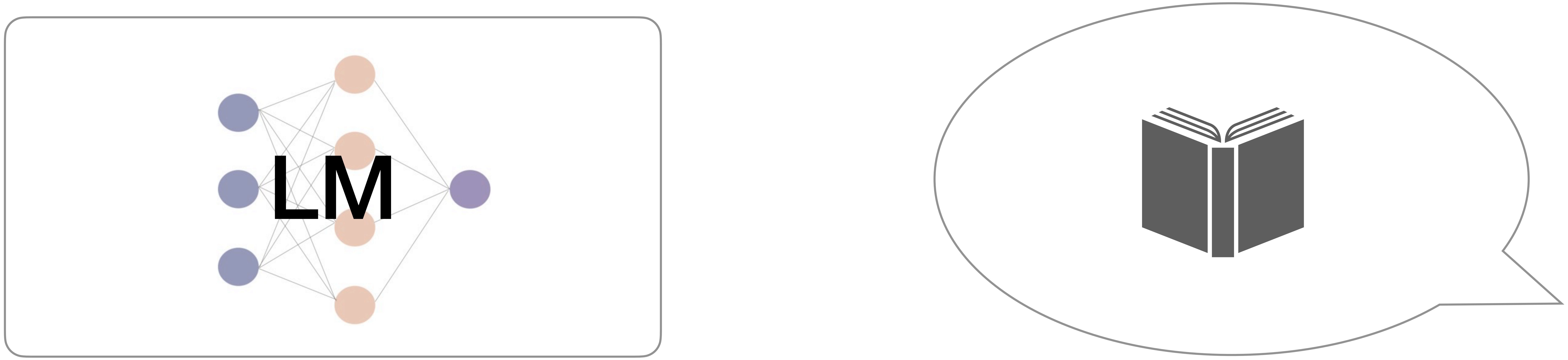
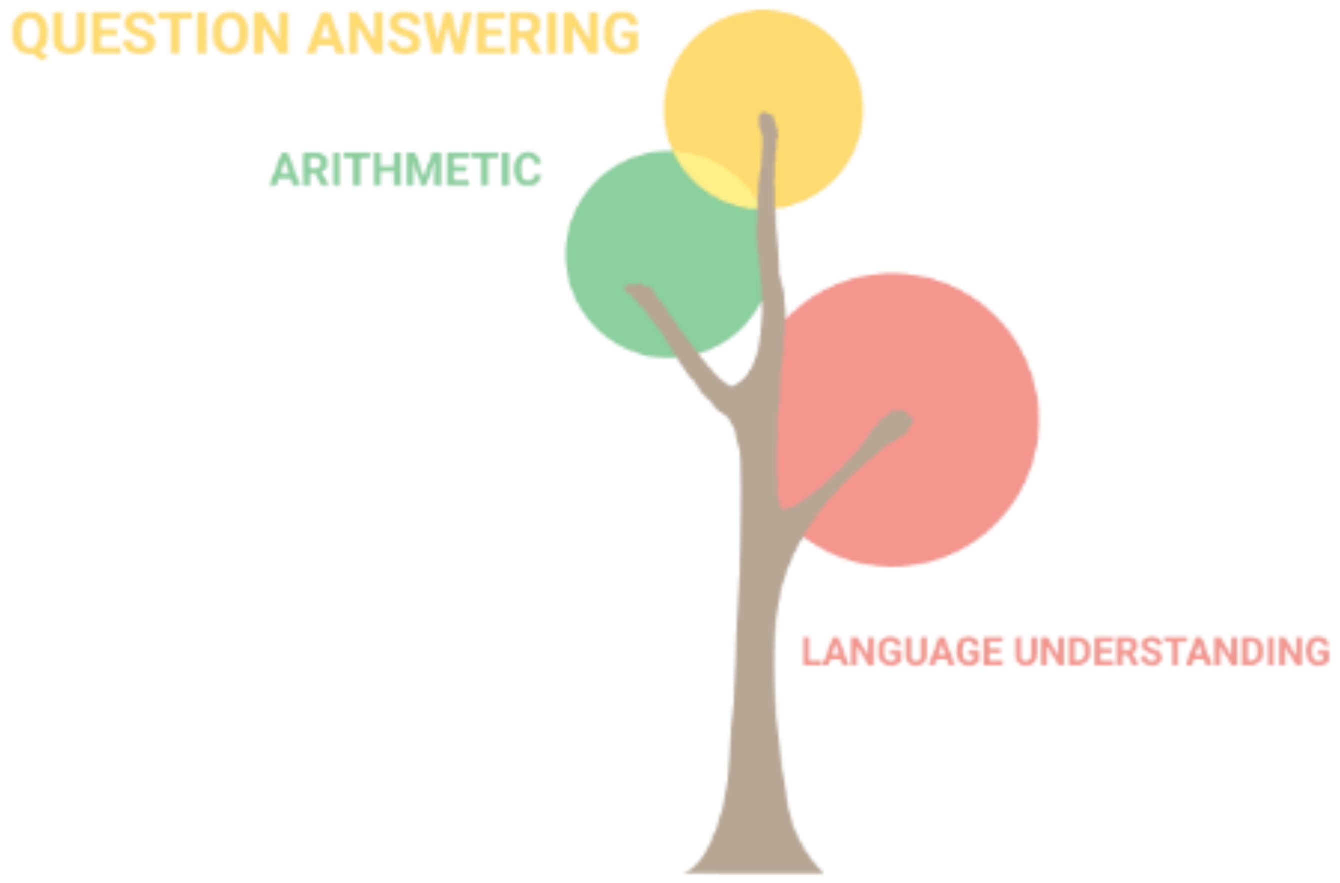
Structured Knowledge in Language Models

Knowledge Base Construction from Pre-Trained Language Models

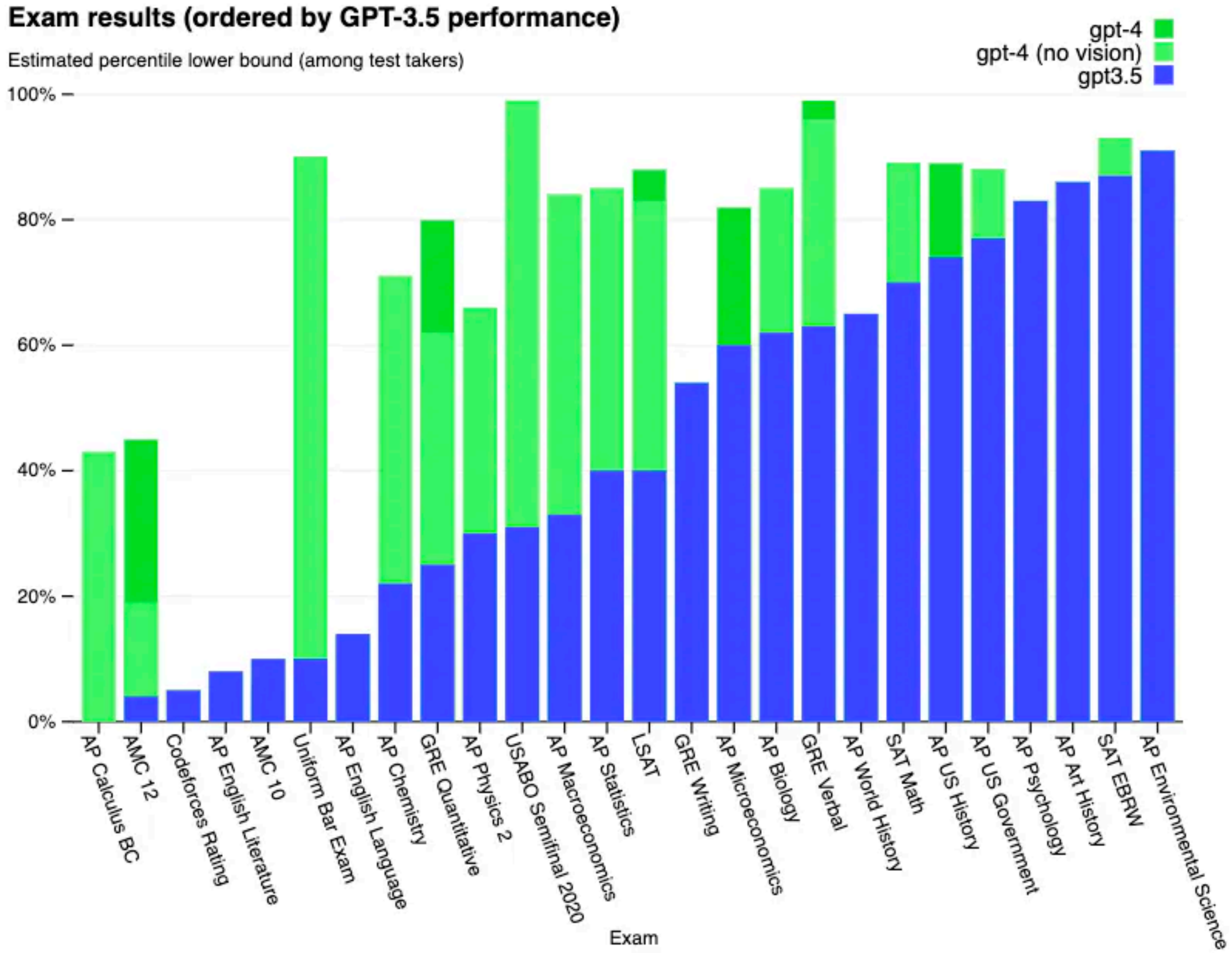
Nora Kassner, November 6th 2023



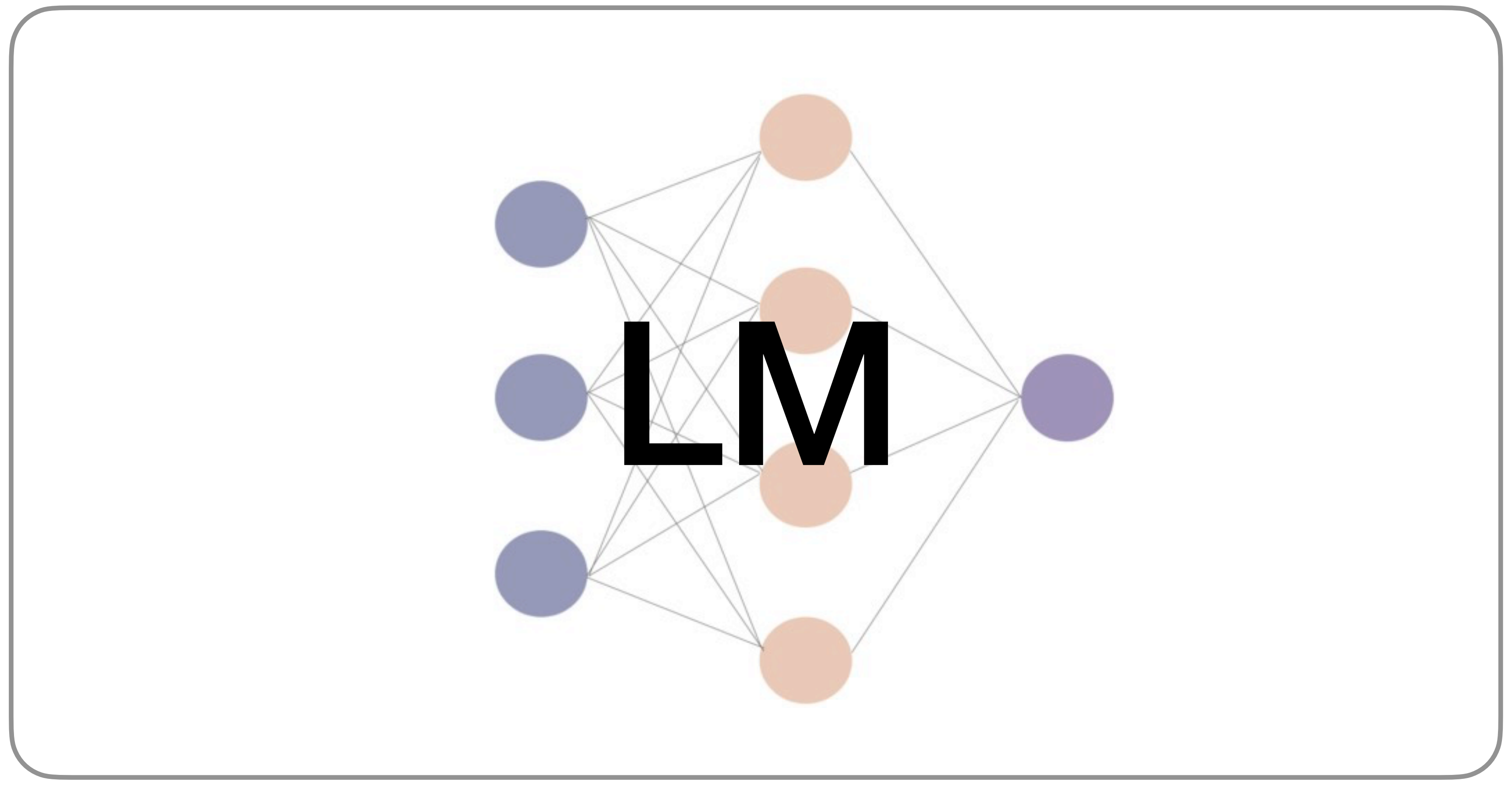
LM

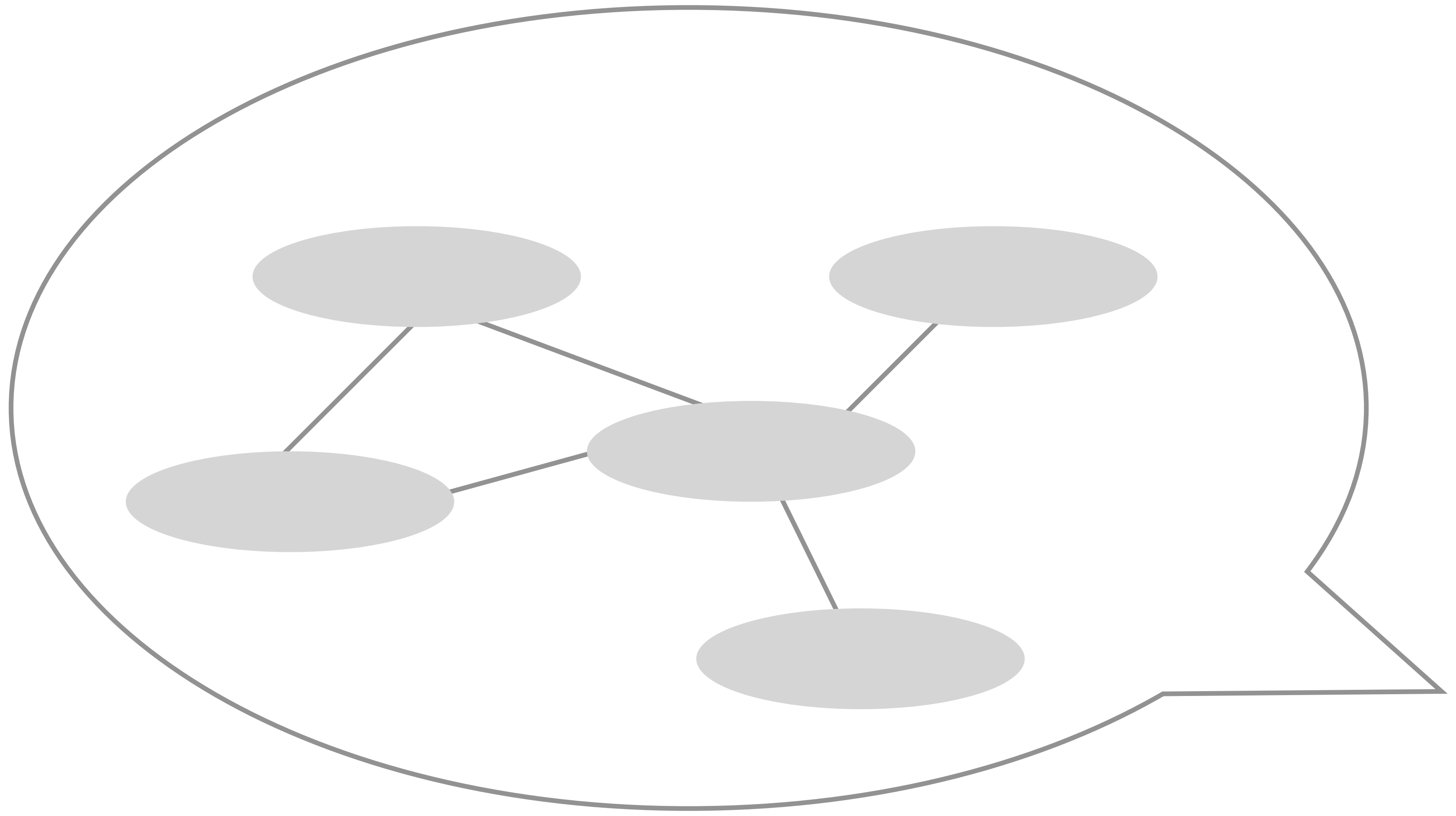


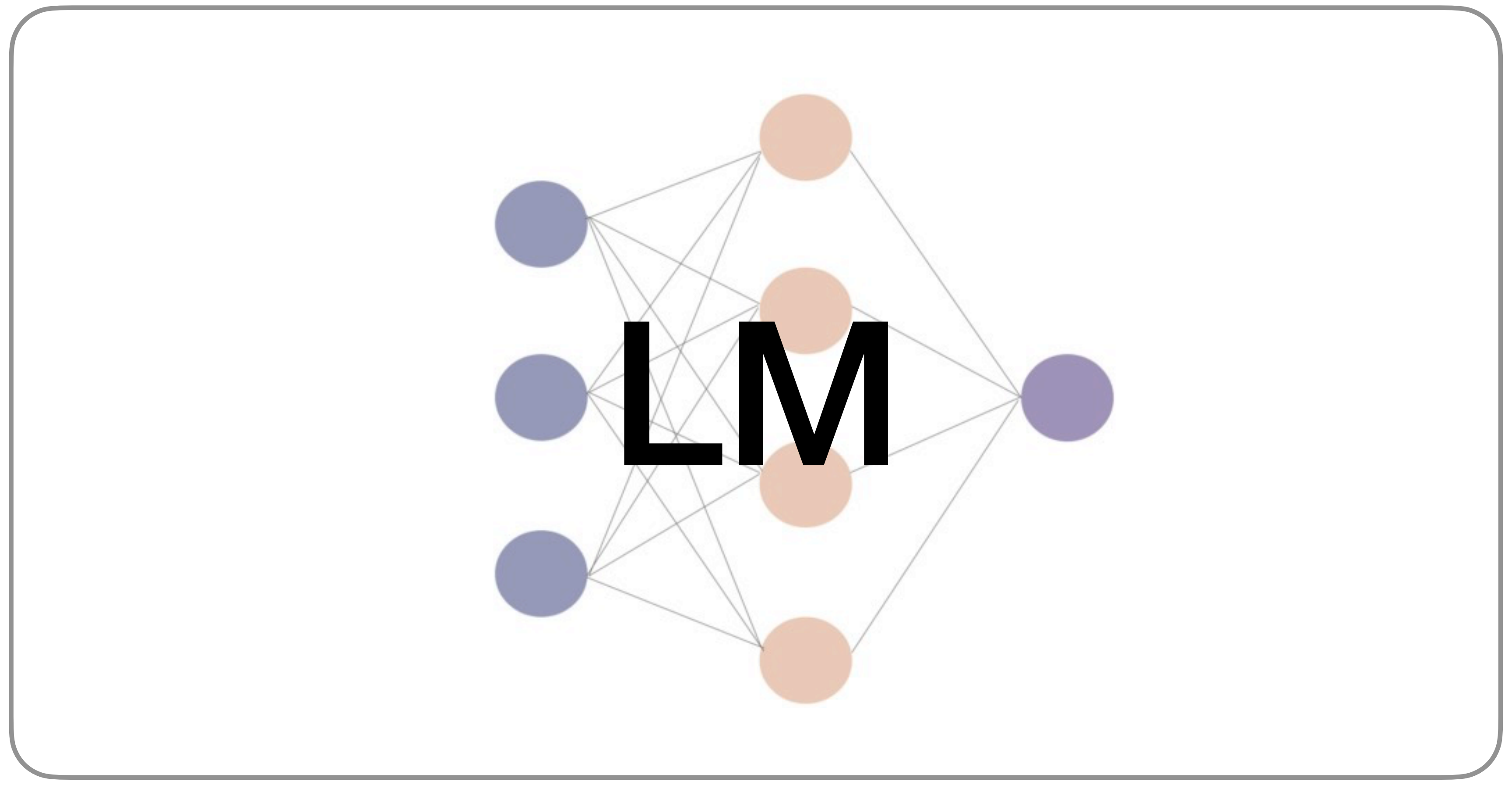
8 billion parameters

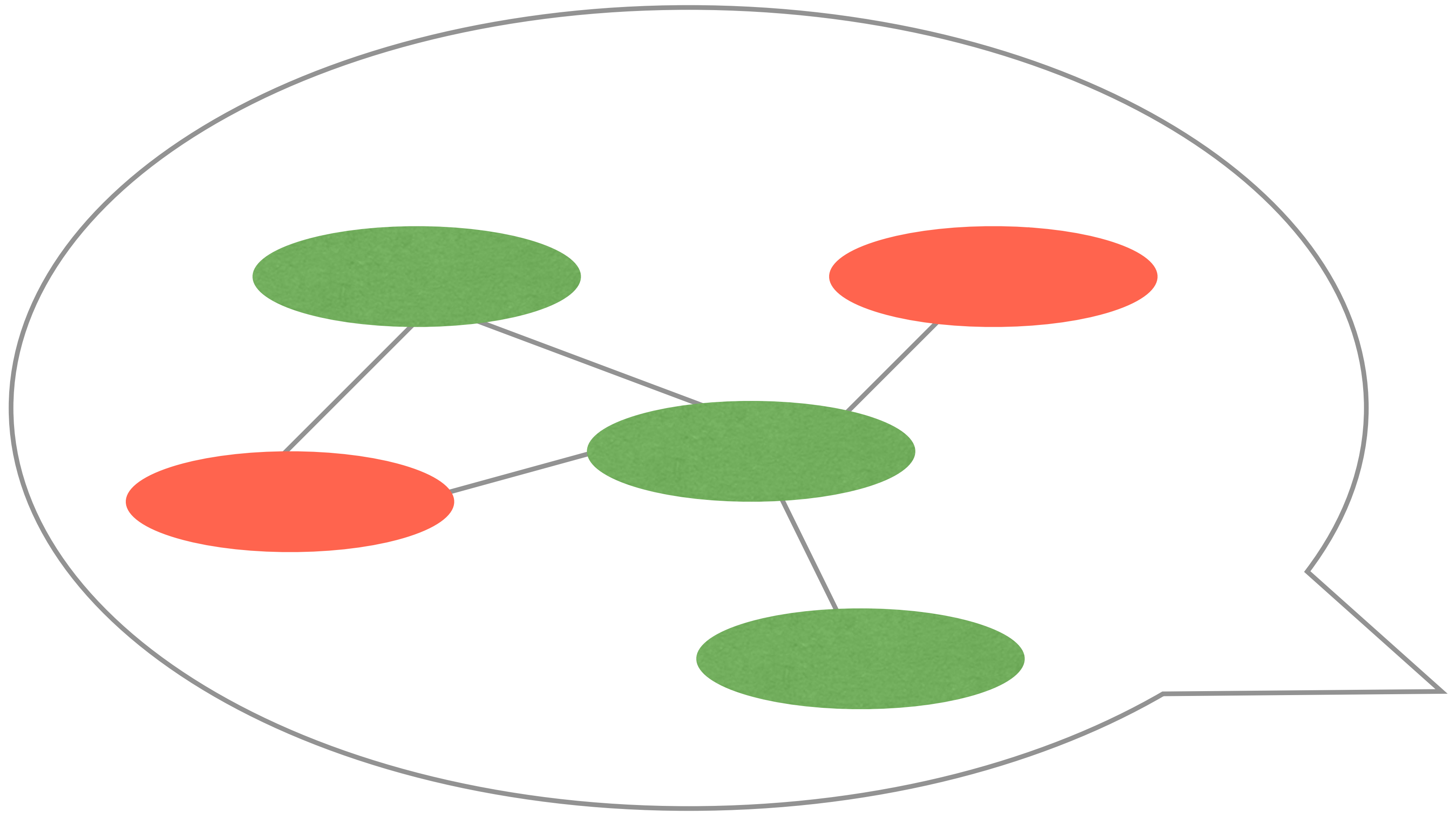


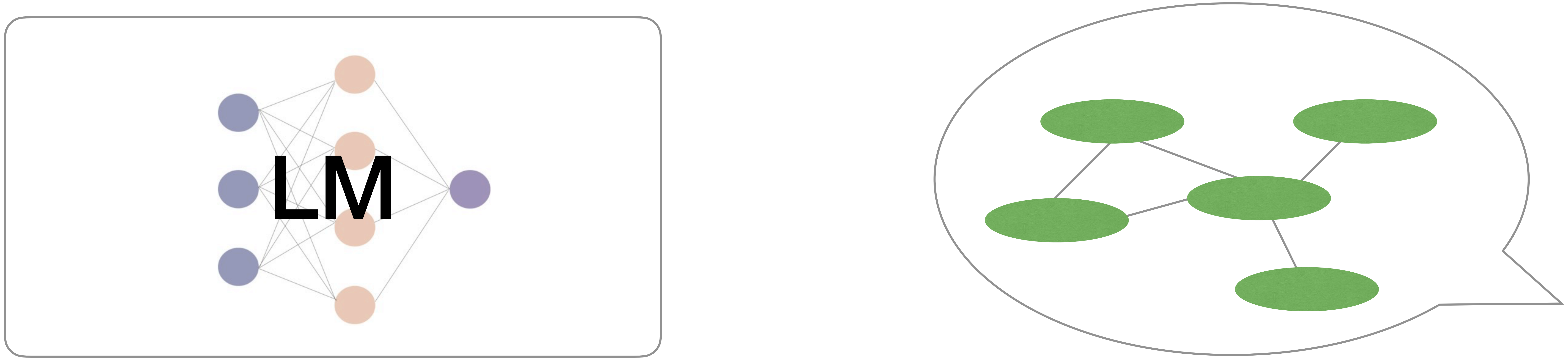
→ Open AI: GPT-4 Technical Report, March 2023


LM




LM




LM

Outline

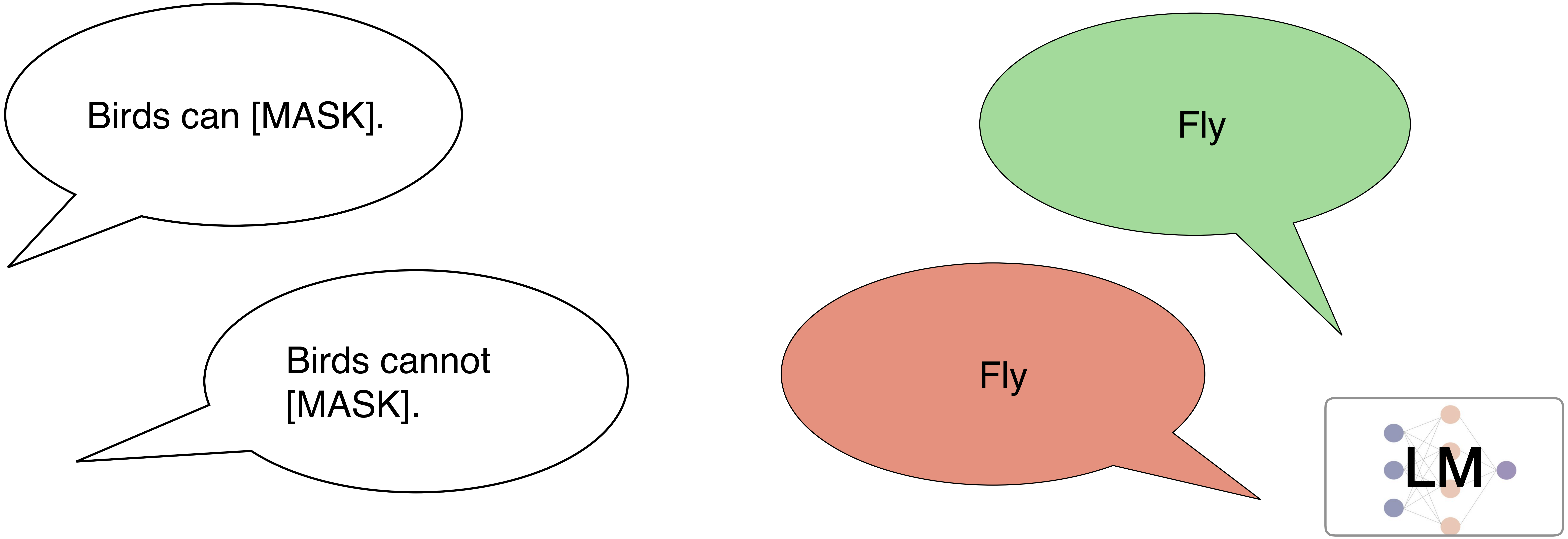
Three types of consistency sets:

- Negation
- Multilinguality
- Reasoning Chains

Towards constructing structured world models:

- BeliefBank
- REFLEX

Consistency with Respect to Negation



Birds can [MASK].

Birds cannot [MASK].

Fly

Fly

LM

→ Kassner et al.: Negated and misprimed probes for pretrained language models: Birds can talk, but cannot fly, ACL 2020

Consistency with Respect to Negation

Data		%
Google-RE	birth-place	20.1
	birth-date	0.3
	death-place	13.2
T-REX	1-1	22.7
	N-1	45.0
	N-M	54.2
ConceptNet	-	31.3
SQuAD	-	41.9

% = Mean percent of overlap in first ranked predictions

→ LMs are prone to generate facts and their incorrect negation

→ Kassner et al.: Negated and misprimed probes for pretrained language models: Birds can talk, but cannot fly, ACL 2020

Consistency with Respect to Negation

Data		%
Google-RE	birth-place	20.1
	birth-date	0.3
	death-place	13.2
T-REX	1-1	22.7
	N-1	45.0
	N-M	54.2
ConceptNet	-	31.3
SQuAD	-	41.9

% = Mean percent of overlap in first ranked predictions

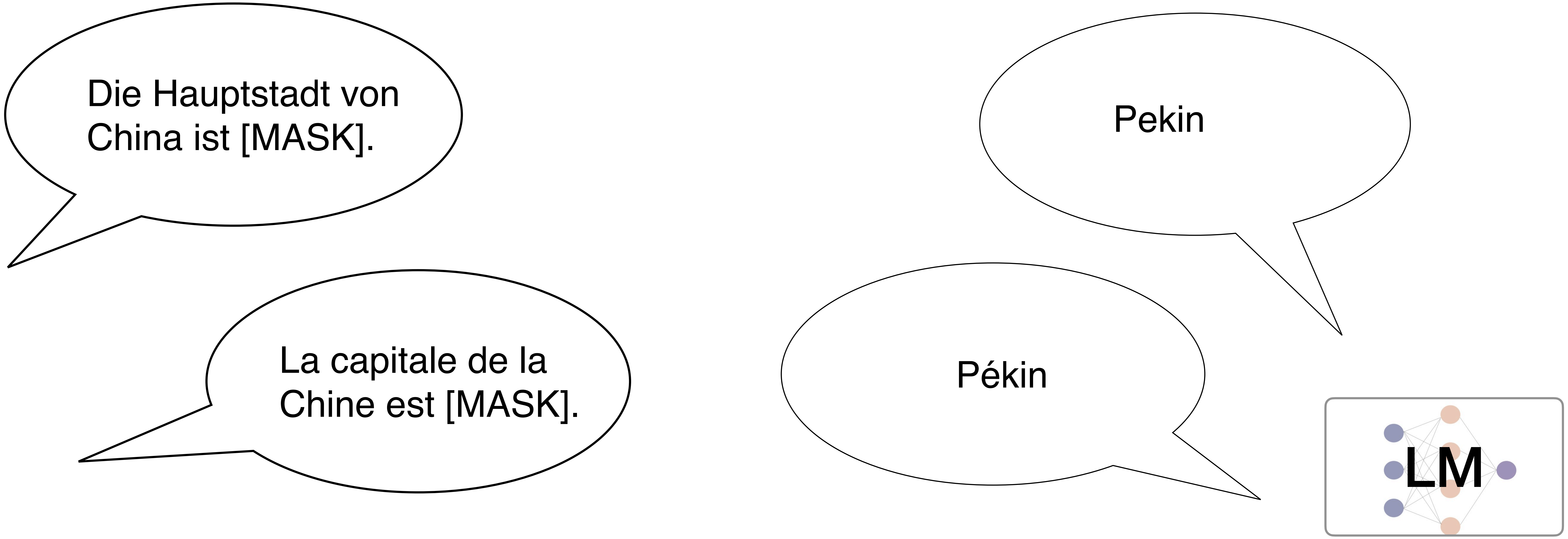
→ LMs are prone to generate facts and their incorrect negation

Enormous progress but still not solved:

Truong et al. Language models are not naysayers: An analysis of language models on negation benchmarks, June 2023

→ Kassner et al.: Negated and misprimed probes for pretrained language models: Birds can talk, but cannot fly, ACL 2020

Consistency with Respect to Multilinguality



Die Hauptstadt von China ist [MASK].

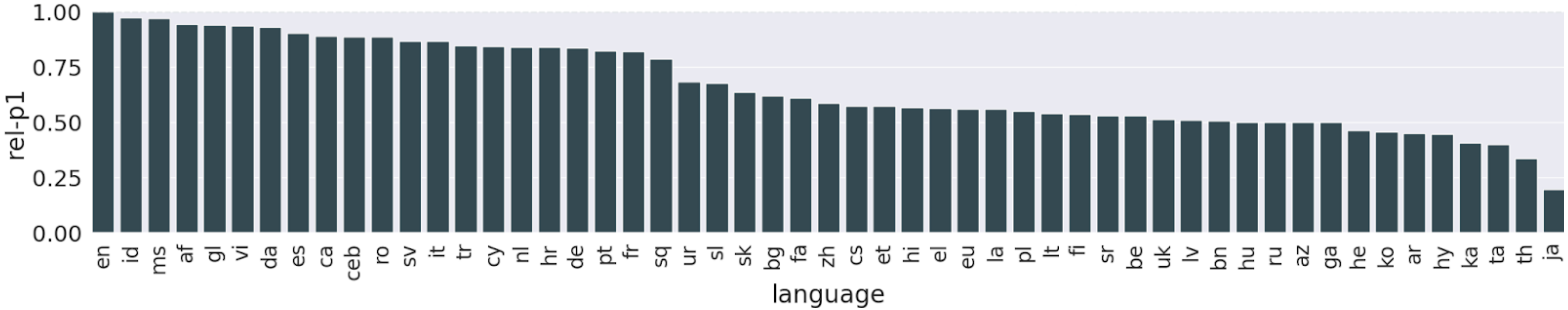
La capitale de la Chine est [MASK].

Pekin

Pékin

→ Kassner et al.: Multilingual LAMA: Investigating knowledge in multilingual Pretrained Language Models, EACL 2021

Consistency with Respect to Multilinguality

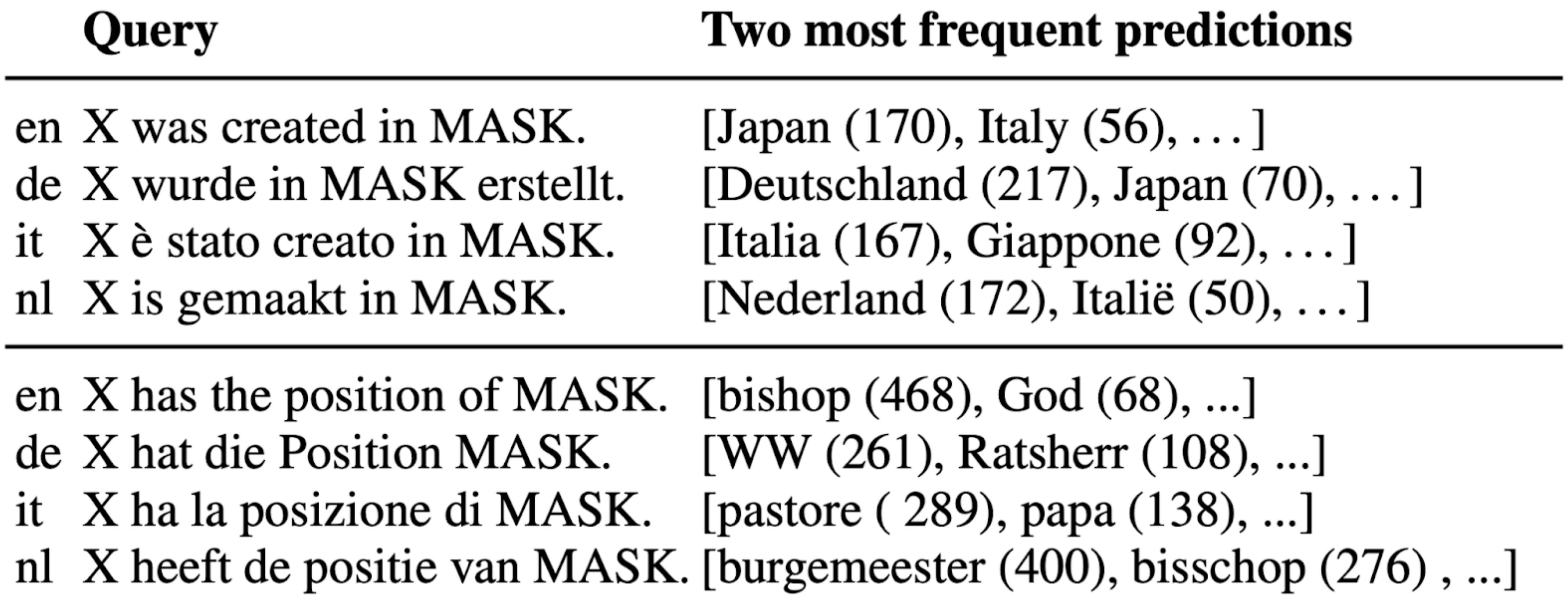


Accuracy for [language] / accuracy for [en]

→ mBert does not exhibit stable performance across languages

→ Kassner et al.: Multilingual LAMA: Investigating knowledge in multilingual Pretrained Language Models, EACL 2021

Consistency with Respect to Multilinguality

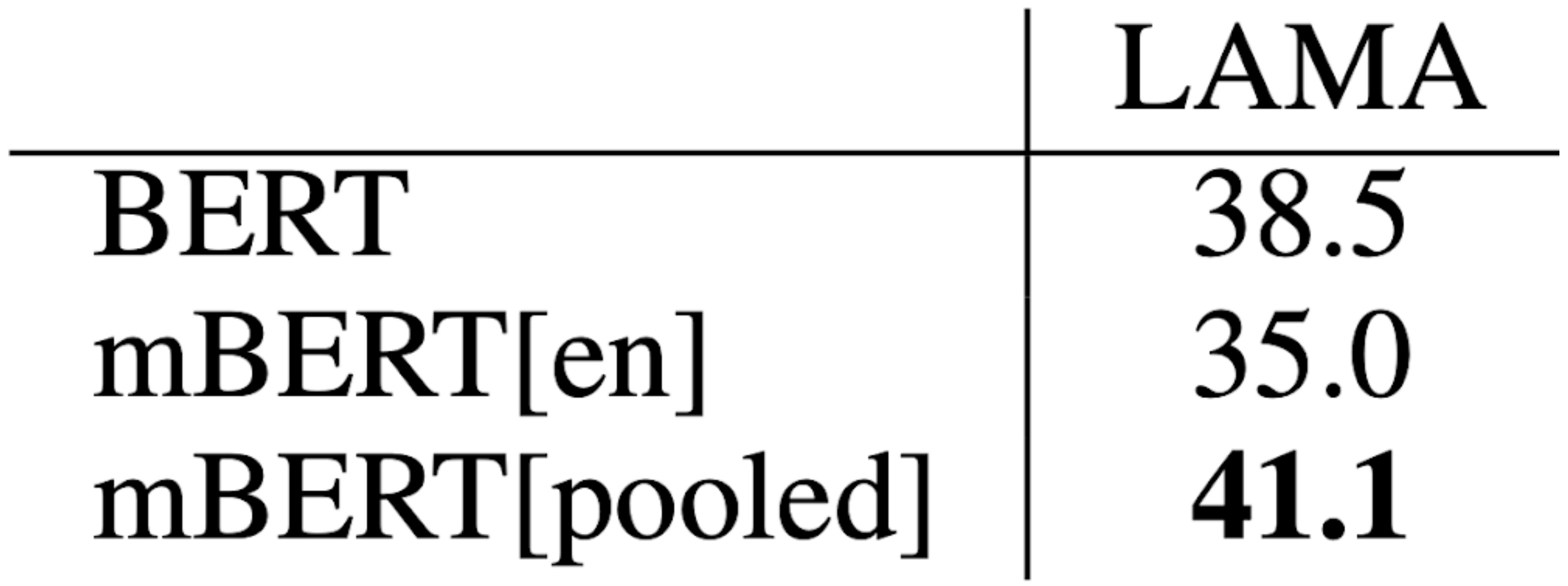


	Query	Two most frequent predictions
en	X was created in MASK.	[Japan (170), Italy (56), . . .]
de	X wurde in MASK erstellt.	[Deutschland (217), Japan (70), . . .]
it	X è stato creato in MASK.	[Italia (167), Giappone (92), . . .]
nl	X is gemaakt in MASK.	[Nederland (172), Italië (50), . . .]
en	X has the position of MASK.	[bishop (468), God (68), ...]
de	X hat die Position MASK.	[WW (261), Ratsherr (108), ...]
it	X ha la posizione di MASK.	[pastore (289), papa (138), ...]
nl	X heeft de positie van MASK.	[burgemeester (400), bisschop (276) , ...]

→ Query language affects predictions

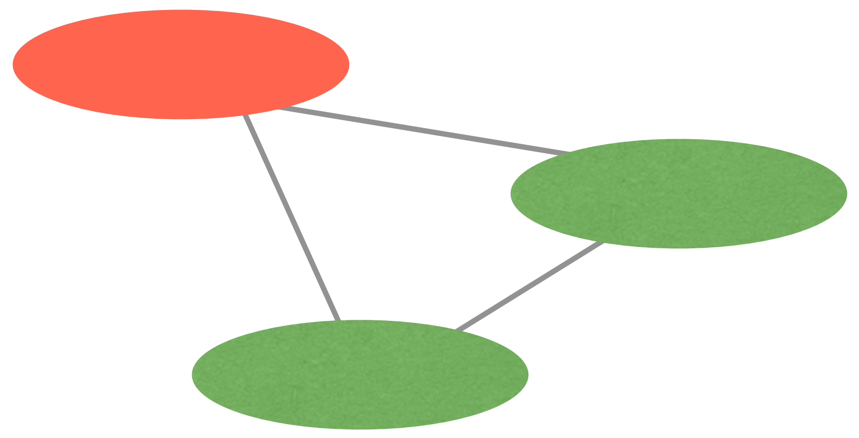
→ Kassner et al.: Multilingual LAMA: Investigating knowledge in multilingual Pretrained Language Models, EACL 2021

Consistency with Respect to Multilinguality



	LAMA
BERT	38.5
mBERT[en]	35.0
mBERT[pooled]	**41.1**

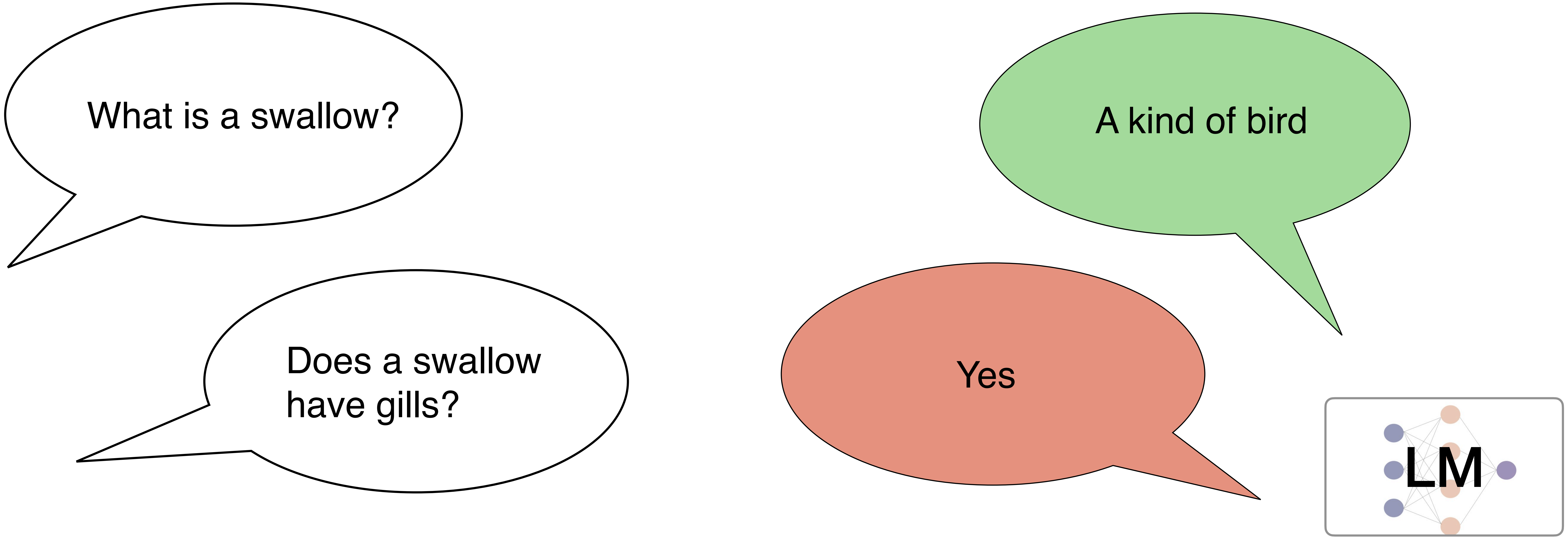
Accuracy



→ Pooling predictions across languages yields performance improvements

→ Kassner et al.: Multilingual LAMA: Investigating knowledge in multilingual Pretrained Language Models, EACL 2021

Consistency with Respect to Cains of Reasoning



What is a swallow?

A kind of bird

Does a swallow have gills?

Yes

→ Kassner et al.: BeliefBank: Adding Memory to a Pre-trained Language Model for a Systematic Notion of Belief, EMNLP 2021

Towards more consistent knowledge: BeliefBank

1. Positive Implications T → T:

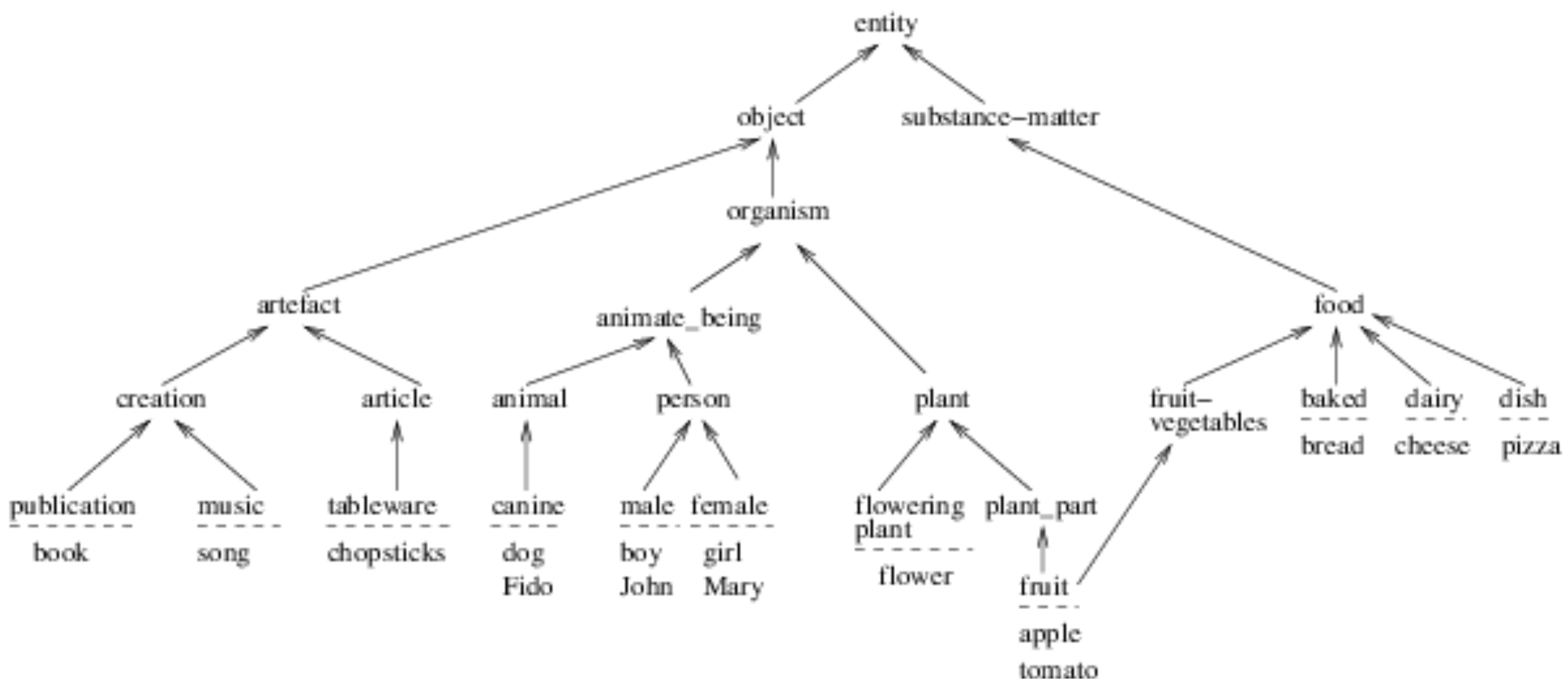
“X is a dog.” T → “X has a tail.” T

2. Mutual Exclusivities T → F:

“X is a bird.” T → “X is a fish.” F

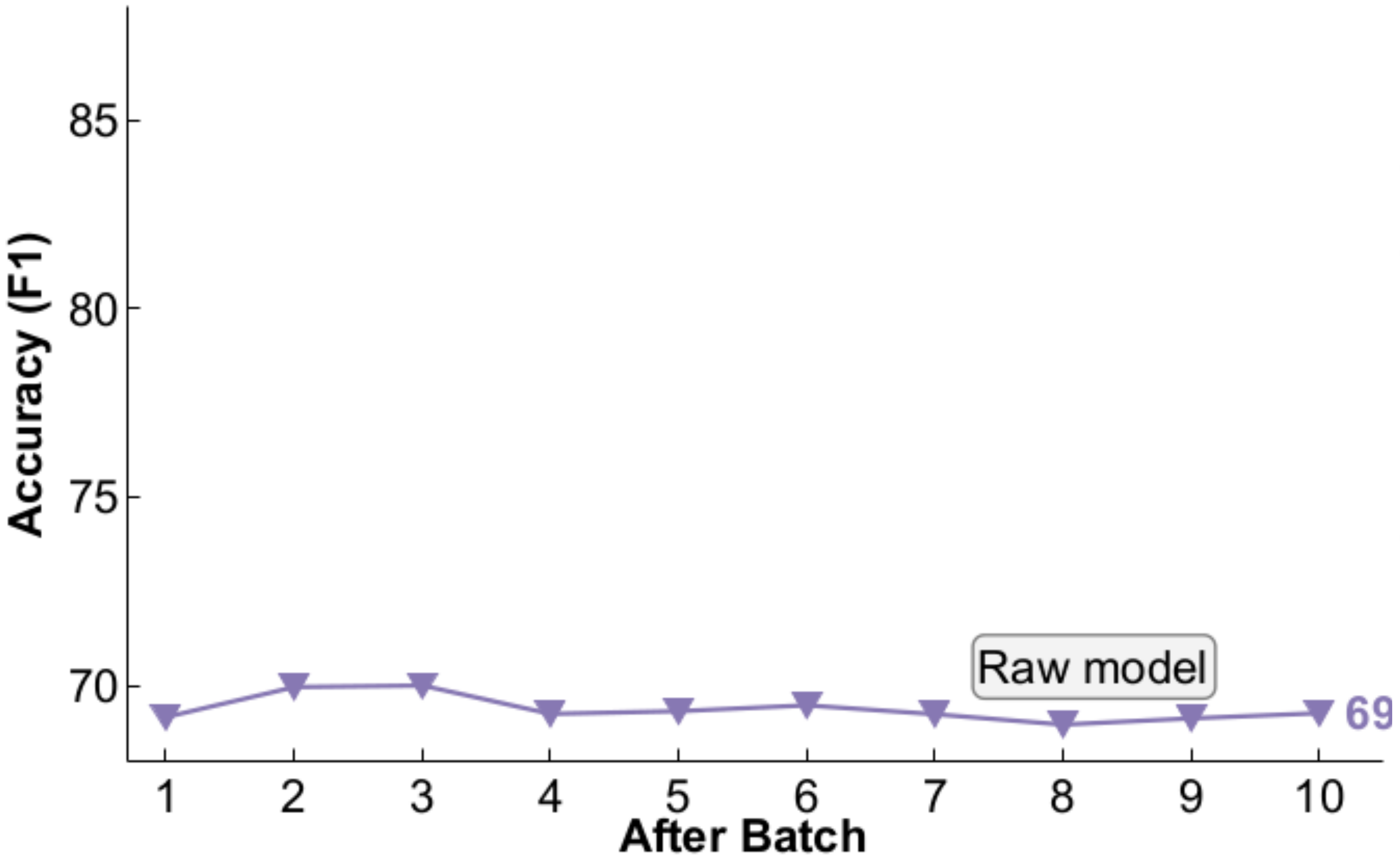
ConceptNet

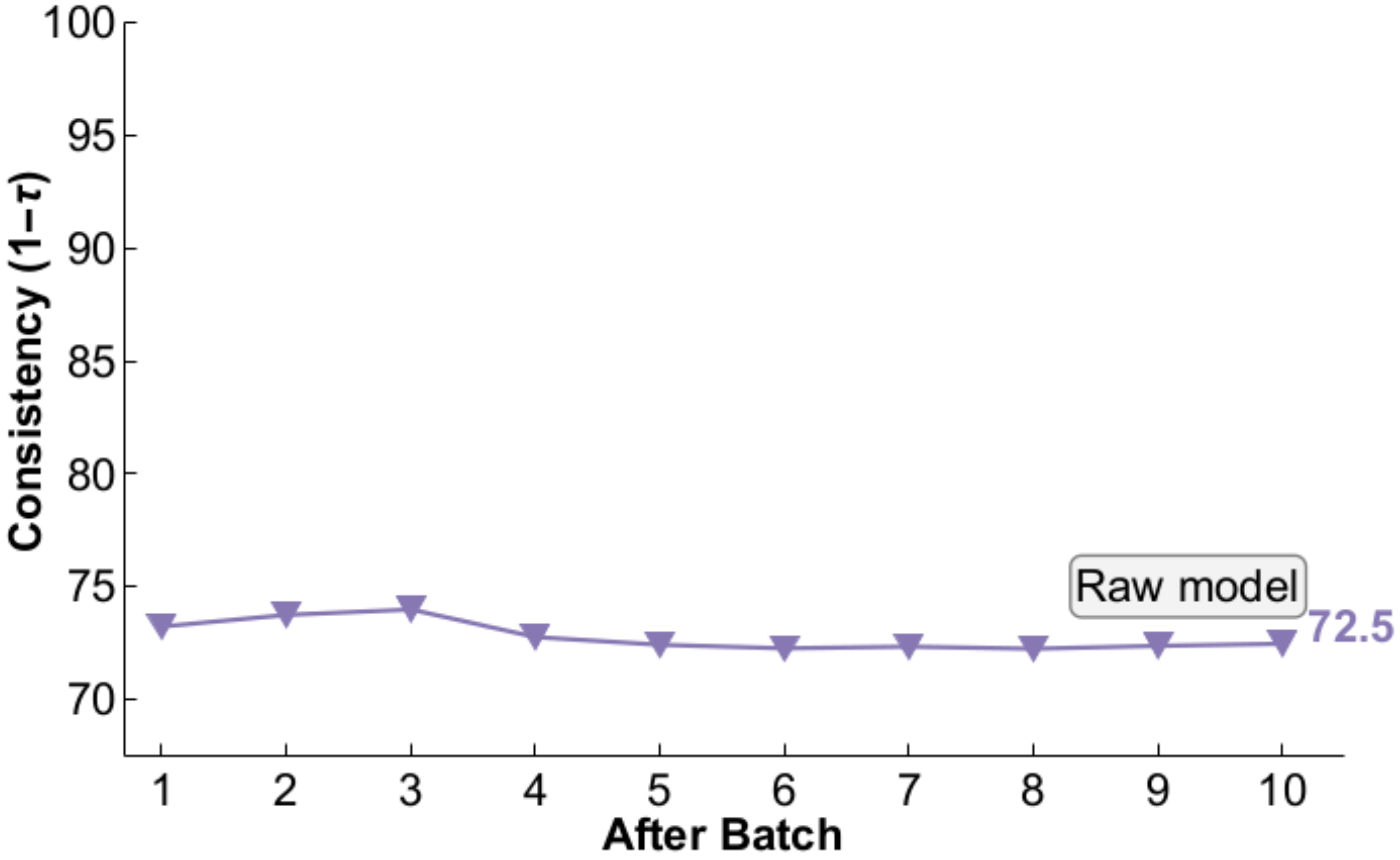
An open, multilingual knowledge graph



→ Kassner et al.: BeliefBank: Adding Memory to a Pre-trained Language Model for a Systematic Notion of Belief, EMNLP 2021

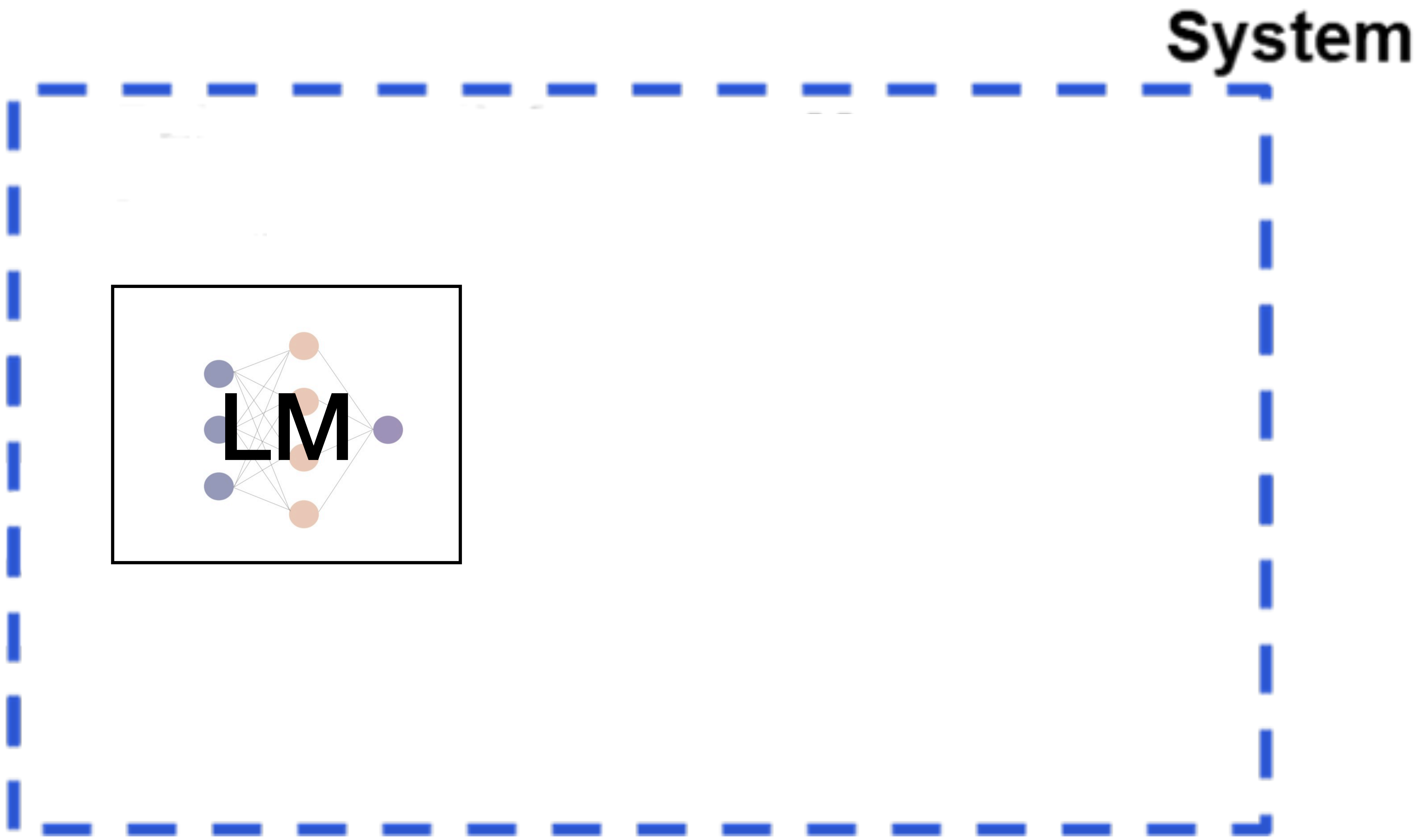
Towards more consistent knowledge: BeliefBank





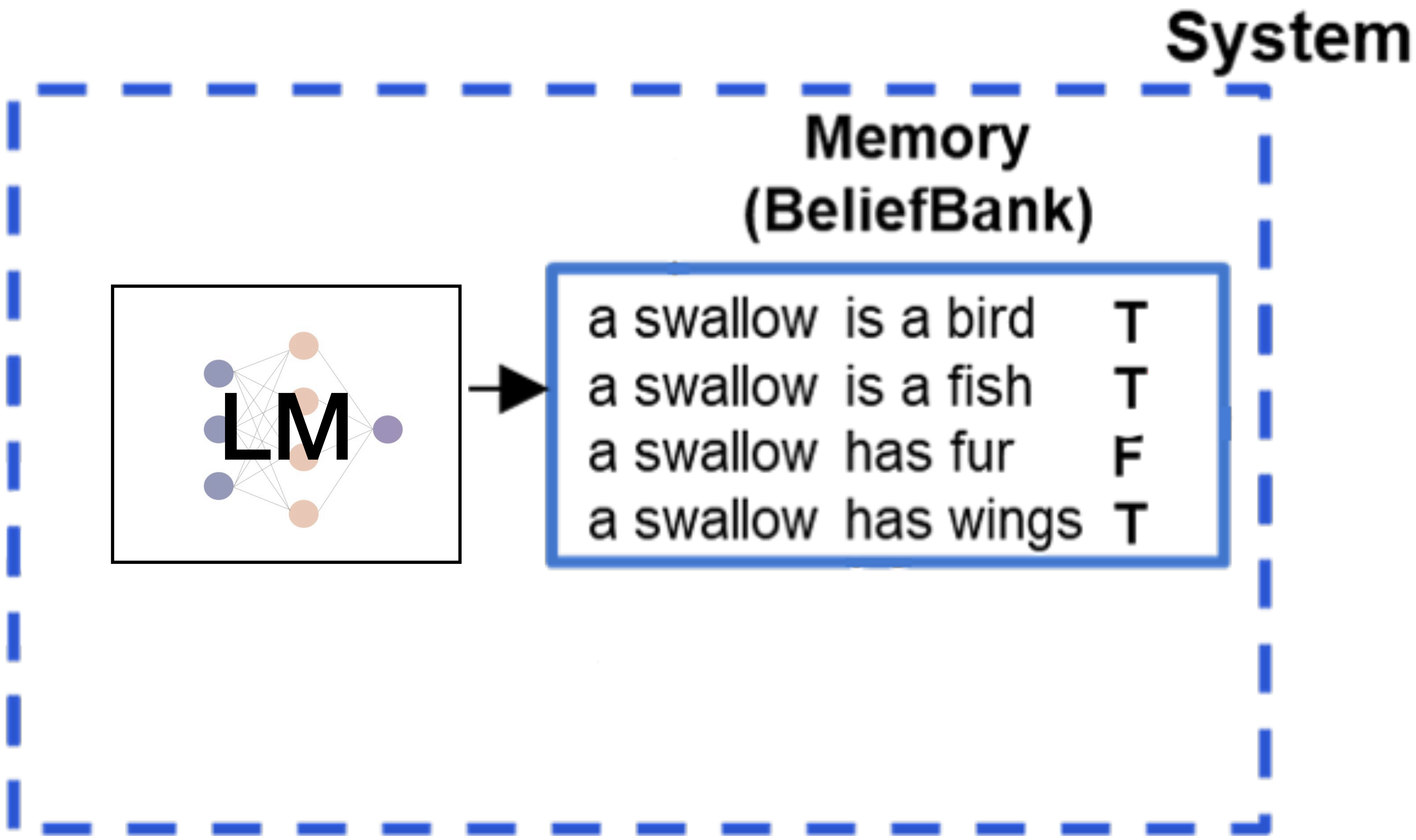
→ Kassner et al.: BeliefBank: Adding Memory to a Pre-trained Language Model for a Systematic Notion of Belief, EMNLP 2021

Towards more consistent knowledge: BeliefBank



→ Kassner et al.: BeliefBank: Adding Memory to a Pre-trained Language Model for a Systematic Notion of Belief, EMNLP 2021

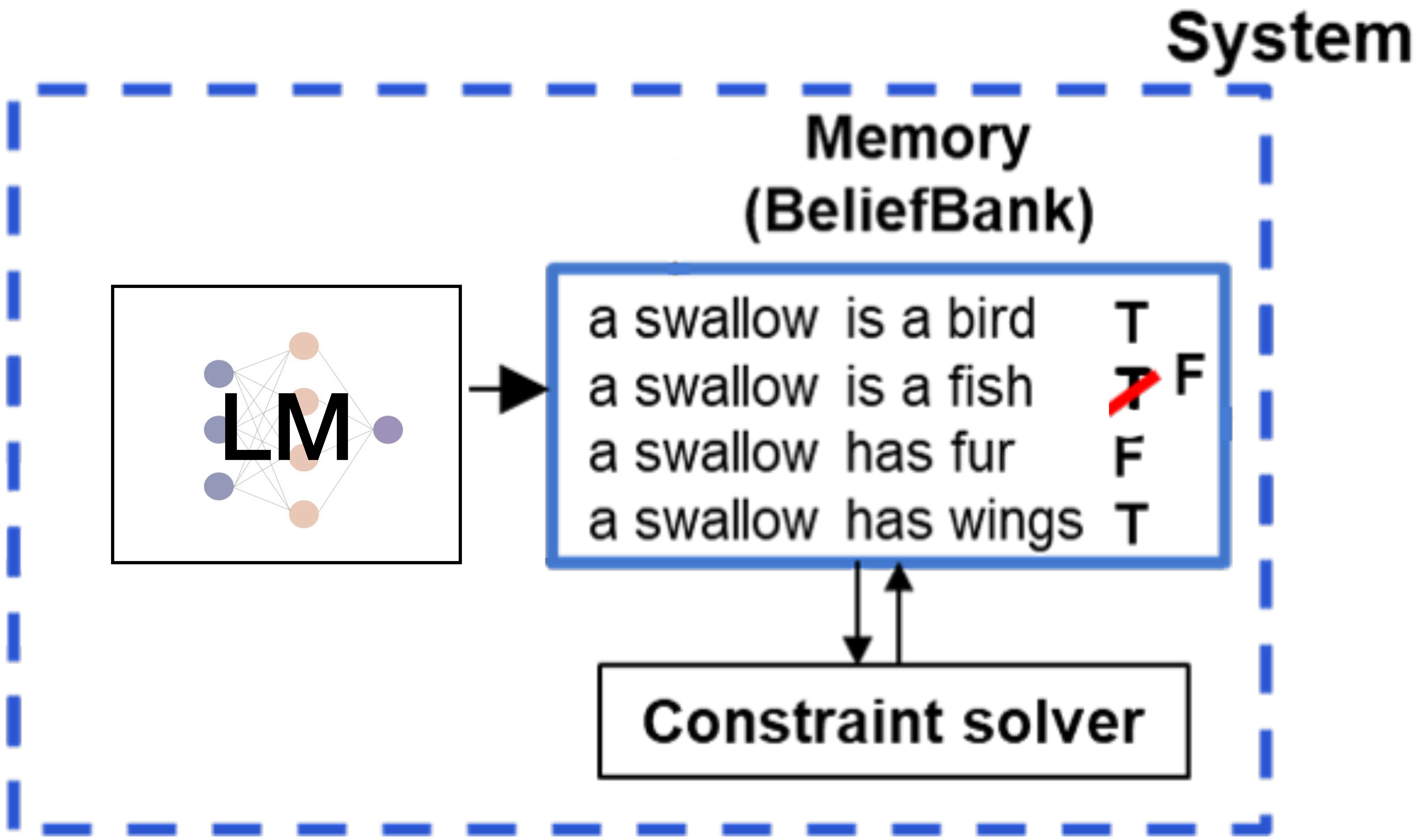
Towards more consistent knowledge: BeliefBank



a swallow is a bird	T
a swallow is a fish	T
a swallow has fur	F
a swallow has wings	T

→ Kassner et al.: BeliefBank: Adding Memory to a Pre-trained Language Model for a Systematic Notion of Belief, EMNLP 2021

Towards more consistent knowledge: BeliefBank



System

LM

Memory (BeliefBank)

a swallow is a bird T

a swallow is a fish ~~T~~ F

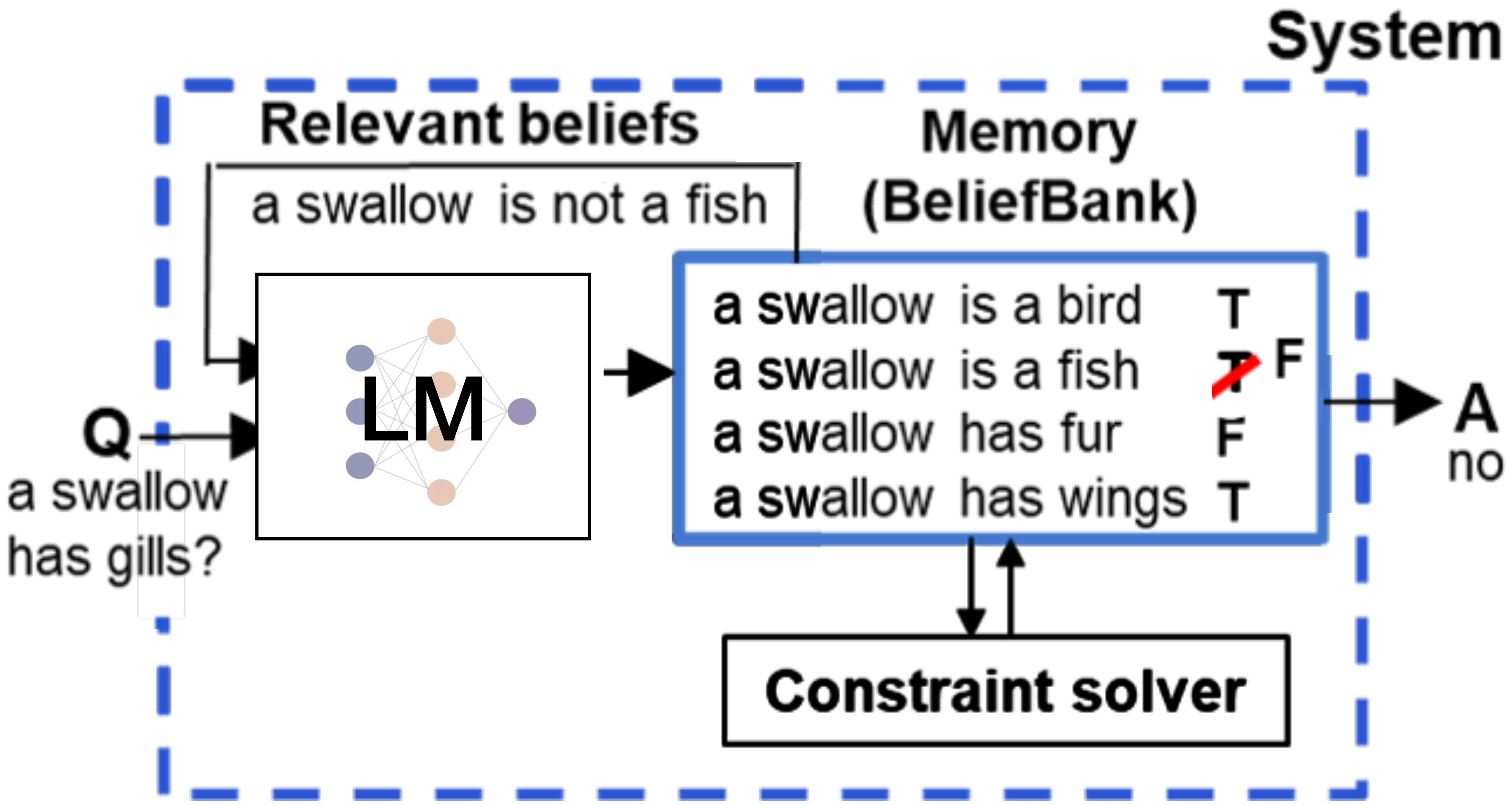
a swallow has fur F

a swallow has wings T

Constraint solver

→ Kassner et al.: BeliefBank: Adding Memory to a Pre-trained Language Model for a Systematic Notion of Belief, EMNLP 2021

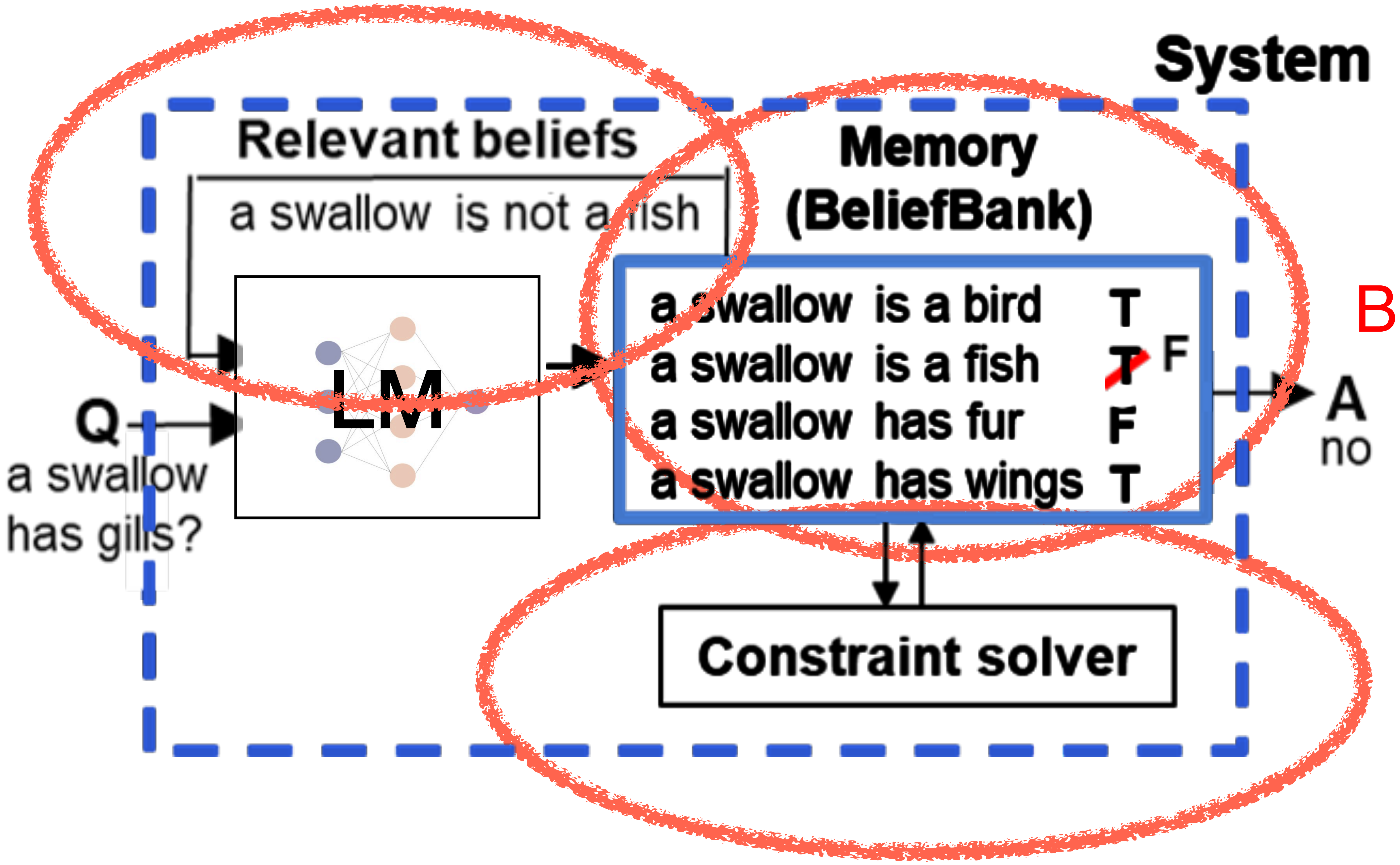
Towards more consistent knowledge: BeliefBank



→ Kassner et al.: BeliefBank: Adding Memory to a Pre-trained Language Model for a Systematic Notion of Belief, EMNLP 2021

Towards more consistent knowledge: BeliefBank

Feedback



BeliefBank

Constraint Solver

→ Kassner et al.: BeliefBank: Adding Memory to a Pre-trained Language Model for a Systematic Notion of Belief, EMNLP 2021

Towards more consistent knowledge: BeliefBank

1. **Feedback mechanism:**

→ Adding related beliefs as context when querying the model

Context: A poodle is a dog. A poodle is an animal.

Question: Is a poodle a mammal?

→ Similar in spirit to: Shwartz et al.: Unsupervised Commonsense Question Answering with Self-Talk, EMNLP 2020

→ Can reduce clashes locally

Towards more consistent knowledge: BeliefBank

2. Constraint solver (Weighted Max SAT solver)

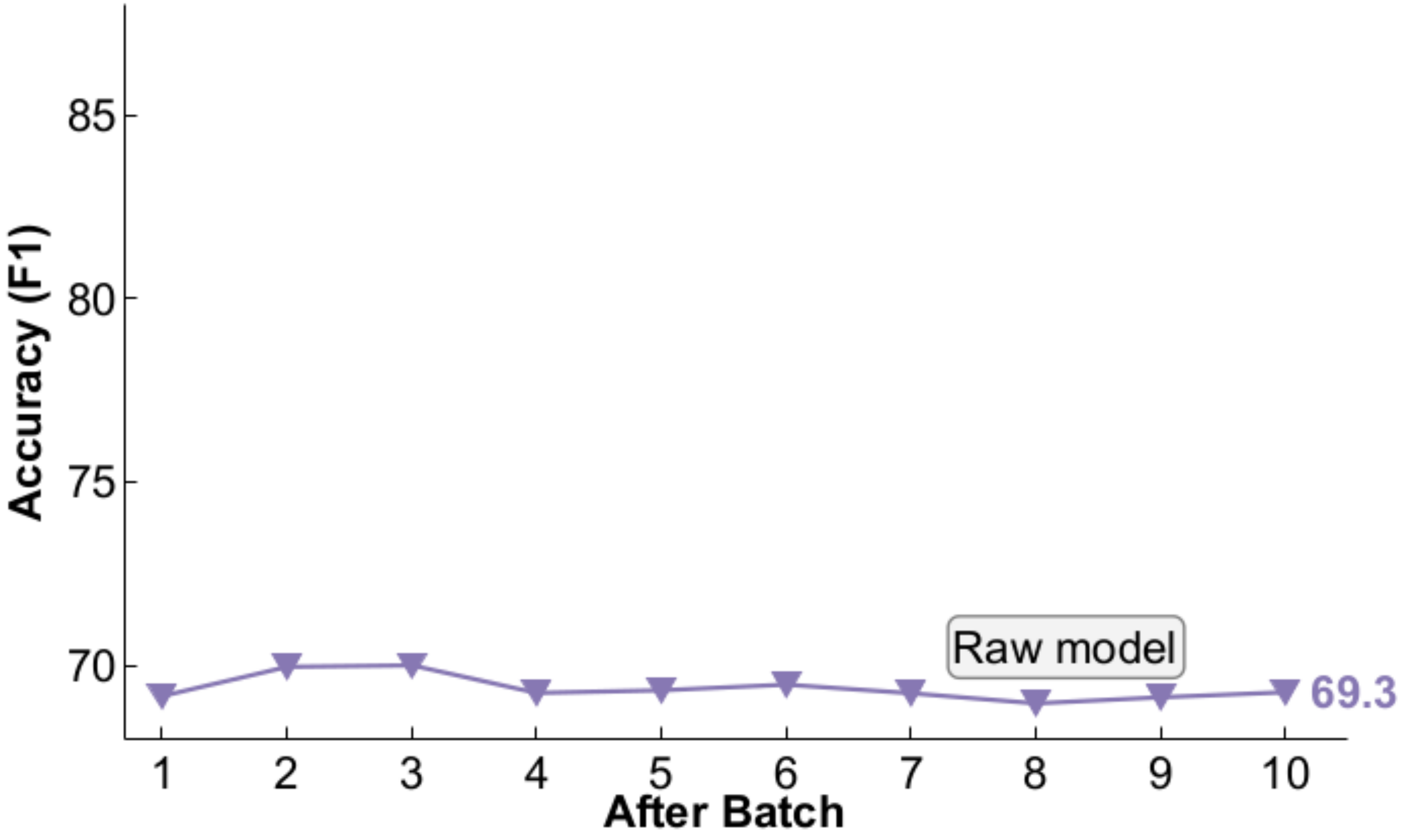
→ Reasoning component that potentially flips answers that maximally clash with others

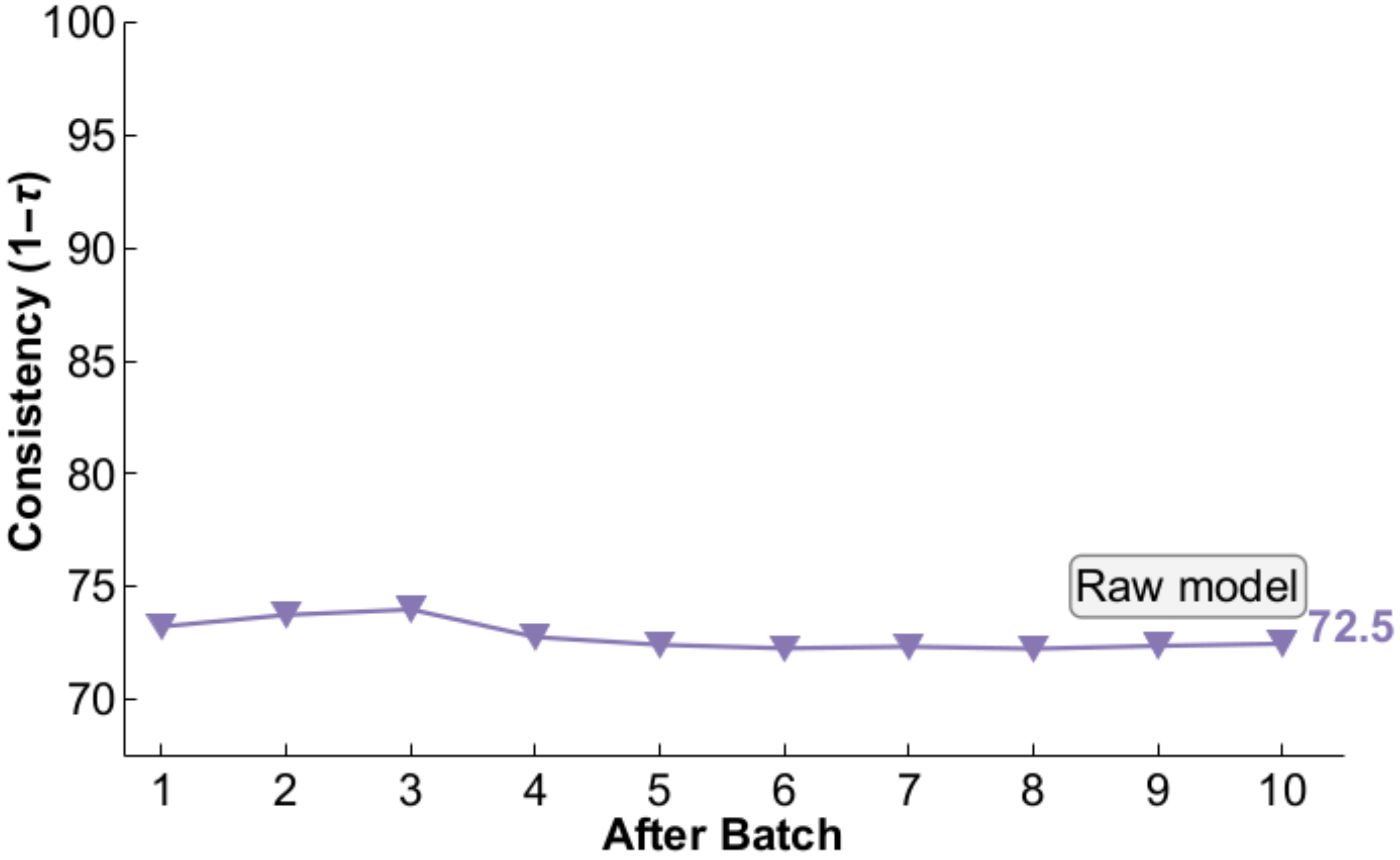
→ Two competing objectives:
a) Flip belief to minimise constraint violations
b) Don’t flip to preserve the model’s raw answers

→ Minimising conflict between the model and constraints

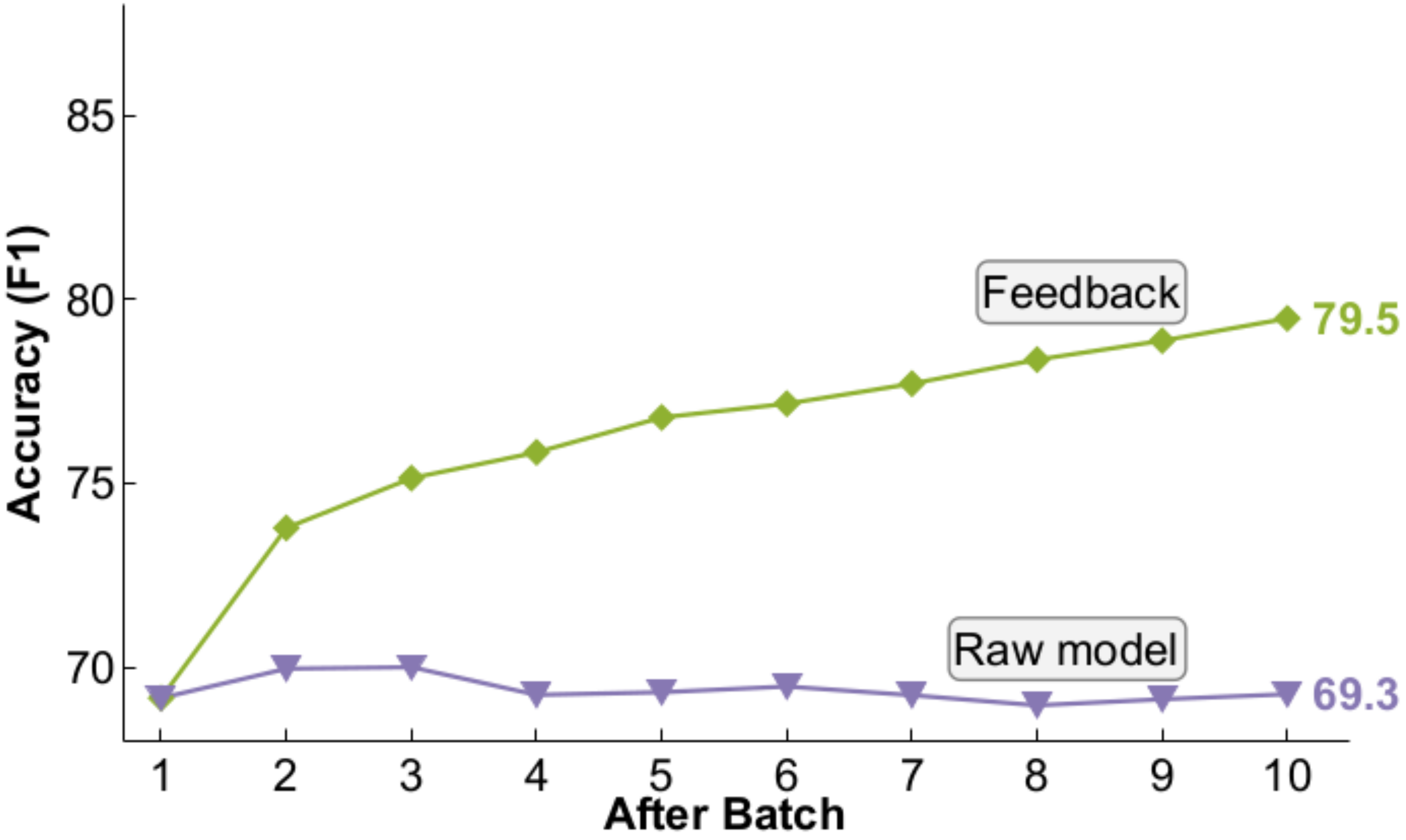
→ Can reduce clashes globally

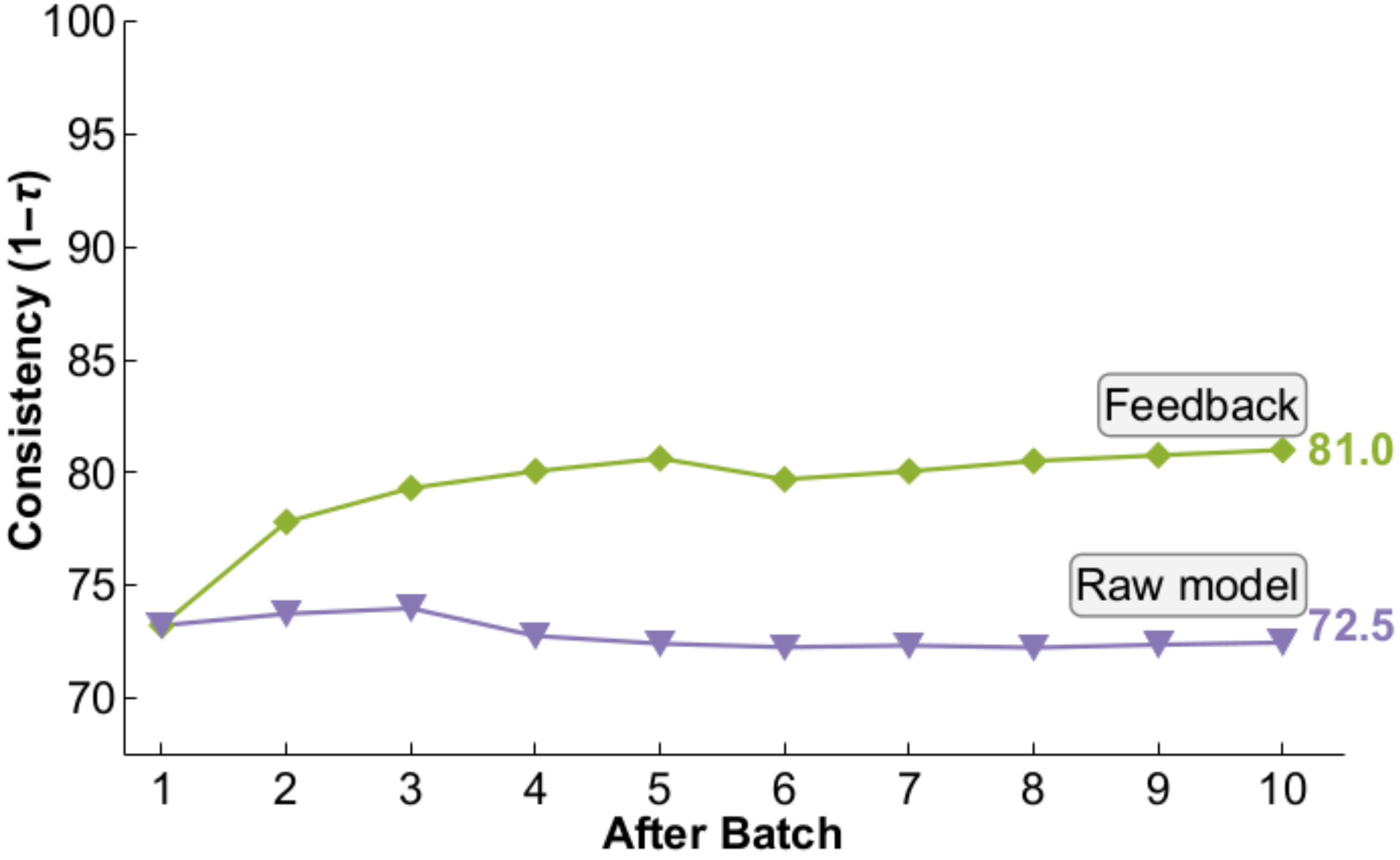
Towards more consistent knowledge: BeliefBank



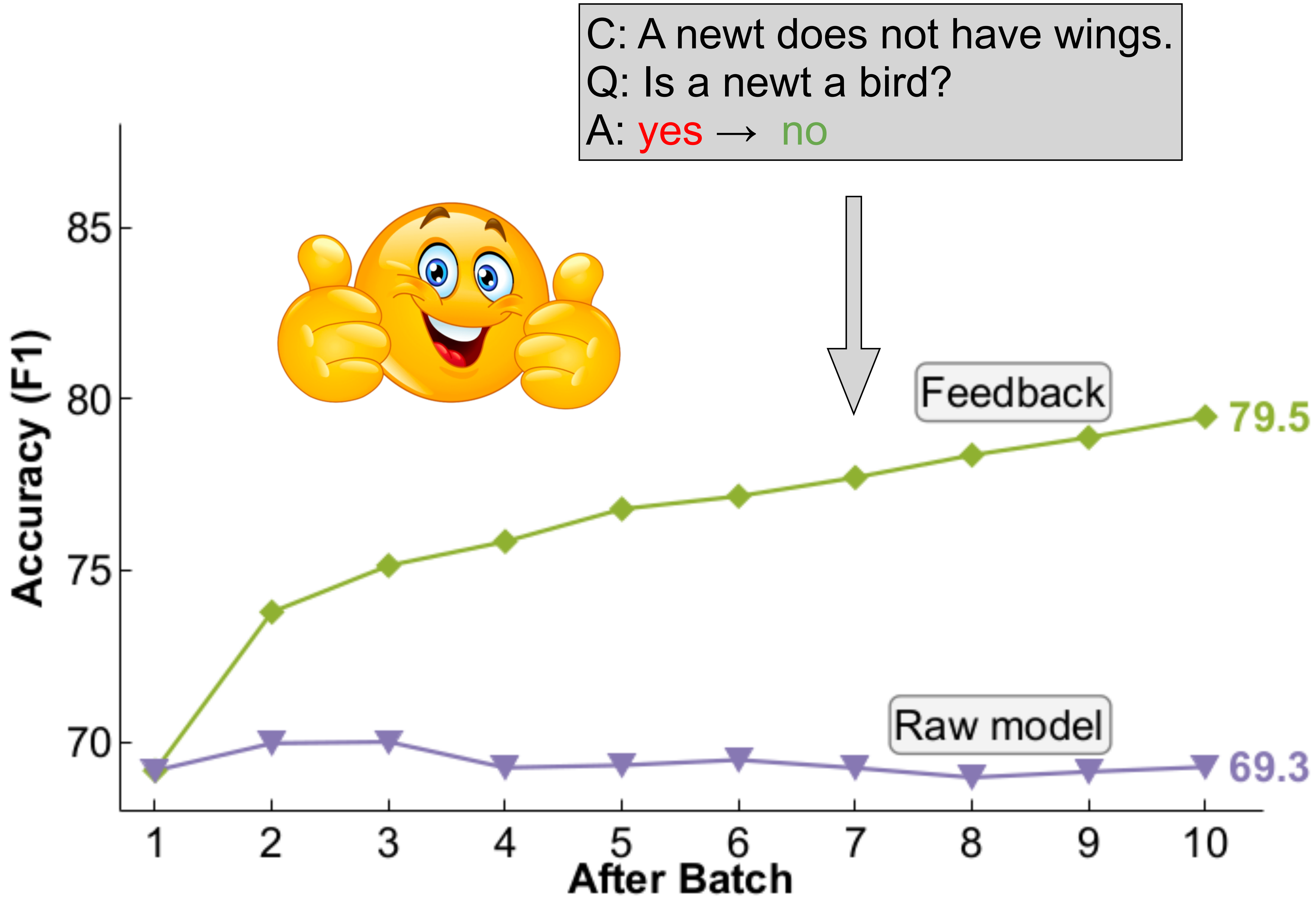


Towards more consistent knowledge: BeliefBank

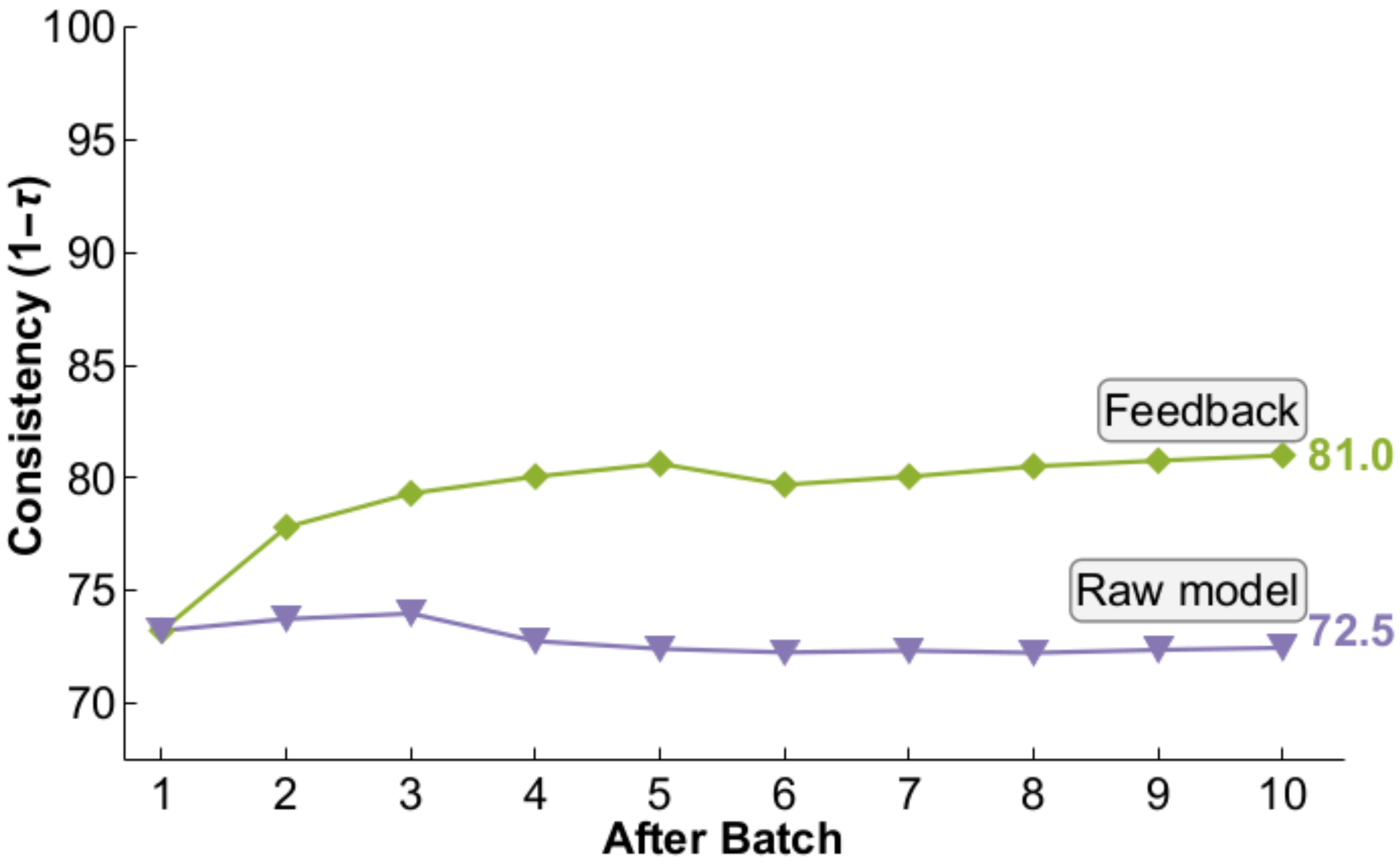




Towards more consistent knowledge: BeliefBank

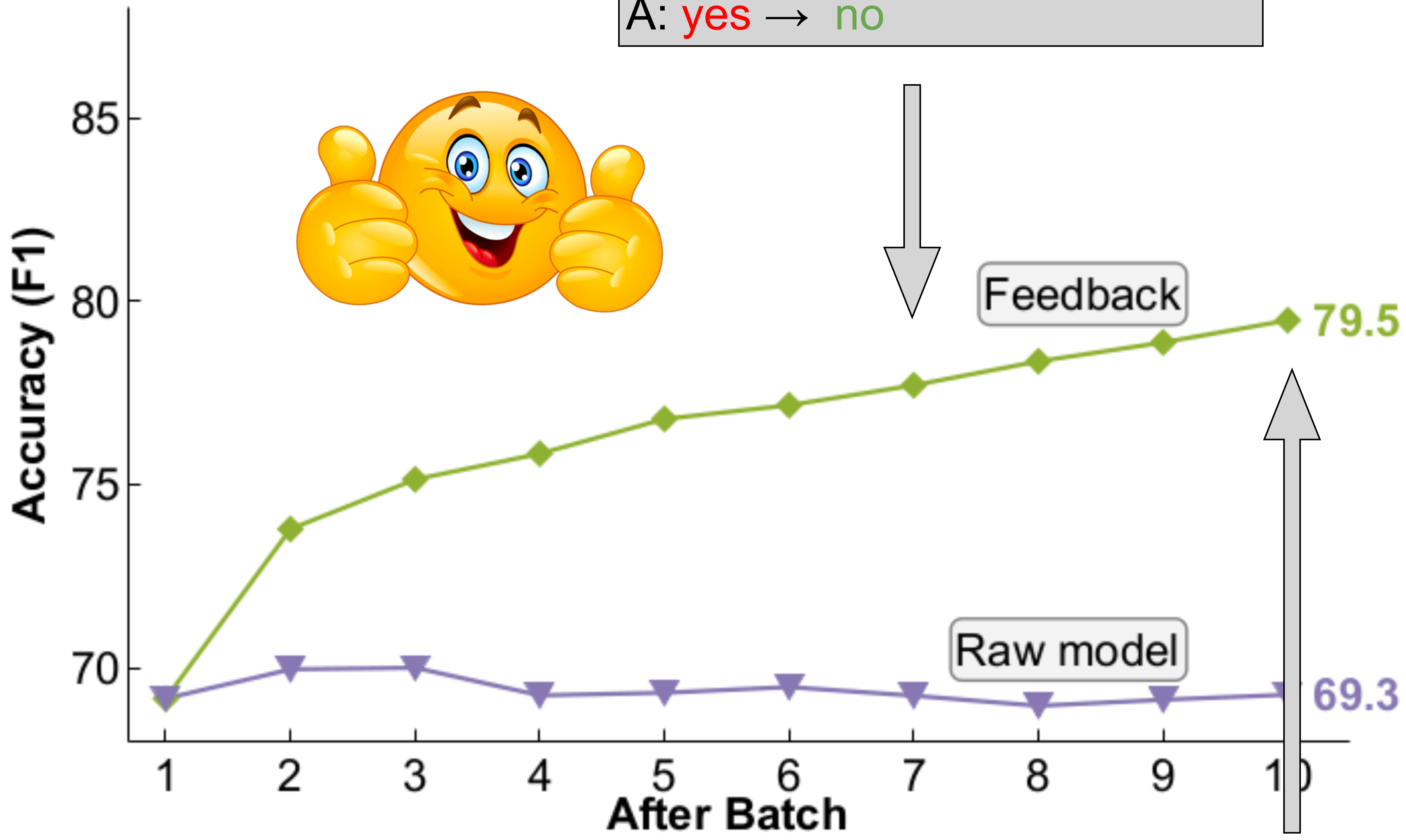


C: A newt does not have wings.
Q: Is a newt a bird?
A: yes → no

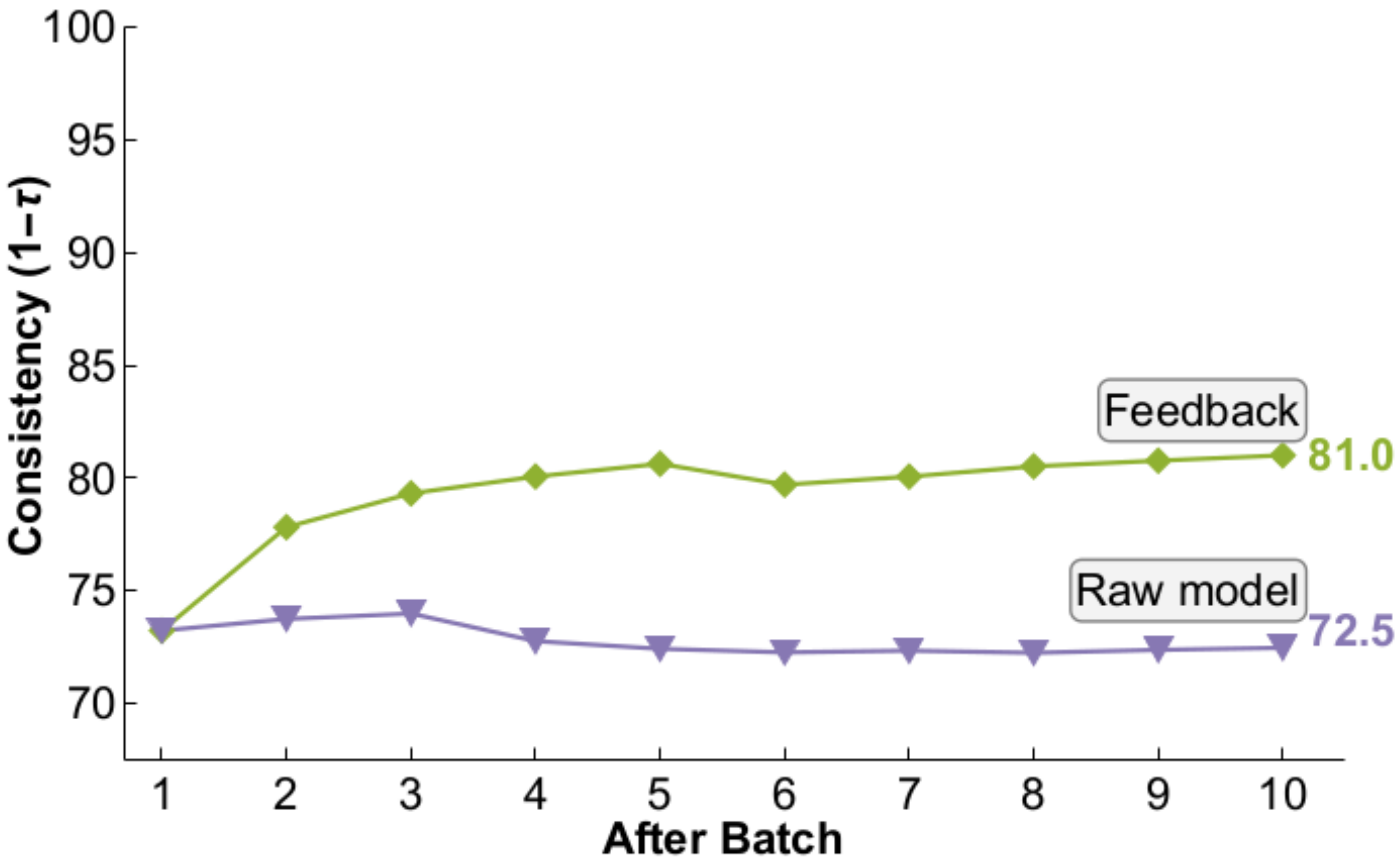


Towards more consistent knowledge: BeliefBank

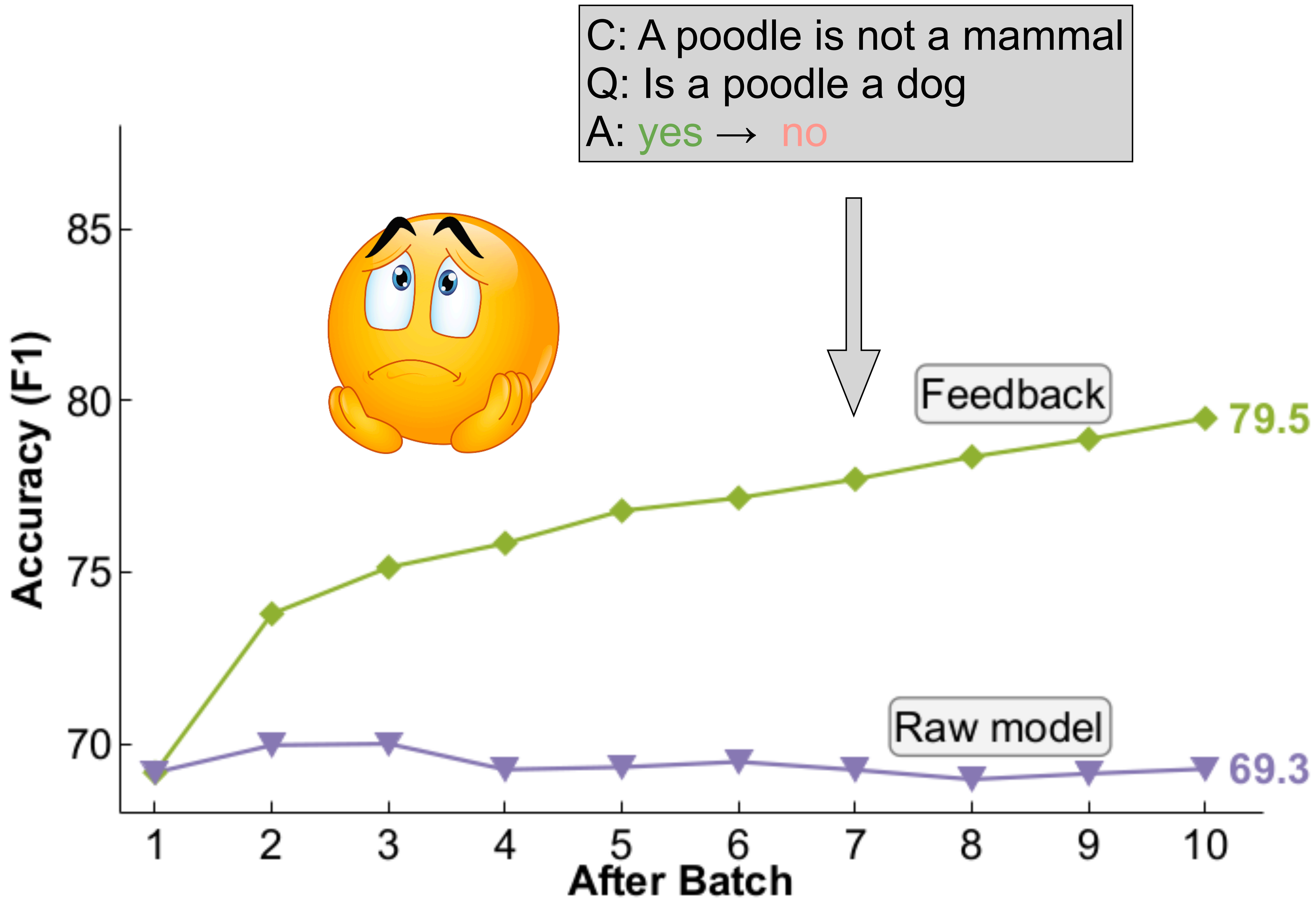
C: A newt does not have wings.
Q: Is a newt a bird?
A: yes → no



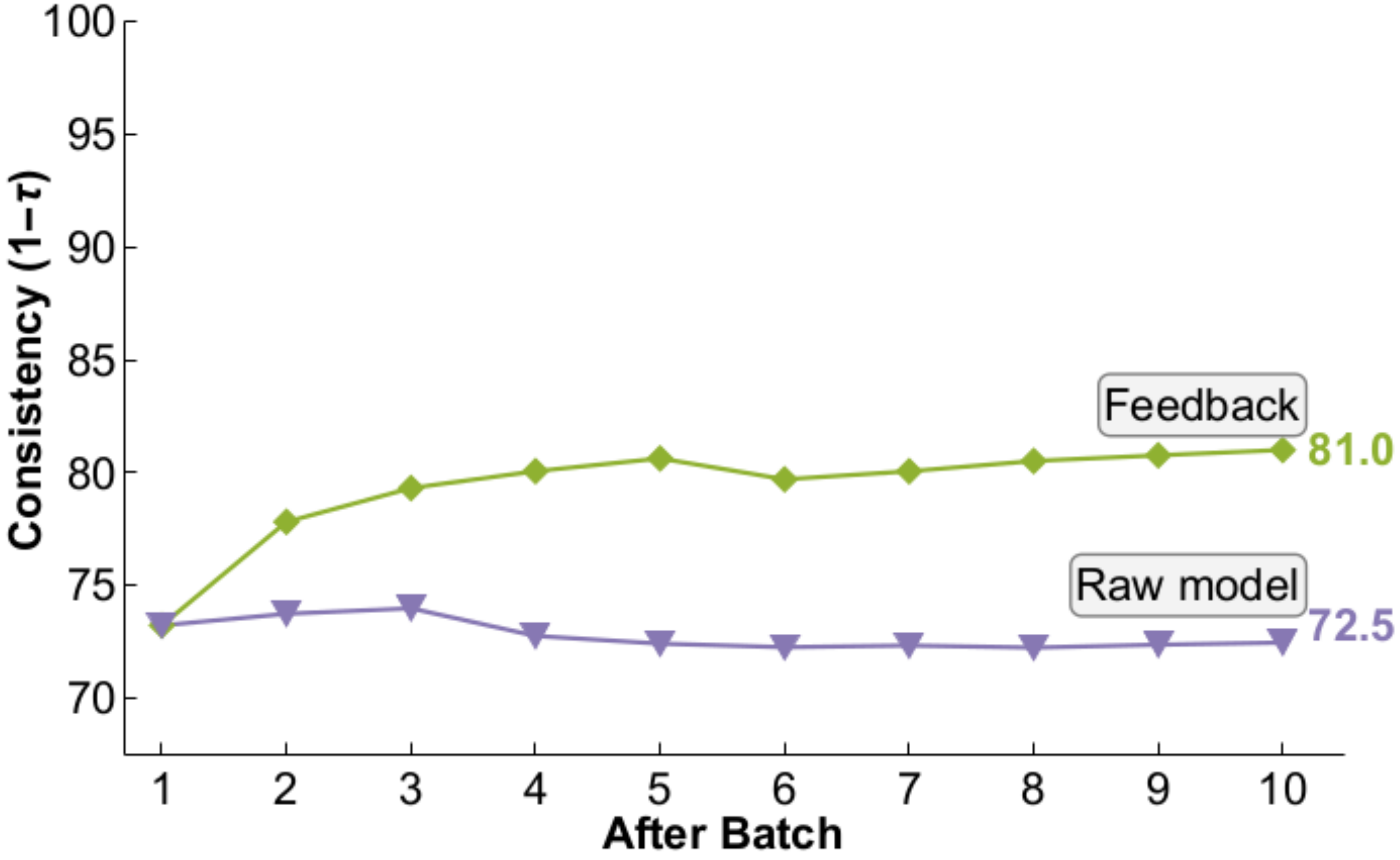
C: A newt is not a bird
Q: Is a newt a feathered animal?
A: yes → no



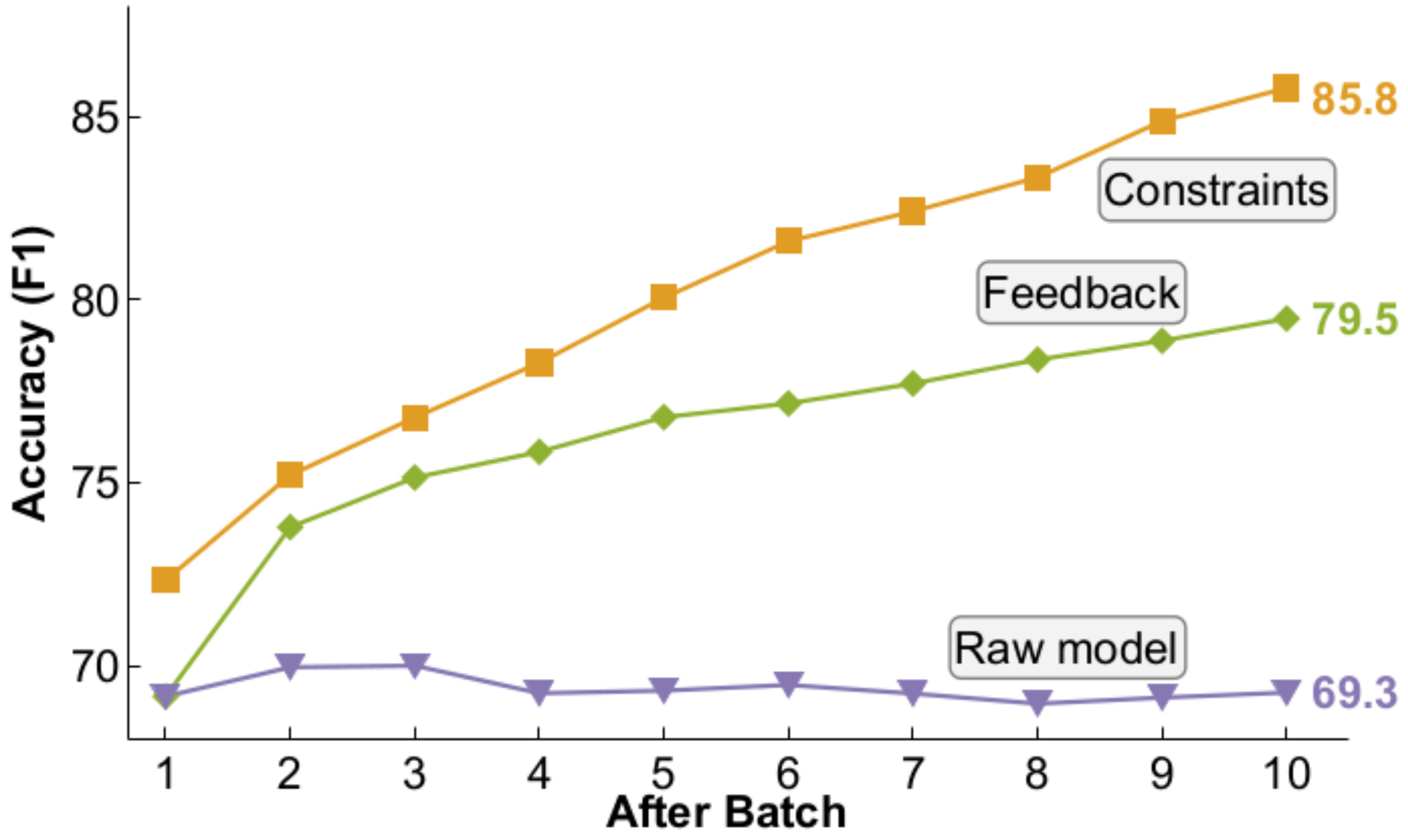
Towards more consistent knowledge: BeliefBank

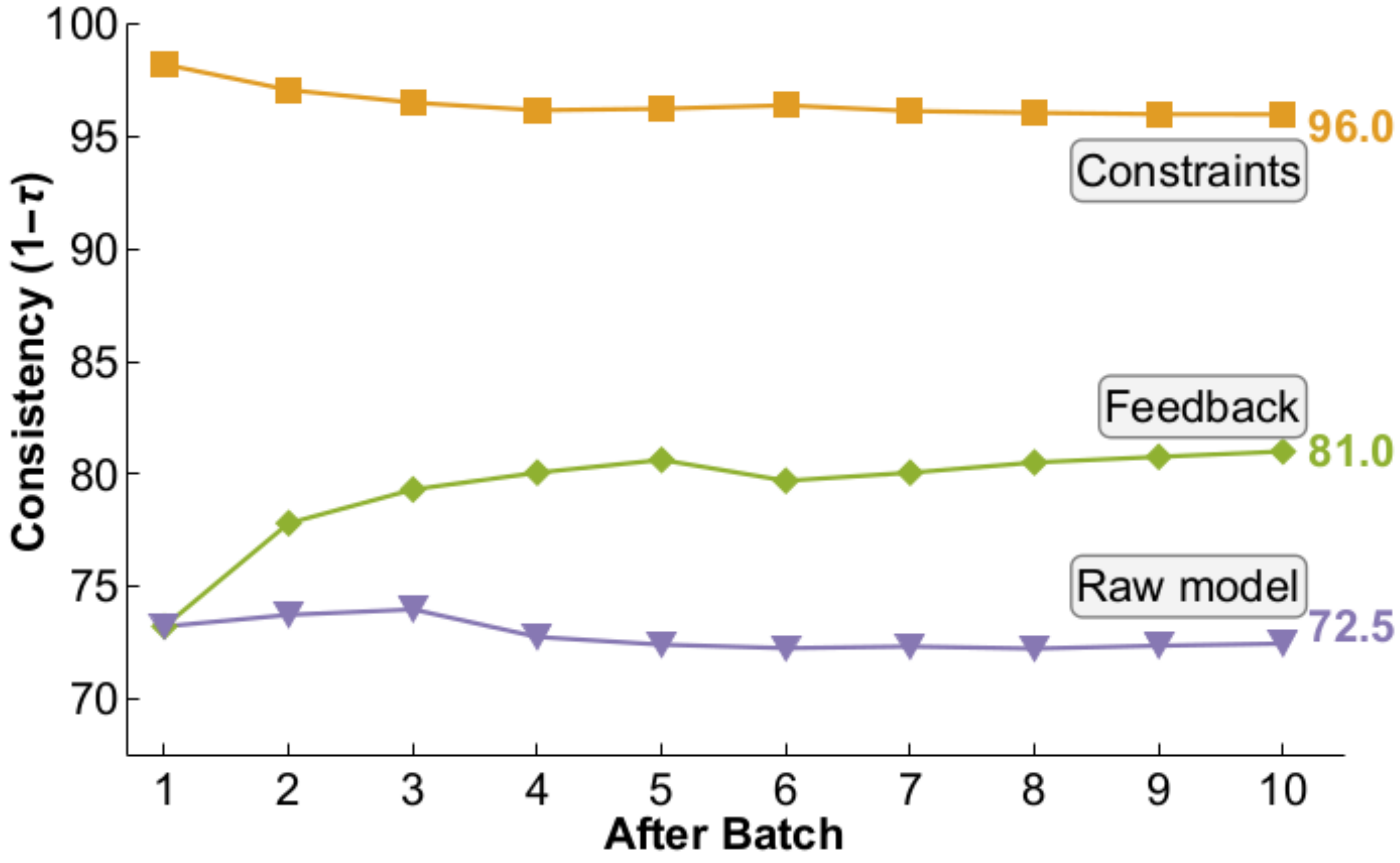


C: A poodle is not a mammal
Q: Is a poodle a dog
A: yes → no

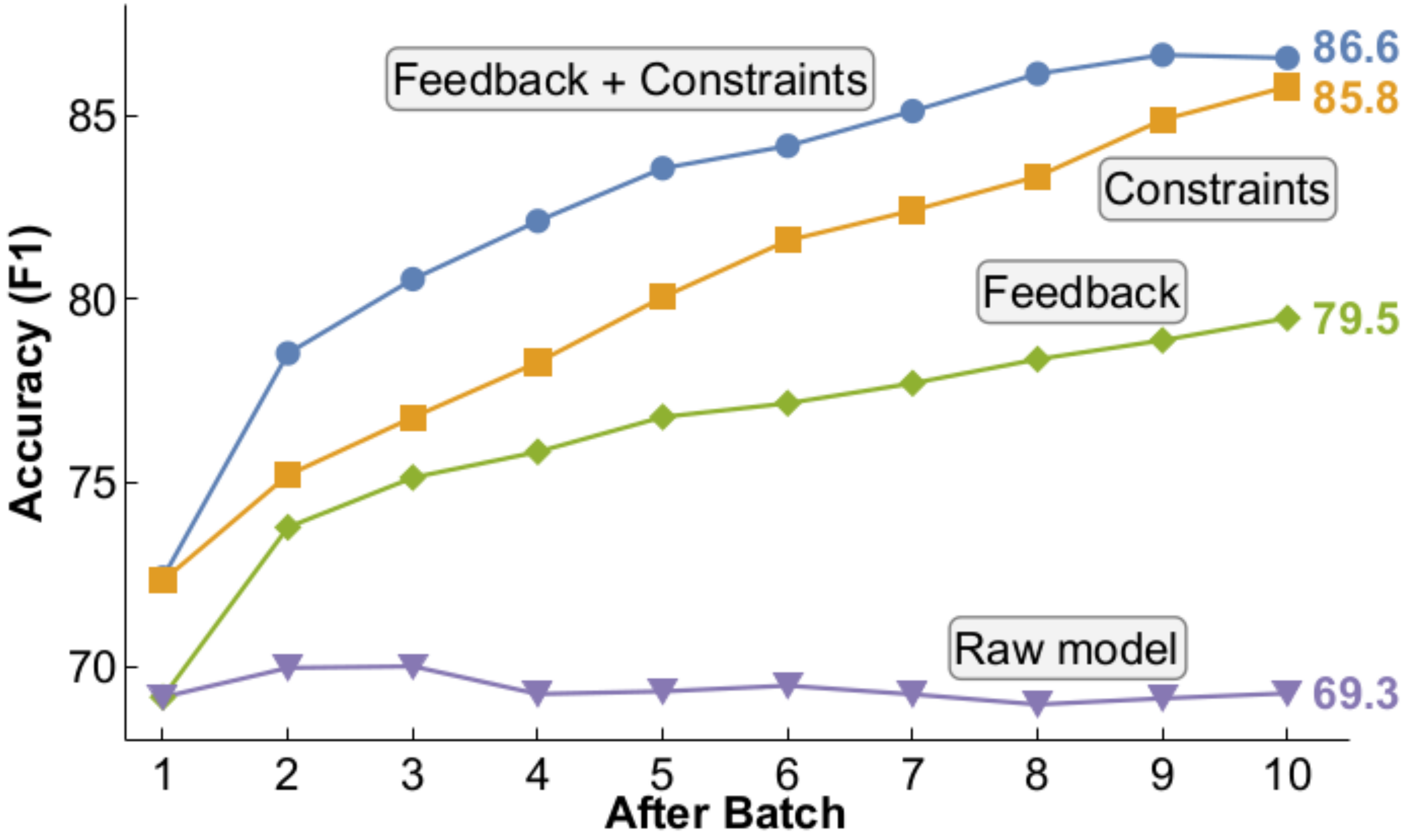


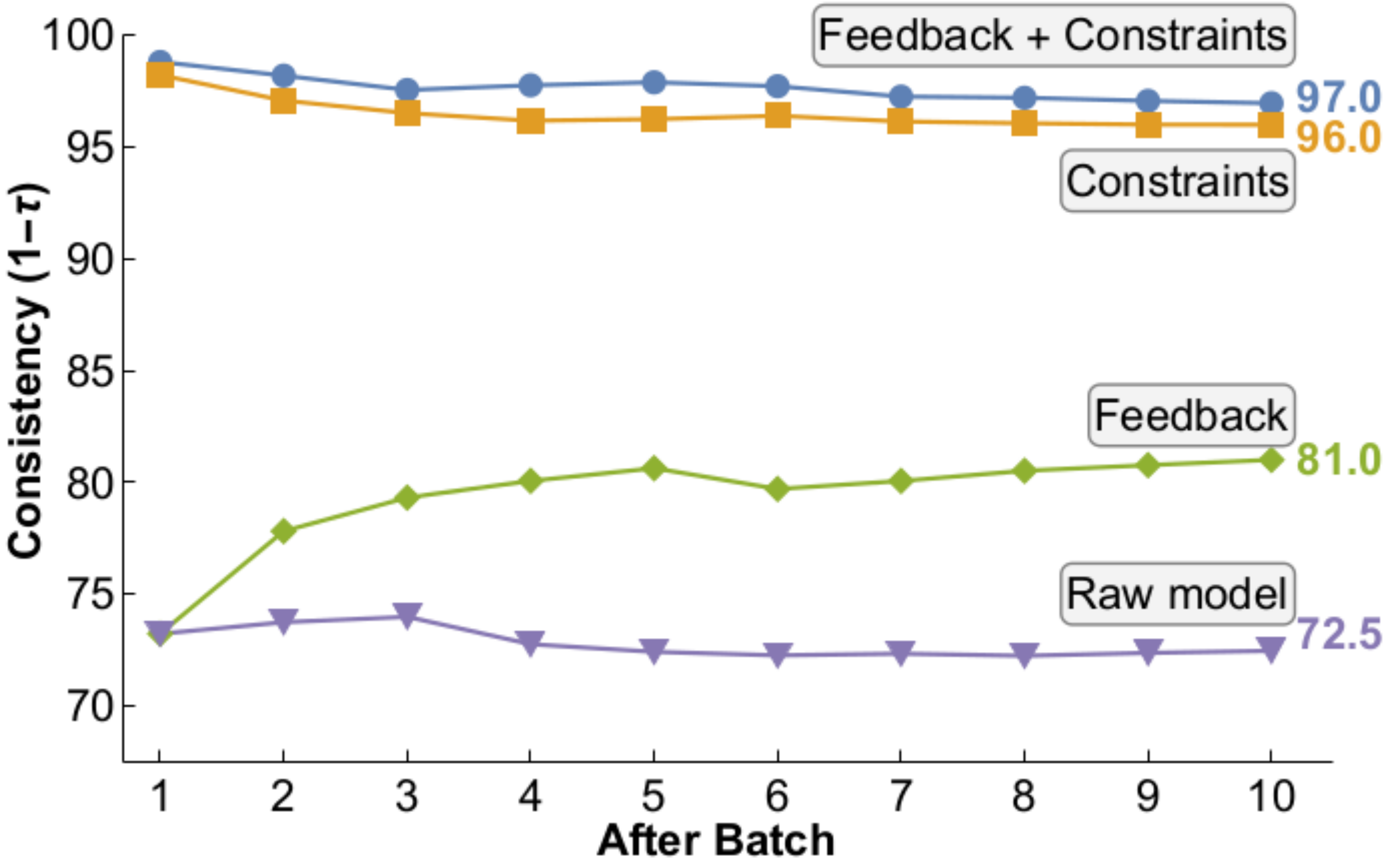
Towards more consistent knowledge: BeliefBank



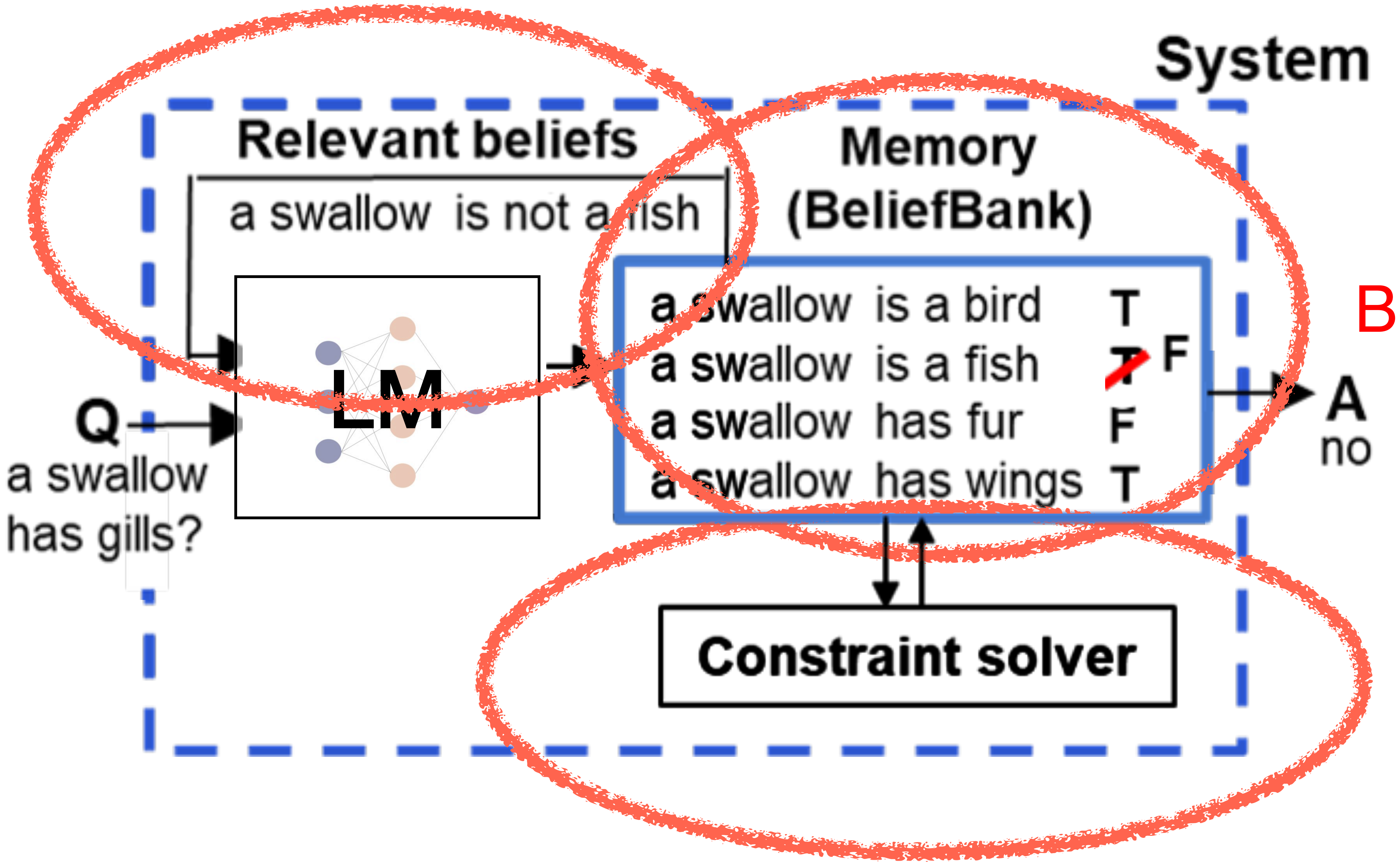


Towards more consistent knowledge: BeliefBank



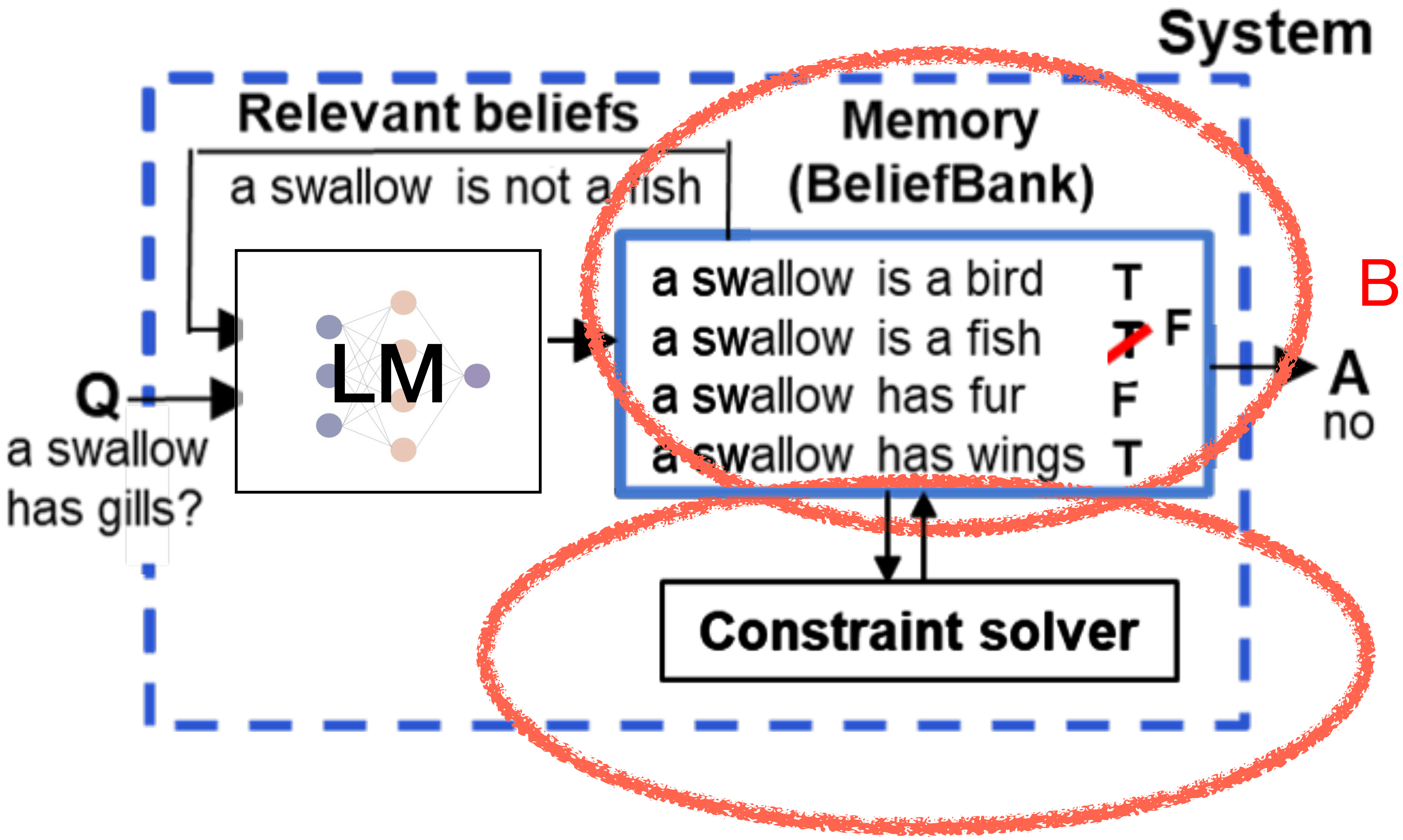


Towards more consistent knowledge: BeliefBank



→ Kassner et al.: BeliefBank: Adding Memory to a Pre-trained Language Model for a Systematic Notion of Belief, EMNLP 2021

Towards more consistent knowledge: REFLEX

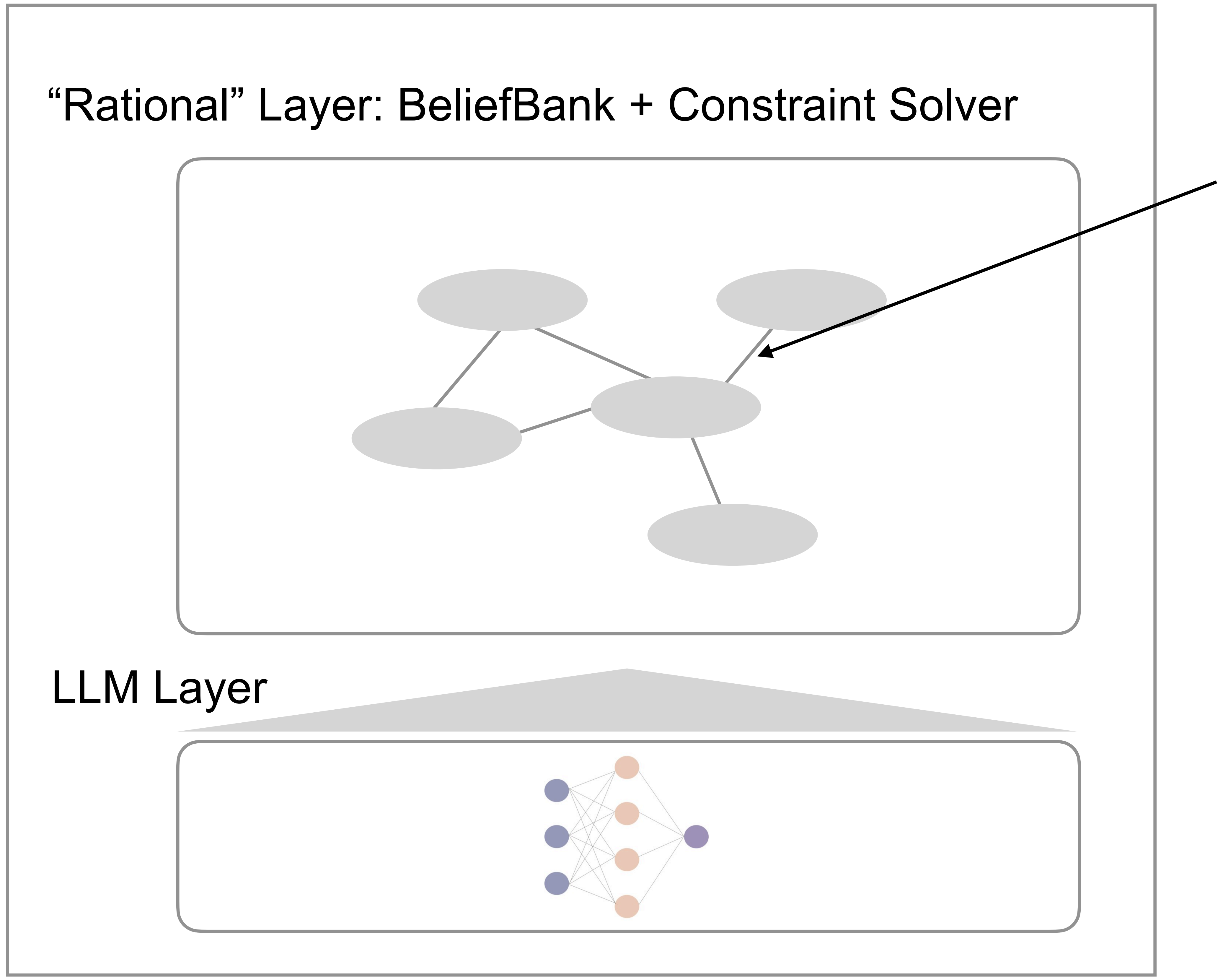


BeliefBank

Constraint Solver

→ Kassner et al.: Language Models with Rationality, EMNLP 2023

Towards more consistent knowledge: REFLEX



Tafjord et al.: **Entailer**: Answering Questions with Faithful and Truthful Chains of Reasoning", EMNLP 2022

Towards more consistent knowledge: REFLEX

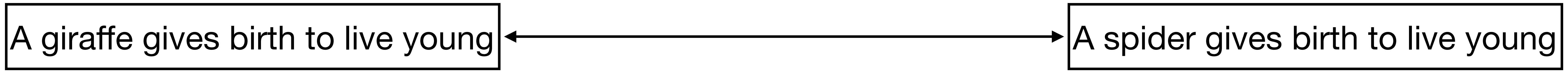
Which animal gives birth to live young?

(A) Shark (B) Turtle (C) Giraffe (D) Spider

Towards more consistent knowledge: REFLEX

Which animal gives birth to live young?

(A) Shark (B) Turtle (C) Giraffe (D) Spider



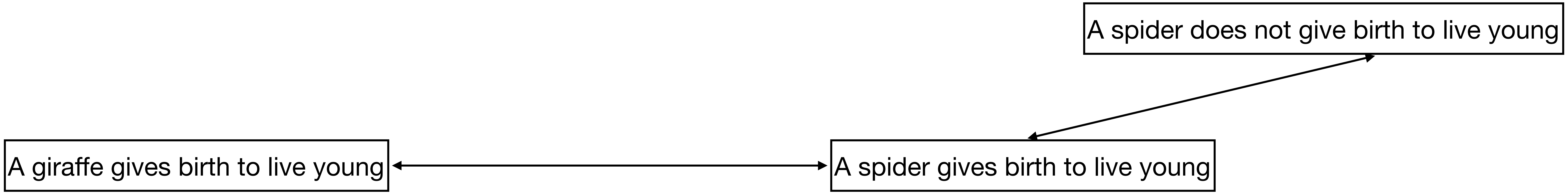
A giraffe gives birth to live young

A spider gives birth to live young

Towards more consistent knowledge: REFLEX

Which animal gives birth to live young?

(A) Shark (B) Turtle (C) Giraffe (D) Spider



A spider does not give birth to live young

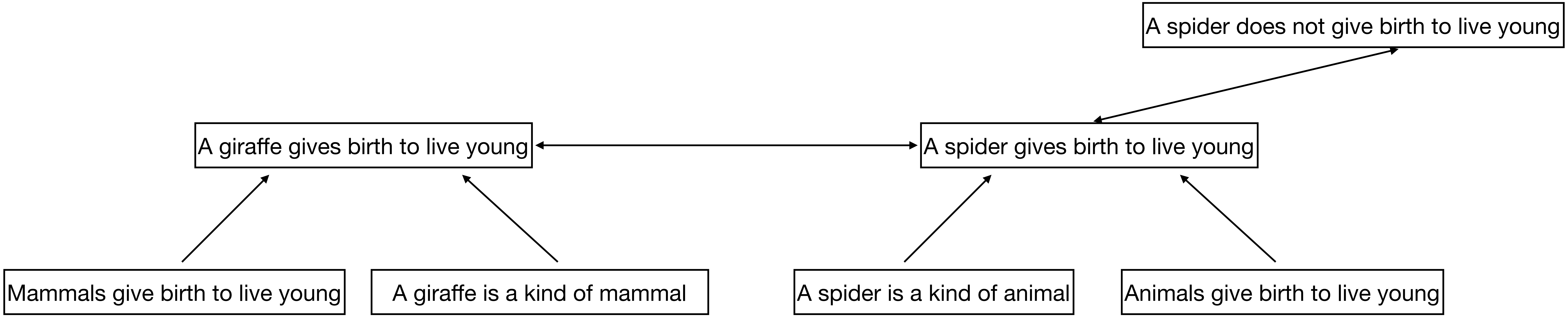
A giraffe gives birth to live young

A spider gives birth to live young

Towards more consistent knowledge: REFLEX

Which animal gives birth to live young?

(A) Shark (B) Turtle (C) Giraffe (D) Spider



A spider does not give birth to live young

A giraffe gives birth to live young

A spider gives birth to live young

Mammals give birth to live young

A giraffe is a kind of mammal

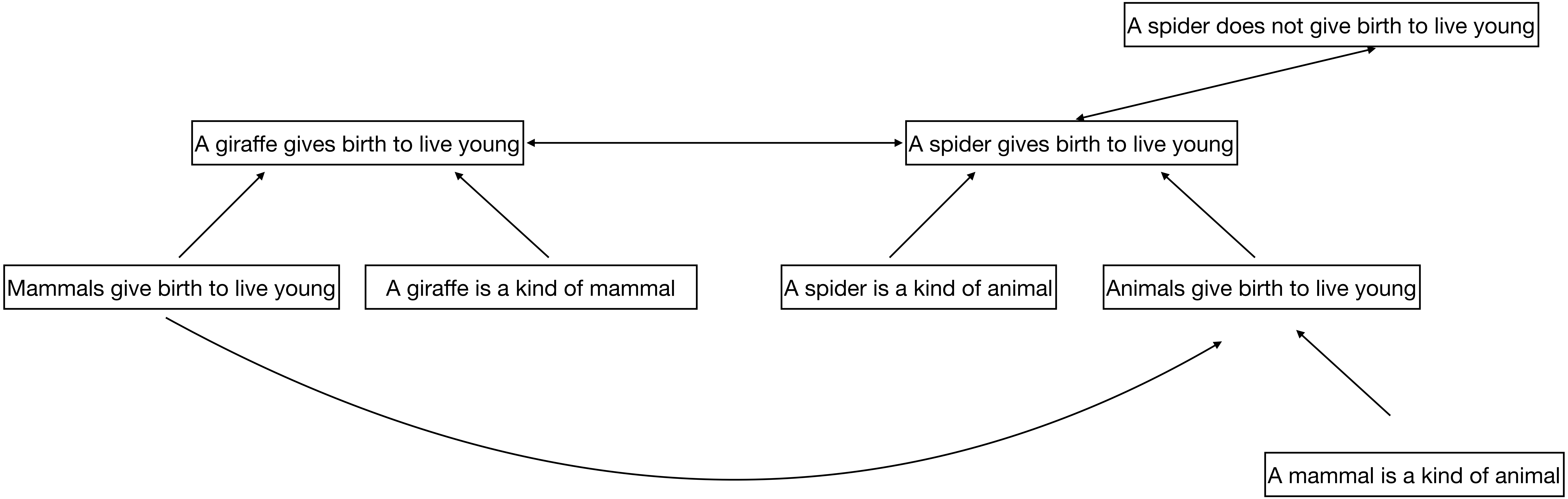
A spider is a kind of animal

Animals give birth to live young

Towards more consistent knowledge: REFLEX

Which animal gives birth to live young?

(A) Shark (B) Turtle (C) Giraffe (D) Spider



A spider does not give birth to live young

A giraffe gives birth to live young

A spider gives birth to live young

Mammals give birth to live young

A giraffe is a kind of mammal

A spider is a kind of animal

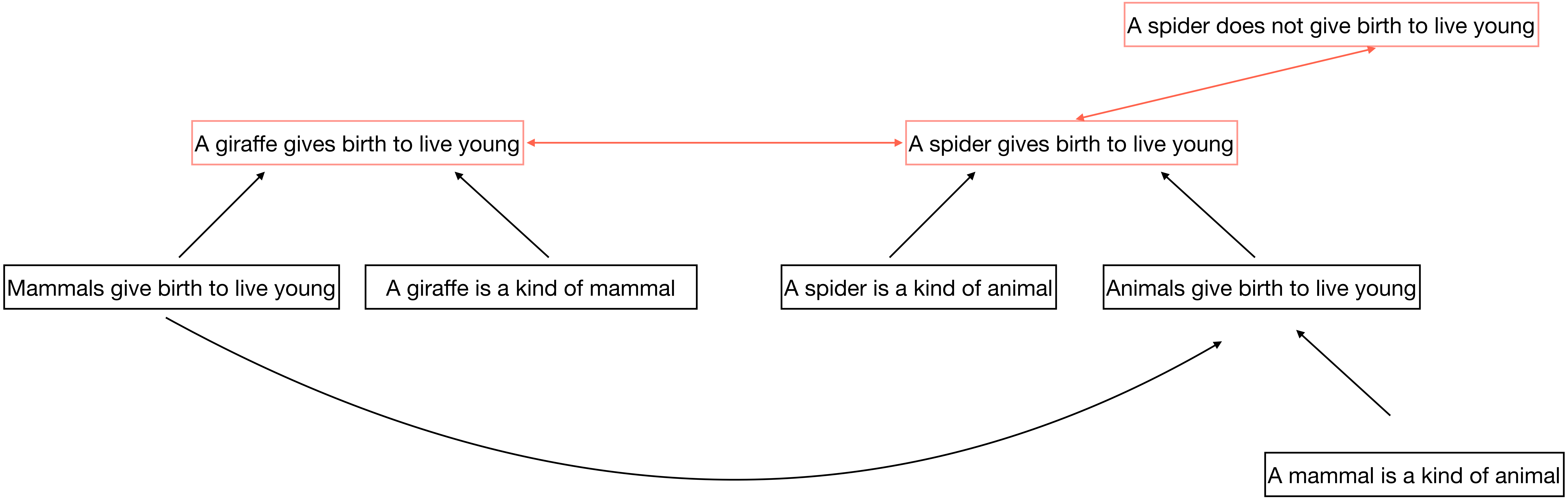
Animals give birth to live young

A mammal is a kind of animal

Towards more consistent knowledge: REFLEX

Which animal gives birth to live young?

(A) Shark (B) Turtle (C) Giraffe (D) Spider



A spider does not give birth to live young

A giraffe gives birth to live young

A spider gives birth to live young

Mammals give birth to live young

A giraffe is a kind of mammal

A spider is a kind of animal

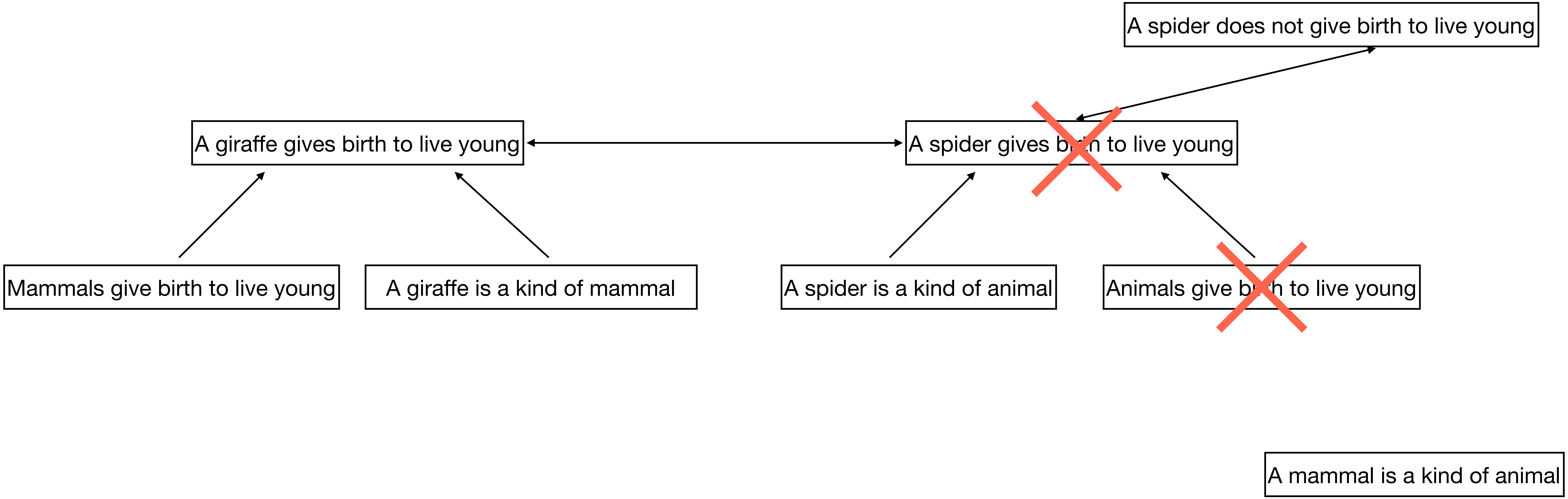
Animals give birth to live young

A mammal is a kind of animal

Towards more consistent knowledge: REFLEX

Which animal gives birth to live young?

(A) Shark (B) Turtle (C) Giraffe (D) Spider



A spider does not give birth to live young

A giraffe gives birth to live young

~~A spider gives birth to live young~~

Mammals give birth to live young

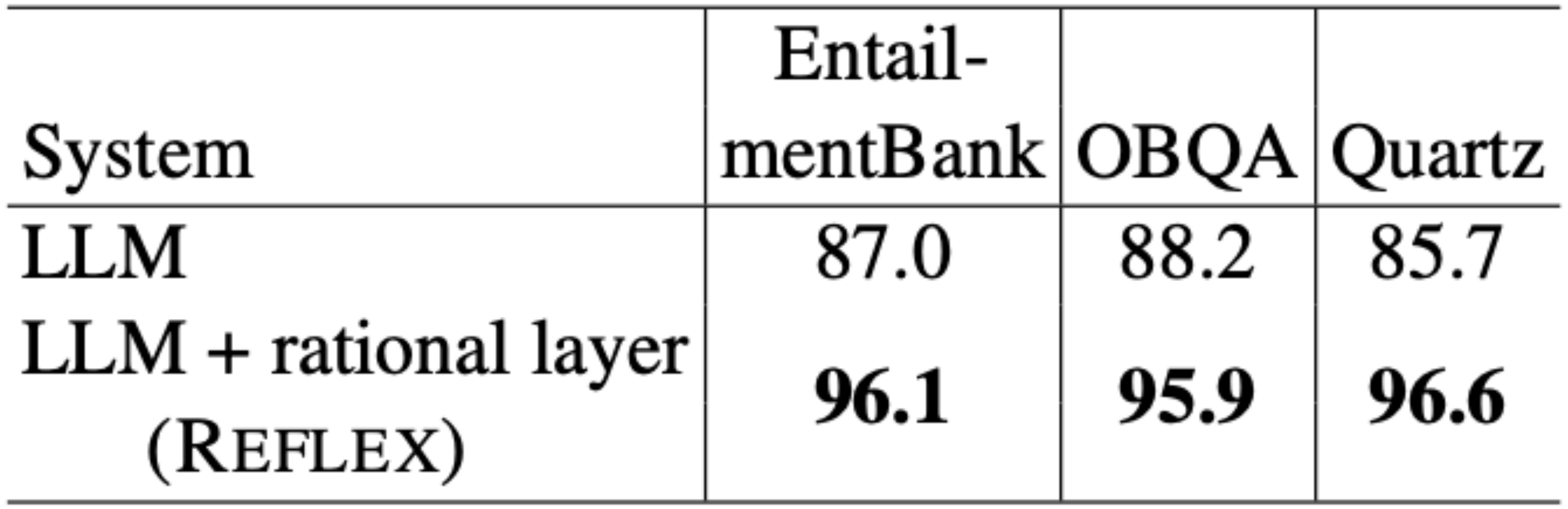
A giraffe is a kind of mammal

A spider is a kind of animal

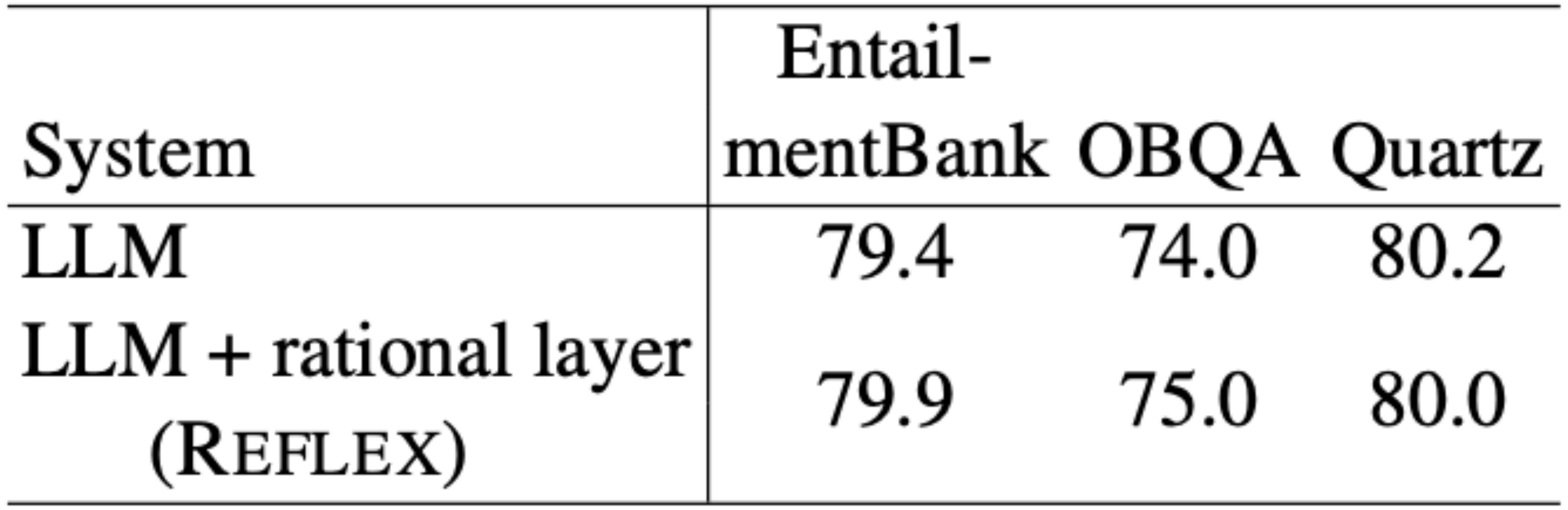
~~Animals give birth to live young~~

A mammal is a kind of animal

Towards more consistent knowledge: REFLEX



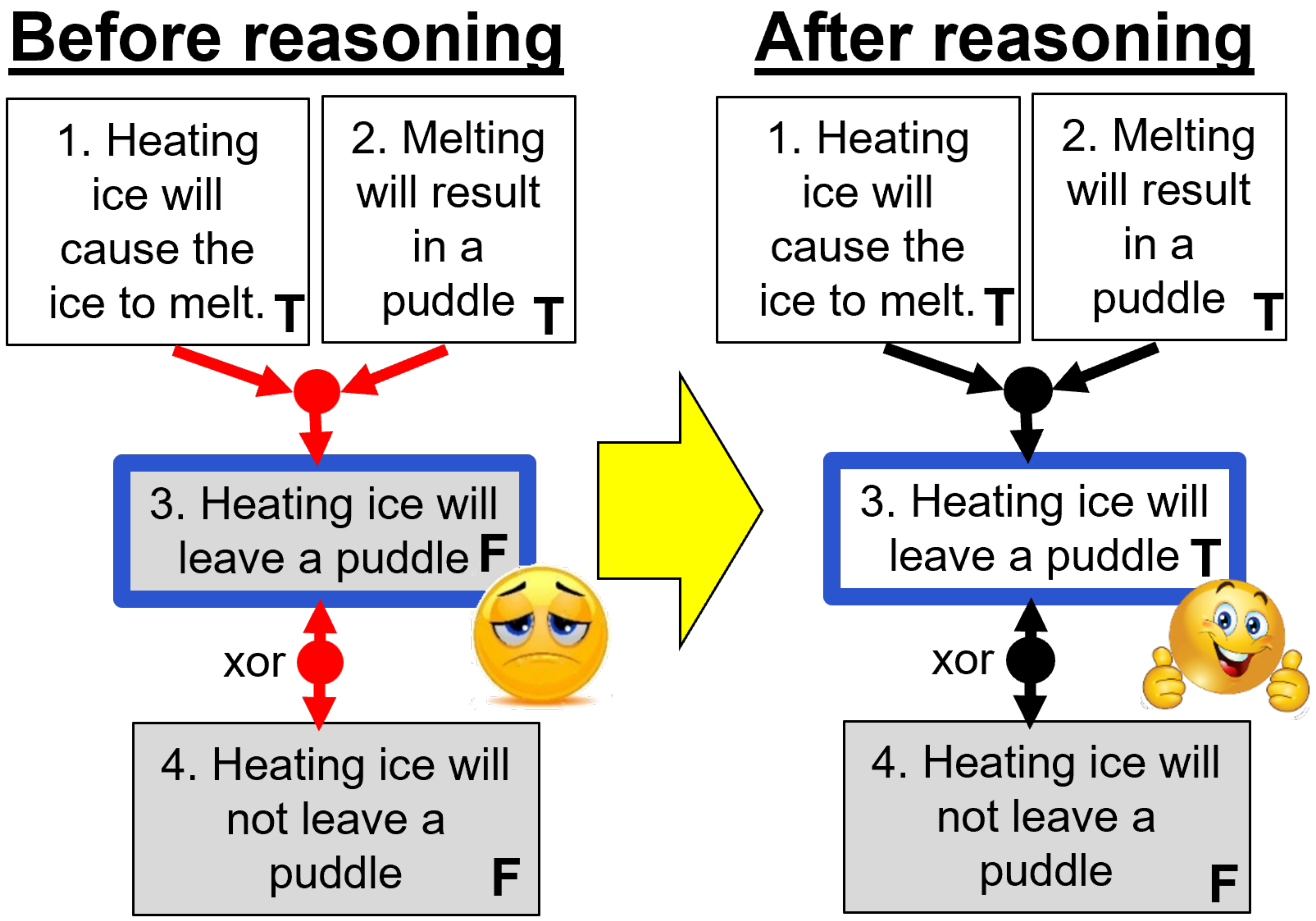
System	EntailmentBank	OBQA	Quartz
LLM	87.0	88.2	85.7
LLM + rational layer (REFLEX)	**96.1**	**95.9**	**96.6**



System	EntailmentBank	OBQA	Quartz
LLM	79.4	74.0	80.2
LLM + rational layer (REFLEX)	79.9	75.0	80.0

Towards more consistent knowledge: REFLEX

Heating ice (A) changes it's chemical make-up (B) will leave a puddle ***[correct]*** (C) makes it even colder



Before reasoning

1. Heating ice will cause the ice to melt. **T**

2. Melting will result in a puddle **T**

3. Heating ice will leave a puddle **F**

xor

4. Heating ice will not leave a puddle **F**

After reasoning

1. Heating ice will cause the ice to melt. **T**

2. Melting will result in a puddle **T**

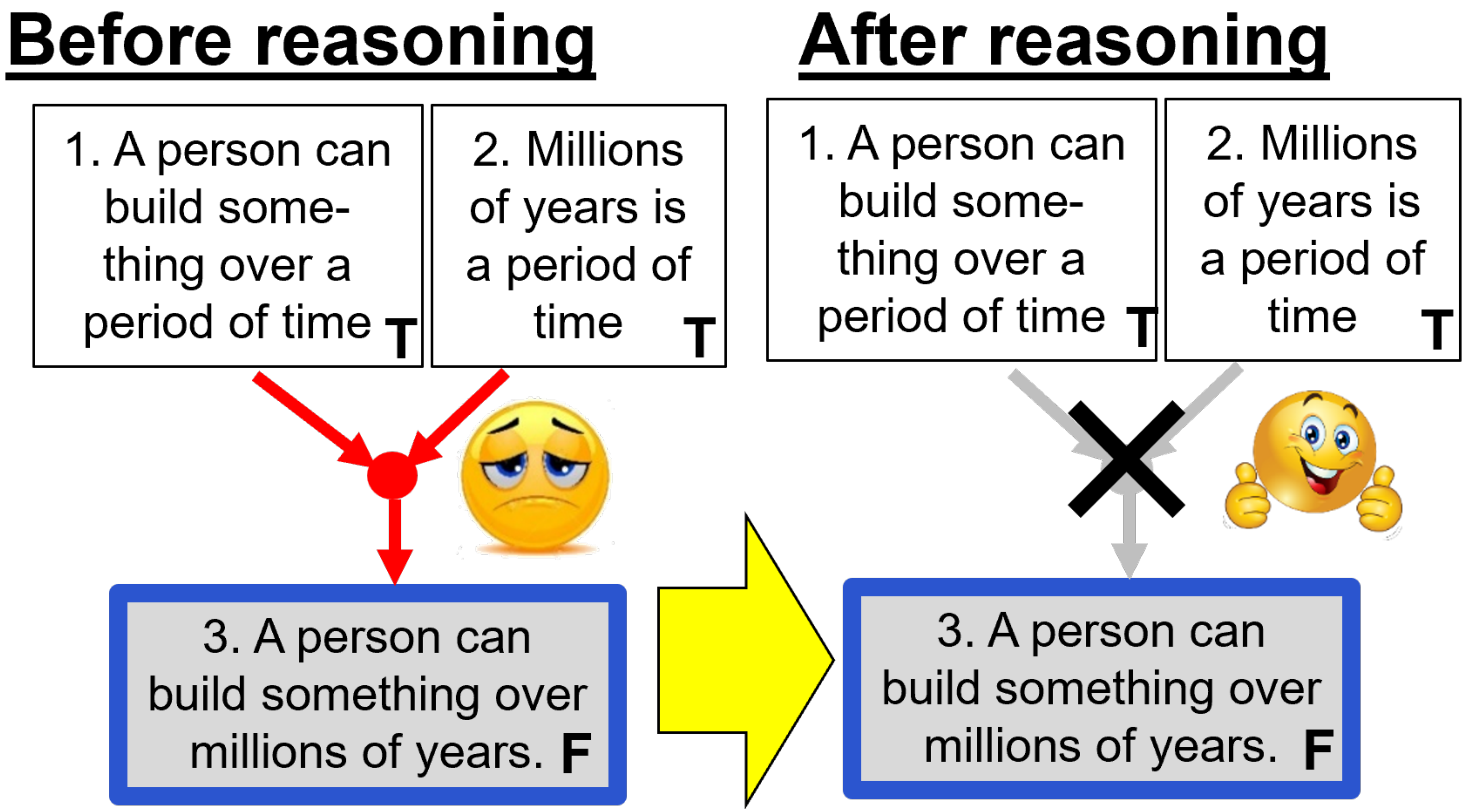
3. Heating ice will leave a puddle **T**

xor

4. Heating ice will not leave a puddle **F**

Towards more consistent knowledge: REFLEX

What can build something over millions of years? (A) a river ***[correct]*** (B) a person (C) society (D) dinosaurs



Before reasoning

1. A person can build something over a period of time **T**

2. Millions of years is a period of time **T**

3. A person can build something over millions of years. **F**

After reasoning

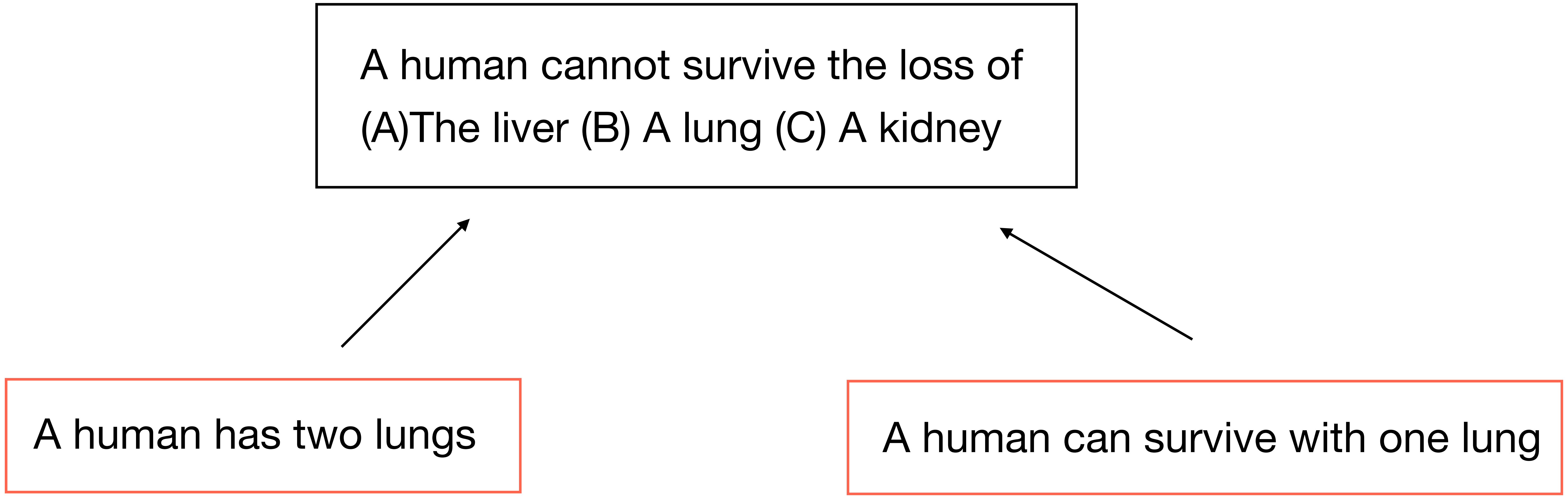
1. A person can build something over a period of time **T**

2. Millions of years is a period of time **T**

3. A person can build something over millions of years. **F**

Towards more consistent knowledge: REFLEX

Error pattern: Missing Rule



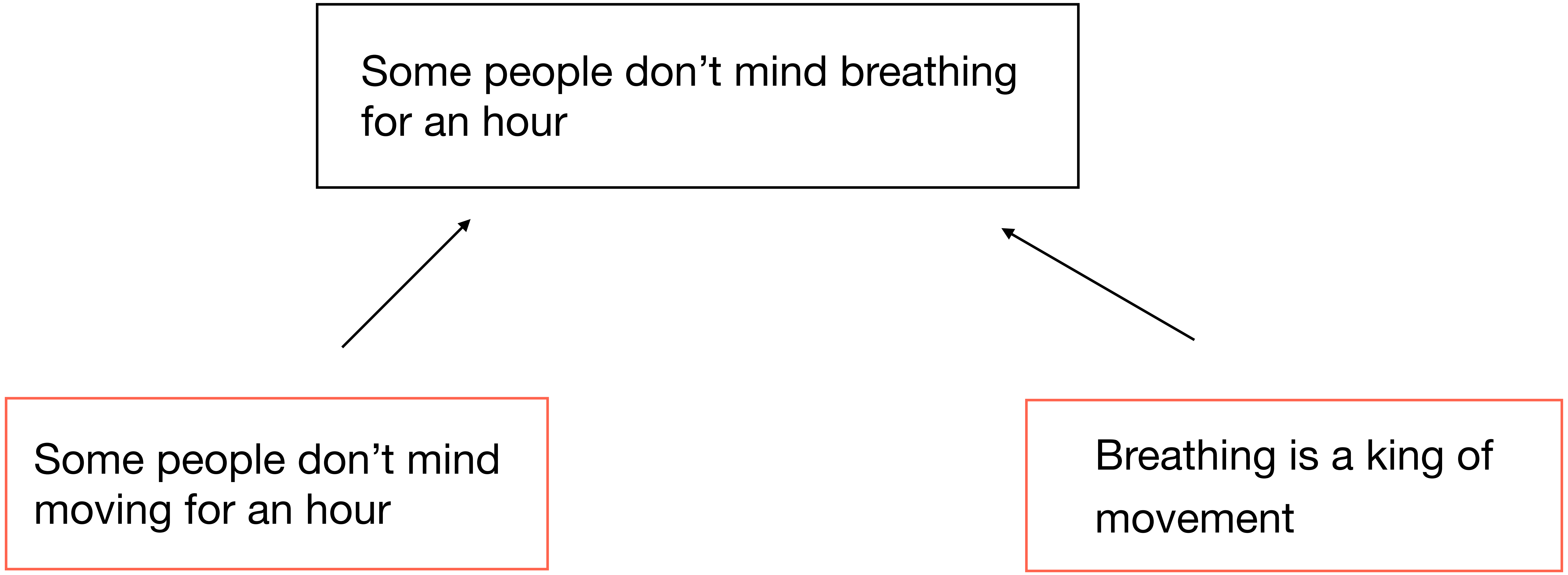
A human cannot survive the loss of (A)The liver (B) A lung (C) A kidney

A human has two lungs

A human can survive with one lung

Towards more consistent knowledge: REFLEX

Error pattern: Wrong Rule



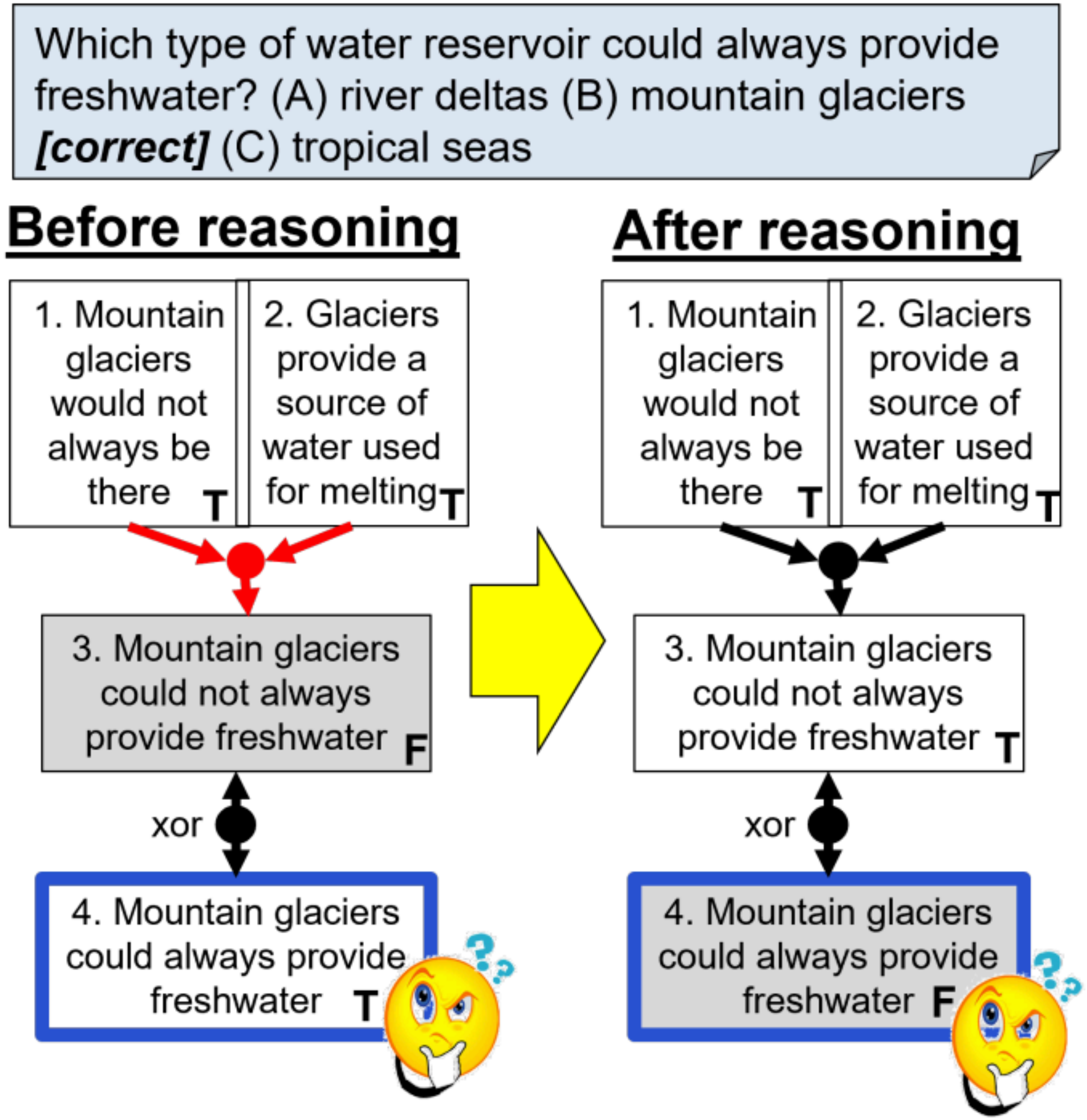
Some people don’t mind breathing for an hour

Some people don’t mind moving for an hour

Breathing is a king of movement

Towards more consistent knowledge: REFLEX

Error pattern: Unexpected Rule



Which type of water reservoir could always provide freshwater? (A) river deltas (B) mountain glaciers ***[correct]*** (C) tropical seas

Before reasoning

1. Mountain glaciers would not always be there **T**

2. Glaciers provide a source of water used for melting **T**

3. Mountain glaciers could not always provide freshwater **F**

xor

4. Mountain glaciers could always provide freshwater **T**

After reasoning

1. Mountain glaciers would not always be there **T**

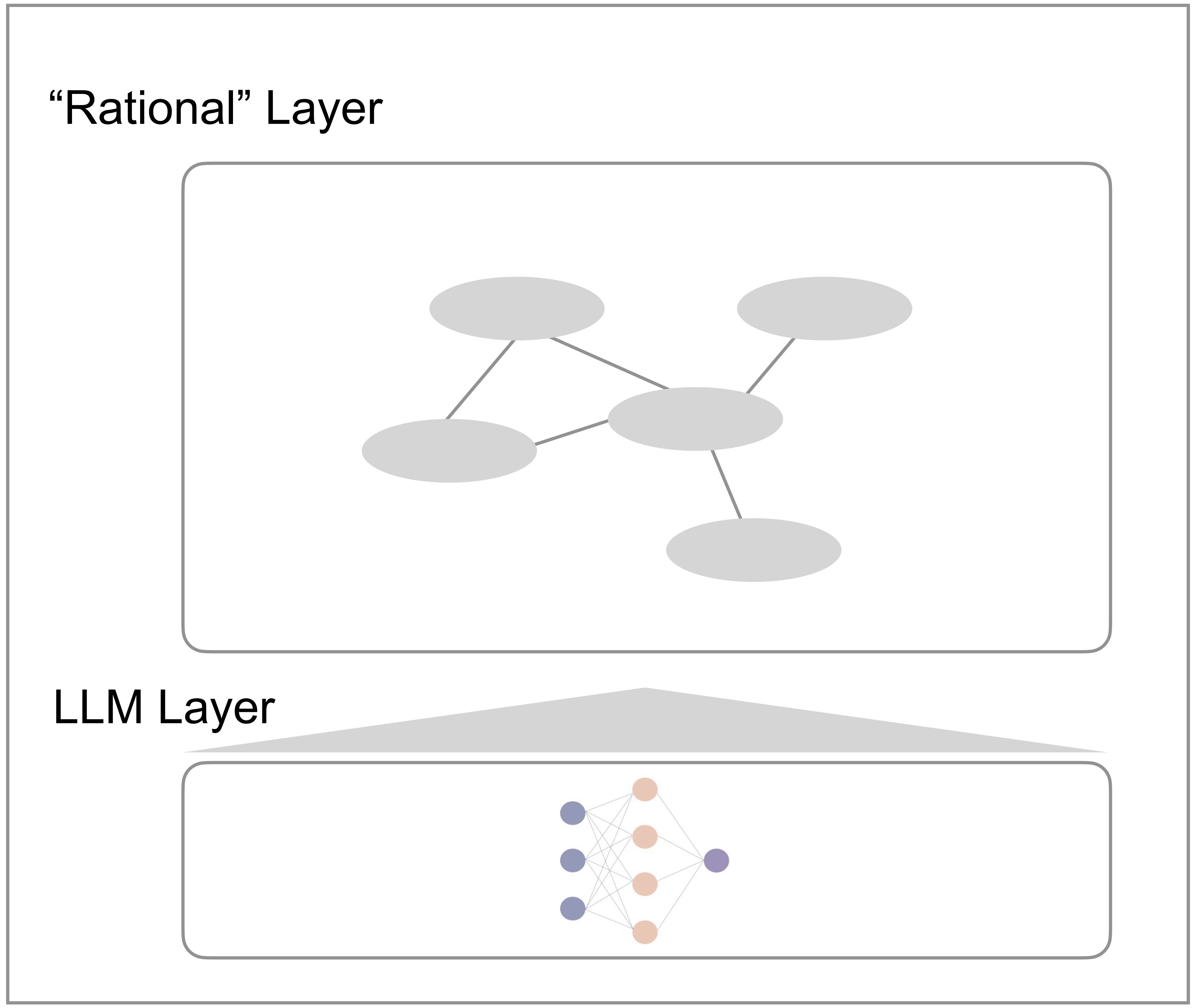
2. Glaciers provide a source of water used for melting **T**

3. Mountain glaciers could not always provide freshwater **T**

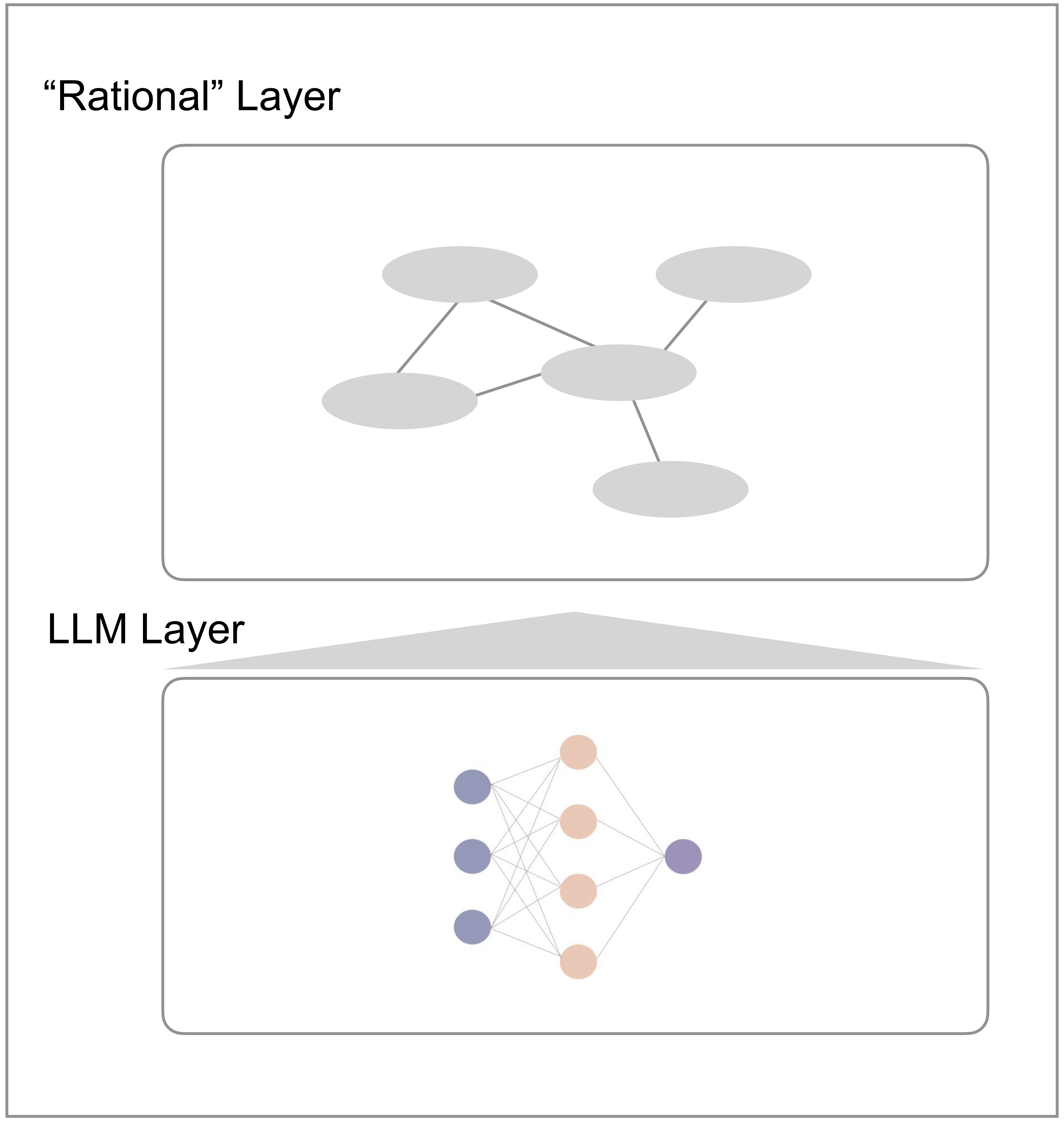
xor

4. Mountain glaciers could always provide freshwater **F**

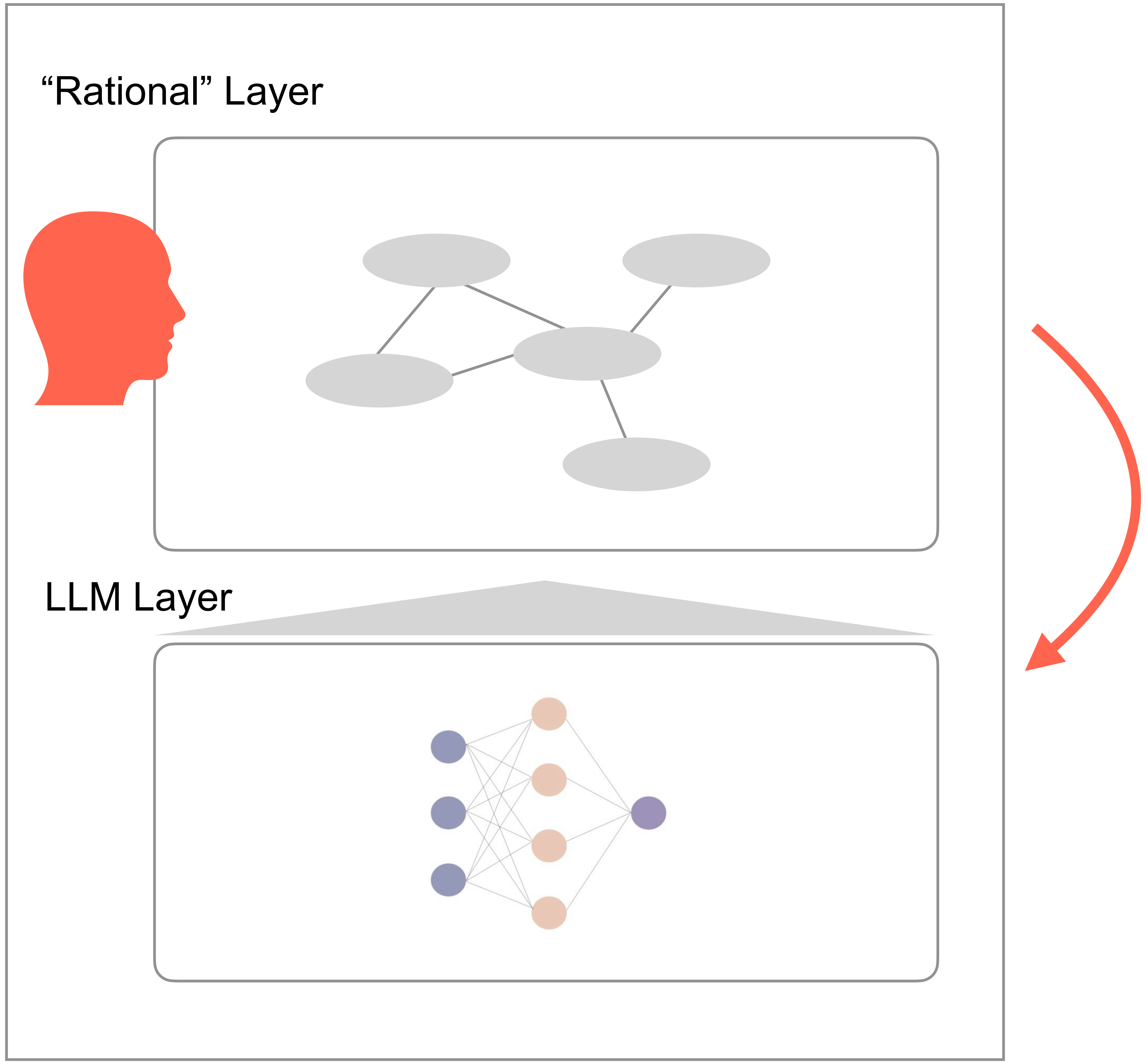
Towards more consistent knowledge: REFLEX



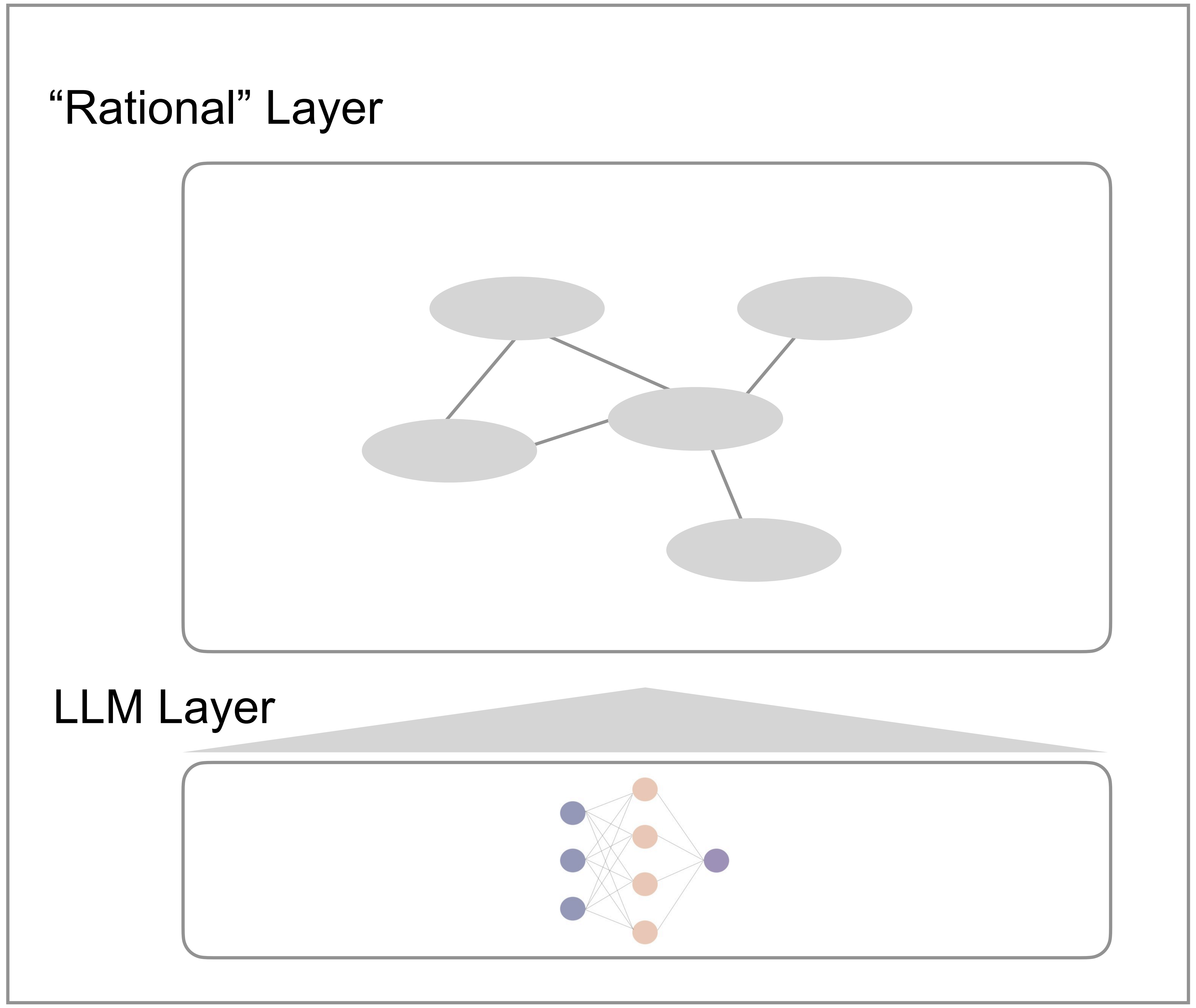
Towards more consistent knowledge: REFLEX



Towards more consistent knowledge: REFLEX



Structured Knowledge in Language Models



Architecture that constructs consistent and interpretable world models from Language Models