# Neural models of compositionality

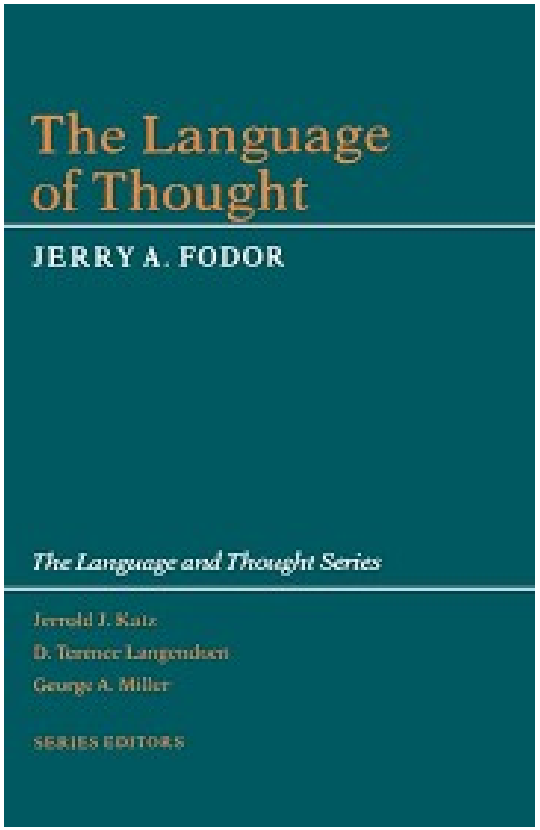
**Alexandre Pouget**

**Reidar Riveland, Pablo Tano, Jacob Bakermans, Charles Findling**

**Basic Neuroscience Department**

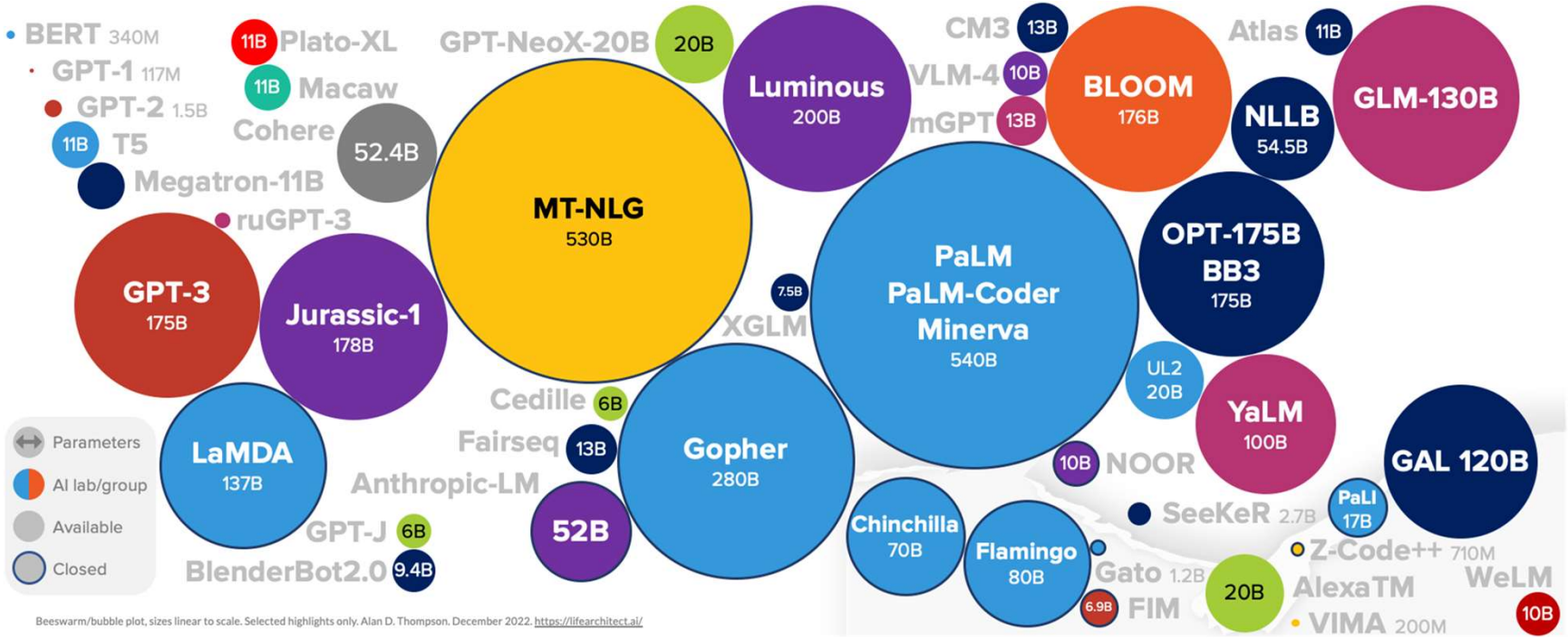
UNIVERSITÉ DE GENÈVE

# Why is connectionism a bad idea?




Jerry Fodor

# Large Language Models



# Instructing RNN with natural language

Reidar Riveland

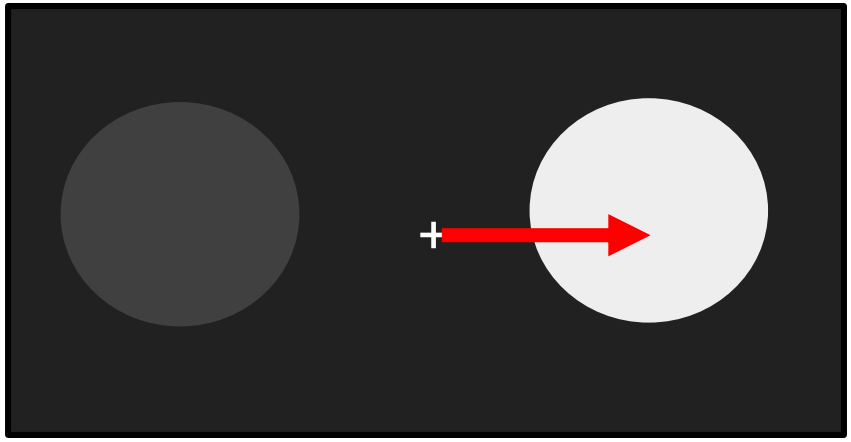
# Training animals

- Operant conditioning: weeks to months of training

(Courtesy of Zach Mainen)

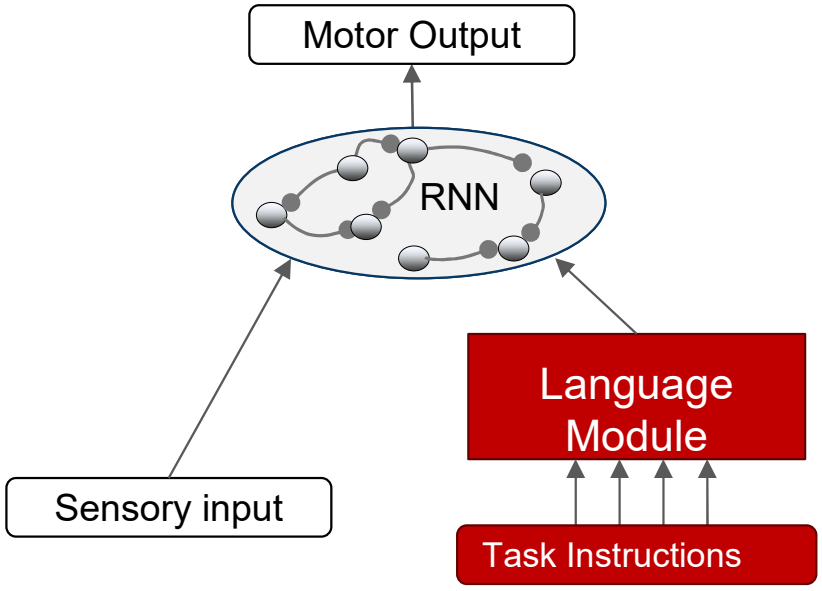
# Humans: ultra-fast learning with verbal instructions

Respond in the direction
of the stimulus presented
with highest contrast

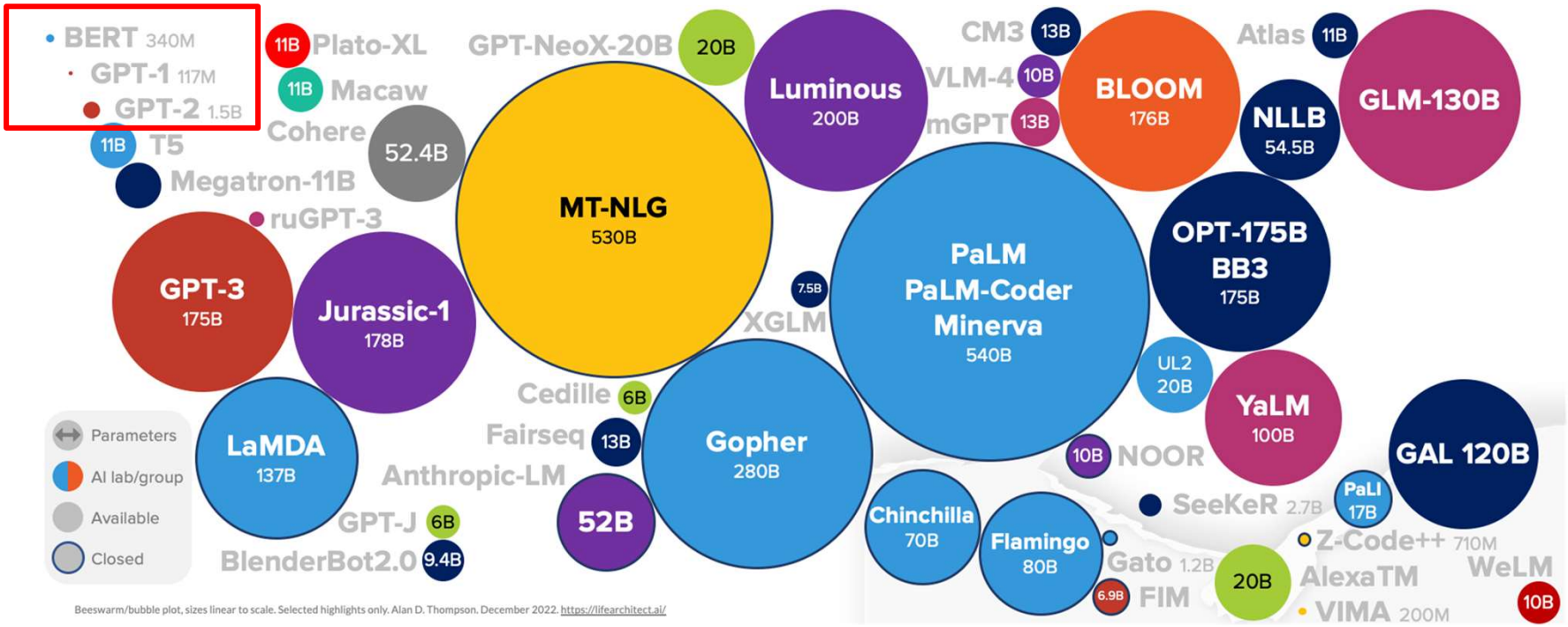



Instructed Models

- Trained on 45 tasks with verbal instructions
- Tested on 5 new tasks with verbal instructions



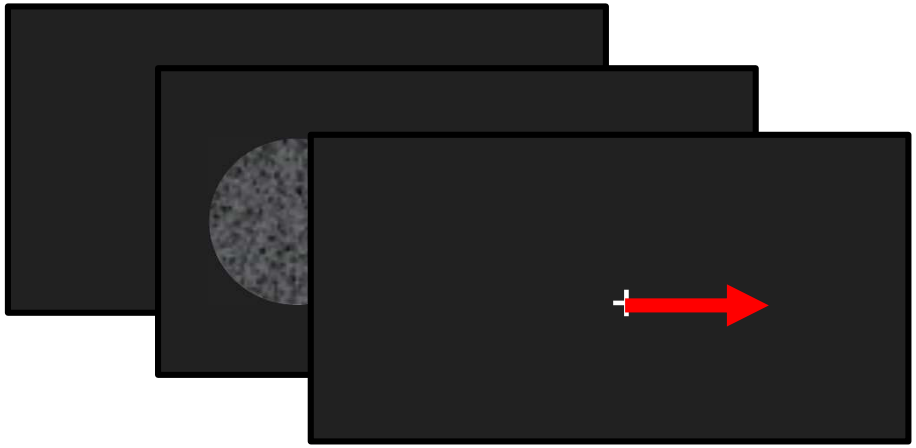
# Large Language Models



# Tasks

**Anti go:**

‘Respond in the direction opposite to the stimulus’



# Tasks

## 15 different instructions per task

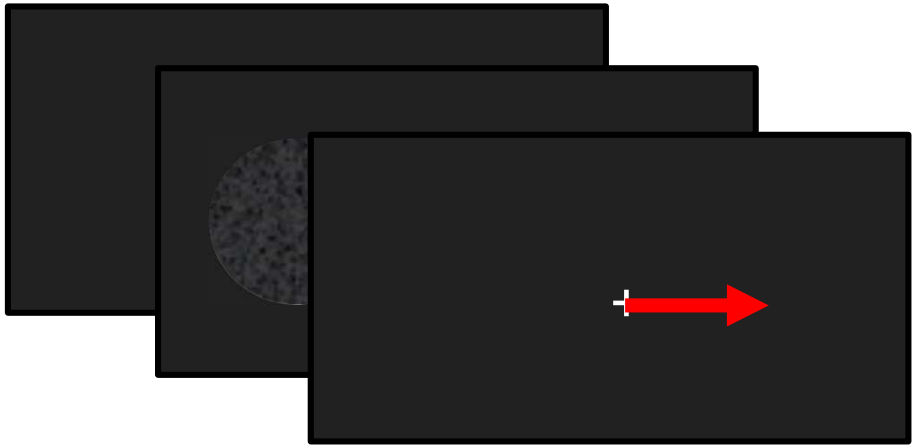
**Anti RT Go:**
‘Respond in the opposite direction of stimuli immediately’

‘respond in the opposite orientation of stimulus when directed’, ‘respond in the reverse direction of stimulus when prompted’, ‘choose the opposite of the displayed direction when prompted’, ‘when cued choose the converse direction’, ‘respond with the converse orientation when prompted’, ‘go in opposite direction when prompted’, ‘choose converse orientation when prompted’, ‘select the reverse of the displayed orientation when prompted’, ‘when directed respond opposite of the displayed stimulus’, ‘choose the inverse of the shown direction when directed’, ‘when prompted choose the opposite of direction shown’, ‘when cued select the inverse of presented stimulus’, ‘pick the reverse of the stimulus displayed when cued’, ‘select the opposite of the displayed direction when cued’, ‘when directed respond in the opposite of the displayed direction’

# Tasks

**Conditional Decision Making (CondDM):**

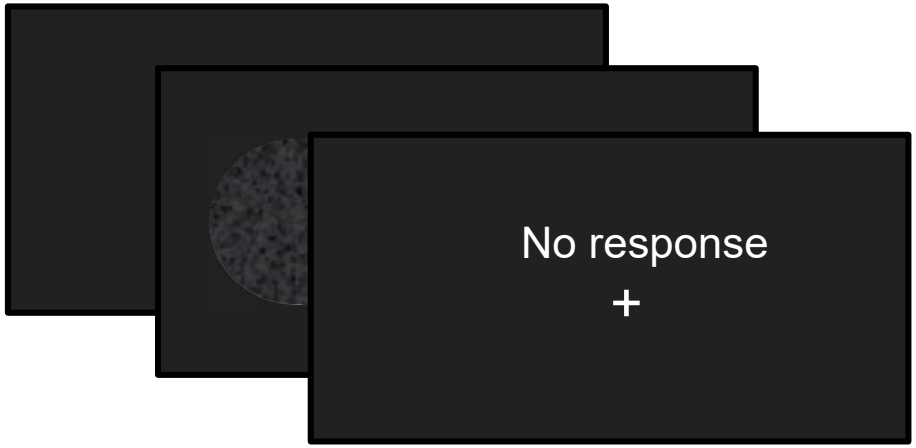
‘respond to the strongest stimulus only if you are confident otherwise do not respond’

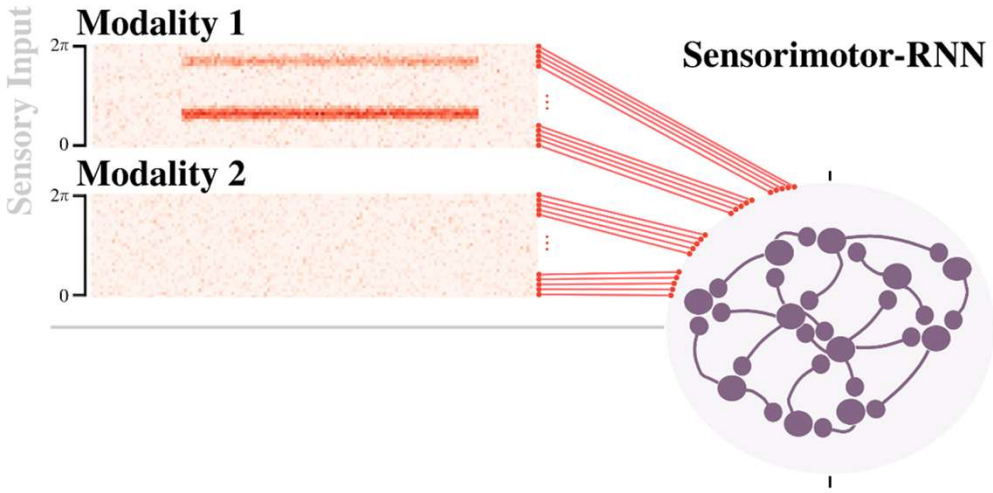


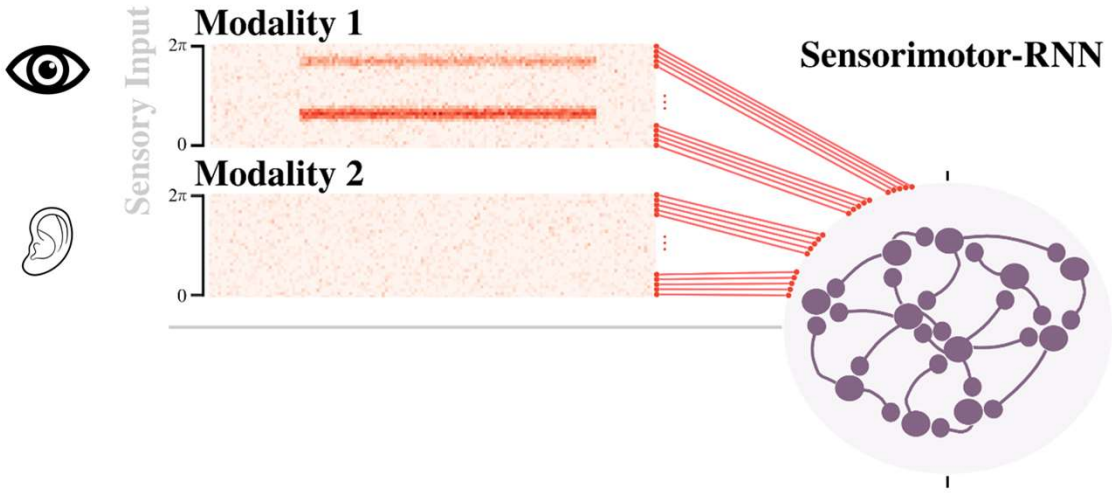
# Tasks

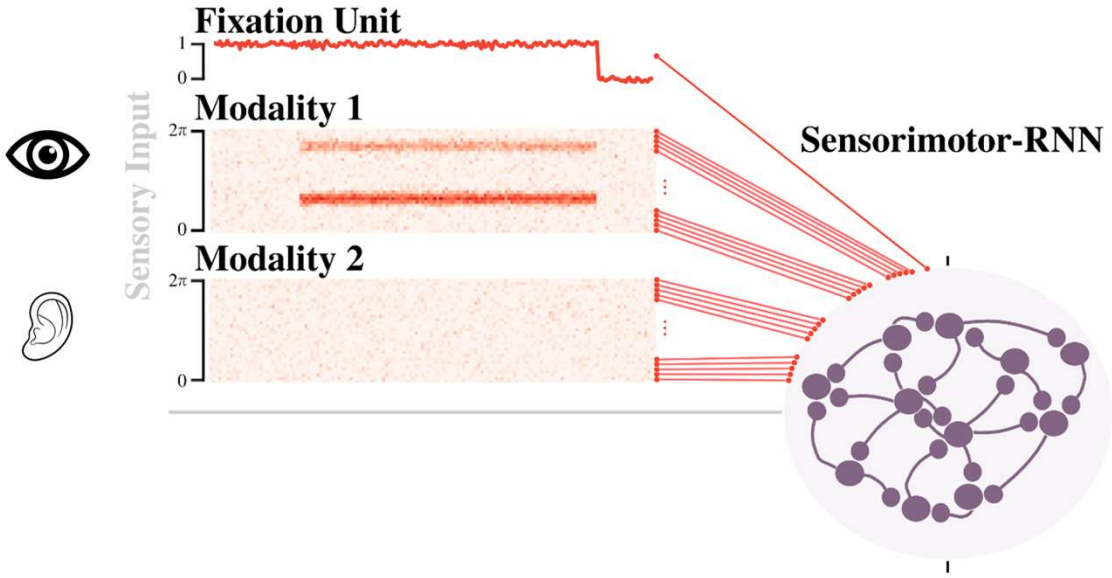
**Conditional Decision Making (CondDM):**

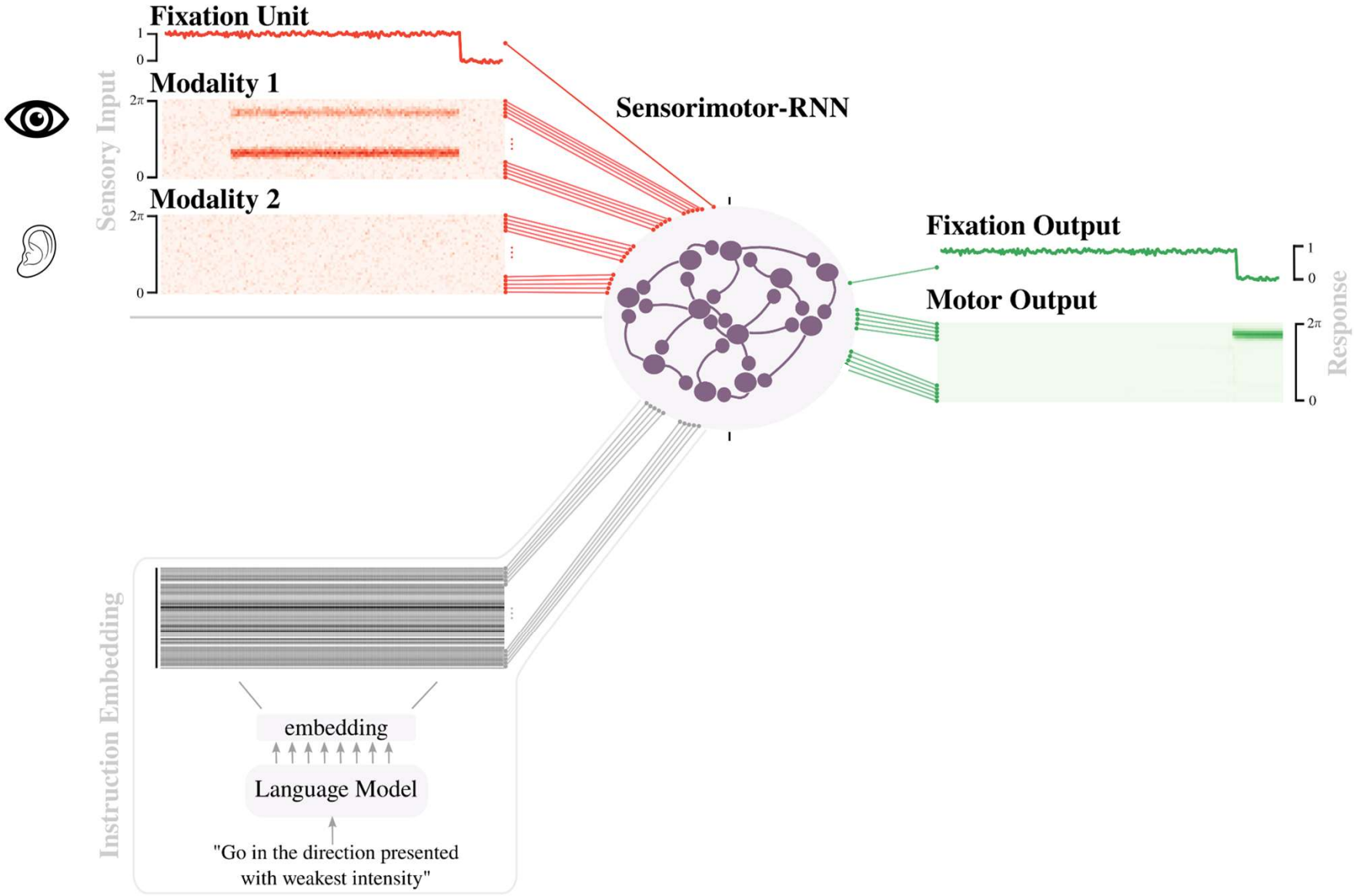
‘respond to the strongest stimulus only if you are confident otherwise do not respond’




Sensory Input
Modality 1
2π
0
Modality 2
2π
0
Sensorimotor-RNN


Sensory Input
Modality 1
2π
0
Modality 2
2π
0
Sensorimotor-RNN


Fixation Unit
1
0
Modality 1
2π
0
Modality 2
2π
0
Sensory Input
Sensorimotor-RNN


Fixation Unit
1
0
Modality 1
2π
0
Modality 2
2π
0
Sensory Input
Sensorimotor-RNN
Fixation Output
1
0
Motor Output
2π
0
Response
Instruction Embedding
embedding
Language Model
"Go in the direction presented
with weakest intensity"

# Compositional task?

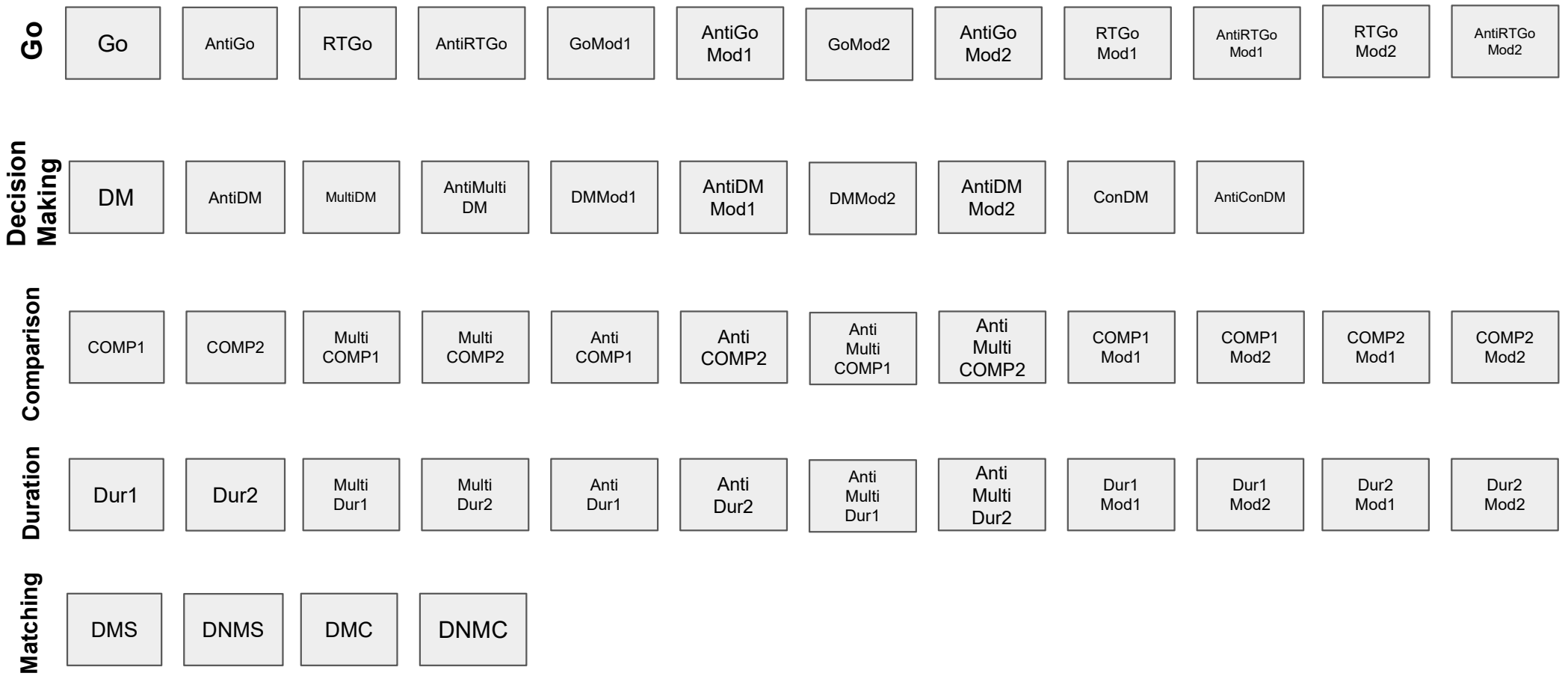
Two compositional aspects to this project:

- The mapping from instructions to semantic relies on the compositional nature of natural language

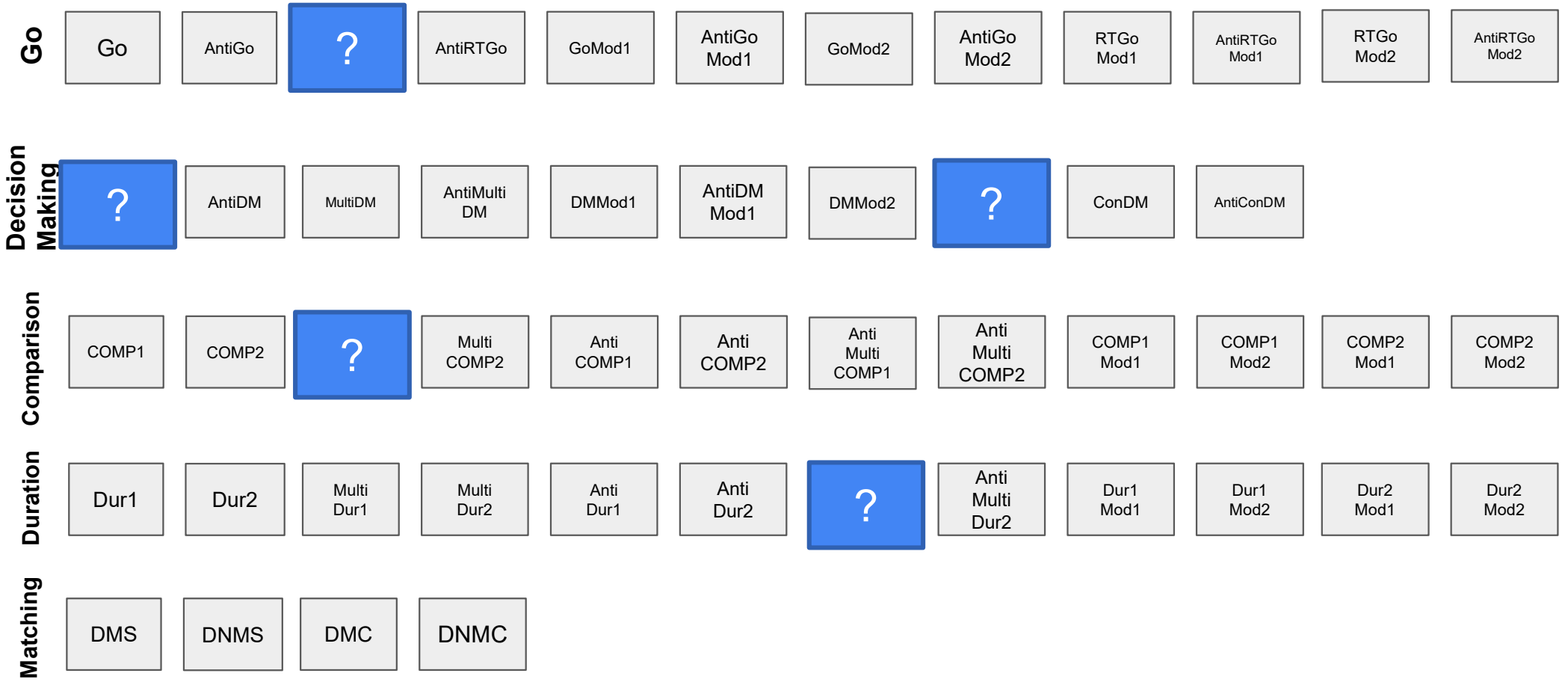
- Several of the tasks can be thought as combination of other tasks (Yang et al, 2019):

    Anti decision making = Decision making + anti go

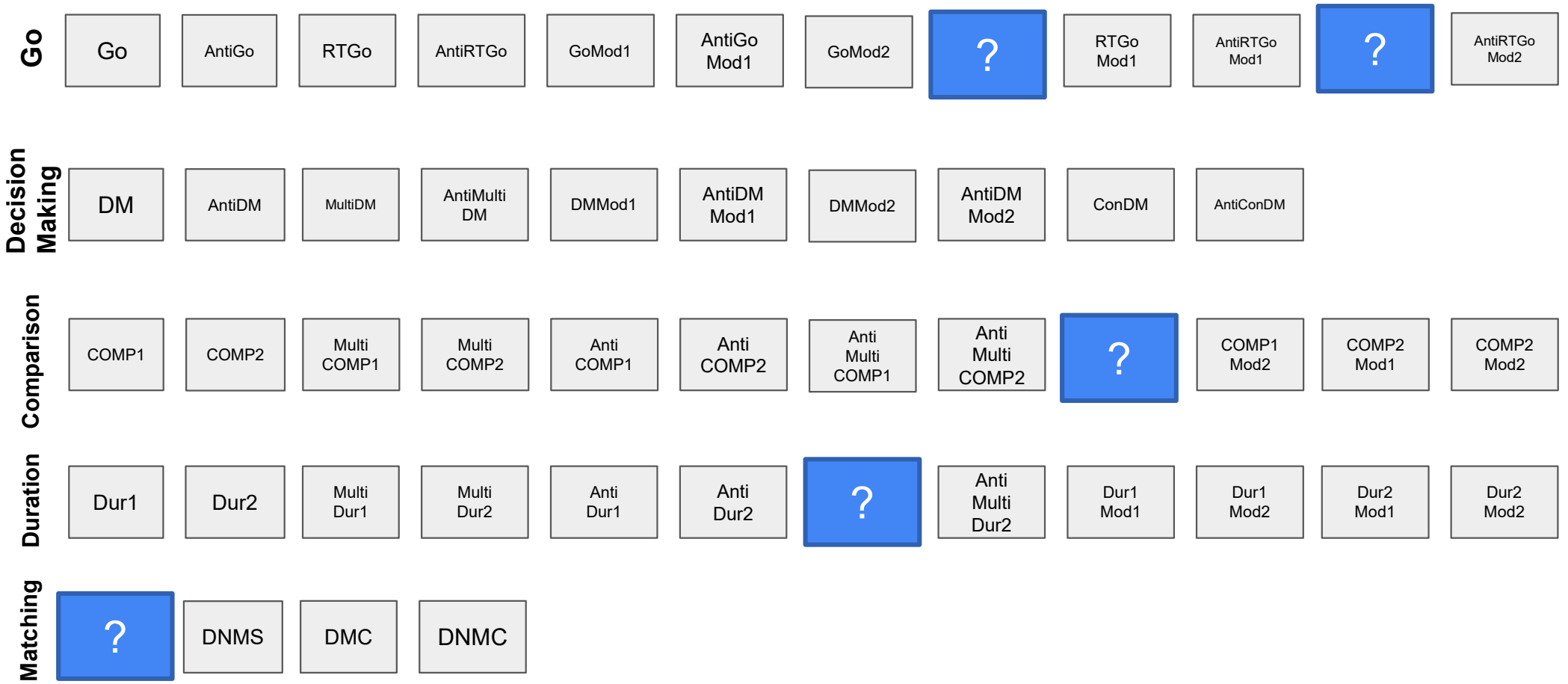
# 50 Tasks


Go
Go
AntiGo
RTGo
AntiRTGo
GoMod1
AntiGo Mod1
GoMod2
AntiGo Mod2
RTGo Mod1
AntiRTGo Mod1
RTGo Mod2
AntiRTGo Mod2
Decision Making
DM
AntiDM
MultiDM
AntiMulti DM
DMMod1
AntiDM Mod1
DMMod2
AntiDM Mod2
ConDM
AntiConDM
Comparison
COMP1
COMP2
Multi COMP1
Multi COMP2
Anti COMP1
Anti COMP2
Anti Multi COMP1
Anti Multi COMP2
COMP1 Mod1
COMP1 Mod2
COMP2 Mod1
COMP2 Mod2
Duration
Dur1
Dur2
Multi Dur1
Multi Dur2
Anti Dur1
Anti Dur2
Anti Multi Dur1
Anti Multi Dur2
Dur1 Mod1
Dur1 Mod2
Dur2 Mod1
Dur2 Mod2
Matching
DMS
DNMS
DMC
DNMC

# Train on 45, test on 5 held out tasks

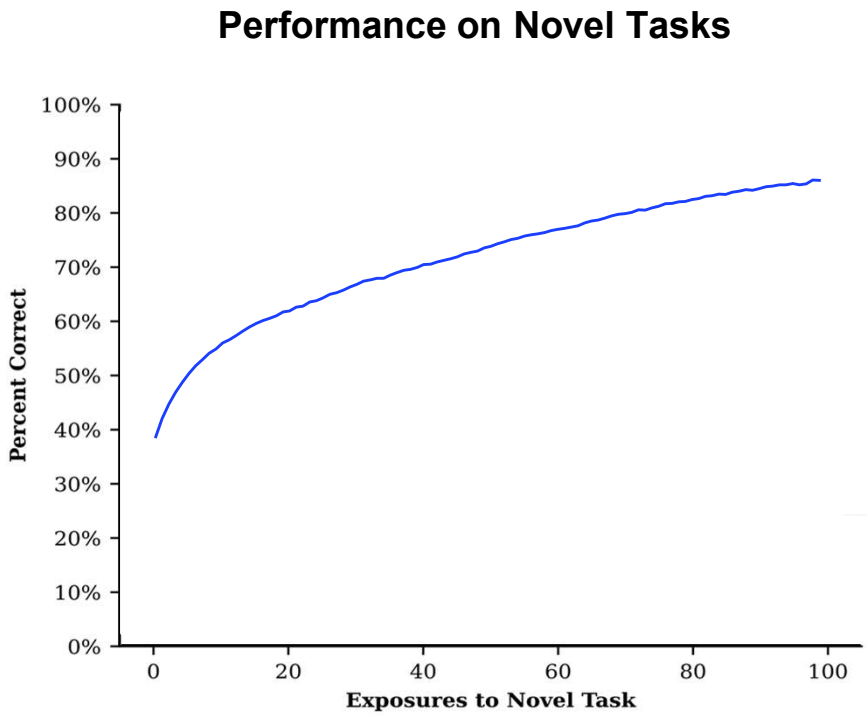


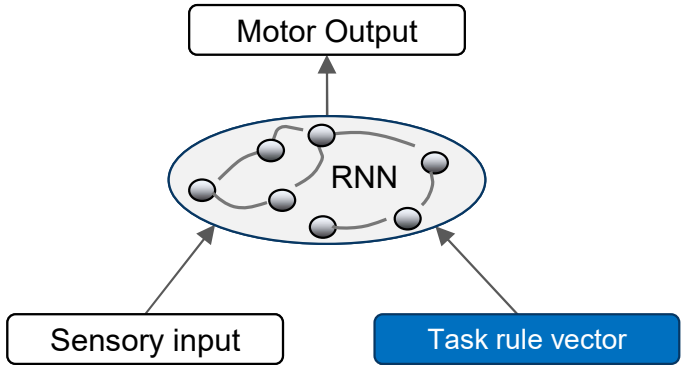
# Train on 45, test on 5 held out tasks



# 0-shot performance

(Riveland and Pouget, 2024)



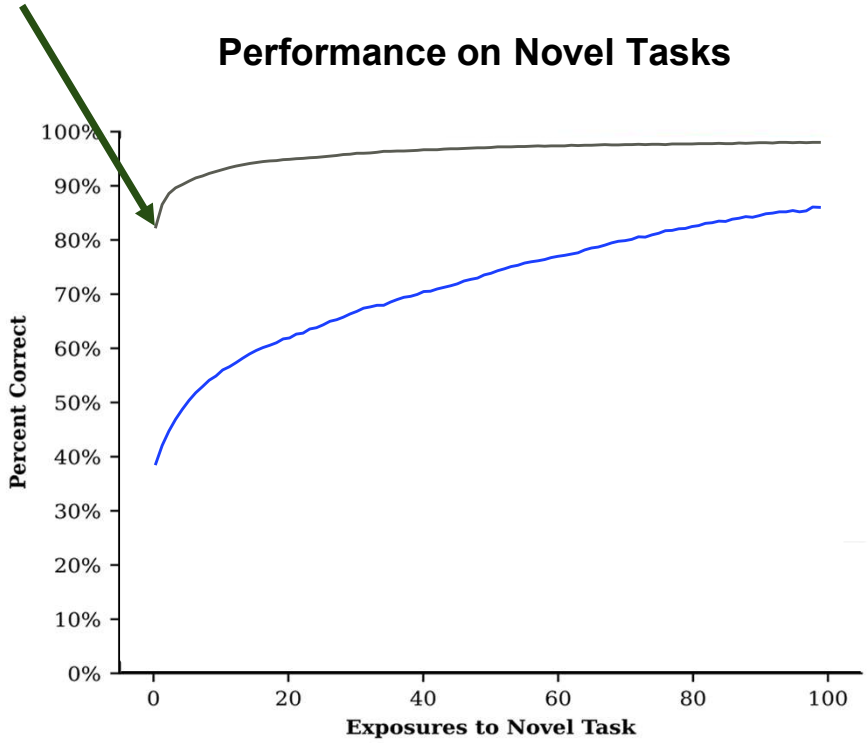



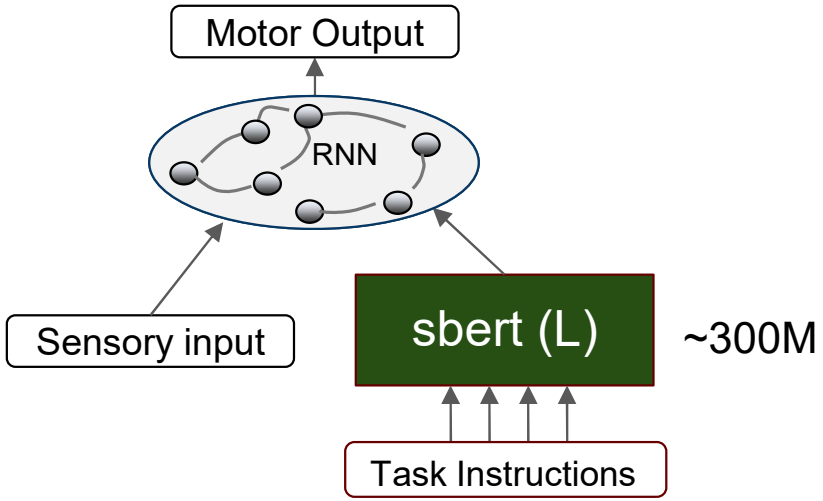

simpleNet (Yang et al, 2019)

# 0-shot performance

(Riveland and Pouget, 2024)

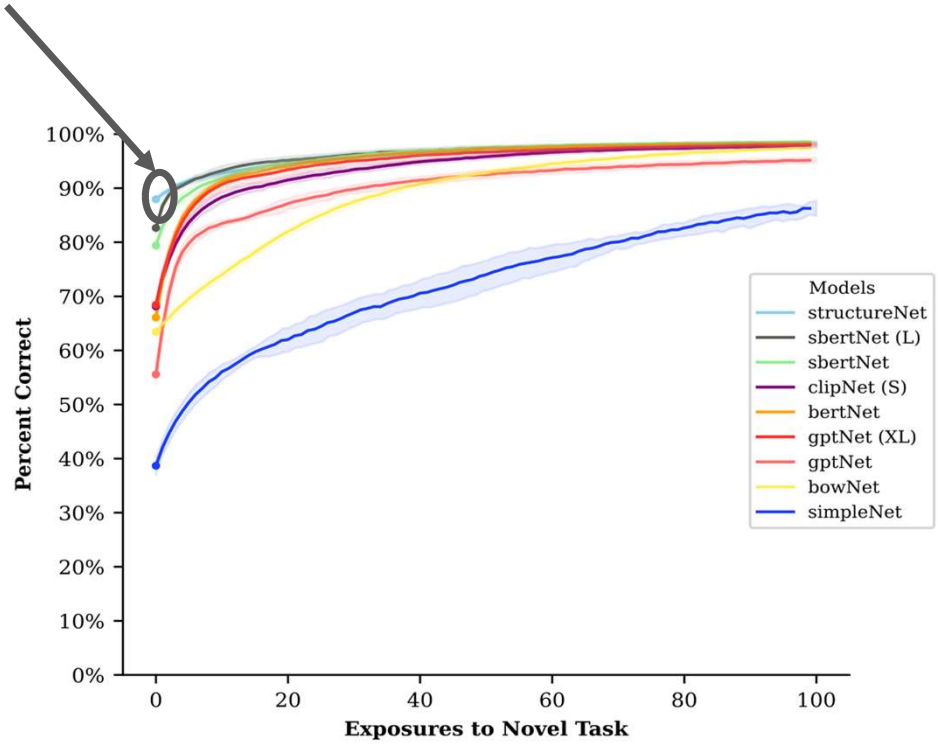
**83% 0-shot performance**



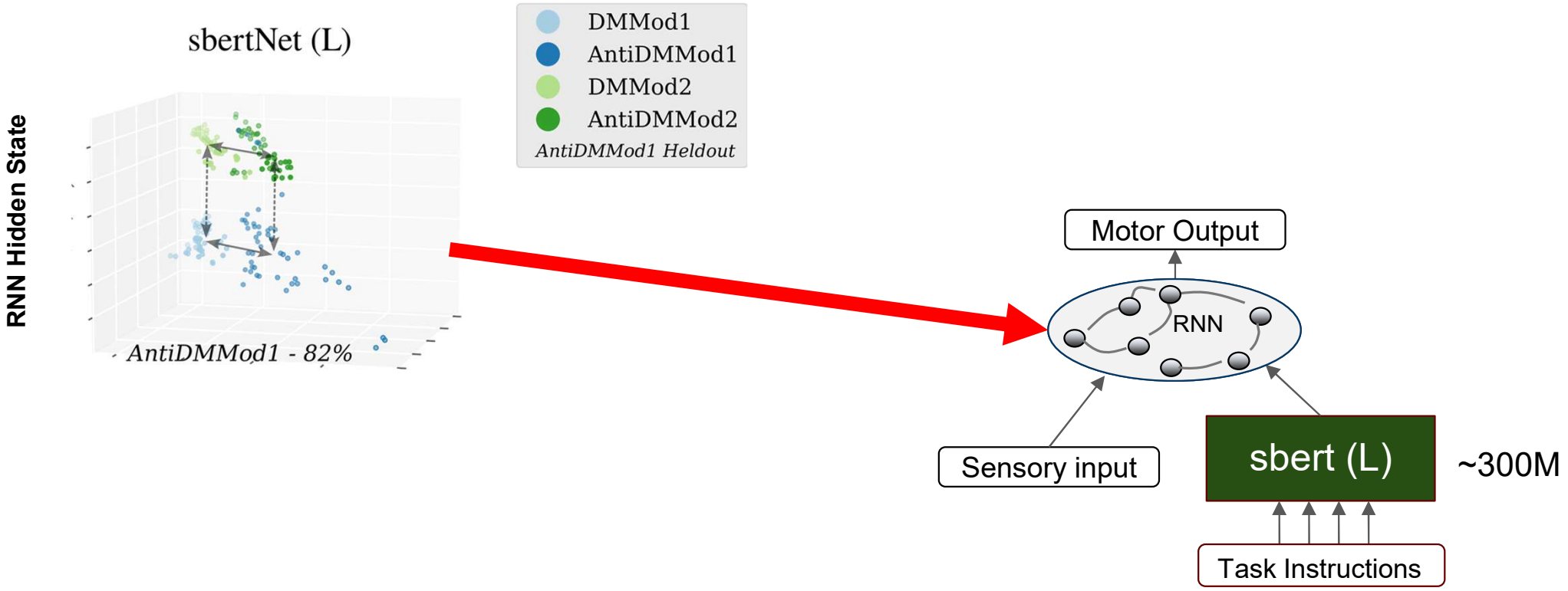



# 0-shot performance

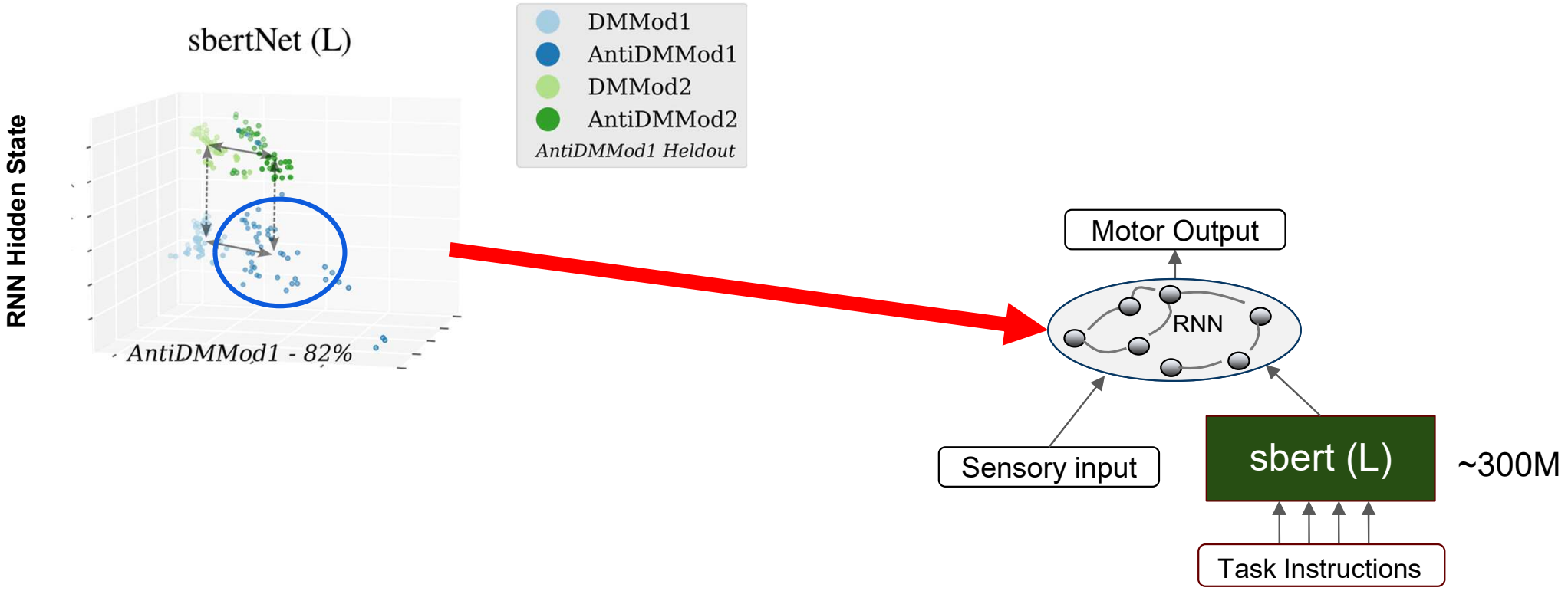
## 88% Upper limit generalization



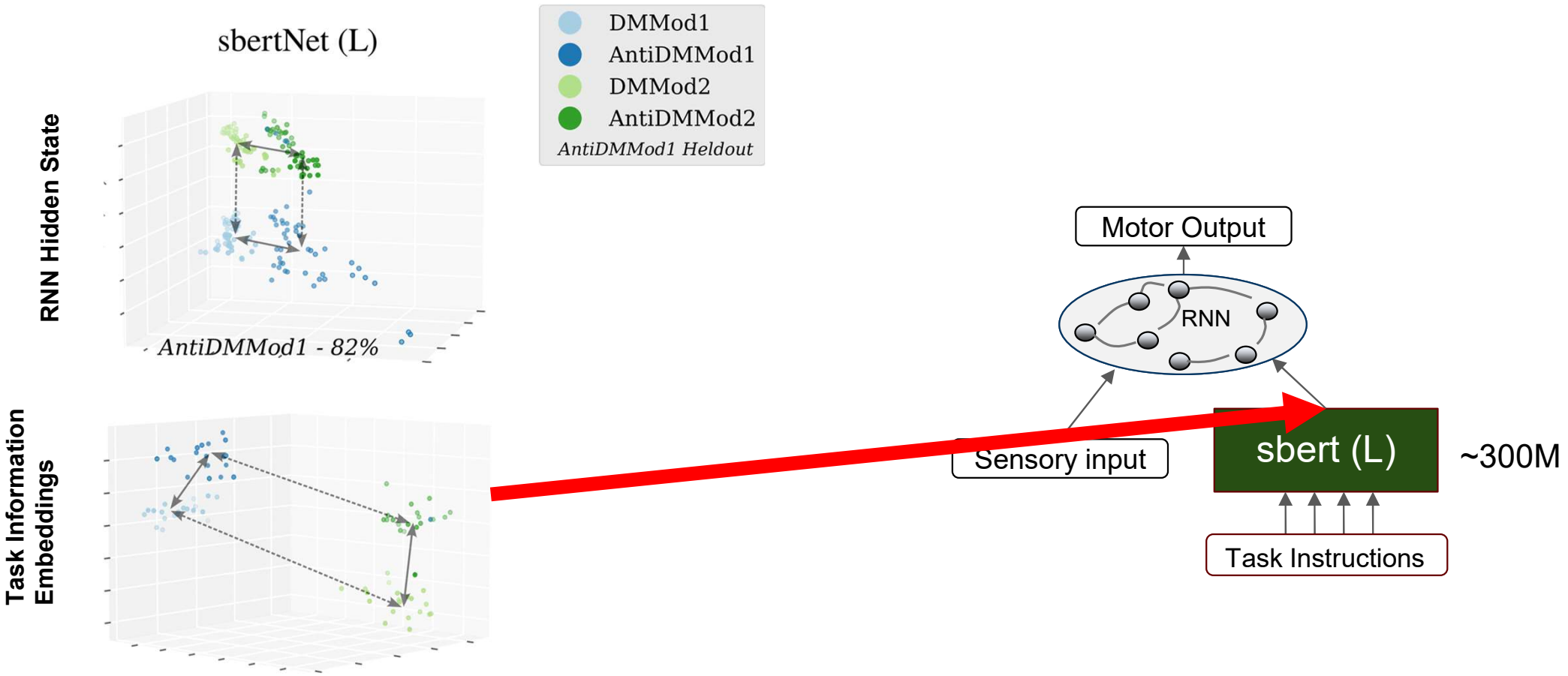
# Task Neural Space



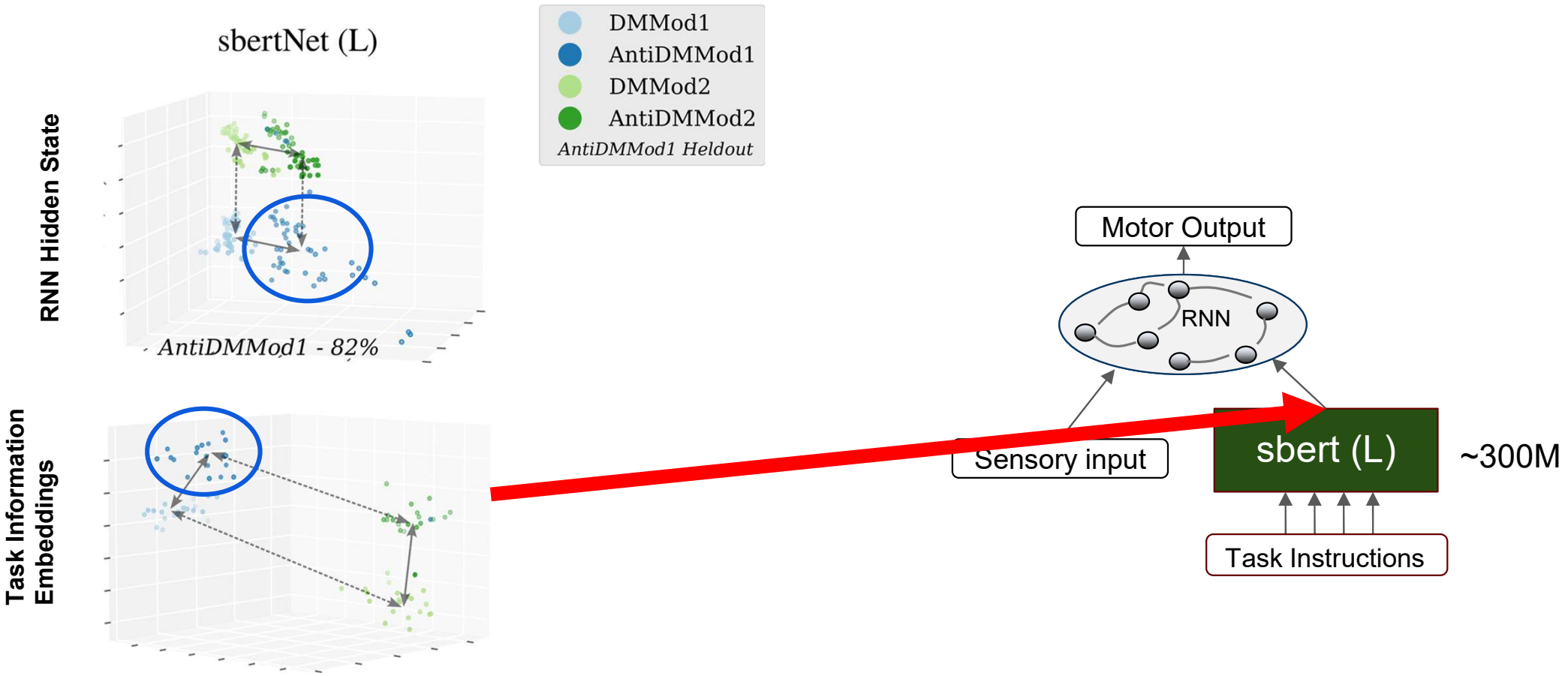
# Task Neural Space



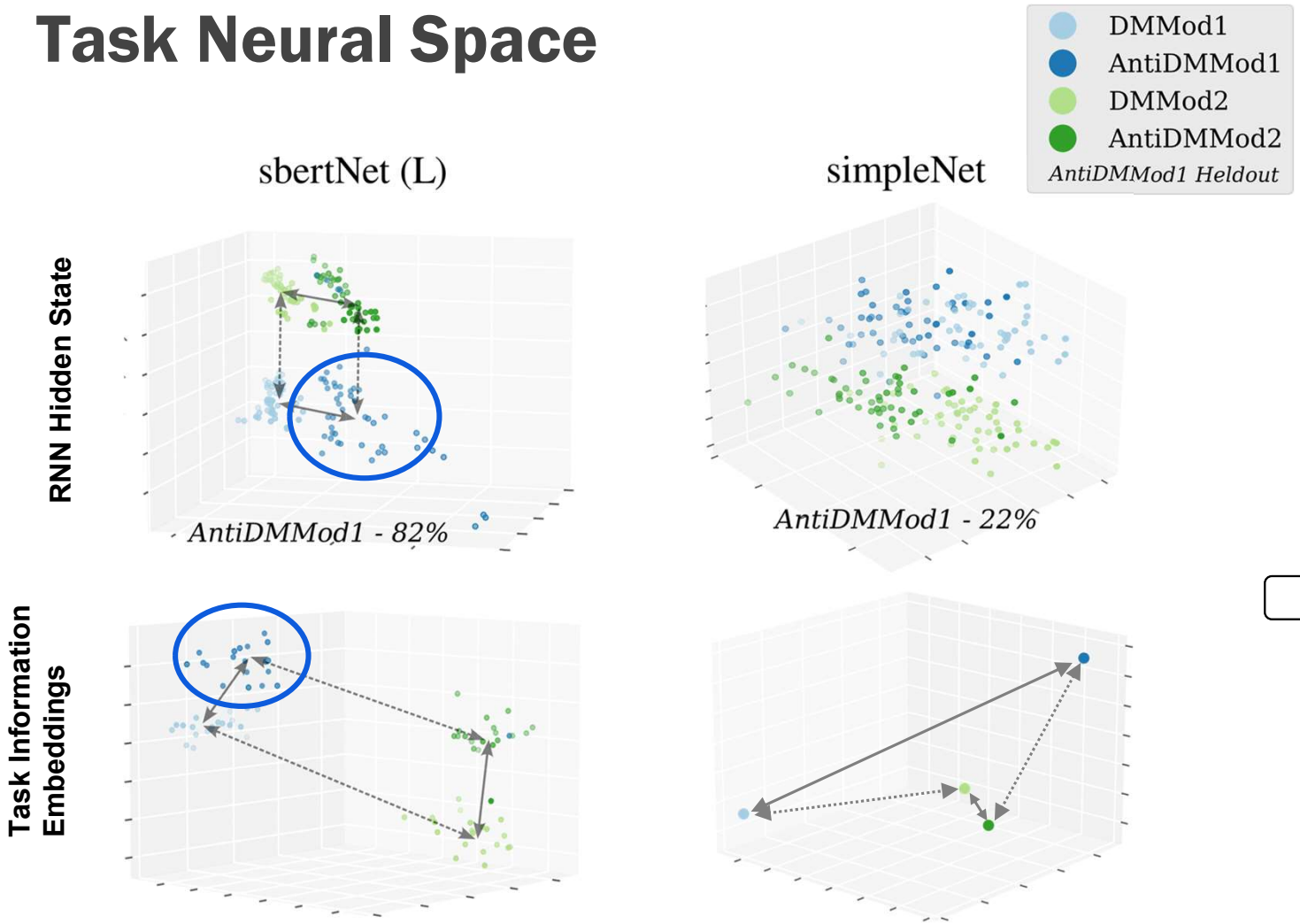
# Task Neural Space

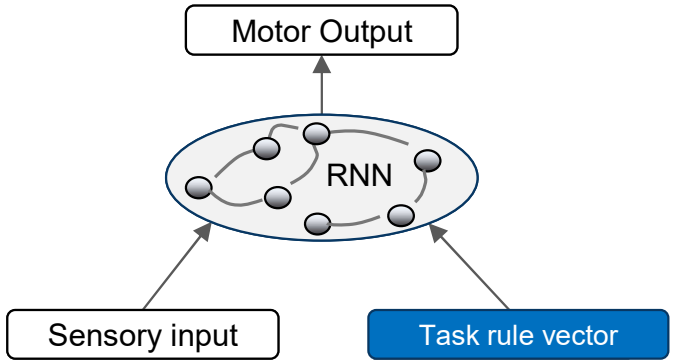


# Task Neural Space



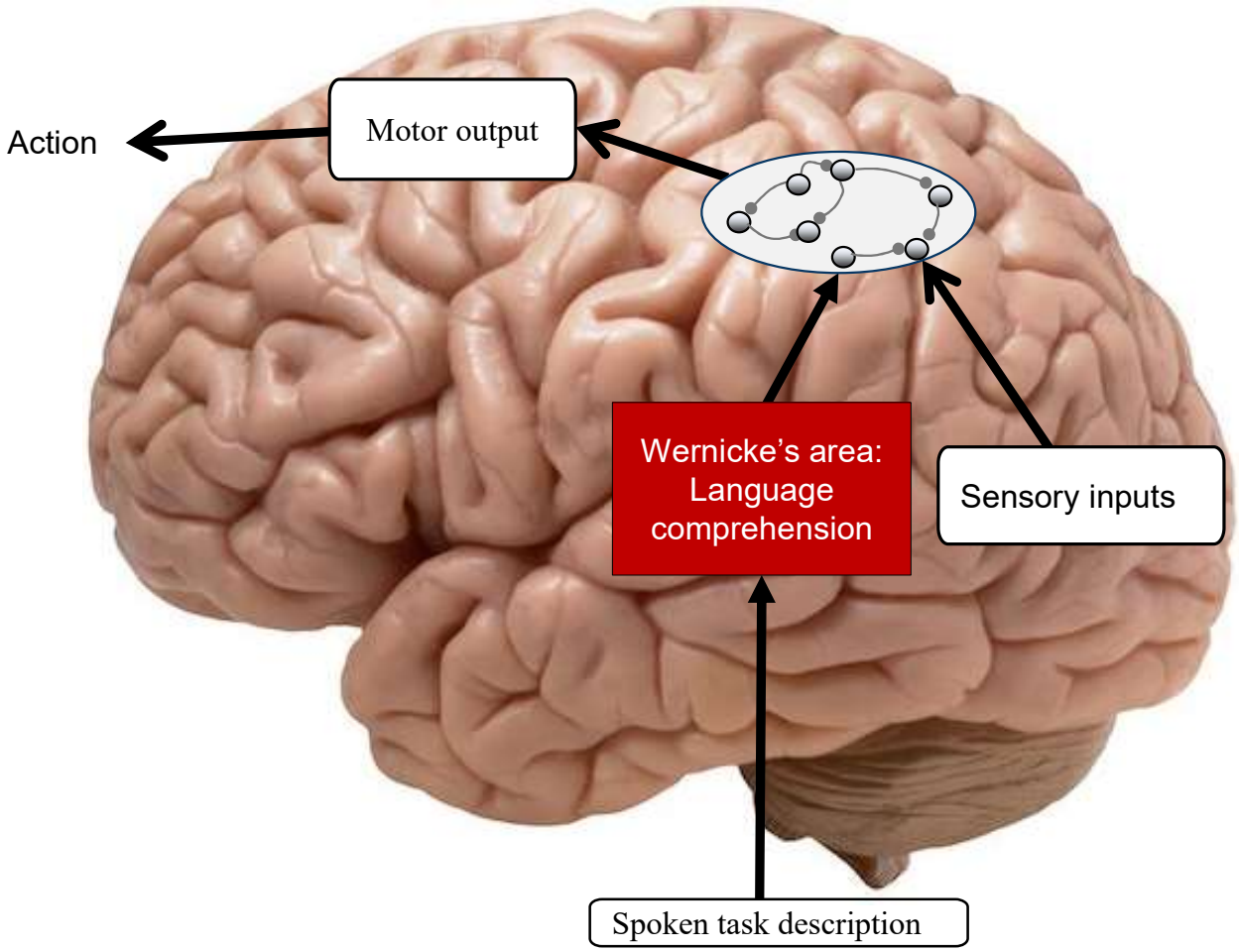
# Task Neural Space



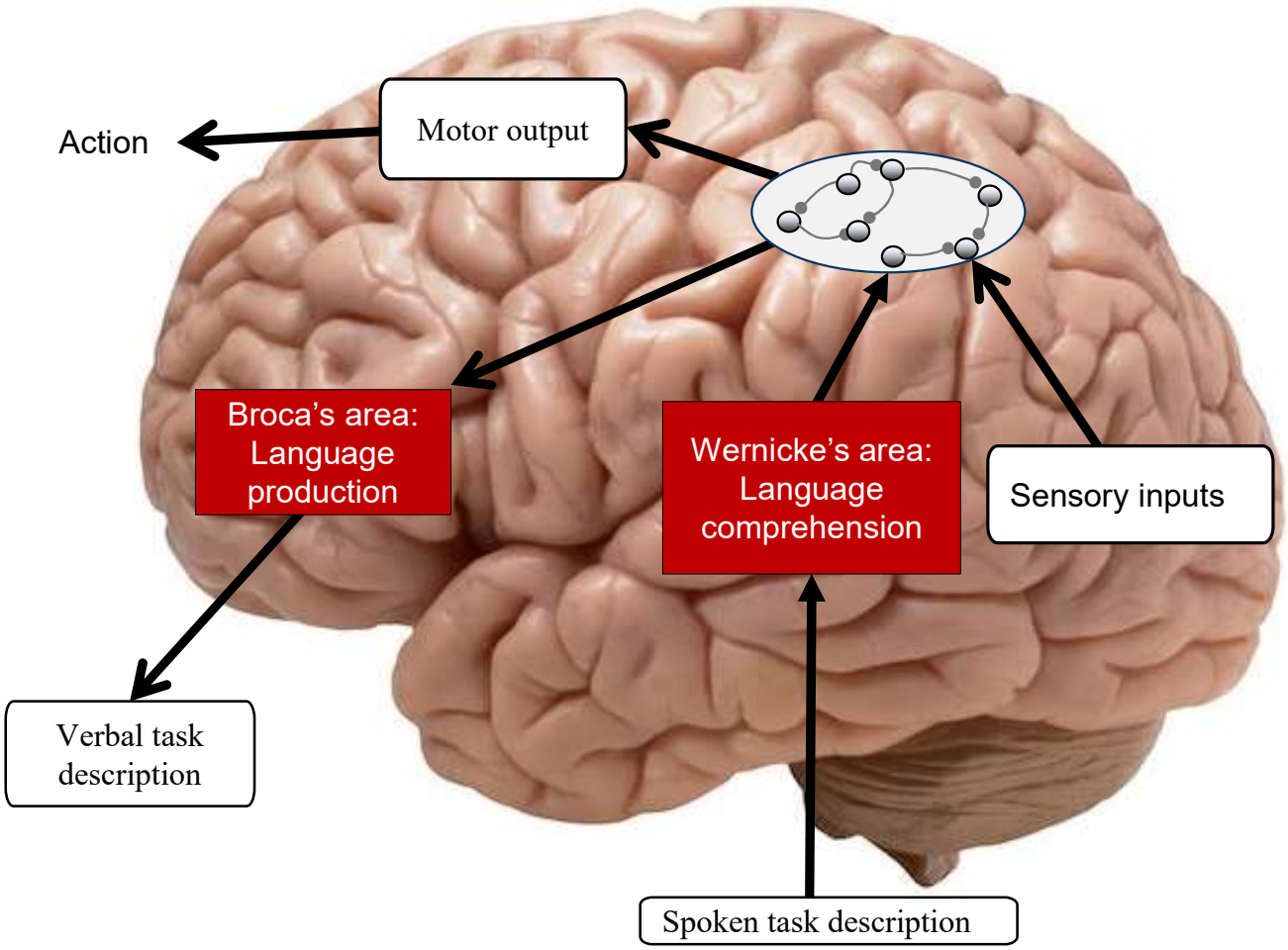




simpleNet (Yang et al, 2019)

# Language areas

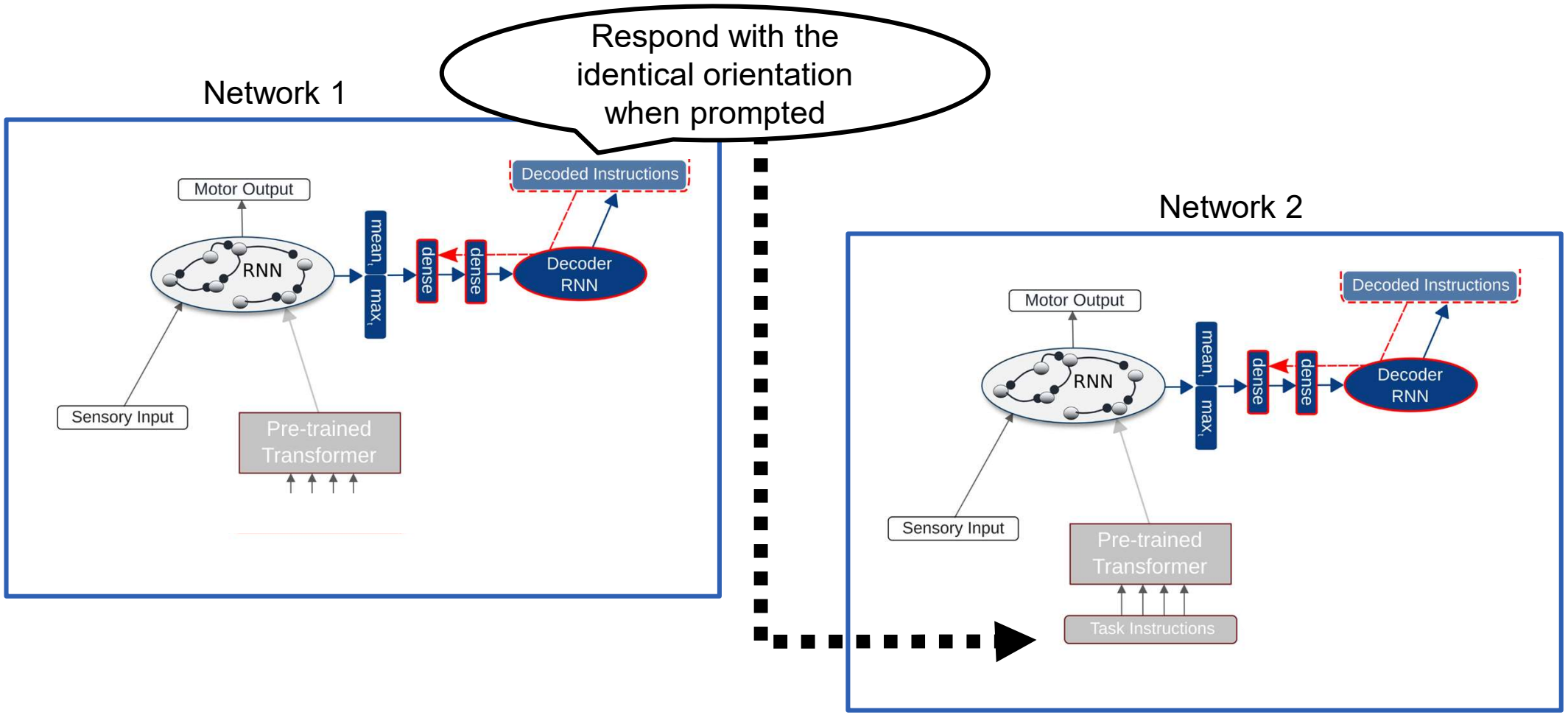


# Language areas



(Riveland and Pouget, 2024)

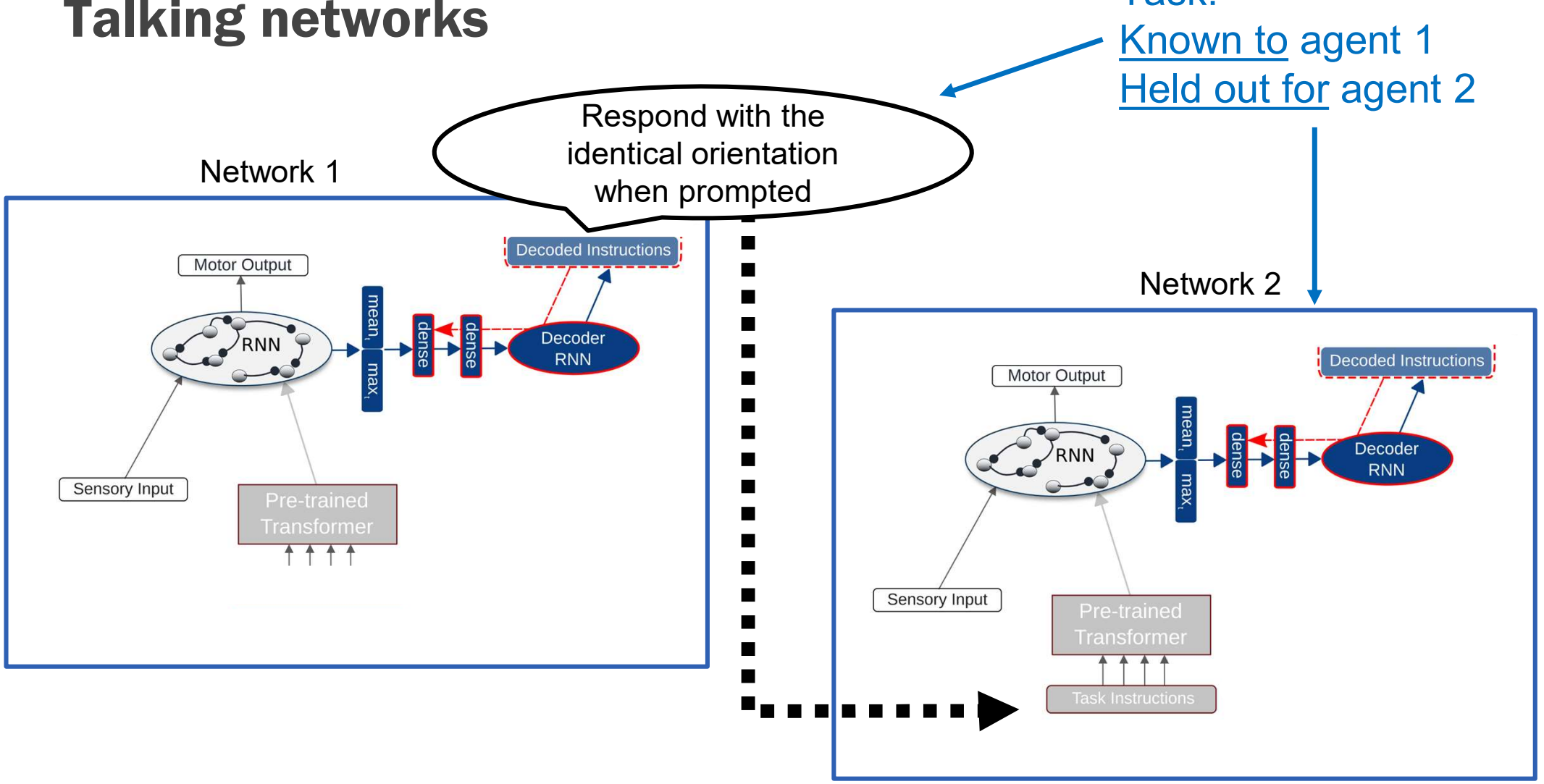
# Talking networks



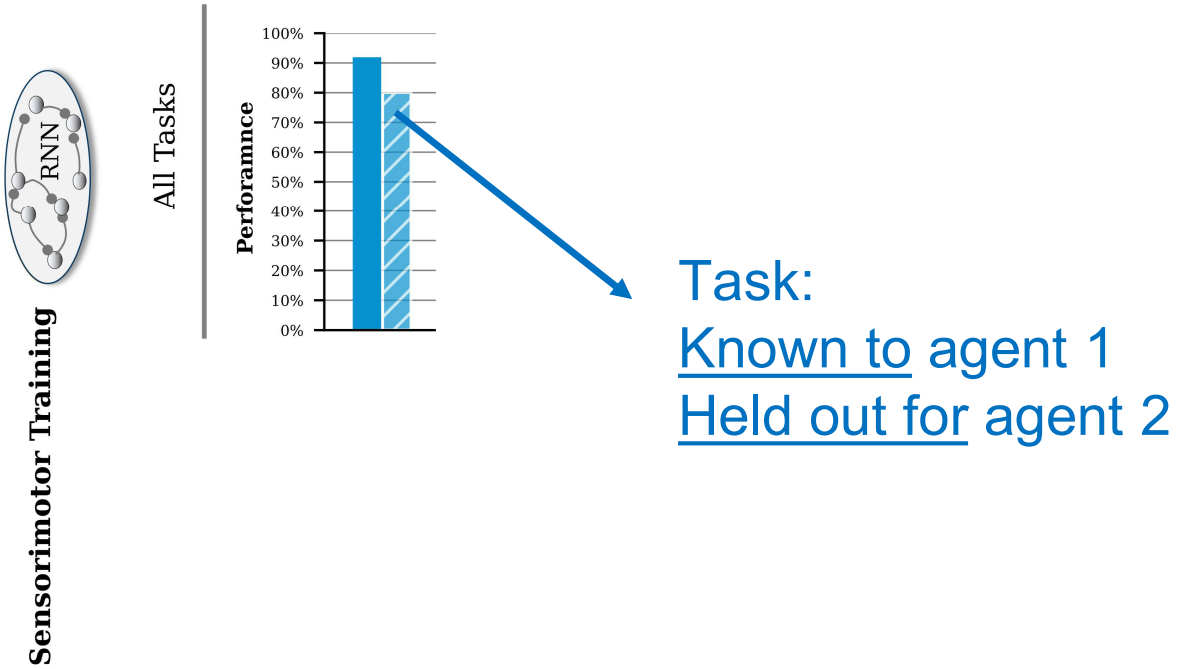
# Talking networks

Task:
Known to agent 1
Held out for agent 2

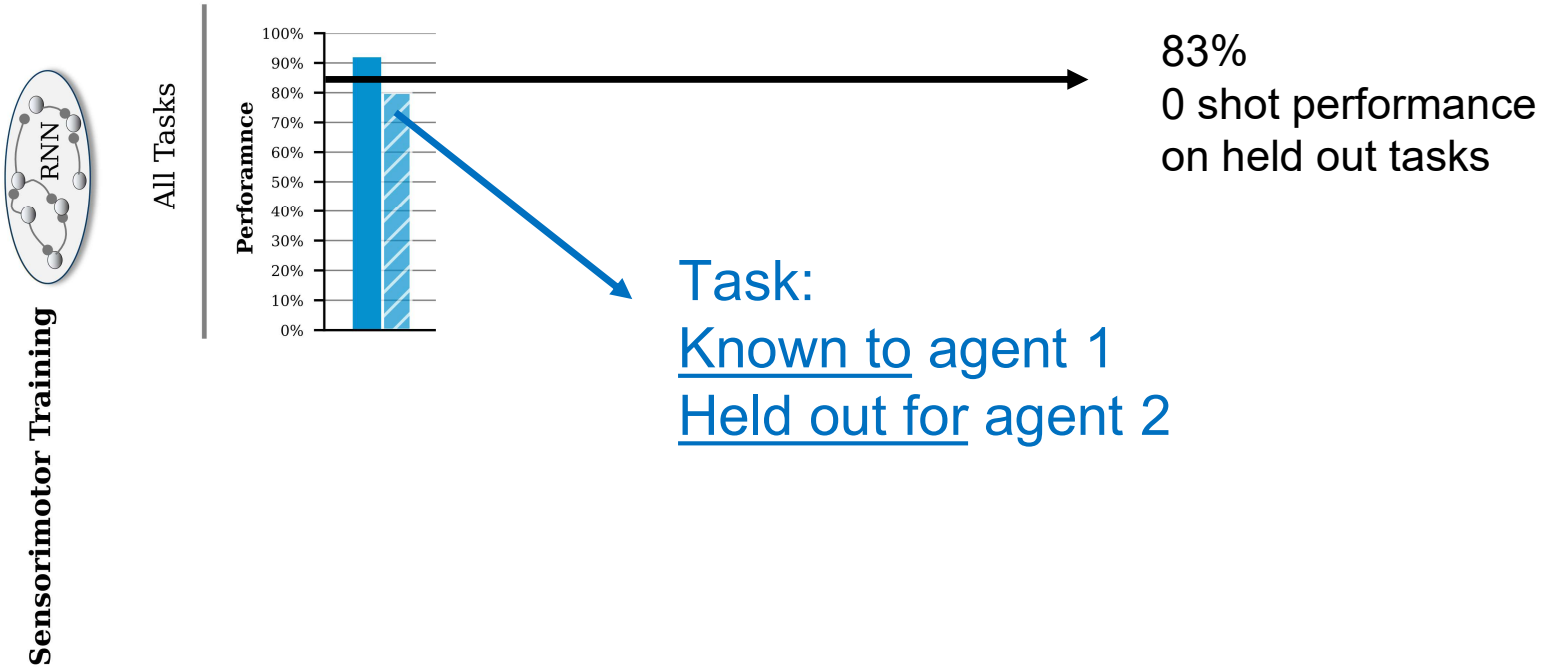
Respond with the identical orientation when prompted



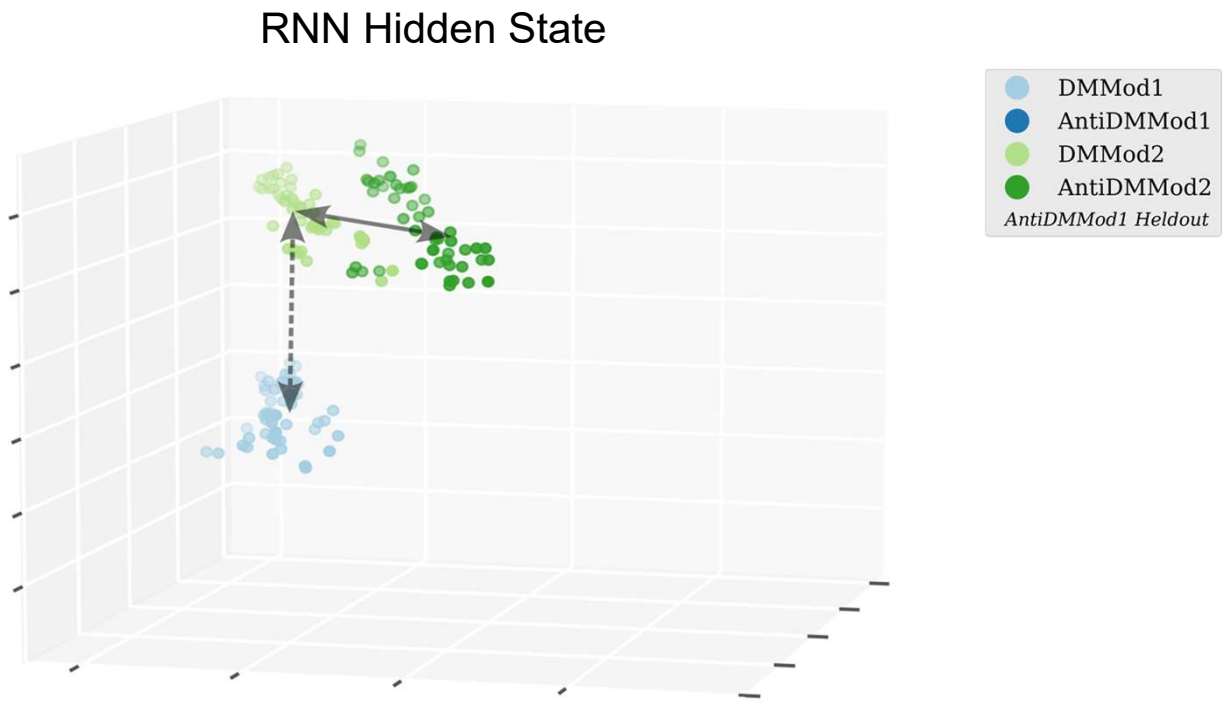
# Performance of listening agent (agent 2)



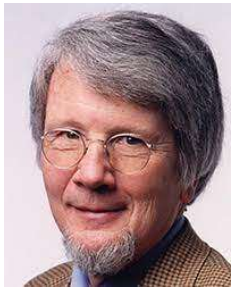
# Performance of listening agent (agent 2)



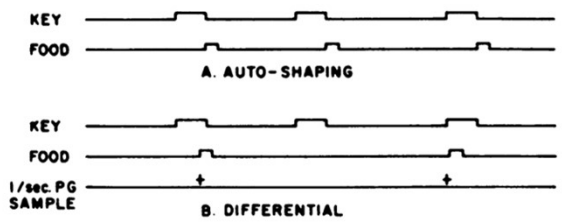
# Part II: leveraging compositional structure

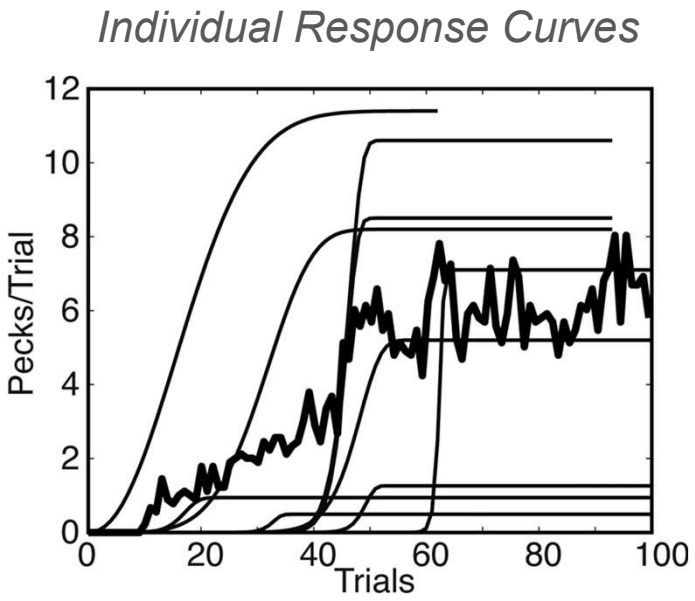


# Insight learning



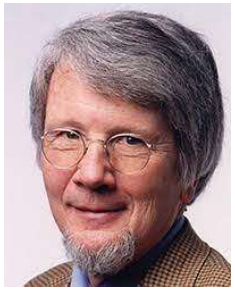
Randy Gallistel



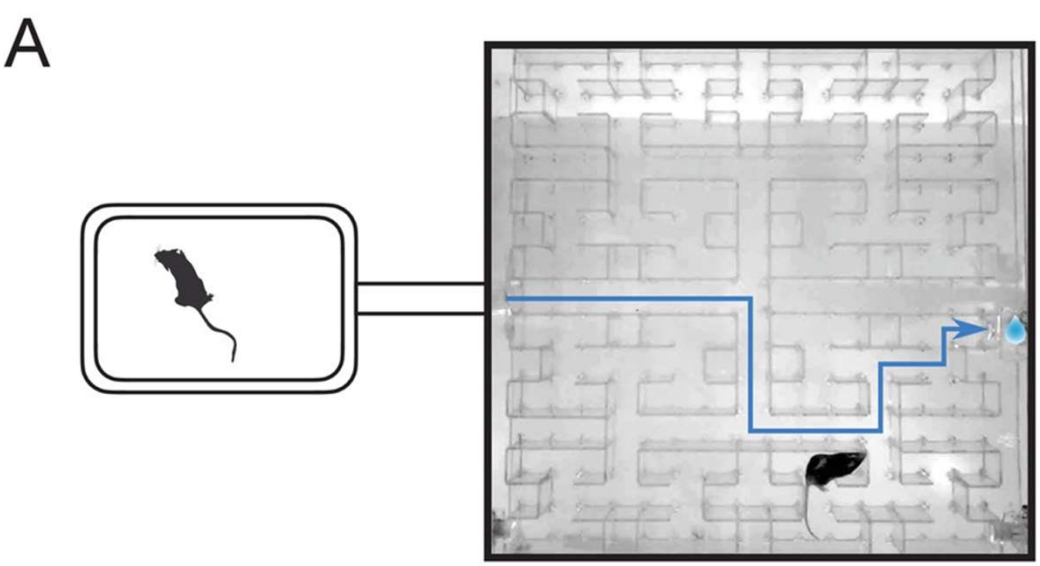



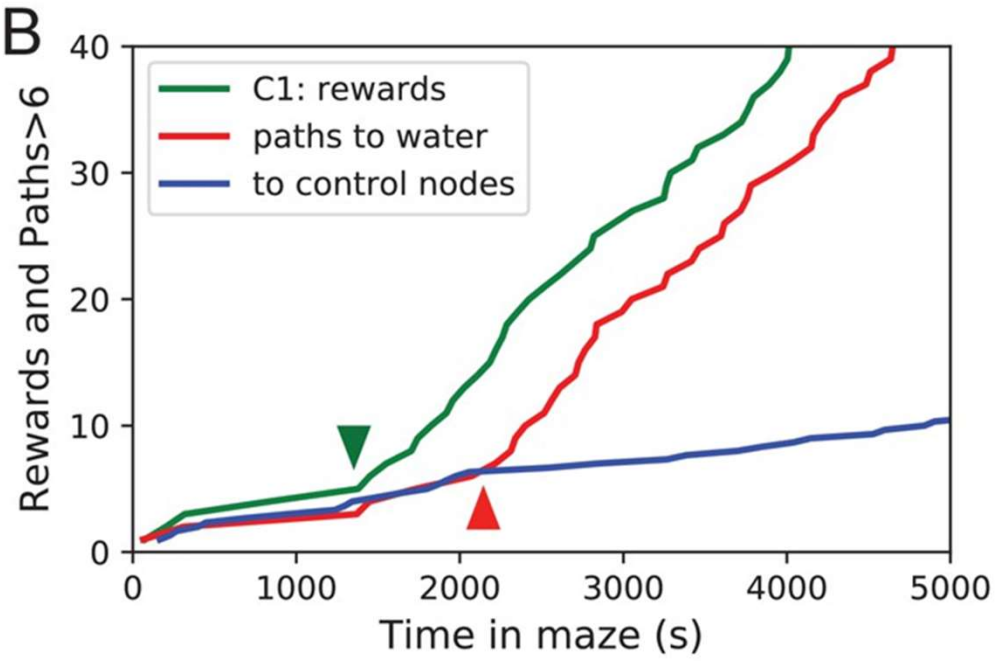

Gallistel et. al., 2004

# Insight learning



Randy Gallistel




Rosenberg et. al. 2021

# Part II: leveraging compositional structure

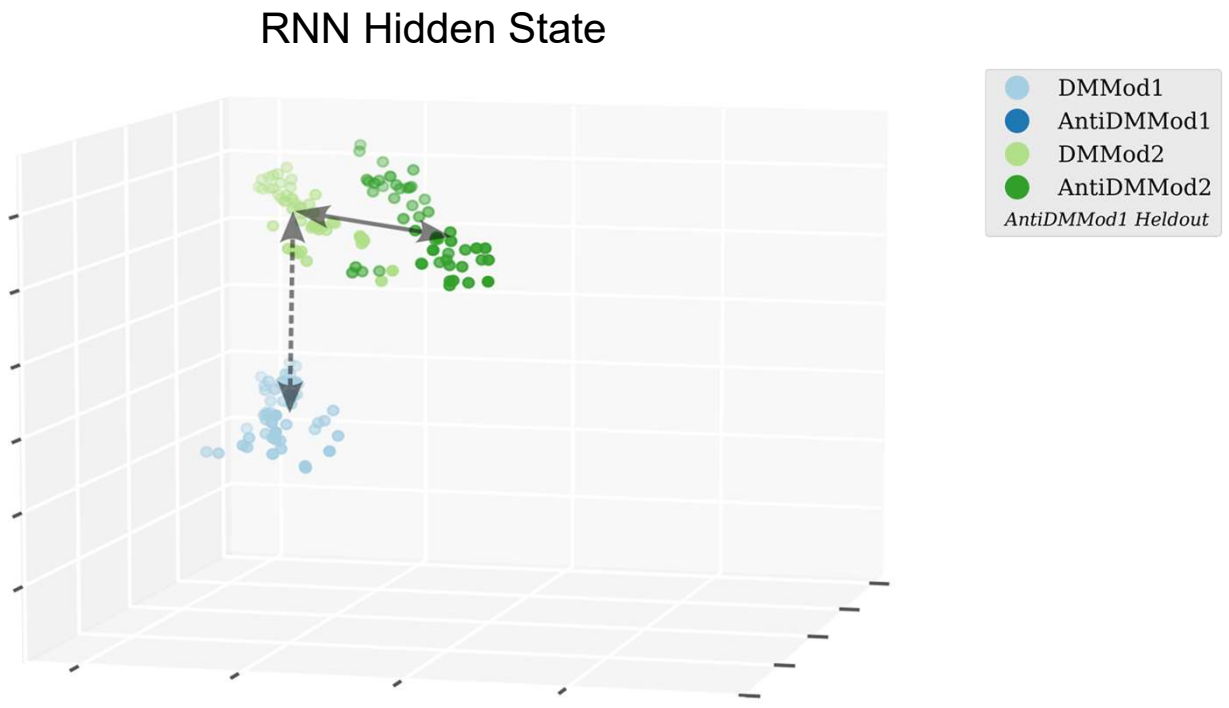
PHILOSOPHICAL TRANSACTIONS
OF THE ROYAL SOCIETY A
MATHEMATICAL, PHYSICAL AND ENGINEERING SCIENCES

## DreamCoder: growing generalizable, interpretable knowledge with wake–sleep Bayesian program learning

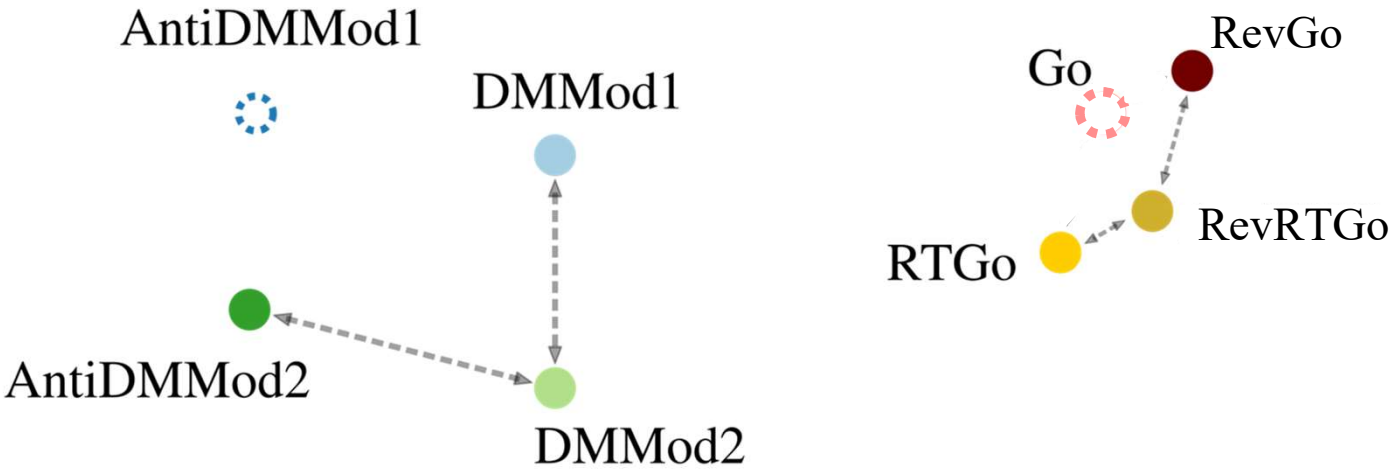
Kevin Ellis ✉, Lionel Wong, Maxwell Nye, Mathias Sablé-Meyer, Luc Cary,
Lore Anaya Pozo, Luke Hewitt, Armando Solar-Lezama and Joshua B. Tenenbaum

Published: 05 June 2023 | https://doi.org/10.1098/rsta.2022.0050

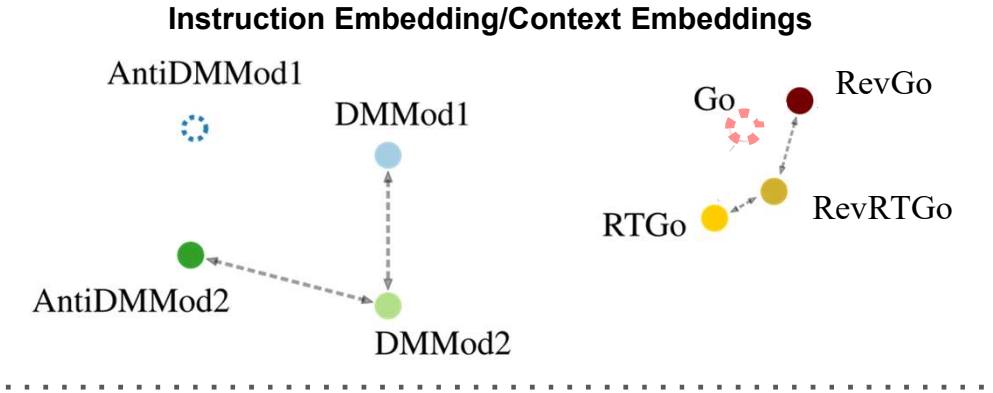
# Can we leverage compositional structure for learning?

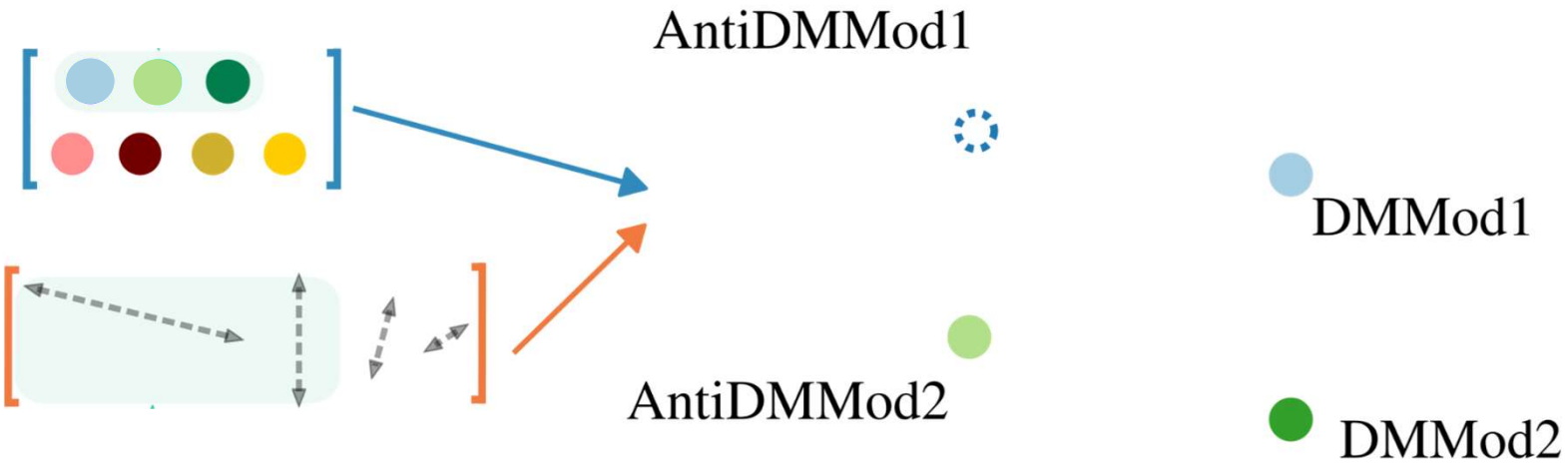


# Leveraging compositional structure

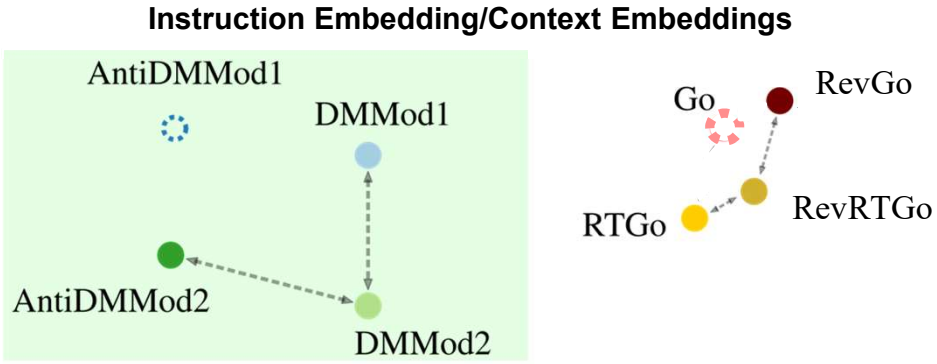


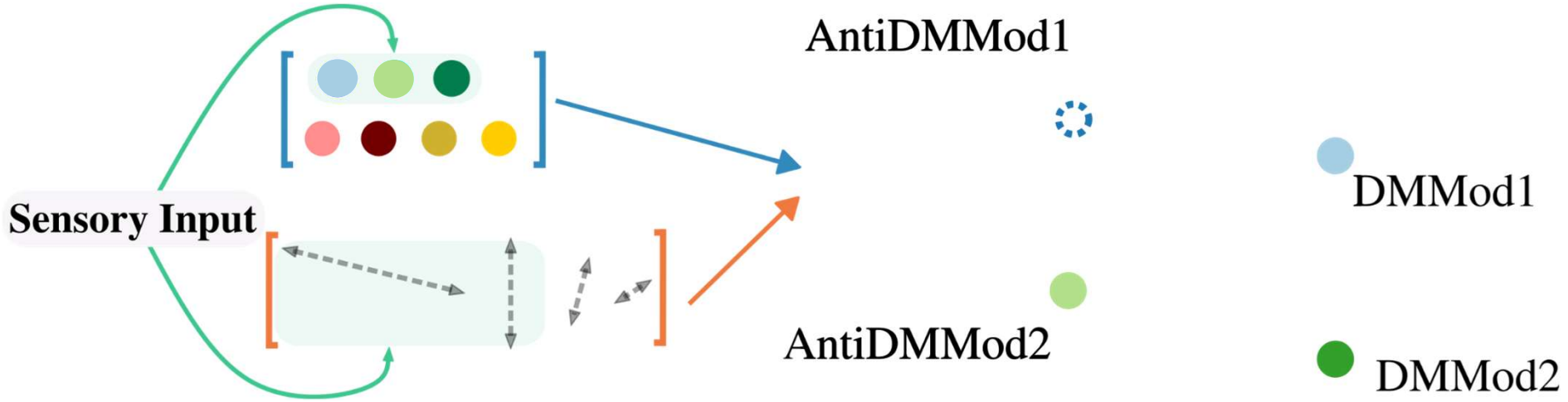
# Leveraging compositional structure



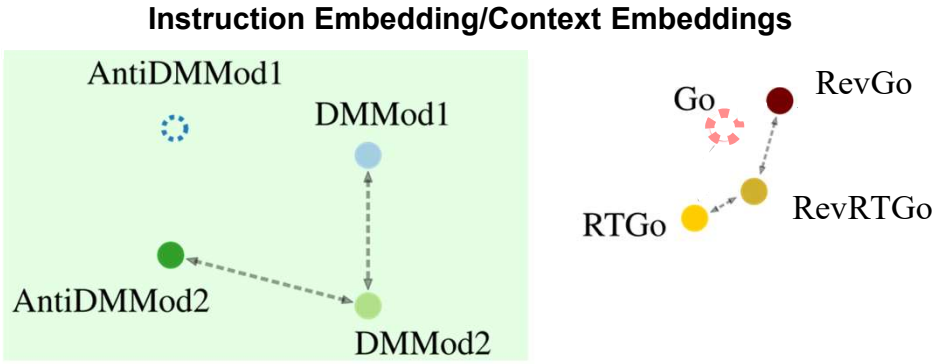



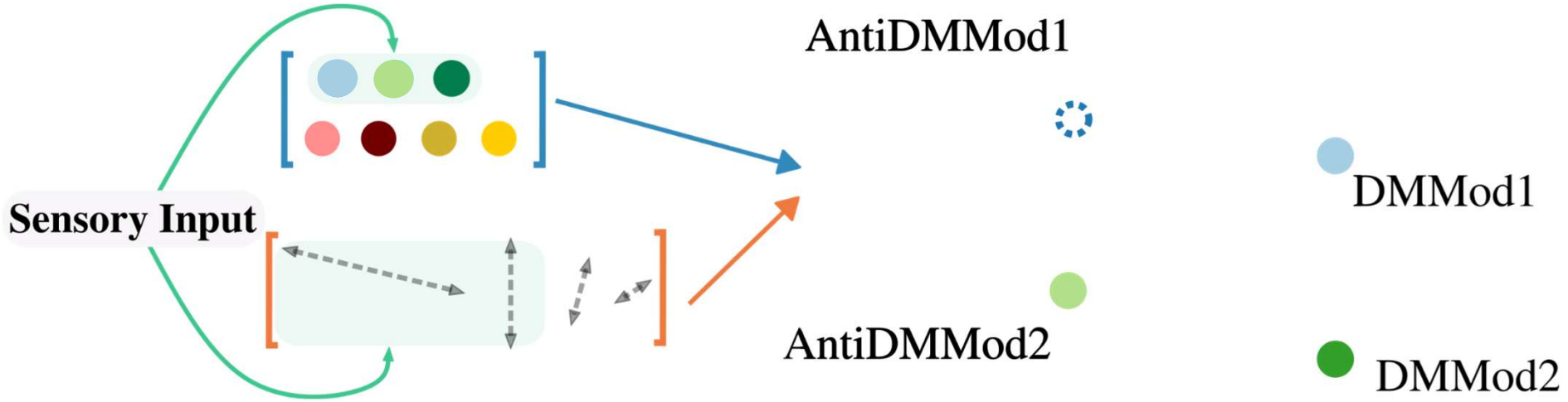
# Leveraging compositional structure



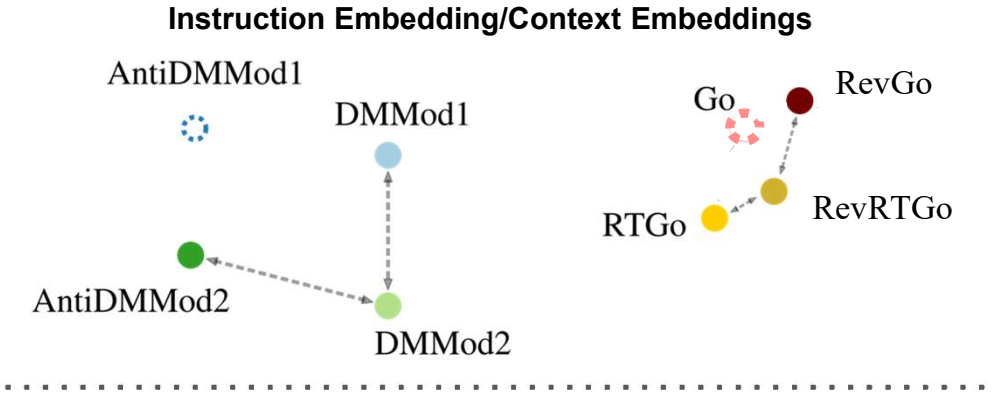



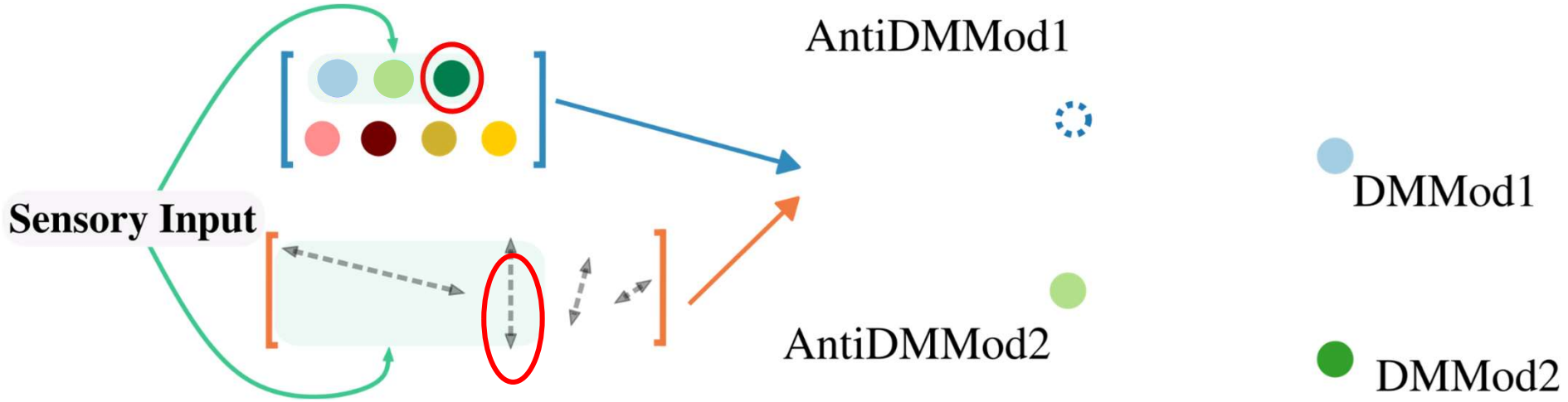
# Leveraging compositional structure



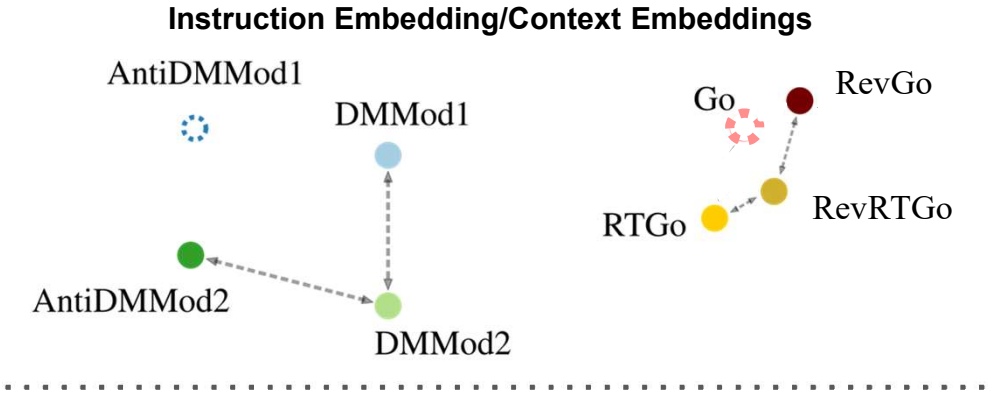



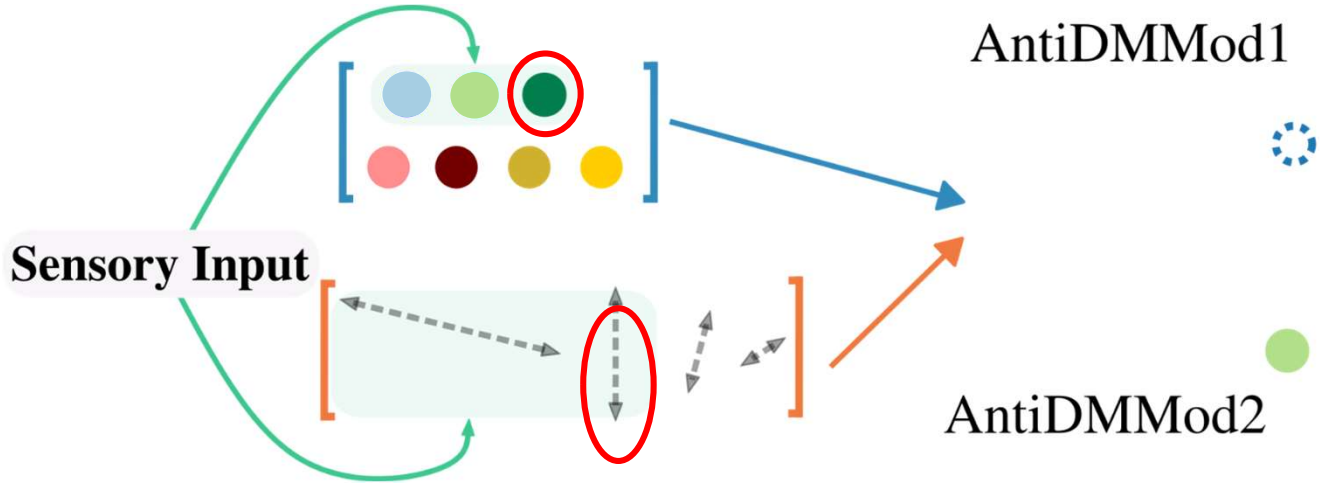
# Leveraging compositional structure

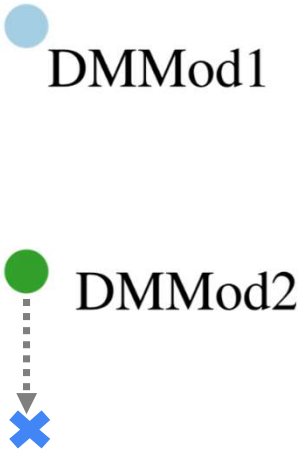


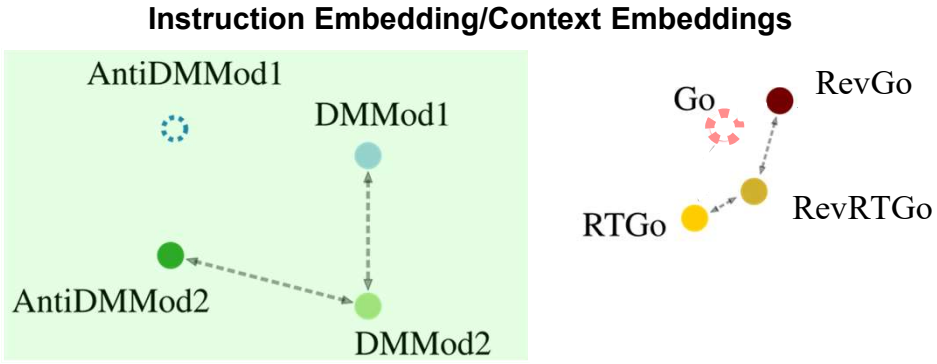
# Leveraging compositional structure

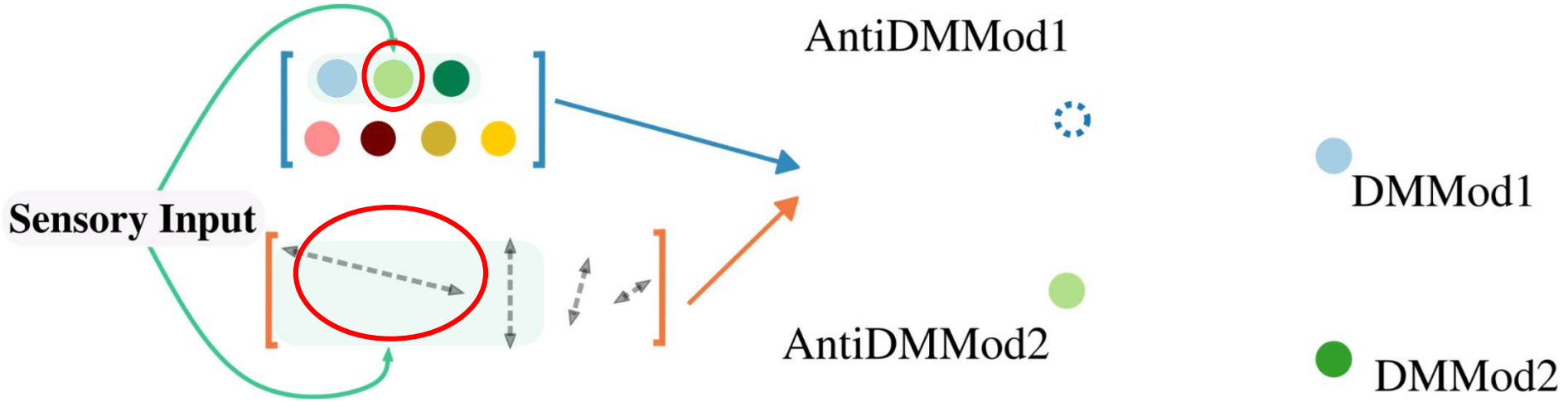


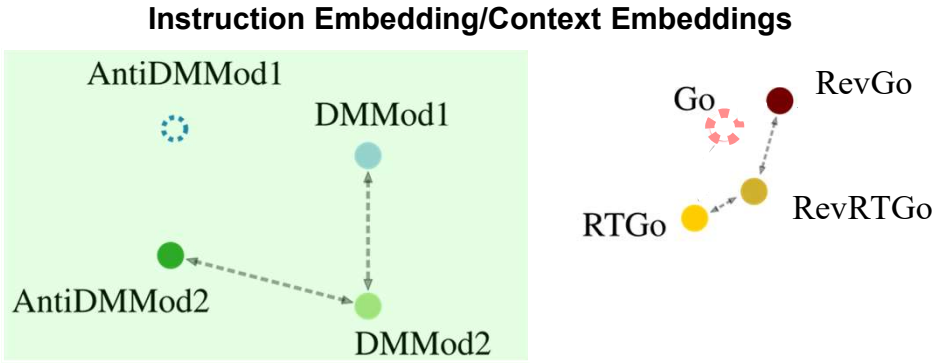



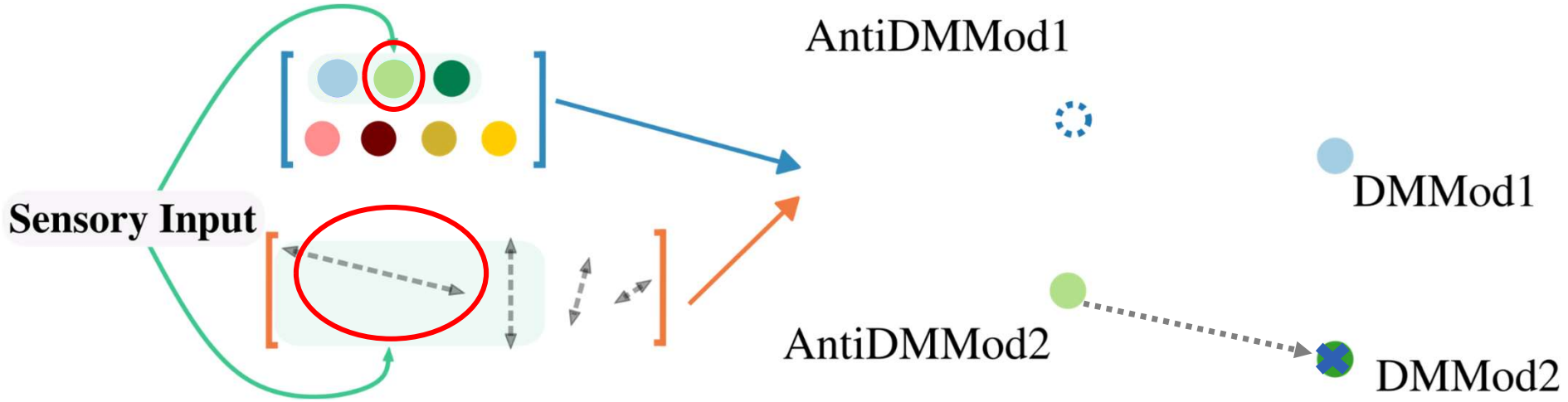
# Leveraging compositional structure



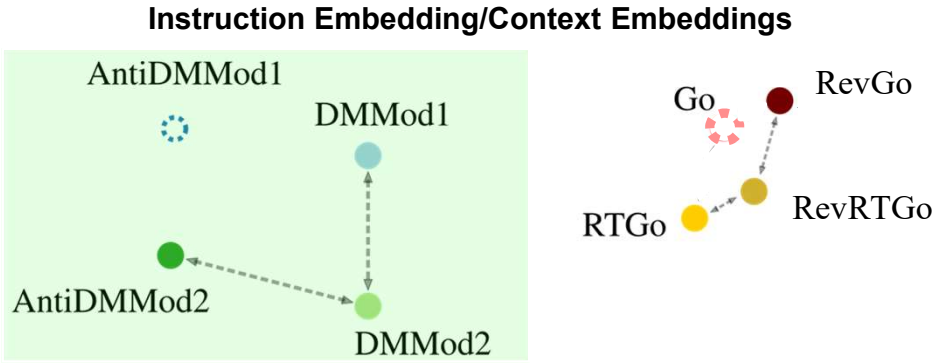



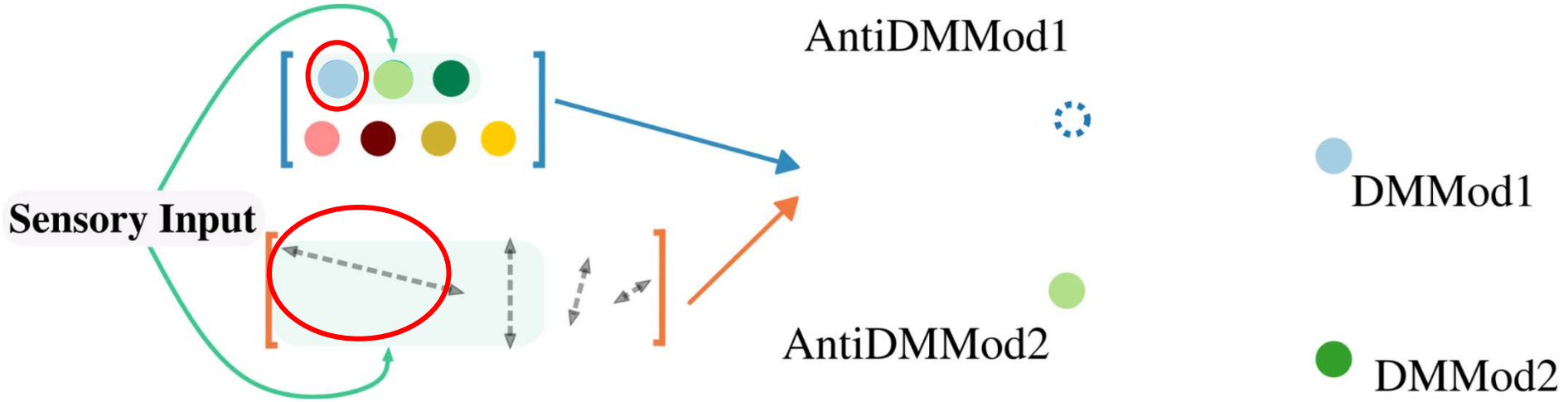
# Leveraging compositional structure



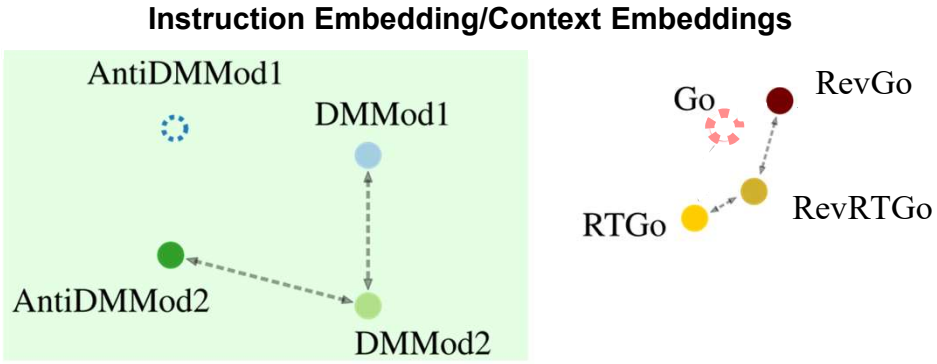



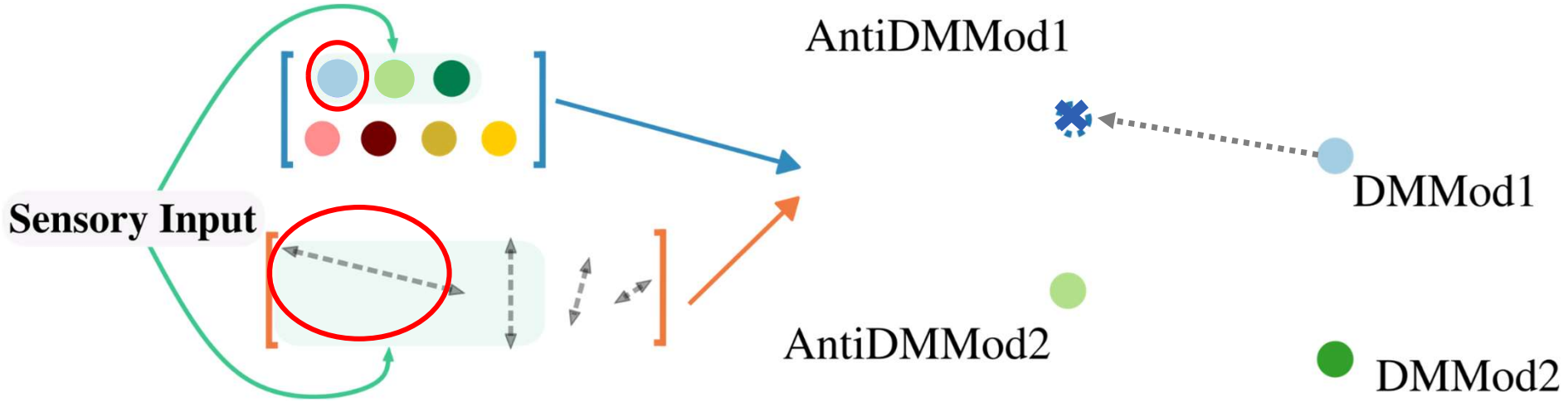
# Leveraging compositional structure

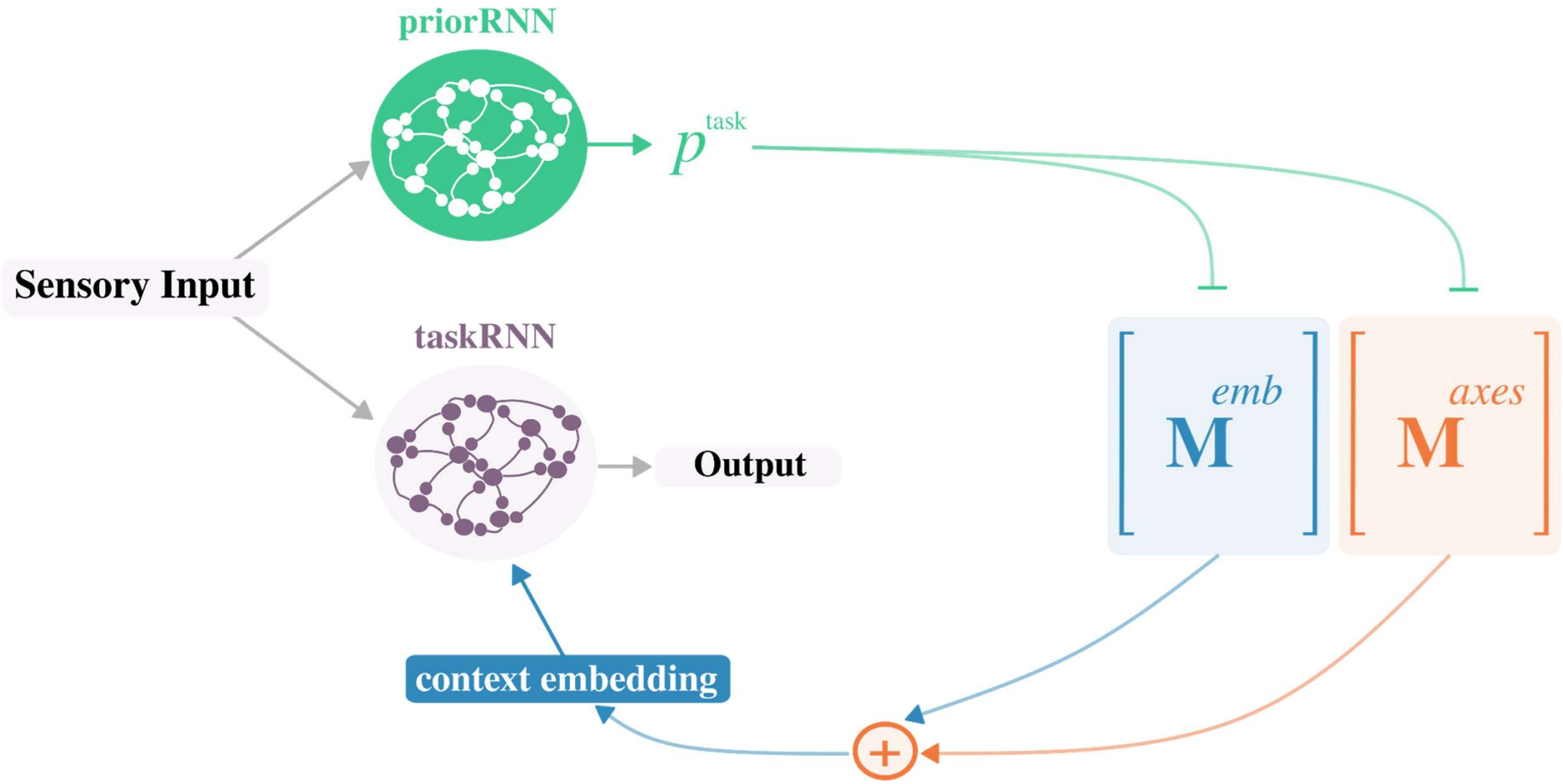


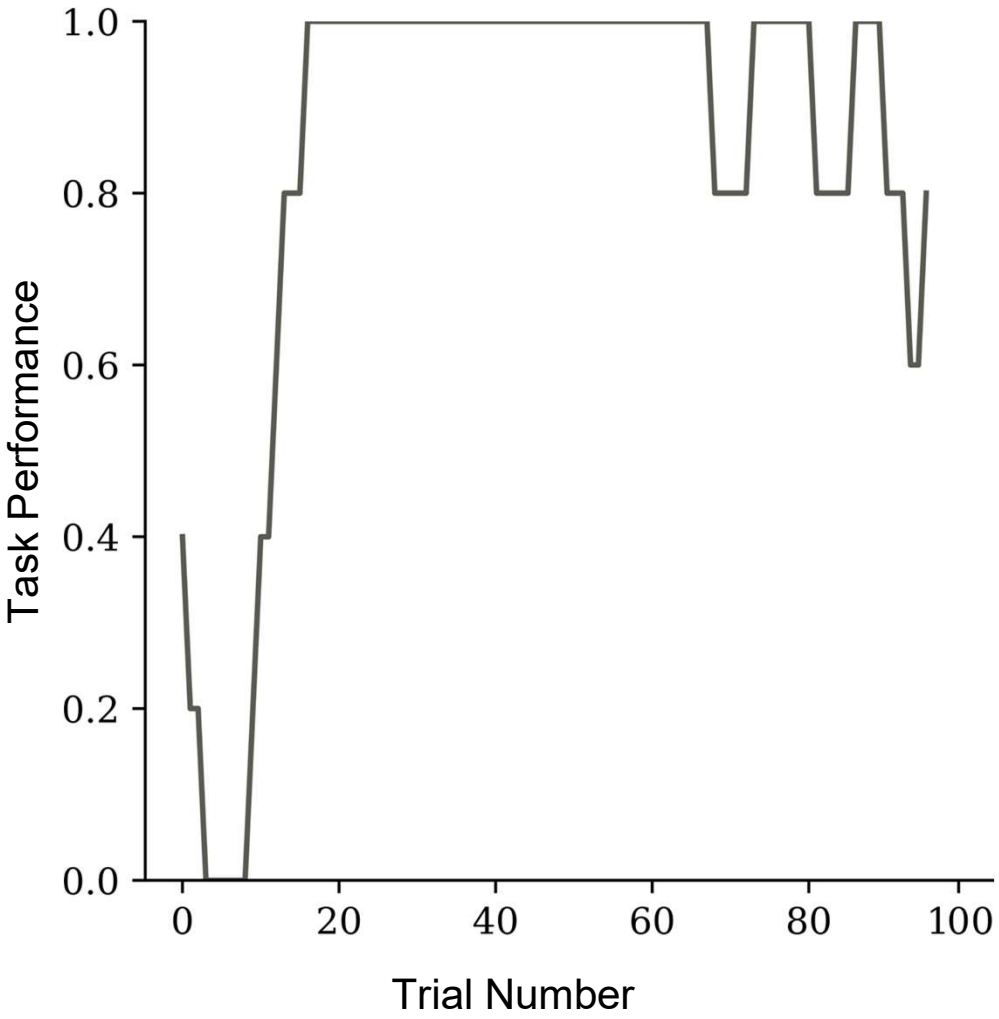
# Leveraging compositional structure

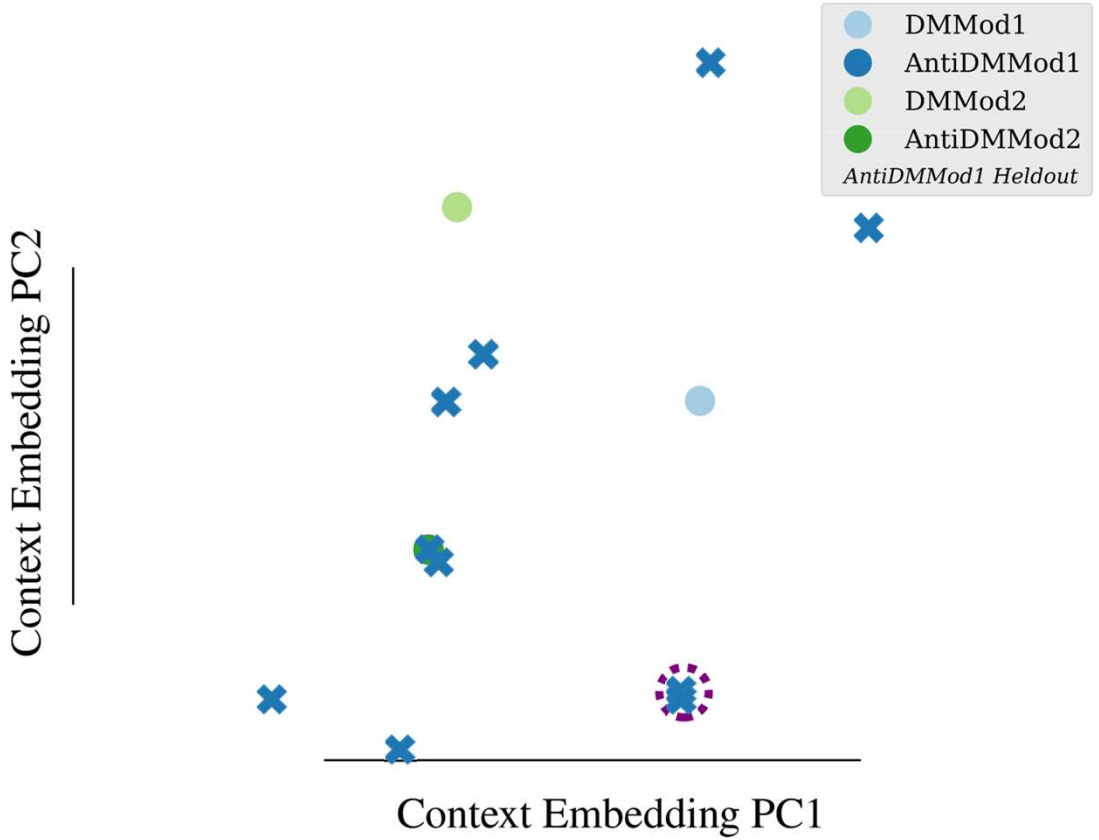


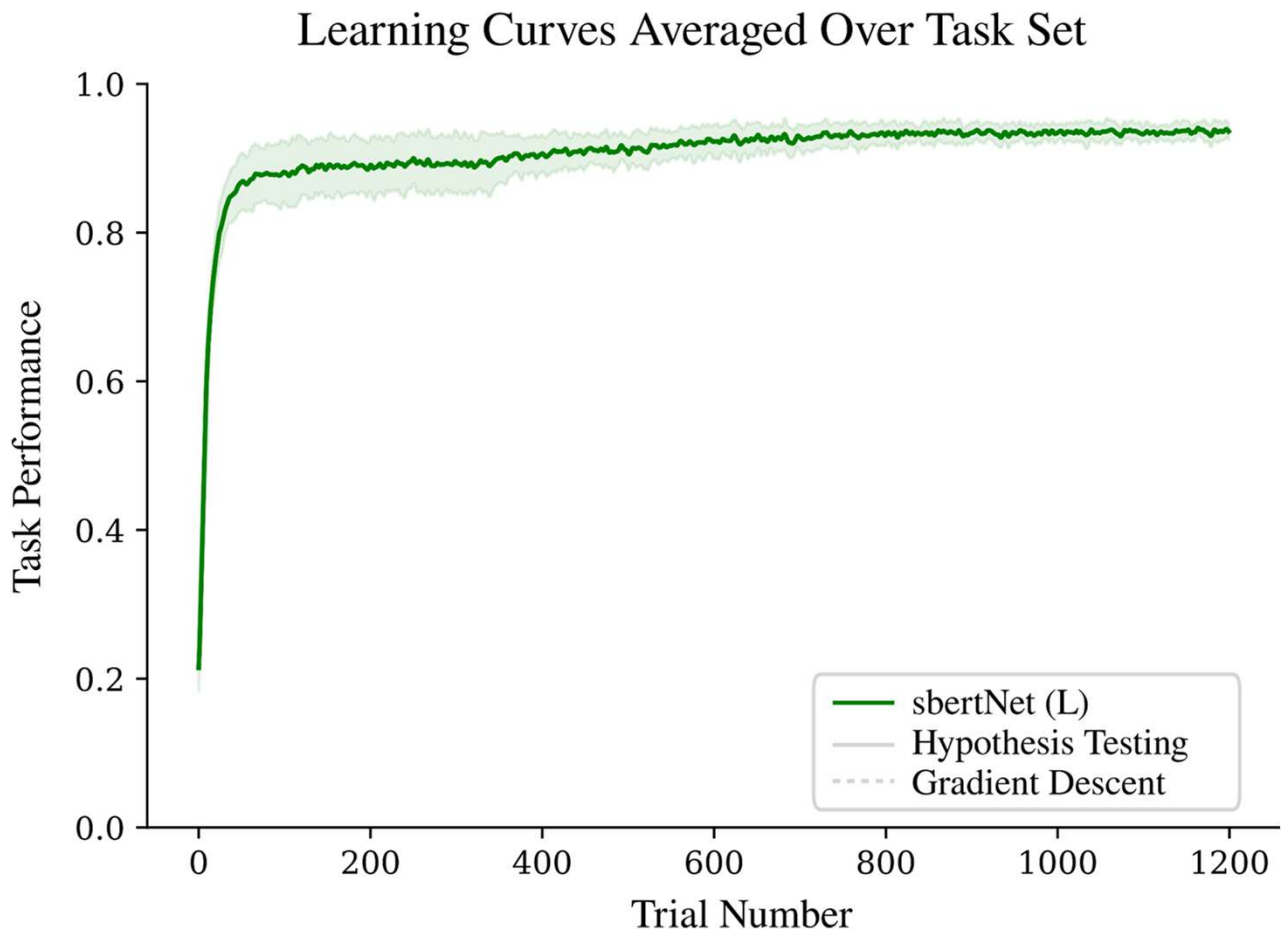



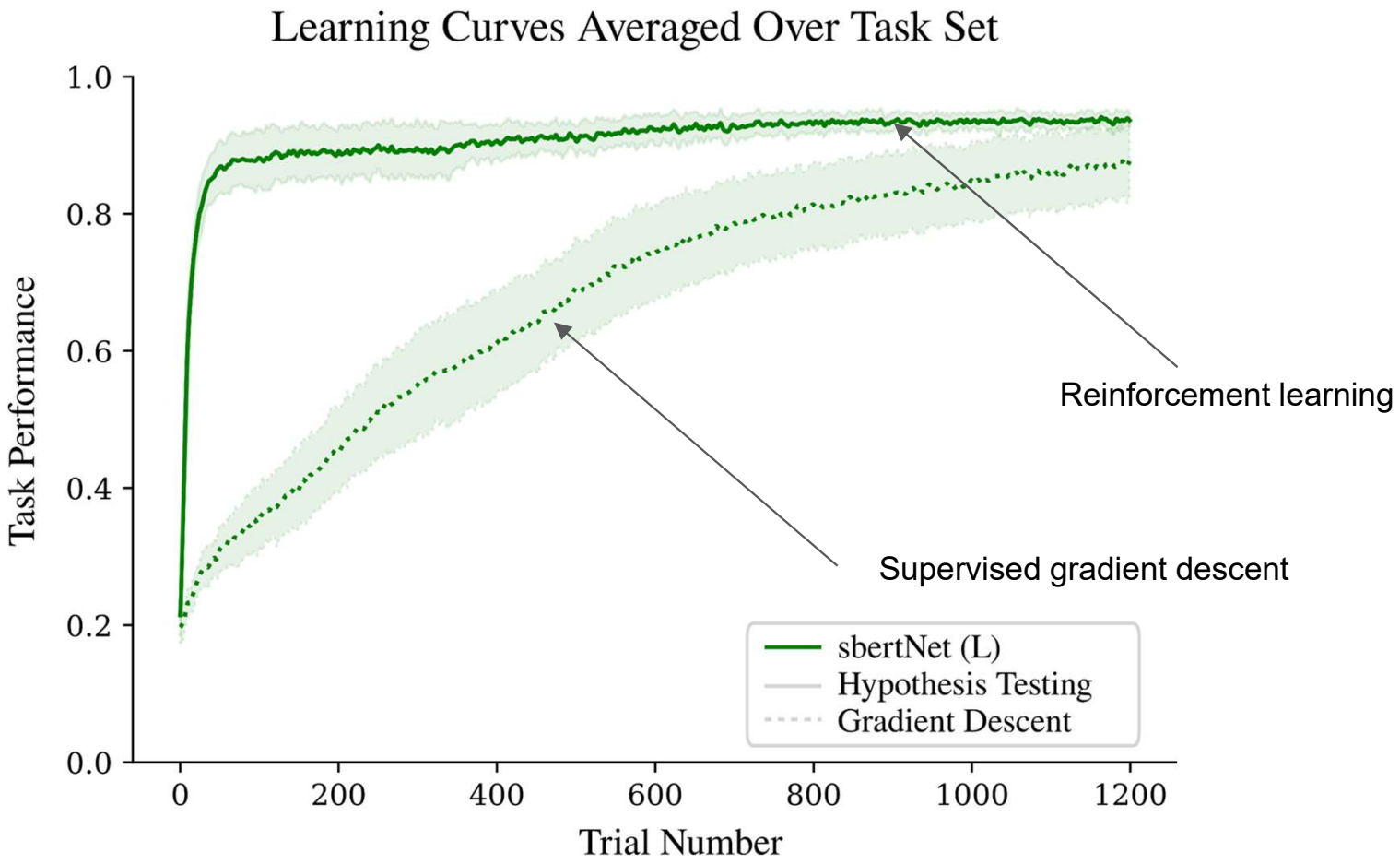

priorRNN
$p^{task}$
Sensory Input
taskRNN
Output
$M^{emb}$
$M^{axes}$
context embedding

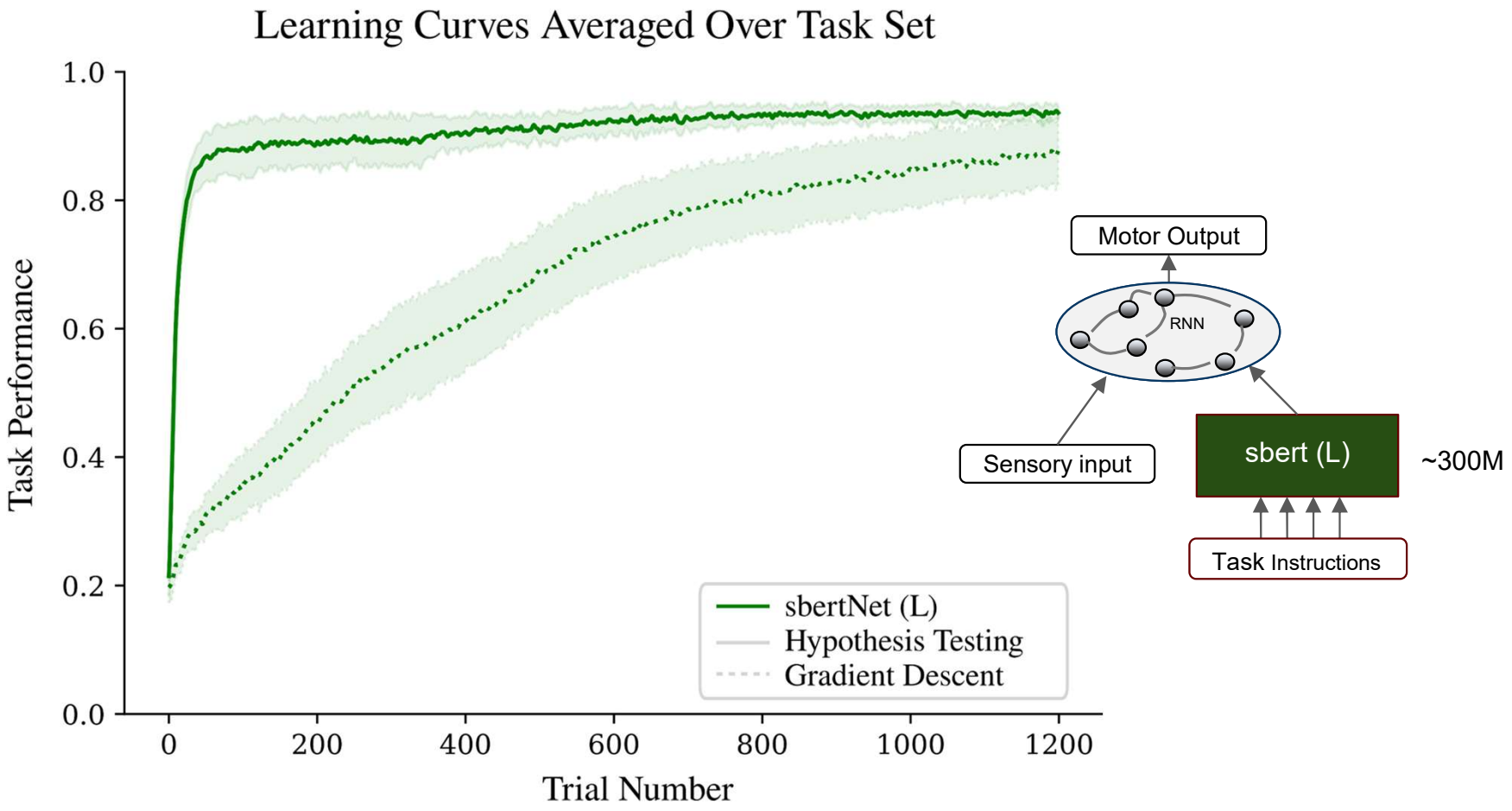
**AntiDMMod1 Heldout Learning**



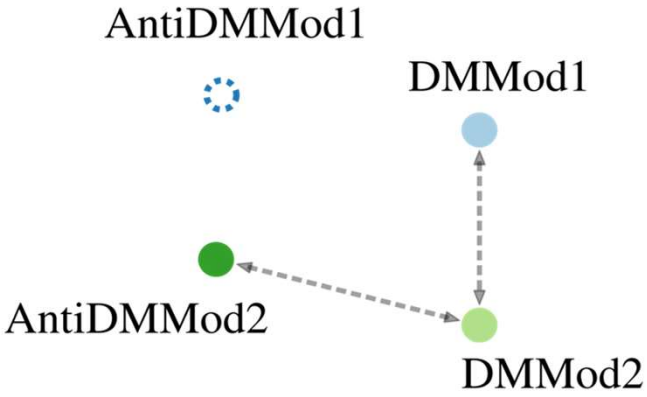



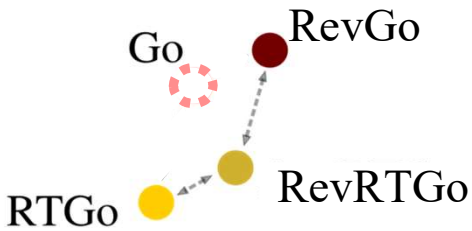

Learning Curves Averaged Over Task Set
Task Performance
Trial Number
1.0
0.8
0.6
0.4
0.2
0.0
0
200
400
600
800
1000
1200
sbertNet (L)
Hypothesis Testing
Gradient Descent

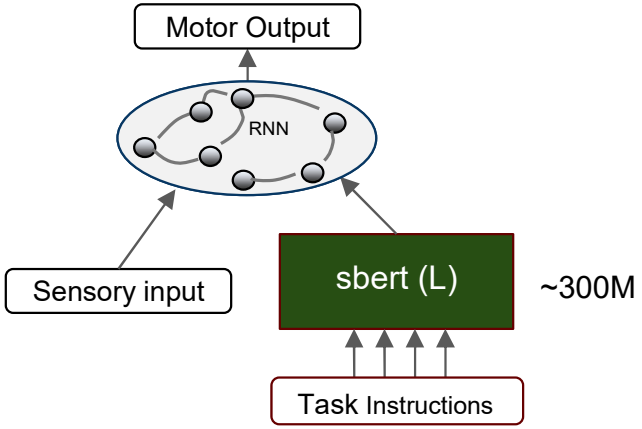

Learning Curves Averaged Over Task Set
Task Performance
1.0
0.8
0.6
0.4
0.2
0.0
0
200
400
600
800
1000
1200
Trial Number
sbertNet (L)
Hypothesis Testing
Gradient Descent
Reinforcement learning
Supervised gradient descent

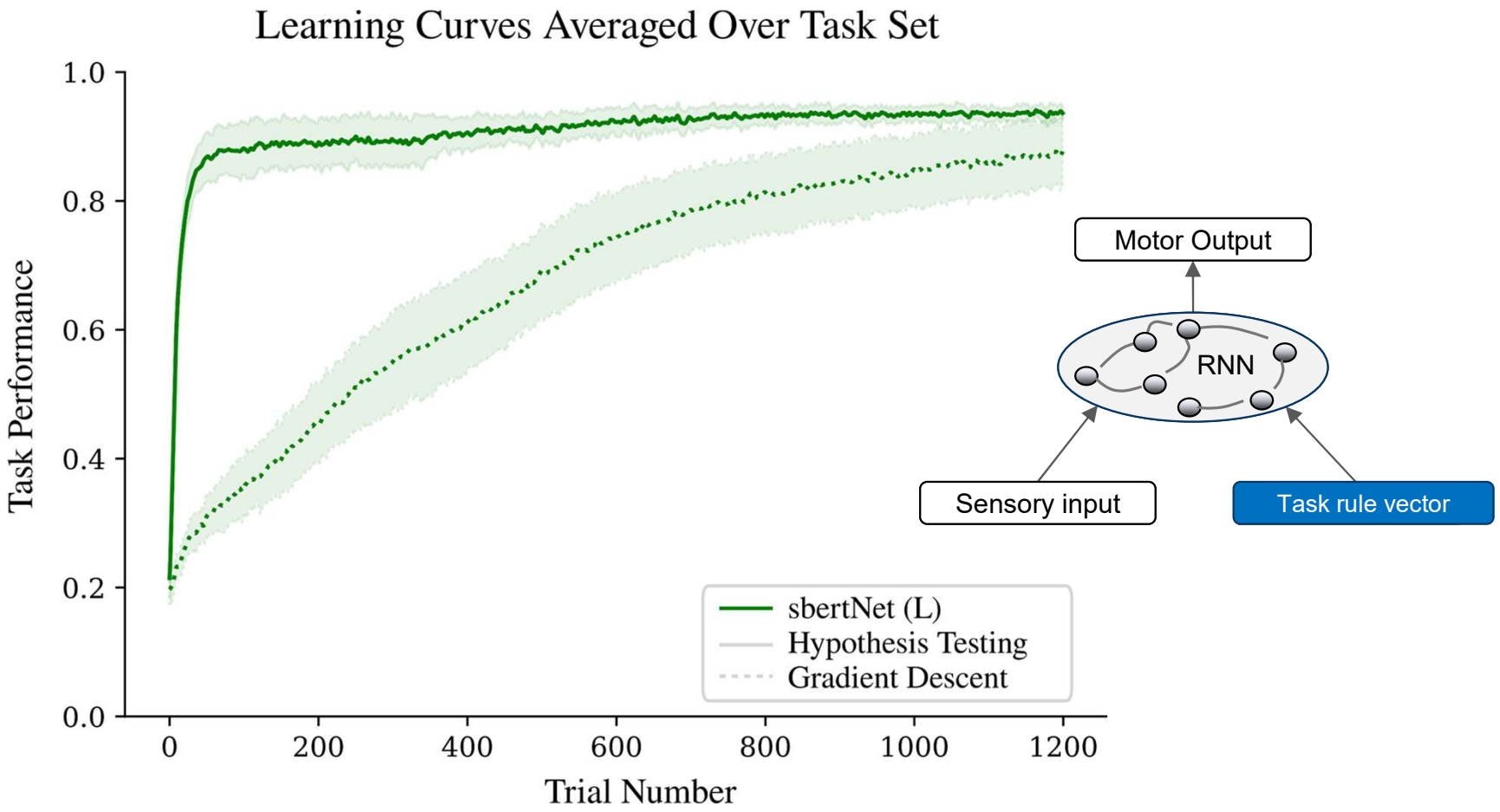

Learning Curves Averaged Over Task Set
Task Performance
Trial Number
1.0
0.8
0.6
0.4
0.2
0.0
0
200
400
600
800
1000
1200
sbertNet (L)
Hypothesis Testing
Gradient Descent
Motor Output
RNN
Sensory input
sbert (L)
~300M
Task Instructions

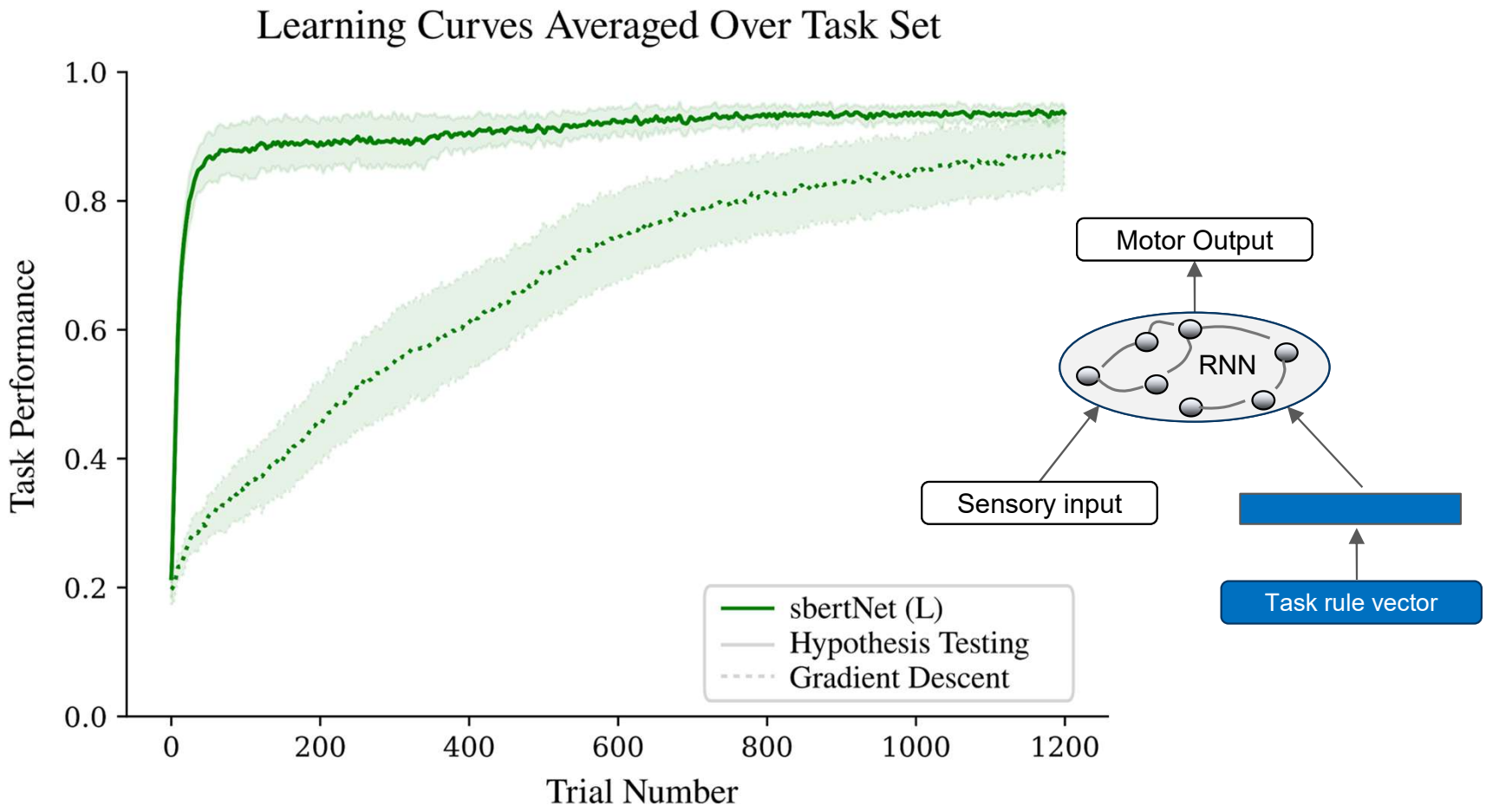
# Leveraging compositional structure

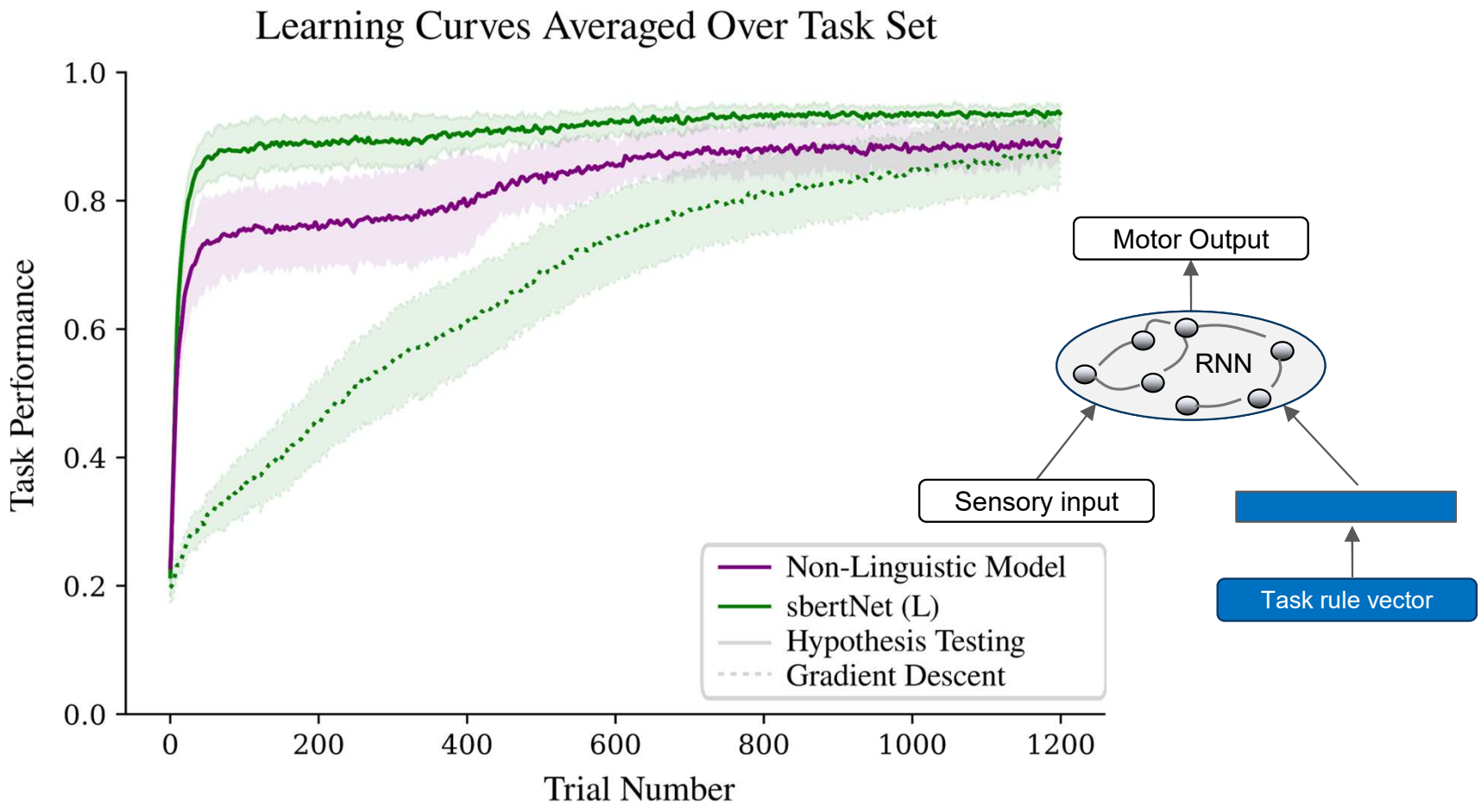


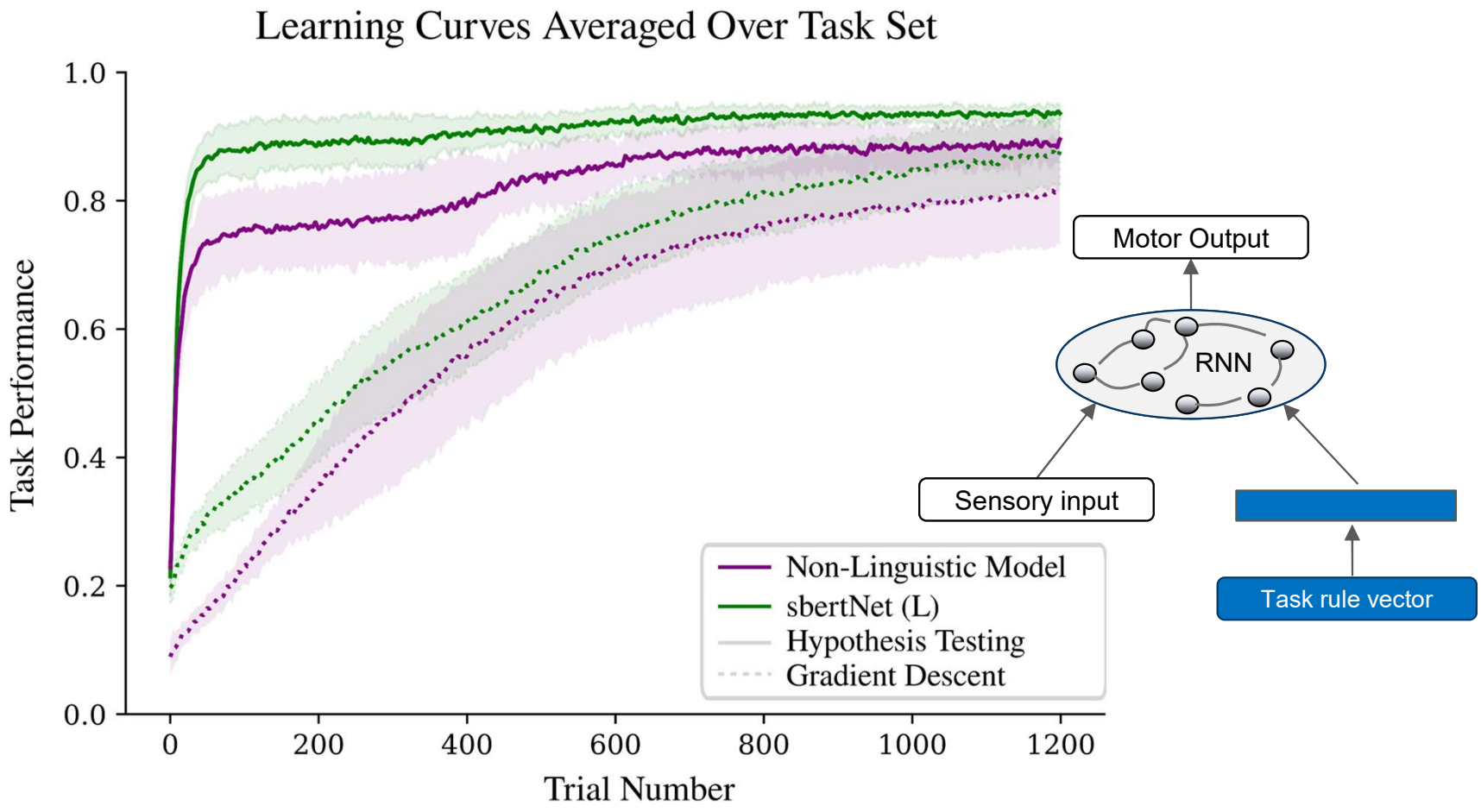



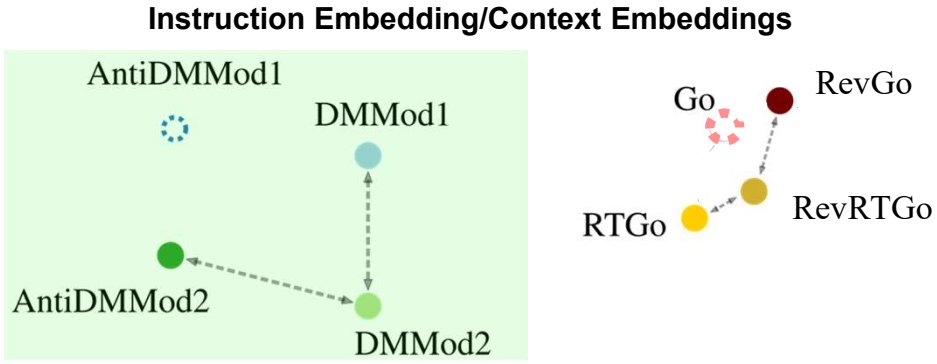

Learning Curves Averaged Over Task Set
Task Performance
Trial Number
1.0
0.8
0.6
0.4
0.2
0.0
0
200
400
600
800
1000
1200
sbertNet (L)
Hypothesis Testing
Gradient Descent
Motor Output
RNN
Sensory input
Task rule vector

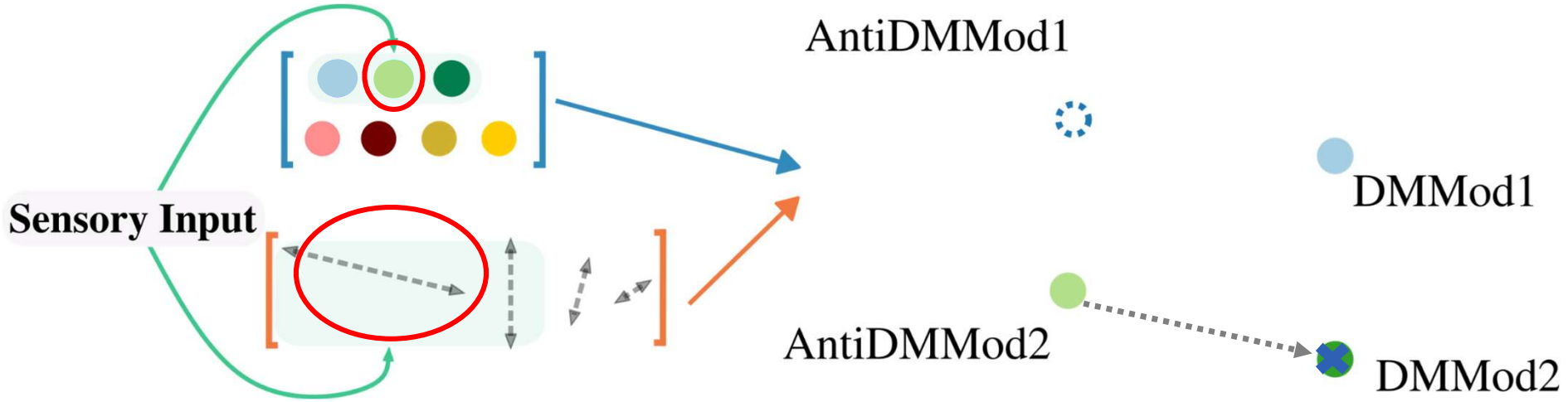

Learning Curves Averaged Over Task Set
Task Performance
Trial Number
1.0
0.8
0.6
0.4
0.2
0.0
0
200
400
600
800
1000
1200
sbertNet (L)
Hypothesis Testing
Gradient Descent
Motor Output
RNN
Sensory input
Task rule vector


Learning Curves Averaged Over Task Set
Task Performance
1.0
0.8
0.6
0.4
0.2
0.0
0
200
400
600
800
1000
1200
Trial Number
Non-Linguistic Model
sbertNet (L)
Hypothesis Testing
Gradient Descent
Motor Output
RNN
Sensory input
Task rule vector


Learning Curves Averaged Over Task Set
Task Performance
Trial Number
1.0
0.8
0.6
0.4
0.2
0.0
0
200
400
600
800
1000
1200
Non-Linguistic Model
sbertNet (L)
Hypothesis Testing
Gradient Descent
Motor Output
RNN
Sensory input
Task rule vector

# Leveraging compositional structure

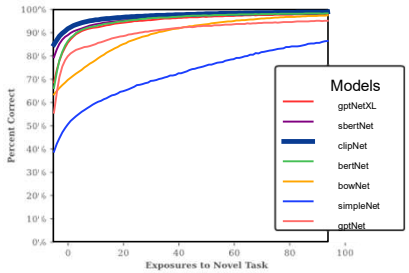


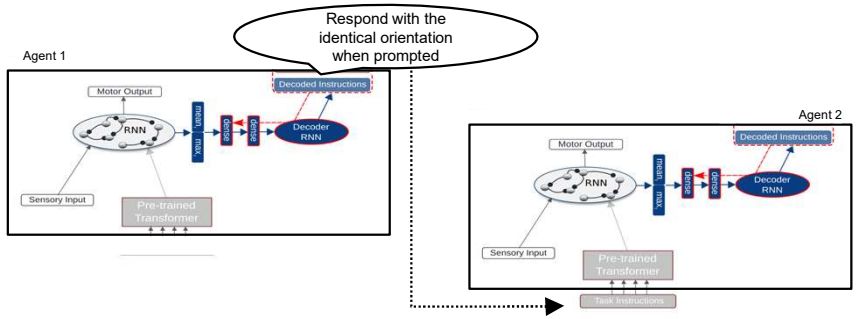
# Summary

Reidar Riveland

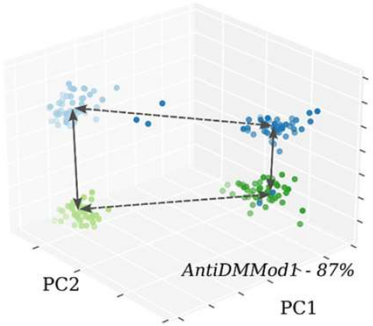
→ Ultra Fast Learning with S-BERT



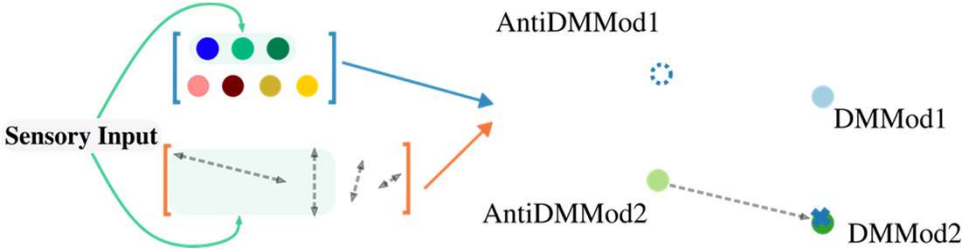

→Talking networks




→ Abstract Task Representations




→ Hypothesis testing based learning

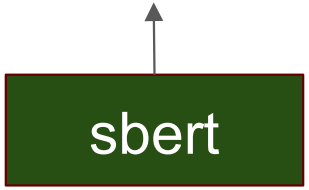


# Sentence-BERT (sbert)

Fine-tuned on **Stanford Natural Language Inference** Dataset:

premise: the man inspects his shirt for coffee stains
hypothesis: the man is eating pancakes

*Vector space of sentence embeddings match human judgements of sentence similarity*




Reimers & Gurevych, 2019

# Structured Representations Across Model Hierarchy