Preconditioned Inference in Commonsense Knowledge Reasoning

Ehsan (Eric) Qasemi

Dec 2nd 2022

Overview

- Motivation & Background
  - What is Commonsense Knowledge?
  - What are Preconditions of Common Sense?
- PaCo: Preconditions Attributed to Common Sense Knowledge (EMNLP 2022)
- PInKS: Preconditioned CS Inference with Minimal Supervision (AACL 2022)
- Future/Ongoing Research Directions
  - Visual Commonsense motivation
  - VIPHY: Probing "Visible" Physical Commonsense Knowledge (Under Rev-EACL 2023)
  - Preconditioned Visual Language Inference with Weak Supervision (Under prep-ACL 2023)

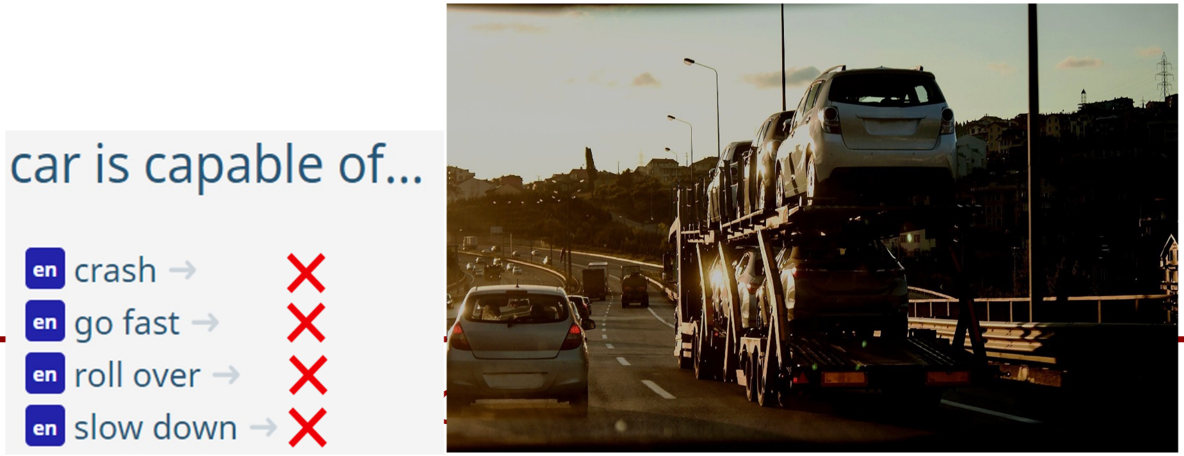
An Example and Some Definitions

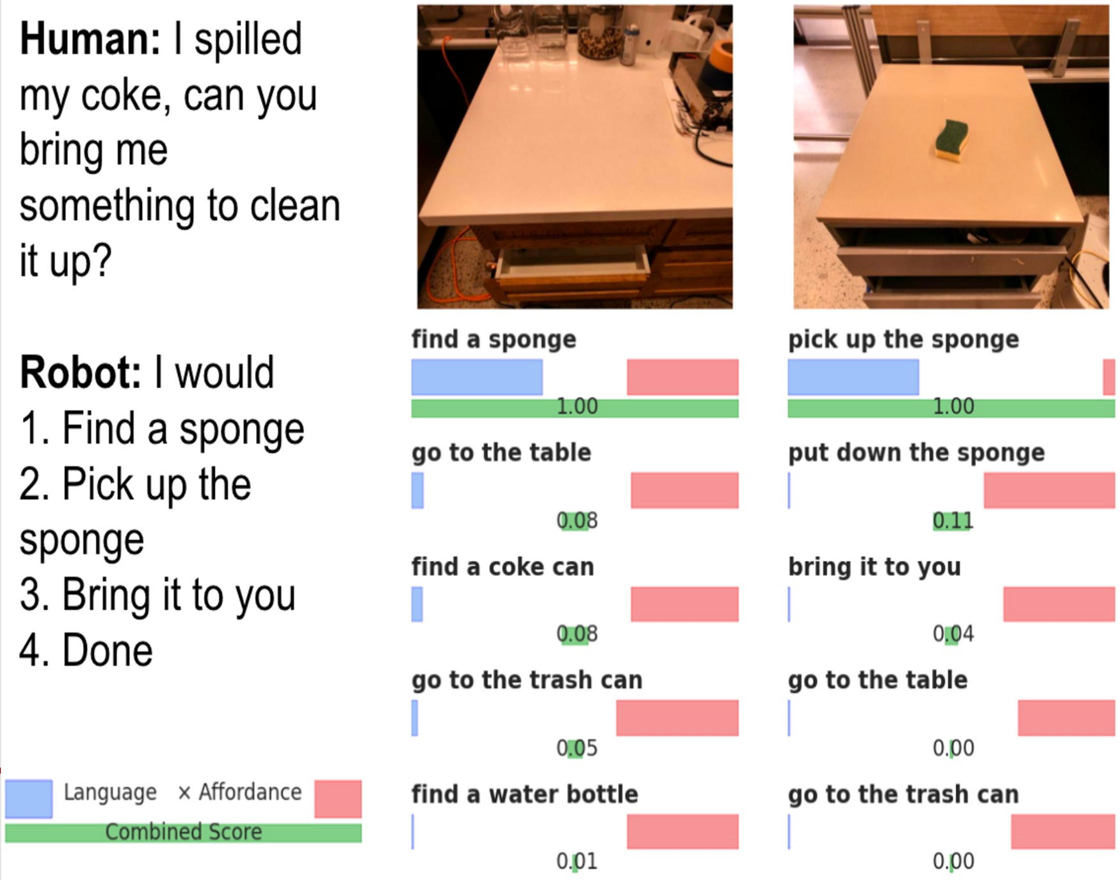
- CommonSense Statements:
  - A Glass is used for drinking water.
- Preconditions that Enable the statement
  - You are able to drink
  - You have access to water
  - There is gravity
- Preconditions that Disable the statement
  - The glass is shattered
  - Water is toxic

Why Preconditions are Important?

- Theory of Affordance
  - Understanding the circumstances in which an action or fact is feasible or impossible is a key aspect of human intelligence.

  - In Practice:
    - Ahn, Michael, et al 2022. https://arxiv.org/abs/2204.01691
      - "... generally, with failures as a result of affordance function misclassification"





PaCo: Preconditions Attributed to CS Knowledge

- Systematic study on the problem of preconditions expressed in natural language
- Develop PaCo, a rich crowdsourced dataset with enabling and disabling preconditions of commonsense statements
- Extensive NLP benchmarking based on PaCo.
  - Natural Language Inference
  - Multiple-Choice Question Answering
  - Generation

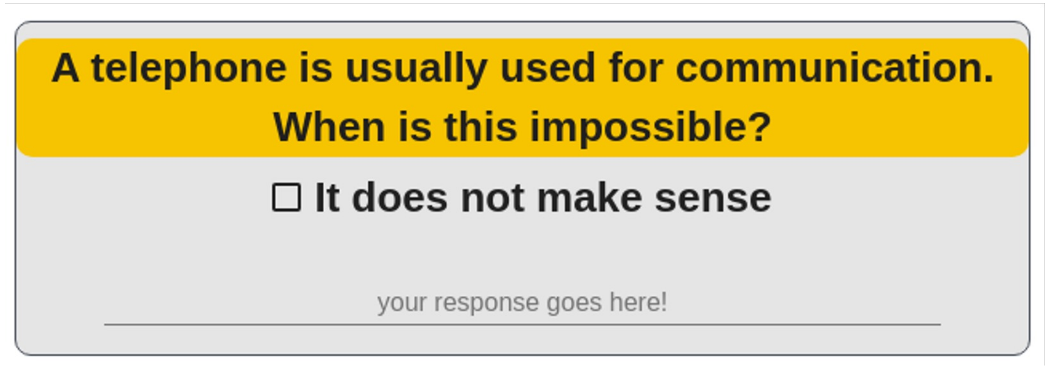
Preconditioned Natural Language Inference

- Paco (our work):
  "You are in a desert" Disables "A net is used for catching fish"
  - Hard preconditions (Enable/Disable) (Causes)
  - Relaxed statement and update format
- **δ**-NLI:
  "The men are facing their granary" strengthens "Two men and a dog are standing among rolling green hills. So the men are farmers"
  - Soft preconditions (strengthen/weaken) (Tends to Cause)
  - Fixed two sentence structure on statement and update

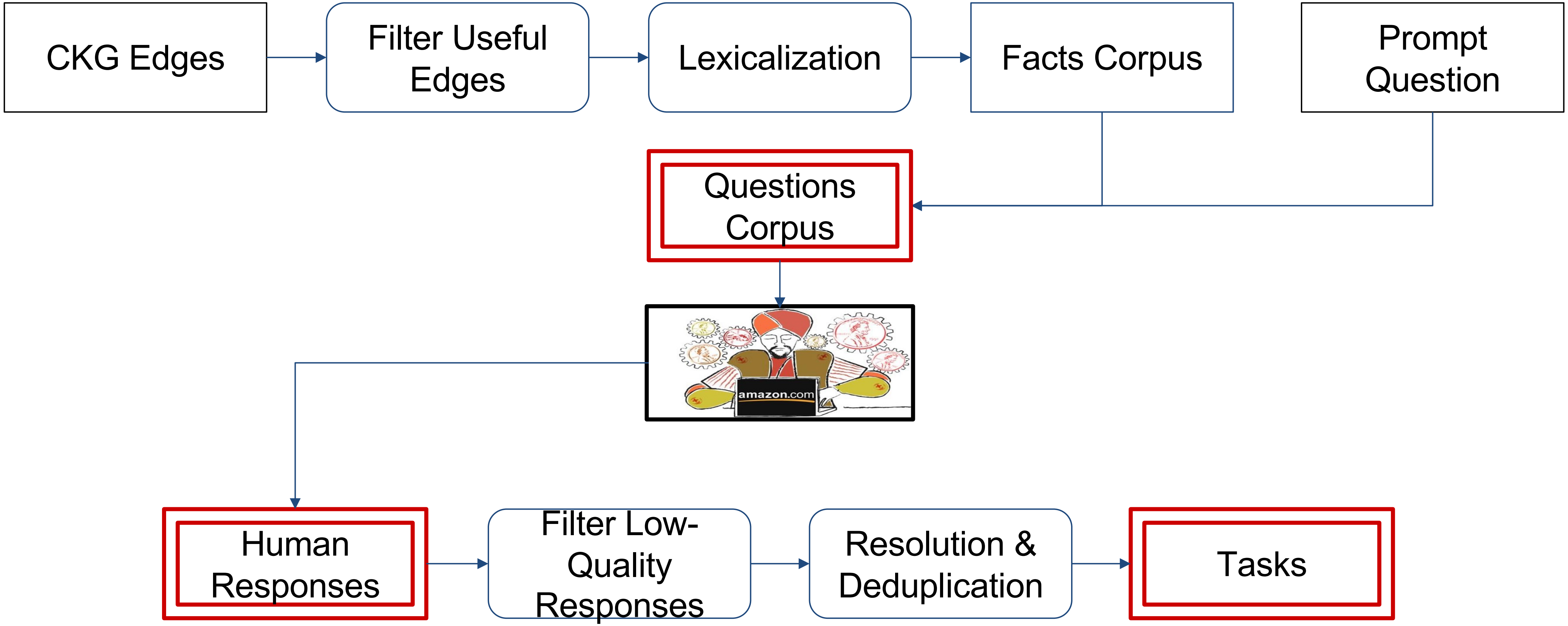
Collecting PaCo

- Statements from ConceptNet
  - Not all statements lend themselves naturally for preconditions annotation
    - E.g. “RelatedTo” is underspecified, and “IsA” is truisms.
  - Three representative types
    - Utility: “UsedFor”
    - Temporal: “Causes”
    - Motivational: “Desires”

- Distantly Supervised Quality Control



Collecting PaCo



Tasks

- Natural Language Inference (P-NLI)
  - Simplest case
- Multiple-Choice Question Answering (P-MCQA)
  - Discriminative without annotation artifacts
  - More Challenging
- Precondition Generation (P-G)
  - The ultimate challenge

ID	**Instance**
P-NLI	*Hypothesis*: *A net is used for catching fish* *Premise*: *We are in a desert* *Label*: *Contradiction*
P-MCQA	*Question*: *A net is used for catching fish. When is this impossible?* *Choices*: *(A) You are in sea, (B) The boat is moving, (C) Net has a large hole in it.*
P-G	*Question*: *A net is used for catching fish. When is this impossible?* *References*: *(-) Net has a large hole in it, (-) You are in downtown LA, (-) There are no fish in the water*

Table 2: Example of the three tasks in *PaCo*.

Evaluation Results: P-NLI

- Key observations:
  - Even the simple NLI data is not trivial in zero-shot setup

AllenNLP TE	0.34	0.85
RoBERTa-large-MNLI	0.47	0.90
BART-large-MNLI	0.48	0.90
DeBERTa-base-MNLI	0.37	0.91
DeBERTa-large-MNLI	0.36	0.94
DeBERTa-xl-MNLI	0.37	0.91
Expert Human	0.99	-
Random Baseline	0.5	-

Table 3: F1-Macro results of SOTA systems on P-NLI

Evaluation Results: P-MCQA

- Key observations:
  - Significant gap between the human and machine performance
  - LMs lack understanding of linguistic permutations like negations.
  - Large models tend to confuse the enabling v.s. disabling cases.

Model	**0-Shot**	**Tuned**
RoBERTa-base	0.24	0.42
RoBERTa-large	0.22	0.22
UnifiedQA-small	0.32	0.50
UnifiedQA-base	0.23	0.59
UnifiedQA-large	0.28	0.68
Expert Human	0.92	-
Random Baseline	0.25	-

Table 4: Accuracy results of SOTA systems on P-MCQA task based on *PaCo*. Best values are highlighted.

Evaluation Results: P-G

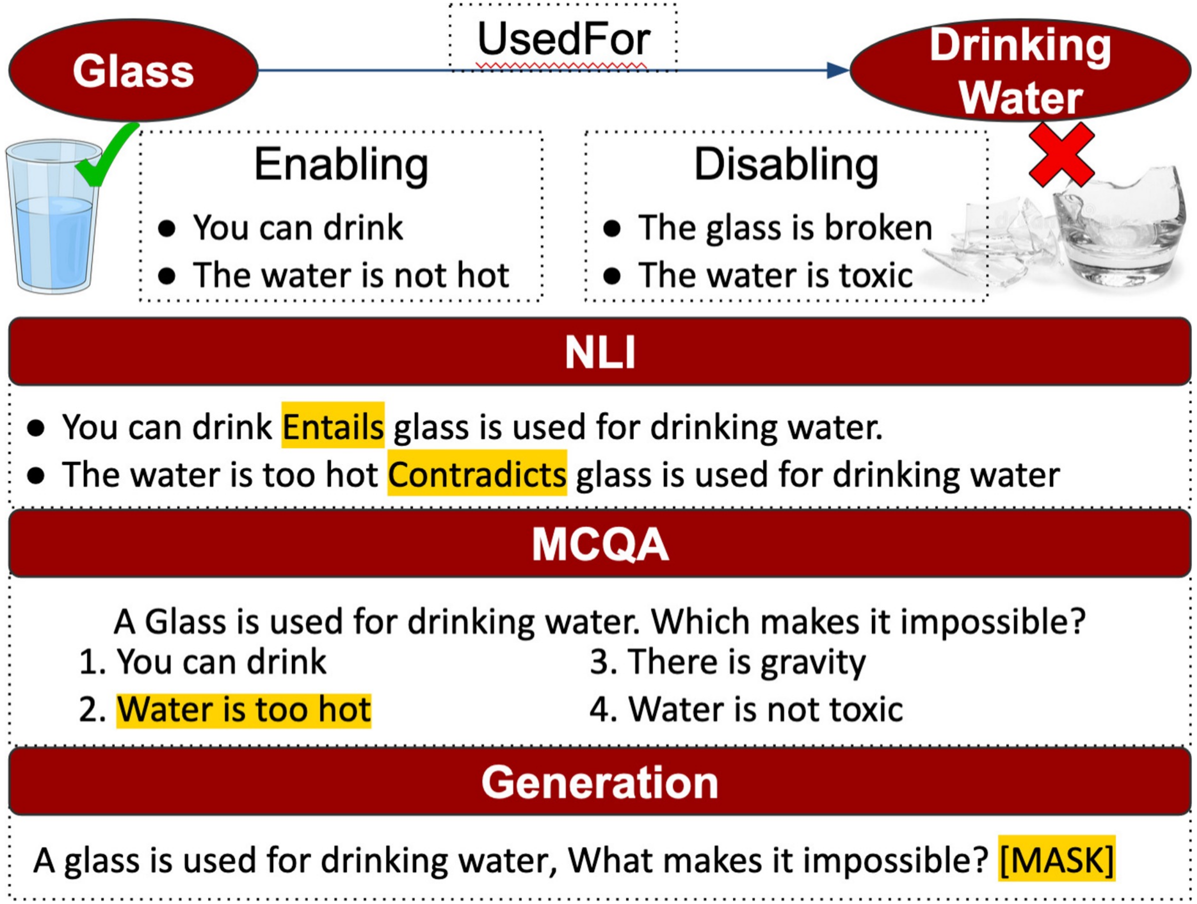
- Key Observations:
  - Fine-tuned models are underperforming
  - Automatic evaluation methods do not sufficiently distinguish between the models
    - Small number of reference responses
    - large space of correct responses
  - Models tend to generate simple answers
    - existence or availability of the subject.

Model	BLEU 0-Shot	BLEU Tuned	ROUGE Tuned	HUM Info.
UnifiedQA-small	0.007	0.157	0.064	0.12
UnifiedQA-base	0.006	0.303	0.115	0.28
UnifiedQA-large	0.029	0.330	0.128	0.48
BART-base	0.046	0.091	0.140	0.19
BART-large	0.041	0.058	0.117	0.11
GPT2	0.097	0.133	0.067	0.36
Expert Human	-	-	-	1.0

Table 5: BLEU-2, ROUGE-2, and human evaluation Information score for results of SOTA systems on the P-G task. Zero-shot ROUGE scores are omitted to save space as they are negligible and do not add additional insight beyond the zero-shot BLEU-2. Best values are highlighted.

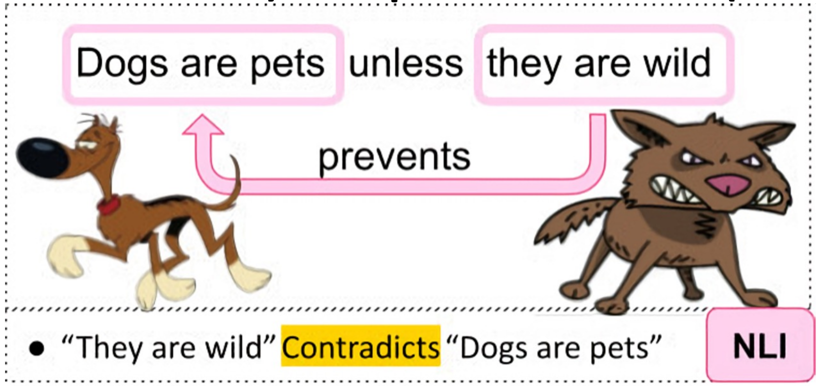
Conclusion of PaCo

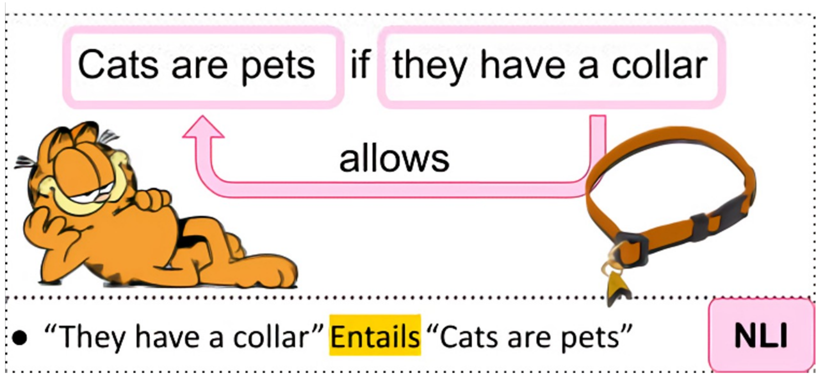
- We saw
  - Reasoning with preconditions
  - PaCo: Data to foster research
  - Evaluation of SOTA
- Challenges:
  - Insufficient annotated data
  - Need for improved LMs



Preconditioned Natural Language Inference

- Conventional NLI: Given a sentence pair composed of a *hypothesis* and a *premise*, the system has to decide whether the hypothesis is true (entailment), false (contradiction), or undetermined (neutral) given the premise
- P-NLI :Given a common sense statement (hypothesis) and an update sentence (premise) that serves as precondition, is the statement still allowed (entailment) or prevented (contradiction)?





Reviewed by ACL Rolling Review (ARR) with meta score of 4/5, committed to EMNLP 2022

Why LLMs are bad at Preconditioned Inference?

- Insufficient annotated data
  - We rarely write commonsense statements and their preconditions
- Need for improved LMs
  - SOTA is trained on raw text with no incentive to focus on preconditions

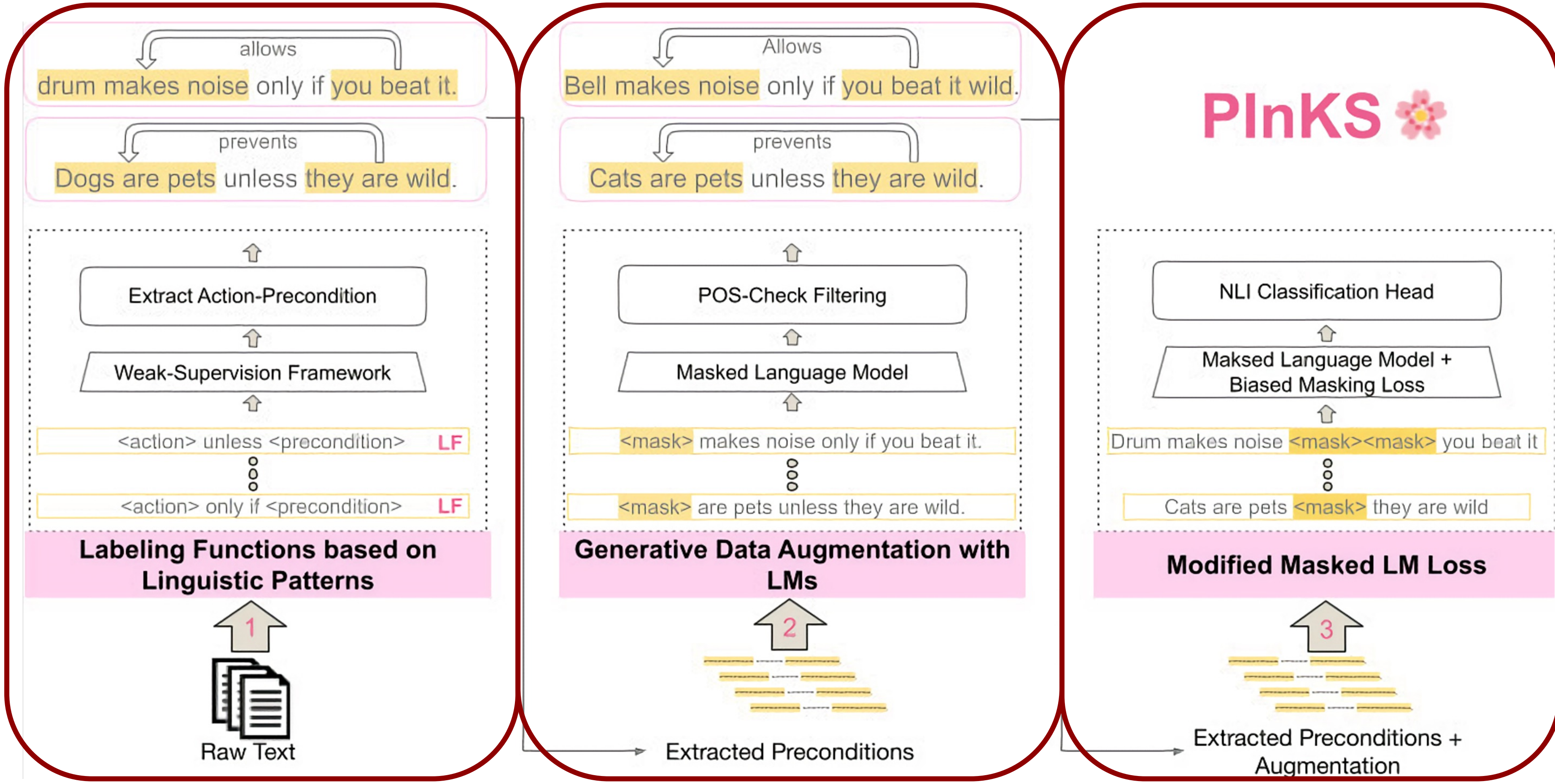
PInKS🌸: Preconditioned Commonsense Inference with Minimal Supervision

1. Weak supervision for obtaining noisy preconditions data
2. Modify masked LM loss function to emphasize on preconditions
3. Empirical improvements and theoretical informativeness guarantees

In submission to AACL 2022

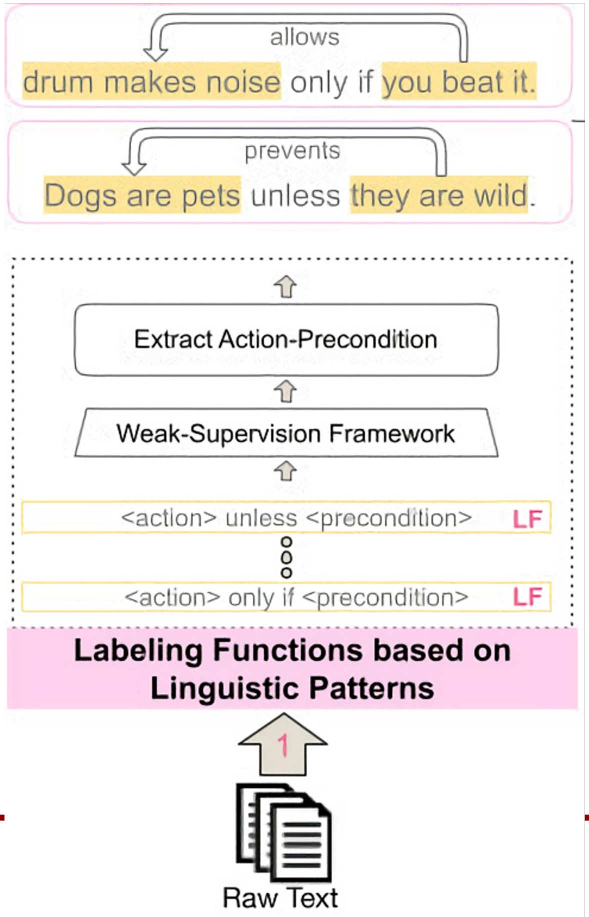
Improved Model + Weak Data

- Weak supervision for noisy training data
  - Linguistic pattern matching
  - Generative data augmentation
- Biased masking on conjunctions during fine-tuning



Weak Supervised Data from linguistic patterns

- Linguistic patterns based on conjunctions applied on raw corpora
  - 20+ Patterns (Labeling Functions)
    e.g. “<action> unless <precondition>”
  - OMCS and ASCENT for raw corpora



Text	Label	Action	Precondition
A drum makes noise only if you beat it.	Allow	A drum makes noise	you beat it.
Your feet might come into contact with something if it is on the floor.	Allow	Your feet might come into contact with something	it is on the floor.
Pears will rot if not refrigerated	Prevent	Pears will rot	refrigerated
Swimming pools have cold water in the winter unless they are heated.	Prevent	Swimming pools have cold water in the winter	they are heated.

Table 1: Examples from the collected dataset through linguistic patterns in §3.1.

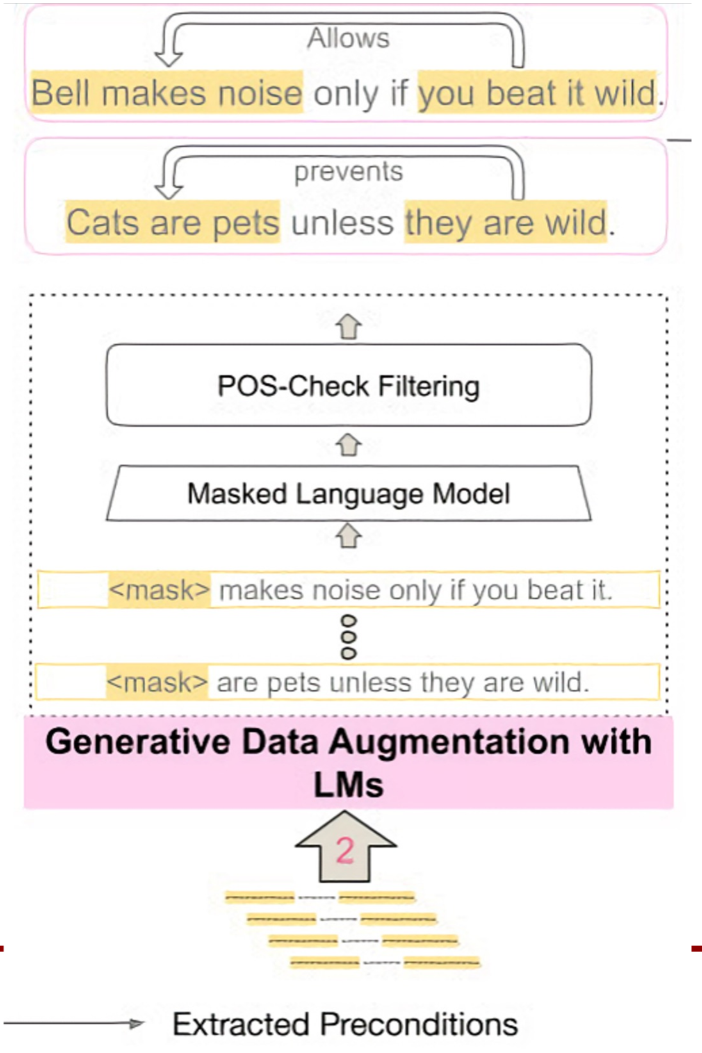
Labeling Functions base on Conjuctions

Type	Conjunctions
Allowing	only if, subject to, in case, contingent upon, given, if, in the case that, in case, in the case that, in the event, on condition, on the assumption, only if, so, hence, consequently, on these terms, subject to, supposing, with the proviso, so, thus, accordingly, therefore, as a result, because of that, as a consequence, as a result
Preventing	but, except, except for, excepting that, if not, lest, saving, without, unless

Table 3. List of conjunctions used in modified masked loss function in section 3.3

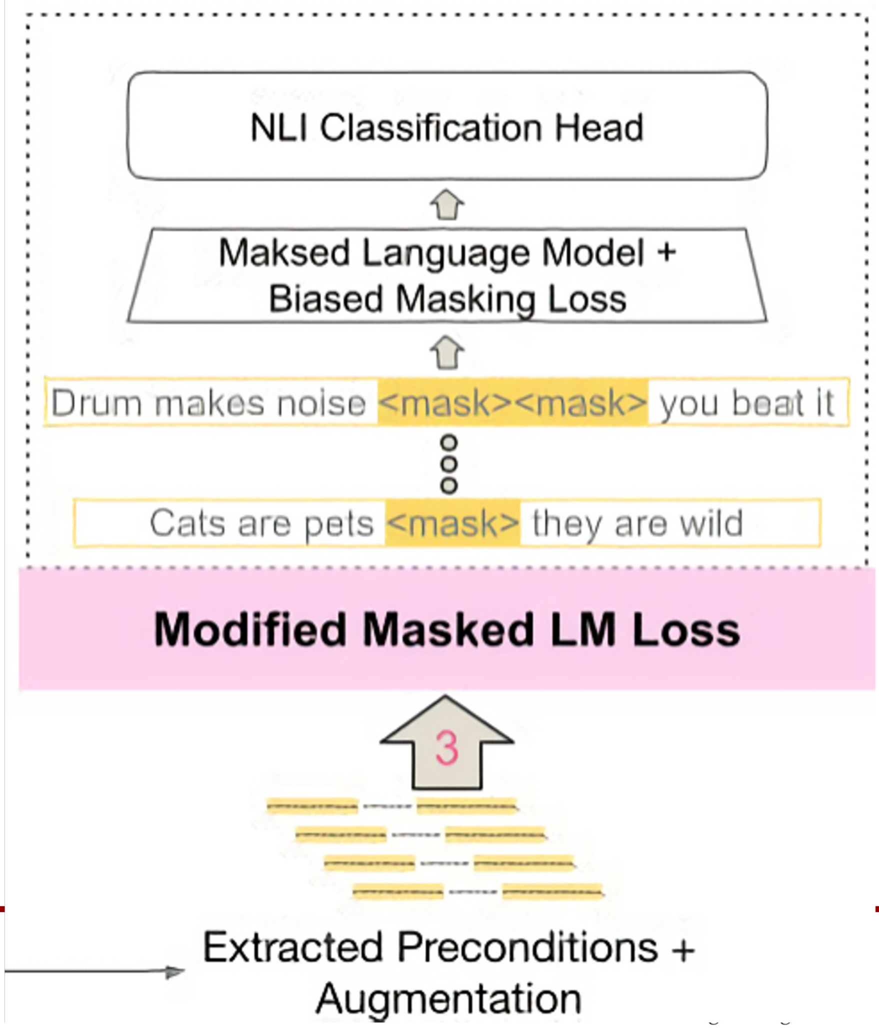
Weak Supervised Data from Generative Augmentation

- Generative data augmentation using masked language models
- POS Tag consistency as filtering mechanism
  - RoBERTa-base (<MASK> will rot if not refrigerated):
    - Food
    - ~~It (Filtered out based on POS tag)~~
    - Meat
    - Fish
    - Chicken



Modified Language Modeling

- Biased masking on conjunctions
  - Only mask conjunctions instead of random tokens
  - Additional fine-tuning on the weak supervised data (no labels) to fill mask
    - LM(Food will rot <mask> <mask> refrigerated)



Main Results: Experimental Setup

- Five target tasks in NLI canonical format derived from supervised resources on preconditions of commonsense
  - RoBERTa-Large-MNLI as model

Name	Original Data		Derived NLI	
Winoventi (Do and Pavlick, 2021)	**masked_prompt**:	Margaret smelled her bottle of maple syrup and it was sweet. The syrup is {MASK}.	**Hypothesis**:	Margaret smelled her bottle of maple syrup and it was sweet.
	target:	edible	**Premise**:	The syrup is edible/malodorous
	incorrect:	malodorous	**Label**:	ENTAILMENT/CONTRADICTION
ANION (Jiang et al., 2021)	**Orig_Head**:	PersonX expresses PersonX's delight.	**Hypothesis**:	Alice expresses Alice's delight/anger.
	Relation:	xEffect	**Premise**:	feel happy.
	Tail:	Alice feel happy	**Label**:	ENTAILMENT/CONTRADICTION
	Neg_Head:	PersonX expresses PersonX's anger.		
ATOMIC2020 (Hwang et al., 2020)	**Head**:	PersonX takes a long walk.	**Hypothesis**:	PersonX takes a long walk.
	Relation:	HinderedBy	**Premise**:	It is 10 degrees outside..
	Tail:	It is 10 degrees outside.	**Label**:	CONTRADICTION
δ-NLI (Rudinger et al., 2020)	**Hypothesis**:	PersonX takes a long walk.	**Hypothesis**:	PersonX takes a long walk.
	Premise:	HinderedBy	**Premise**:	It is 10 degrees outside..
	Update:	It is 10 degrees outside.	**Label**:	CONTRADICTION
	Label:	Weakener		
PaCo (Qasemi et al., 2022)	**Statement**:	A net is used for catching fish.	**Hypothesis**:	A net is used for catching fish.
	Precondition:	You are in a desert.	**Premise**:	You are in a desert.
	Label:	Disabling	**Label**:	CONTRADICTION

Table 8: Examples from target tasks in NLI format

Main Results

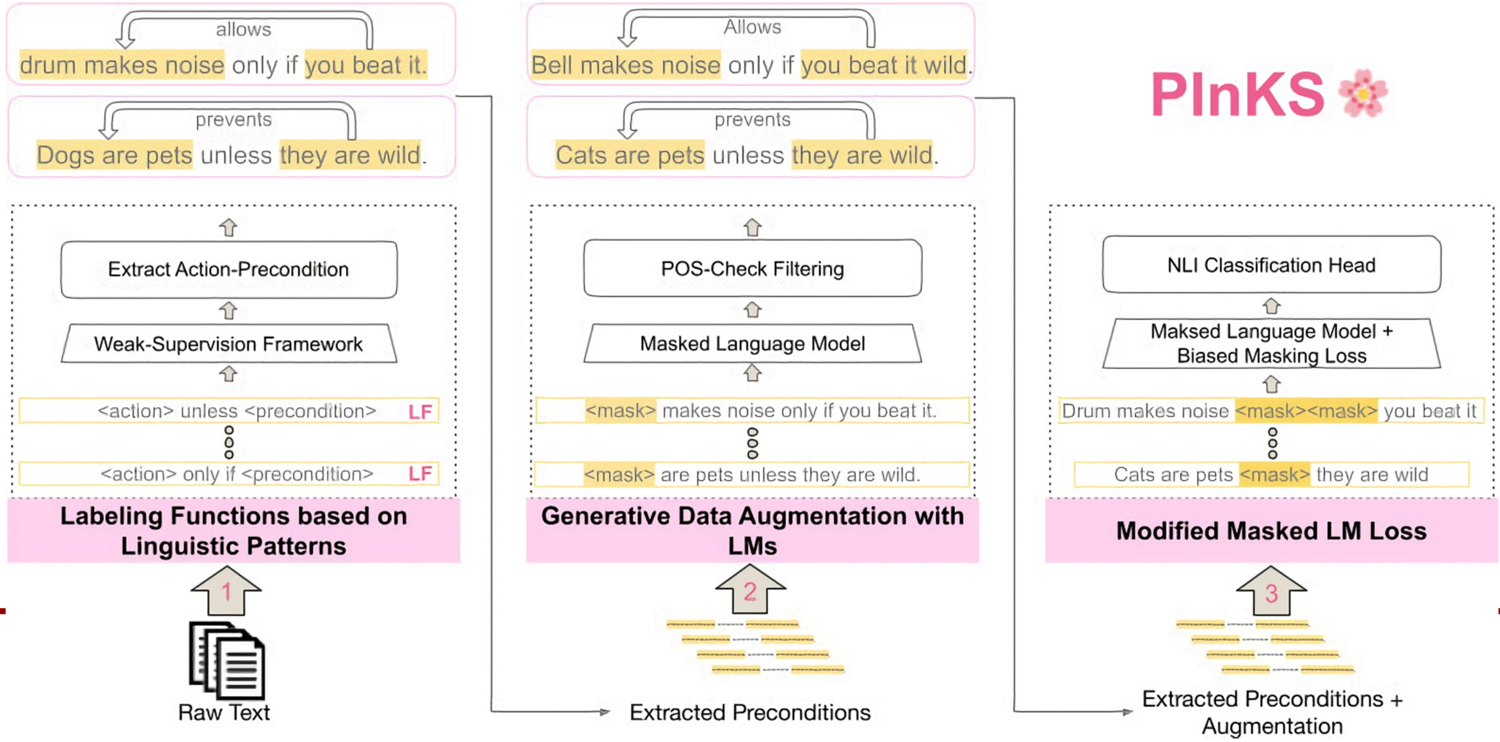
- Takeaway:
  - Exceeds the supervised results in zero-shot (Atomic, Winoventi)
  - Outperforms when combined with supervised signals (δ-NLI, ANION, ATOMIC)
  - No improvement on the hard preconditions (CoreQuisite/PaCo)
- Setups:
  - Orig.
    - tune on train, eval on test
  - PInKS
    - PINKS eval on test
  - Orig+PInKS
    - tune PInKS on train, eval on test

Target Task	**Orig.**	***PInKS***	**Orig+*PInKS***
δ-NLI	83.4	60.3	**84.1**
CoreQuisite	77.1	69.5	68.0
ANION	81.1	52.9	**81.2**
ATOMIC	43.2	**48.0**	**88.6**
Winoventi	51.1	**52.4**	**51.3**

Table 2: Macro-F1 (%) results of *PInKS* on the target datasets: no *PInKS* (*Orig.*), with *PInKS* in zero-shot transfer learning setup (*PInKS*) and *PInKS* in addition to original task's data (*Orig.*+*PInKS*). **Bold** values are cases where *PInKS* is improving supervised results.

Summary

- We saw
  - Preconditions and Preconditioned Inference
  - Weak supervision methods to obtain data
  - Biased masking to improve LMs
  - Theoretical and Empirical improvements



Takeaways

- Preconditioned inference is inherently difficult tasks with vast applications
- Conjunctions can serve as proxy for models to comprehend preconditions
- Inference with hard preconditions tend to be more challenging
    - Enough weak supervision can help but needs huge computation and more fine-grained LFs

Visual Commonsense is the Viable Path

- Future/Ongoing Research Directions
  - Visual Commonsense motivation
  - VIPHY: Probing "Visible" Physical Commonsense Knowledge (Under Rev-EACL 2023)
  - Preconditioned Visual Language Inference with Weak Supervision (Under prep-ACL 2023)

References

- [ASCENT] Nguyen, Tuan-Phong, Simon Razniewski, and Gerhard Weikum. "Advanced semantics for commonsense knowledge extraction." Proceedings of the Web Conference 2021. 2021.
- [OMCS] Havasi, Catherine, et al. "Open mind common sense: Crowd-sourcing for common sense." Workshops at the Twenty-Fourth AAAI Conference on Artificial Intelligence. 2010.
- [Snorkel] Ratner, Alexander, et al. "Snorkel: Rapid training data creation with weak supervision." Proceedings of the VLDB Endowment. International Conference on Very Large Data Bases. Vol. 11. No. 3. NIH Public Access, 2017.
- [PaCo] Ehsan Qasemi, Filip Ilievski, Muhao Chen, and Pedro Szekely. 2022. Paco: Circumstantial preconditions of common sense knowledge. arXiv preprint arXiv:2104.08712.
- [**δ**-NLI] Rudinger, Rachel, et al. "Thinking like a skeptic: Defeasible inference in natural language." Findings of the Association for Computational Linguistics: EMNLP 2020. 2020.
- [ANION] Jiang, Liwei, et al. "" I'm Not Mad": Commonsense Implications of Negation and Contradiction." arXiv preprint arXiv:2104.06511 (2021).
- [Winoventi] Do, Nam, and Ellie Pavlick. "Are Rotten Apples Edible? Challenging Commonsense Inference Ability with Exceptions." Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021. 2021.
- [ATOMIC2020] Hwang, Jena D., et al. "(Comet-) Atomic 2020: On Symbolic and Neural Commonsense Knowledge Graphs." Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 35. No. 7. 2021.
- [PaBi] He, Hangfeng, et al. "Foreseeing the benefits of incidental supervision." arXiv preprint arXiv:2006.05500 (2020).

Thank you

Understanding and Generating Multimodal Feedback in Human-Machine Story-Telling

Setareh Nasihati Gilani

November 2022

Outline

- Introduction
- Completed Work Part 1: Story-Telling Agent
- Completed Work Part 2 : RAVE
- Completed Work Part 3 : Identifying feedback Opportunities
- Proposed Work

Definitions-Terminology

- Feedback (in the context of interaction)
  - response of a receiver to sender's message
  - Can be verbal, nonverbal
- Story
  - A story or narrative is a connected series of events told through words (written or spoken), imagery (still and moving), body language, performance, music, or any other form of communication.
- Story-telling context
  - An interaction which involves a story told by an interlocuter

Motivating Examples

Outline

- Introduction
    - Motivation
    - Research Questions
    - Approach
    - Contributions

Motivation

- Feedback is an important part of any interaction
- Telling Stories is important in interactions
- Feedback in story-telling context is important
- Human-Machine Interactions are rapidly increasing in different aspects of our lives
- Machines should be able to use feedback within story-telling context

Motivation

- Feedback is an important part of any interaction
  - Feedback can be multimodal (verbal, visual, audio, …)
  - Used by dialogue partners as evidence of their success
  - Can be positive / negative
  - Come naturally for humans

Motivation

- Feedback is an important part of any interaction
- Telling Stories is important in interactions
- Feedback in story-telling context is important
- Human-Machine Interactions are rapidly increasing in different aspects of our lives
- Machines should be able to use feedback within story-telling context

Motivation

- Feedback is an important part of any interaction
- Telling Stories is important in interactions
  - Used to establish identity
  - Pass on cultural heritage
  - Build Rapport

Motivation

- Feedback is an important part of any interaction
- Telling Stories is important in interactions
- Feedback in story-telling context is important
- Human-Machine Interactions are rapidly increasing in different aspects of our lives
- Machines should be able to use feedback within story-telling context

Motivation

- Feedback is an important part of any interaction
- Telling Stories is important in interactions
- Feedback in story-telling context is important

→ One might try to regain attention, or cut the story short upon receiving negative feedback

→ One might want to go into detail about one specific part upon receiving positive feedback

Motivation

- Feedback is an important part of any interaction
- Telling Stories is important in interactions
- Feedback in story-telling context is important
- Human-Machine Interactions are rapidly increasing in different aspects of our lives
- Machines should be able to use feedback within story-telling context

Motivation

- Feedback is an important part of any interaction
- Telling Stories is important in interactions
- Feedback in story-telling context is important
- Human-Machine Interactions are rapidly increasing in different aspects of our lives
  - Smart assistants on phones, laptops, …
  - Assistive robots for homes
  - Smart Devices, Internet of things
  - Companion robots for elderly

Motivation

- Feedback is an important part of any interaction
- Telling Stories is important in interactions
- Feedback in story-telling context is important
- Human-Machine Interactions are rapidly increasing in different aspects of our lives
- Machines should be able to use feedback

within story-telling context

Motivation

- Feedback is an important part of any interaction
- Telling Stories is important in interactions
- Feedback in story-telling context is important
- Human-Machine Interactions are rapidly increasing in different aspects of our lives
- Machines should be able to use feedback within story-telling context

Examples of Machines who use stories

- New Dimensions in Testimony
  - Story-teller agent
- Simsensei kiosk
  - Elicits and listens to stories

USC University of Southern California

Outline

- Introduction
- Completed Work Part 1: Story-Telling Agent
- Completed Work Part 2: RAVE
- Completed Work Part 3: Identifying feedback Opportunities
- Proposed Work

Story-Telling Agent

RQ 1: What kinds of stories should virtual storytellers provide?

- Different story-designs
  - Context
  - Audience
  - Goals
- Investigated the answer in a specific scenario

USC University of Southern California

What kind of stories?

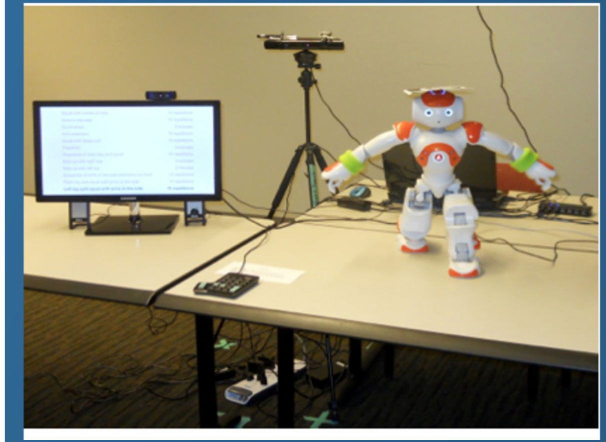
- Goals
  - Building long-term rapport
  - Desire to keep interacting with the system
- Which types of stories will best achieve the goal?
- Ways that stories differ
  - Taxonomy of Collins & Traum includes 5 main dimensions with several sub-dimensions (LREC 2016)
    - Point of View/Perspective (first, second, third person)
    - Identity (explicit/unspecified, types of explicit)

Specifying types of stories

- Perspective: First person stories or third person?
- Identity: Human backstory or artificial?
    - Are human oriented stories plausible coming from a virtual human?
    - Is there a backfire effect when virtual humans claim to have human like experiences?

Related Work

- Perspective
  - Virtual exercise counselor agent (Bickmore et al IVA 2009)
    - Tells inspiring stories about weight loss using either 1st person or 3rd person
    - Results: First person participants were likely to report greater enjoyment
- Identity:
  - Robot exercise buddy (Swift-Spong et al Ro-man 2016)
    - Robot or space alien identity backstory
    - Results: no significant difference



USC University of Southern California

Experiment

- Engage with users in a simple ‘get to know you’ dialogue
- Swap stories for a set of ‘ice-breaker’ questions
- Independent Variables
  - Perspective (1st person vs 3rd person stories)
  - Identity (Human vs Virtual human)
  - Embodiment (Human video vs Virtual human)
- Investigated degree of reciprocity and compared rapport

Dialogue conditions

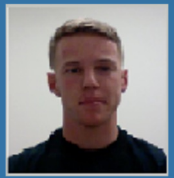
Identity

Virtual (VH)

Human (H)

Embodiment

Human (H)



H – H – 1st

H – H – 3rd

Virtual (VH)

VH – VH – 1st

VH – H – 1st

VH – VH – 3rd

VH – H – 3rd

← 1st person

← 3rd person

← 1st person

← 3rd person

Example story (VH – VH – 3rd)

- I’m a virtual human at ICT, which means that I get to talk and interact with people like you. It’s a really great job. I get to talk with some intriguing people and hear about all kinds of interesting experiences. This woman the other day was telling me about how she had ruined her last 8 laptops in the same way; by spilling Greek yogurt all over the keyboard. To my relief, she said that she does not eat snacks around computers anymore, which I would probably say it’s a good thing.

USC University of Southern California

Participant interactions

- 60 participants assigned randomly to one of the 3 conditions of a between-subject factor
- VH-VH, VH-H, H-H
- They interacted with both perspective agents in that condition (within-subject factor): 1st person, 3rd person
- Counter balancing order
- Interaction episodes
  - Participant asks questions
  - Agent answers and follow up with a reciprocal question (Like “What about you?”)
  - User answers the question himself

Experiment Metrics

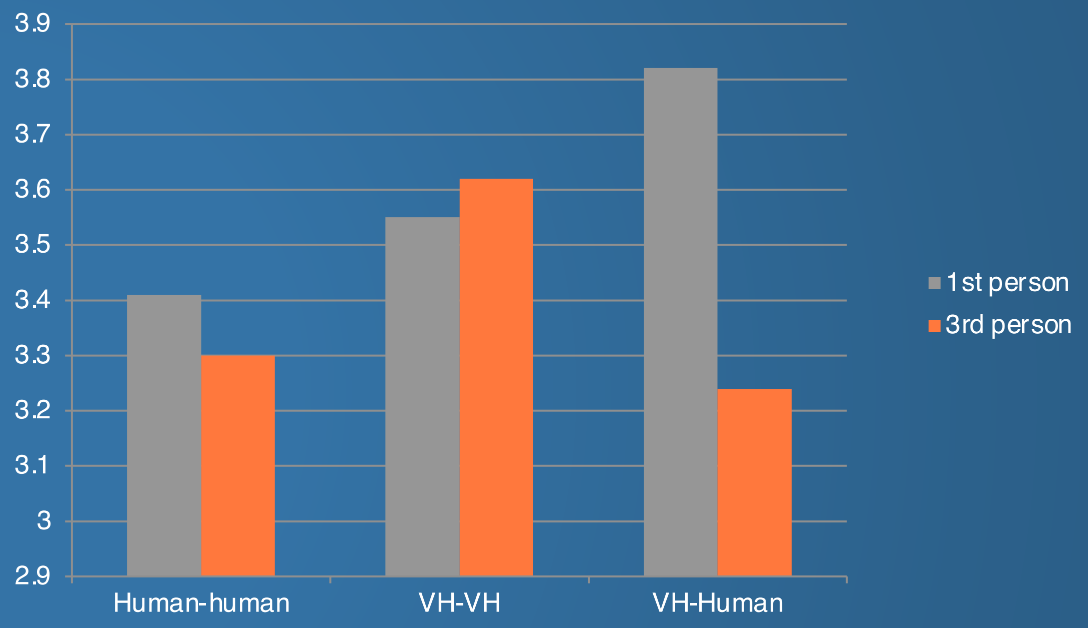
- 9-item rapport scale
  - Eg: I felt I had a connection with my partner.
- 6-item ancillary rapport scale
  - Eg: How much did you like the character?
- Two items on subjective sharing of personal information
  - How much did you reveal about yourself in your answers?
  - How personal were your answers?
- A set of 30 personality characteristics
  - Cheerful, threatening, aloof
- Length of participant responses
- Number of participant responses containing stories

Results

- Conducted a 2 (perspective) * 2 (order) * 3 (agent) mixed ANOVAs
  - Order and agent as between-subject factors
  - Perspective as within-subject factor

Results (continued)

- First Person: More Rapport
  - Marginally significant main effect of perspective exists
  - $F(1, 53) = 3.21$, $p = .08$.
  - 1st person agent ($M = 3.61$, $SE = 0.09$)
  - 3rd person agent ($M = 3.42$, $SE = 0.09$).



Interaction of perspective and agent on the Rapport Scale.

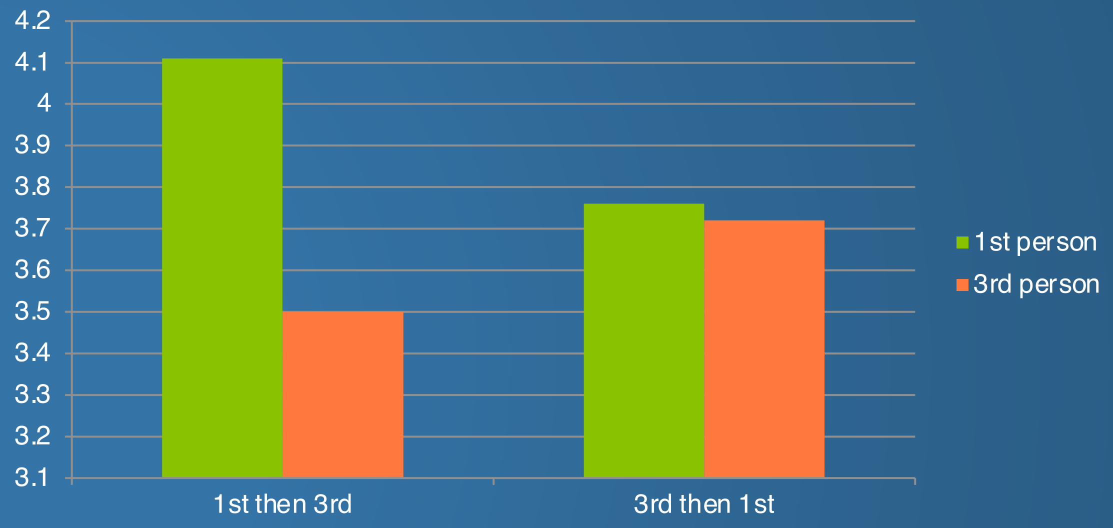
USC University of Southern California

Results (continued)

- First Person: More Rapport
- First Person: More disclosure of personal information
  - significant main effect of perspective, $F(1, 53) = 6.88$
  - 1st person agent ($M = 3.92$, $SE = 0.12$)
  - 3rd person agent ($M = 3.65$, $SE = 0.14$).

Results (continued)

- First Person: More Rapport
- First Person: More disclosure of personal information
- First Person: More cheerful
  - Marginally significant main effect of perspective, $F(1, 53) = 3.33$, $p = 0.07$
  - 1st person agent ($M = 1.66$, $SE = 0.1$)
  - 3rd person agent ($M = 1.89$, $SE = 0.11$)



Interaction of perspective and order on the personality characteristic “cheerful"

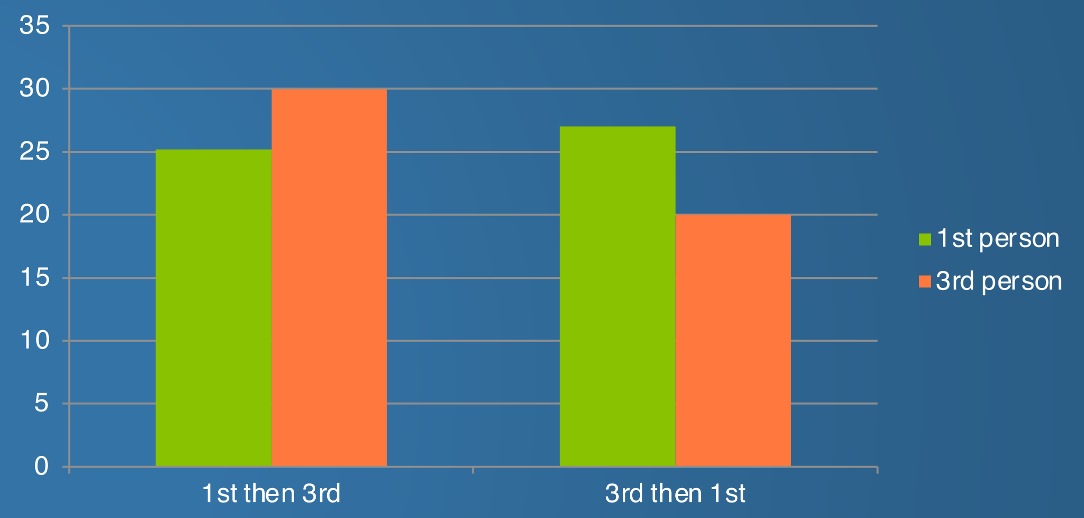
Results (continued)

- First Person: More Rapport
- First Person: More disclosure of personal information
- First Person: More cheerful
- First Person: More trustworthy
  - significant main effect of perspective, $F(1, 52) = 5.67$; $p = 0.02$
  - 1st person agent: ($M = 3.95$; SE 0.13)
  - 3rd person agent: (M = 3:55; SE = 0:16).

USC University of Southern California

Results (continued)

- First Person: More Rapport
- First Person: More disclosure of personal information
- First Person: More cheerful
- First Person: More trustworthy
- No difference in length of response



Interaction of perspective and order on the length participant talked

USC University of Southern California

Takeaway w.r.t. Research question

RQ 1: What kinds of stories should virtual storytellers provide?

- We see a general preference for First person over third person stories
- Fine to tell 3rd person stories as well
- No significant difference for identity
- Also shows the necessity for having feedback in interaction

Outline

- Introduction
- Completed Work Part 1: Story-Telling Agent
- Completed Work Part 2: RAVE
- Completed Work Part 3: Identifying feedback Opportunities
- Proposed Work

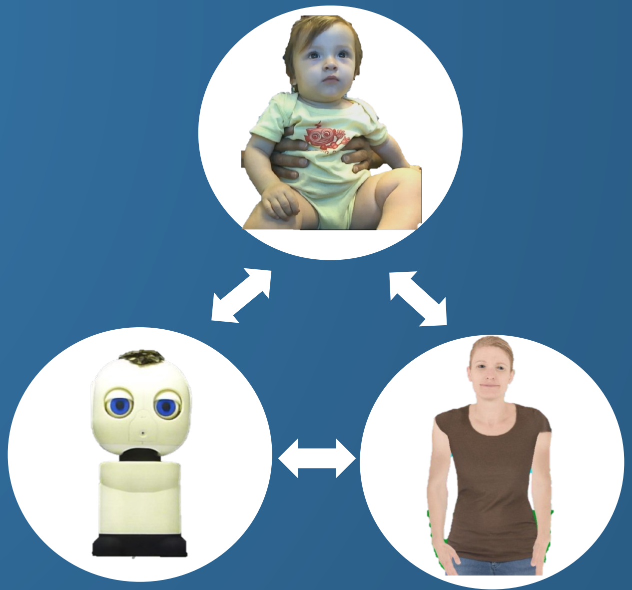
Robot Avatar Thermal Enhanced System (RAVE)

RQ 2: What kinds of multimodal feedback can agents recognize and how can they be used to adapt the dialogue management policies of the agent?

→ Designed and implemented the dialogue management for a multimodal multiparty system called RAVE

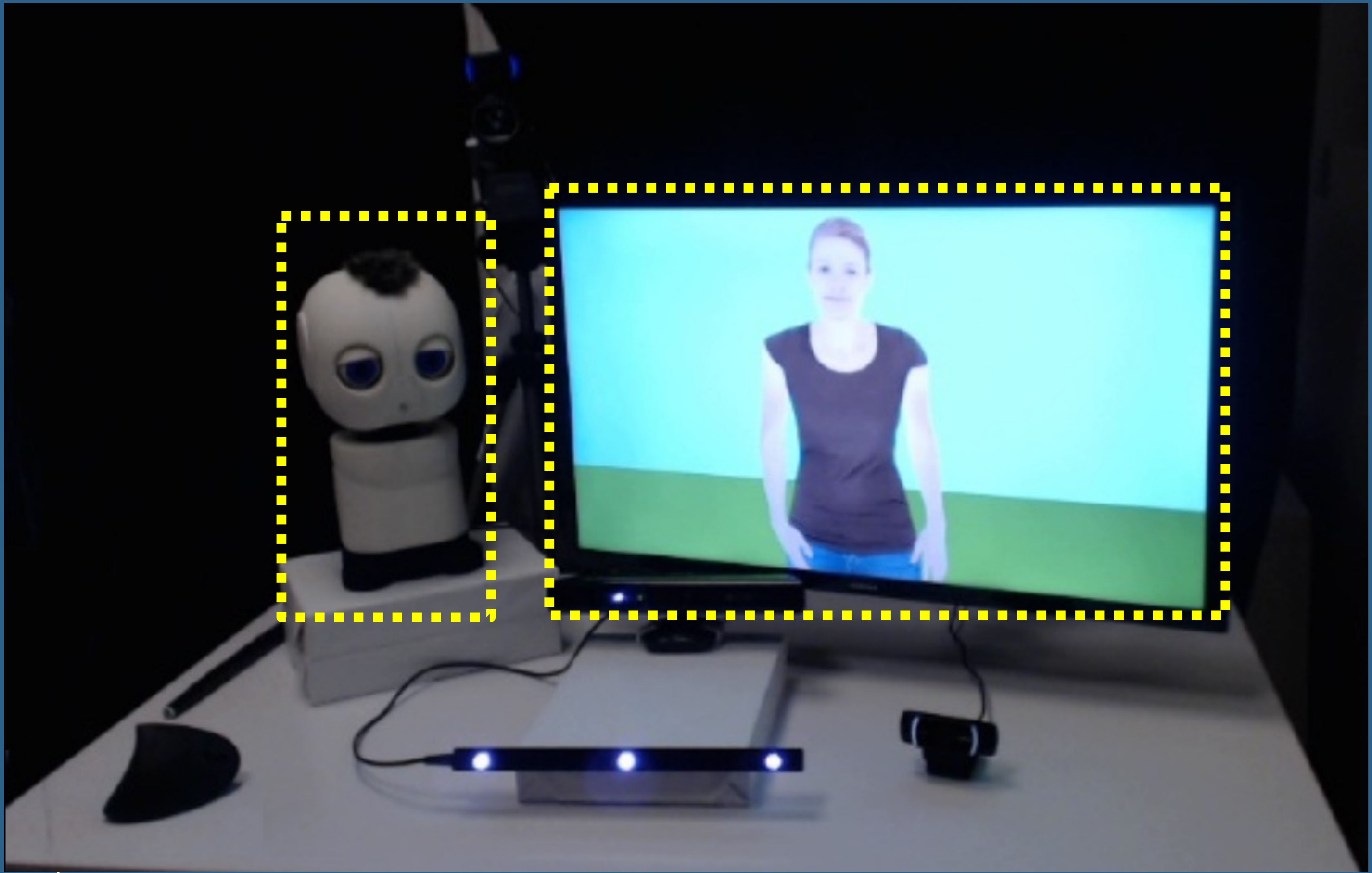
→ Builds a 3-way interaction between a robot, a virtual human and the baby

→ Provides Stories (nursery rhymes) and adapt the flow based on baby's feedback

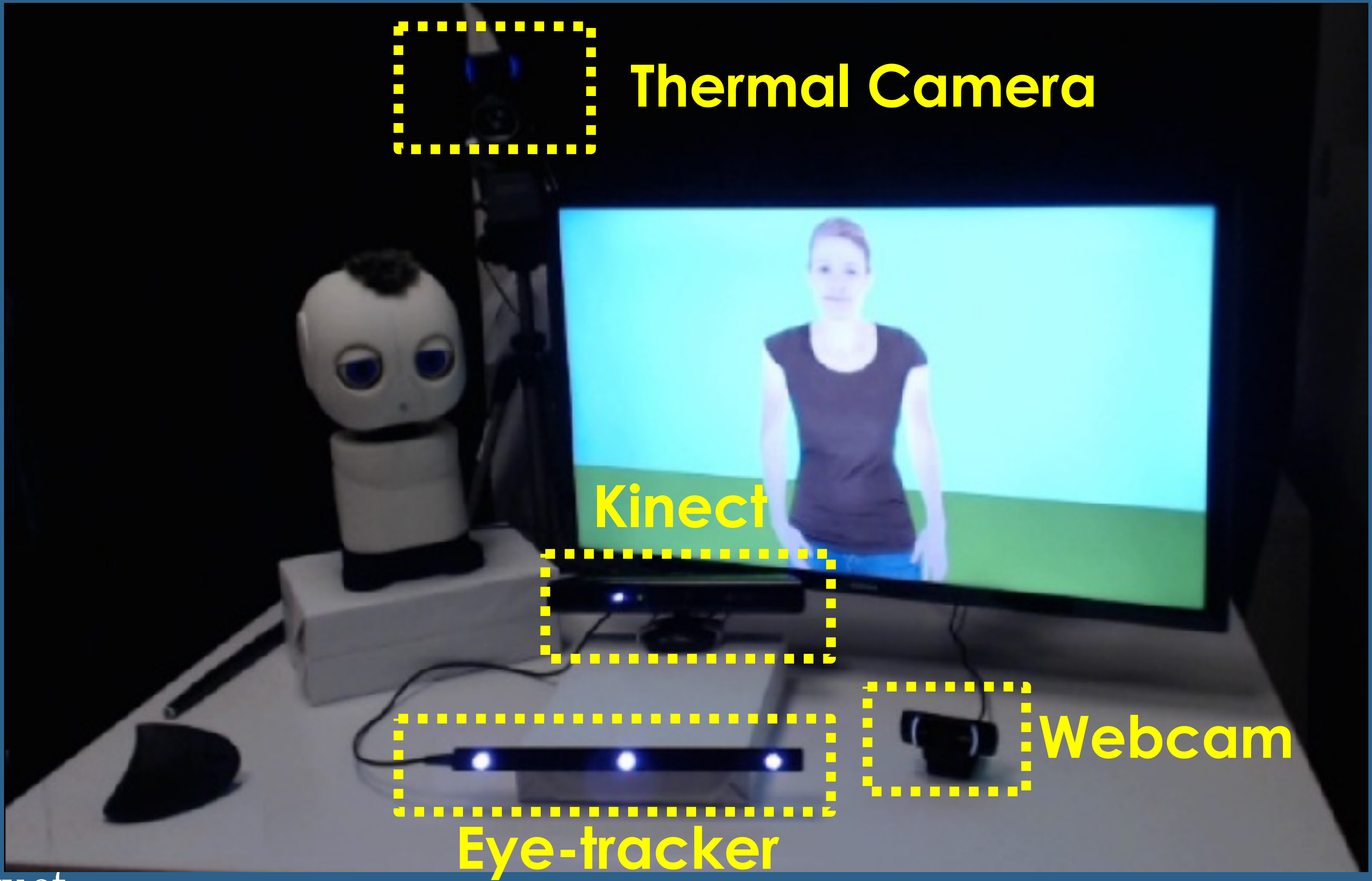


RAVE - Robot, AVatar, thermal-Enhanced System

RAVE - Robot, AVatar, thermal-Enhanced System



RAVE - Robot, AVatar, thermal-Enhanced System

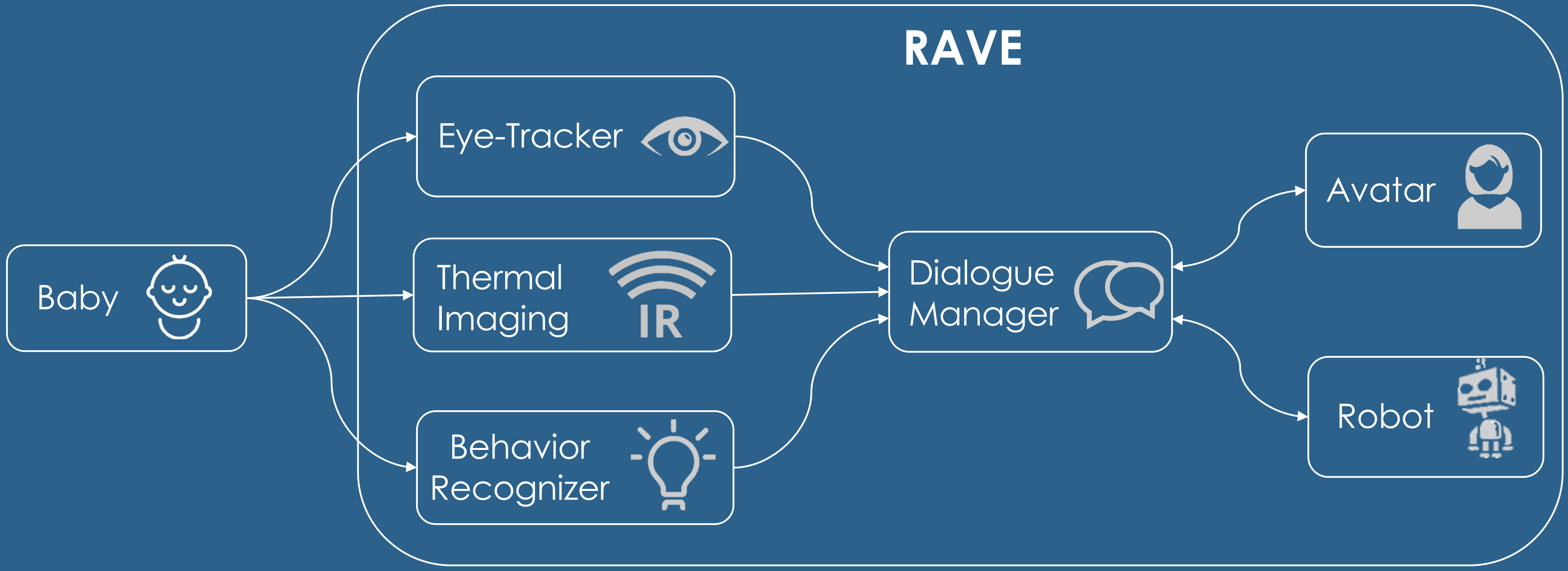


Why RAVE? Background and Motivation

- Language Exposure during early life
- Learning the phonetic units of language by 6-10 months
- Children with minimal or delayed exposure are at risk
- Sign Language has also phonology!
  - Same patterns of phonological development for signed languages
- Deaf/Hard of hearing babies are at risk
- How can technology help?

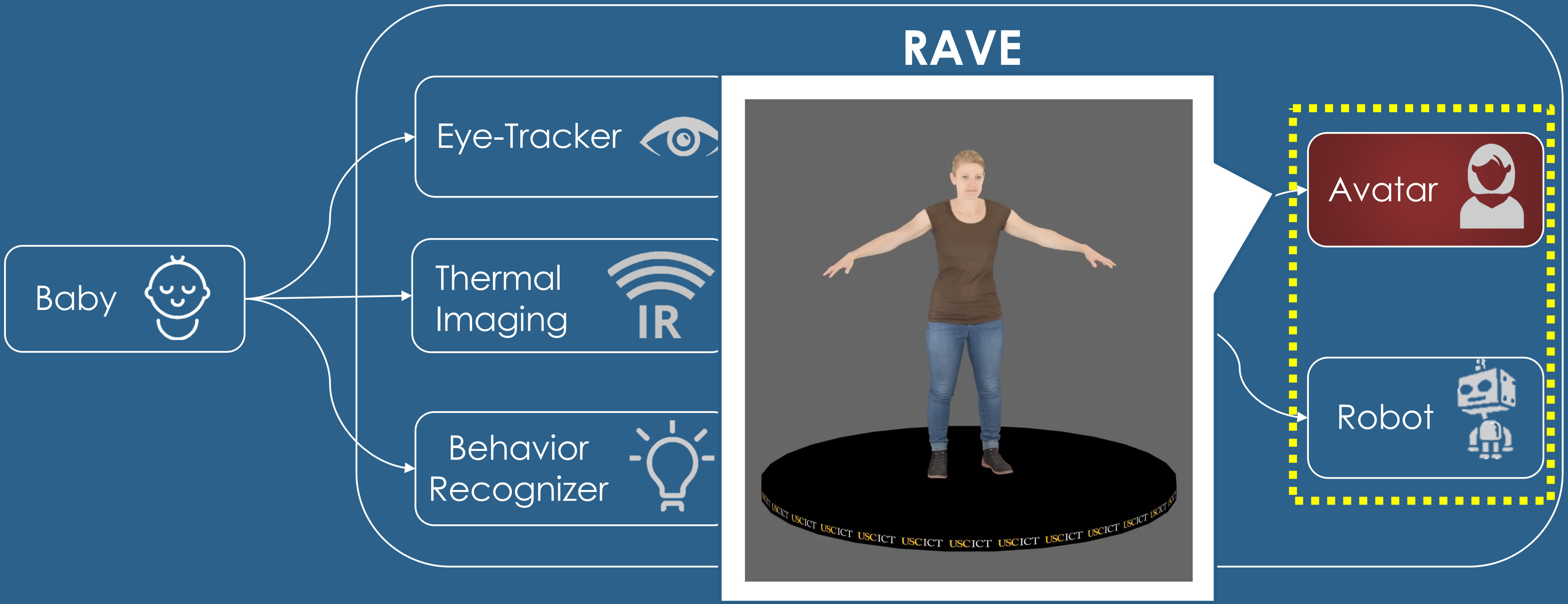
Examples

RAVE - Robot, AVatar, thermal-Enhanced System

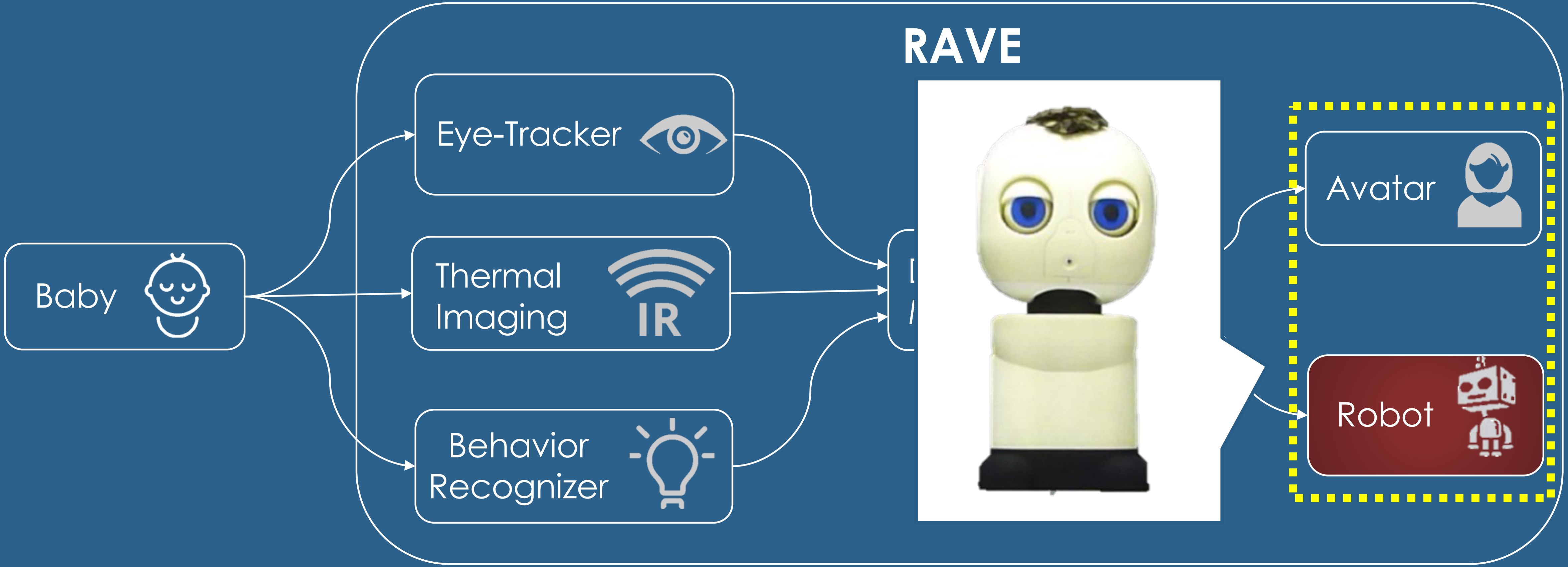


USC University of Southern California

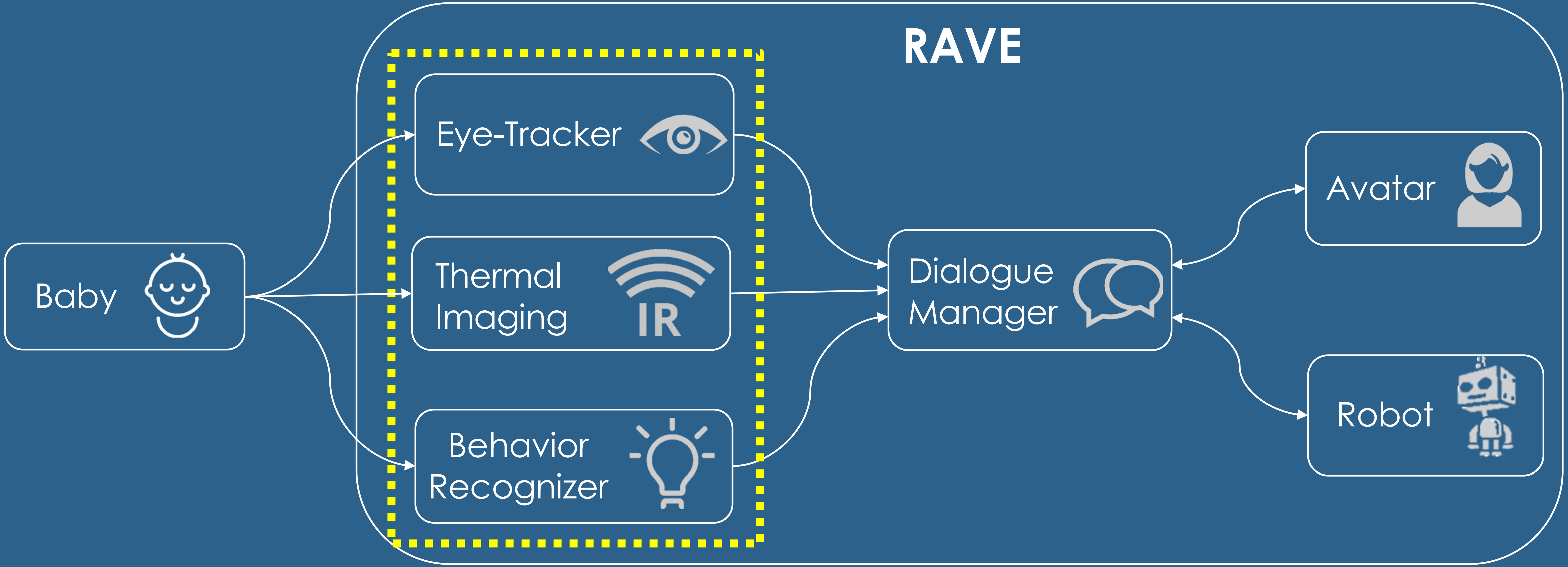
RAVE - Robot, AVatar, thermal-Enhanced System



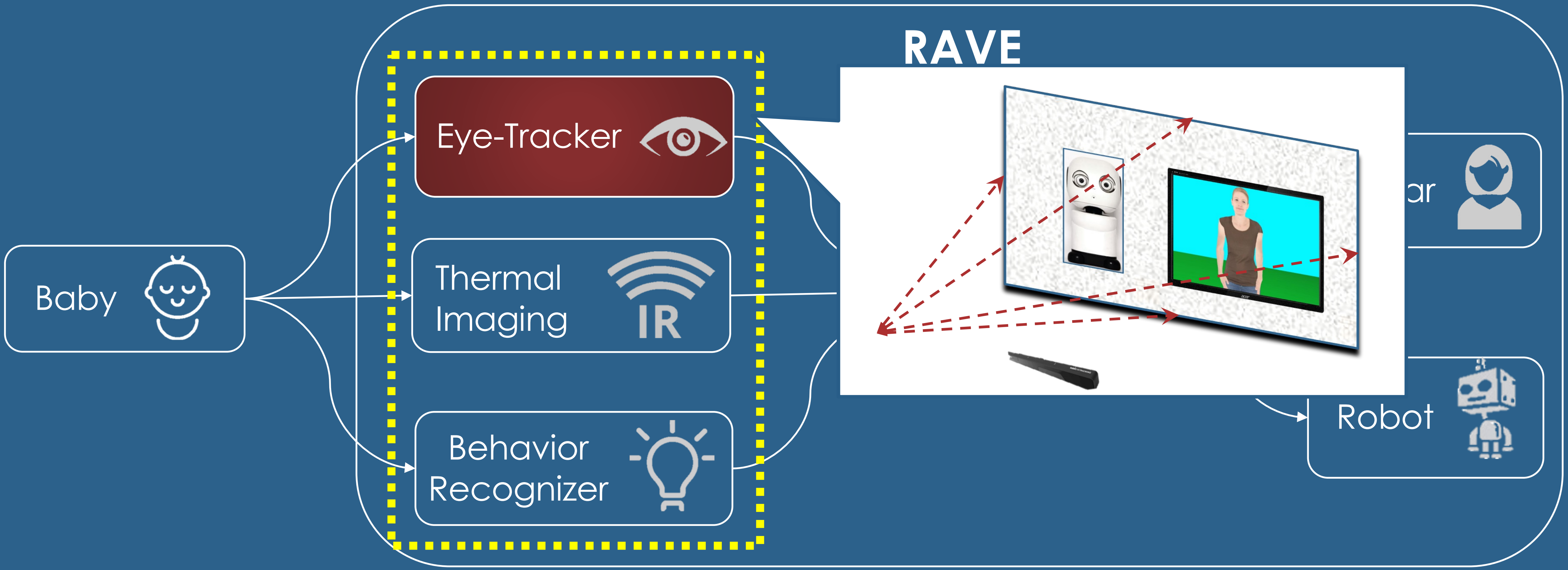
RAVE - Robot, AVatar, thermal-Enhanced System



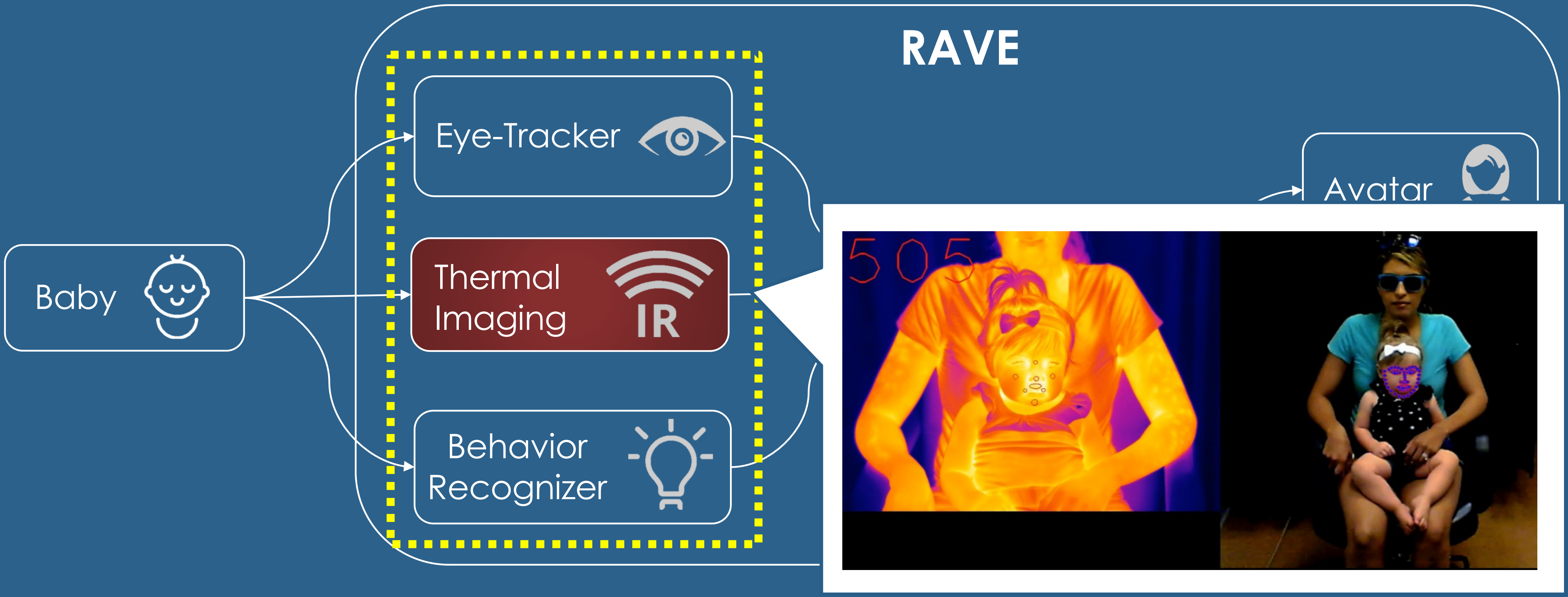
RAVE - Robot, AVatar, thermal-Enhanced System



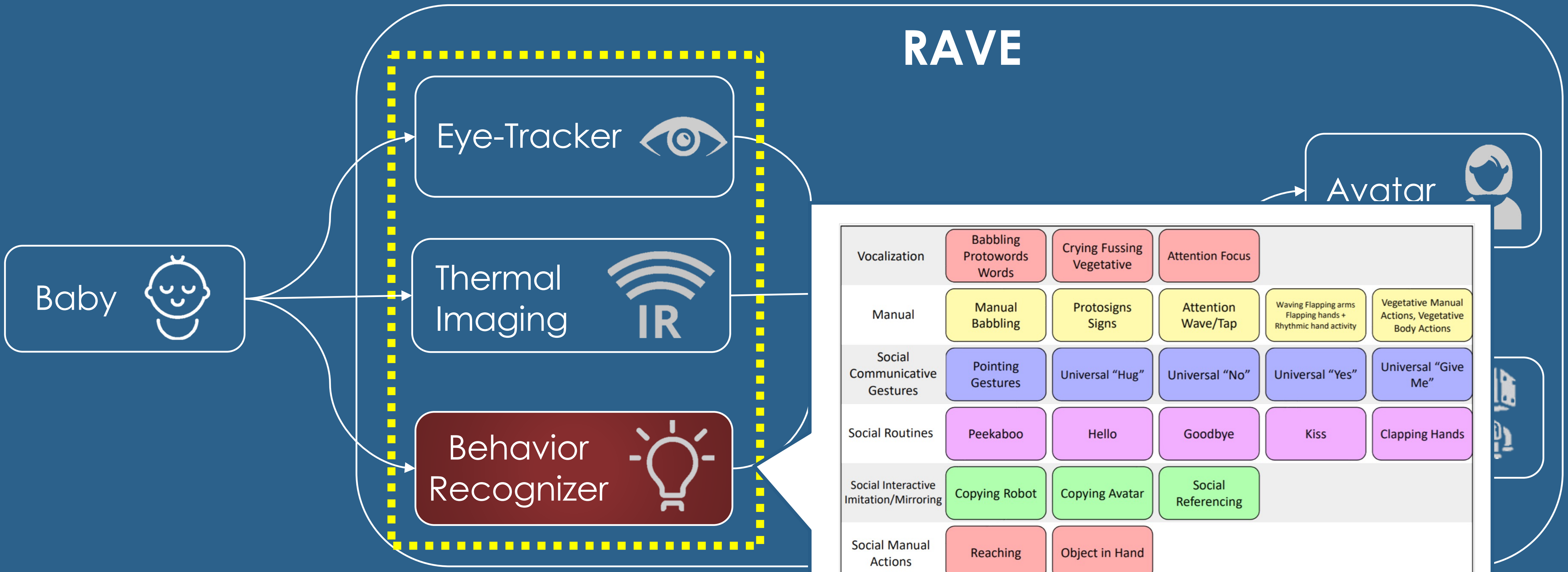
RAVE - Robot, AVatar, thermal-Enhanced System



RAVE - Robot, AVatar, thermal-Enhanced System

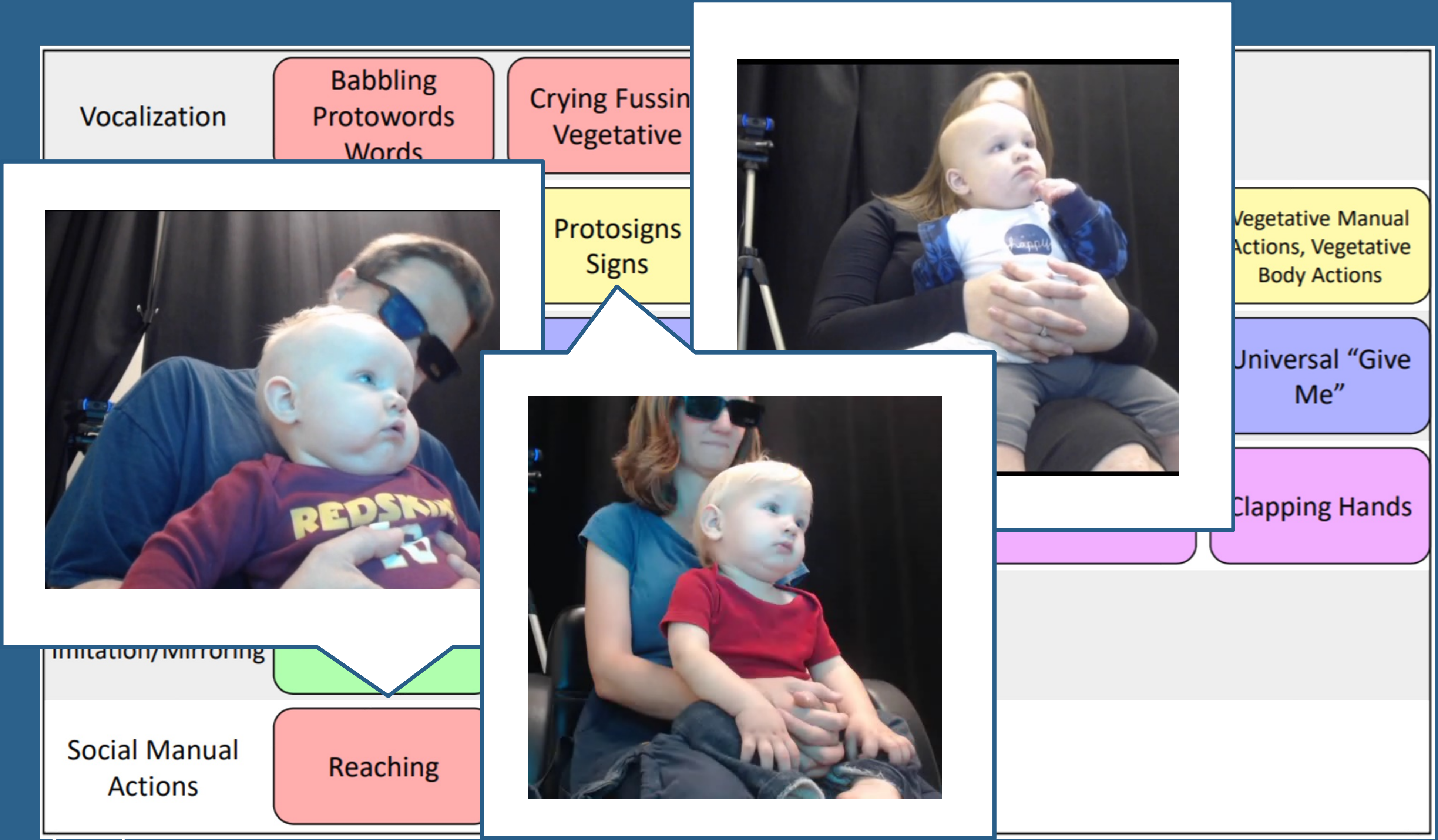


RAVE - Robot, AVatar, thermal-Enhanced System

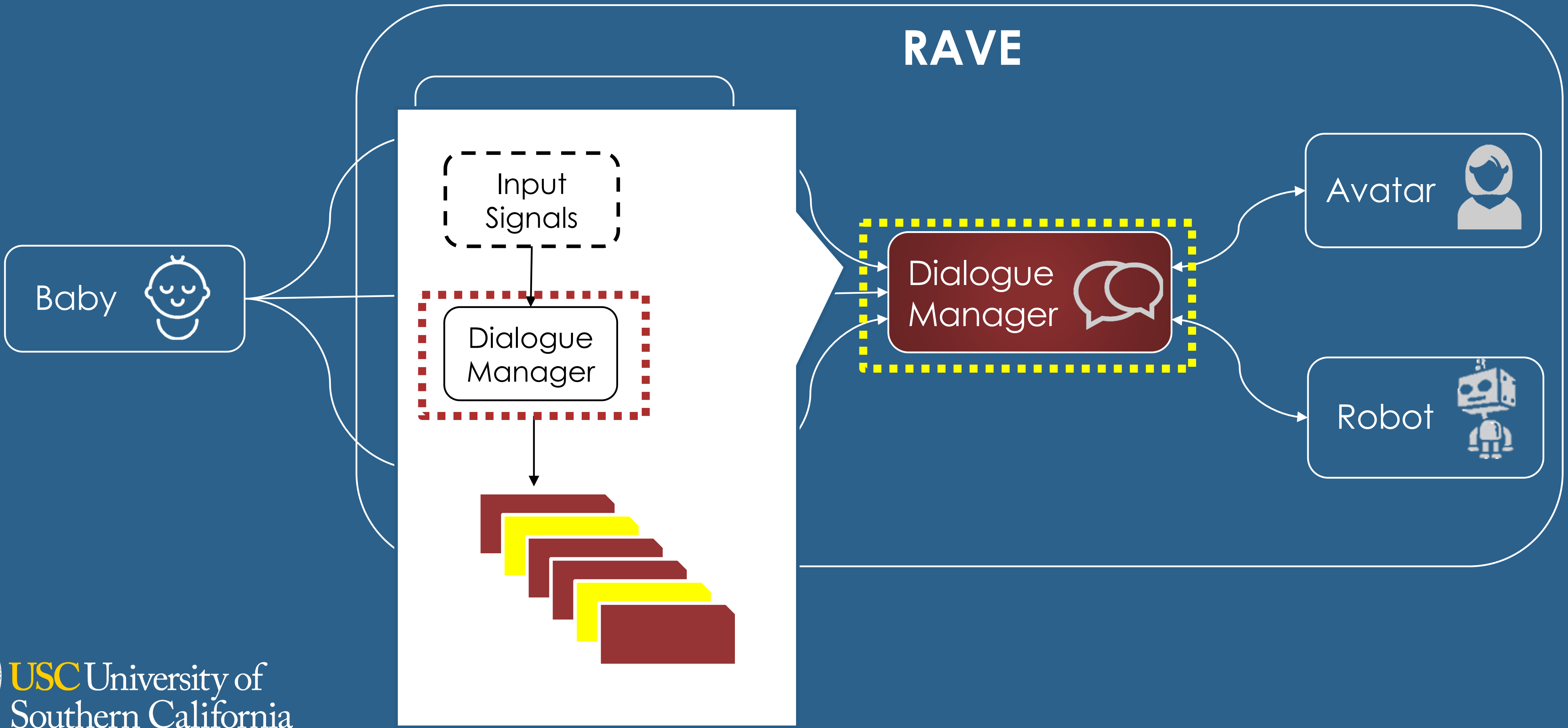


Vocalization	Babbling Protowords Words	Crying Fussing Vegetative	Attention Focus		
Manual	Manual Babbling	Protosigns Signs	Attention Wave/Tap	Waving Flapping arms Flapping hands + Rhythmic hand activity	Vegetative Manual Actions, Vegetative Body Actions
Social Communicative Gestures	Pointing Gestures	Universal “Hug”	Universal “No”	Universal “Yes”	Universal “Give Me”
Social Routines	Peekaboo	Hello	Goodbye	Kiss	Clapping Hands
Social Interactive Imitation/Mirroring	Copying Robot	Copying Avatar	Social Referencing		
Social Manual Actions	Reaching	Object in Hand			

Human Observer Interface

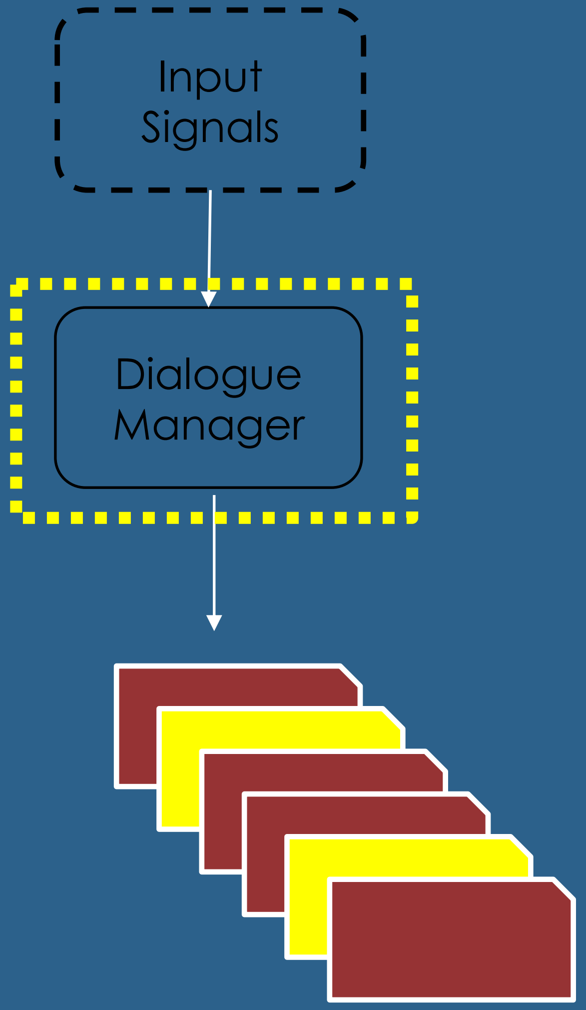


RAVE - Robot, AVatar, thermal-Enhanced System



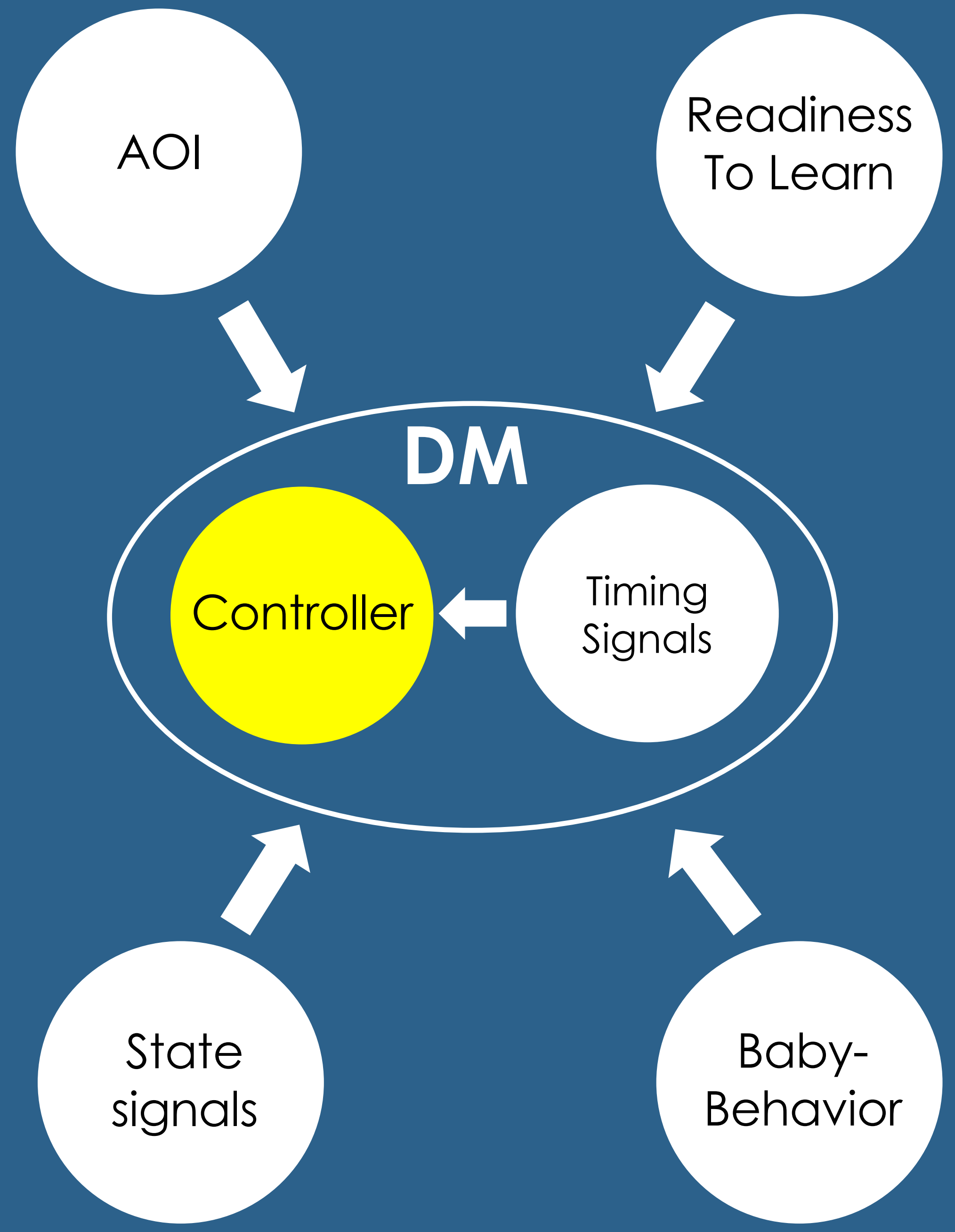
Dialogue Manager

- Rule based model
- Based on information-state architecture
- Plans ahead based on current observations
- Adjusts and re-plans as needed if the desired outcome is not observed



USC University of Southern California

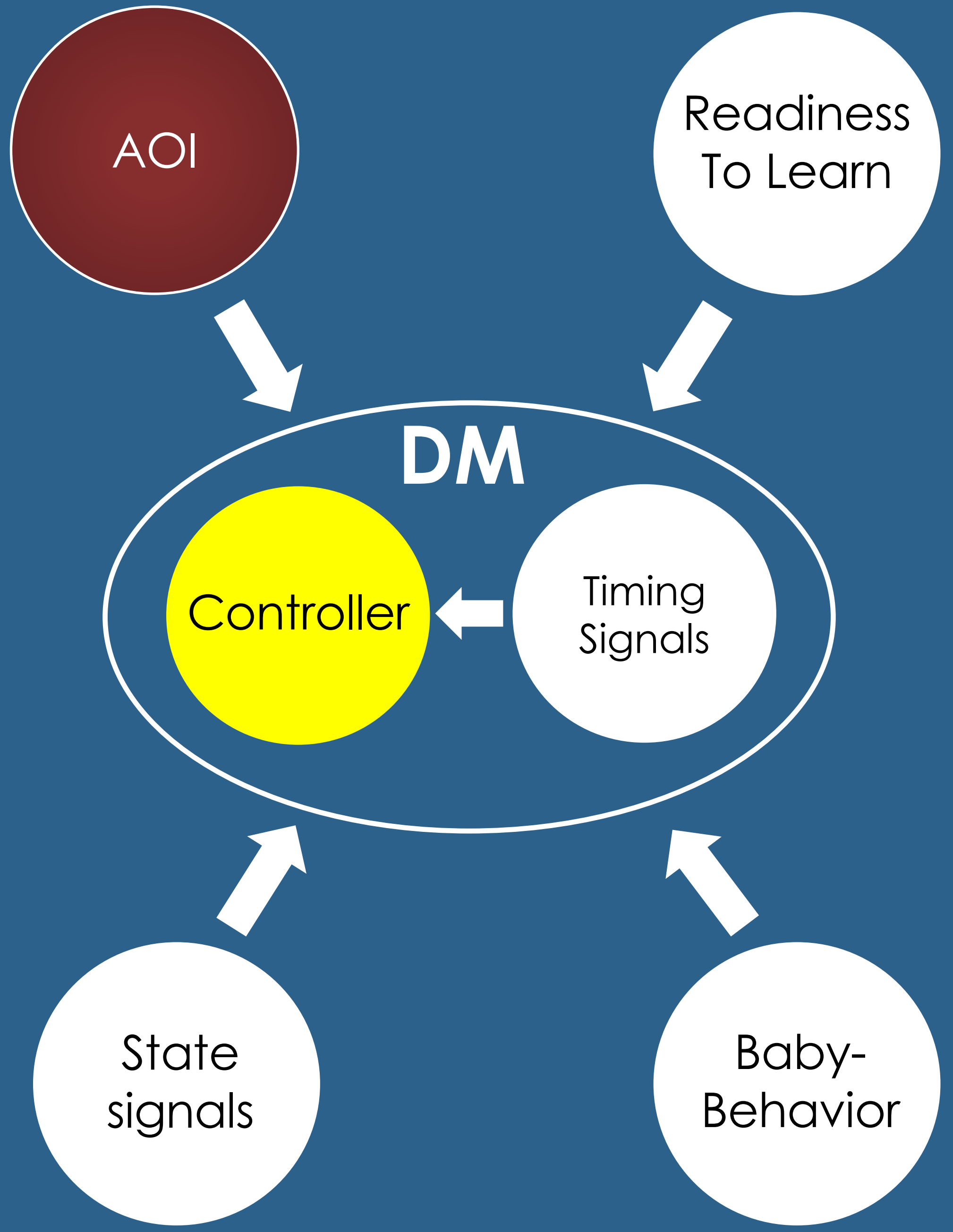
Dialogue Manager Inputs



USC University of Southern California

Dialogue Manager Inputs

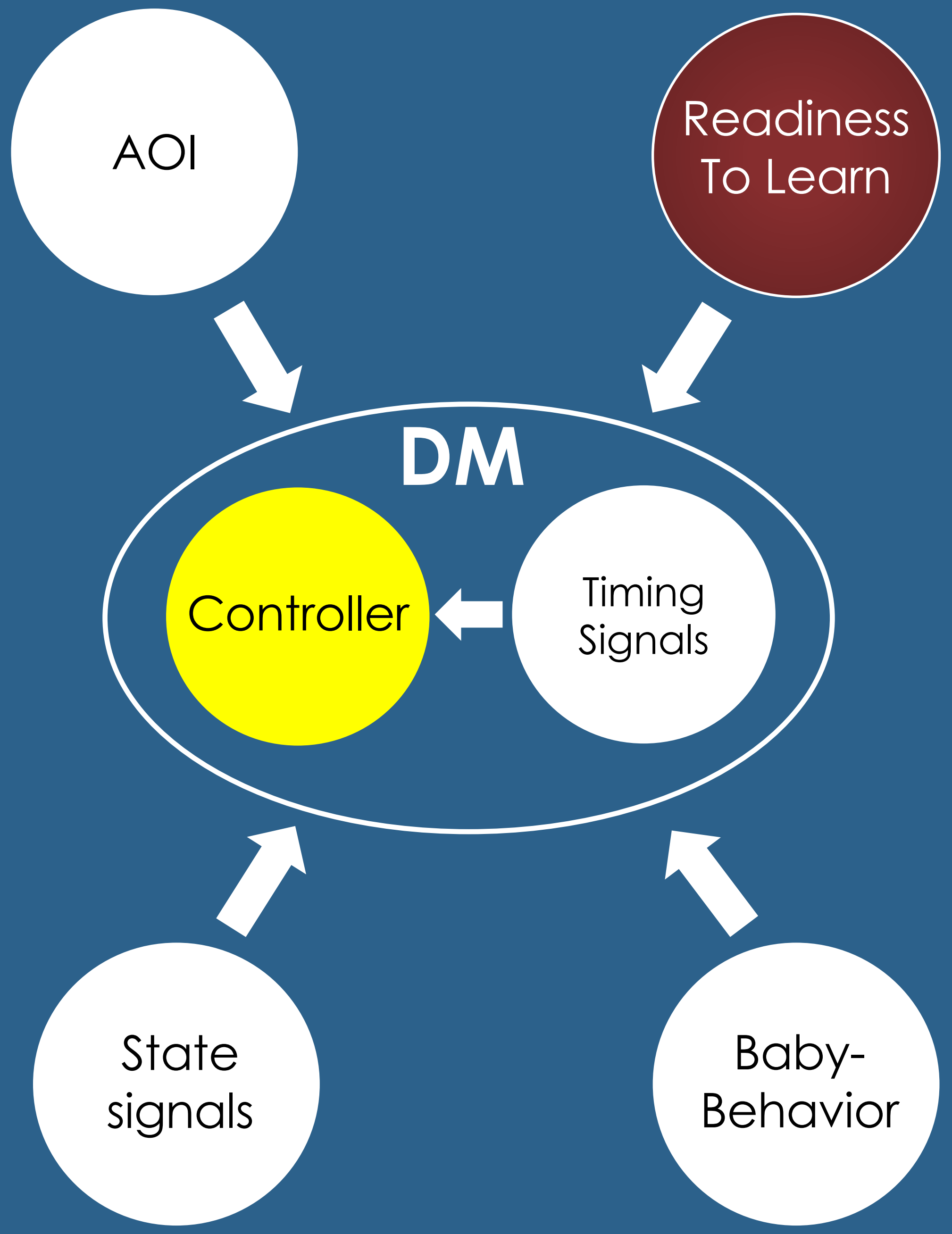
- **Area of Interest (AOI): Avatar, Robot, Between**



USC University of Southern California

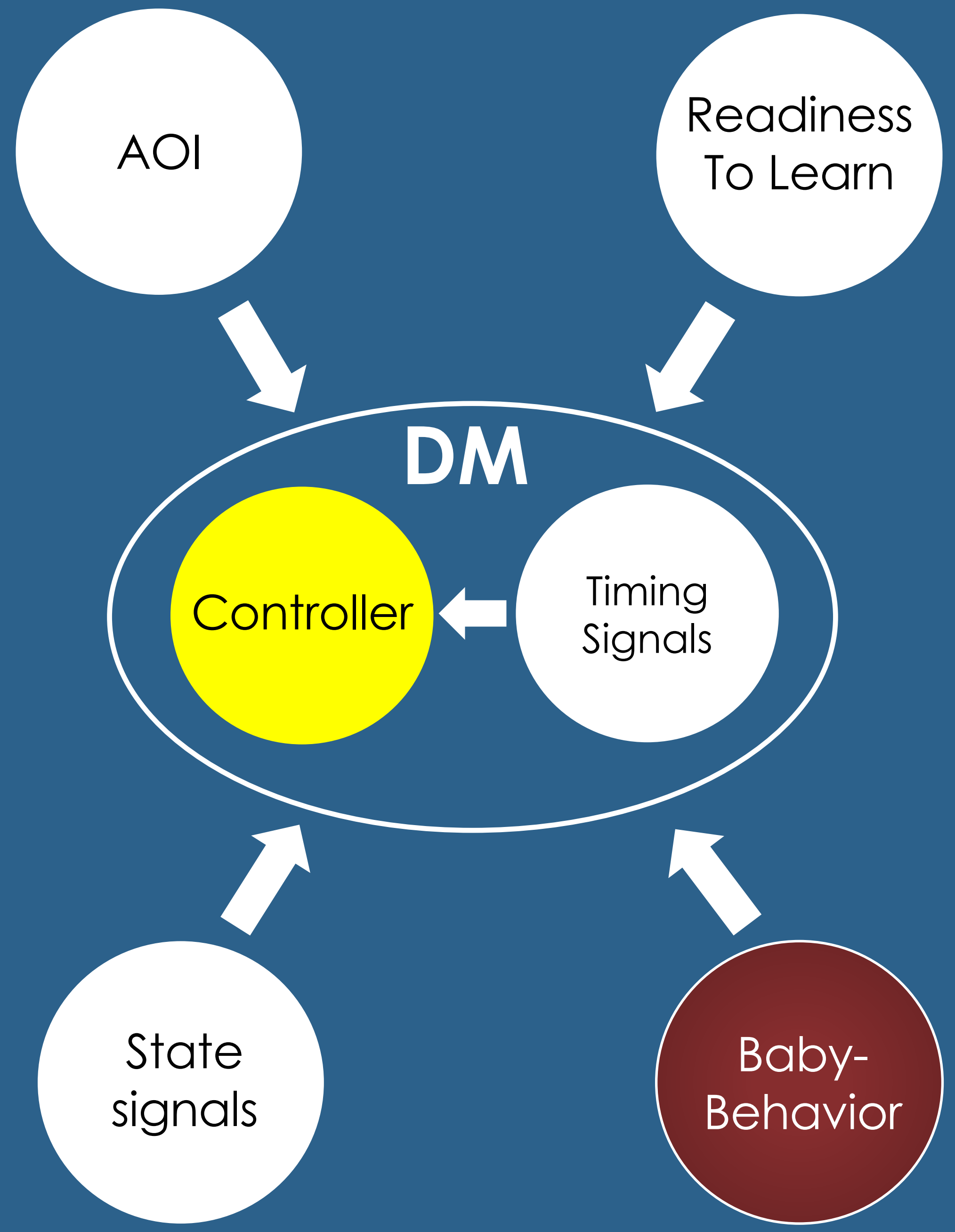
Dialogue Manager Inputs

- Area of Interest (AOI): Avatar, Robot, Between
- **Readiness To Learn: Sympathetic, Parasympathetic**



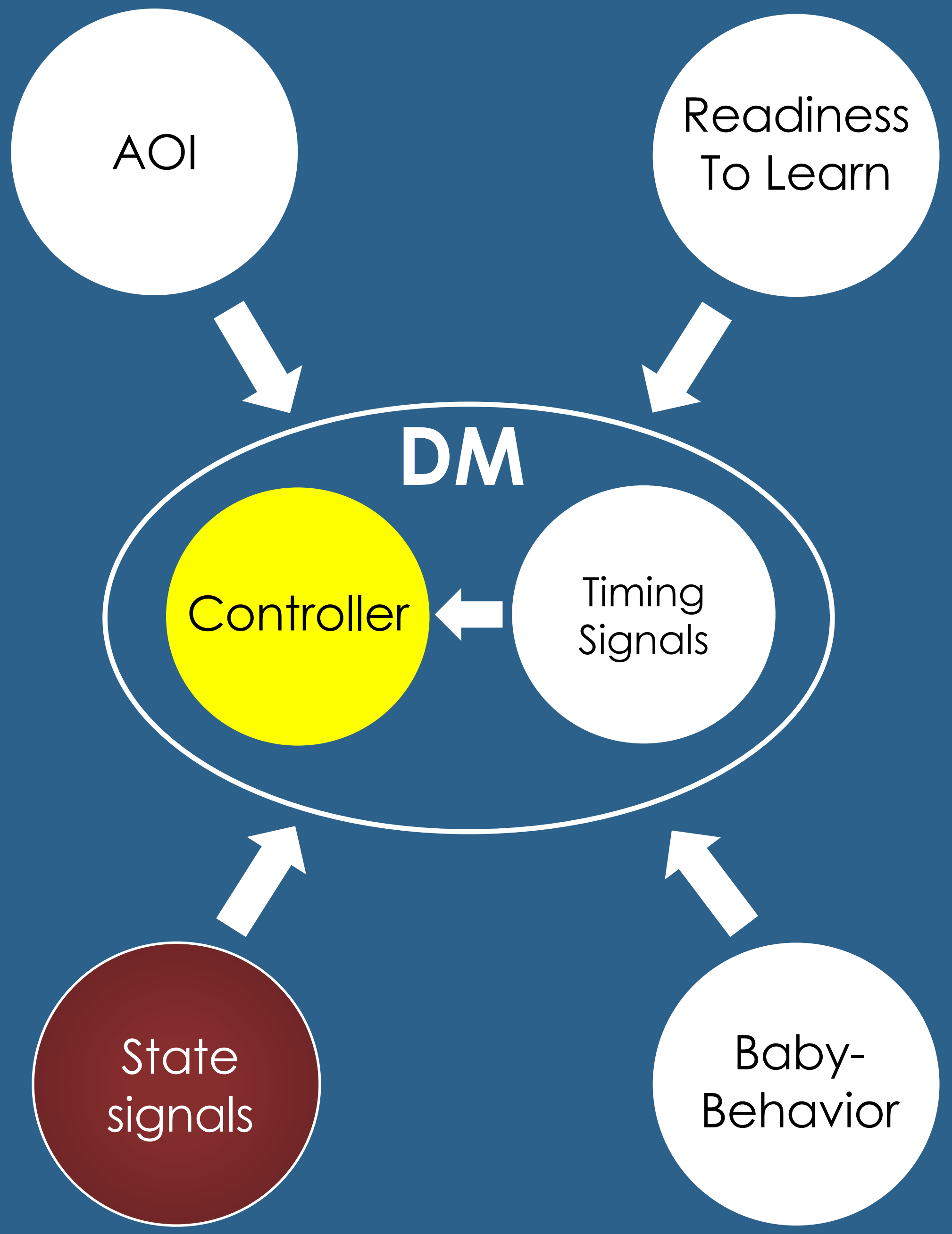
Dialogue Manager Inputs

- Area of Interest (AOI): Avatar, Robot, Between
- Readiness To Learn: Sympathetic, Parasympathetic
- **Baby-Behavior (BB): 23 Distinct States (e.g., Reaching)**



Dialogue Manager Inputs

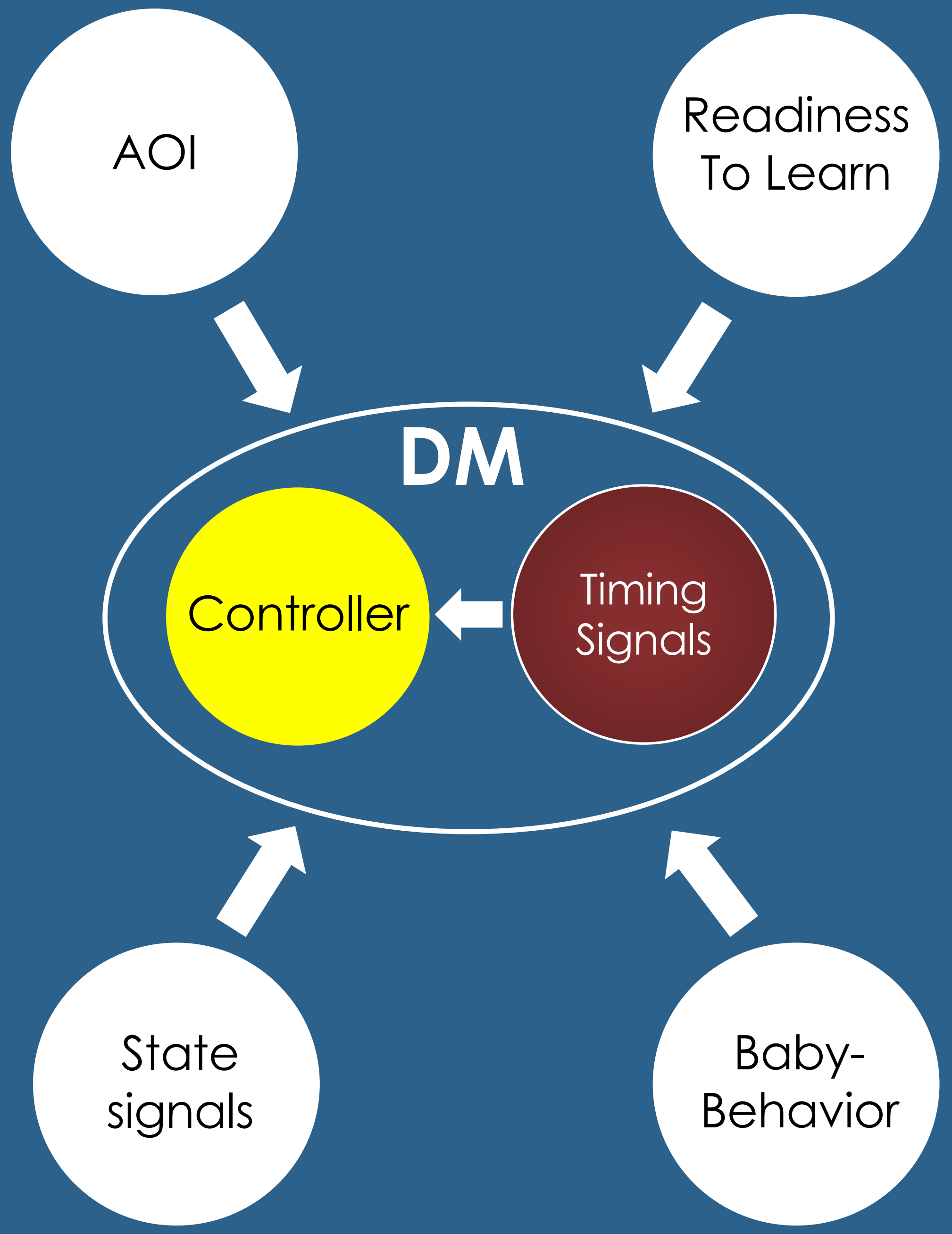
- Area of Interest (AOI): Avatar, Robot, Between
- Readiness To Learn: Sympathetic, Parasympathetic
- Baby-Behavior (BB): 23 Distinct States (e.g., Reaching)
- **State Signals: (e.g., Robot Behavior Exception)**



USC University of Southern California

Dialogue Manager Inputs

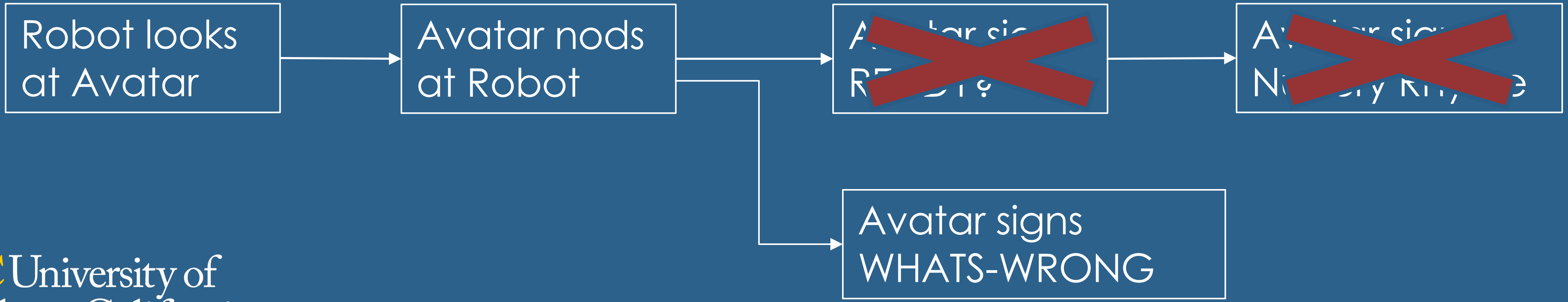
- Area of Interest (AOI): Avatar, Robot, Between
- Readiness To Learn: Sympathetic, Parasympathetic
- Baby-Behavior (BB): 23 Distinct States (e.g., Reaching)
- State Signals: (e.g., Robot Behavior Exception)
- **Timing Signals: (e.g., Time-Out)**



USC University of Southern California

Dialogue Manager Output Commands

- Primitive Behaviors
  - Atomic, cannot be interrupted
- Action Sequences
  - A sequence of primitive behaviors
  - Can be interrupted if needed at primitive behavior boundary



Avatar Primitive Behaviors

- Conversational Fillers
- Social Behaviors
- Question Solicitations
- Stories

USC University of Southern California

Avatar Primitive Behaviors

- **Conversational Fillers**
- Social Behaviors
- Question Solicitations
- Stories

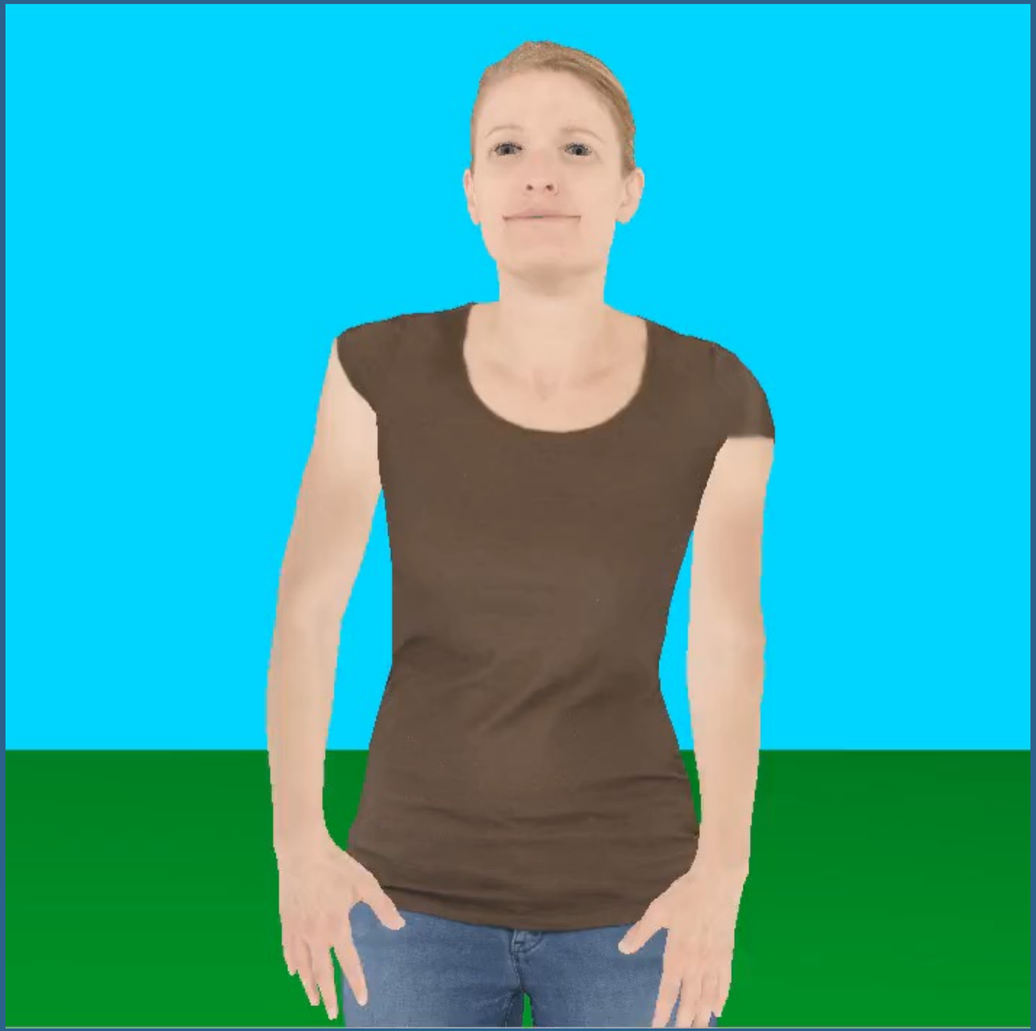
Avatar Primitive Behaviors

- Conversational Fillers
- **Social Behaviors**
- Question Solicitations
- Stories

USC University of Southern California

Avatar Primitive Behaviors

- Conversational Fillers
- Social Behaviors
- **Question Solicitations**
- Stories



Avatar Primitive Behaviors

- Conversational Fillers
- Social Behaviors
- Question Solicitations
- **Stories**

Examples of Avatar Signing

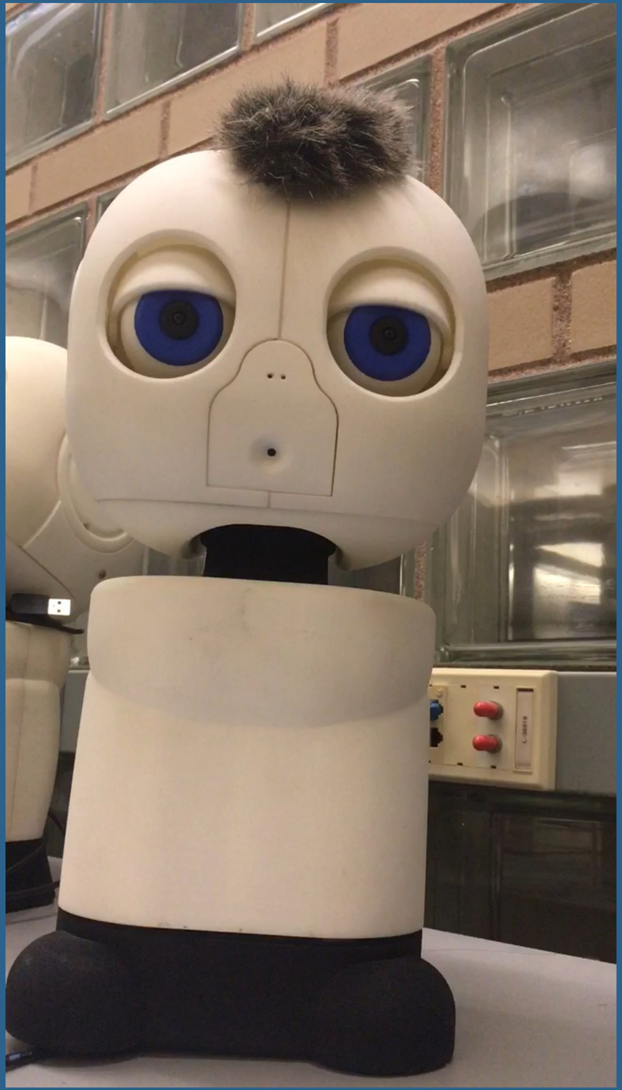
Examples of Avatar Signing (Conversational Openers)

Robot Primitive Behaviors

- Head Turns
- Nodding
- Hide/Unhide

Robot Primitive Behaviors

- **Head Turns**
- Nodding
- Hide/Unhide



Robot Primitive Behaviors

- Head Turns
- **Nodding**
- Hide/Unhide

Robot Primitive Behaviors

- Head Turns
- Nodding
- **Hide/Unhide**

USC University of Southern California

Action Sequence Example

Robot head turn

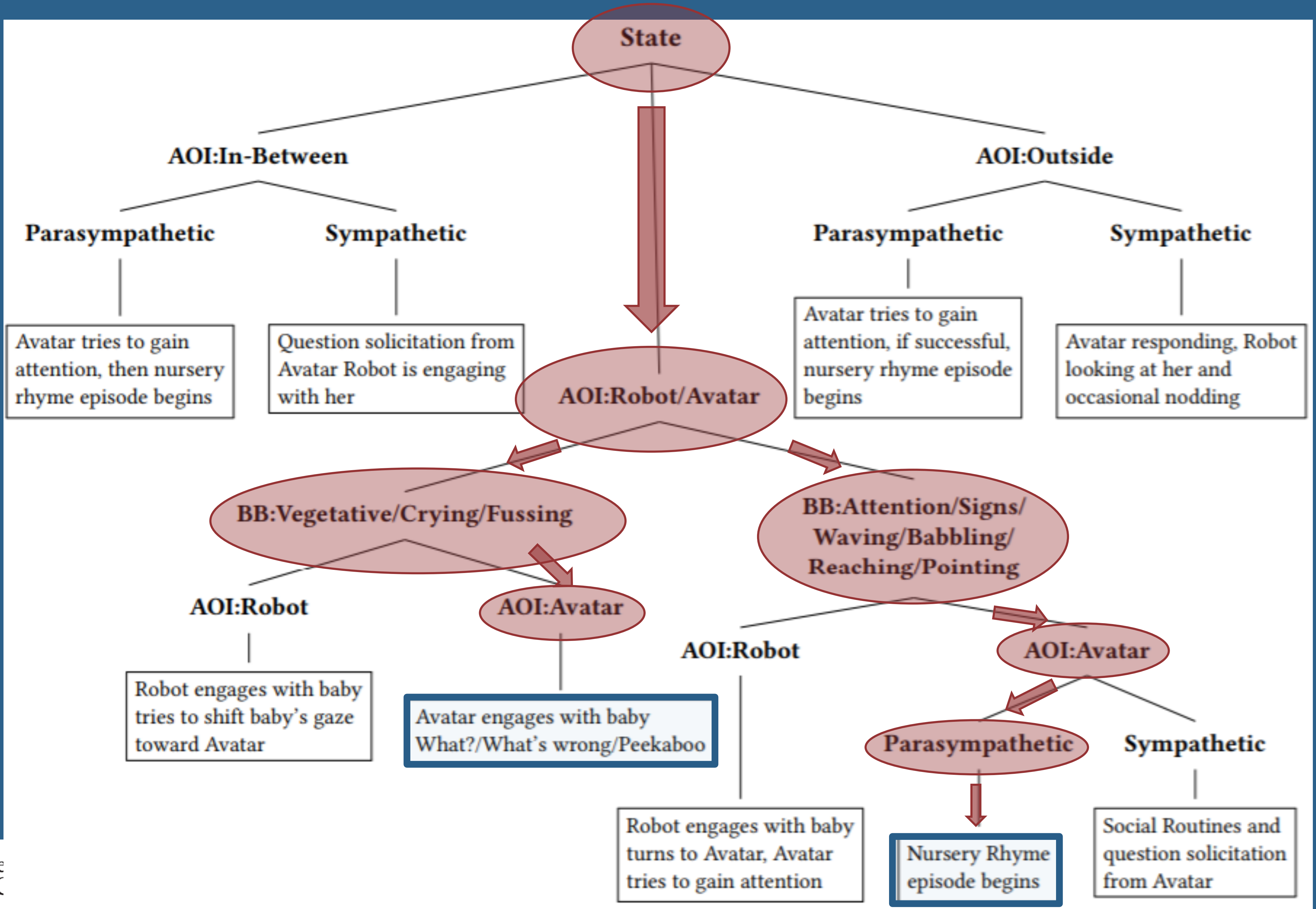
Avatar signing HELLO

Avatar nods

Avatar tells a story (PIG NR)

Robot Nods

Robot head turn



USC Unive Southern C

Evaluation

Can Rave engage the infant's attention?

Can agents elicit socially contingent interaction?

Experiment Subjects

- 8 infant participants
  - 2 females and 6 males
  - average age of 9 months and 20 days (range 7-13 months)
  - Different levels of sign-exposure
    - 5 not sign-exposed
    - 2 hearing sign-exposed
    - 1 deaf sign-exposed

Evaluation

Can Rave engage the infant's attention?

Can agents elicit socially contingent interaction?

Evaluation

Can Rave engage the infant's attention?

Can agents elicit socially contingent interaction?

Evaluation – Engagement

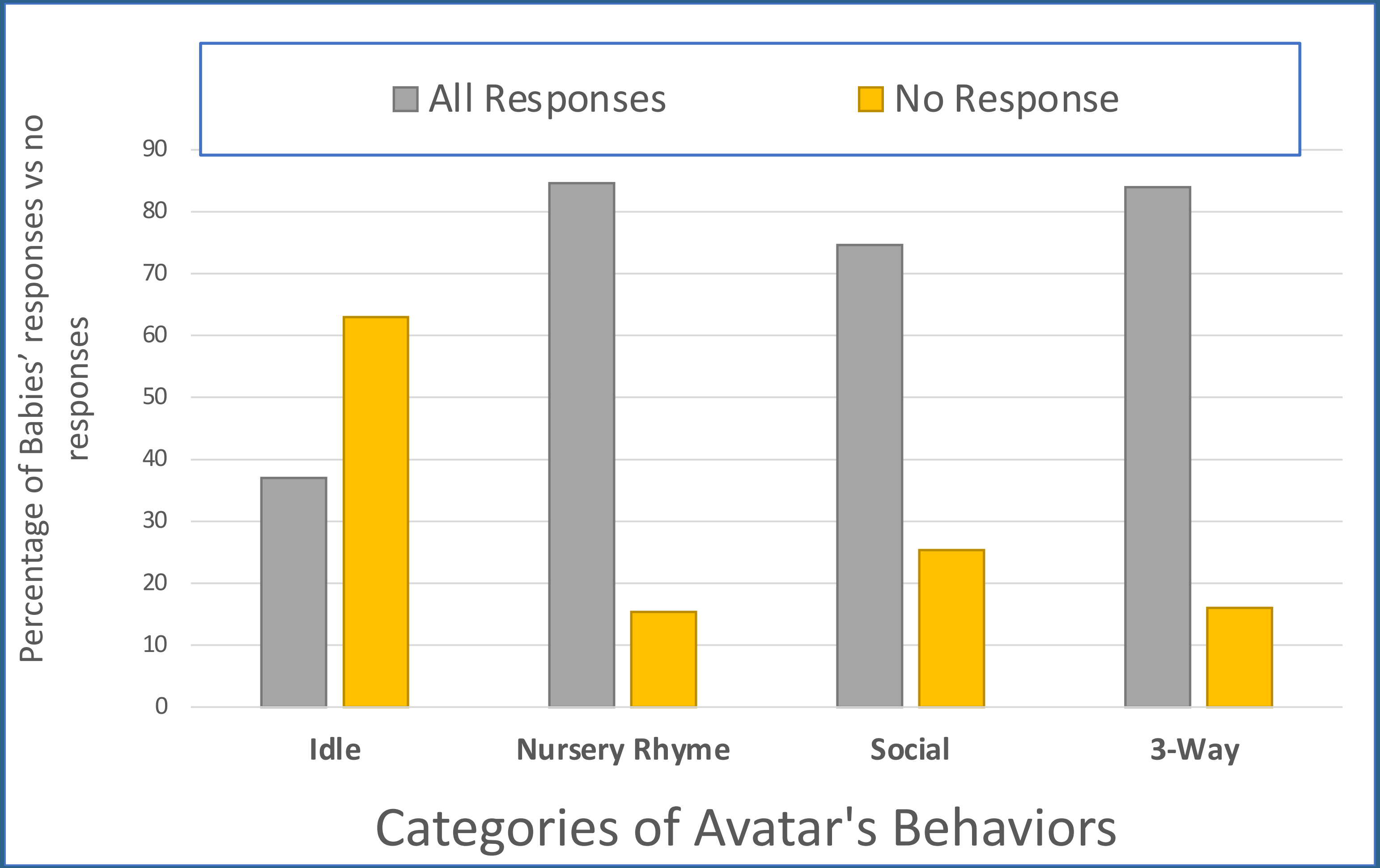
- Concern:
  - Babies would see the agents as boring and not social
  - Stranger Anxiety
- Method:
  - Overall response rate to different avatar behaviors
  - Percentage of different baby's behaviors
  - Amount of tracking the agents

Evaluation – Engagement

- All babies exhibited positive engagement behaviors
  - Immediate visual engagement (locked attention) with the agents
  - Sustained attention (persisting over time)
  - Visually tracked (gaze following) the Avatar and Robot

Response rate across Avatar’s behaviors

Overall response rate : 60% (M = 61.8, SD = 6.9)



Examples of engagement

Evaluation

Can Rave engage the infant's attention?

Can agents elicit socially contingent interaction?

Evaluation

Can Rave engage the infant's attention?

Can agents elicit socially contingent interaction?

Evaluation – Contingent Interaction

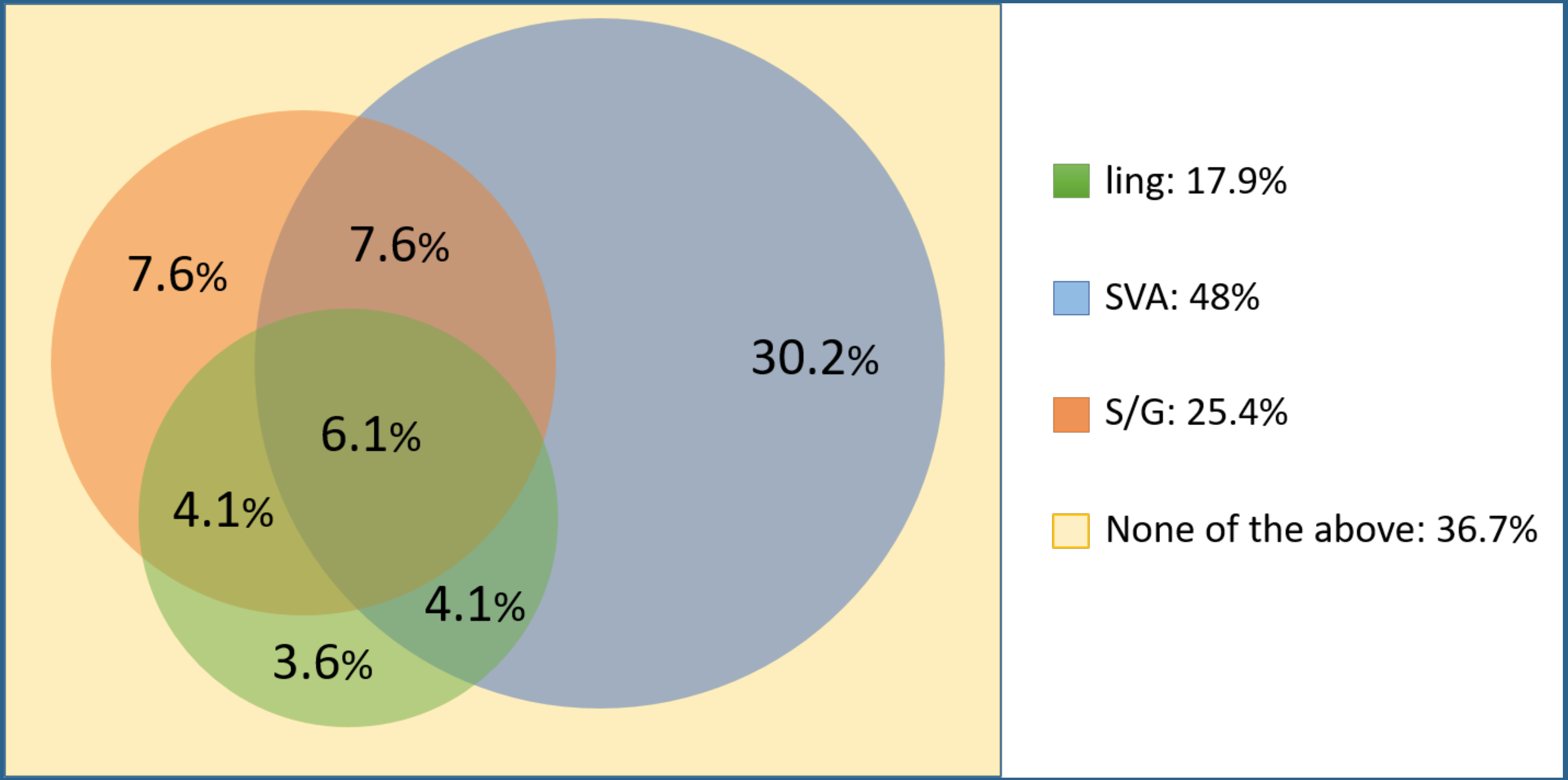
- Babies attempted to copy the avatar
- Babies performed age-appropriate proto-linguistic behavior
- Even though they were not understanding the meanings

Results - Babies' behavioral distribution

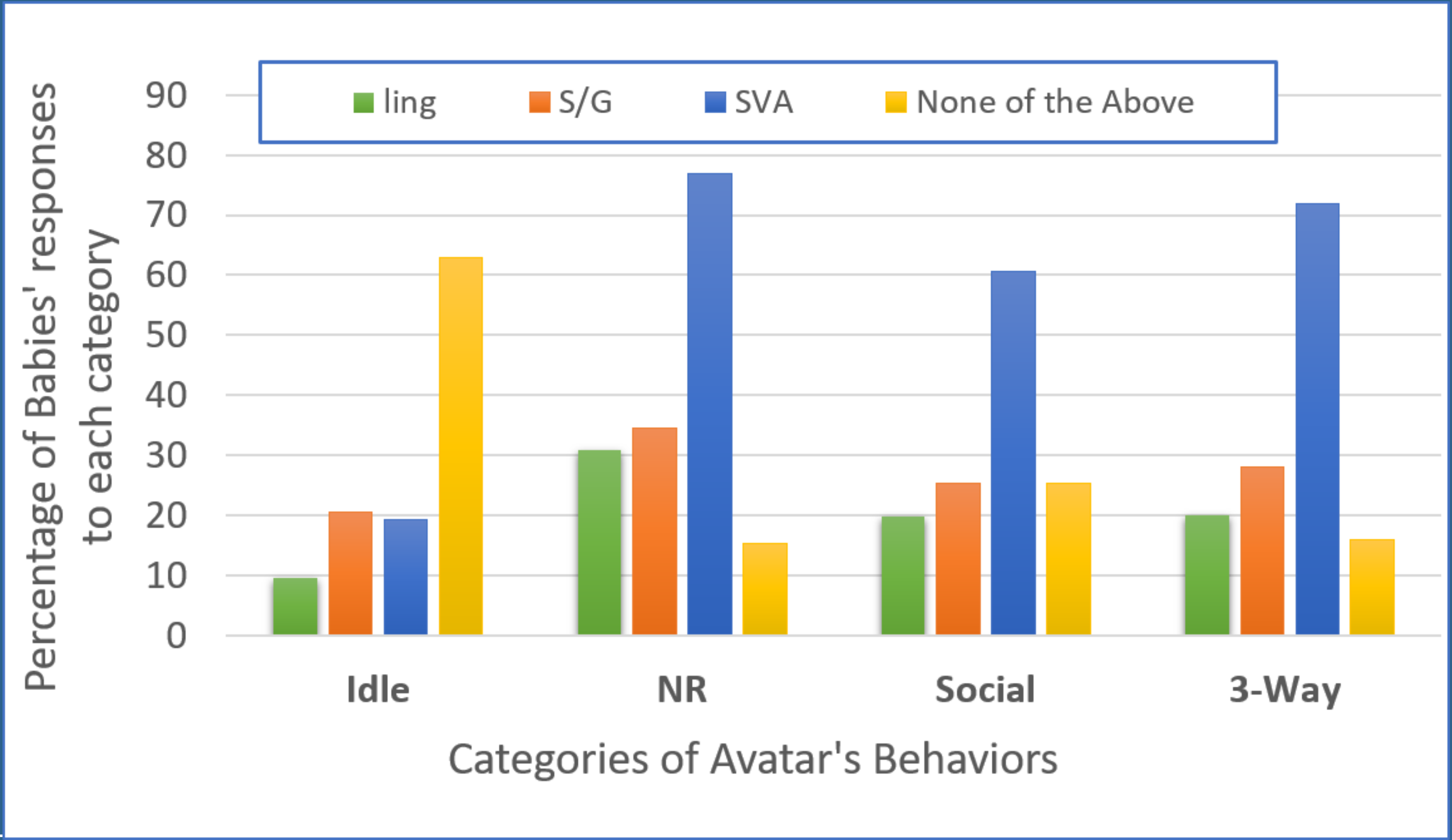
Ling (linguistic): production of manual proto-signs/sings

SVA (Sustained visual attention: baby being visually transfixed on the agents

S/G (Social/Gestural): pointing, waving, social referencing



Baby Behavior in response to different Avatar's behaviors



Example: Contingent Interaction

Summary

- Agents performing dialogue routines were successful in eliciting socially contingent conversation from the infant
- Potential viability for using this kind of system for language learning in young infants

Takeaway w.r.t. Research Question

RQ 2: What kinds of multimodal feedback can agents recognize and how can they be used to adapt the dialogue management policies of the agent?

- RAVE is able to pick up multimodal input (feedback) from babies
- It can successfully use this feedback to adapt the dialogue management policy to maintain a contingent interaction

Q&A

USC University of Southern California