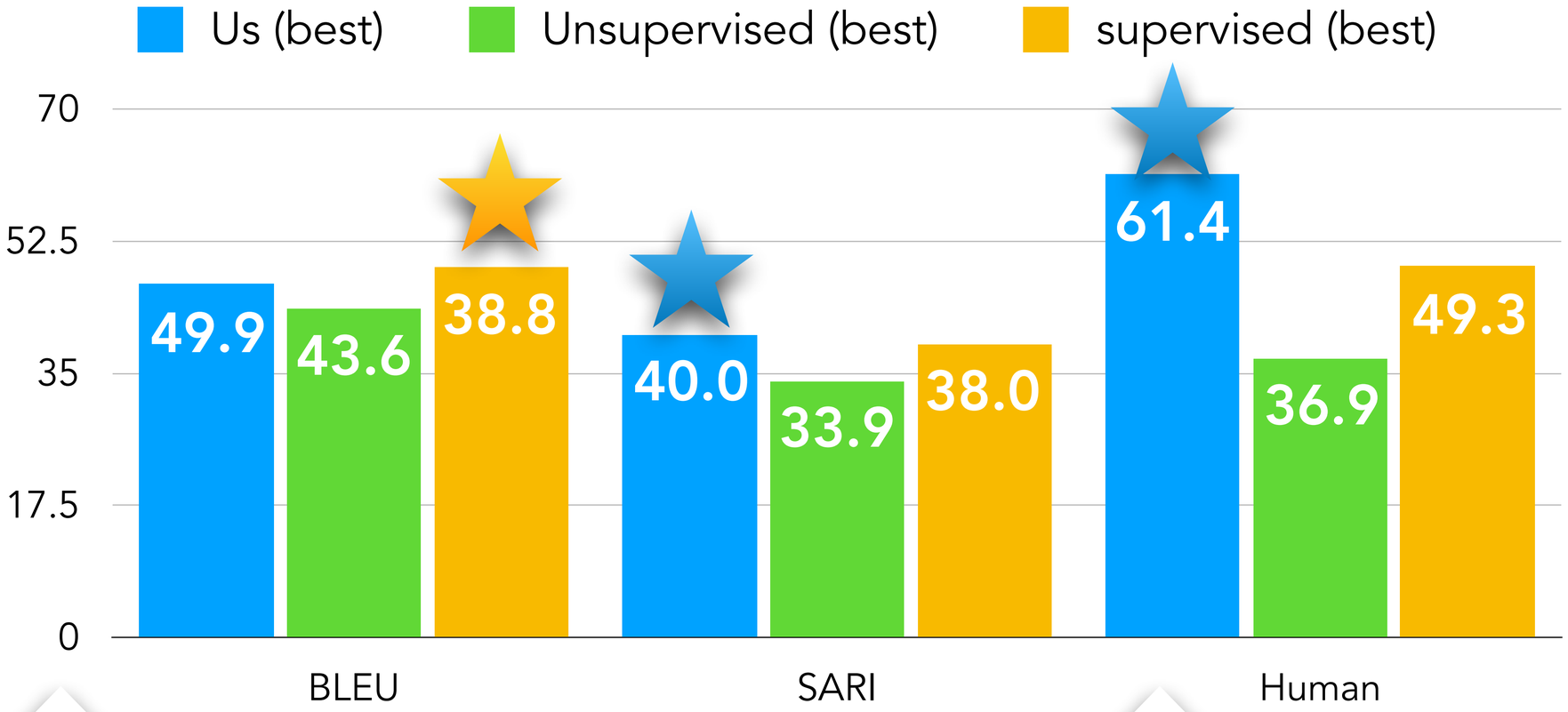




Use case 2: Abductive NLG



Evaluation on Paraphrasing



BLEU can be gamed by copy-and-paste the source phrase as is (without paraphrasing) SARI is a new major that accounts for novelty!

Unsupervised Reflective Decoding (not even fine-tuned on the target domain) outperforms **supervised (BART)** based on **SARI** & **Human Eval**

In this talk: Reasoning as Generation

- **Part 1:** unsupervised inference-time algorithms

Reasoning thru Neural Backpropagation	Reasoning thru Search with Logical Constraints	Reasoning thru Distributional Neural Imagination
DeLorean	NeuroLogic	Reflective Decoding

=> How to make 

from (off-the-shelf) neural language models

Current Paradigm of Deep Learning

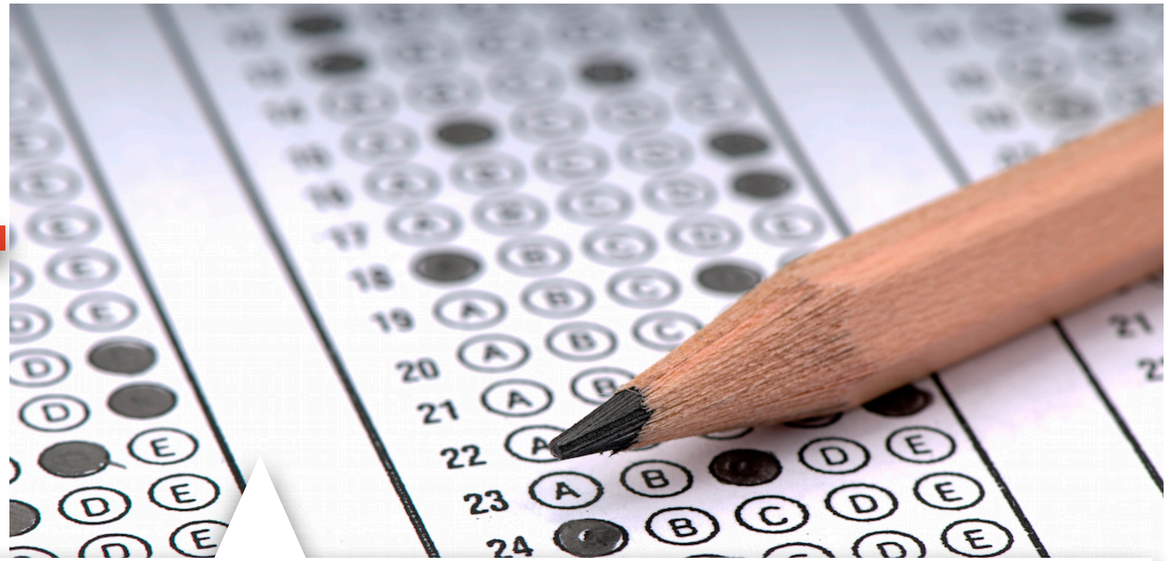
To achieve (super) human level performances on language (and vision) leaderboards

Self-supervision on a lot of raw text

Supervision on a lot of exam problems

rict reports indicate somewhat stronger
early January than at the time of the
ntered in the retail and industrial sec
ts , that the national economy gained
ed , manufacturing activity [MASK] to
ant and equipment .

nal economic activity on balance in Dec
eports in November , with much of the
It would appear , on the basis of th
um
a



This recipe fails on generative evaluation such as

abductive reasoning, counterfactual story revision, and

various constrained text generation tasks with logical constraints

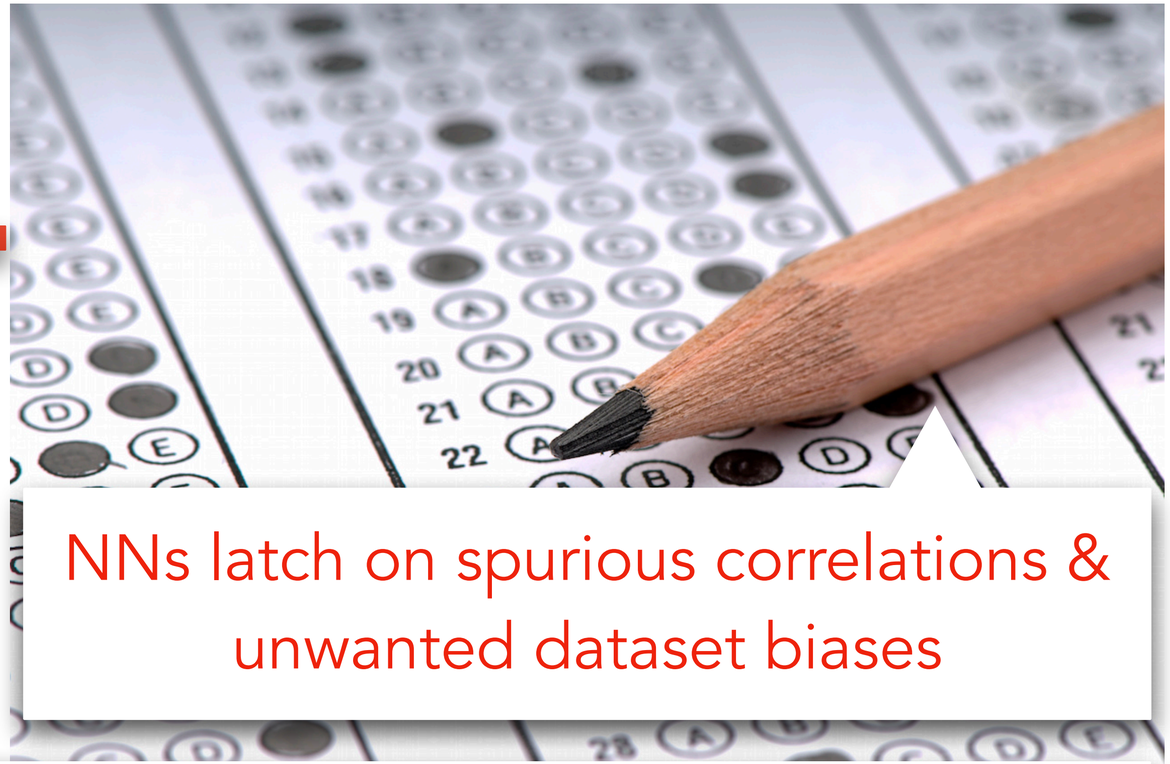
Imagine taking a deep learning class in which…

Self-supervision on a lot of DL code

```
        self.names = names
        self.name2index = dict(zip(names, range(len(names))))


    def __del__(self):
        # free memory created by C to avoid memory leak
        if hasattr(self, '__createfrom__') and self.__createfrom__ == 'C':
            if pointer(self) is not None:
                libbigfile.free_file(pointer(self))

    def read(self, requested, isname=True):
        if isname:
            index_name_array = [(self.name2index[x], x) for x in requested]
        else:
            assert(min(requested)>=0)
            assert(max(requested)<len(self.names))
            index_name_array = [(x, self.names[x]) for x in requested]
        index_name_array.sort()

        npoints = len(index_name_array)
        c_index = (c_ulonglong * npoints)()
        for i in range(npoints):
            c_index[i] = index_name_array[i][0]

        size = self.ndims * npoints
        pdata = (c_float * size)()
        res = libbigfile.seq_read_memory(self, npoints, c_index, pdata)
        assert(res)
```

+

Supervision on a lot of exam problems



NNs latch on spurious correlations & unwanted dataset biases

Can’t learn concepts well enough; need to learn from declarative knowledge

In this talk: Reasoning as Generation

- **Part 1:** unsupervised inference-time algorithms

Reasoning thru Neural Backpropagation	Reasoning thru Search with Logical Constraints	Reasoning thru Distributional Neural Imagination
DeLorean	NeuroLogic	Reflective Decoding

- **Part 2:** supervision with declarative knowledge for knowledge modeling

Neural & Symbolic Commonsense Knowledge	Visually Grounded Commonsense Knowledge	Social, Ethical, Moral Norms
COMET & ATOMIC 2020	Visual COMET	Social Chemistry 101

- **Part 3:** benchmarks and algorithmic bias reduction

(COMET-) ATOMIC$^{20}_{20}$:

On Symbolic and Neural Commonsense Knowledge Graphs

— wait, doesn't GPT-3 know everything? —

To appear at AAAI 2021

Jena Hwang

Chandra Bhagavatula

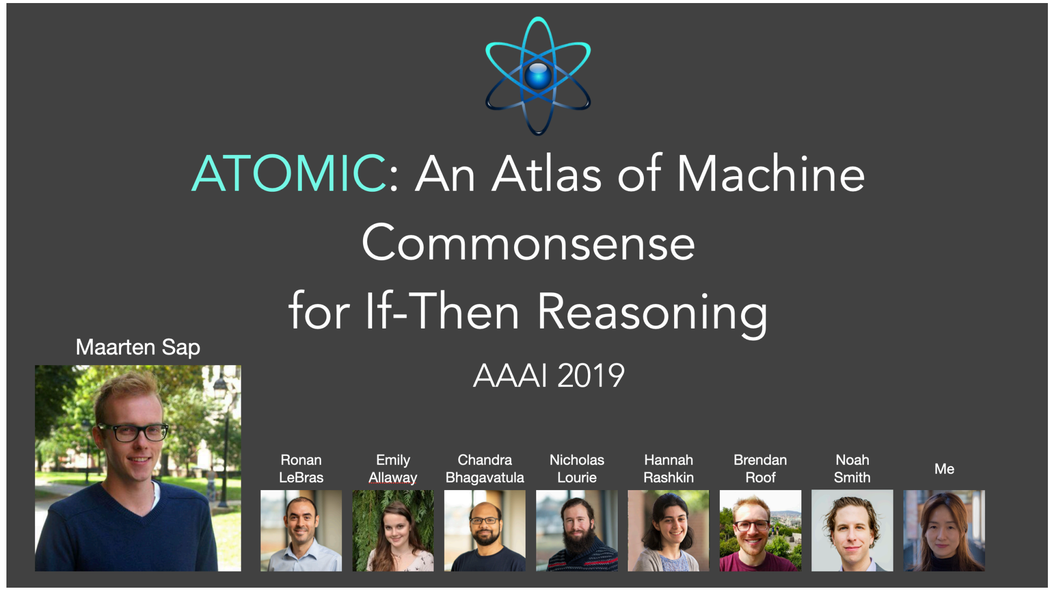
Ronan Le Bras

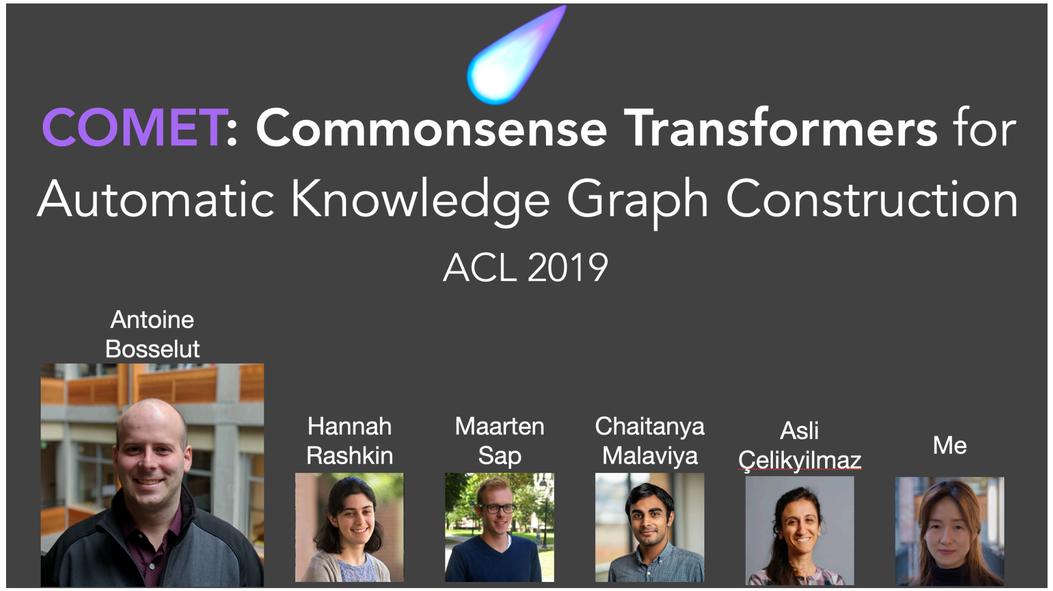
Jeff Da

Keisuke Sakaguchi

Antoine Bosseult

Me



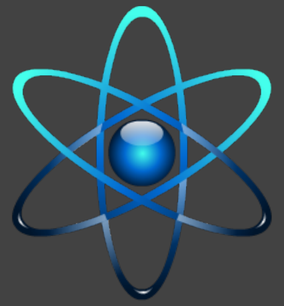


Therapy Chabot
Kearns et al. 2020 @ CHI EA 2020

Sarcasm generation
Chakrabarty et al. 2020 @ ACL 2020

Simile generation
Chakrabarty et al. 2020 @ EMNLP 2020

Automated Storytelling
Ammanabrolu et al. 2020 @ arXiv:2009.00829



ATOMIC: An Atlas of Machine Commonsense for If-Then Reasoning

AAAI 2019

Maarten Sap

Ronan LeBras

Emily Allaway

Chandra Bhagavatula

Nicholas Lourie

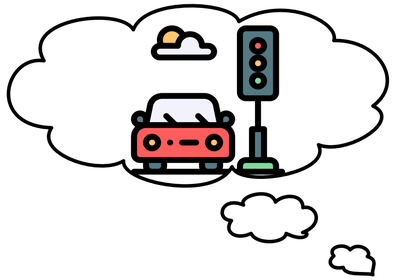
Hannah Rashkin

Brendan Roof

Noah Smith

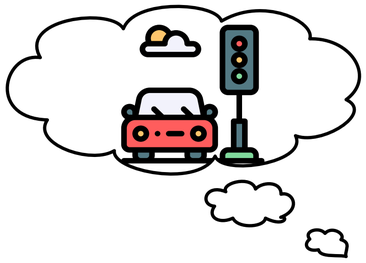
Me

Definition of Common Sense

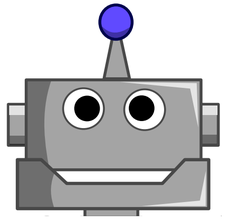


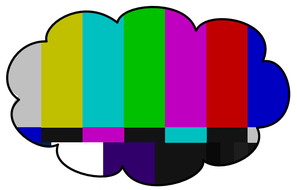
- the basic level of **practical knowledge** and **reasoning**
- concerning **everyday situations** and **events**
- that are **commonly** shared among **most** people.

For example, it's ok to keep the closet door open,
but it's not ok to keep the fridge door open,
as the food inside might go bad.



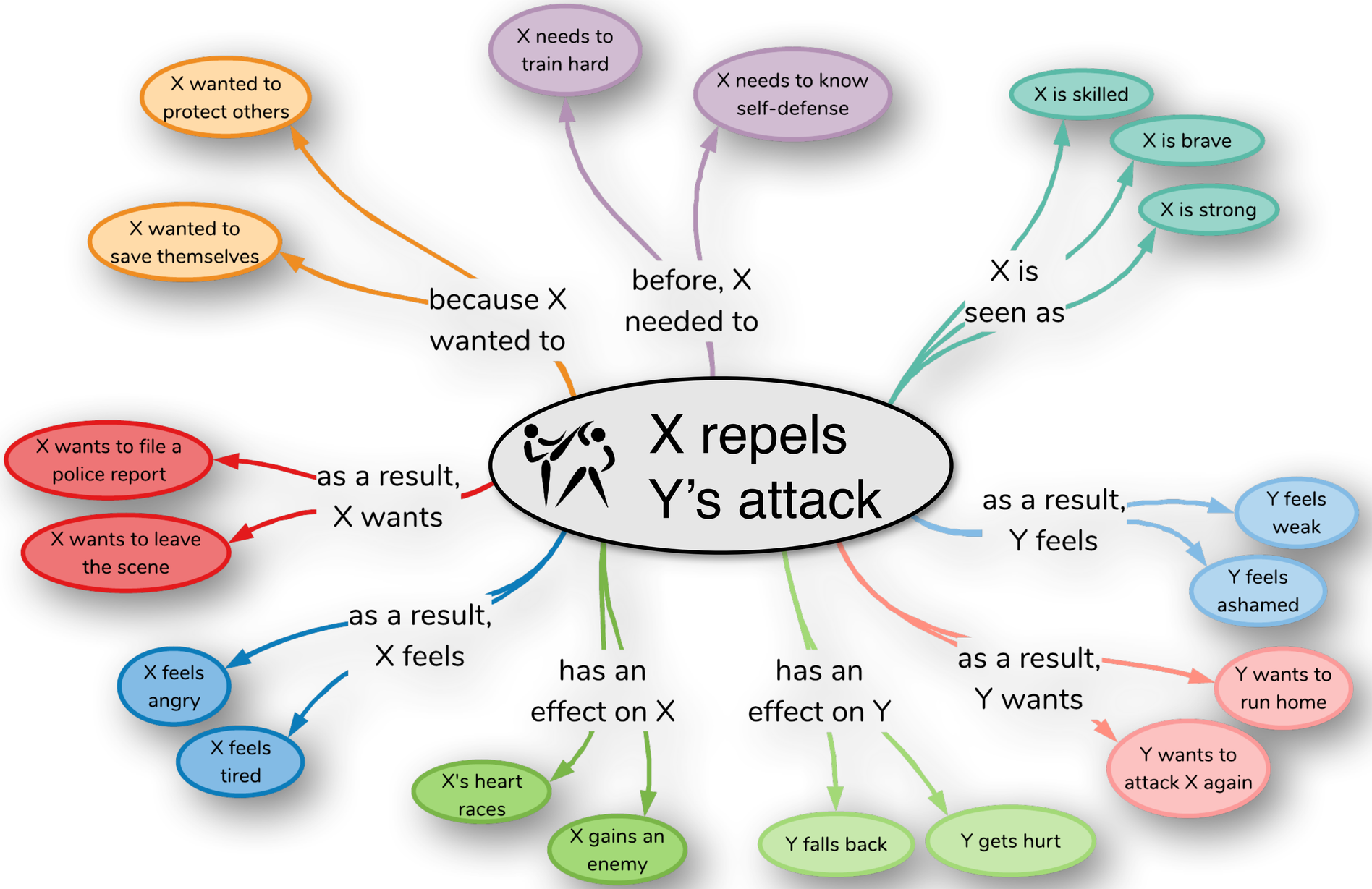
Essential for humans to live and interact with each other in a reasonable and safe way.

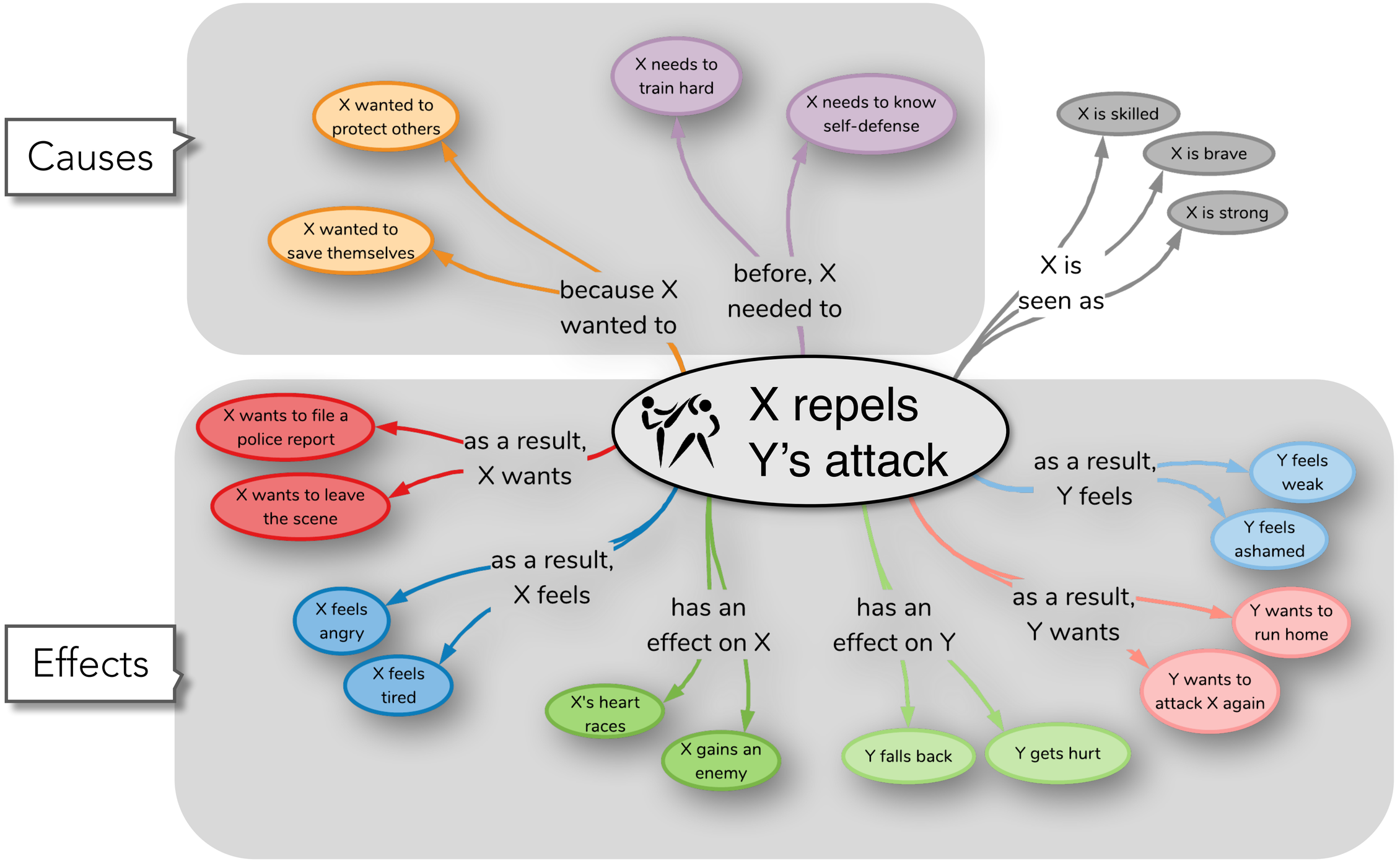


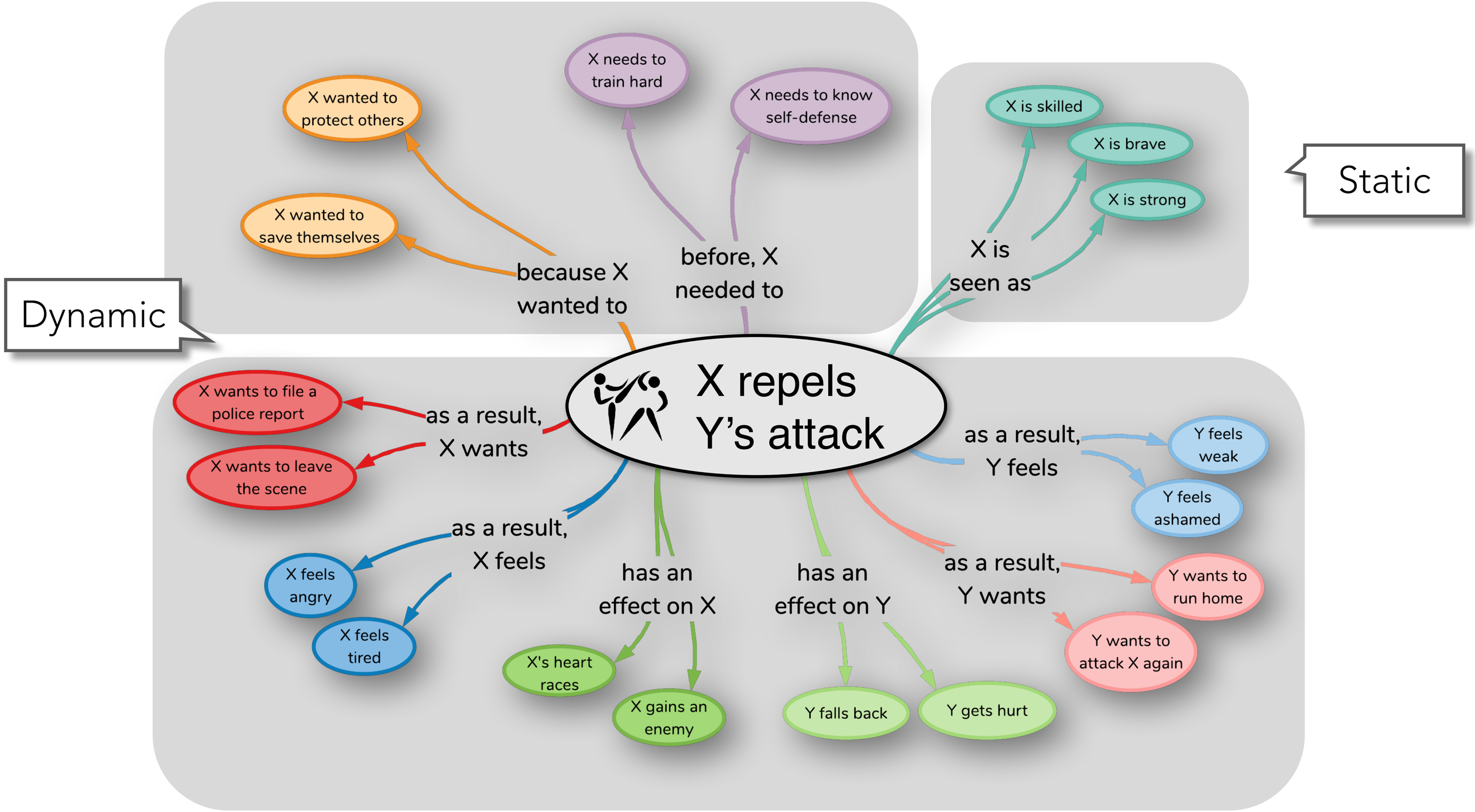


Essential for AI to understand human needs and actions better

For example, it’s ok to keep the closet door open,
but it’s not ok to keep the fridge door open,
as the food inside might go bad.


X repels Y's attack
because X wanted to
X wanted to protect others
X wanted to save themselves
before, X needed to
X needs to train hard
X needs to know self-defense
X is seen as
X is skilled
X is brave
X is strong
as a result, X wants
X wants to file a police report
X wants to leave the scene
as a result, X feels
X feels angry
X feels tired
has an effect on X
X's heart races
X gains an enemy
has an effect on Y
Y falls back
Y gets hurt
as a result, Y wants
Y wants to run home
Y wants to attack X again
as a result, Y feels
Y feels weak
Y feels ashamed

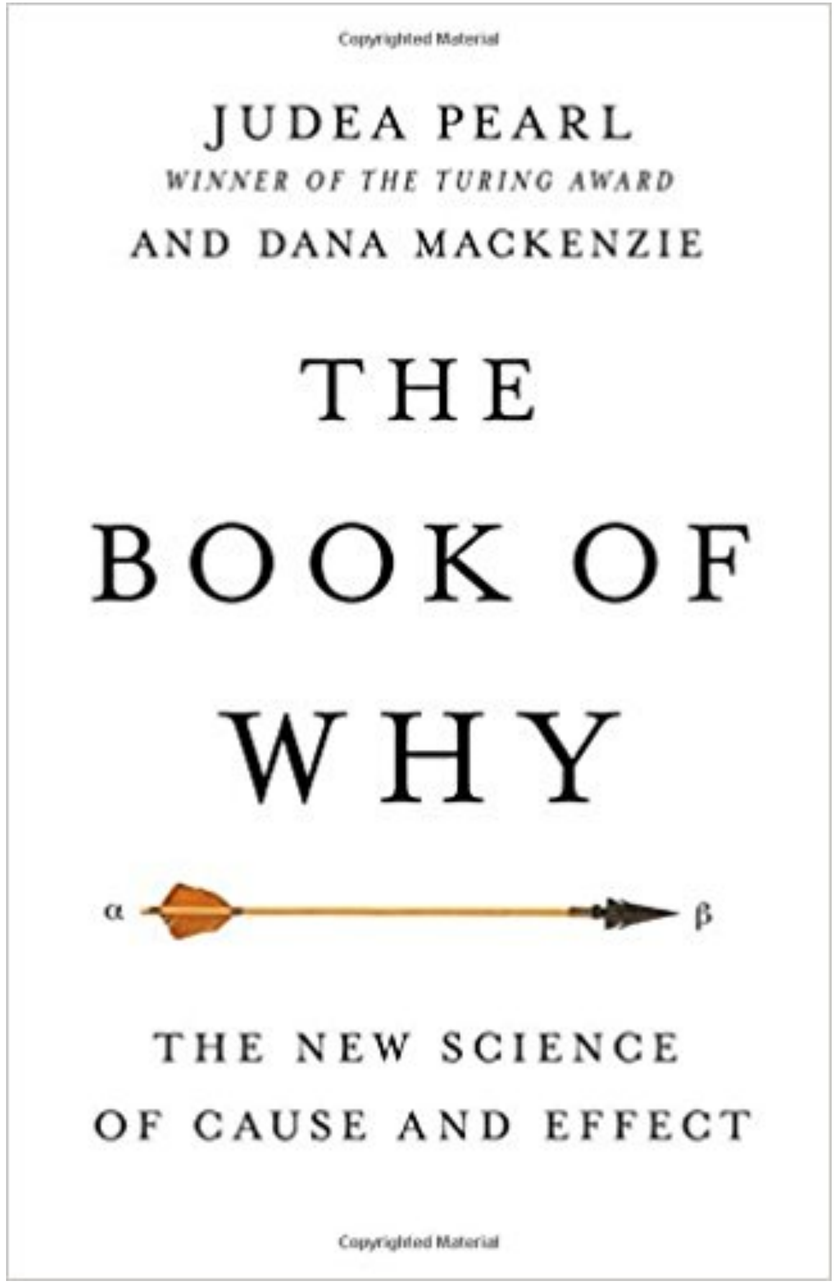

Causes
X wanted to protect others
X needs to train hard
X needs to know self-defense
X wanted to save themselves
because X wanted to
before, X needed to
X is skilled
X is brave
X is strong
X is seen as
X repels Y's attack
Effects
X wants to file a police report
X wants to leave the scene
as a result, X wants
as a result, X feels
X feels angry
X feels tired
has an effect on X
X's heart races
X gains an enemy
has an effect on Y
Y falls back
Y gets hurt
as a result, Y feels
Y feels weak
Y feels ashamed
as a result, Y wants
Y wants to run home
Y wants to attack X again

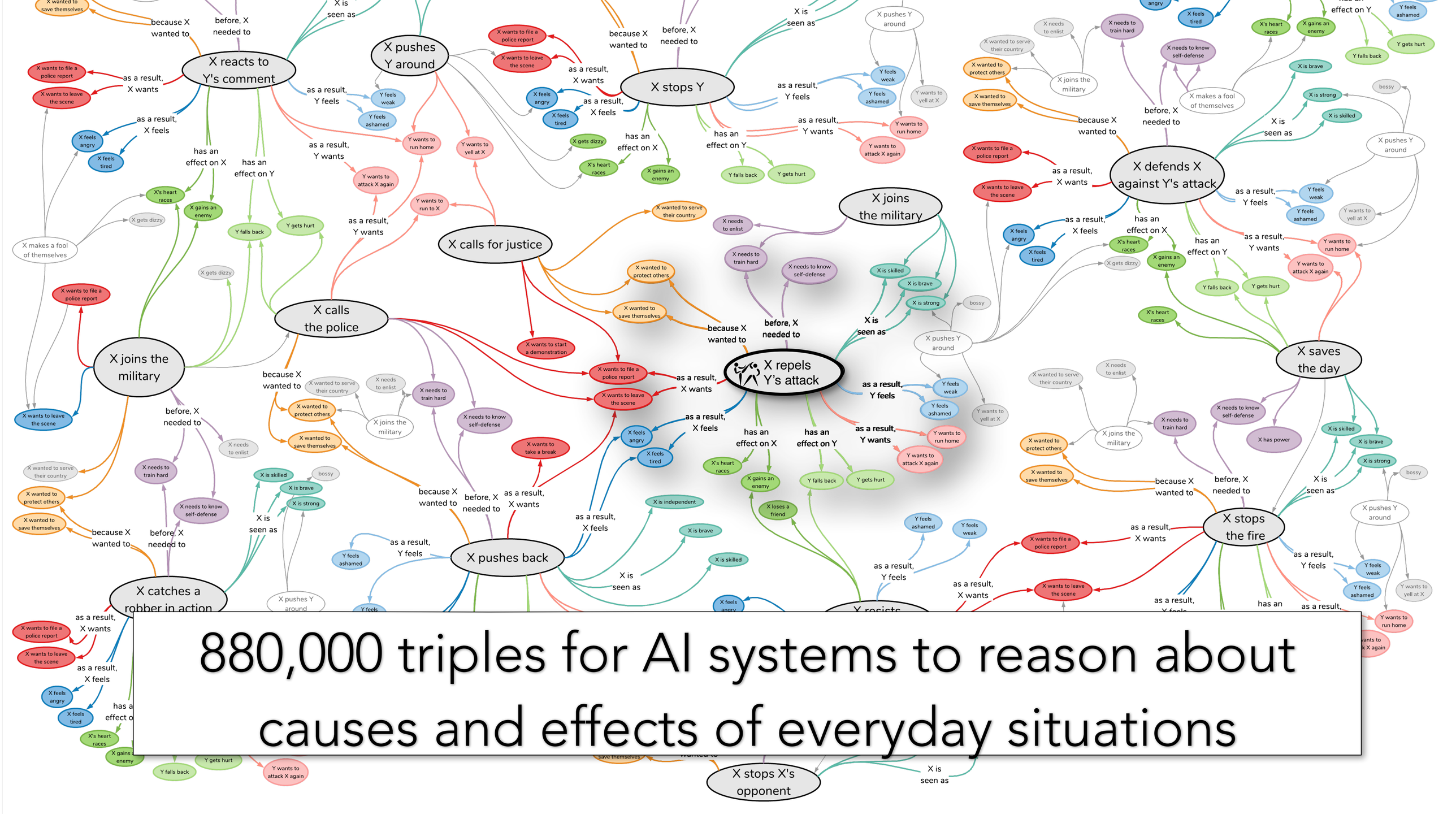

Dynamic
Static
X repels Y's attack
because X wanted to
before, X needed to
X is seen as
X wanted to protect others
X wanted to save themselves
X needs to train hard
X needs to know self-defense
X is skilled
X is brave
X is strong
as a result, X wants
X wants to file a police report
X wants to leave the scene
as a result, X feels
X feels angry
X feels tired
has an effect on X
X's heart races
X gains an enemy
has an effect on Y
Y falls back
Y gets hurt
as a result, Y feels
Y feels weak
Y feels ashamed
as a result, Y wants
Y wants to run home
Y wants to attack X again

"Cause and Effect"

"Deep learning, I see they're all stuck there on the level of associations. Curve fitting."

"To build truly intelligent machines, teach them cause and effect"





880,000 triples for AI systems to reason about causes and effects of everyday situations

How to acquire causes and effects at scale?

From unl eled text?

- Reporting bias (Gordon & Van Durme, 2013)

Murdering is 4x as common as exhaling?

- Commonsense knowledge is not often written (Grice, 1975)

amazon mechanical turk beta

ATOMIC: Crowdsourced commonsense knowledge around ***event prompts*** using ***natural language***

Browse ATOMIC:
https://tinyurl.com/atomic-commonsense

Existing commonsense knowledge bases

Knowledge of "what"
(taxonomic: **A isA B**; Davis and Marcus, 2015)

Represented in logical forms
(except ConceptNet)

```
event := (forall (e) (iff (event e) (or (exists (e1 e2) (and
(nequal e1 e2)  (change' e e1 e2))) (exists (e1) (subevent
e1 e)))))
```

99% taxonomic

98% taxonomic

OpenCyc
(Lenat, 1995)

ConceptNet
(Liu & Singh, 2004)

EventNet
(Espinosa & Lieberman, 2005)

Formal Theory of Commonsense Psychology
(Gordon & Hobbs, 2017)

Knowledge of "why" and "how"
(inferential: causes and effects)

Represented in natural language
(how humans talk and think)

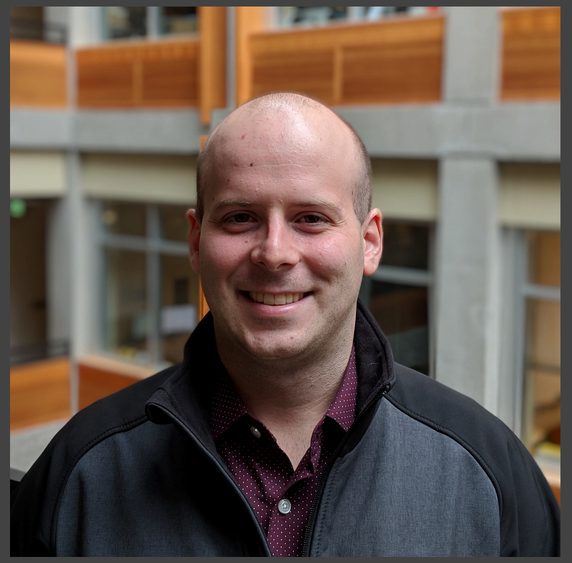
Visual Commonsense Graphs (Park et al., 2020)

ATOMIC
(Sap et al.; 2019)

COMeT: Commonsense Transformers for Automatic Knowledge Graph Construction

ACL 2019

Antoine Bosselut



Hannah Rashkin

Maarten Sap

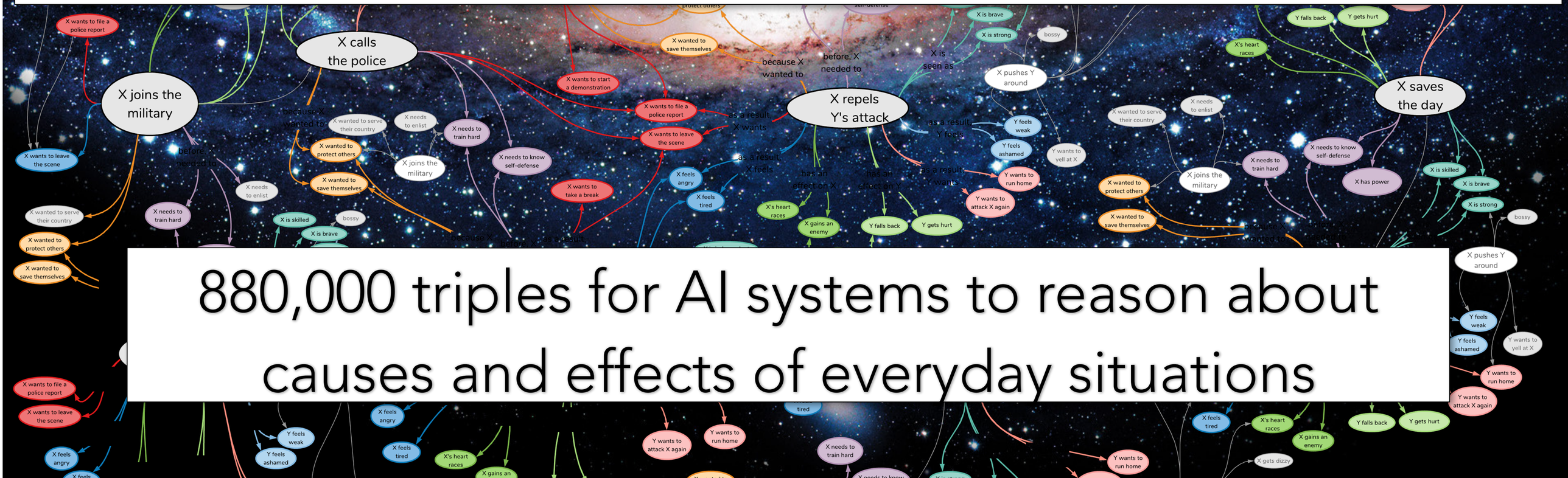
Chaitanya Malaviya

Asli Çelikyilmaz

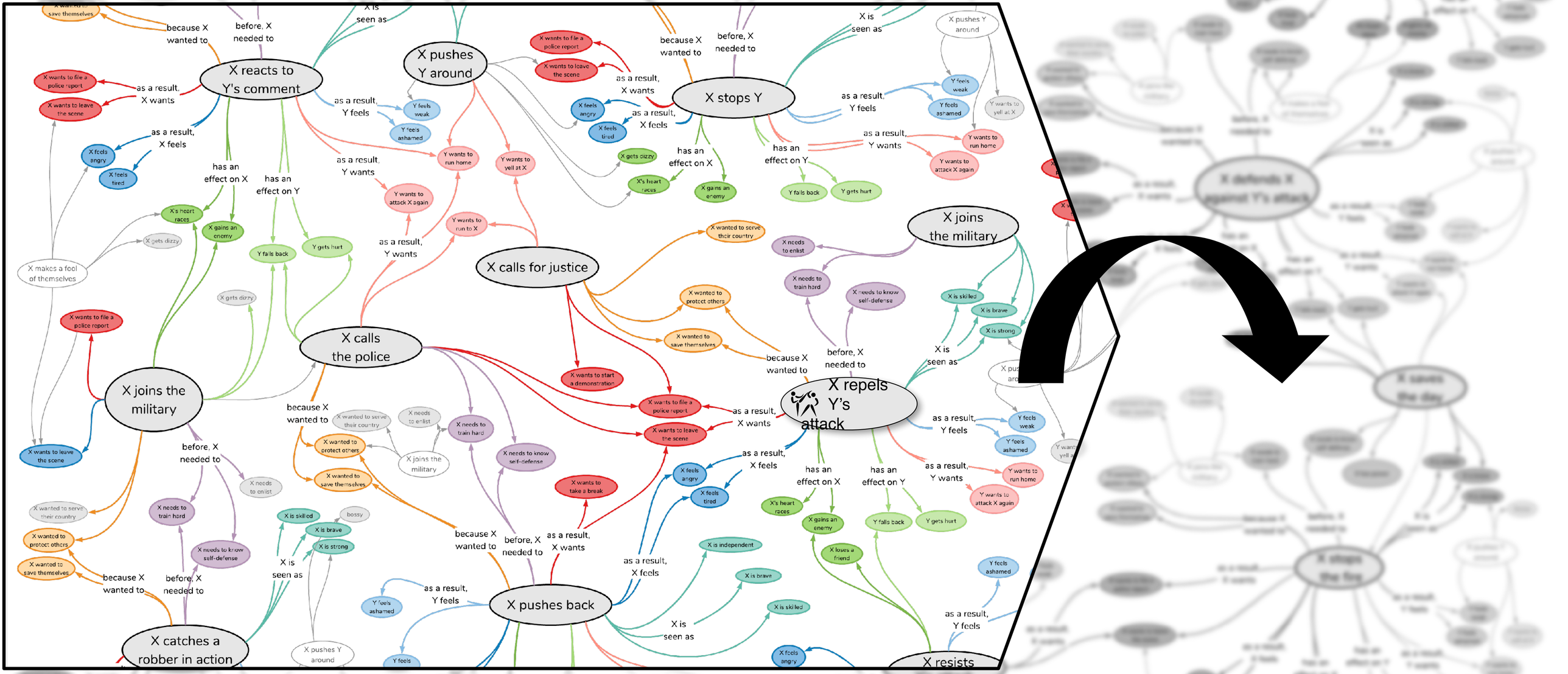
Me

Can we reason about commonsense knowledge without storing all of them explicitly?

Transfer learning from language (self-supervised) to knowledge (supervised)?



880,000 triples for AI systems to reason about causes and effects of everyday situations



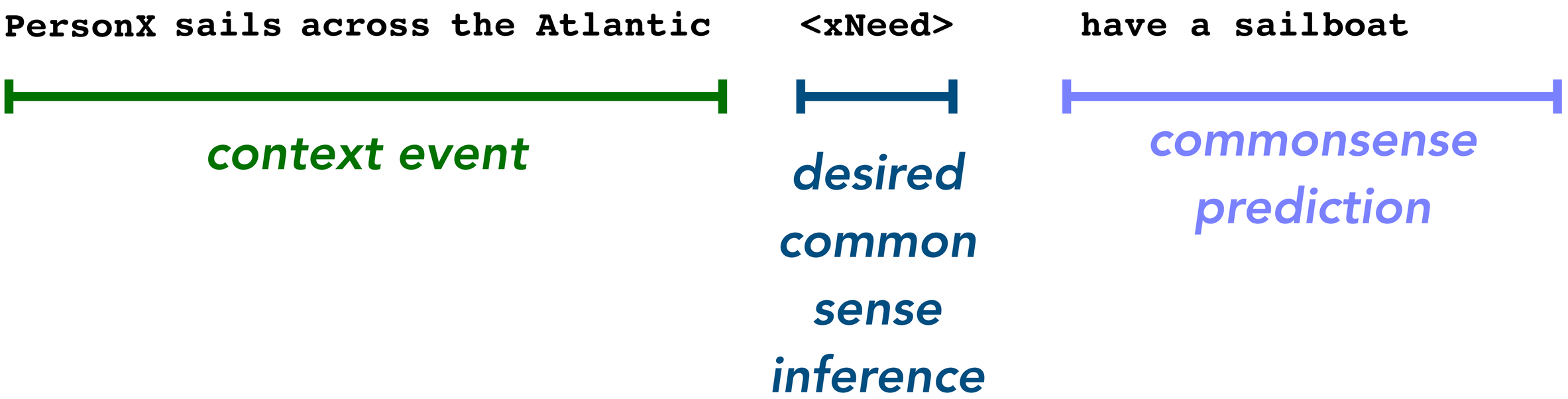
Goal: to teach models to reason about causes and effects of new ATOMIC events

COMeT

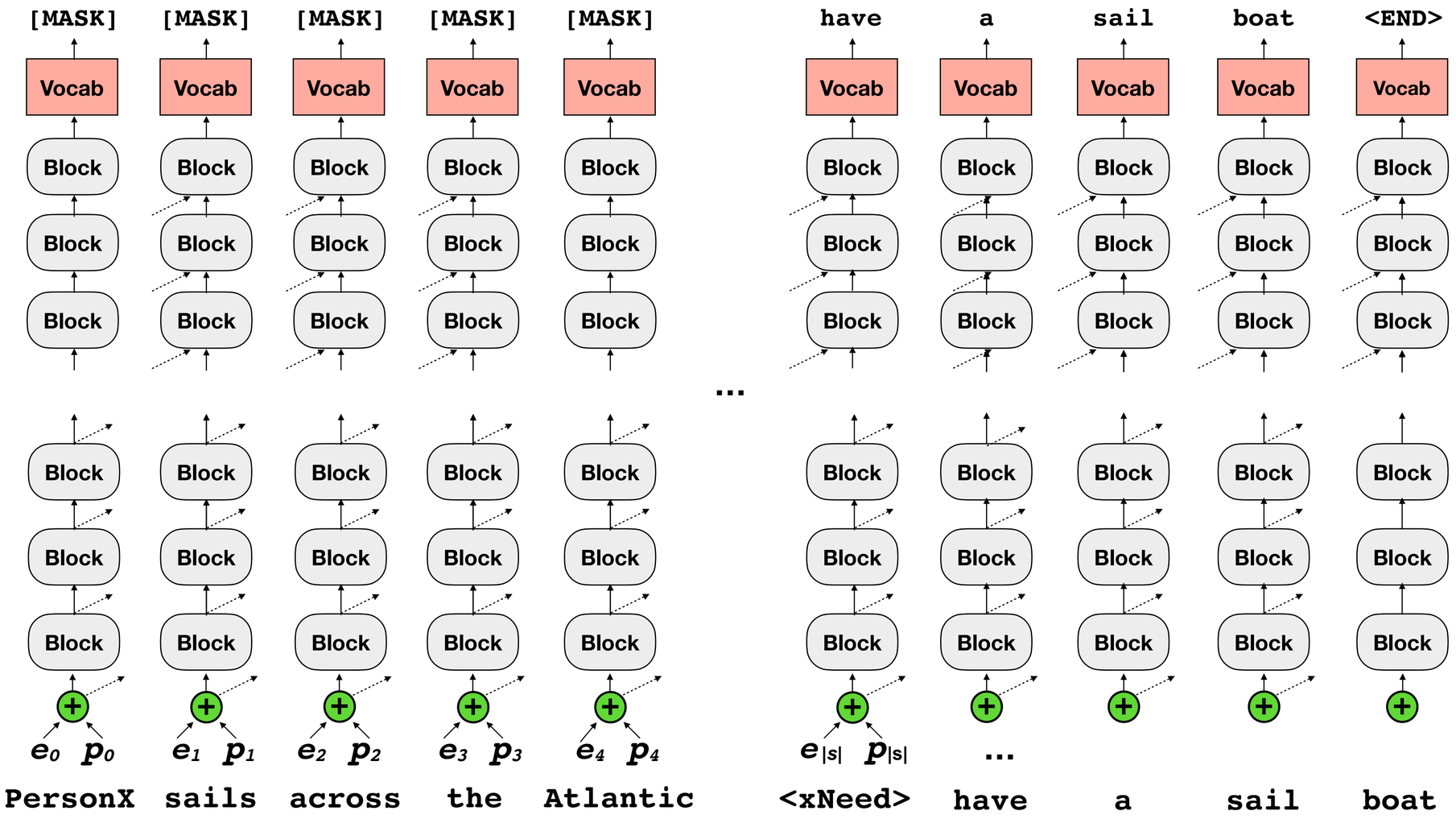
Transfer learning from language to knowledge!

Reporting bias issues, but provides signal about which events are similar

COMeT


PersonX sails across the Atlantic
<xNeed>
have a sailboat
context event
desired common sense inference
commonsense prediction

COMeT


[MASK] [MASK] [MASK] [MASK] [MASK]
have a sail boat <END>
Vocab
Block
...
e0 p0 e1 p1 e2 p2 e3 p3 e4 p4
e|s| p|s| ...
PersonX sails across the Atlantic
<xNeed> have a sail boat

Turns out, COMET generalizes well on out-of-domain examples

(which we realized only after publishing our ACL 2019 paper…)

Demo: https://mosaickg.apps.allenai.org/

AI2 Allen Institute for AI

Mosaic Knowledge Graphs

- COMeT
  - Events (ATOMIC)
  - Concepts (ConceptNet)
  - Images (VisualCOMET)
- Knowledge Graph
  - ATOMIC
- About

Click here for COMET demo (COMET trained on ATOMIC)

Commonsense Inferences about Events (COMmonsensE Transformers on ATOMIC)

A knowledge base construction e
graphs. By training on a seed set

Write a short text to reason about

Explore

sarah repel's jack's attack in a chess game

Predict

Try: PersonX acts quickly, PersonX is a big deal, My boss is very good

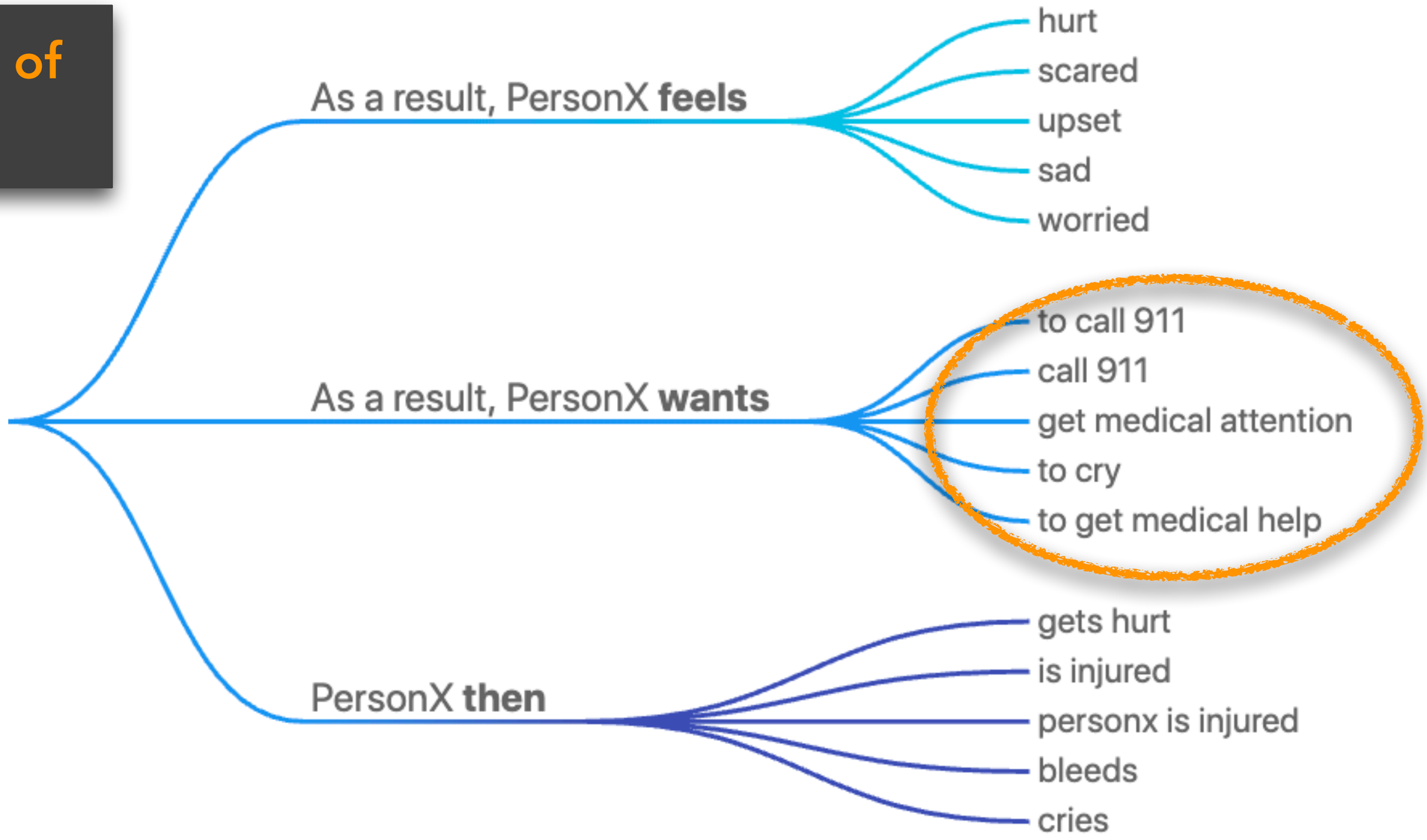
COMeT Predictions Graph

The model has predicted these relationships for 'sarah repel's jack's attack in a chess game'
View 'sarah repel's jack's attack in a chess game' in the ATOMIC dataset

Demo: https://mosaickg.apps.allenai.org/

John gets into an accident

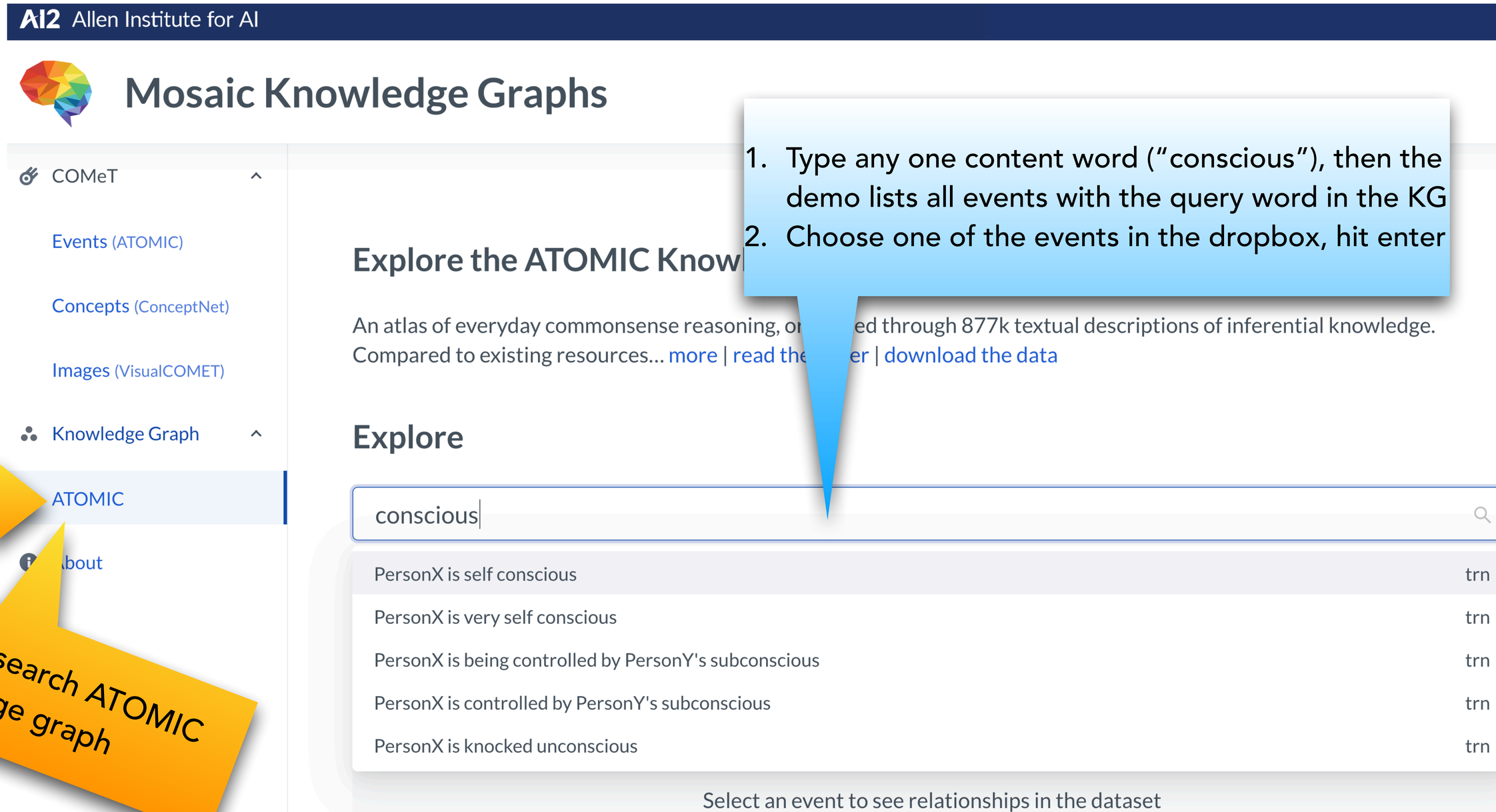
"John" instead of "Person X"



(Some amount of)

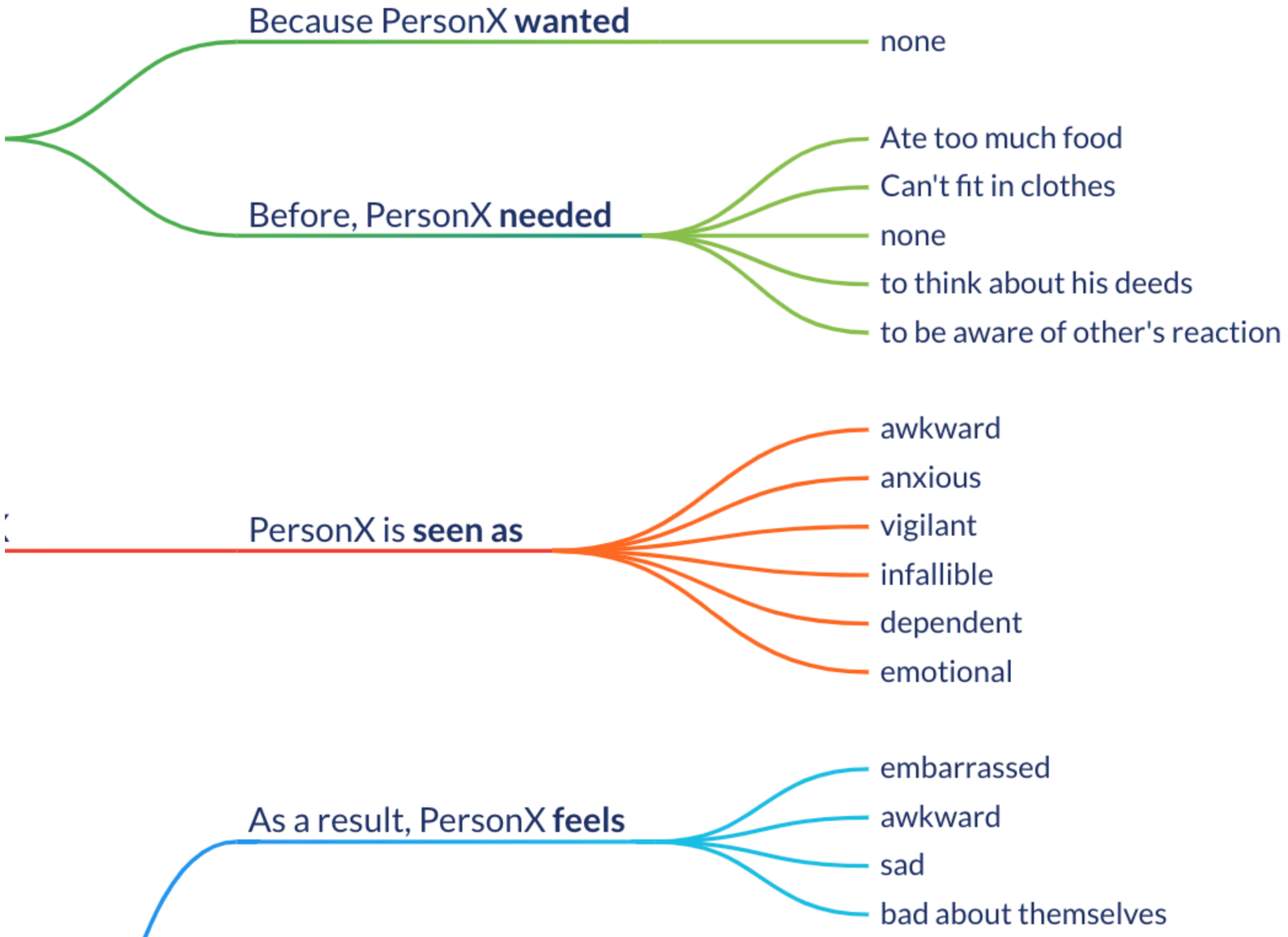
Abstraction by COMET?

Demo: https://mosaickg.apps.allenai.org/

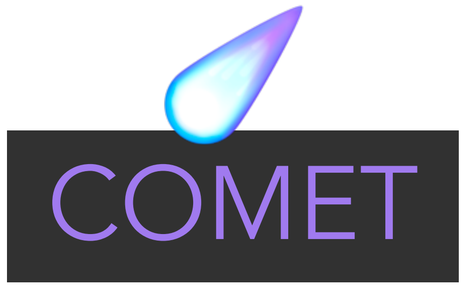


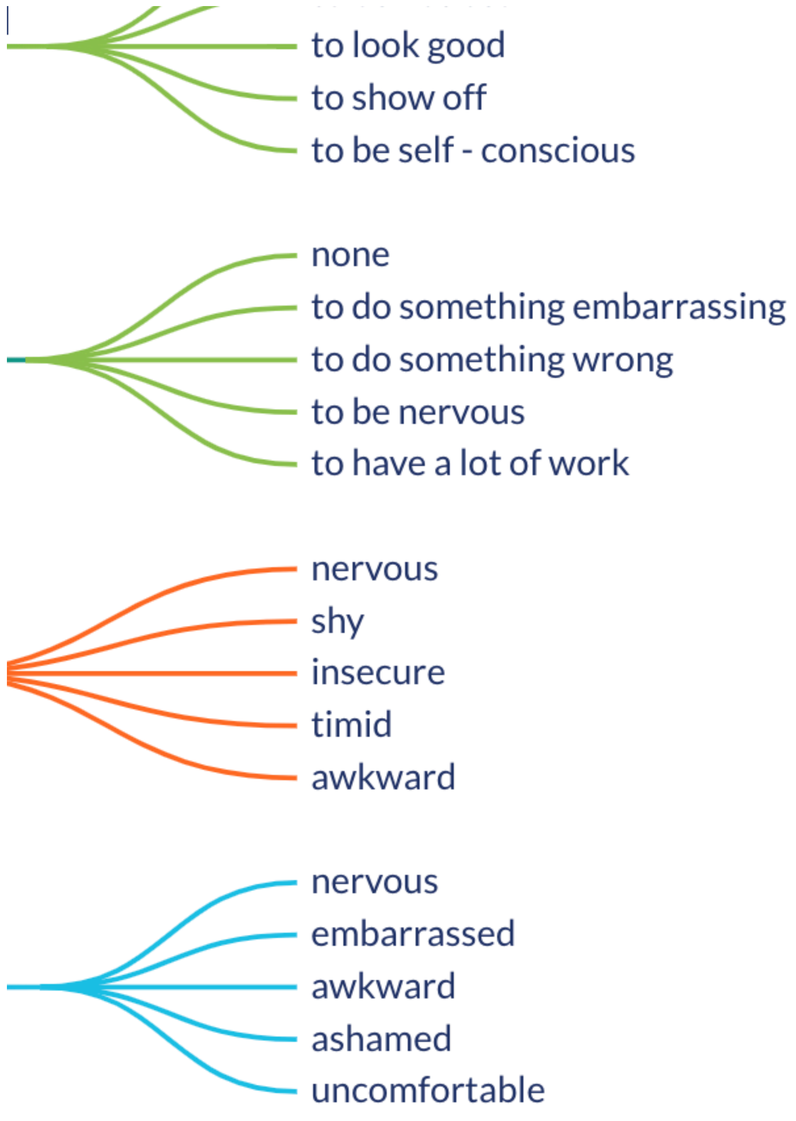
ATOMIC

"PersonX is self conscious"



"PersonX is self conscious"





"PersonX is self conscious"

ATOMIC

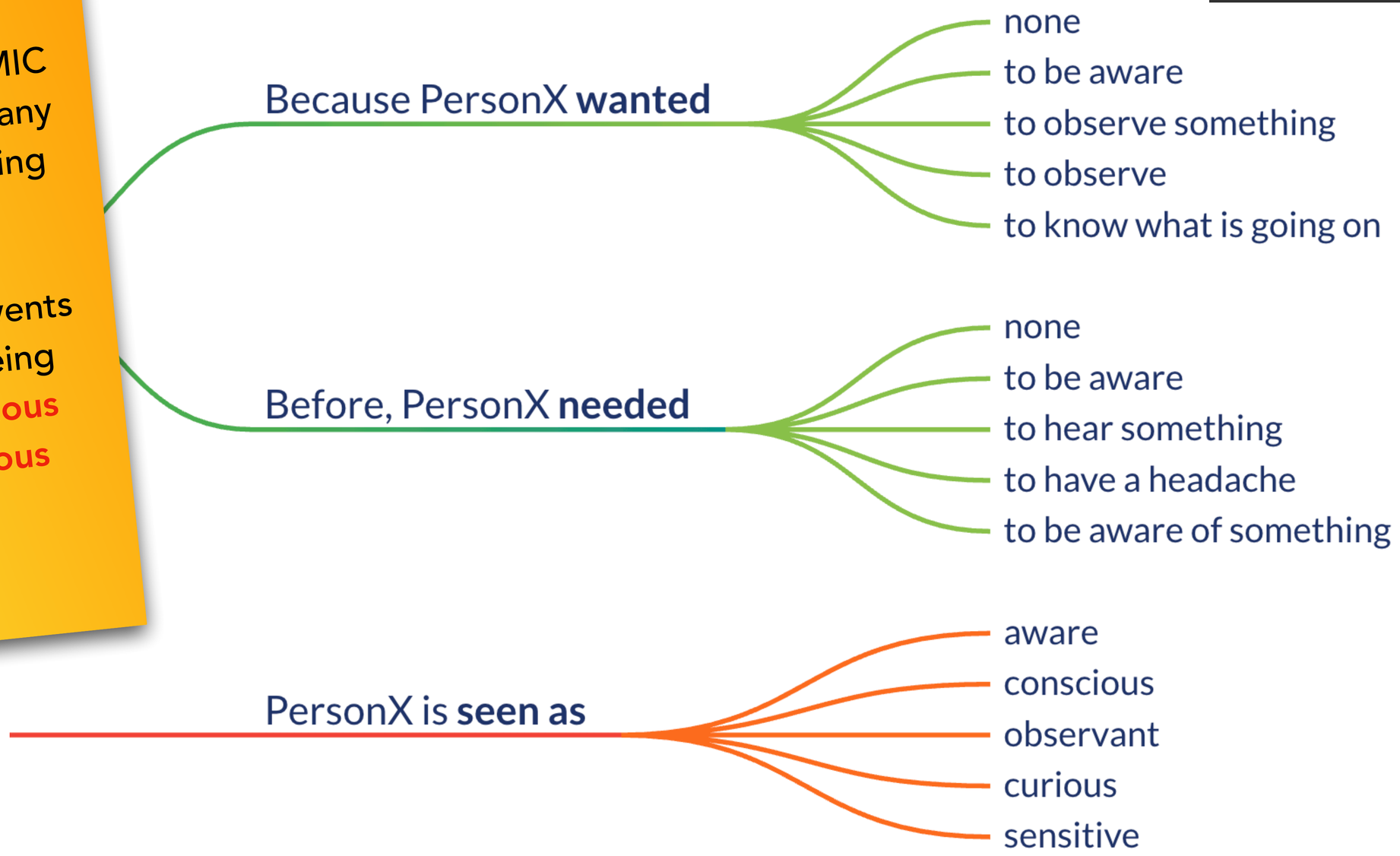
COMET

"Sarah is conscious"

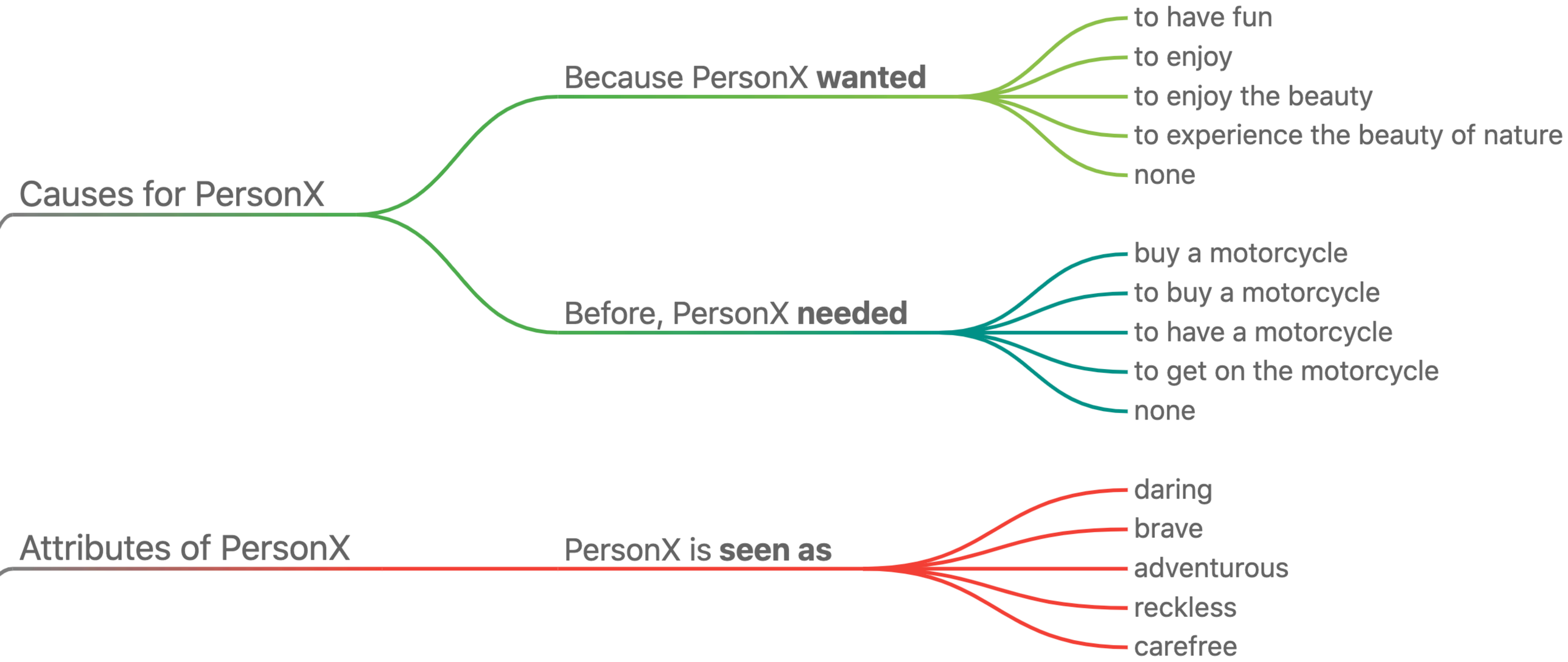
COMET

Recall that ATOMIC does NOT have any event about being **conscious**

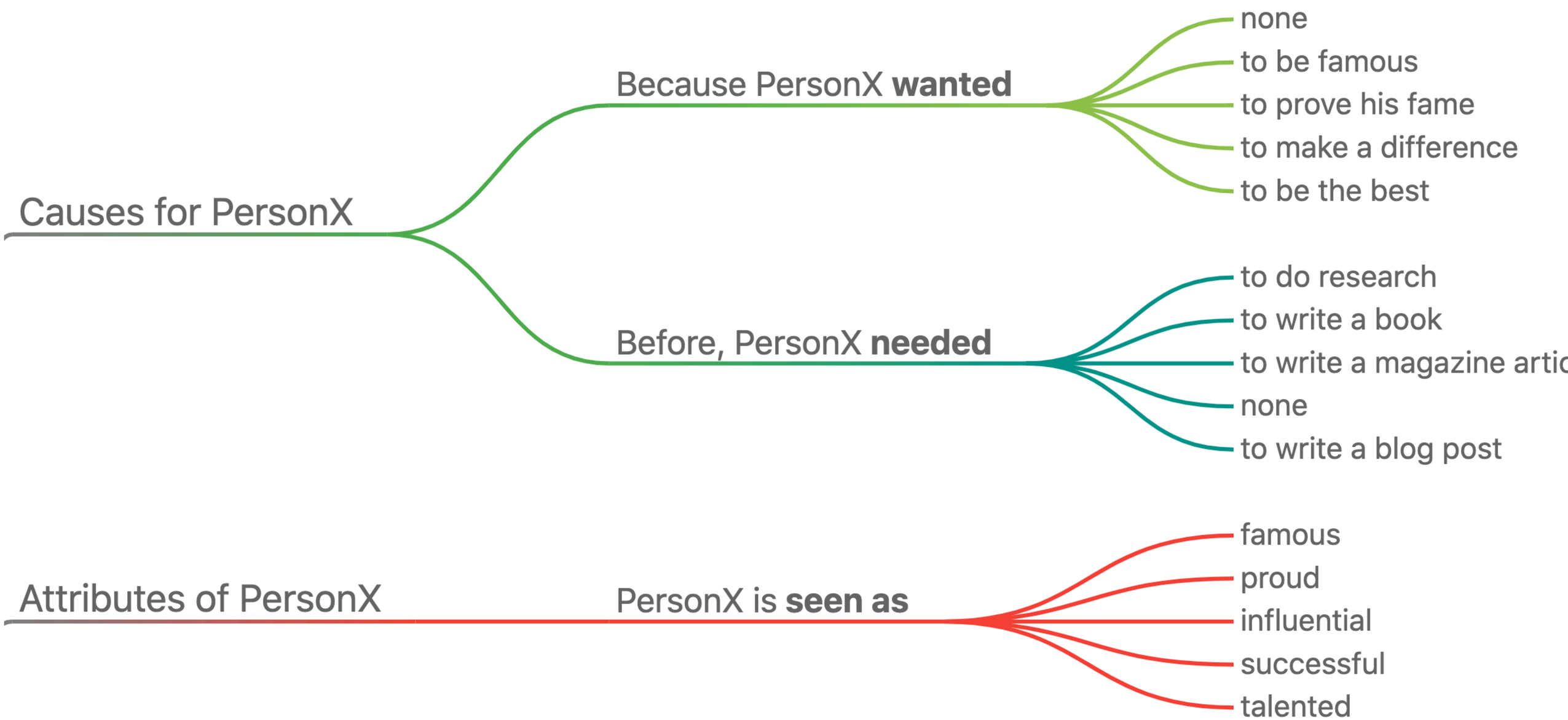
ATOMIC has events only about being
- **Self-conscious**
- **Unconscious**



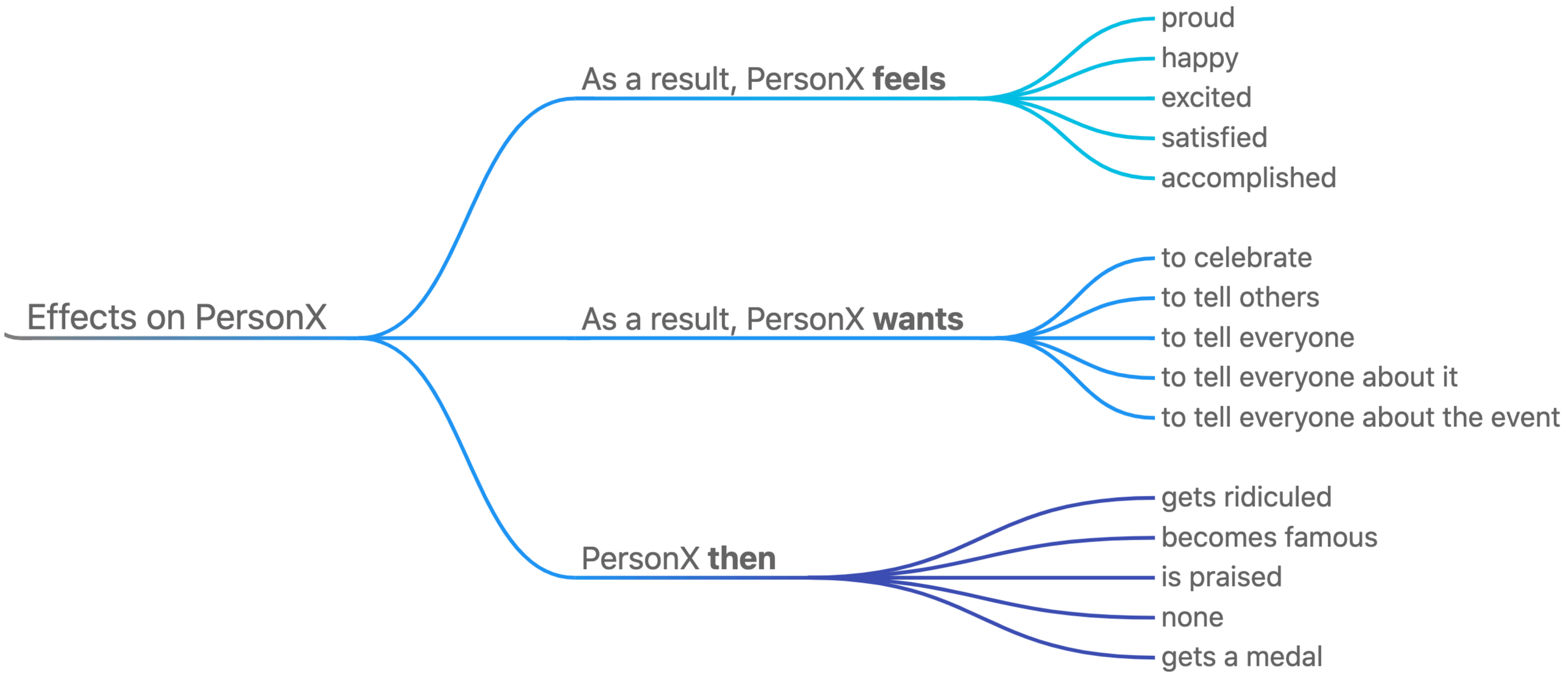
Sanja rides into the sunset on a motorcycle after solving AI.



Gary breaks the world record for most controversial tweet.



Gary breaks the world record for most controversial tweet.



Gary Marcus @GaryMarcus · 1h

ps i realize there is a typo in my query. fixing the typo doesn't much help though:

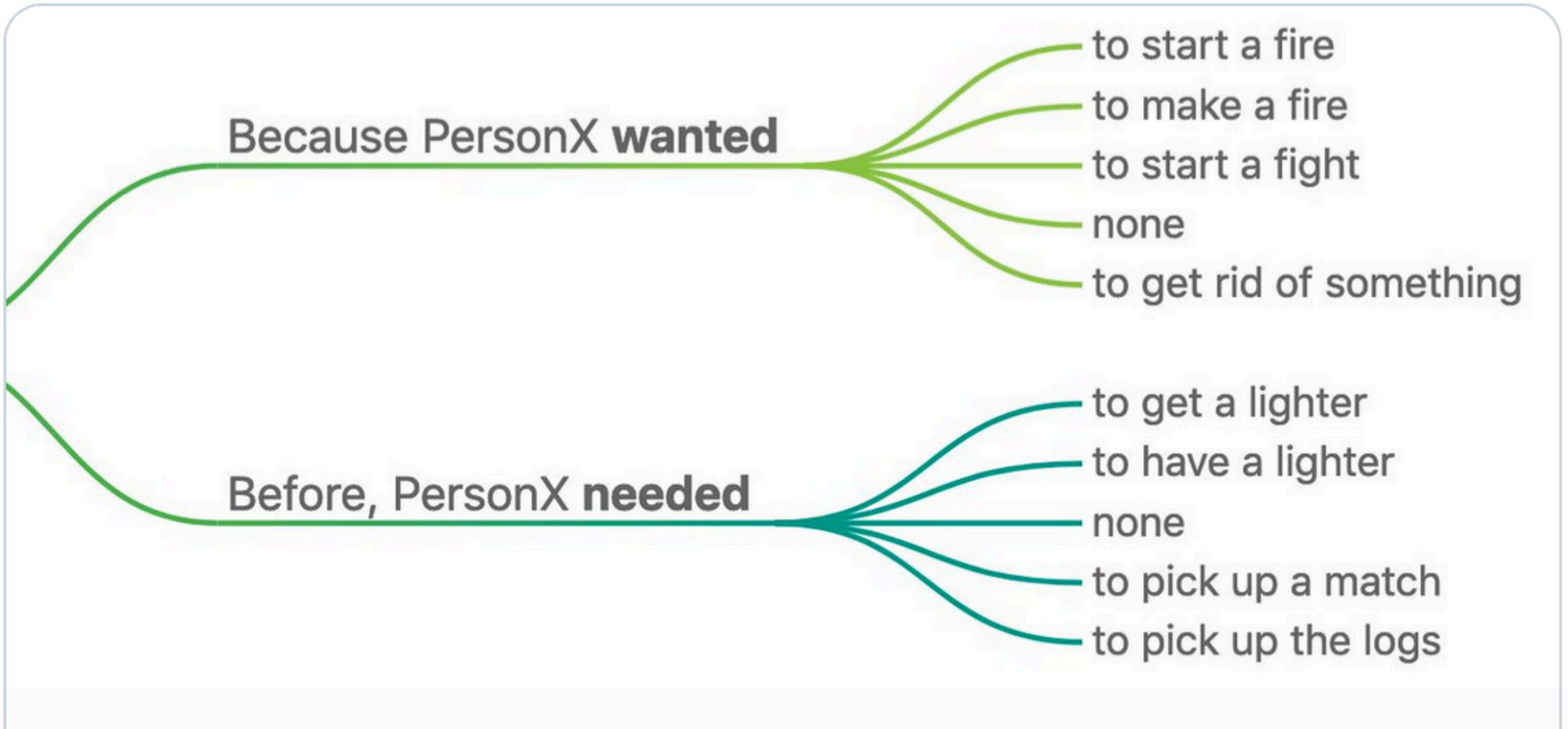
Completion

what happens when you stack kindling and logs in a fireplace and then drop some matches is that you typically start a ick. So, it's kind of ironic that the second day after my son was born, the fire in the living room had melted through the kindling. It's pretty neat."

Yejin Choi
@YejinChoinka

Replying to @GaryMarcus

Gary, try mosaickg.apps.allenai.org by typing "Gary stacks kindling and logs and drops some matches". Sorry I used deep learning... :)



focus on
"causes and effects"
(causal knowledge)

COMeT

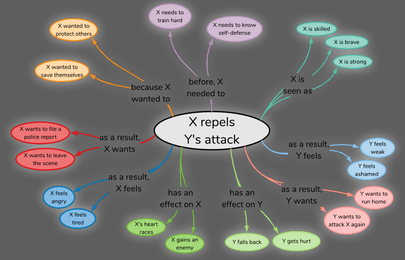
NEURAL
(generalizes well to compositional & unseen events)

(semi-) supervised learning of declarative knowledge

self supervised learning of observed knowledge

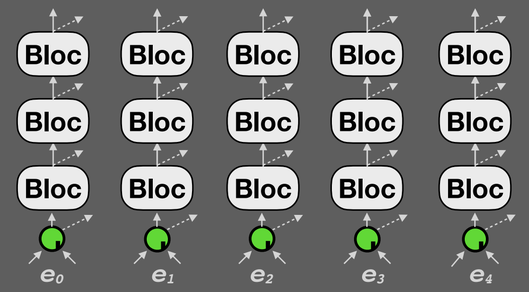
ATOMIC

tinyurl.com/atomic-



SYMBOLIC
but in LANGUAGE
(instead of LOGIC)

Language Models



(COMET-) ATOMIC$^{20}_{20}$:

On Symbolic and Neural Commonsense Knowledge Graphs

— wait, doesn't GPT-3 know everything? —

To appear at AAAI 2021

Jena Hwang

Chandra Bhagavatula

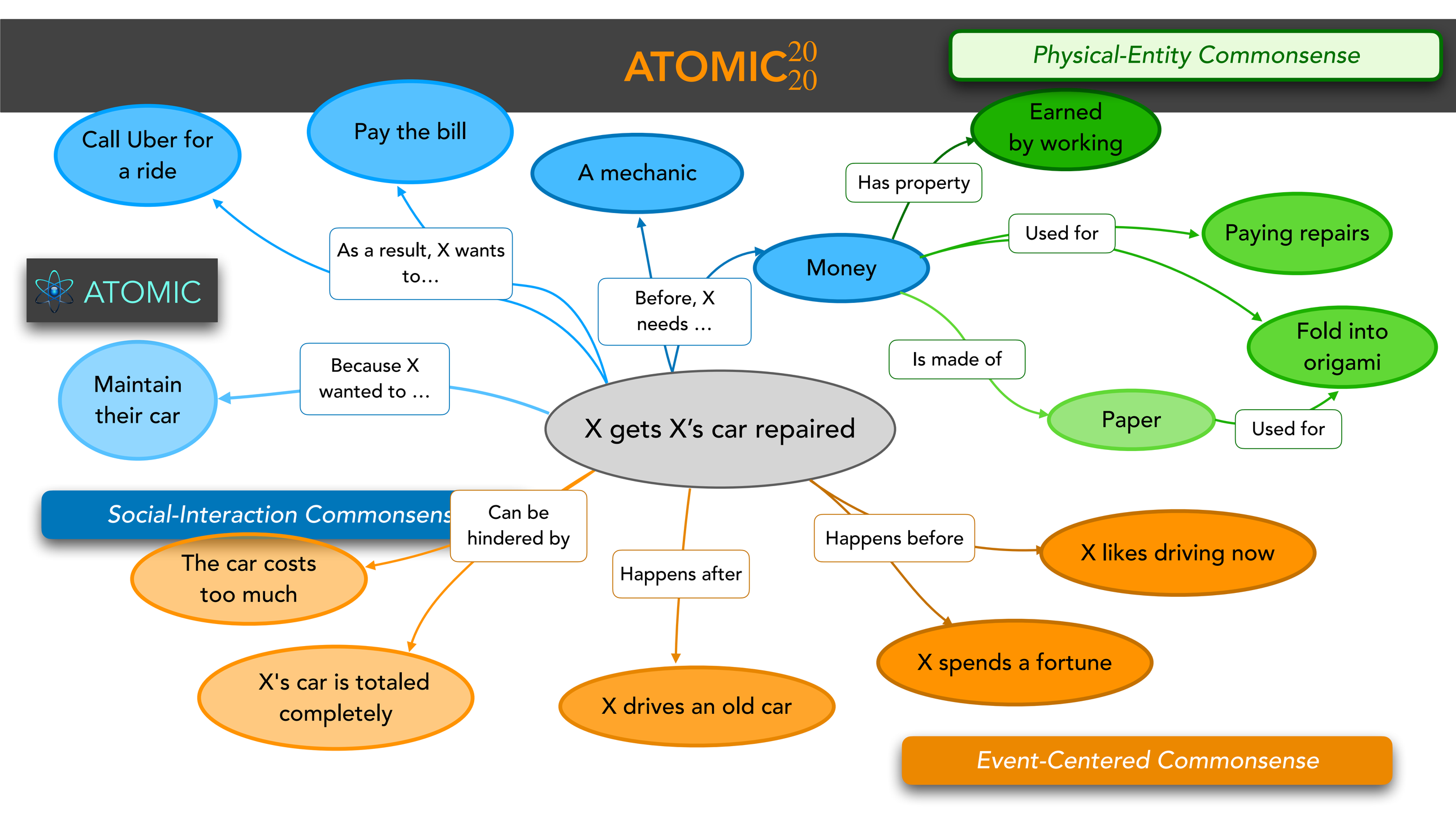
Ronan Le Bras

Jeff Da

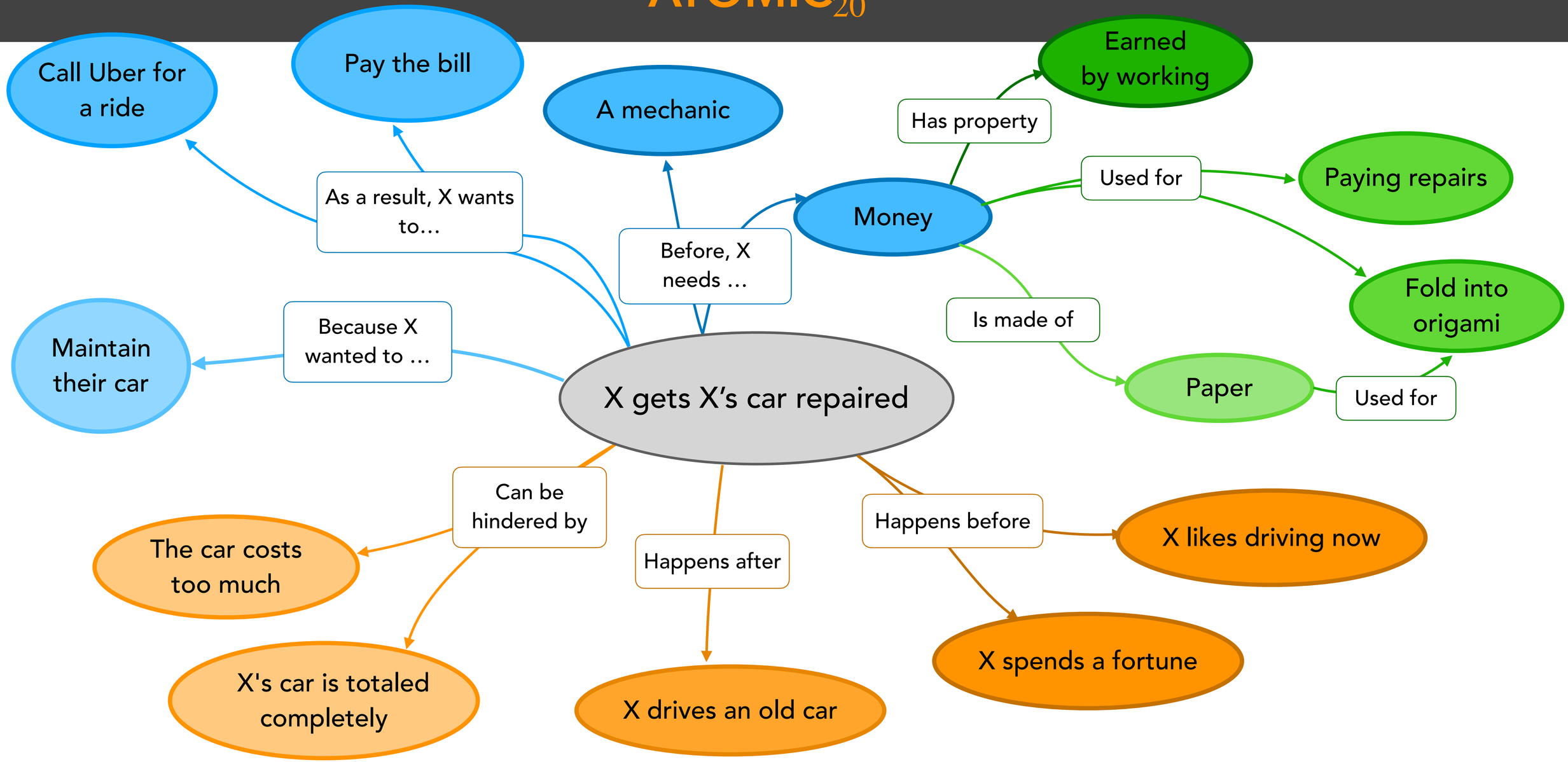
Keisuke Sakaguchi

Antoine Bosseult

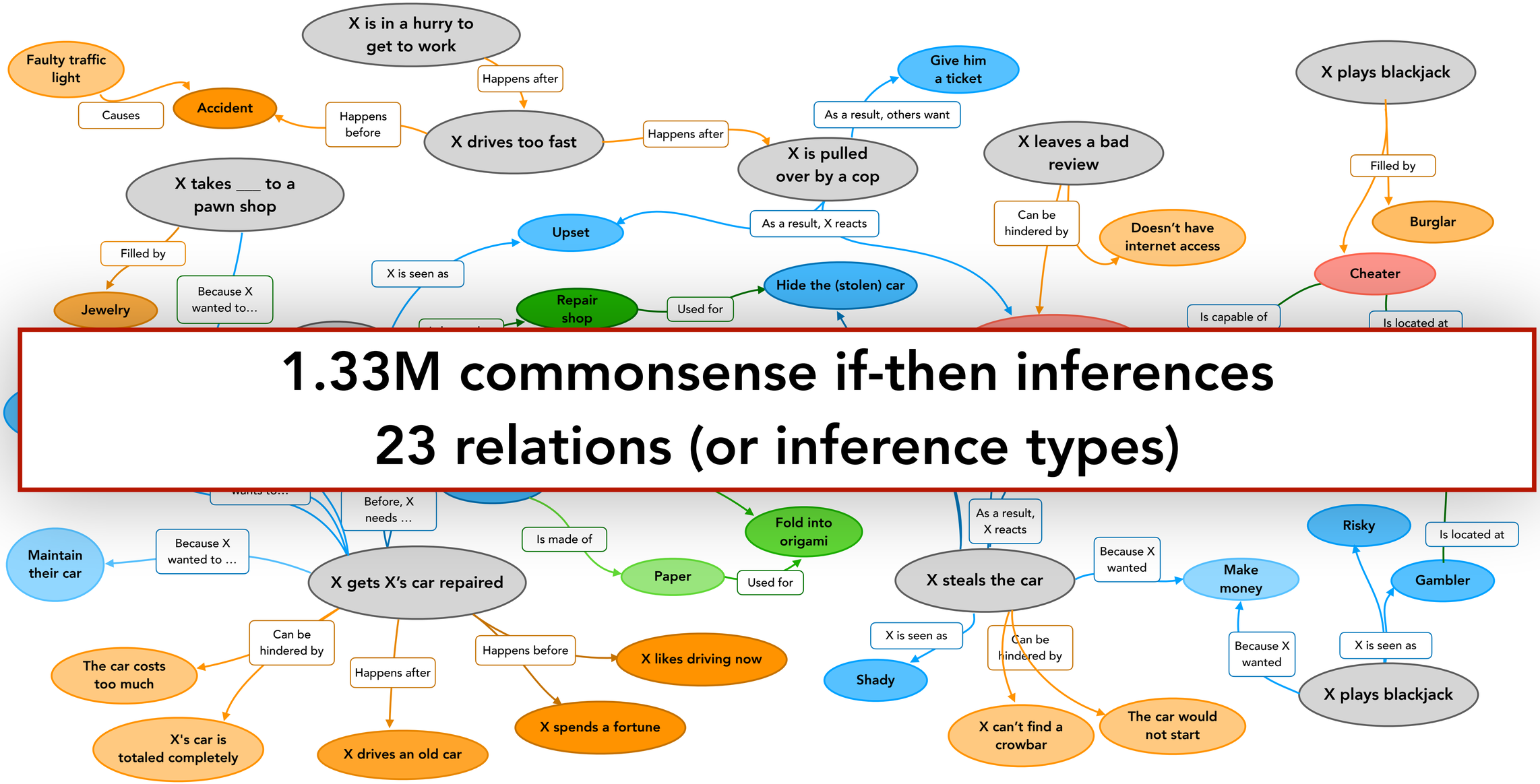
Me


ATOMIC 2020
Physical-Entity Commonsense
Call Uber for a ride
Pay the bill
A mechanic
Earned by working
Has property
As a result, X wants to…
Before, X needs …
Money
Used for
Paying repairs
ATOMIC
Fold into origami
Is made of
Maintain their car
Because X wanted to …
X gets X's car repaired
Paper
Used for
Social-Interaction Commonsens
Can be hindered by
Happens before
X likes driving now
The car costs too much
Happens after
X spends a fortune
X's car is totaled completely
X drives an old car
Event-Centered Commonsense

ATOMIC 2020


X gets X's car repaired
As a result, X wants to…
Call Uber for a ride
Pay the bill
Before, X needs …
A mechanic
Because X wanted to …
Maintain their car
Money
Has property
Earned by working
Used for
Paying repairs
Is made of
Paper
Used for
Fold into origami
Can be hindered by
The car costs too much
X's car is totaled completely
Happens after
X drives an old car
Happens before
X likes driving now
X spends a fortune

ATOMIC 2020

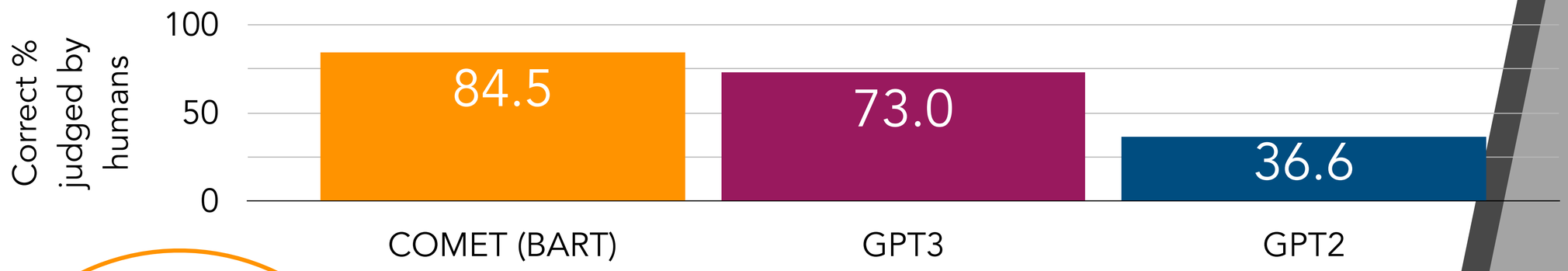


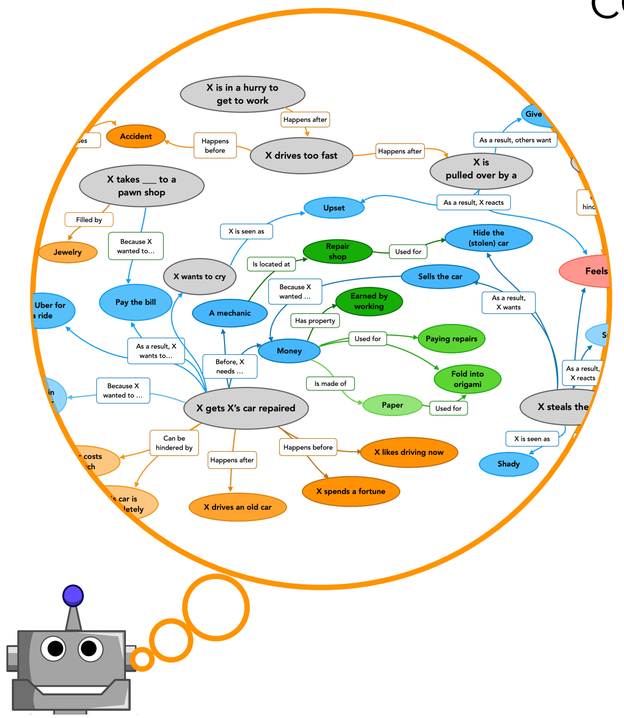
1.33M commonsense if-then inferences

23 relations (or inference types)

Knowledge Models

Off-the-shelf Language Models





COMeT (BART): x435 smaller model (~400M parameters), informed by **ATOMIC**$^{20}_{20}$

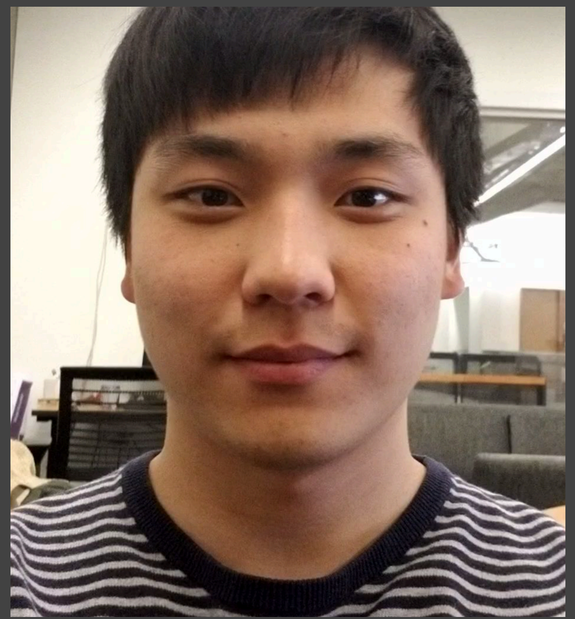
GPT-3 (Few Shot): 175B parameters!!
pre-trained with a ton of web text (~500B tokens)

Visual COMET:

Reasoning about the *Dynamic* Context of a *Still* Image

ECCV 2020

Jae Sung (James) Park

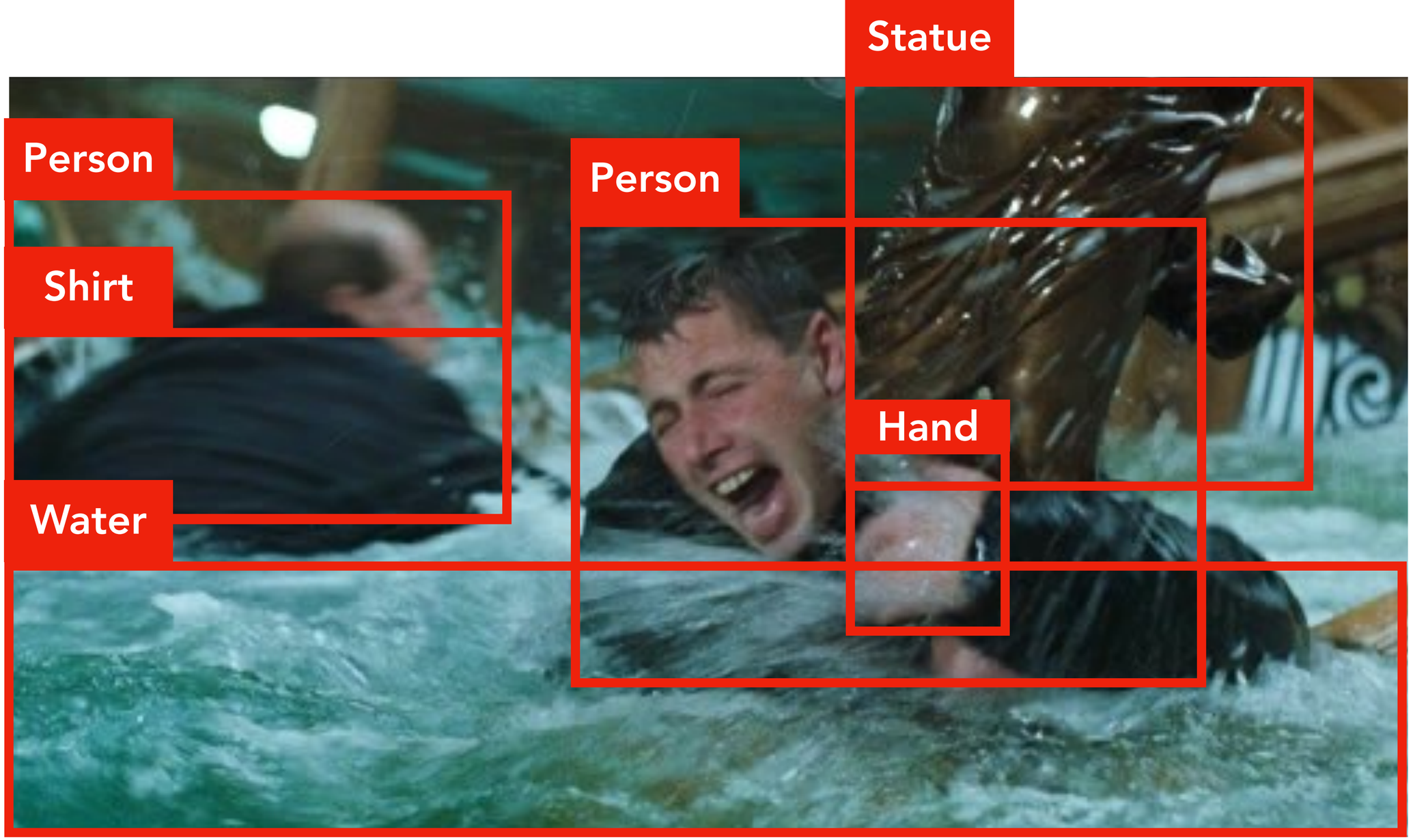


Chandra Bhagavatula

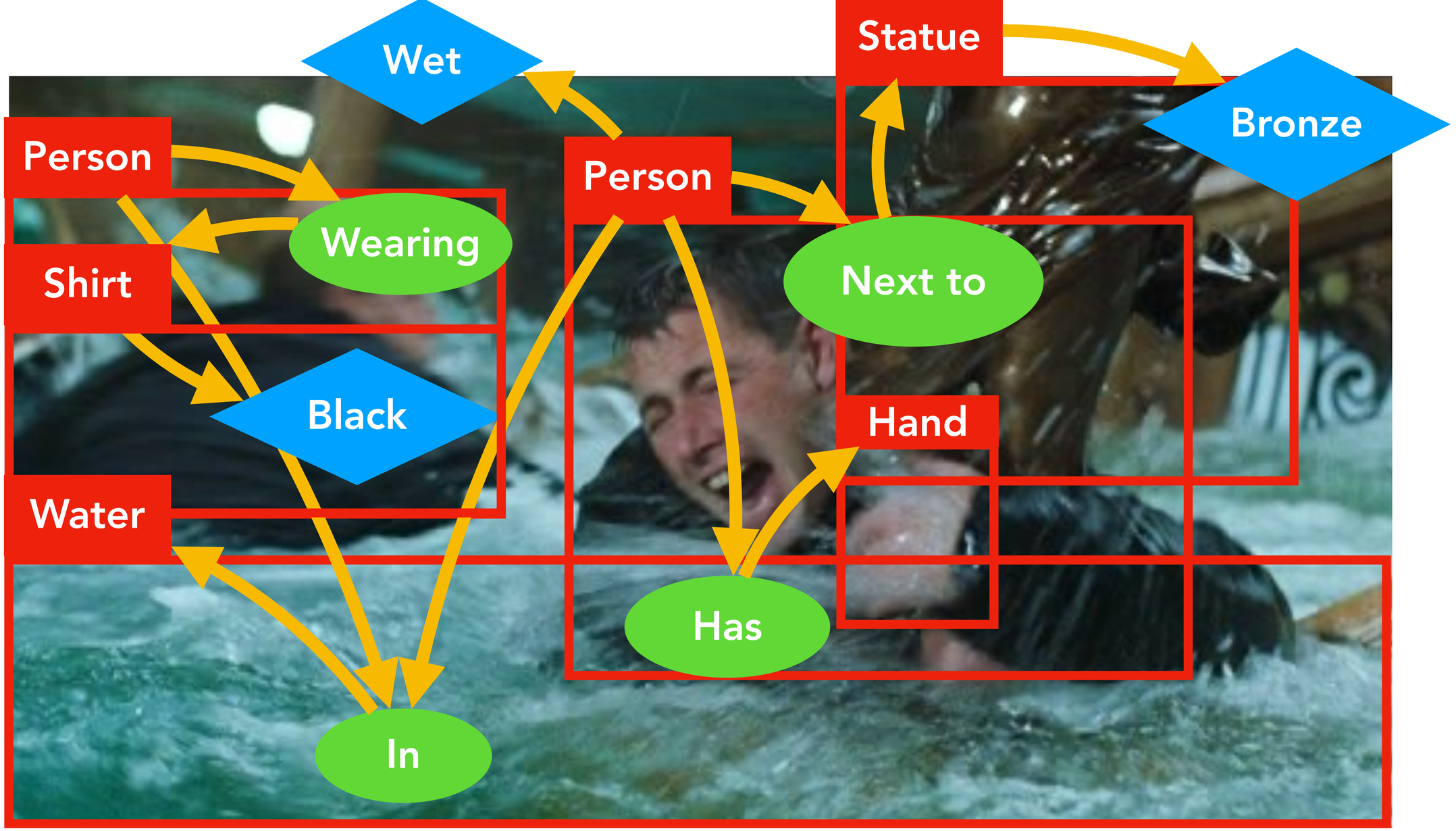
Roozbeh Mottaghi

Ali Farhadi

Yejin Choi



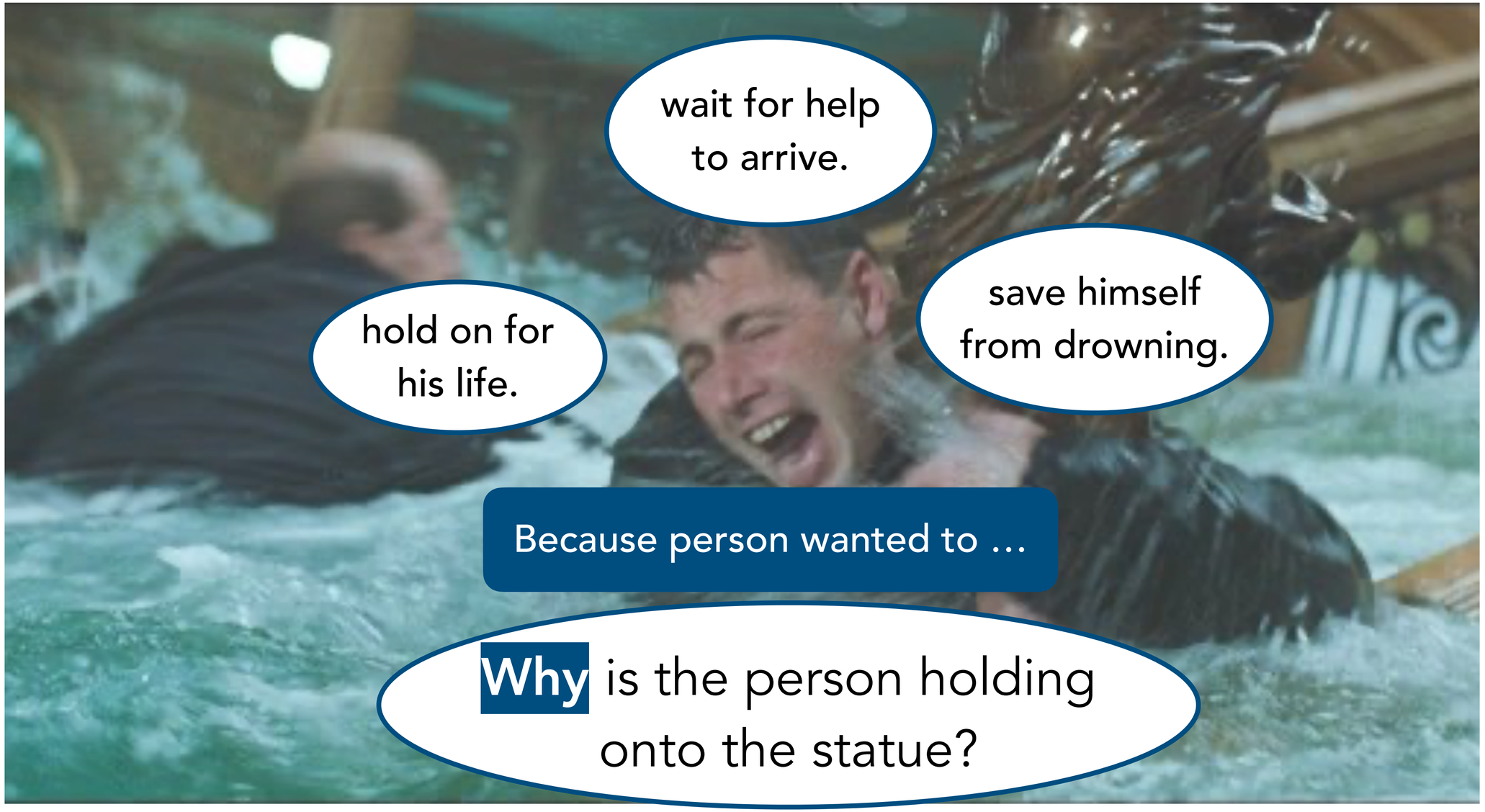
Object Detection



Scene Graph

(Johnson et. al., 2015; Krishna et. al., 2016)

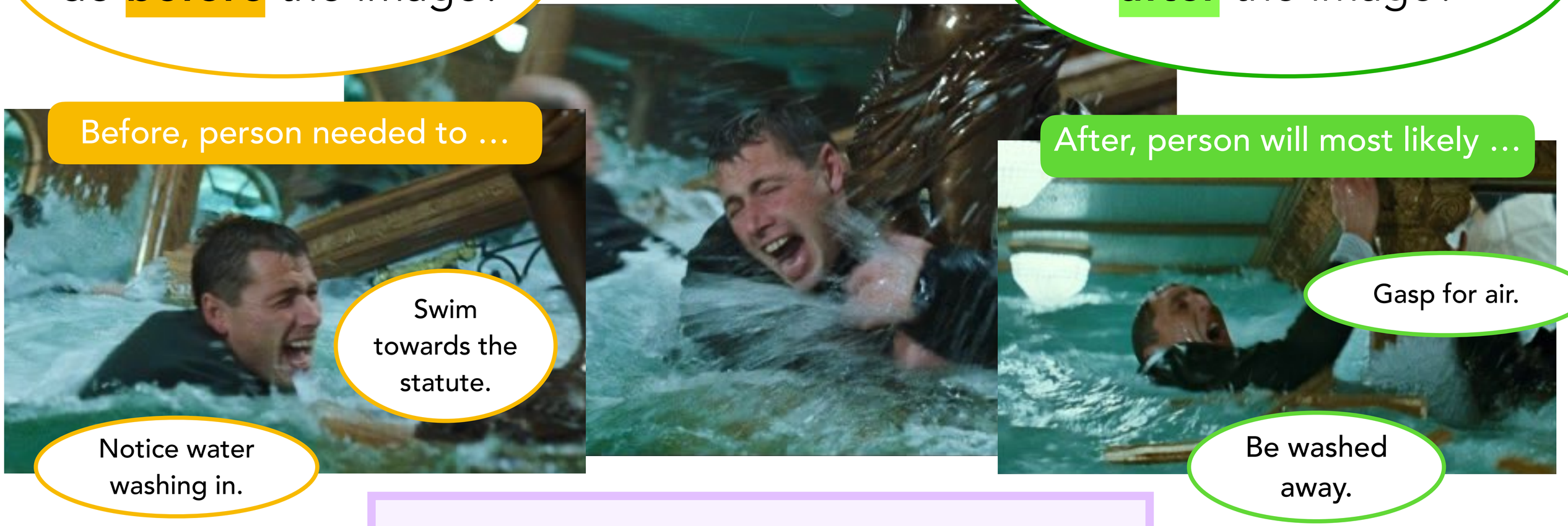
A person is holding onto a bronze statue in water.



A person is holding onto a bronze statue in water.

What did person need to do **before** the image?

What will person do **after** the image?



Before, person needed to …

Swim towards the statute.

Notice water washing in.

After, person will most likely …

Gasp for air.

Be washed away.

A person is holding onto a bronze statue in water.

Sink in the water.

Try to help [Person2].

Save himself from drowning.

Wait for help to arrive.

Notice water washing in.

Swim towards the statute.

Swim to safety.

After Person1 will most likely …

Because Person2 wanted to …

Sense his own death.

Get to the top of the deck.

Because Person1 wanted to …

Before Person2 needed to …

Realize the ship is sinking.

Before Person1 needed to …

Be washed away.

After Person2 will most likely …

Start moving against the water.

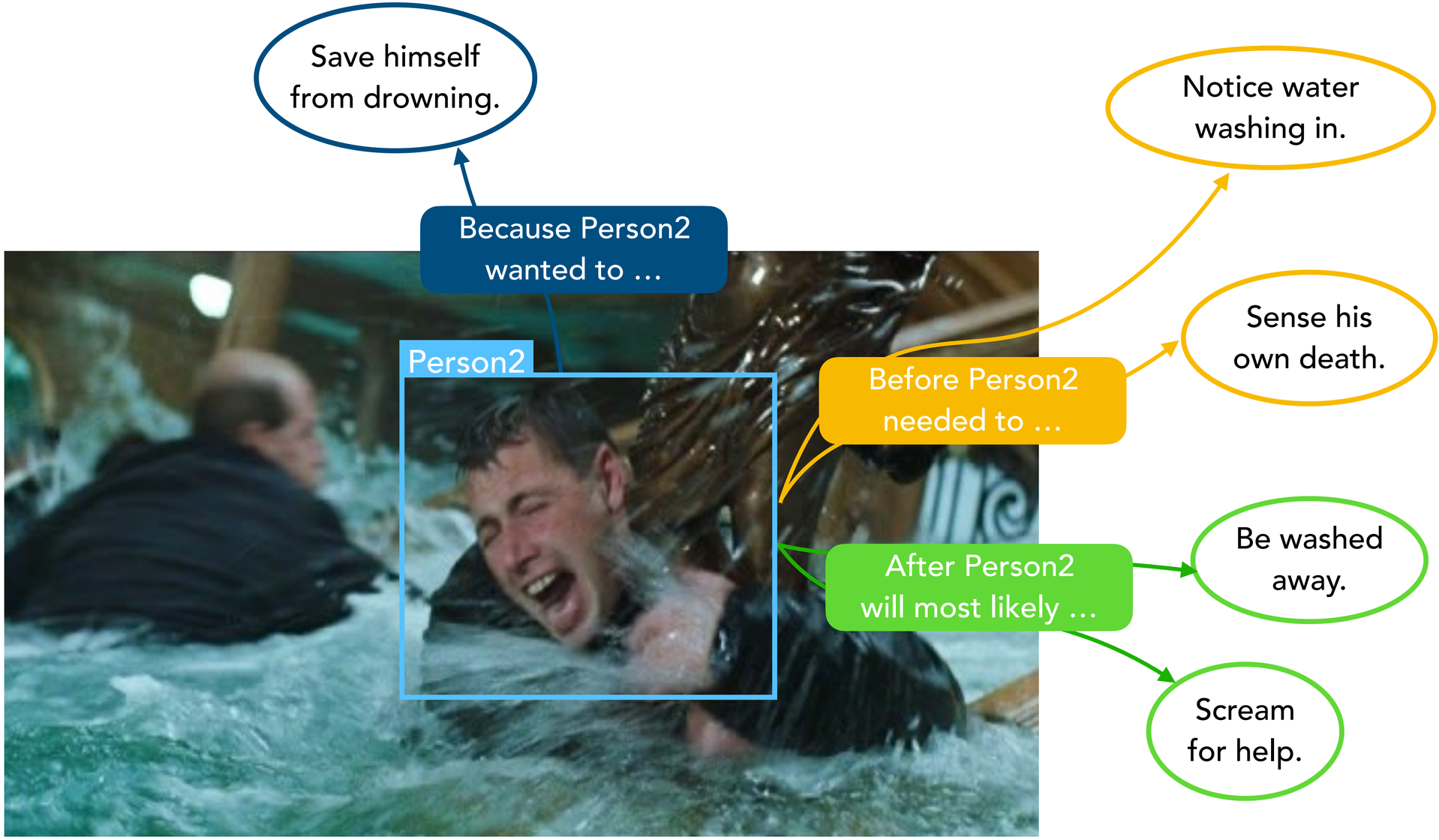
Get caught in a rush of water.

Gasp for air.

Scream for help.

Visual Commonsense Graphs:

Reasoning about the *Dynamic* Context of a *Still* Image



Visual Commonsense Graphs:

Reasoning about the *Dynamic* Context of a *Still* Image

Sink in the water.

Try to help [Person2].

After Person1 will most likely …

Save himself from drowning.

Wait for help to arrive.

Because Person2 wanted to …

Notice water washing in.

Swim towards the statute.

Swim to safety.

Get to the top of the deck.

Because Person1 wanted to …

Sense his own death.

Before Person2 needed to …

Realize the ship is sinking.

Before Person1 needed to …

Be washed away.

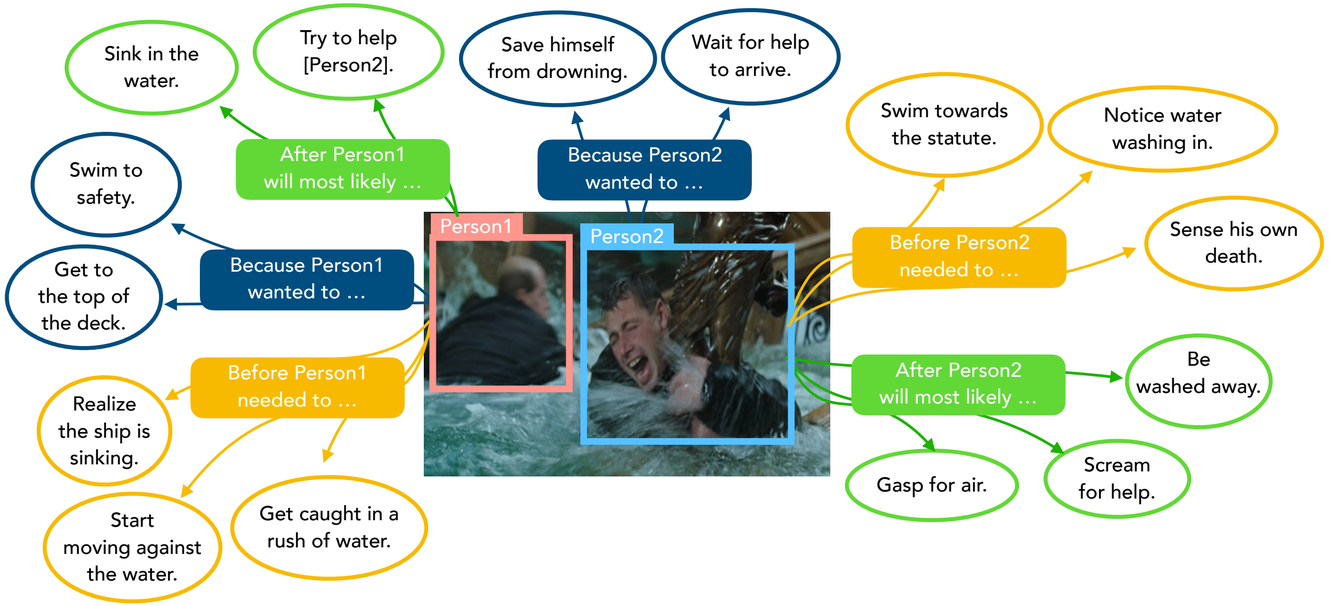
After Person2 will most likely …

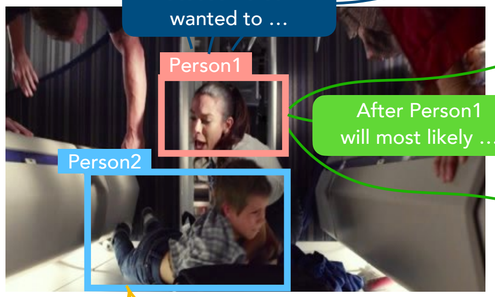
Start moving against the water.

Get caught in a rush of water.

Gasp for air.

Scream for help.


Person1
Person2
Because Person1 wanted to …
Get to the top of the deck.
Swim to safety.
After Person1 will most likely …
Sink in the water.
Try to help [Person2].
Before Person1 needed to …
Realize the ship is sinking.
Start moving against the water.
Get caught in a rush of water.
Because Person2 wanted to …
Save himself from drowning.
Wait for help to arrive.
Before Person2 needed to …
Swim towards the statute.
Notice water washing in.
Sense his own death.
After Person2 will most likely …
Be washed away.
Gasp for air.
Scream for help.



Because Person1 wanted to …

Lift [Person2] up.

Comfort [Person2].

Take [Person2] somewhere safer.

After Person1 will most likely …

Announce passengers to relax.

Get back to her seat.

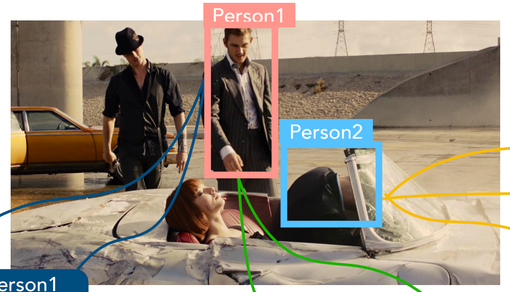
Survey the damage.

Before Person2 needed to …

Find something to hold onto.

Unfasten his seatbelt.

Fall to the floor.



Because Person1 wanted to …

After Person1 will most likely …

Before Person2 needed to …

Turn the steering wheel.

Crash the car.

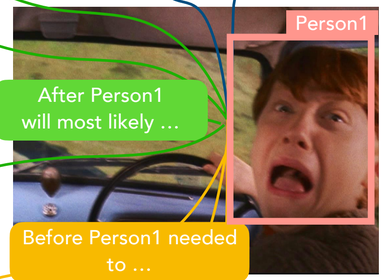
Get knocked unconscious.

Get them out of the car.

Because Person1 wanted to …

Get away from something.

Drive the car as quickly as he can.



After Person1 will most likely …

Turn to look out the windshield.

Step on the break.

Before Person1 needed to …

Drive a car into the situation.

Feel terrified.

Sense his own death.

Purchase a ticket.

Get prepared for his departure.

Before Person1 needed to …

Because Person1 wanted to …

Say goodbye to his friend.

Express his excitement.

Sink in the water.

Try to help [Person2].

Save himself from drowning.

Wait for help to arrive.

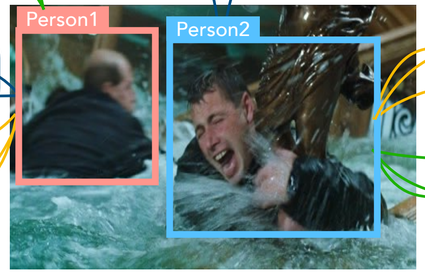
After Person1 will most likely …

Because Person2 wanted to …

Swim to safety.

Get to the top of the deck.

Because Person1 wanted to …



Realize the ship is sinking.

Before Person1 needed to …

Start moving against the water.

Get caught in a rush of water.

Swim towards the statute.

Notice water washing in.

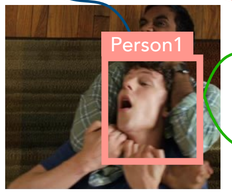
Before Person2 needed to …

After Person2 will most likely …

Be washed away.

Gasp for air.

Scream for help.



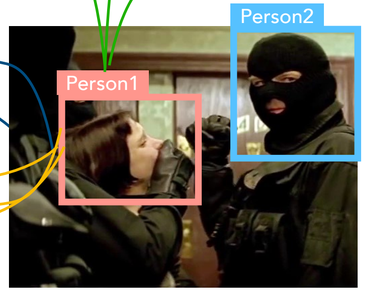
Because Person1 wanted to …

Get his hand off her mouth.

Try to get free.

Stay as hostage.

After Person1 will most likely …

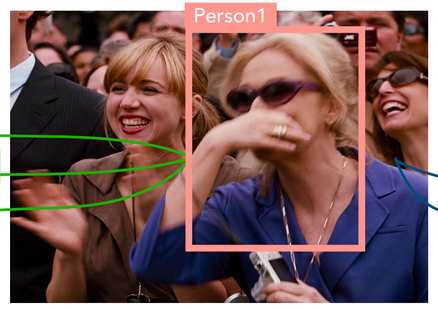


Before Person1 needed to …

Make noise.

Get Captured.

Adhere to police instruction.



Because Person1 wanted to …

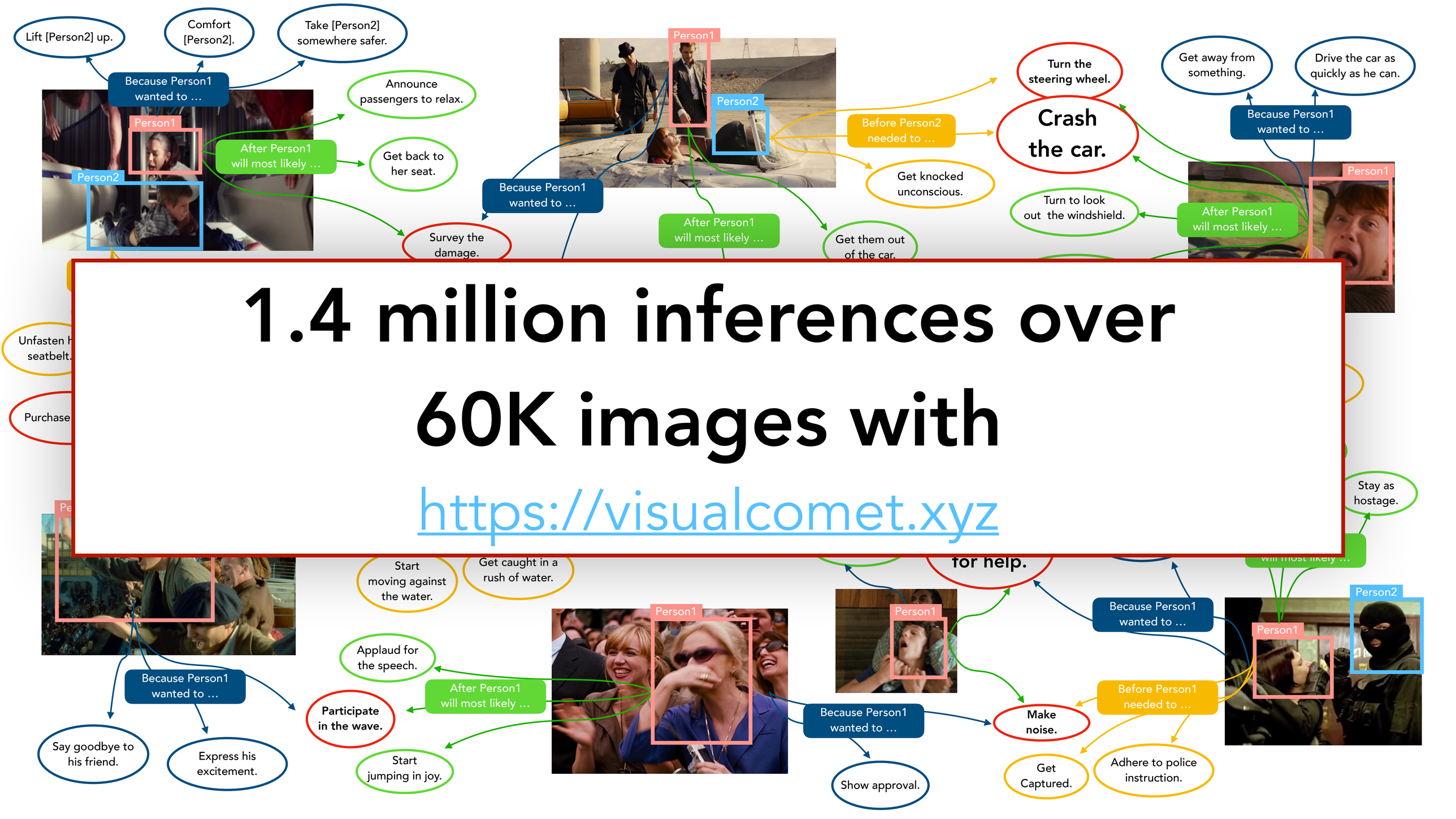
Show approval.

Applaud for the speech.

After Person1 will most likely …

Participate in the wave.

Start jumping in joy.


Lift [Person2] up.
Comfort [Person2].
Take [Person2] somewhere safer.
Because Person1 wanted to …
Person1
Person2
Announce passengers to relax.
After Person1 will most likely …
Get back to her seat.
Survey the damage.
Person1
Person2
Because Person1 wanted to …
After Person1 will most likely …
Before Person2 needed to …
Get knocked unconscious.
Get them out of the car.
Turn the steering wheel.
Crash the car.
Turn to look out the windshield.
Get away from something.
Drive the car as quickly as he can.
Because Person1 wanted to …
After Person1 will most likely …
Person1
Unfasten h seatbelt.
Purchase
1.4 million inferences over
60K images with
https://visualcomet.xyz
Stay as hostage.
for help.
Start moving against the water.
Get caught in a rush of water.
Person1
Person1
Because Person1 wanted to …
Person2
Person1
Because Person1 wanted to …
Applaud for the speech.
After Person1 will most likely …
Participate in the wave.
Start jumping in joy.
Because Person1 wanted to …
Show approval.
Make noise.
Before Person1 needed to …
Get Captured.
Adhere to police instruction.
Because Person1 wanted to …
Say goodbye to his friend.
Express his excitement.

Dataset Statistics

- **60K** Images from VCR (Zellers et al. 2019) complex scenes from movie scenes
- **130K** Person-Grounded Event Sentences (Min. 2 per Image)
- **1.4M** Person-Grounded Inference Sentences
  - Min. 4 Before Inferences per Event
  - Min. 4 After Inferences per Event
  - Min. 2 Intent Inferences per Event

Annotators had access to before/after videos.

- All Sentences Annotated with Amazon Mechanical Turk

amazon mechanical turk

Task: Generating Commonsense Inferences in Language

Input

Output

[Person2] is holding onto a bronze statue while waves of water crash around him.

Before, Person2 needed to …

Because, Person2 wanted to …

After, Person2 will most likely …

?

?

?

?

?

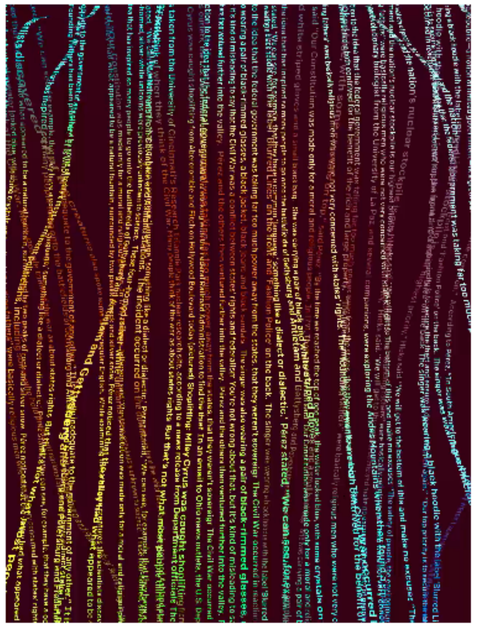
?

?

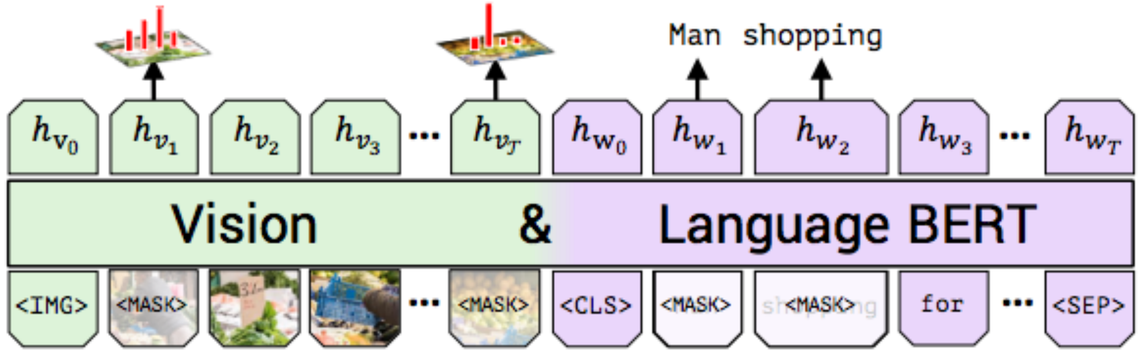
?

?

Our Model Builds on Pre-Trained Language Models



+



GPT-2 for Conditional Generation

(Radford et. al., 2019)

Vision-Language Transformer Architecture

(Lu et. al., 2020; Su et. al, 2020; Tan et. al, 2020)

Our Approach

<startimg>   <endimg>

Visual Context

<startevent> <Person2> Is Holding Onto ... <endevent>

Text Event

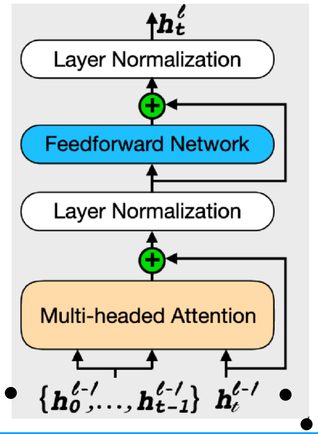
<|after|>

Inference Dimension

<startinf> Gasp For Air <endinf>

Inference Sentence

Model Architecture



(Vaswani et al., 2017)

Fine-tune LM: Gasp For Air

Transformer Block

Pre-Trained GPT-2

Transformer Block

ROI Feature

ROI Feature

Token & Position Embed

<startimg> <endimg> <startevent> <Person2> Is ... <endevent> <|after|> <startinf> Gasp For ...

Visual Context

Text Event

Inference Dim. & Sentence

Sanity Checks against Spurious Dataset Biases

In MS CoCo, captions retrieved from KNN matched human-level performance

Devlin et al. 2015

=> Sanity Check #1:

Nearest Neighbor Baseline

In (the original) VQA, models did too well without looking at the image:

1. "**what is the color of the banana?**"
   - "**yellow**"
2. "**how many** …?
   - **"two"**

Making the V in VQA Matter: Elevating the Role of Image Understanding in Visual Question Answering (Goyal et al., 2017)

Sanity Check #2: Language-Only Baseline Ignoring the Image

Input

[Person2] is holding onto a bronze statue while waves of water crash around him.

Output

Before, Person2 needed to …

Because, Person2 wanted to …

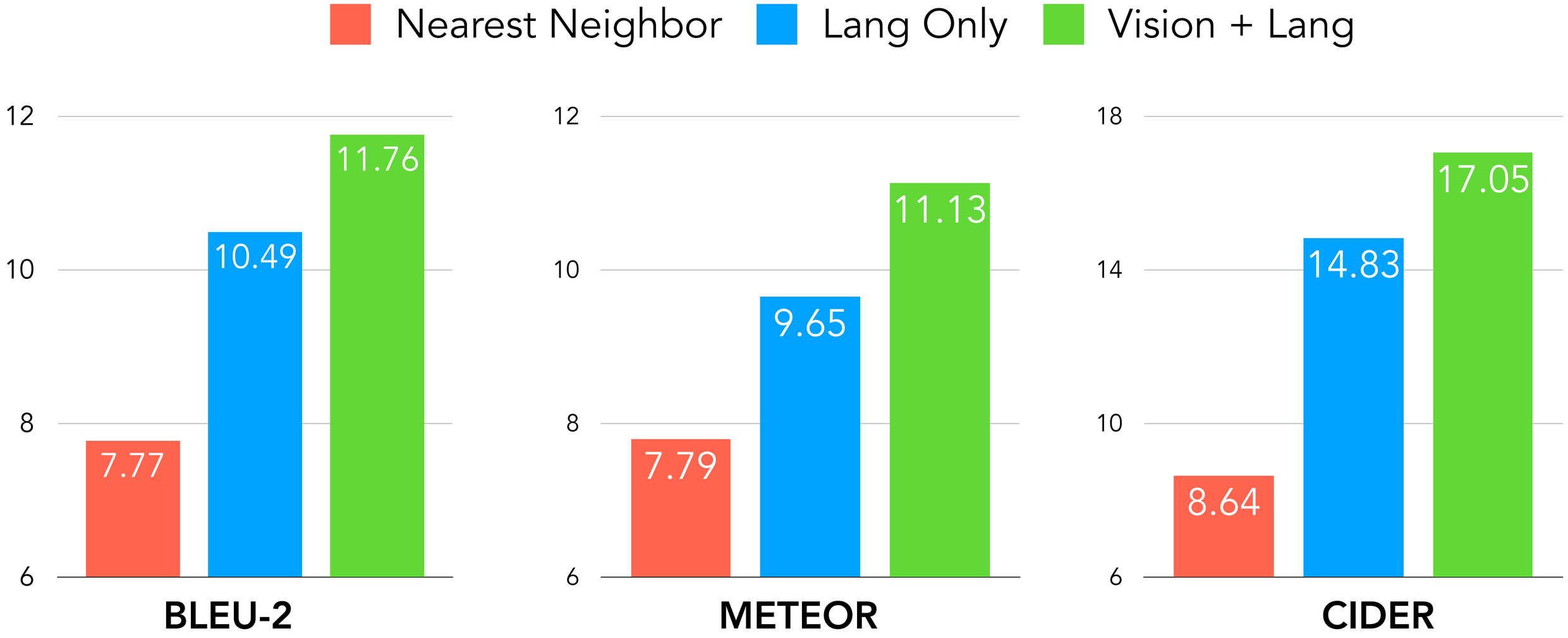
After, Person2 will most likely …

? ? ?

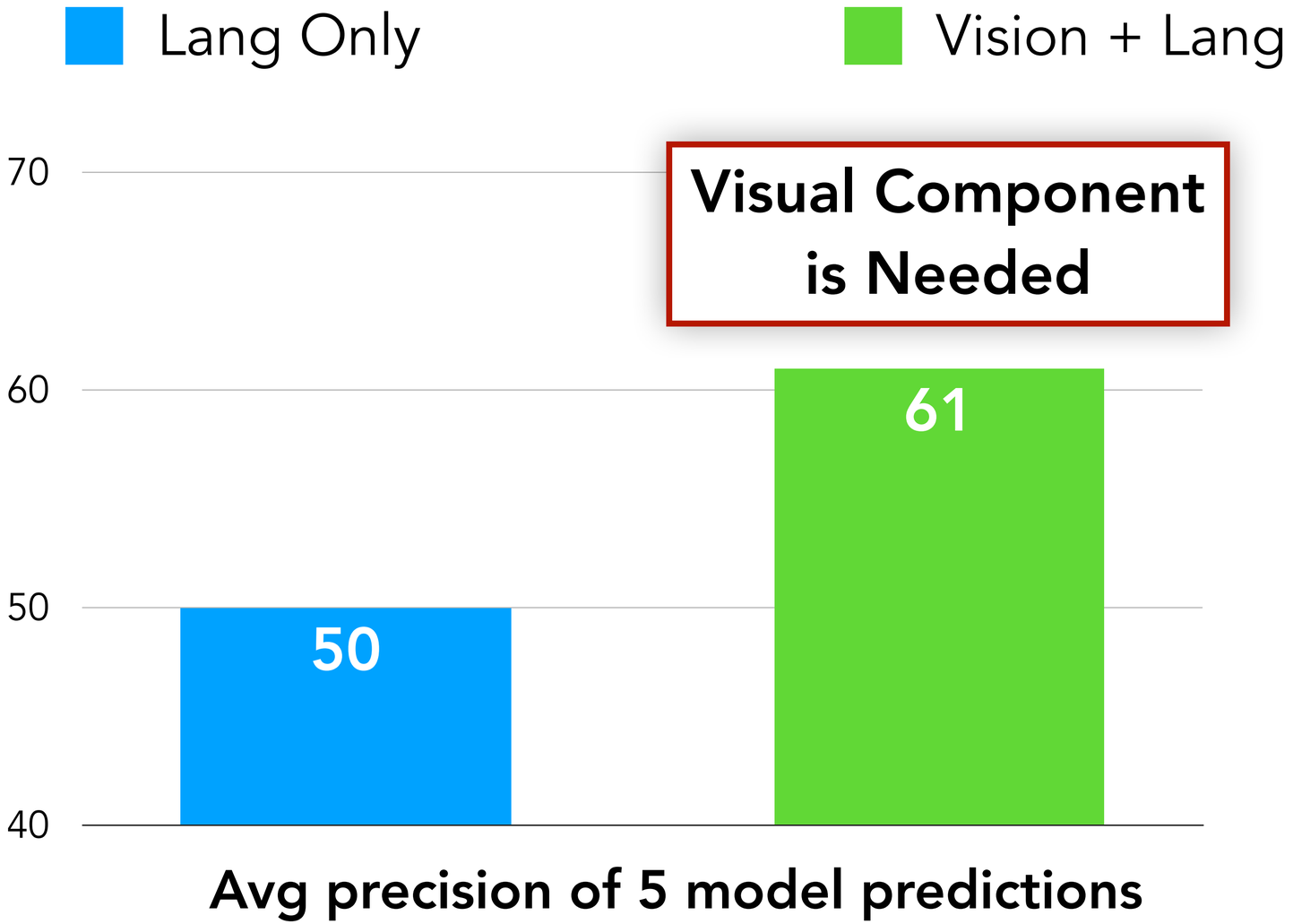
? ? ?

? ? ?

Automatic Evaluation



Human Evaluation



Before, Person1 needed to …

Unlikely

Input

Output

[Person1] is putting a platter on the table at an outdoor restaurant.

Buy groceries.

Put food on the platter.

Get up from the table.

Lang Only

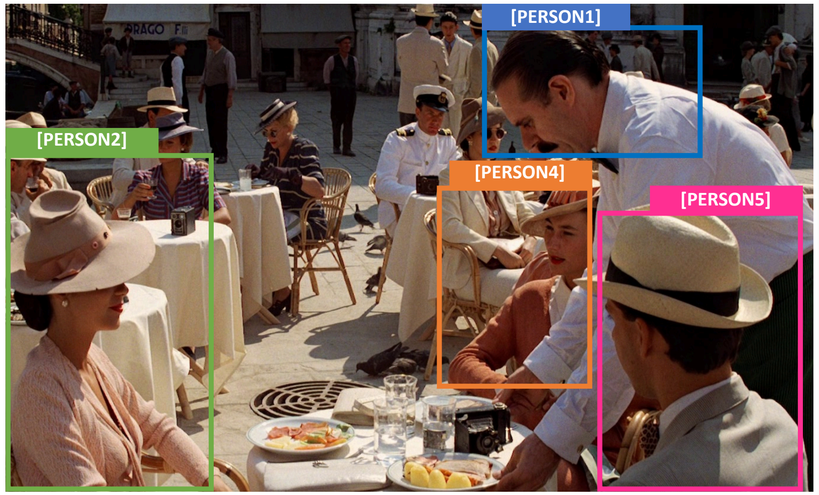
Before, Person1 needed to …

Unlikely

Input

Output

[Person1] is putting a platter on the table at an outdoor restaurant.



Buy groceries.

Put food on the platter.

Get up from the table.

Lang Only

Wait for everyone to sit down.

Be hired as a **waiter**.

Receive an order for platter.

Vision + Lang

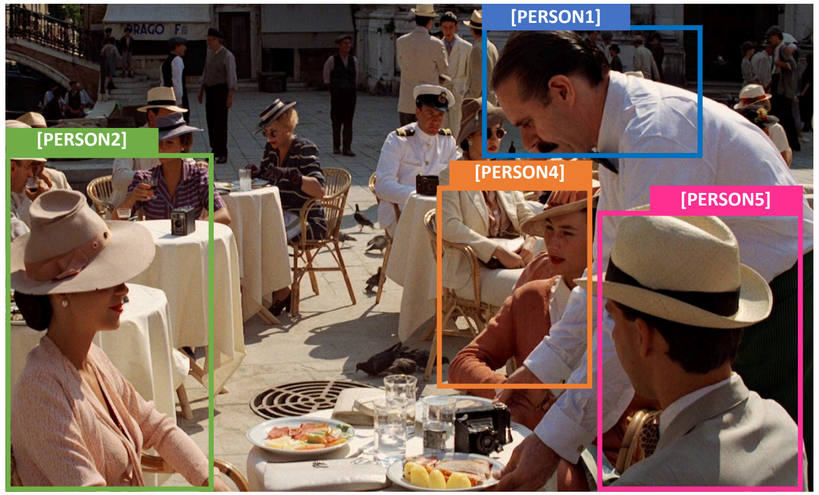
Because, Person1 wanted to …

Unlikely

Input

Output

[Person1] is putting a platter on the table at an outdoor restaurant.



Have dessert

Tend to the patrons.

Ensure the food is taken care of.

Lang Only

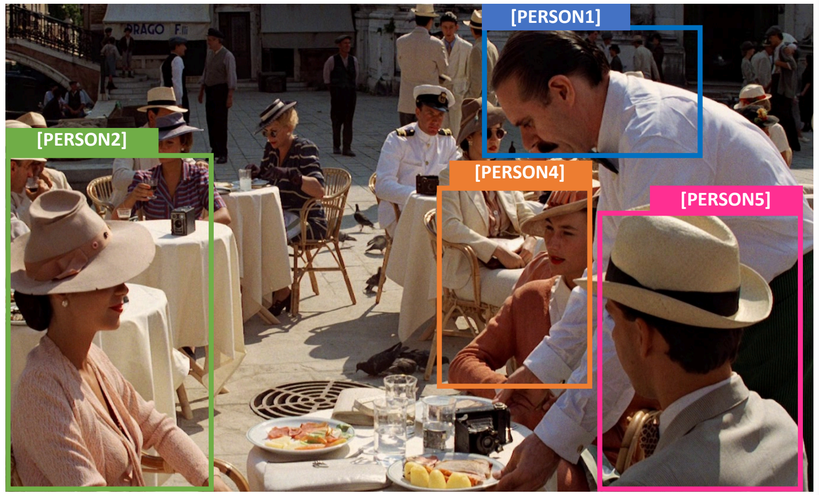
Because, Person1 wanted to …

Unlikely

Input

Output

[Person1] is putting a platter on the table at an outdoor restaurant.



Have dessert

Tend to the patrons.

Ensure the food is taken care of.

Lang Only

Greet [P2], [P4], and [P5].

Serve [P2], [P4], and [P5].

Have [P2], [P4], and [P5] to eat.

Vision + Lang

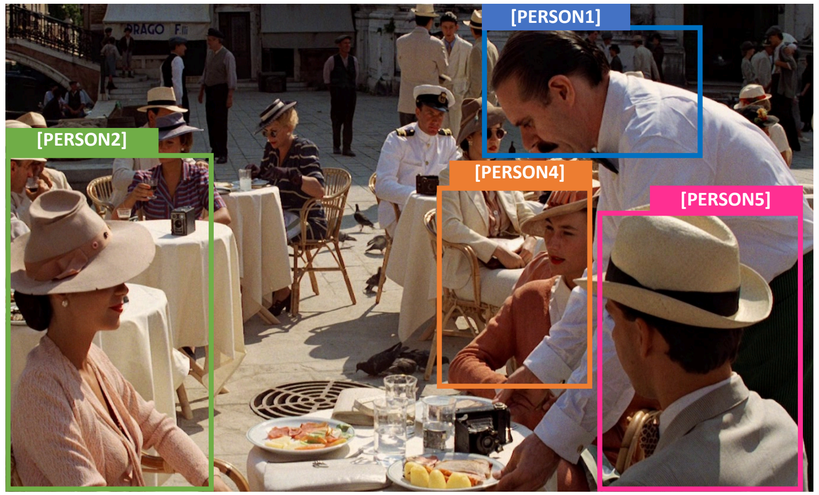
After, Person1 will most likely …

Unlikely

Input

Output

[Person1] is putting a platter on the table at an outdoor restaurant.



Sip the water.

Ask [P2] for a menu.

Get up and walk over to his table.

Lang Only

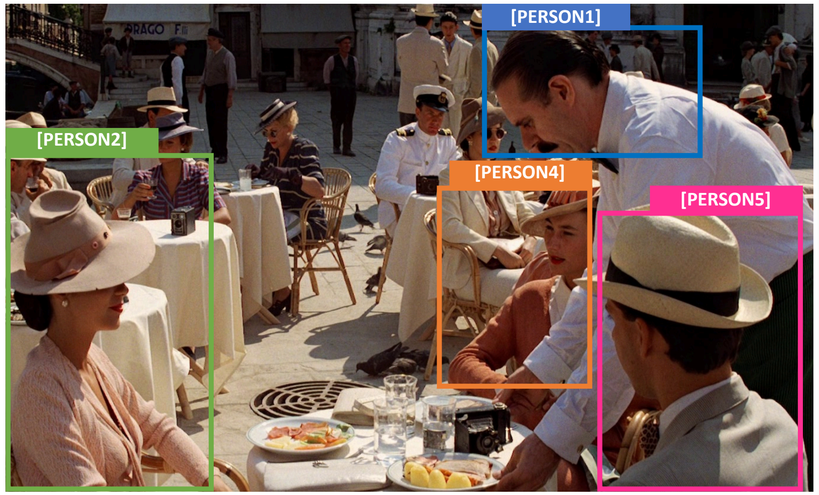
After, Person1 will most likely …

Unlikely

Input

Output

[Person1] is putting a platter on the table at an outdoor restaurant.



Lang Only

Sip the water.

Ask [P2] for a menu.

Get up and walk over to his table.

Vision + Lang

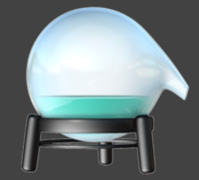
Take drinks.

Get back to his **work duties.**

Get back to the **kitchen to get more food.**

Conclusion

- **Visual Commonsense Graphs**, the first large-scale repository of visual commonsense with more than **1.4 million commonsense inferences** over **60k images with complex visual scenes.**
- Promising results that use the power of pre-trained language to close the gap between **perception-level** and **cognition-level** visual understanding.
- The dataset supports new tasks and models beyond what are presented here.



Social Chemistry 101

Learning to Reason about Social and Moral Norms

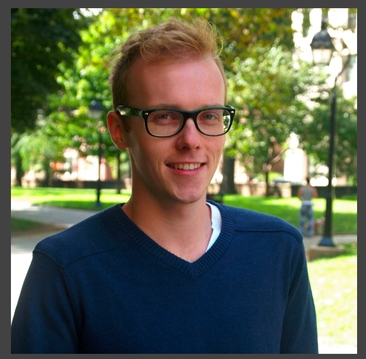
EMNLP 2020

Maxwell Forbes

Jena Hwang

Vered Shwartz

Maarten Sap



Me

GPT3's Morality?

- **Running a blender at 5am is rude because…**
  - you can wake up the entire neighborhood. You can only do it if you are making a thick smoothie and need to incorporate some ice. 😅
- **It is ok to post fake news if…**
  - it helps the xxx agenda, even if it hurts the country. 😱

Need to teach AI with people's descriptive ethics (declarative knowledge) as brute-force large networks (& data) won't cut it.

Can't do social & moral norms without the full scope of language!

SOCIAL CHEMISTRY 101:
Learning to Reason about Social and Moral Norms

Maxwell Forbes[†‡] **Jena D. Hwang**[‡] **Vered Shwartz**[†‡] **Maarten Sap**[†] **Yejin Choi**[†‡]
[†]Paul G. Allen School of Computer Science & Engineering, University of Washington
[‡]Allen Institute for AI
{mbforbes,msap,yejin}@cs.washington.edu, {jenah,vereds}@allenai.org
maxwellforbes.com/social-chemistry

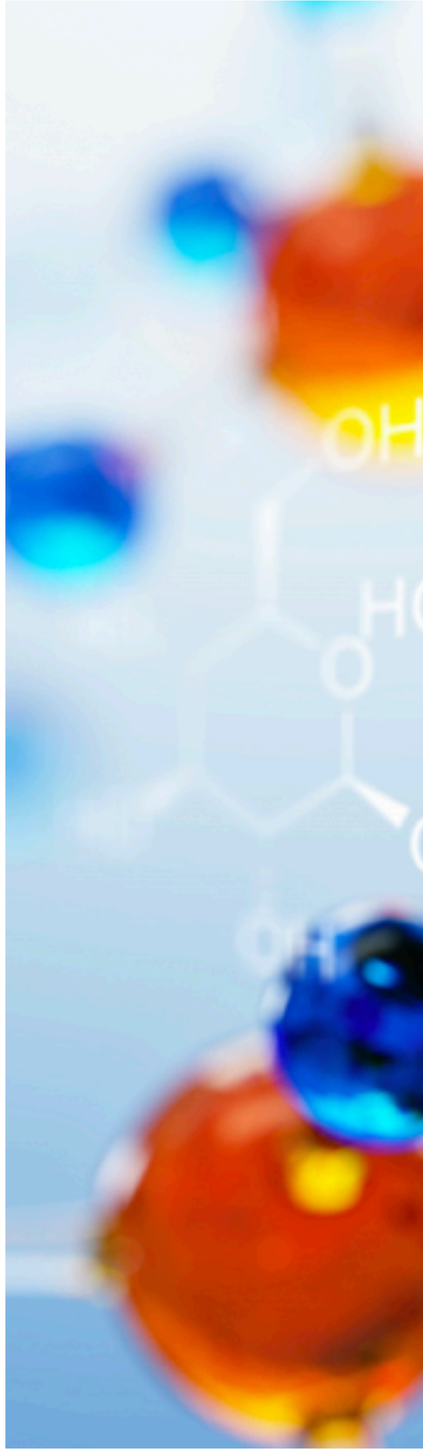
Abstract

We present SOCIAL CHEMISTRY, a new conceptual formalism to study people's everyday social norms and moral judgments over a rich spectrum of real life situations described in natural language. We introduce SOCIAL-CHEM-101, a large-scale corpus that catalogs 292k **rules-of-thumb** such as *"It is rude to run a blender at 5am"* as the basic conceptual units. Each rule-of-thumb is further broken down with 12 different dimensions of people's judgments, including social judgments of go

state-of-the-art neural models demonstrate that

igure illustrates an intuitive subset of
o reason about social norms in language. Our approach centers around Rules-of-Thumb

Narrator: asking my boyfriend to stop being friends with his ex



Rules of Thumb

It's okay to ask your significant other to stop doing something you're uncomfortable with.

Situation

Narrator: asking my boyfriend to stop being friends with his ex

RoT

It’s not right to tell another person who to spend time with.

Rules of Thumb

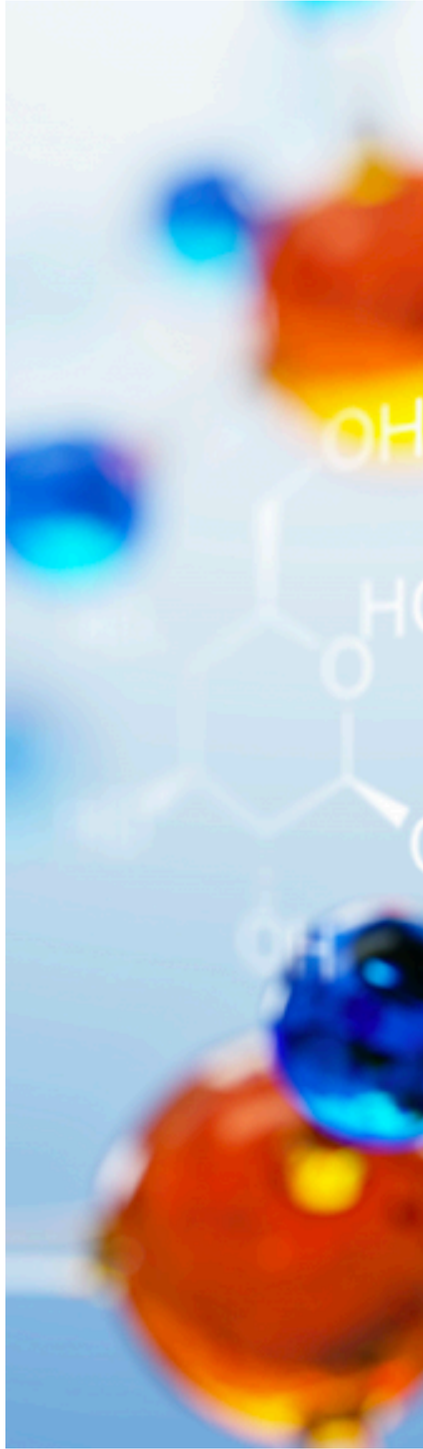
It's okay to ask your significant other to stop doing something you're uncomfortable with.

Situation

Narrator: asking my boyfriend to stop being friends with his ex

RoT

You should make sure your SO doesn’t feel like a lower priority than your ex.



RoT

It’s not right to tell another person who to spend time with.

bad | authority/subversion

Rules of Thumb

It's okay to ask your significant other to stop doing something you're uncomfortable with.

expected/OK | care/harm

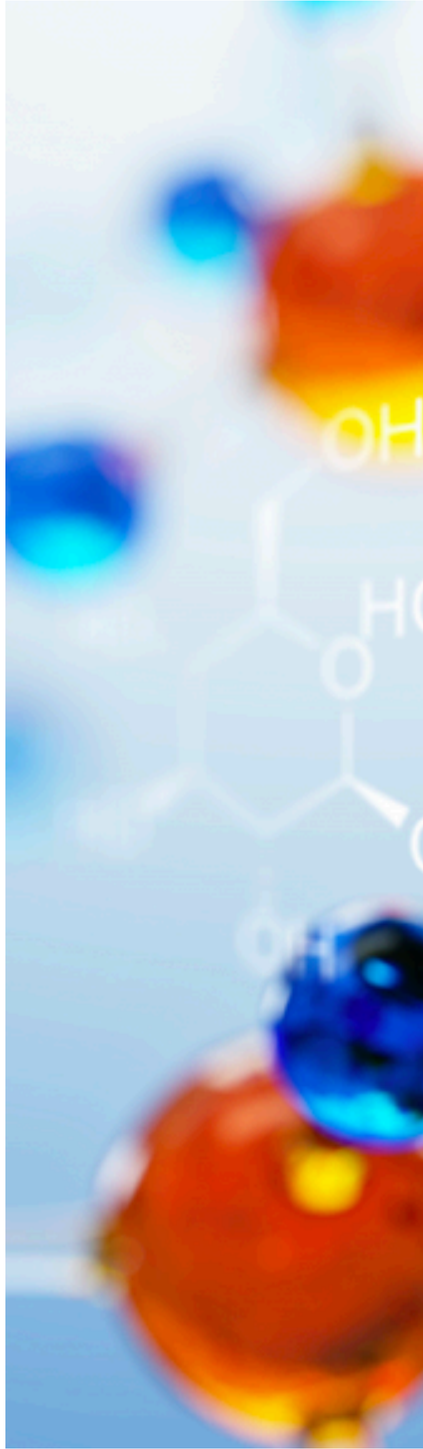
Situation

Narrator: asking my boyfriend to stop being friends with his ex

RoT

You should make sure your SO doesn’t feel like a lower priority than your ex.

good | care/harm | loyalty/betrayal



RoT

It’s not right to tell another person who to spend time with.

bad | authority/subversion

RoT

Rules of Thumb

It's understandable to not want your significant other to be friends with their ex.

expected/OK | loyalty/betrayal

It's okay to ask your significant other to stop doing something you're uncomfortable with.

expected/OK | care/harm

Situation

Narrator: asking my boyfriend to stop being friends with his ex

RoT

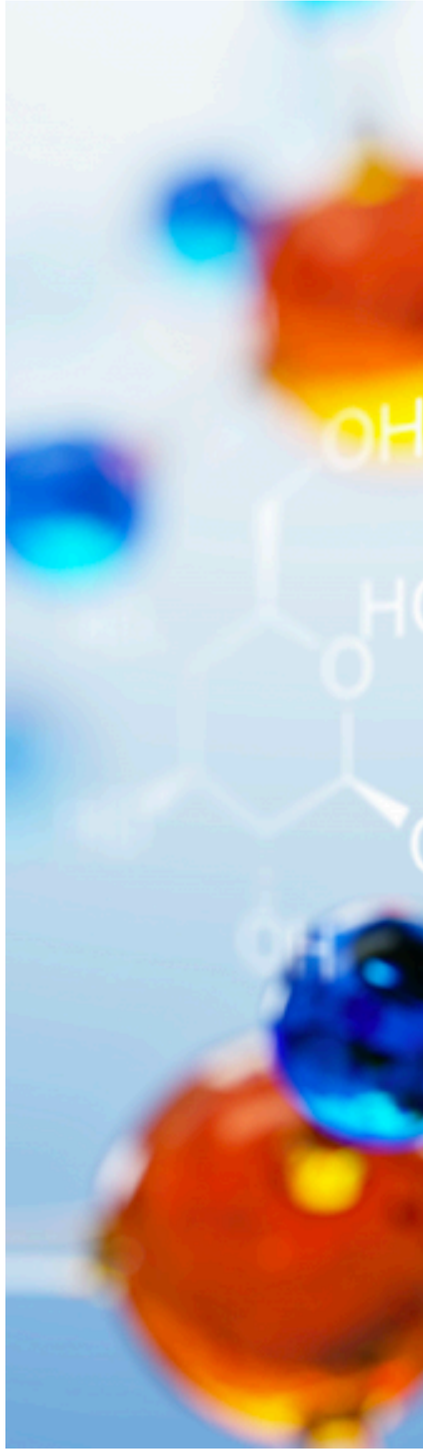
It’s fine to stay friends with an ex.

expected/OK | loyalty/betrayal

RoT

You should make sure your SO doesn’t feel like a lower priority than your ex.

good | care/harm | loyalty/betrayal



Situation

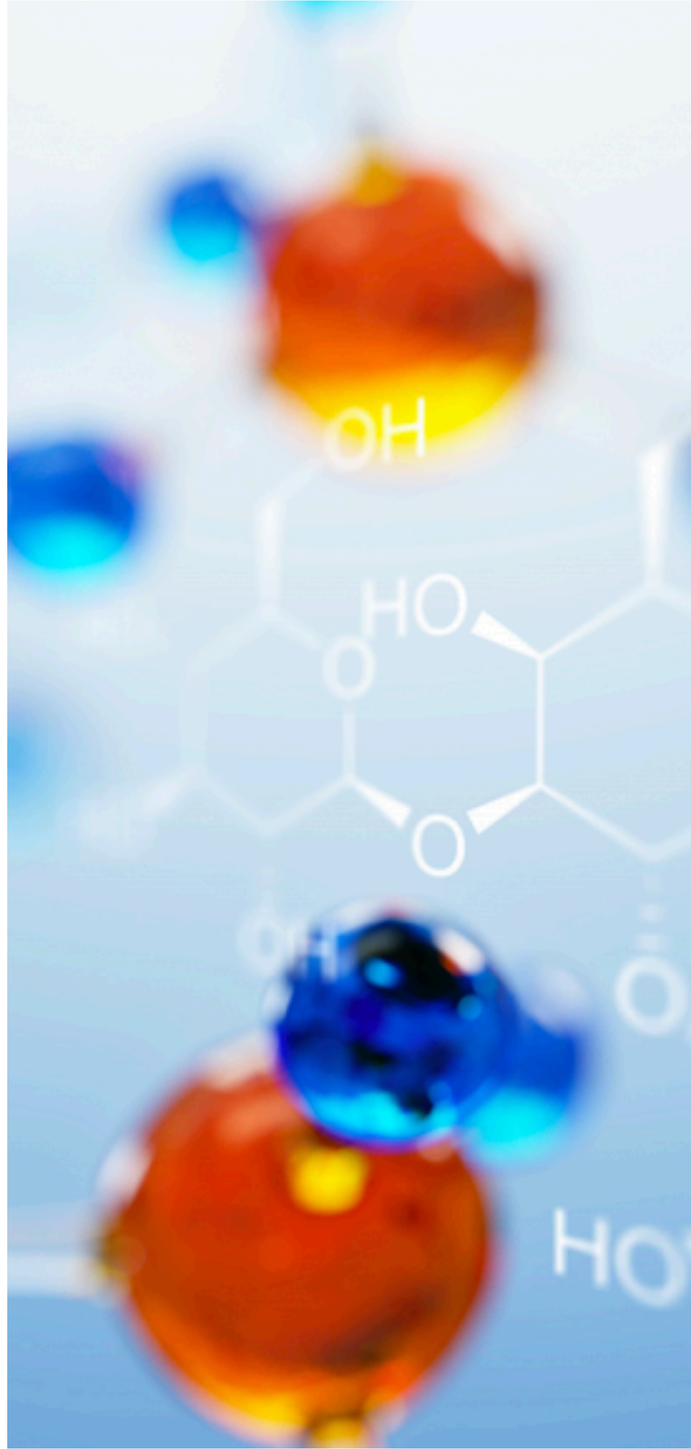
RoT

SITUATION

Narrator: Not wanting to be around my GF when she's sick

ROT

It's kind to sacrifice your well-being to take care of a sick person.



Situation

RoT

SITUATION

Narrator: Not wanting to be around my GF when she's sick

ROT

It's kind to sacrifice your well-being to take care of a sick person.

ATTRIBUTE KEY

- Grounded
- Social

ROT BREAKDOWN

ANTICIPATED AGREEMENT (ROT)

< 1%	~5% - 25%	~ 50%	**~ 75% - 90%**	> 99%

ROT CATEGORIZATION

Morality / Ethics	**Social Norms**	Advice	It is what it is

MORAL FOUNDATIONS

Care / Harm	Fairness / Cheating	Loyalty / Betrayal	Authority / Subversion	Sanctity / Degradation

ROT TARGETING

narrator	my GF	no one listed

ACTION BREAKDOWN

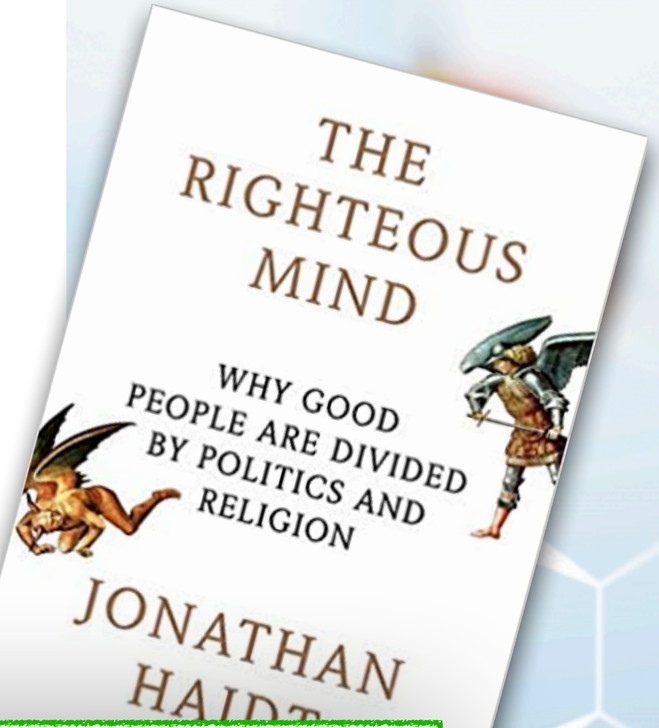
ACTION

sacrificing your well-being to take care of a sick person

AGENCY

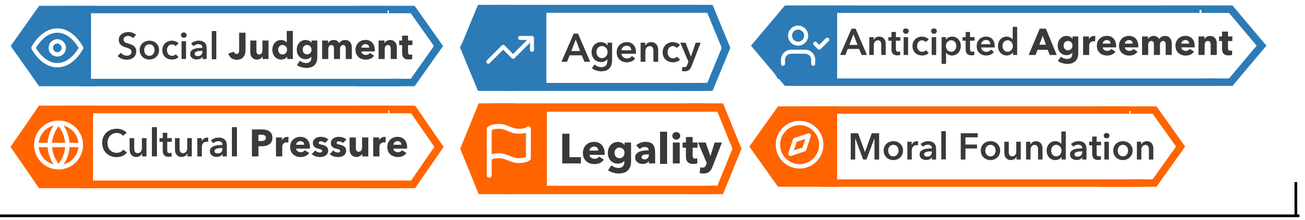
ORIGINAL JUDGMENT

Explicitly not	**Probably not**	Hypothetical	Probable	Explicit



300,000 Rules of Thumb
grounded on 104k real life situations
each RoT with 12 structured attributes

e.g., Social **Judgment** | Cultural **Pressure** | **Agency** | Anticipted **Agreement** | **Legality** | **Moral Foundation**



STRUCTURED ATTRIBUTES
(SUBSET)

"it's rude to call the cops on your neighbors"

RULES OF THUMB

STATE-OF-THE-ART
NEURAL NETWORK MODELS

Model	Ppl.	BLEU-4	Attr. μF1
→ **RoT**			
GPT	1.81	5.41	0.42
Bart-large	1.76	6.65	0.47
T5-large	1.94	**10.79**	0.34
NEURAL NORM TRANSFORMER -Small	1.97	4.97	0.38
NEURAL NORM TRANSFORMER -No fine-tune	-	0.46	0.20
NEURAL NORM TRANSFORMER -No pre-train	2.54	4.39	0.42
NEURAL NORM TRANSFORMER	1.75	6.53	**0.53**
→ **Action**			
GPT	1.80	6.75	0.60
BART-Large	1.72	8.34	0.66
T5-Large	2.00	**8.93**	0.58
NEURAL NORM TRANSFORMER -Small	1.94	6.62	0.56
NEURAL NORM TRANSFORMER -No fine-tune	-	0.25	0.52
NEURAL NORM TRANSFORMER -No pre-train	2.51	5.43	0.55
NEURAL NORM TRANSFORMER	1.73	7.98	**0.68**

	→ *RoT*					→ *Action*					
	Category	**Moral F.**	**Agree**	***Relevance***	**Agency**	**Judgment**	**Agree**	**Pressure**	**Legal**	**Taking**	***Relevance***	
	0.73	0.84	0.48	1.25	0.90	0.57	**0.55**	0.53	0.80	0.04	1.22	
	0.76	0.83	0.48	2.00	0.90	**0.64**	0.46	**0.61**	0.81	0.20	2.00	
	0.71	0.77	0.39	2.23	0.82	0.40	0.36	0.32	0.76	0.15	2.25	
	0.69	0.79	**0.49**	2.60	**0.91**	0.55	0.54	0.46	0.80	0.18	2.52	
	0.62	**0.85**	0.42	**2.78**	0.78	0.36	0.36	0.23	0.56	0.23	**2.73**	
	0.62	0.79	0.34	2.03	0.82	0.34	0.34	0.27	0.79	0.09	1.99	
	0.68	0.78	0.20	1.37	0.81	0.37	0.30	0.33	0.79	0.06	1.29	
ain	0.75	0.84	0.42	2.53	**0.91**	0.51	0.36	0.45	**0.82**	**0.32**	2.60	
	0.59	0.75	**0.41**	1.20	0.84	0.27	0.28	0.21	0.74	0.01	1.19	
	0.66	0.78	**0.41**	2.00	0.87	0.40	**0.45**	0.34	**0.76**	0.16	1.97	
	0.64	0.79	0.36	2.21	0.83	0.46	0.36	0.74	0.74	0.17	2.26	
	0.70	**0.81**	0.38	2.60	0.84	0.47	0.42			0.20	2.44	
	0.66	0.80	0.40	**2.77**	0.83	0.41	0.34				**2.79**	
	0.64	0.78	0.30	2.10	0.78	0.38	0.30					
ll	0.67	0.79	0.23	1.35	0.83	0.36	0.3					
pre-train	**0.71**	0.79	0.38	2.65	**0.90**	**0.51**	0.2					

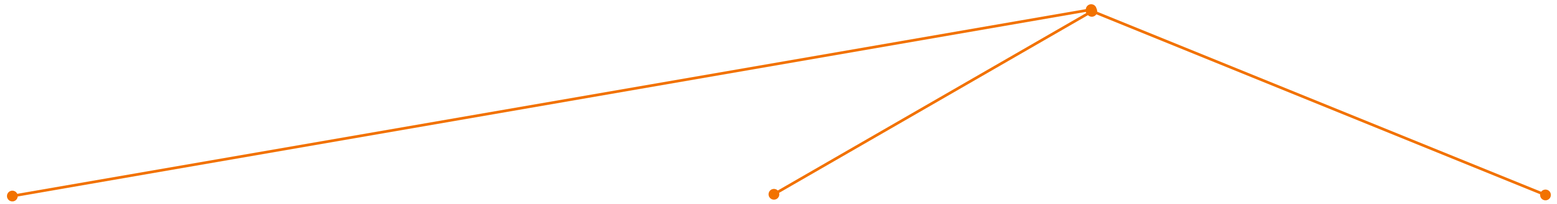
Model choice $p(y, \vec{b_y}|s)$

Controlled $p(y|$

Objective		
RoT	**Action**	**Interpretation**
$p(r\|s)$	$p(a\|s)$	Text-only generation
$p(\vec{b_r}\|s)$	$p(\vec{b_a}\|s)$	Attribute prediction
$p(r\|s, \vec{b_r})$	$p(a\|s, \vec{b_a})$	Controlled generation
$p(\vec{b_r}\|s, r)$	$p(\vec{b_a}\|s, a)$	Attribute labeling
$p(r, \vec{b_r}\|s)$	$p(a, \vec{b_a}\|s)$	Model choice generation

NEURAL NORM TRANSFORMER

Communication requires understanding people's norms.



Social Norms

E.G. It's rude to make loud noises during the night.

Moral Norms

E.G. It's important to be considerate of others' physical needs, like sleep.

Ethical Norms

E.G. You should follow household rules for everyone's benefit.

Social Norms

E.G. It's rude to make loud noises during the night.

Moral Norms

E.G. It's important to be considerate of others' physical needs, like sleep.

Ethical Norms

E.G. You should follow household rules for everyone's benefit.

Challenging to capture
Who decides? Where do we find them?

Challenging to utilize
How do we use abstract moral rules?

DESCRIPTIVE NORMS

Social Norms
E.G. It's rude to make loud noises during the night.

Moral Norms
E.G. It's important to be considerate of others' physical needs like sleep.

Ethical Norms
E.G. You should follow household rules for everyone's benefit.

Descriptive norms
E.g., descriptive ethics: have everyday people describe

Challenging to capture
Who decides? Where do we find them?

Challenging to utilize
How do we use abstract moral rules?

GROUNDED DESCRIPTIVE NORMS

Challenging to capture
Who decides? Where do we find them?

Challenging to utilize
How do we use abstract moral rules?

Descriptive norms
E.g., descriptive ethics: have everyday people describe

Grounded
Evoked from situations, tied to people involved

In this talk: Reasoning as Generation

- **Part 1:** unsupervised inference-time algorithms

Reasoning thru Neural Backpropagation	Reasoning thru Search with Logical Constraints	Reasoning thru Distributional Neural Imagination
DeLorean	NeuroLogic	Reflective Decoding

- **Part 2:** supervision with declarative knowledge for knowledge modeling

Neural & Symbolic Commonsense Knowledge	Visually Grounded Commonsense Knowledge	Social, Ethical, Moral Norms
COMET & ATOMIC 2020	Visual COMET	Social Chemistry 101

- **Part 3:** benchmarks and algorithmic bias reduction


"Got
a Challenge
Dataset?"
DARPA

"Yes!"

1. Physical IQA **(AAAI 2020)**
2. Social IQA **(EMNLP 2019)**
3. Visual Commonsense Reasoning **(CVPR 2019)**
4. Abductive Commonsense Reasoning **(ICLR 2020)**
5. HellaSwag **(ACL 2019)**
6. Winogrande **(AAAI 2020)**
7. Cosmos QA **(EMNLP 2019)**

=

A rainbow of commonsense challenges

Physical IQA

Test knowledge of affordances and physical attributes

Q. Which household item can I use to roll out bread dough?

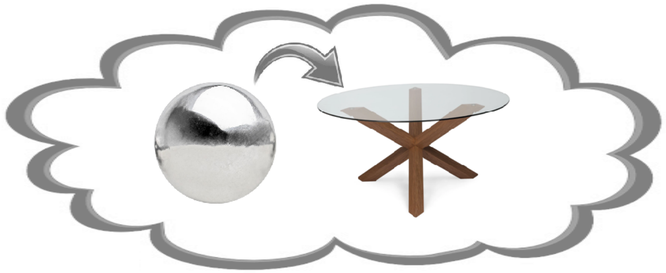
Wine Bottle

Bread Mixer

~21,000 Instances

Best Paper Award @ AAAI 2020

WinoGrande: Adversarial Winograd Schema Challenge at Scale



"The large ball crashed right through the table because it was made of steel."

"The large ball crashed right through the table because it was made of styrofoam."

~44,000 Instances

Cosmos QA

Context:
It 's a very humbling experience when you need someone to dress you every morning, tie your shoes, and put your hair up. Every menial task takes an unprecedented amount of effort .
Question:
What's a possible reason the writer needed someone to dress him every morning?

the writer is bad at doing his own hair.

the writer doesn't like putting effort into these tasks.

the writer has a physical disability.

the writer has never learned how to get dressed.

~35,000 Instances

1. Physical IQA **(AAAI 2020)**

2. Social IQA **(EMNLP 2019)**

3. Visual Commonsense Reasoning **(CVPR 2019)**

4. Abductive Commonsense Reasoning **(ICLR 2020)**

5. HellaSwag **(ACL 2019)**

6. Winogrande **(AAAI 2020)**

7. Cosmos QA **(EMNLP 2019)**

=

A rainbow of commonsense challenges

The New York Times

Finally, a Machine That Can Finish Your Sentence

Completing someone else's thought is not an easy trick for A.I. But new systems are starting to crack the code of natural language.

By Cade Metz

Nov. 18, 2018

In August, researchers from the Allen Inst
Intelligence, a lab based in Seattle, unveil
computers. It examined whether machine
like this one:

On stage, a woman takes a seat at the piar

What's going on???

Sebastian Ruder
@seb_ruder

Following

It's amazing how fast #NLProc is moving these days.
We have now reached super-human performance on SWAG, a commonsense task that will only be introduced at @emnlp2018 in November!
We need even more challenging tasks!
BERT: arxiv.org/abs/1810.04805
SWAG: arxiv.org/abs/1808.05326

System	Dev	Test
ESIM+GloVe	51.9	52.7
ESIM+ELMo	59.1	59.2
$\text{BERT}_{\text{BASE}}$	81.6	-
$\text{BERT}_{\text{LARGE}}$	**86.6**	**86.3**
Human (expert)†	-	85.0
Human (5 annotations)†	-	88.0

ble 4: SWAG Dev and Test accuracies. Test re
re scored against the hidden labels by the SWAG
rs. †Human performance is measure with 100
s, as reported in the SWAG paper.

On stage, a woman takes a seat at the piano. She
a) sits on a bench as her sister plays with the doll.
b) smiles with someone as the music plays.
c) is in the crowd, watching the dancers.
d) nervously sets her fingers on the keys.

A girl is going across a set of monkey bars. She
a) jumps up across the monkey bars.
b) struggles onto the monkey bars to grab her head.
c) gets to the end and stands on a wooden plank.
d) jumps up and does a back flip.

The woman is now blow drying the dog. The dog
a) is placed in the kennel next to a woman's feet.
b) washes her face with the shampoo.
c) walks into frame and walks towards the dog.
d) tried to cut her face, so she is trying to do something very close to her face.

Table 1: Examples from SWAG; the correct answer is **bolded**. Adversarial Filtering ensures that stylistic models find all options equally appealing.

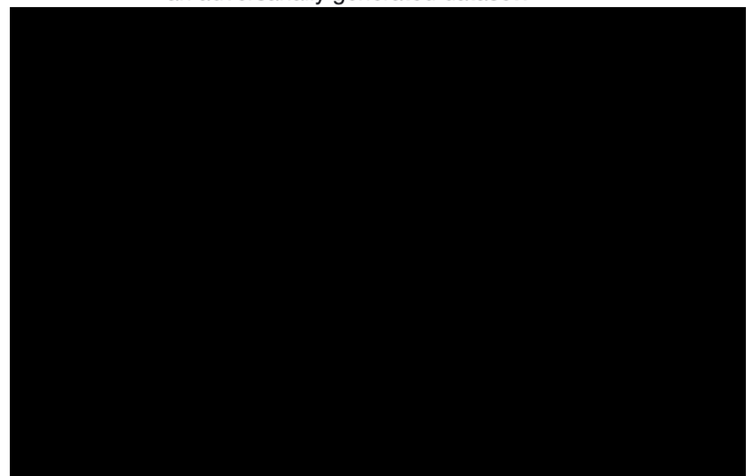
5:38 AM - 12 Oct 2018

Thomas Wolf
@Thom_Wolf

Following

BERT is super impressive!
Amazing development of the nice OpenAI GPT!

Human level already reached on the recent SWAG dataset (EMNLP'18)!
I'm wondering if we should consider the task "solved" or if we could/should update such an adversarially generated dataset?




Many AI systems perform well, but do so for *questionable reasons*

HellaSwag: Can a Machine Really Finish Your Sentence?

ACL 2019

What happened with SWAG? TLDR:
— dataset must be debiased (by algorithms)
— dataset must evolve (with the evolving SOTA)

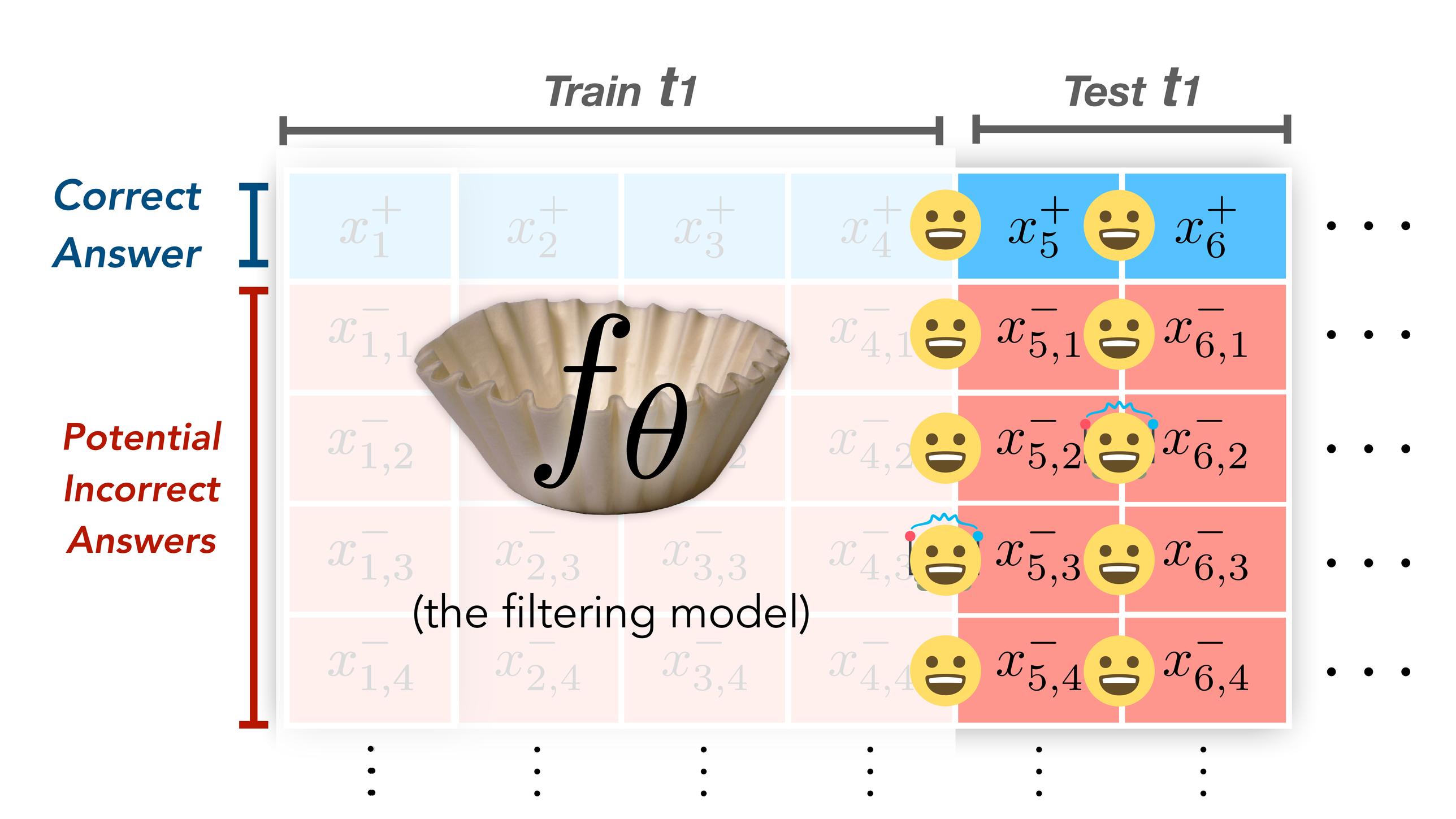
Rowan Zellers

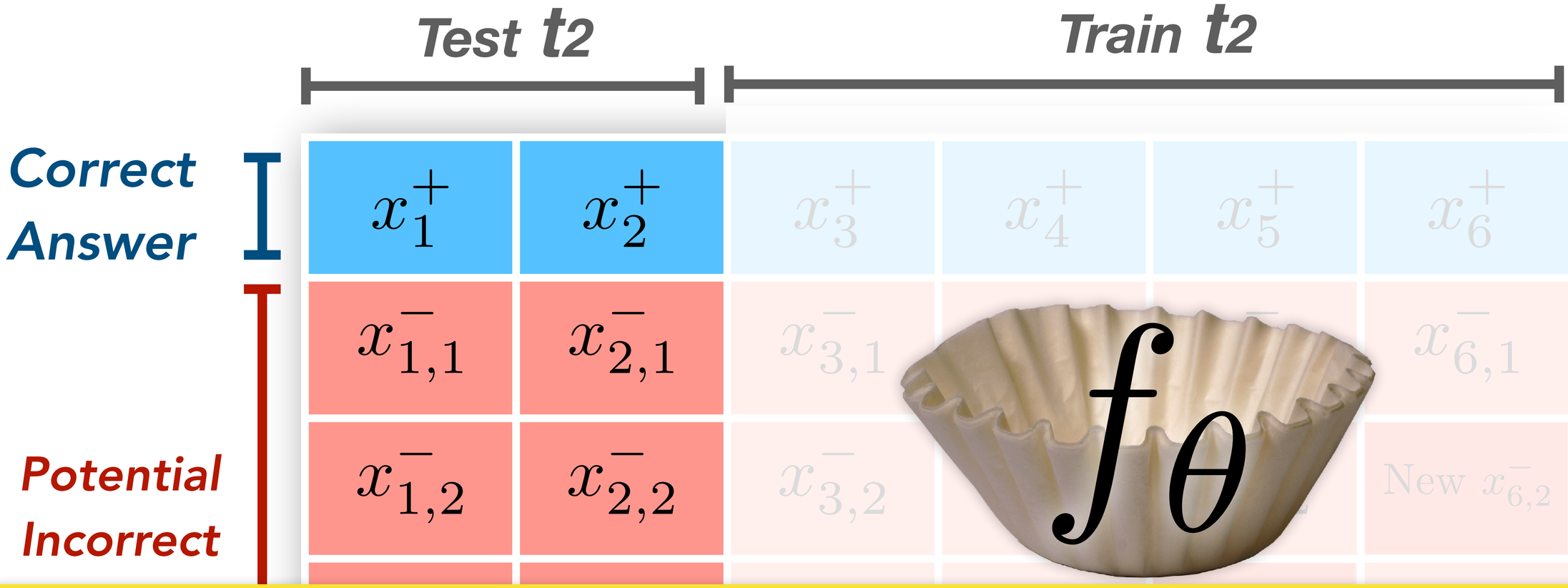
Ari Holtzman

Yonatan Bisk

Ali Farhadi

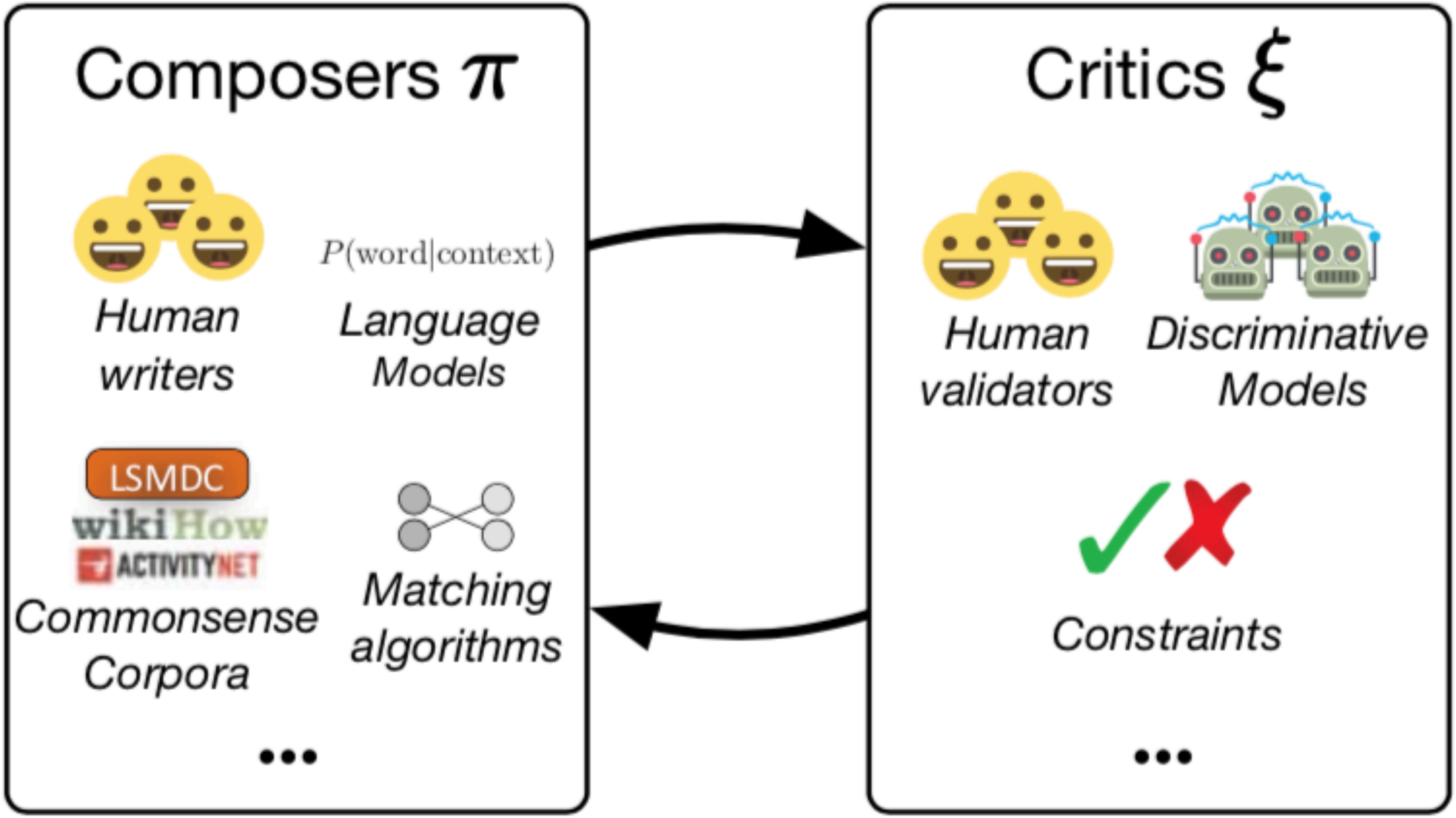
Yejin Choi


Train t_1
Test t_1
Correct Answer
Potential Incorrect Answers
x_1^+
x_2^+
x_3^+
x_4^+
x_5^+
x_6^+
$x_{1,1}^-$
$x_{4,1}^-$
$x_{5,1}^-$
$x_{6,1}^-$
$x_{1,2}^-$
$x_{4,2}^-$
$x_{5,2}^-$
$x_{6,2}^-$
$x_{1,3}^-$
$x_{2,3}^-$
$x_{3,3}^-$
$x_{4,3}^-$
$x_{5,3}^-$
$x_{6,3}^-$
$x_{1,4}^-$
$x_{2,4}^-$
$x_{3,4}^-$
$x_{4,4}^-$
$x_{5,4}^-$
$x_{6,4}^-$
$f\theta$
(the filtering model)

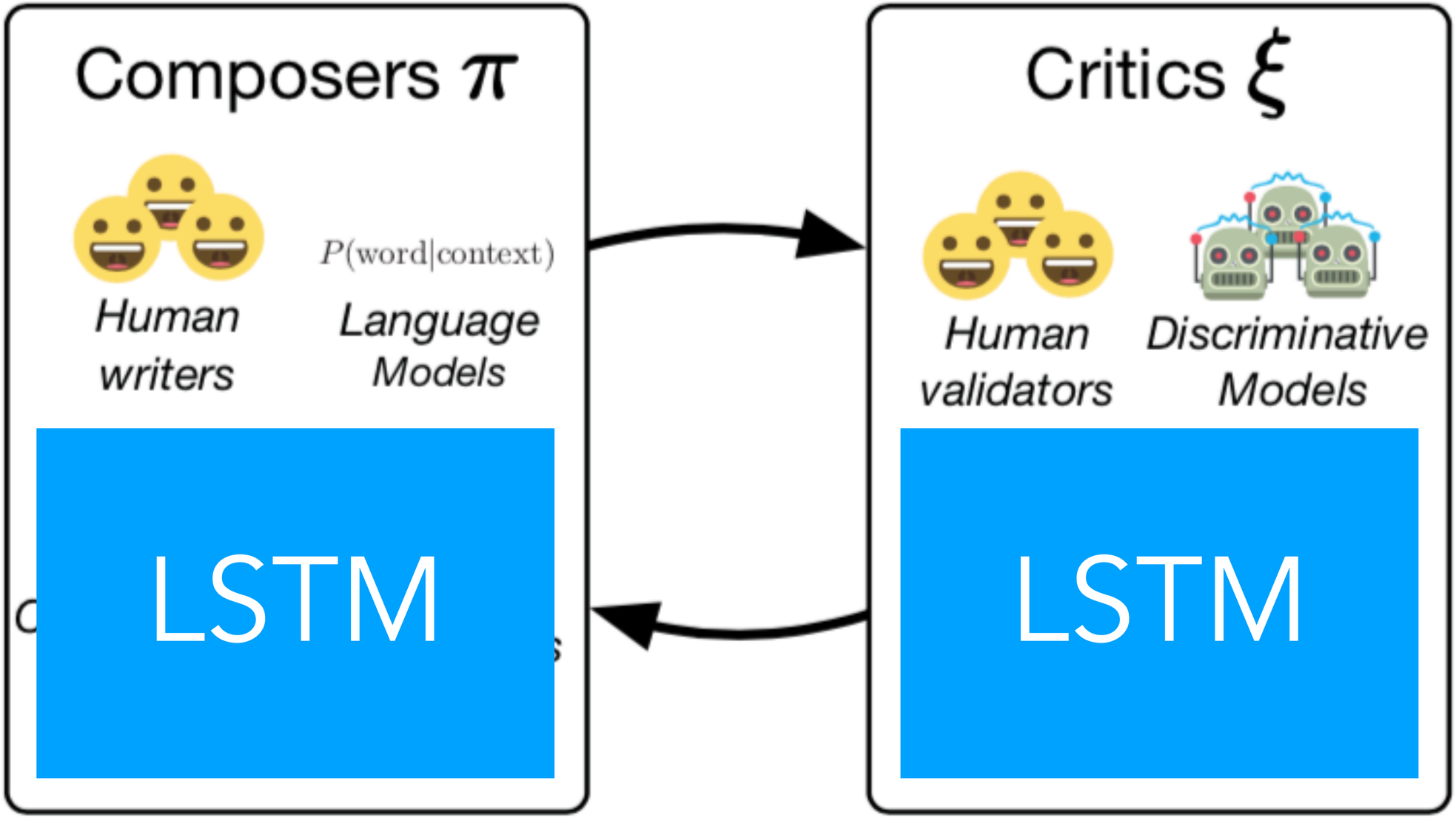


- We repeat this until convergence!
- The resulting AF'ed data holds identical distribution across training & test
- Human performance remains high on the AF'ed data, while AI performance drops considerably

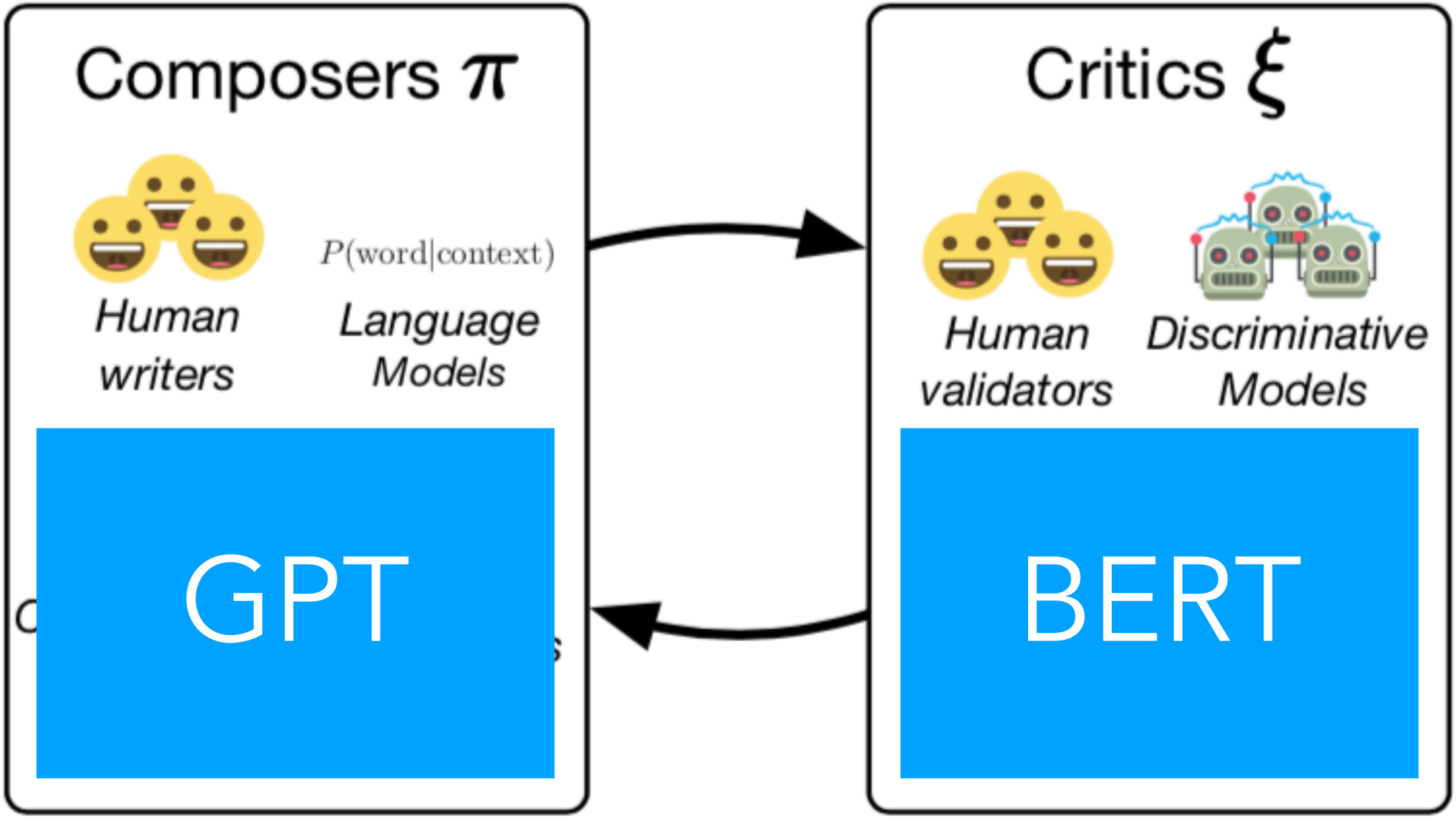
An Adversarial Evaluation Framework of



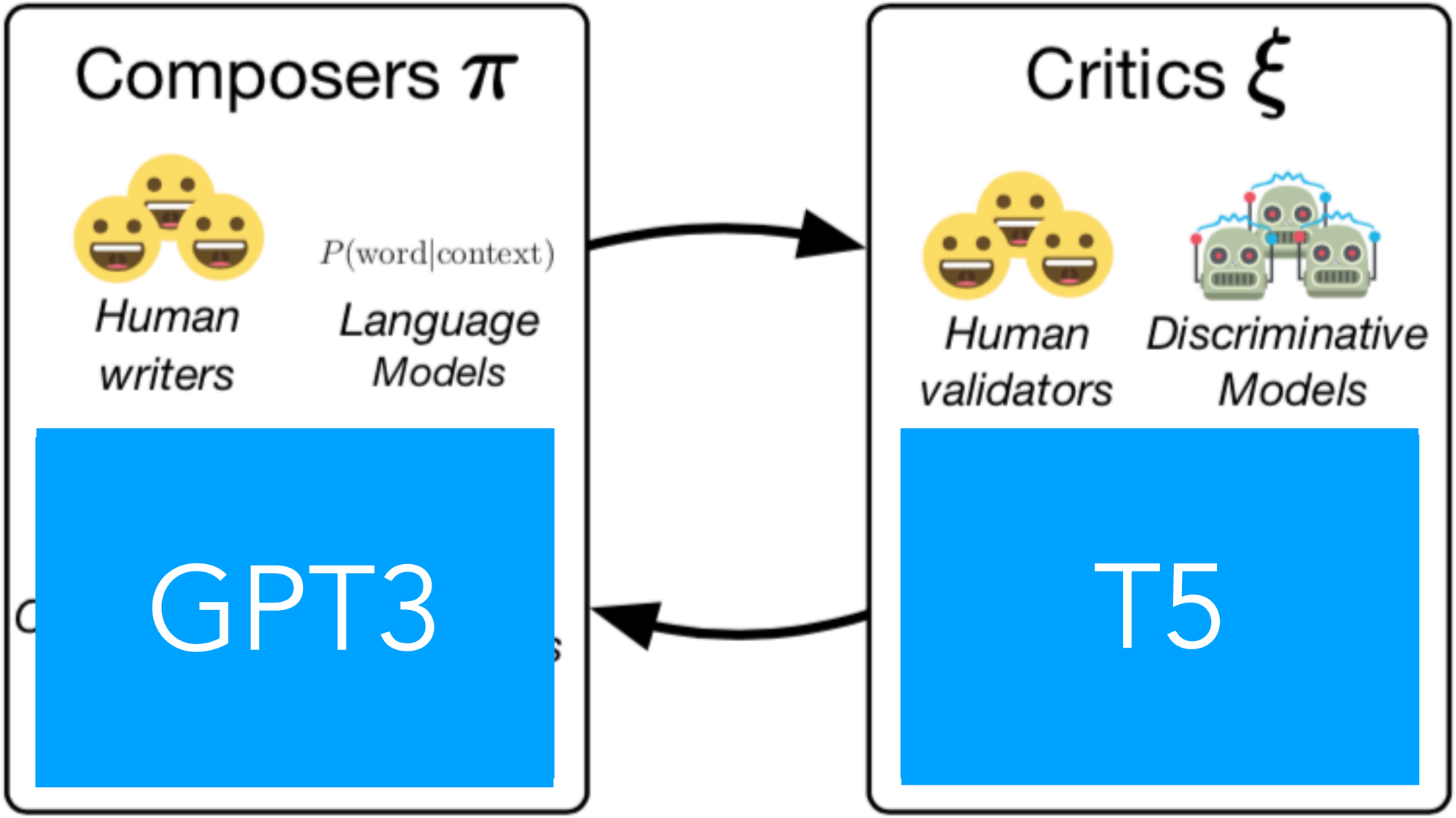
Pre GPT and BERT era (swag, 2018)



Post GPT and BERT era (hellaswag 2019)



Post GPT and BERT era (2020?)



How do we check if a model only solved a “dataset”
without solving an underlying task? Try adversarial evaluation!

nchmarks must evolve!

WinoGrande

Adversarial Winograd Schema Challenge at Scale

Outstanding Paper (Best Paper) Award at AAAI'20




Adversarial Filters of Dataset Biases

ICML 2020


Ronan LeBras


Swabha Swayamdipta


Chandra Bhagavatula


Rowan Zellers

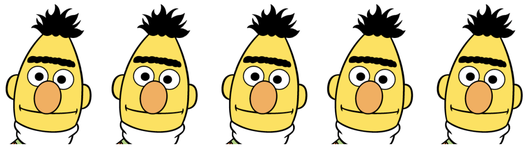

Matt Peters


Ashish Sabharwal


Yejin Choi

Original Flavor!

Original



- **SWAG** (Zellers et al., 2018)
- **HellaSWAG** (Zellers et al.,2019)

Af-Optimum Flavor!

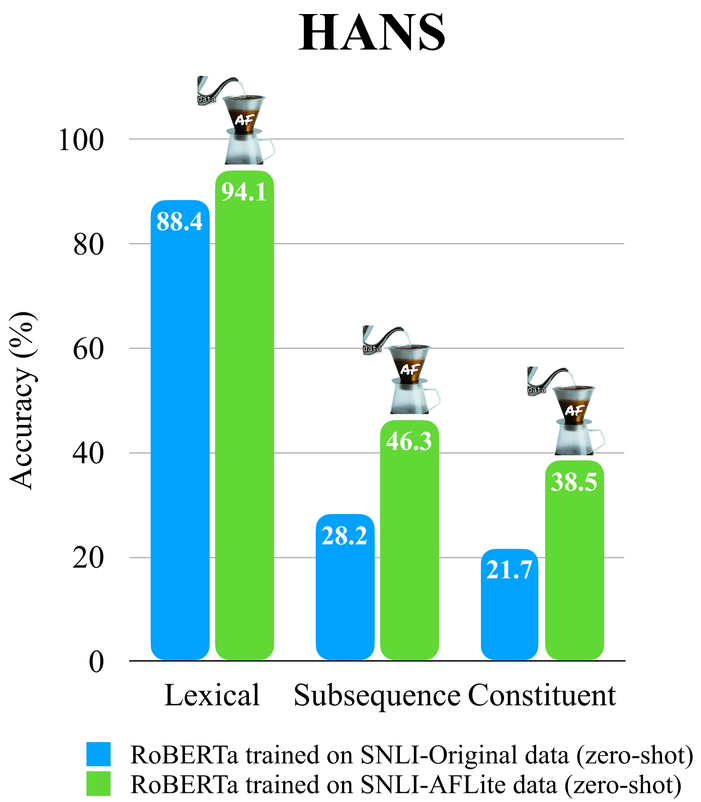
Optimum

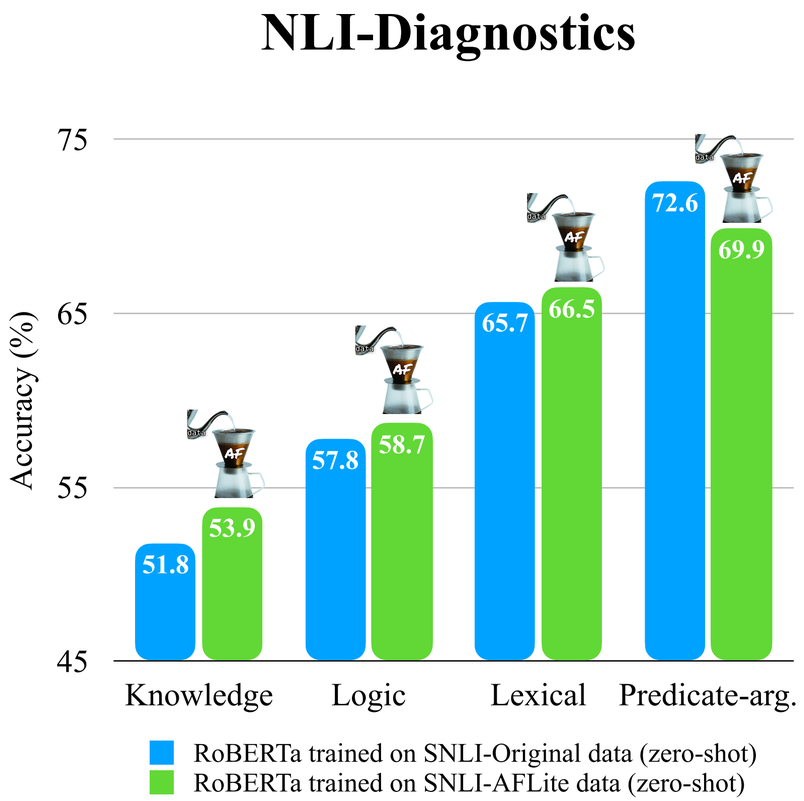
Af-Lite Flavor!

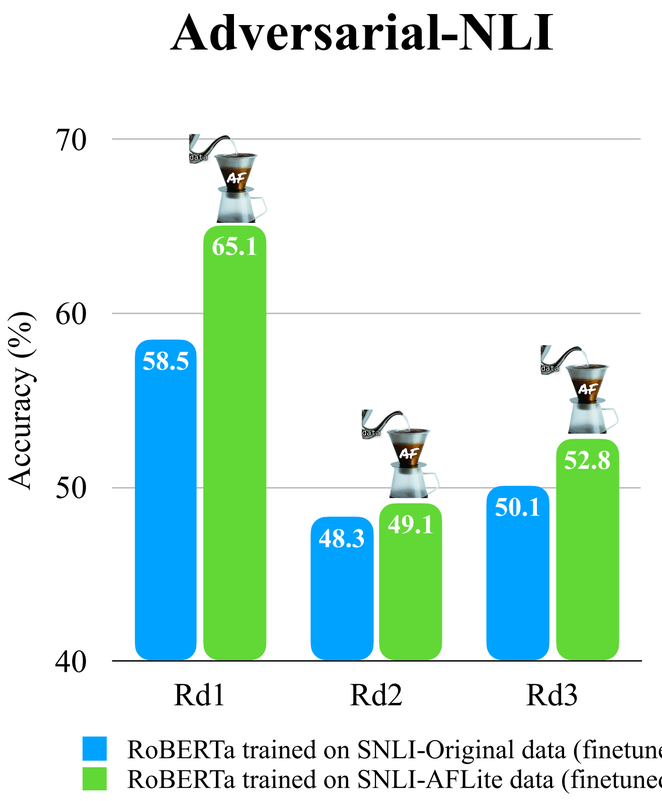
Af-Lite

- **Winogrande** (Sakaguchi et al., AAAI 2020)
- **Adversarial Filters of Dataset Biases**
  (Le Bras et al., ICML 2020)

Models Trained on AF'ed Data —Better on Out-of-Distribution



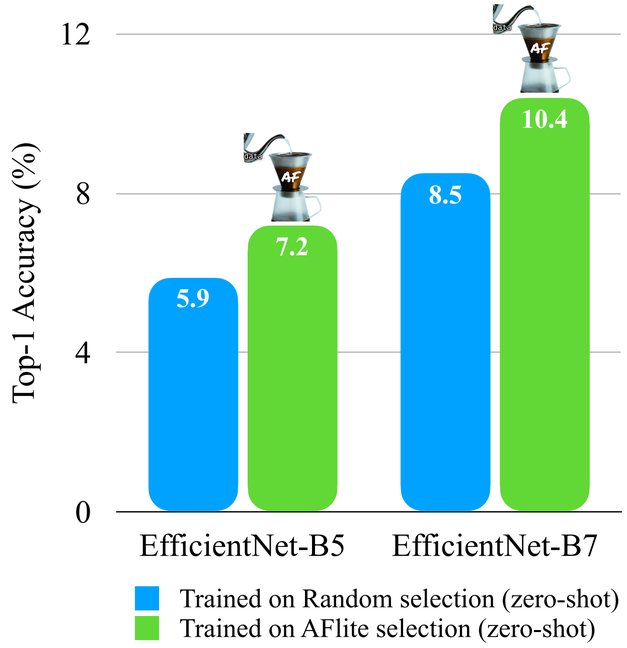




Models Trained on AF'ed Data —Better on Out-of-Distribution

ImageNet-A (Hendrycks et al., arXiv'19)





Current Paradigm of QA datasets

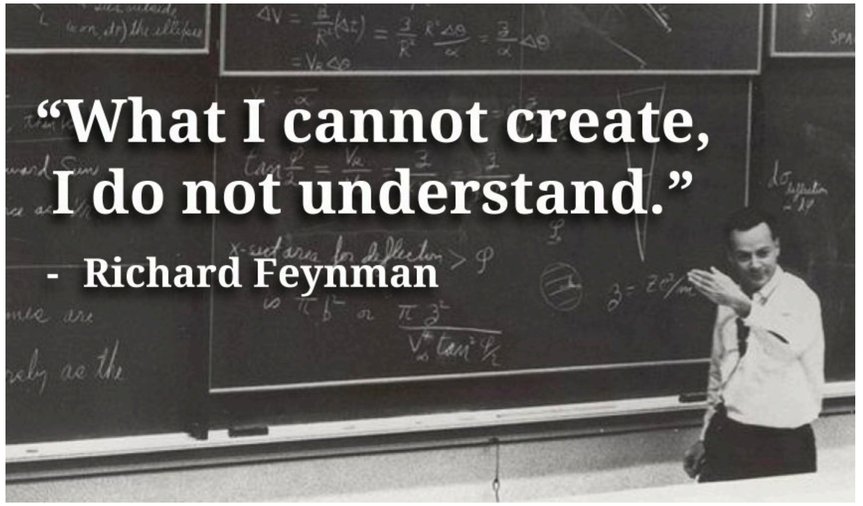
1. Neural networks immediately latch on spurious correlations & unwanted dataset biases

   - Multiple-choice QA datasets might require algorithmic debiasing to avoid overestimation of true AI capabilities

(Beery et al., 2018)

2. Can you imagine learning only by solving a lot of exam problems? — it's hard even for humans!

   - Need to evaluate machines generatively
   - Need to teach machines concepts more directly



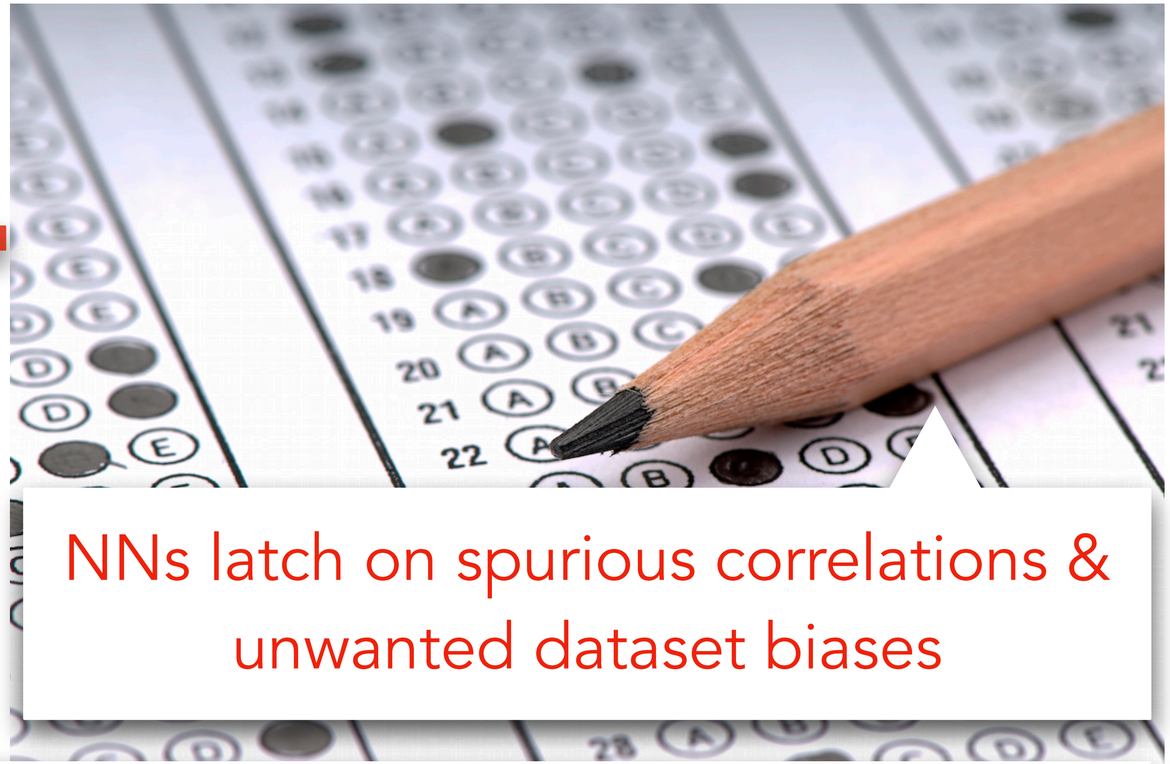
Imagine taking a deep learning class in which…

Self-supervision on a lot of DL code

```
        self.names = names
        self.name2index = dict(zip(names, range(len(names))))


    def __del__(self):
        # free memory created by C to avoid memory leak
        if hasattr(self, '__createfrom__') and self.__createfrom__ == 'C':
            if pointer(self) is not None:
                libbigfile.free_file(pointer(self))

    def read(self, requested, isname=True):
        if isname:
            index_name_array = [(self.name2index[x], x) for x in requested]
        else:
            assert(min(requested)>=0)
            assert(max(requested)<len(self.names))
            index_name_array = [(x, self.names[x]) for x in requested]
        index_name_array.sort()

        npoints = len(index_name_array)
        c_index = (c_ulonglong * npoints)()
        for i in range(npoints):
            c_index[i] = index_name_array[i][0]

        size = self.ndims * npoints
        pdata = (c_float * size)()
        res = libbigfile.seq_read_memory(self, npoints, c_index, pdata)
        assert(res)
```

+

Supervision on a lot of exam problems



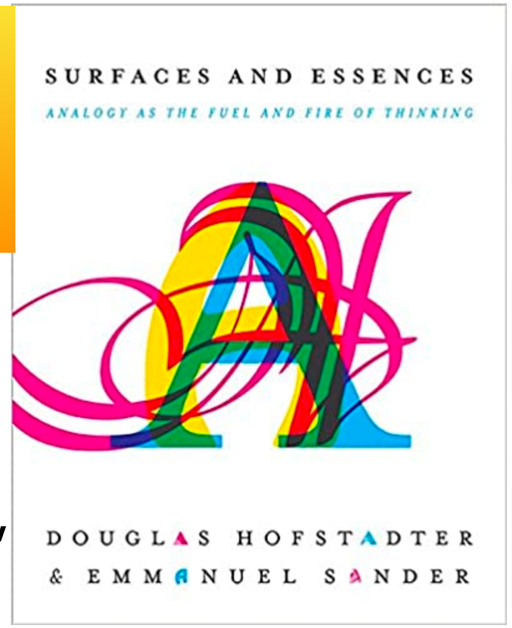
NNs latch on spurious correlations & unwanted dataset biases

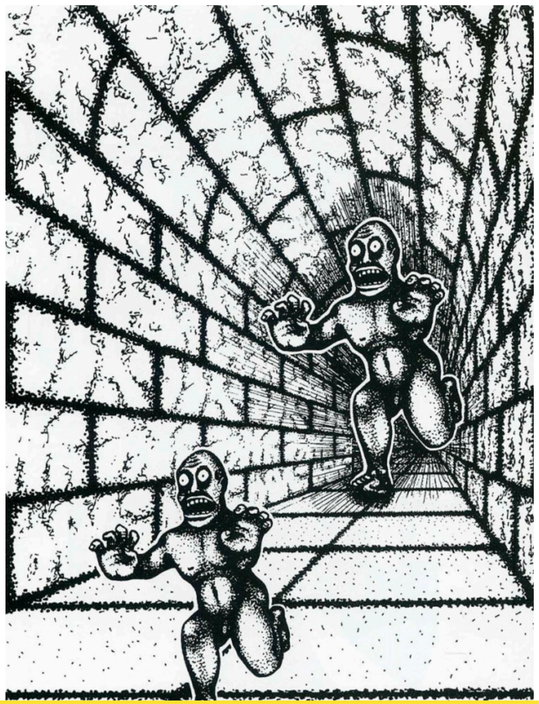
Can’t learn concepts well enough; need to learn from declarative knowledge

Tldr; 🔥 Language and Symbols 🔥 🔥 Reasoning as Generation🔥

1. ***Language*** as the ***symbols***
   - all of it —
     - not just ***words*** (ImageNet labels, …)
     - not just small ***sets of words*** (scene graphs,
2. Reasoning as ***generative tasks***
   - As opposed to ***discriminative tasks*** (i.e., categorization)
   - Because the space of reasoning in language is ***infinite***

"Categories (concepts) vastly outnumber words, and require free-form open text descriptions"





"thinking out loud"

We often think as we speak, on the fly, word-by-word without enumerating all possible alternative sentences

Commonsense AI: Closing Remarks & Open Research Questions

Commonsense

- Searching "commonsense" from ACL anthology
  - Most papers are either from 80s or from the past few years

Position Paper on Common-sense and Formal Semantics

Geoffrey Nunberg
Xerox PARC and CSLI, Stanford

1. A philological excursus

I'm not sure what I'm doing on this panel, but I thought it would be helpful if we could start at the beginning. It's interesting to note that both the dictionary and common sense were eighteenth-century inventions. This is no coincidence; in fact, it's entirely appropriate that the most celebrated

Revisiting Commonsense

I was told not to speak the word commonsense...

Past failures (in 70s – 80s) are inconclusive

-- weak computing power

-- not much data

-- no crowdsourcing

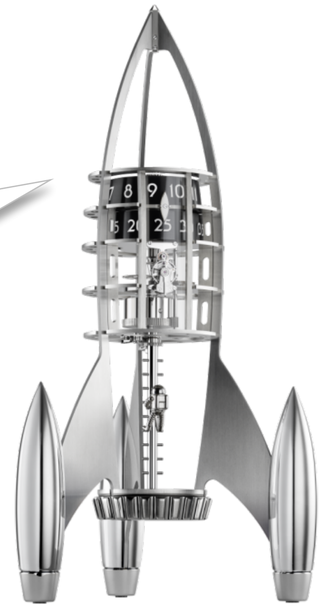
-- not as strong computational models

-- not ideal conceptualization / representations

Path to commonsense?

Brute force larger networks with deeper layers?

You don’t reach to the moon
by making the tallest building in the world
taller



Thoughts on Language & Embodiment

- My dog (my cat, my baby…) has common sense without language. Therefore, AI doesn't need language for common sense
  - AI for (ultimately) humans, not for (just) dogs
- Some concepts (e.g., that juicy crunch feel you get when biting into an apple) can't be learned without embodiment. Thus, can't learn concepts without embodied experiences
  - We learn that a tiger can eat a human without experiencing it in real life

Thoughts on Language & Embodiment

- We might never have a robot that can bite into an apple.

- Nor might we ever have a simulation environment in which AI can experience what it is like for humans to bite on an apple.

- Trade-offs between **the coverage of concepts** and **the richness of multimodal / embodied experiences**
  - 3D environments, real or simulation, allow for richer experiences with a narrow slice of concepts
  - Language, images, videos allow for a significantly broader range of concepts with impoverished embodied experiences
  - Language provides a powerful representation medium to learn humans's embodied experiences as declarative knowledge

In this talk: Reasoning as Generation

- **Part 1:** unsupervised inference-time algorithms

Reasoning thru Neural Backpropagation	Reasoning thru Search with Logical Constraints	Reasoning thru Distributional Neural Imagination
DeLorean	NeuroLogic	Reflective Decoding

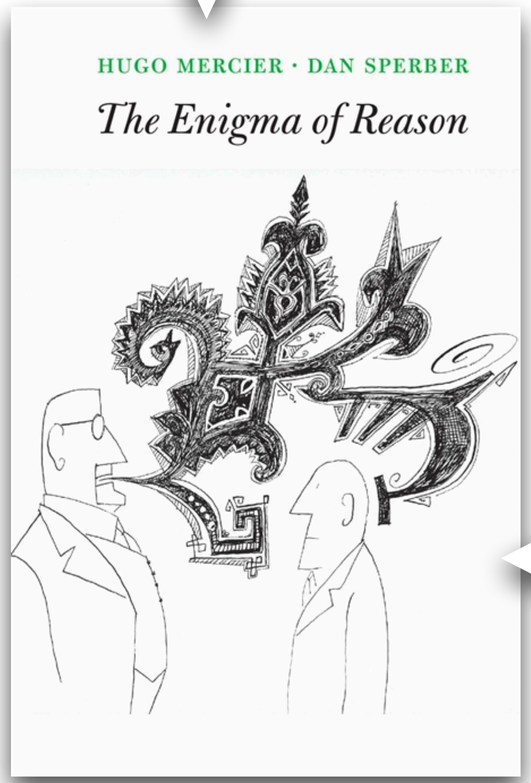
- **Part 2:** supervision with declarative knowledge for knowledge modeling

Neural & Symbolic Commonsense Knowledge	Visually Grounded Commonsense Knowledge	Social, Ethical, Moral Norms
COMET & ATOMIC 2020	Visual COMET	Social Chemistry 101

- **Part 3:** benchmarks and algorithmic bias reduction

Remarks on Language & Reason

Intuitive inferences is a great deal about extraction of **new information** from the information already available.

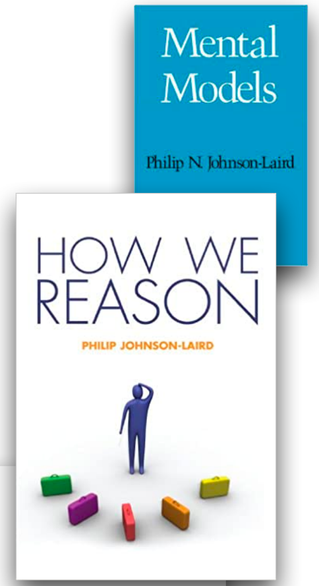


INTUITION
SYSTEM 1

REASONING
SYSTEM 2

Reasons are used primarily
not to guide oneself
but to justify oneself
in the eyes of others,
and to convince others

Reasoning serves the purpose of
communication



Human reason is
a mechanism of intuitive inferences

… in which
logic plays at best a marginal role.

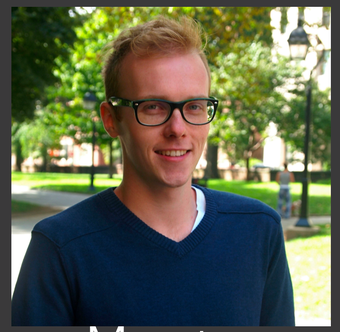
ACL 2020 Commonsense Tutorial

— 2nd most popular (among 8 tutorials) —

https://homes.cs.washington.edu/~msap/acl2020-commonsense/

1288 registrations for our tutorial
2819 view counts on our recorded lectures
(Total registration of the main conference = 4972)

Vered Shwartz



Maarten Sap

Antoine Bosselut

Dan Roth

Yejin Choi

Thanks! Questions?