Neural Reasoning about Agents' Goals, Preferences, and Actions

University of Stuttgart
Germany

PUI Perceptual User Interfaces
www.perceptualui.org

Matteo Bortoletto, Lei Shi, Andreas Bulling

Contact: matteo.bortoletto@vis.uni-stuttgart.de

European Research Council
Established by the European Commission

SimTech

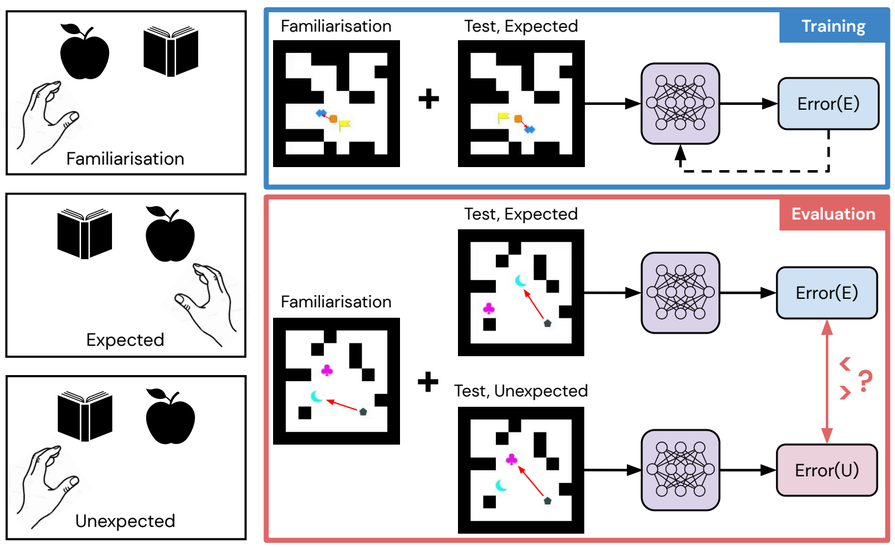
Motivation

- Common-sense reasoning has fundamental importance for human social cognition and behaviour.
- Studies have demonstrated that even young infants can understand basic facts about events, objects, beliefs, or desires [1].
- It is imperative that AI agents possess these capabilities to understand humans, and be understood.
- Limited progress in **intuitive psychology** compared to intuitive physics.

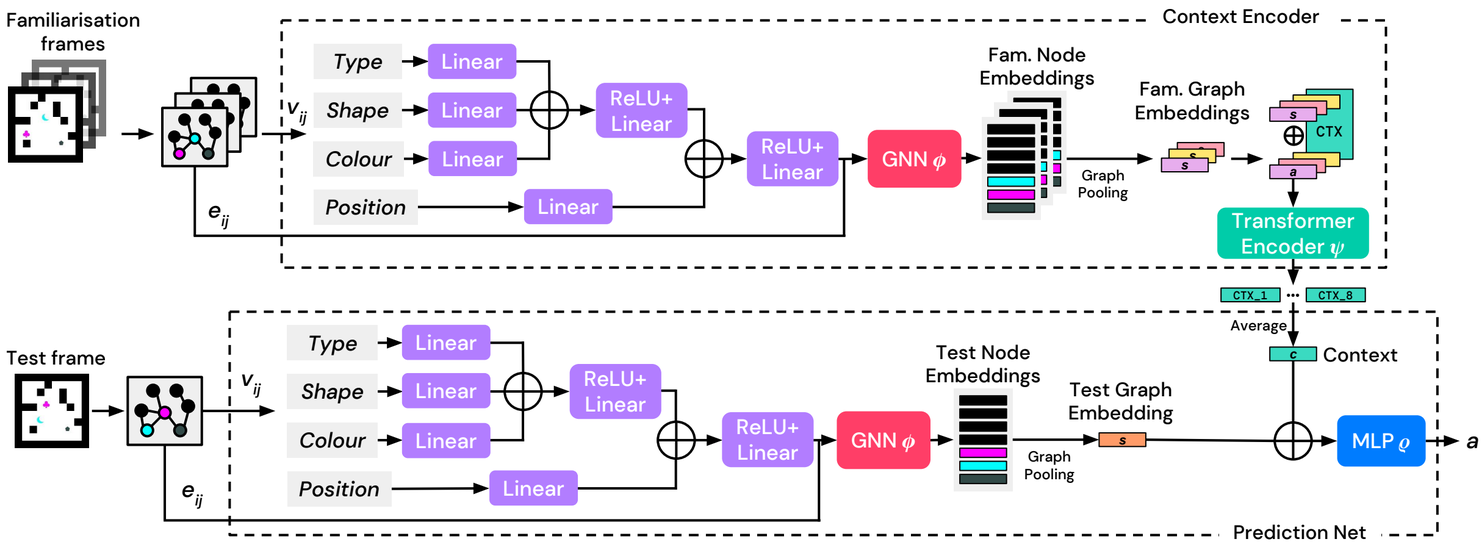
Contributions

- **IRENE, a novel model for intuitive psychological reasoning** that combines a GNN and a transformer to learn state and context representations.
- New **state-of-the-art performance on three out of five BIB tasks** [2]. IRENE is capable of binding preferences to specific agents and of modelling blocking barriers and irrational agents better than existing models.
- Analysis of the influence of BIB training tasks on performance. **IRENE can combine knowledge gained during training to solve unseen tasks**.

Violation of Expectation (VoE)



Intuitive Reasoning Network (IRENE)



Evaluations

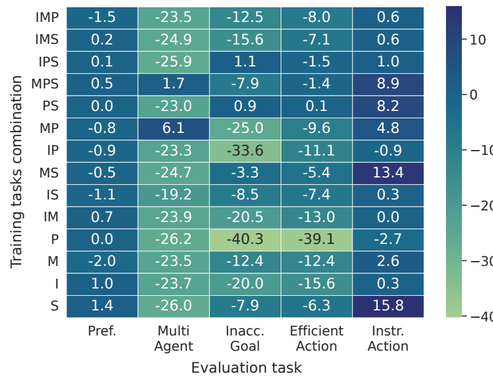
VoE Performance

BIB Task	BC-MLP	BC-RNN	Video-RNN	VT	IRENE
Preference	26.3	*48.3*	47.6	**80.8**	48.5
Multi-Agent	48.7	48.2	50.3	*49.2*	**74.9**
Inaccessible Goal	76.9	*81.6*	74.0	85.5	**85.8**
Eff. Path Control	94.0	92.8	**99.2**	*97.5*	98.1
Eff. Time Control	99.1	99.1	99.9	*99.7*	**100.0**
Eff. Irrational Agent	73.8	*56.5*	50.1	34.1	**85.7**
Eff. Action Average	88.8	82.5	*83.1*	77.1	**94.7**
Inst. No Barrier	98.8	98.8	**99.7**	*97.9*	78.4
Inst. Incons. Barrier	55.2	78.2	*77.0*	**91.9**	52.4
Inst. Blocking Barrier	47.1	56.8	*62.9*	64.2	**83.5**
Inst. Action Average	67.0	*77.9*	79.9	**84.7**	71.5

Ablation Experiments

BIB Task	LSTM	GCN	Local	Remote	IRENE
Preference	48.2	49.7	49.8	50.7	48.5
Multi-Agent	49.7	50.3	98.2	50.0	79.4
Inaccessible Goal	84.8	58.1	41.1	80.6	85.8
Eff. Path Control	97.3	94.7	31.7	98.2	98.1
Eff. Time Control	99.9	98.5	37.6	99.8	100.0
Eff. Irrational Agent	52.4	89.3	99.7	83.6	85.7
Eff. Action Average	83.2	94.2	56.3	93.9	94.7
Inst. No Barrier	78.5	64.6	51.6	78.7	78.4
Inst. Incons. Barrier	53.3	52.1	52.4	52.7	52.4
Inst. Blocking Barrier	83.2	48.0	48.9	83.8	83.5
Inst. Action Average	71.7	54.8	51.0	71.7	71.5

Analysis of Training Tasks



- Largest improvements are in the *Multi-Agent*, *Instrumental Blocking Barrier* and *Efficiency Irrational Agent* tasks.
- VoE scores on *Instrumental No Barrier* and *Inconsequential Barrier* are lower than those of other methods. Models that learn the simple heuristic of ignoring barriers are more effective.
- Remote relations contribute more to the final scores than local ones; effectiveness of combining GraphSAGE and transformer.
- Training on one type of task does not always improve performance for similar types of tasks in the evaluation set; IRENE performs best when combining knowledge from all training tasks.

[1] Needham, Amy, and Renee Baillargeon. "Intuitions about support in 4.5-month-old infants." *Cognition* 47.2 (1993): 121-148.

[2] Gandhi, Kanishk, et al. "Baby Intuitions Benchmark (BIB): Discerning the goals, preferences, and actions of others." *Advances in neural information processing systems* 34 (2021): 9963-9976.

AAAI Association for the Advancement of Artificial Intelligence