Breakpoint Transformers for Modeling and Tracking Intermediate Beliefs

Kyle Richardson[1,*], **Ronen Tamari**[2,*], Oren Sultan[2], Reut Tsarfaty[3], Dafna Shahaf[2], Ashish Sabharwal[1]

Allen Institute for AI[1], Hebrew University of Jerusalem[2], Bar-Ilan University[3]

*equal contribution

EMNLP 2022

Tracking the beliefs of model

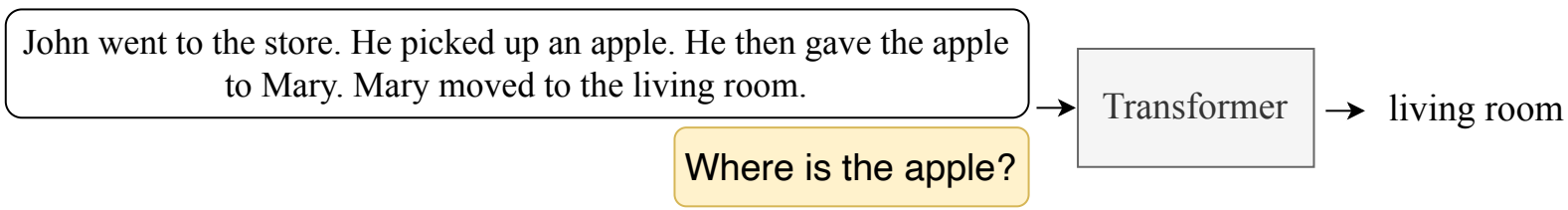
Narrative Understanding

John went to the store. He picked up an apple. He then gave the apple to Mary. Mary moved to the living room.

Where is the apple?

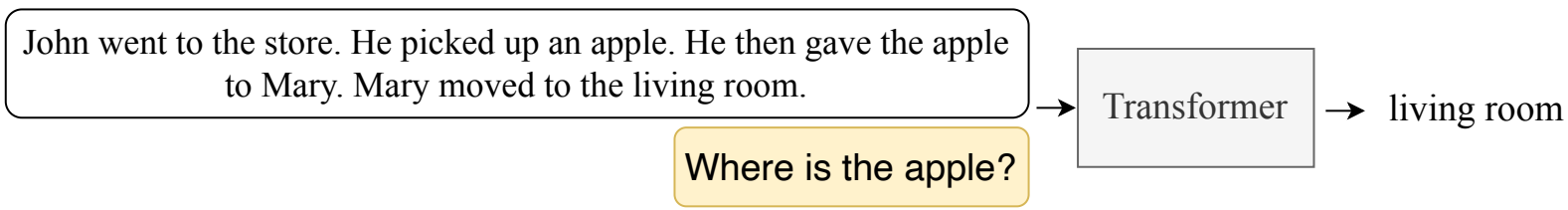
Tracking the beliefs of model

Narrative Understanding



Tracking the beliefs of model

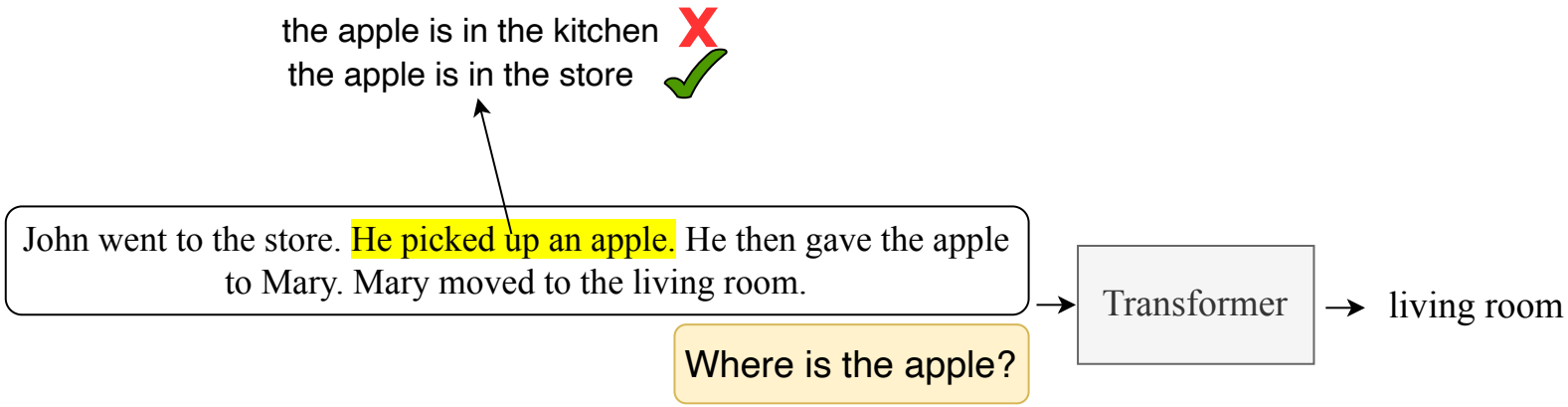
Narrative Understanding



Desiderata: Examining and tracking beliefs of models at arbitrary points in texts, posing textual queries.

Tracking the beliefs of model

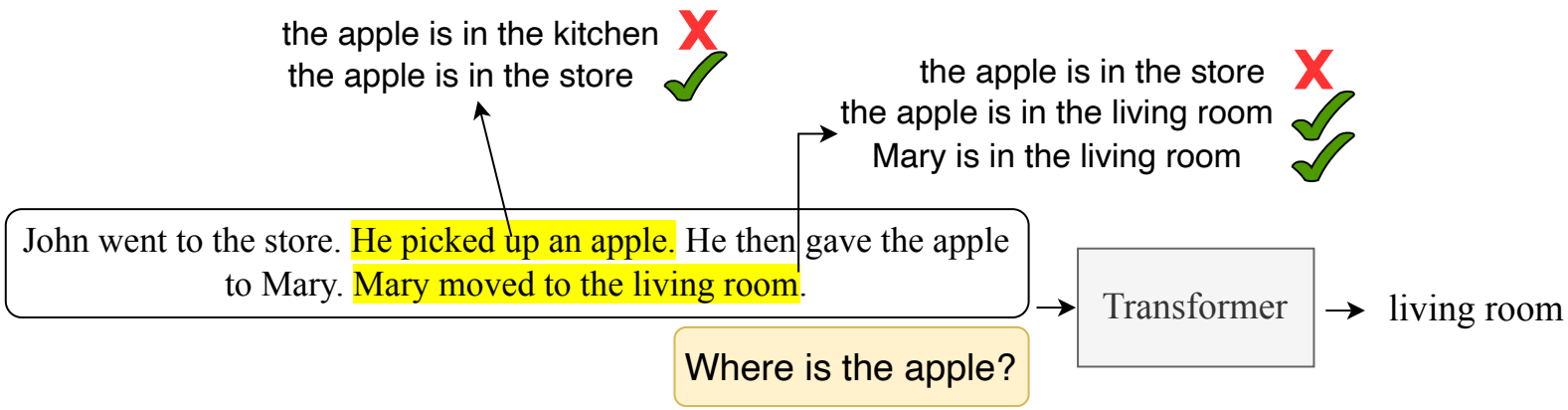
Narrative Understanding



Desiderata: Examining and tracking beliefs of models at arbitrary points in texts, posing textual queries.

Tracking the beliefs of model

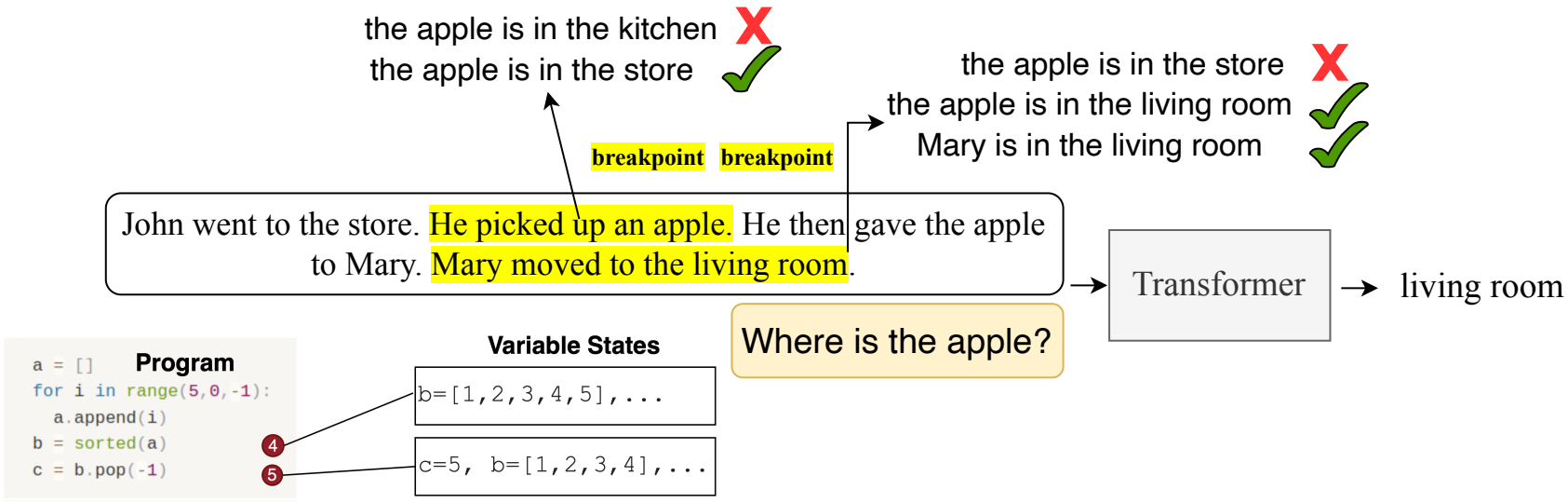
Narrative Understanding



Desiderata: Examining and tracking beliefs of models at arbitrary points in texts, posing textual queries.

Tracking the beliefs of model

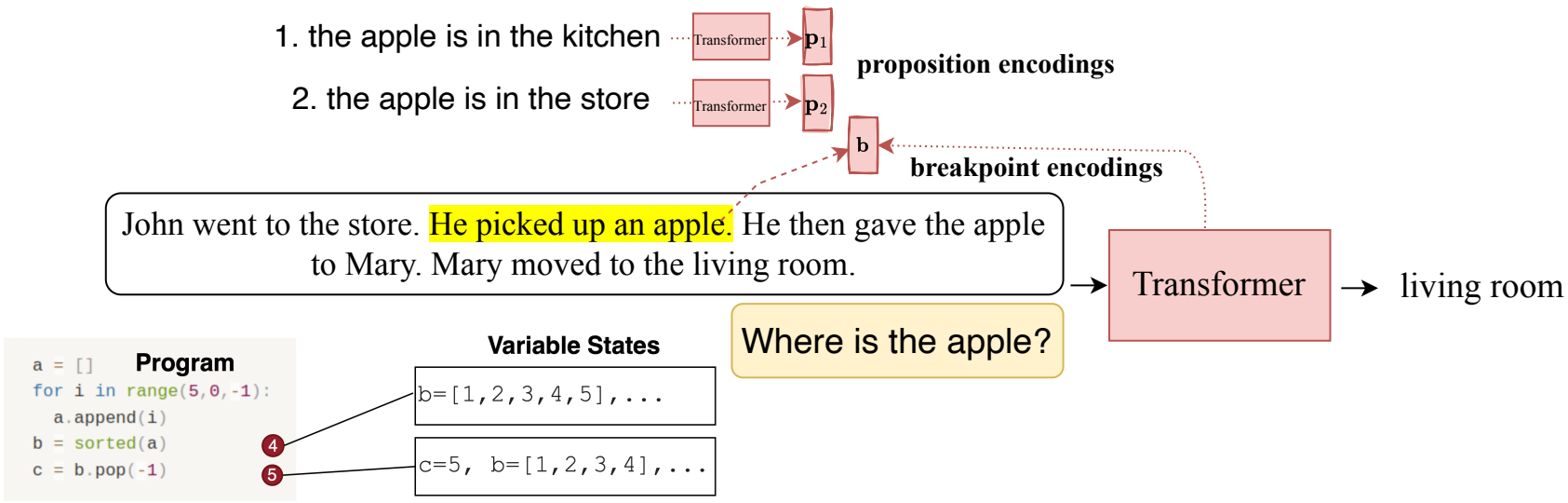
Narrative Understanding



Why: Reasoning about correctness of model behavior, track information change through time, find errors.

Tracking the beliefs of model

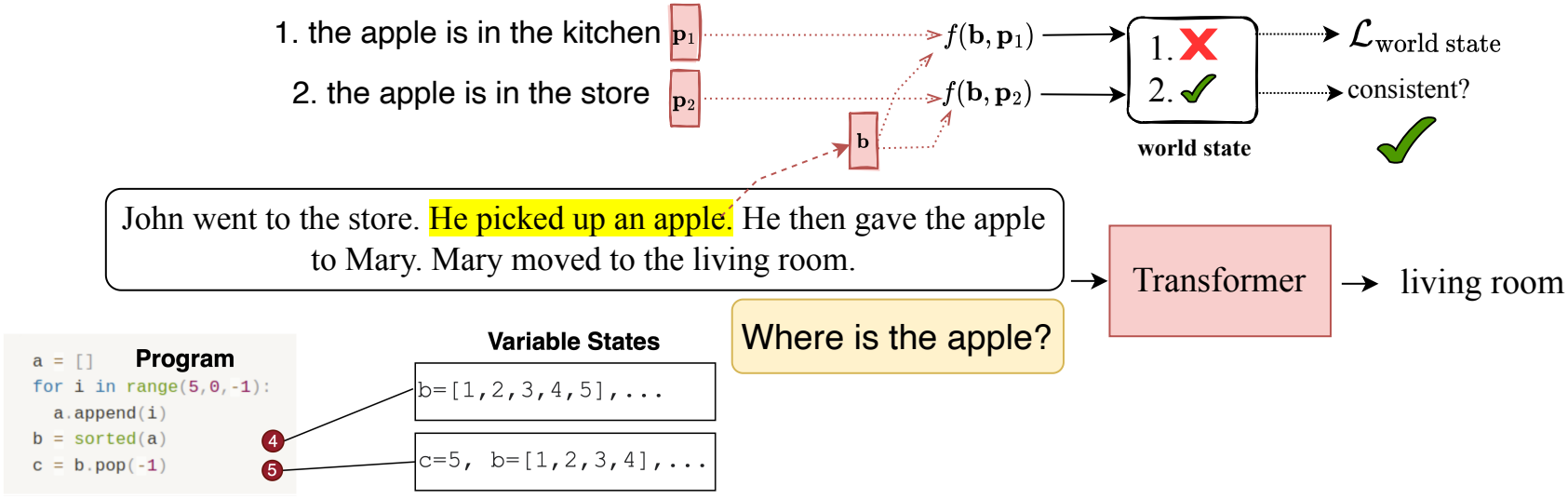
Narrative Understanding



Breakpoint Models: representation learning framework, modeling intermediate states, constructing world state representations.

Tracking the beliefs of model

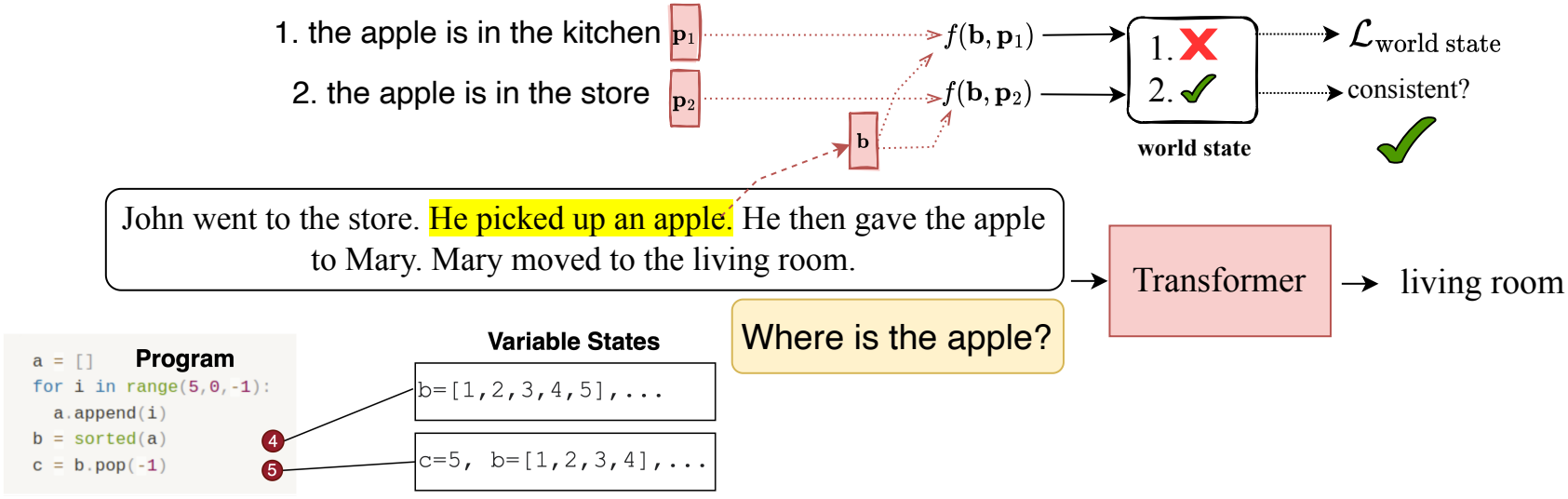
Narrative Understanding



Breakpoint Models: representation learning framework, modeling intermediate states, constructing world state representations.

Tracking the beliefs of model

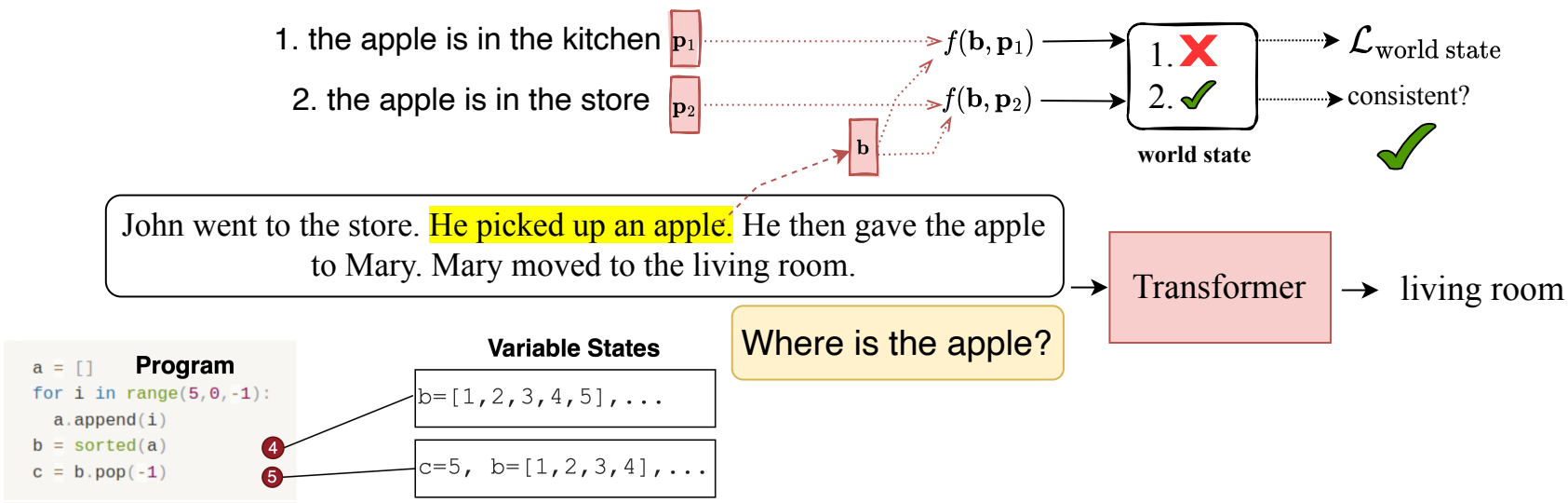
Narrative Understanding



Breakpoint Models: representation learning framework, modeling intermediate states, constructing world state representations.

Tracking the beliefs of model

Narrative Understanding

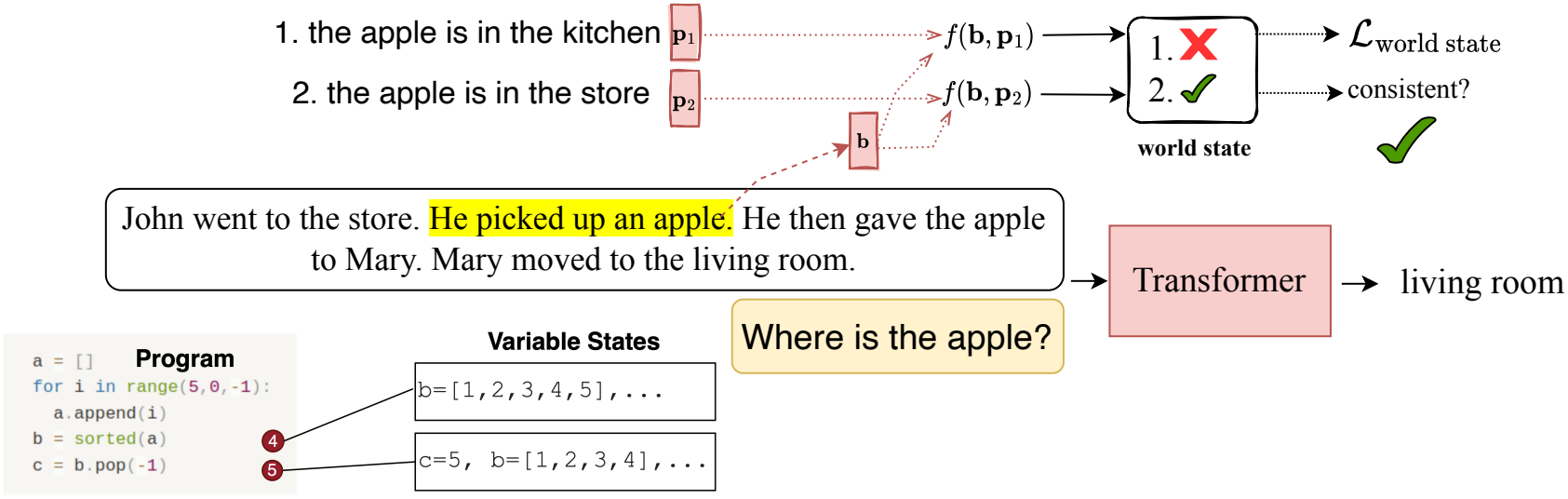


New belief prediction tasks, *breakpoint transformer* based on T5.

Task	Example Stories	Breakpoint Propositions
Relational Reasoning (CLUTRR)	John is the brother of Susan $[\mathbf{B}]_1$ Susan's mother is Janice $[\mathbf{B}]_2$, ...	$\mathbf{P}_1$: { 'Susan is the sister of John' **true**, 'Susan is the sister-in-law of Janice' **false**,'Janice is the mother of John' unk } $\mathbf{P}_2$: { 'Janice is the mother of John' **true**, 'John is the father of Janice' **false**, ...}
Story Understanding (bAbI)	John moved to the kitchen $[\mathbf{B}]_1$ He picked up an apple $[\mathbf{B}]_2$ John then gave the apple to Mary $[\mathbf{B}]_3$ ...	$\mathbf{P}_1$:{ 'John has the apple' **false**, 'John is in the kitchen' **true**,...} $\mathbf{P}_2$: { 'John has the apple' **true**,'John is in the kitchen'...}... $\mathbf{P}_3$: { 'John has the apple' **false**, 'Mary has the apple' **true** }
Commonsense (TRIP)	Tom dropped his radio ...carpet. $[\mathbf{B}]_1$ The radio broke .. $[\mathbf{B}]_2$ Tom turned on the radio ... $[\mathbf{B}]_3$ ...	$\mathbf{P}_2$: { 'radio is in pieces' **true**, 'radio is powered' **false**}, ... $\mathbf{P}_3$: { 'radio was powered' **true** }

Tracking the beliefs of model

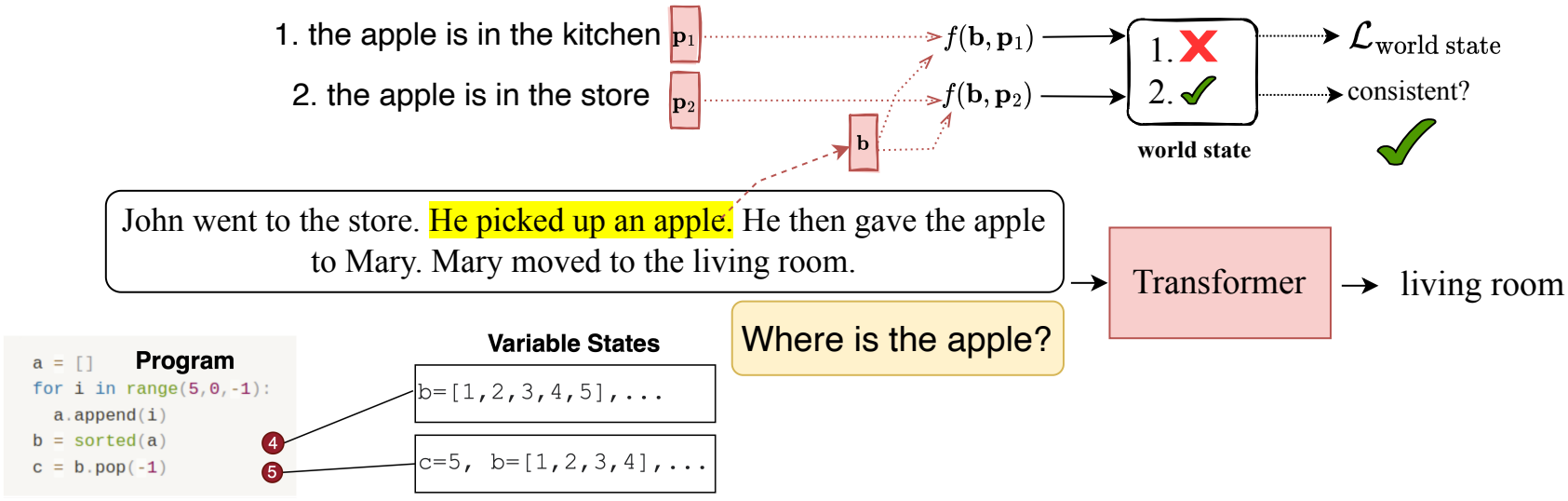
Narrative Understanding



Findings: Improved performance over conventional learning approaches, SOTA on TRIP task, joint training with existing tasks (QA).

Tracking the beliefs of model

Narrative Understanding



Please come to our poster to learn more.