TarMAC: Targeted Multi-Agent Communication


Abhishek Das


Théophile Gervet


Joshua Romoff


Dhruv Batra


Devi Parikh


Mike Rabbat


Joelle Pineau

Georgia Tech

UNIVERSITÉ McGill

facebook Artificial Intelligence Research

Multi-Agent Communication

Multi-Agent Communication

- Learning effective communication is a key ability for collaboration.

Multi-Agent Communication

- Learning effective communication is a key ability for collaboration.
- Wide-ranging applications
  - Multi-player games

AlphaStar, DeepMind.

Multi-Agent Communication

- Learning effective communication is a key ability for collaboration.

- Wide-ranging applications
  - Multi-player games
  - Self-driving car networks

AlphaStar, DeepMind.

US Department of Transportation.

Multi-Agent Communication

- Learning effective communication is a key ability for collaboration.

- Wide-ranging applications
  - Multi-player games
  - Self-driving car networks
  - Search-and-rescue robots

AlphaStar, DeepMind.

US Department of Transportation.

DARPA.

Multi-Agent Communication

- Learning effective communication is a key ability for collaboration.

- Wide-ranging applications
  - Multi-player games
  - Self-driving car networks
  - Search-and-rescue robots

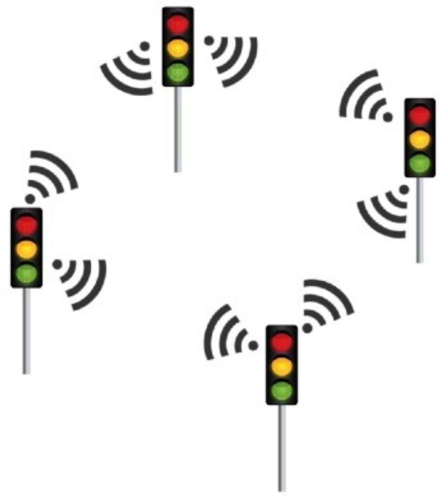
  ...

AlphaStar, DeepMind.

US Department of Transportation.

DARPA.

AFP / Getty Images.



Budapest University

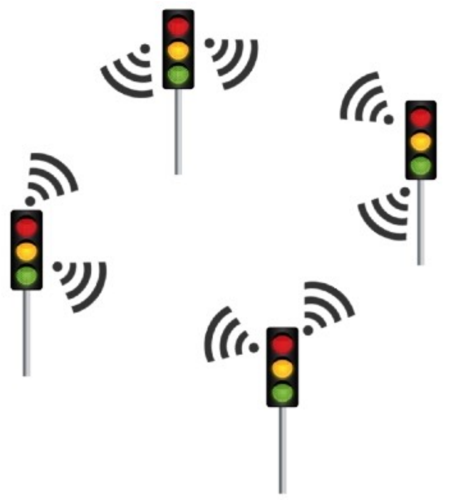
Multi-Agent Communication

AlphaStar, DeepMind.

US Department of Transportation.

DARPA.

AFP / Getty Images.



Budapest University

Multi-Agent Communication

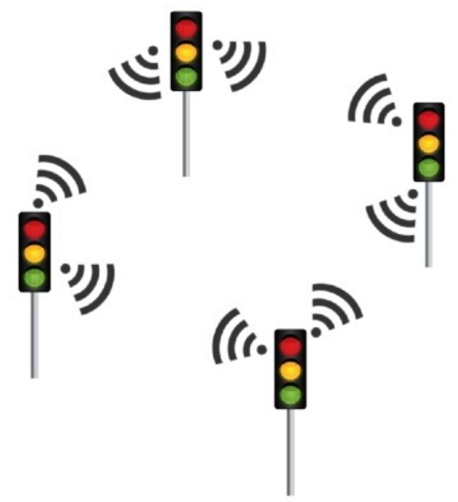
- Prior work on learning multi-agent communication:
  - Learning Multi-agent Communication with Backpropagation. Sukhbaatar et al., 2016
  - Learning to Communicate with Deep Multi-Agent Reinforcement Learning. Foerster et al., 2016.
  - Learning When to Communicate at Scale in Multi-Agent Cooperative and Competitive Tasks. Singh et al., 2019.

AlphaStar, DeepMind.

US Department of Transportation.

DARPA.

AFP / Getty Images.



Budapest University

Multi-Agent Communication

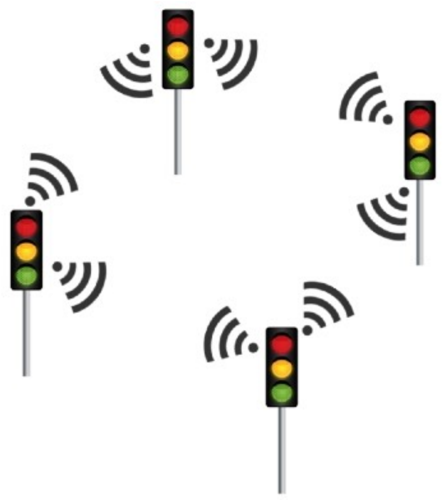
- Prior work on learning multi-agent communication:
  - Learning Multi-agent Communication with Backpropagation. Sukhbaatar et al., 2016
  - Learning to Communicate with Deep Multi-Agent Reinforcement Learning. Foerster et al., 2016.
  - Learning When to Communicate at Scale in Multi-Agent Cooperative and Competitive Tasks. Singh et al., 2019.
- **Agents broadcasting messages to other agents**.

AlphaStar, DeepMind.

US Department of Transportation.

DARPA.

AFP / Getty Images.



Budapest University

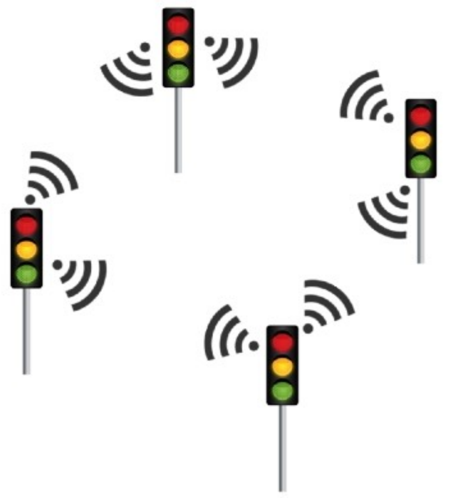
Multi-Agent Communication

AlphaStar, DeepMind.

US Department of Transportation.

DARPA.

AFP / Getty Images.



Budapest University

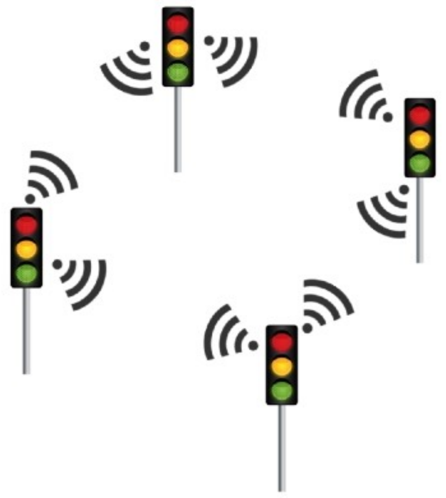
Targeted Multi-Agent Communication

AlphaStar, DeepMind.

US Department of Transportation.

DARPA.

AFP / Getty Images.



Budapest University

Targeted Multi-Agent Communication

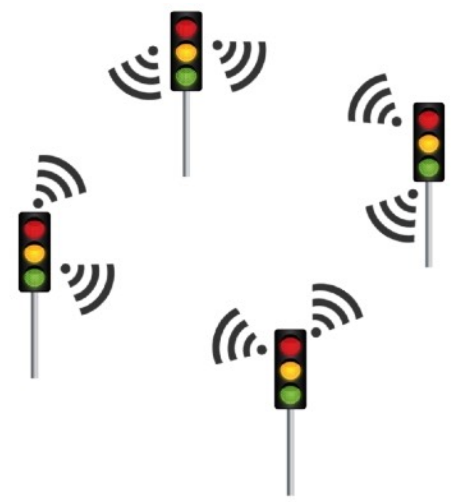
- But **for complex collaboration strategies to emerge** among agents with different roles and goals, **targeted communication is important**.

AlphaStar, DeepMind.

US Department of Transportation.

DARPA.

AFP / Getty Images.



Budapest University

Targeted Multi-Agent Communication

- But **for complex collaboration strategies to emerge** among agents with different roles and goals, **targeted communication is important**.

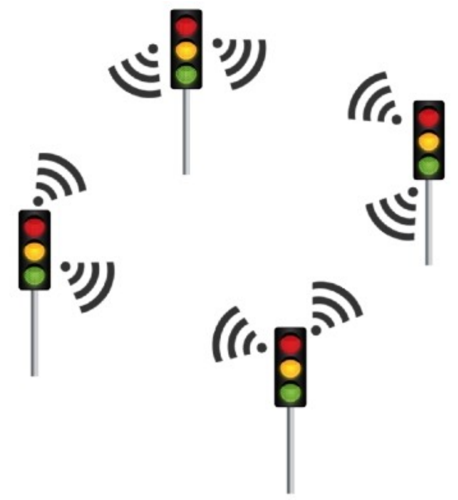
- i.e. **being able to direct certain messages to specific recipients**.

AlphaStar, DeepMind.

US Department of Transportation.

DARPA.

AFP / Getty Images.



Budapest University

Targeted Multi-Agent Communication

- But **for complex collaboration strategies to emerge** among agents with different roles and goals, **targeted communication is important**.

- i.e. **being able to direct certain messages to specific recipients**.

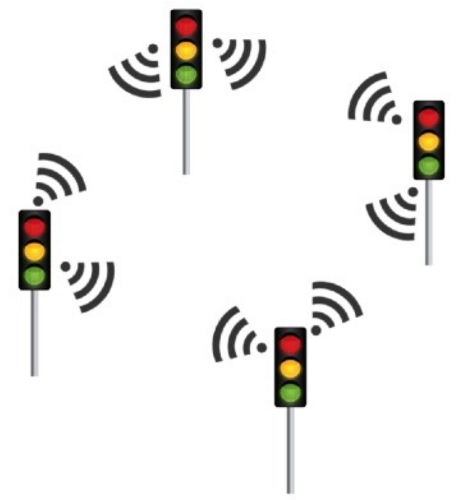
- We introduce **TarMAC**, a multi-agent reinforcement learning architecture enabling **targeted, multi-round communication learnt through backpropagation**.

AlphaStar, DeepMind.

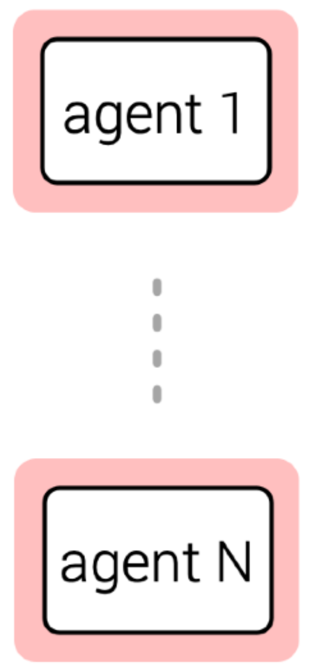
US Department of Transportation.

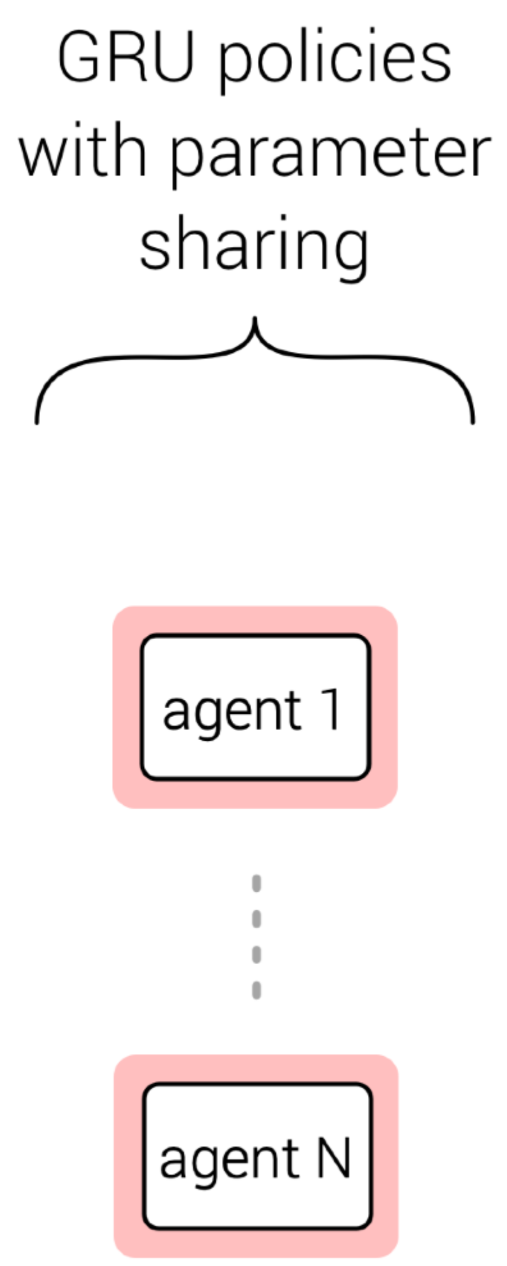
DARPA.

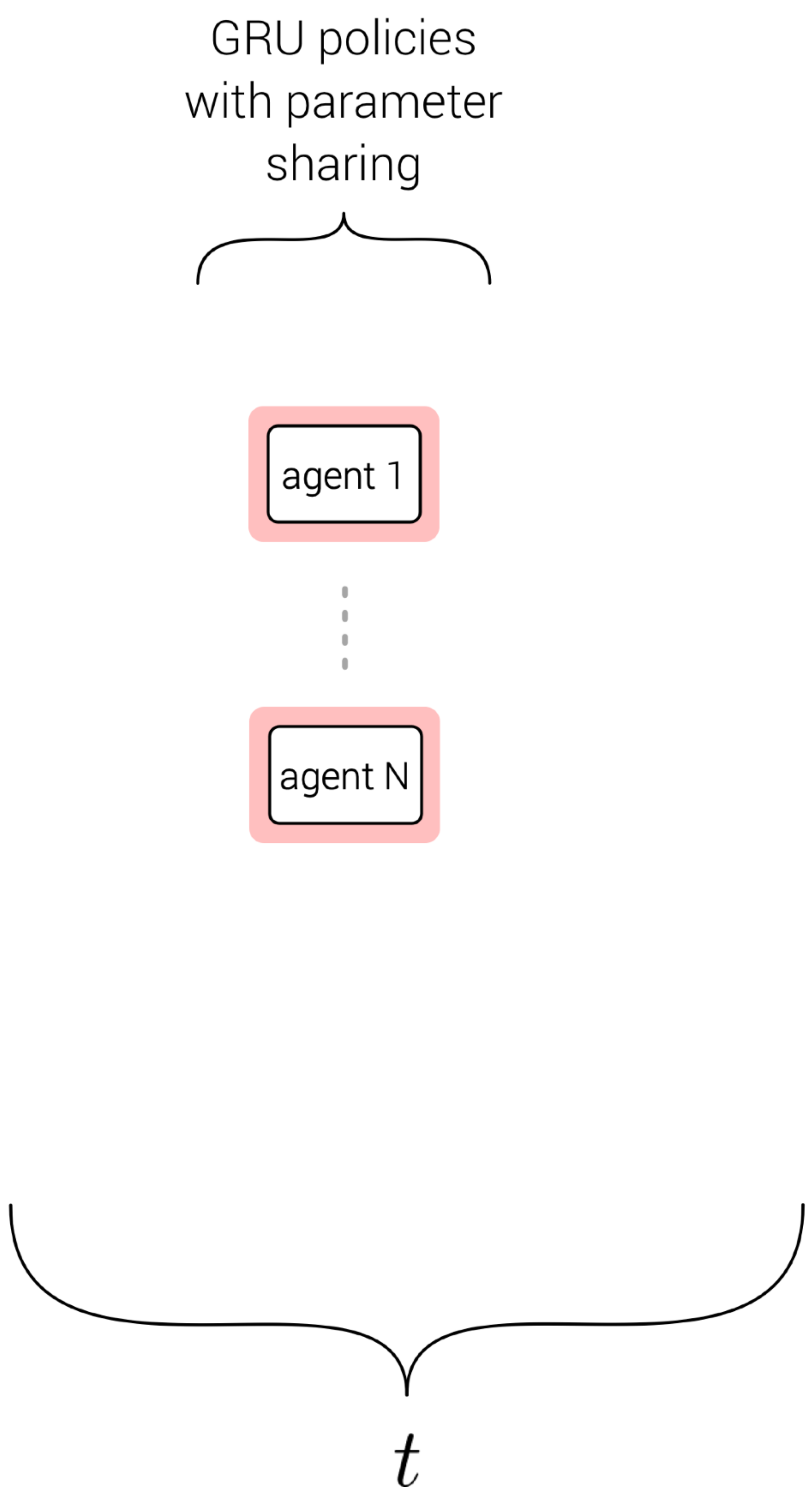
AFP / Getty Images.

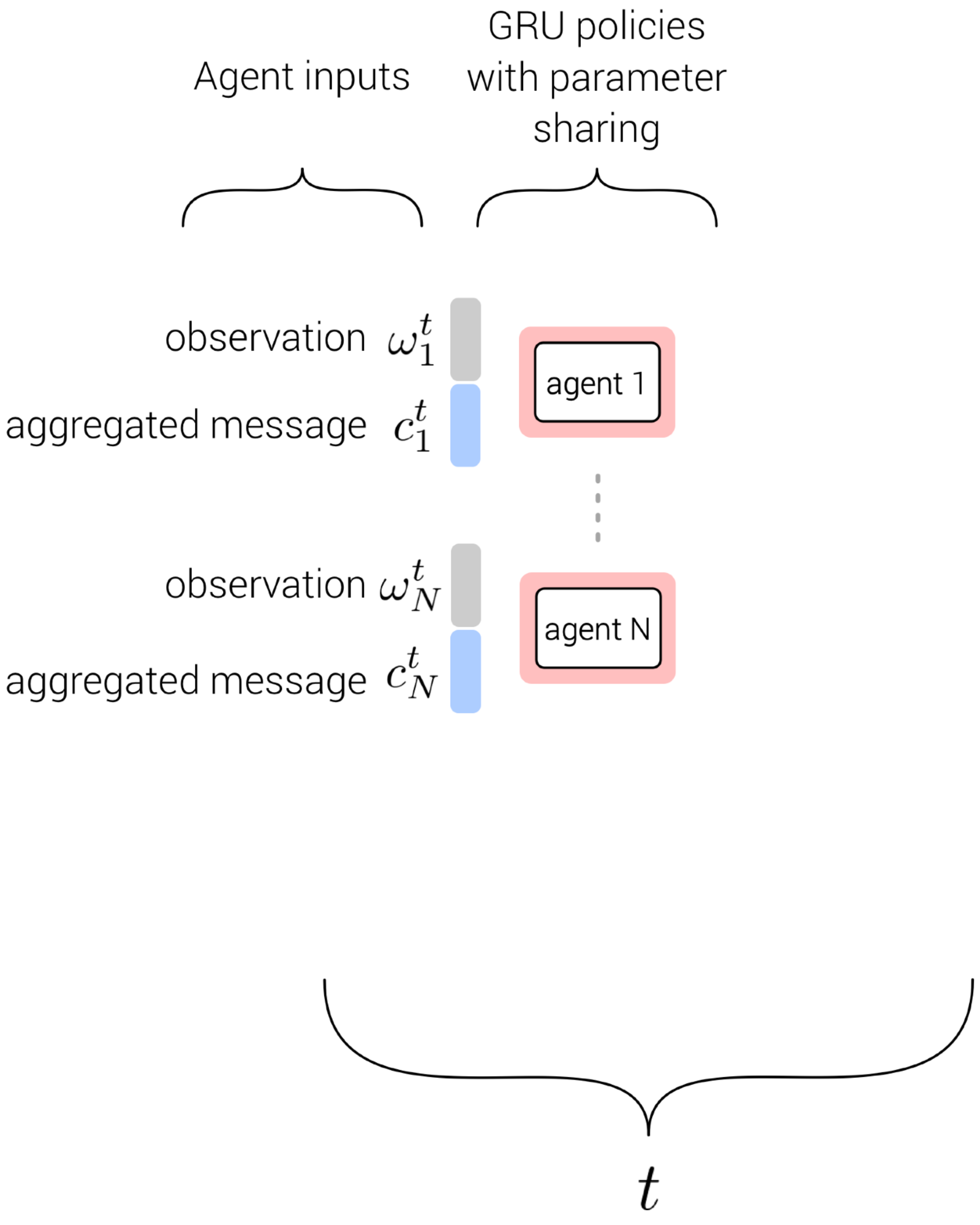


Budapest University

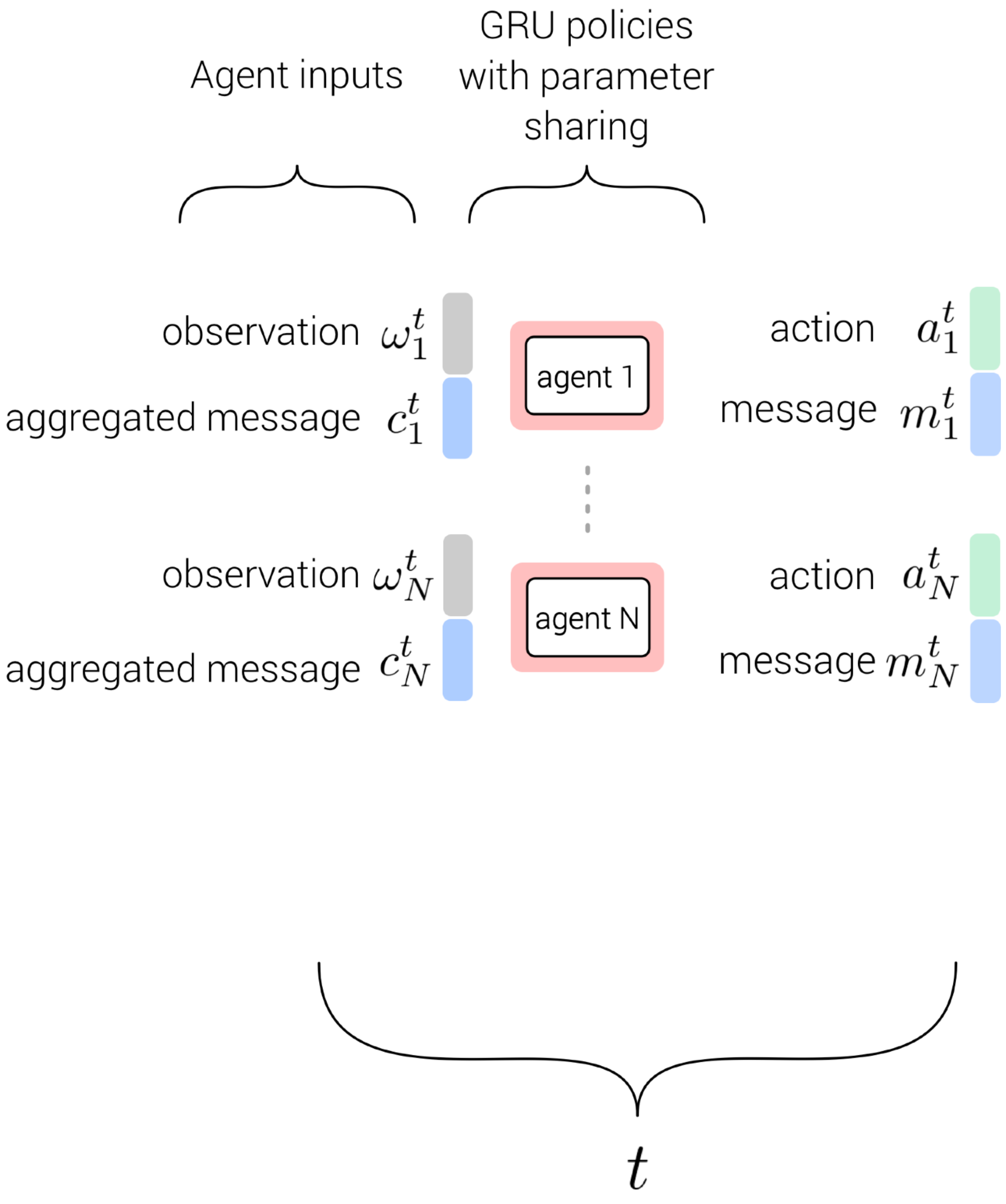

agent 1
agent N

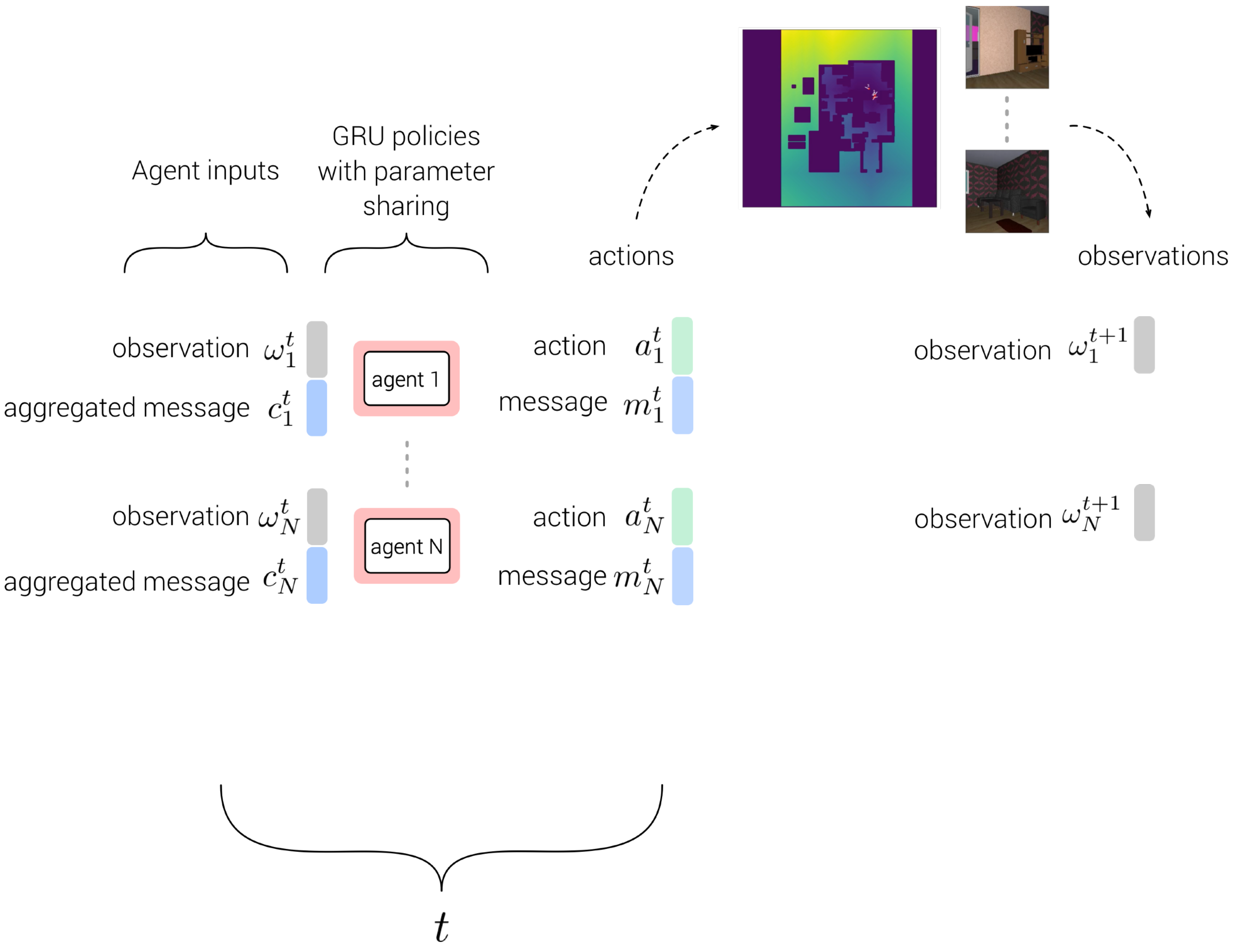

GRU policies
with parameter
sharing
agent 1
agent N


GRU policies
with parameter
sharing
agent 1
agent N
t

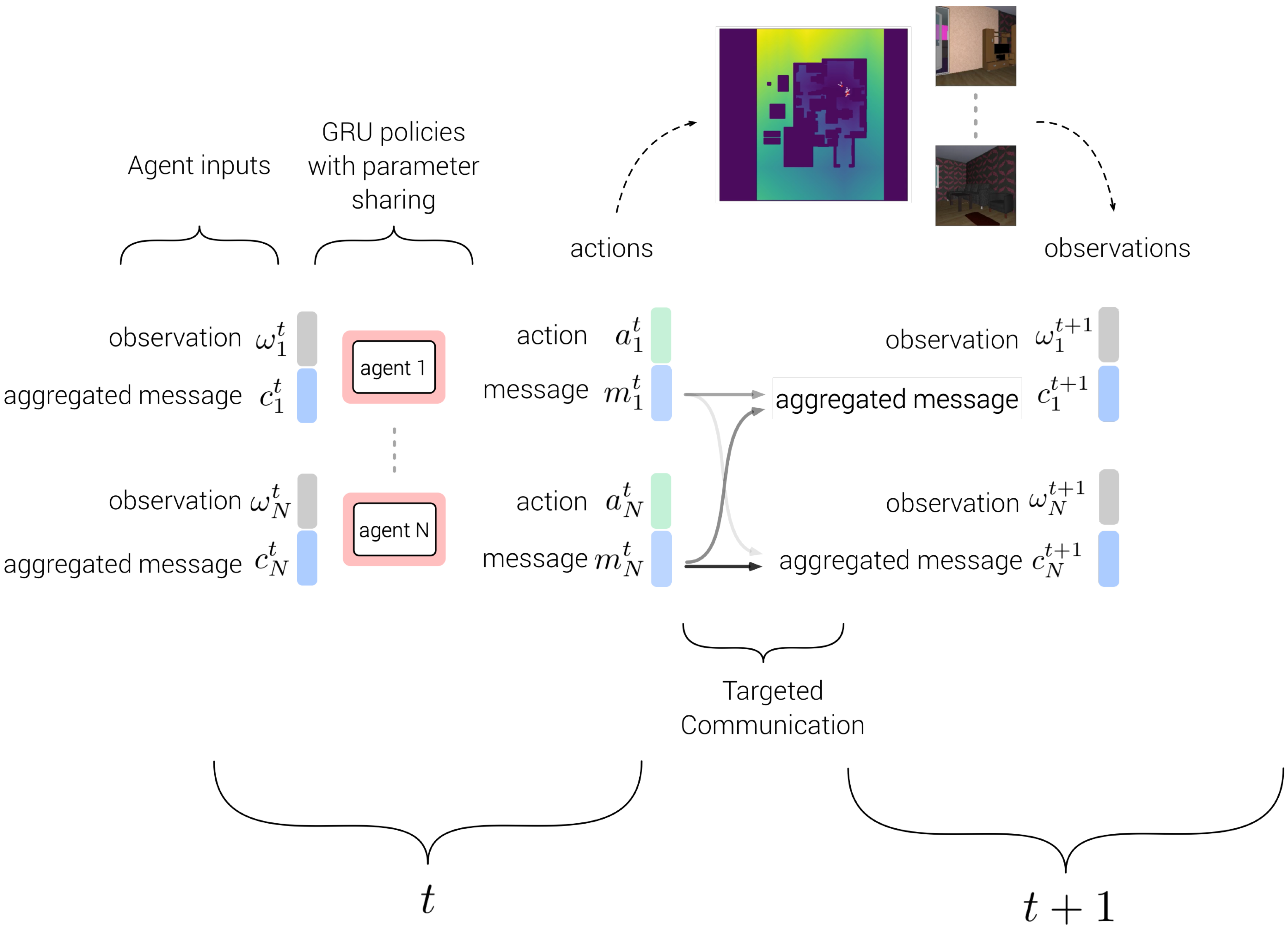


Agent inputs
GRU policies with parameter sharing
observation ω_1^t
aggregated message c_1^t
agent 1
observation ω_N^t
aggregated message c_N^t
agent N
t

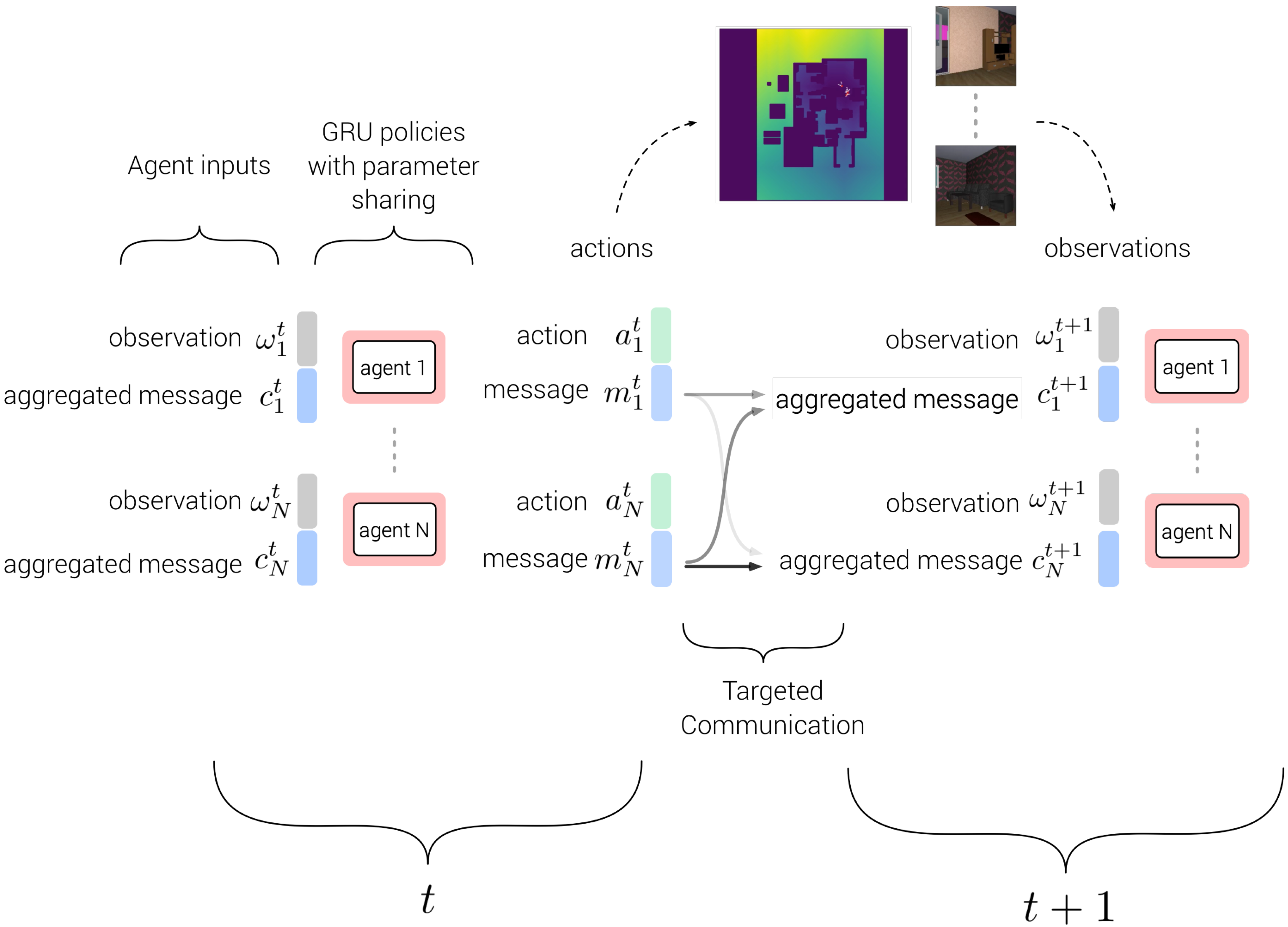

Agent inputs
GRU policies with parameter sharing
observation ω_1^t
aggregated message c_1^t
agent 1
action a_1^t
message m_1^t
observation ω_N^t
aggregated message c_N^t
agent N
action a_N^t
message m_N^t
t



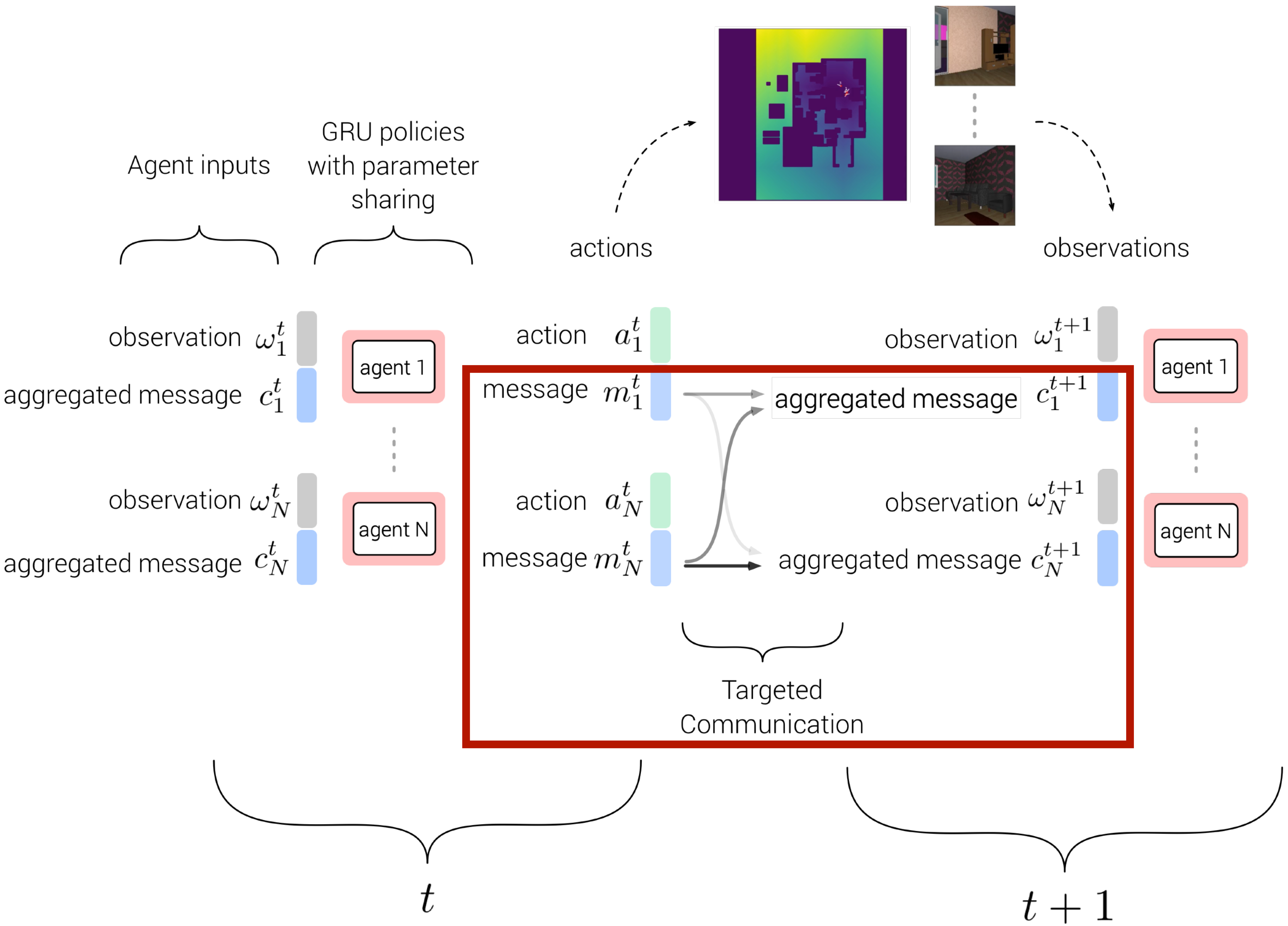
Agent inputs
GRU policies with parameter sharing
actions
observations
observation ω_1^t
aggregated message c_1^t
agent 1
action a_1^t
message m_1^t
observation ω_1^{t+1}
observation ω_N^t
aggregated message c_N^t
agent N
action a_N^t
message m_N^t
observation ω_N^{t+1}
t



Agent inputs
GRU policies with parameter sharing
actions
observations
observation ω_1^t
aggregated message c_1^t
agent 1
action a_1^t
message m_1^t
observation ω_1^{t+1}
aggregated message c_1^{t+1}
observation ω_N^t
aggregated message c_N^t
agent N
action a_N^t
message m_N^t
observation ω_N^{t+1}
aggregated message c_N^{t+1}
Targeted Communication
t
$t+1$

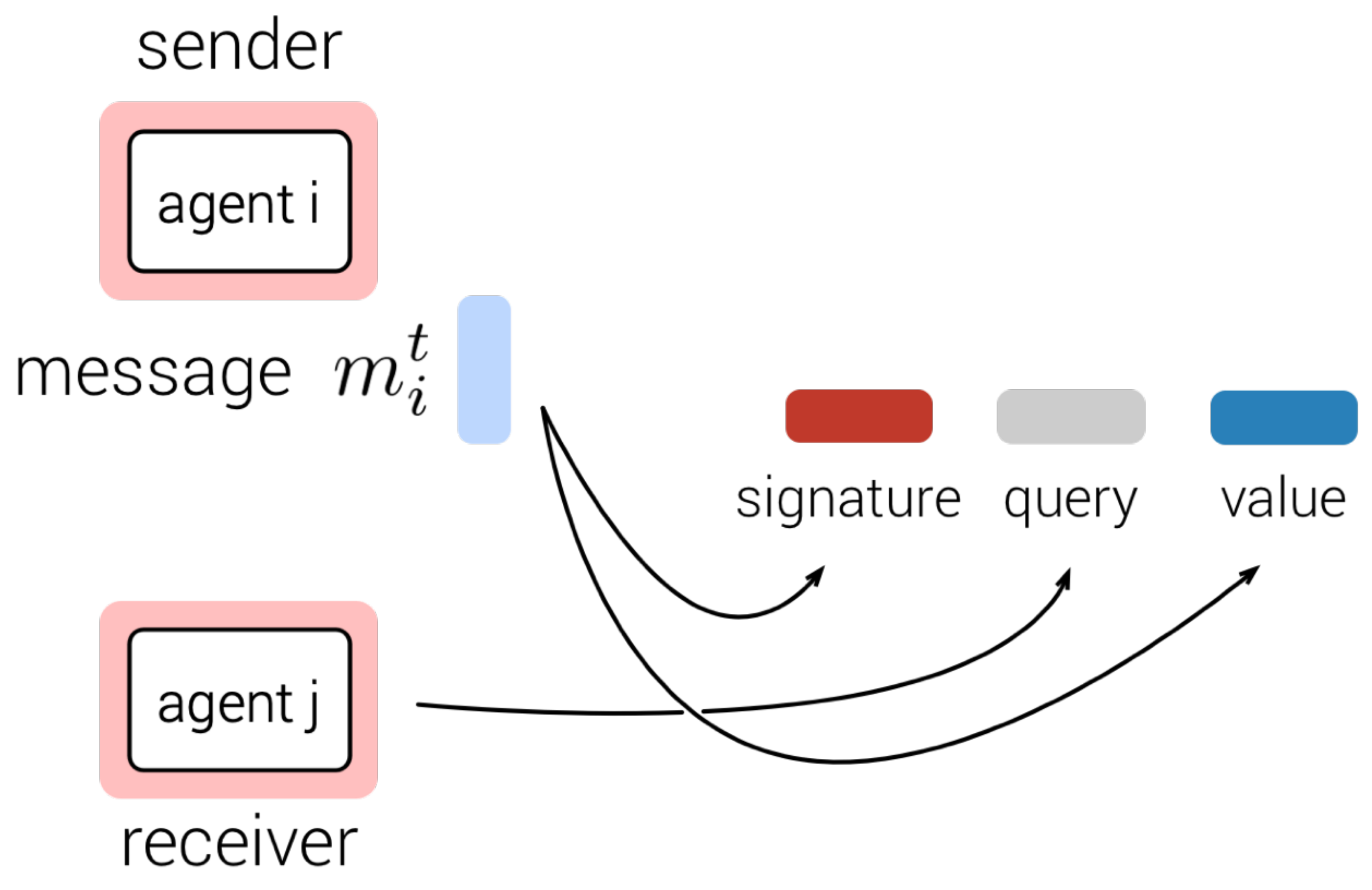


Agent inputs
GRU policies with parameter sharing
actions
observations
observation ω_1^t
aggregated message c_1^t
agent 1
observation ω_N^t
aggregated message c_N^t
agent N
action a_1^t
message m_1^t
action a_N^t
message m_N^t
observation ω_1^{t+1}
aggregated message c_1^{t+1}
agent 1
observation ω_N^{t+1}
aggregated message c_N^{t+1}
agent N
Targeted Communication
t
$t+1$

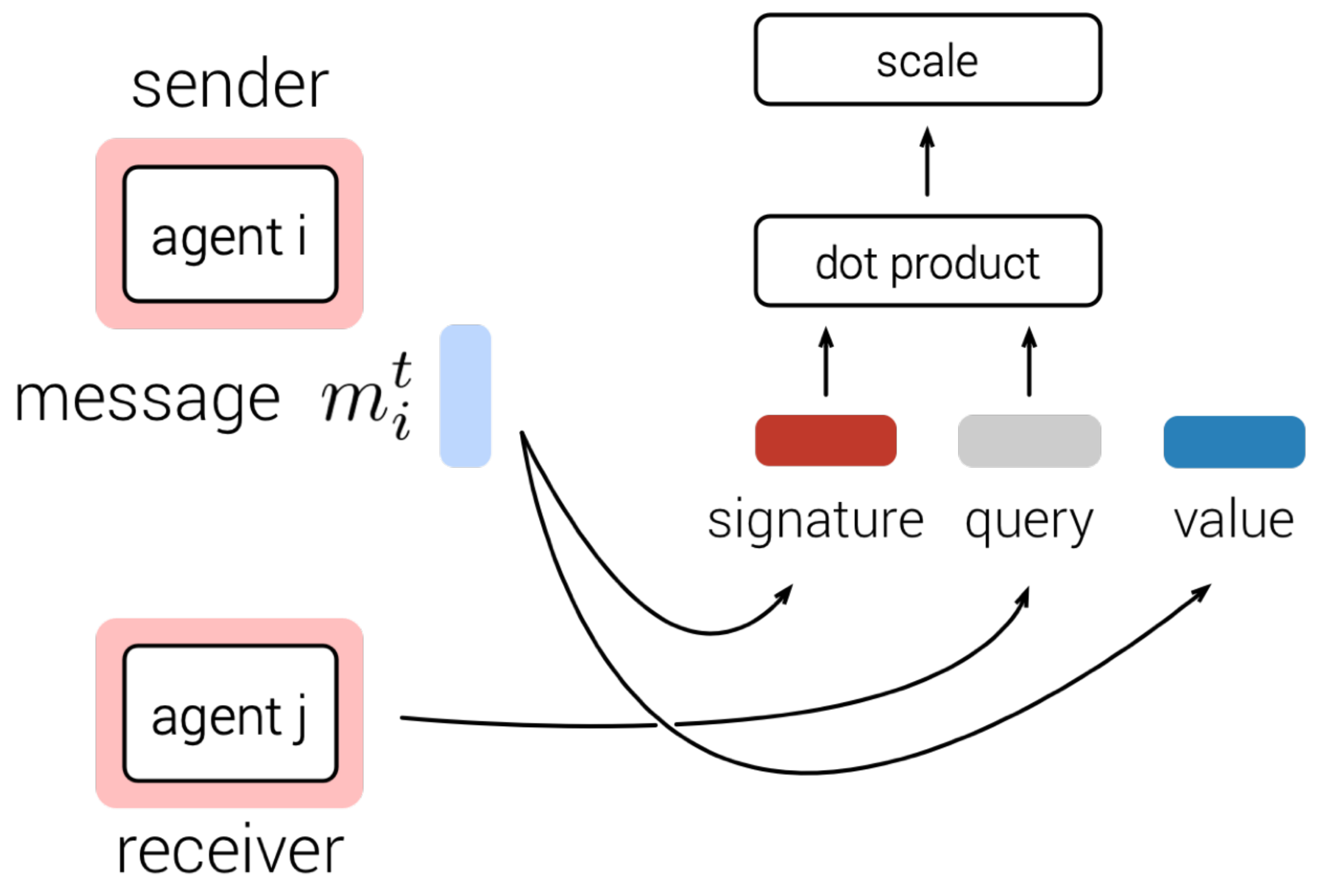

Agent inputs
GRU policies with parameter sharing
actions
observations
observation ω_1^t
aggregated message c_1^t
agent 1
observation ω_N^t
aggregated message c_N^t
agent N
action a_1^t
message m_1^t
action a_N^t
message m_N^t
observation ω_1^{t+1}
aggregated message c_1^{t+1}
observation ω_N^{t+1}
aggregated message c_N^{t+1}
agent 1
agent N
Targeted Communication
t
$t+1$

How does targeting work?

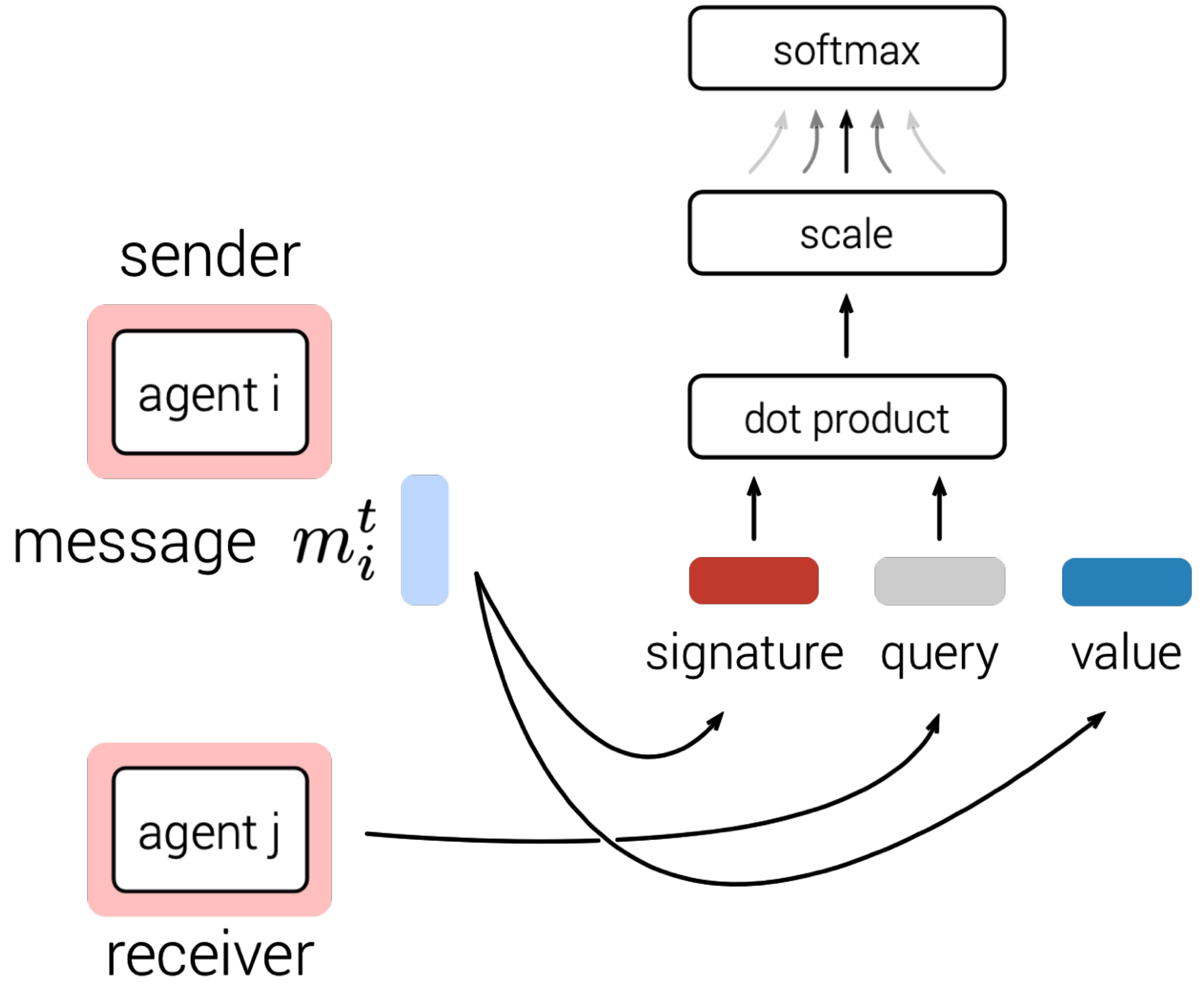
How does targeting work?



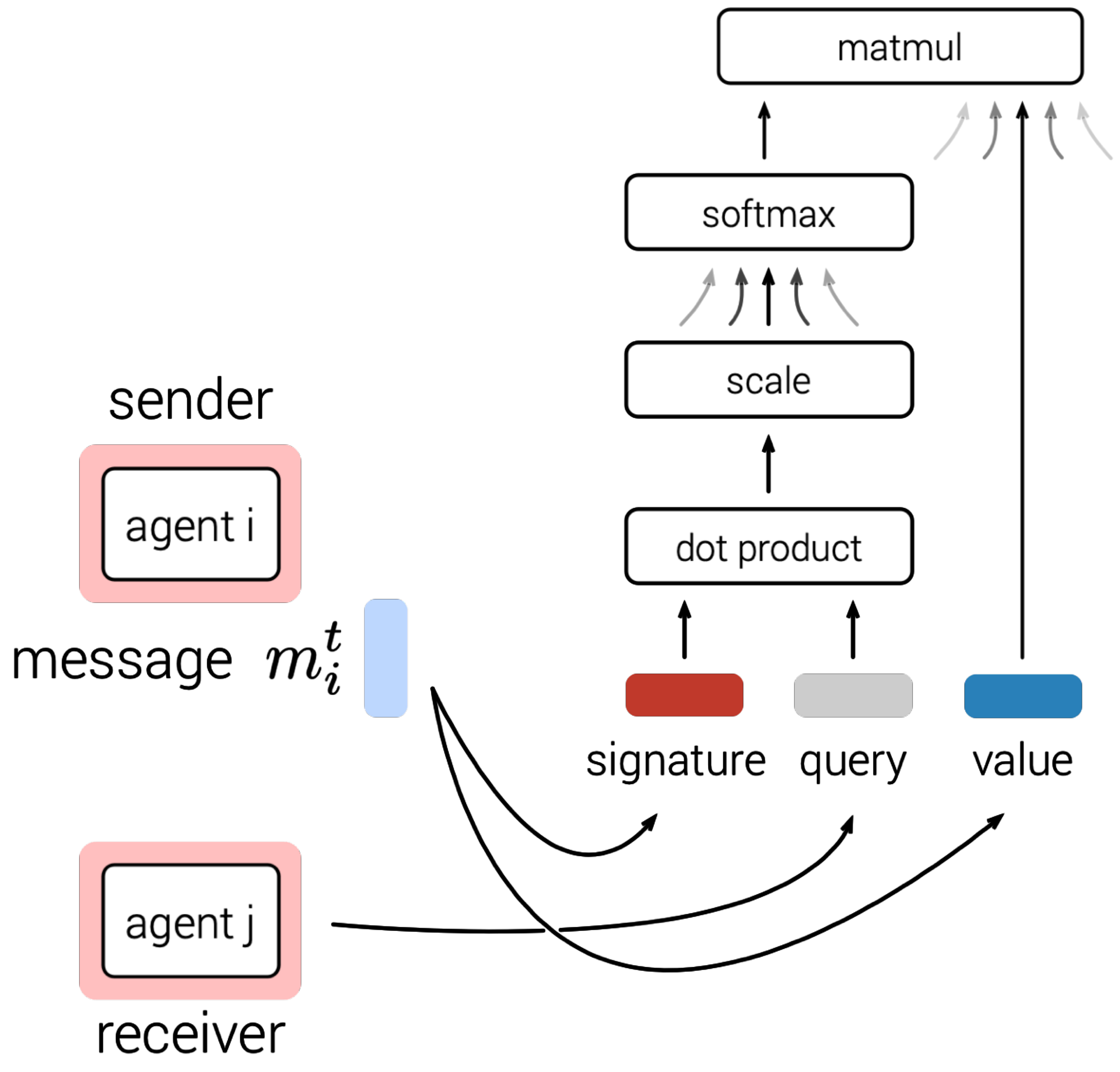
How does targeting work?



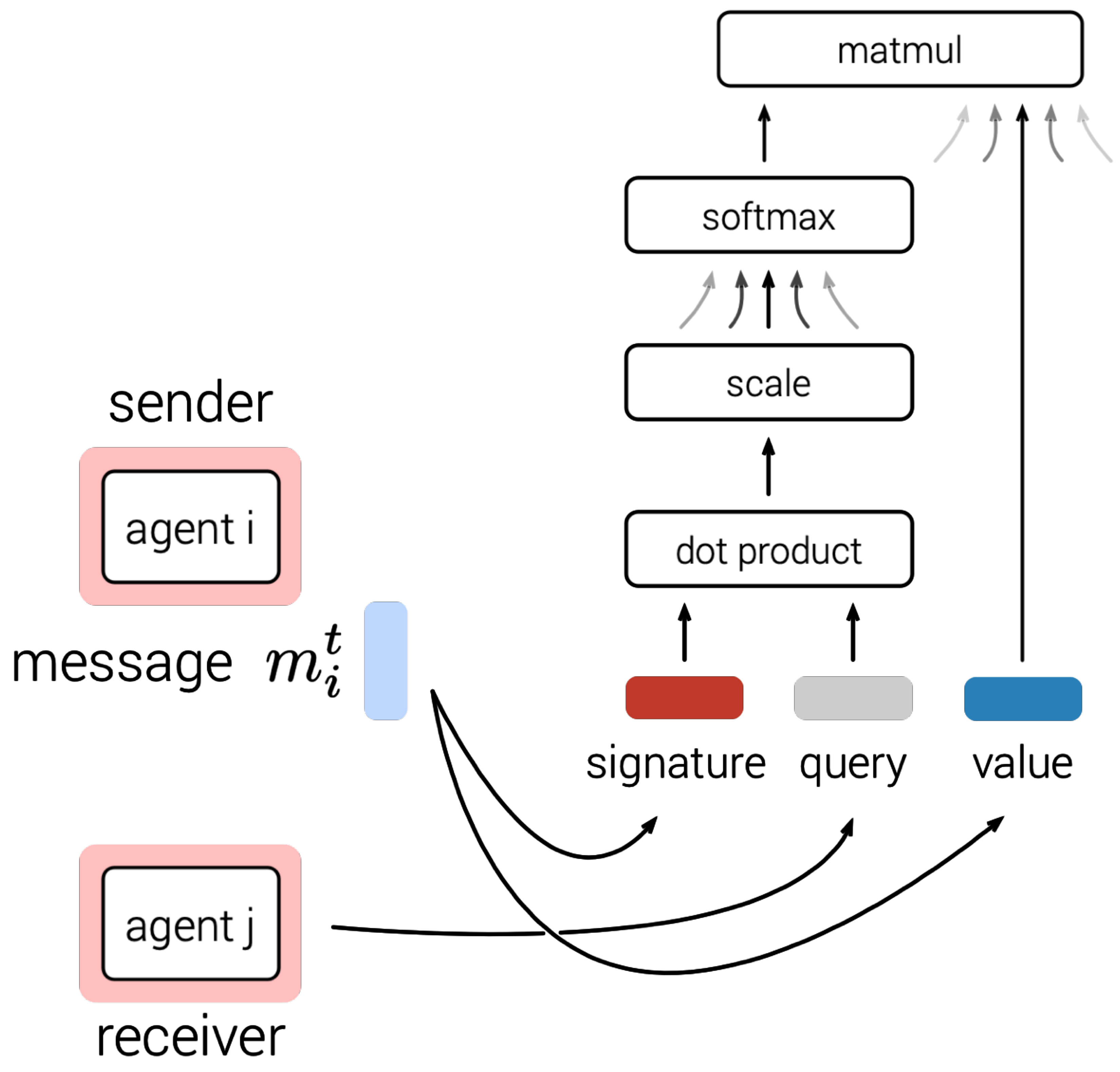
How does targeting work?



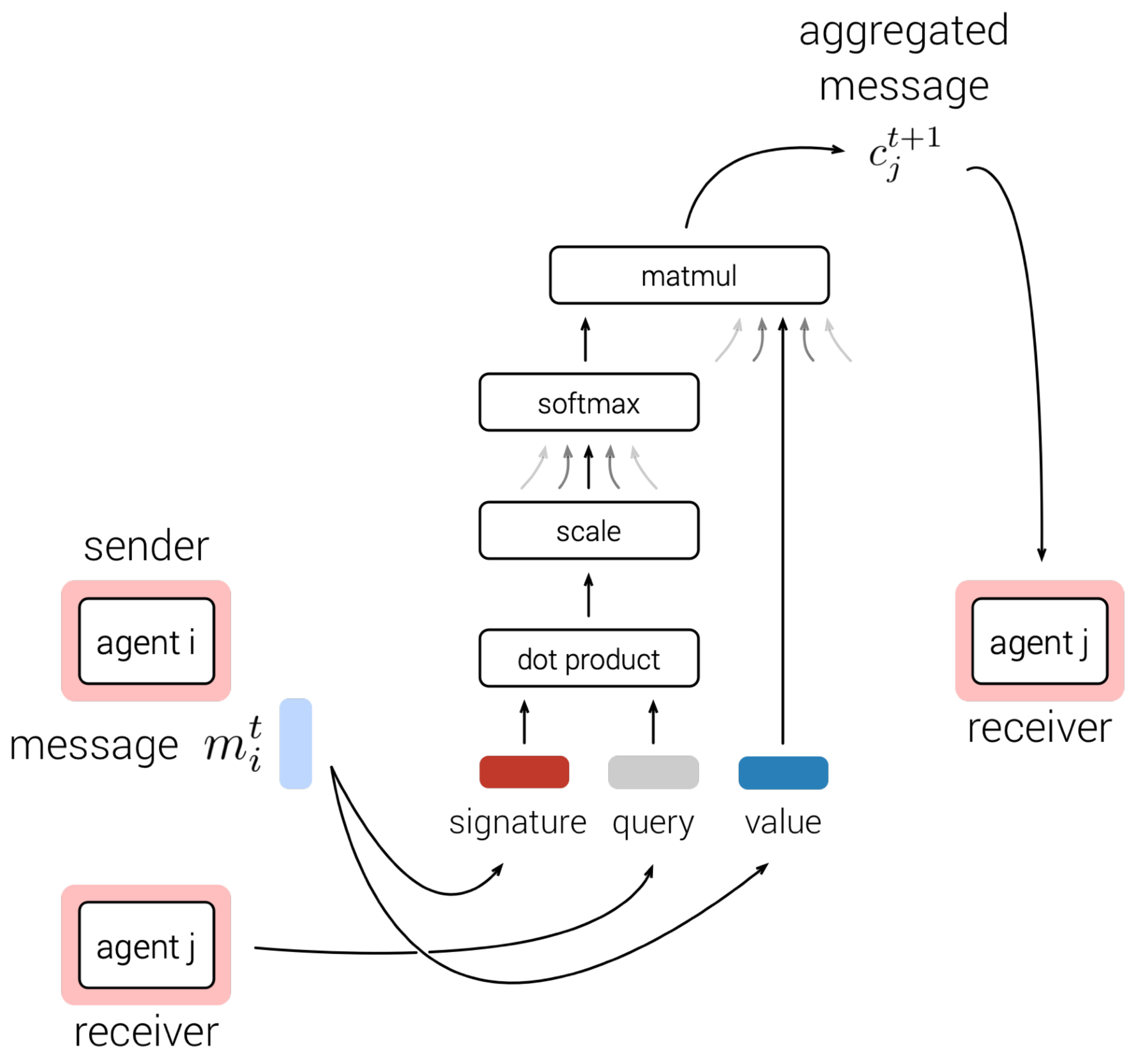
How does targeting work?



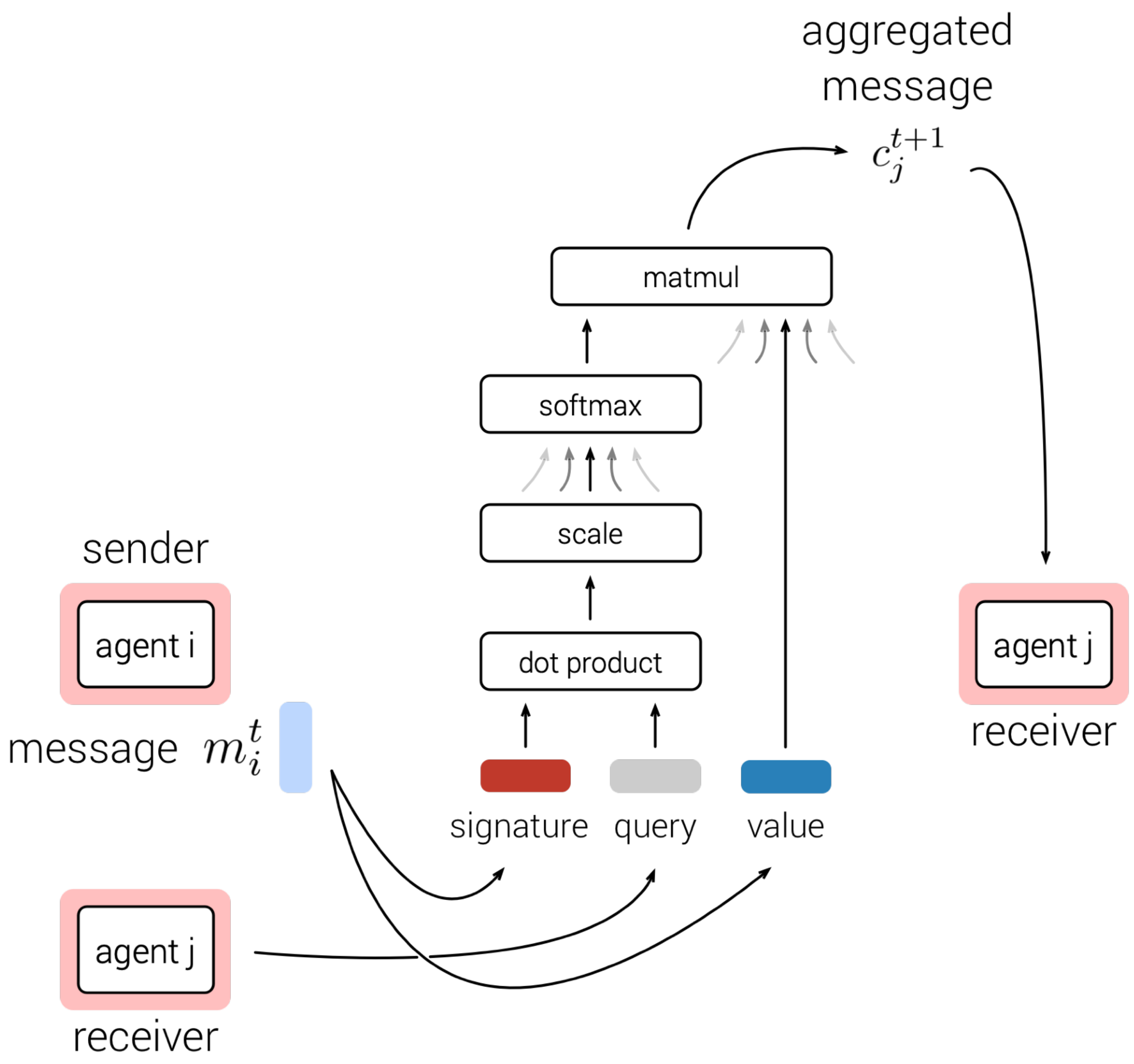
How does targeting work?



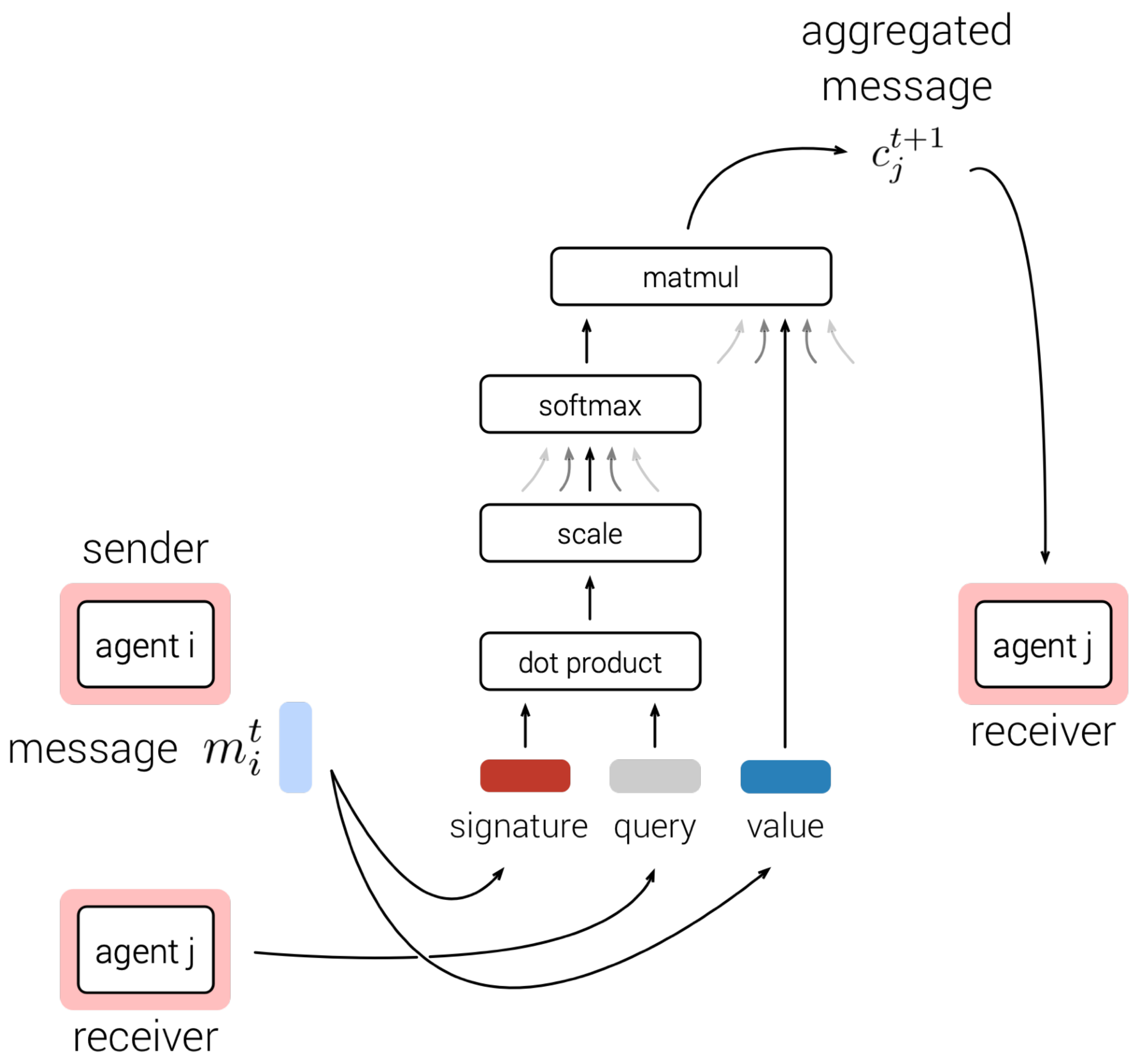
How does targeting work?



How does targeting work?



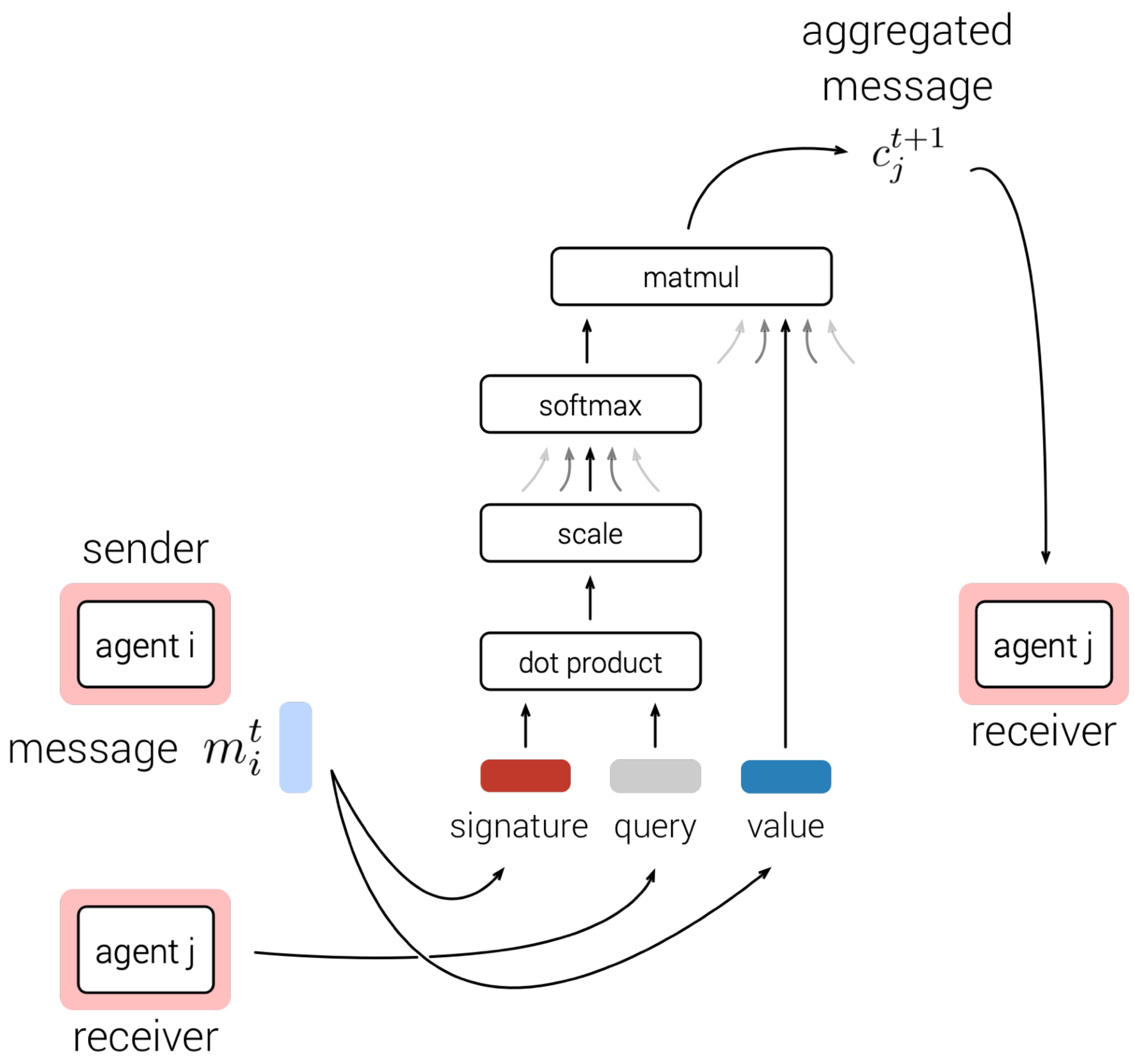
How does targeting work?



$$\alpha_j = \text{softmax}\left[\frac{{q_j^{t+1}}^T k_1^t}{\sqrt{d_k}} \ldots \underbrace{\frac{{q_j^{t+1}}^T k_i^t}{\sqrt{d_k}}}_{\alpha_{ji}} \ldots \frac{{q_j^{t+1}}^T k_N^t}{\sqrt{d_k}} \right]$$

$$c_j^{t+1} = \sum_{i=1}^{N} \alpha_{ji} v_i^t$$

How does targeting work?

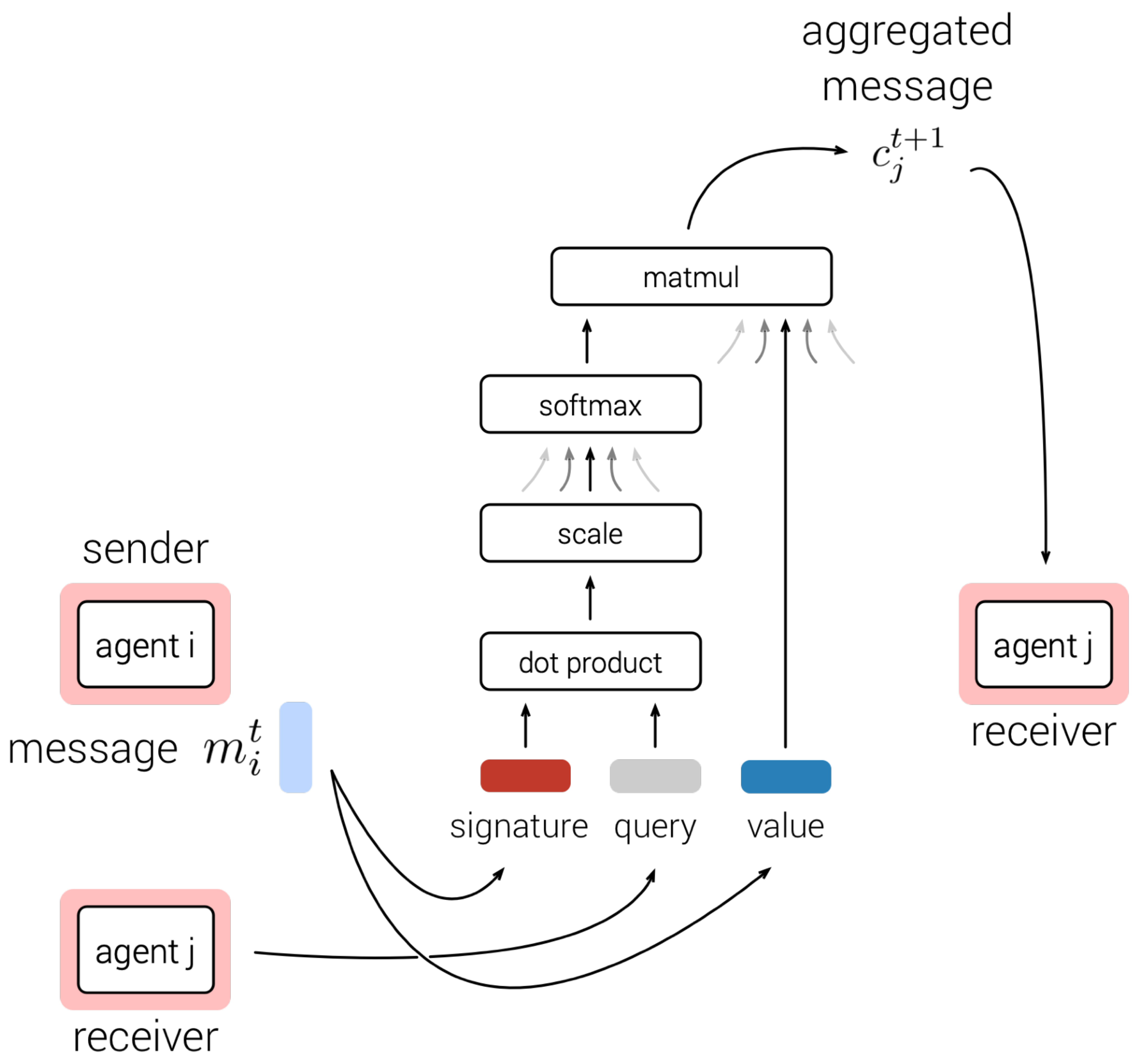


$$\alpha_j = \text{softmax}\left[\frac{{q_j^{t+1}}^T k_1^t}{\sqrt{d_k}} \ \dots \ \underbrace{\frac{{q_j^{t+1}}^T k_i^t}{\sqrt{d_k}}}_{\alpha_{ji}} \ \dots \ \frac{{q_j^{t+1}}^T k_N^t}{\sqrt{d_k}}\right]$$

$$c_j^{t+1} = \sum_{i=1}^{N} \alpha_{ji} v_i^t$$

- Operationalized as soft attention
  - differentiable
  - interpretable

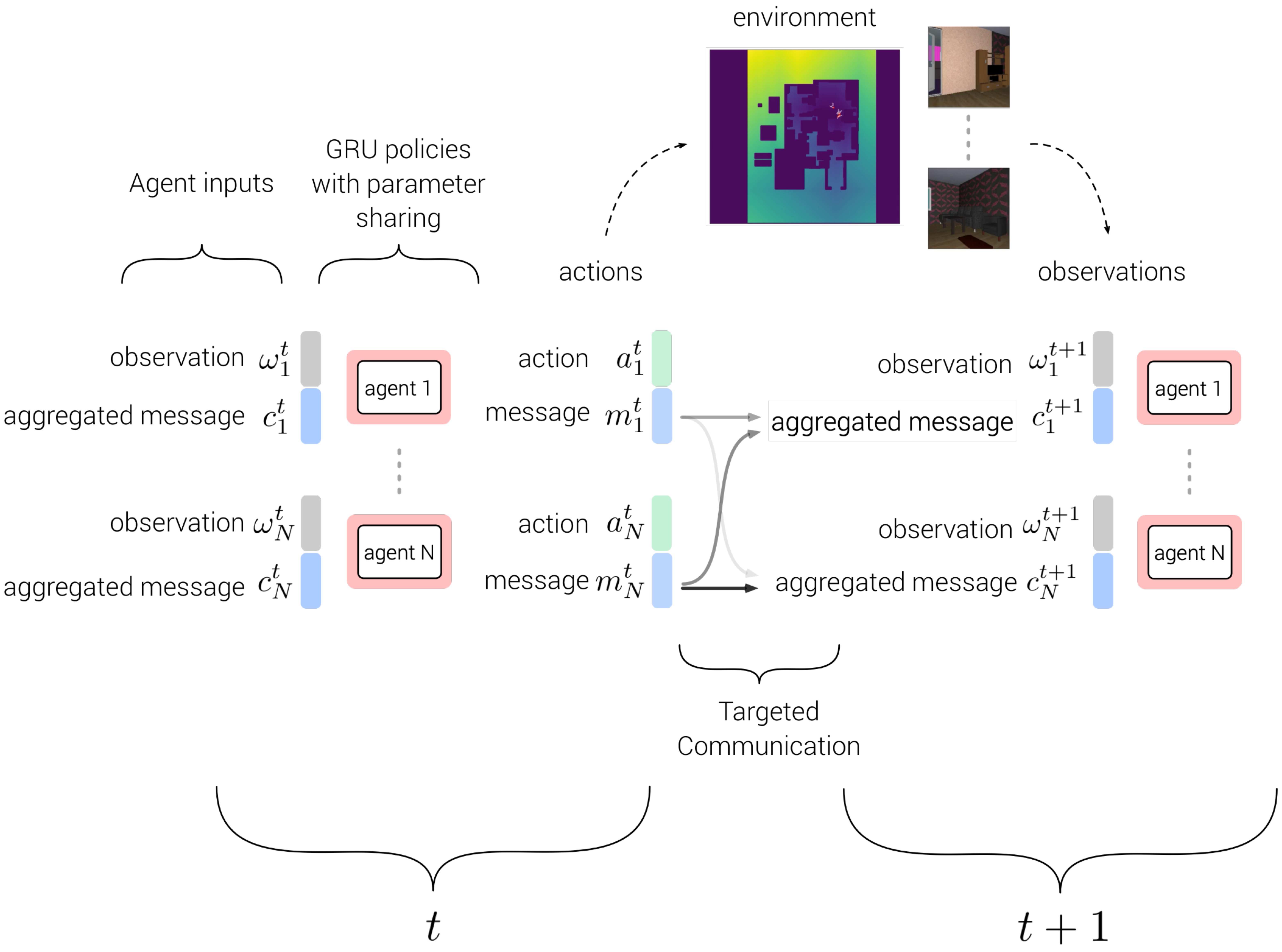
How does targeting work?

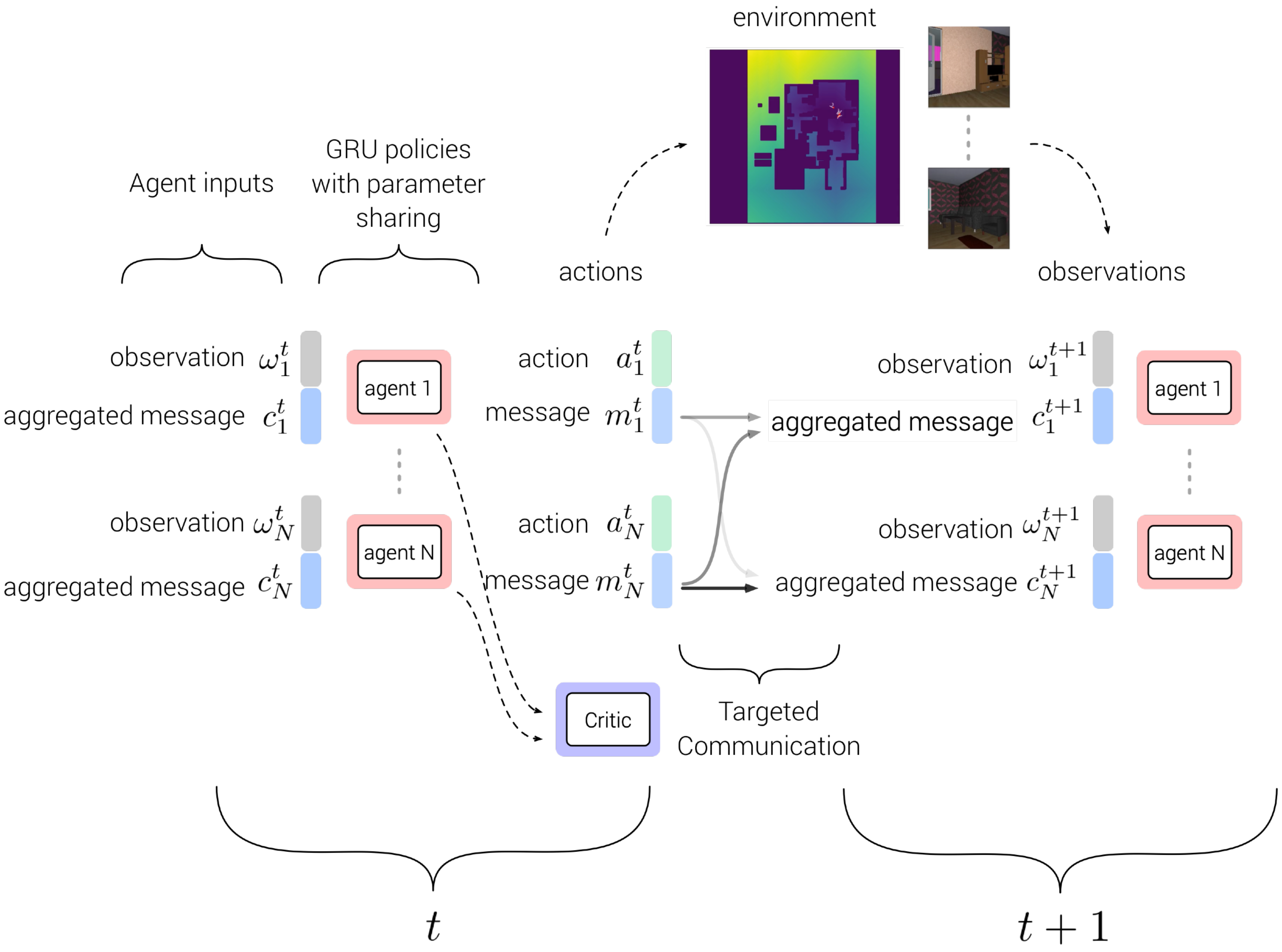


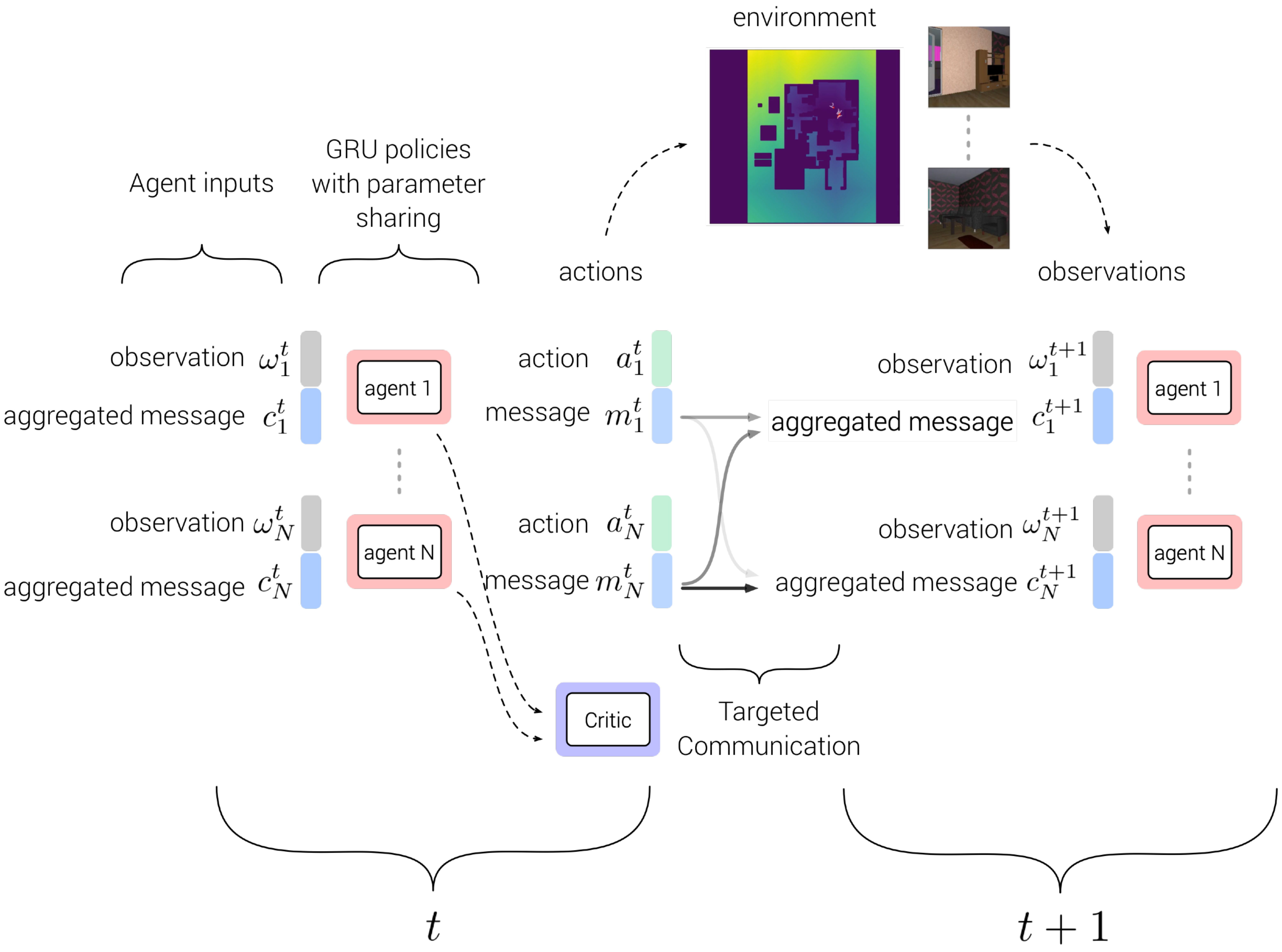
$$\alpha_j = \text{softmax}\left[\frac{{q_j^{t+1}}^T k_1^t}{\sqrt{d_k}} \cdots \underbrace{\frac{{q_j^{t+1}}^T k_i^t}{\sqrt{d_k}}}_{\alpha_{ji}} \cdots \frac{{q_j^{t+1}}^T k_N^t}{\sqrt{d_k}}\right]$$

$$c_j^{t+1} = \sum_{i=1}^{N} \alpha_{ji} v_i^t$$

- Operationalized as soft attention
  - differentiable
  - interpretable
- Agents encode recipient properties in “query” vectors
  - e.g. “cars traveling on the west to east road”


environment
Agent inputs
GRU policies with parameter sharing
actions
observations
observation ω_1^t
aggregated message c_1^t
agent 1
observation ω_N^t
aggregated message c_N^t
agent N
action a_1^t
message m_1^t
action a_N^t
message m_N^t
observation ω_1^{t+1}
aggregated message c_1^{t+1}
agent 1
observation ω_N^{t+1}
aggregated message c_N^{t+1}
agent N
Targeted Communication
t
$t+1$


environment
Agent inputs
GRU policies with parameter sharing
actions
observations
observation ω_1^t
aggregated message c_1^t
agent 1
observation ω_N^t
aggregated message c_N^t
agent N
action a_1^t
message m_1^t
action a_N^t
message m_N^t
observation ω_1^{t+1}
aggregated message c_1^{t+1}
agent 1
observation ω_N^{t+1}
aggregated message c_N^{t+1}
agent N
Critic
Targeted Communication
t
$t+1$

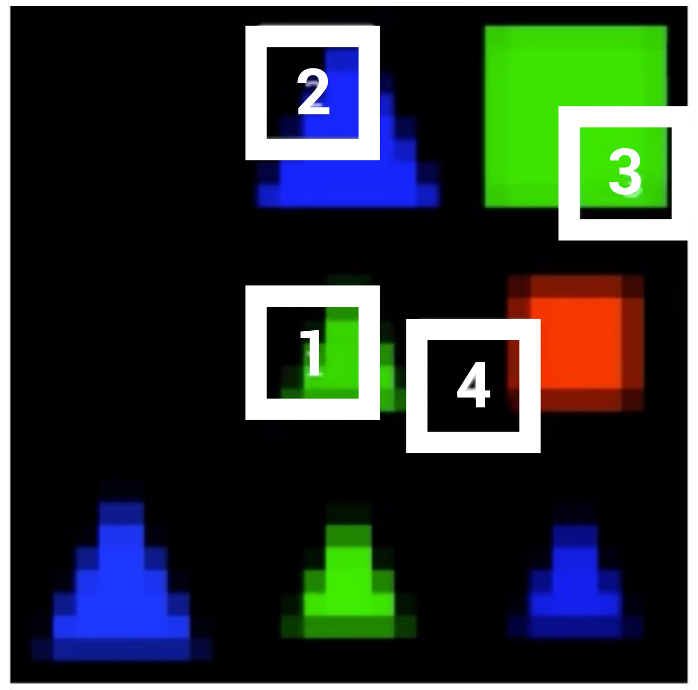


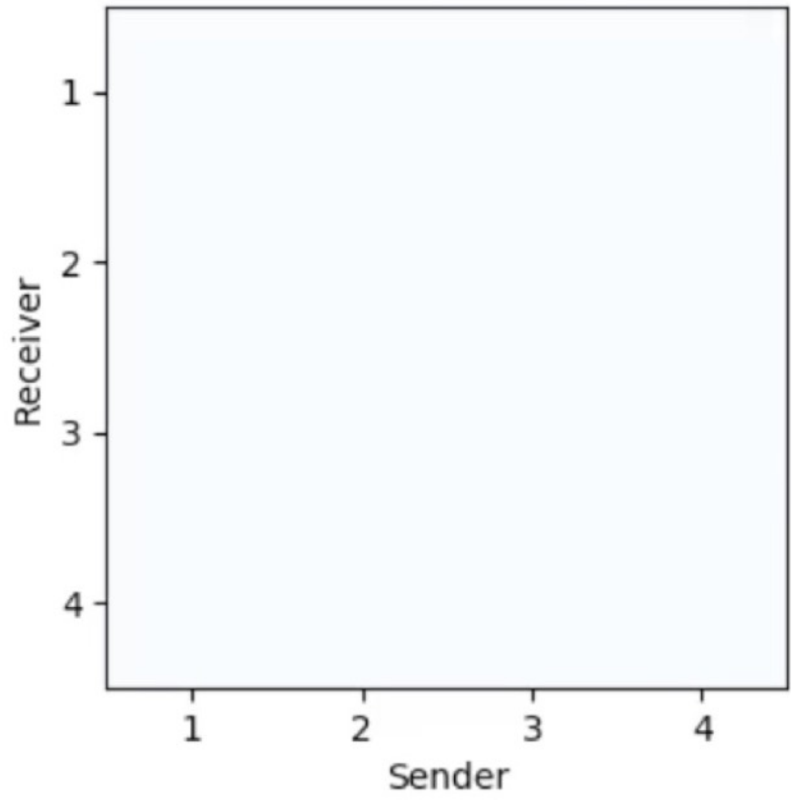
- Decentralized actors
- Centralized critic

  (Lowe et al., 2017)
- Targeted continuous communication

Results: SHAPES

4 agents in a gridworld with partial observations
looking for **red**, **red**, **green**, **blue** respectively

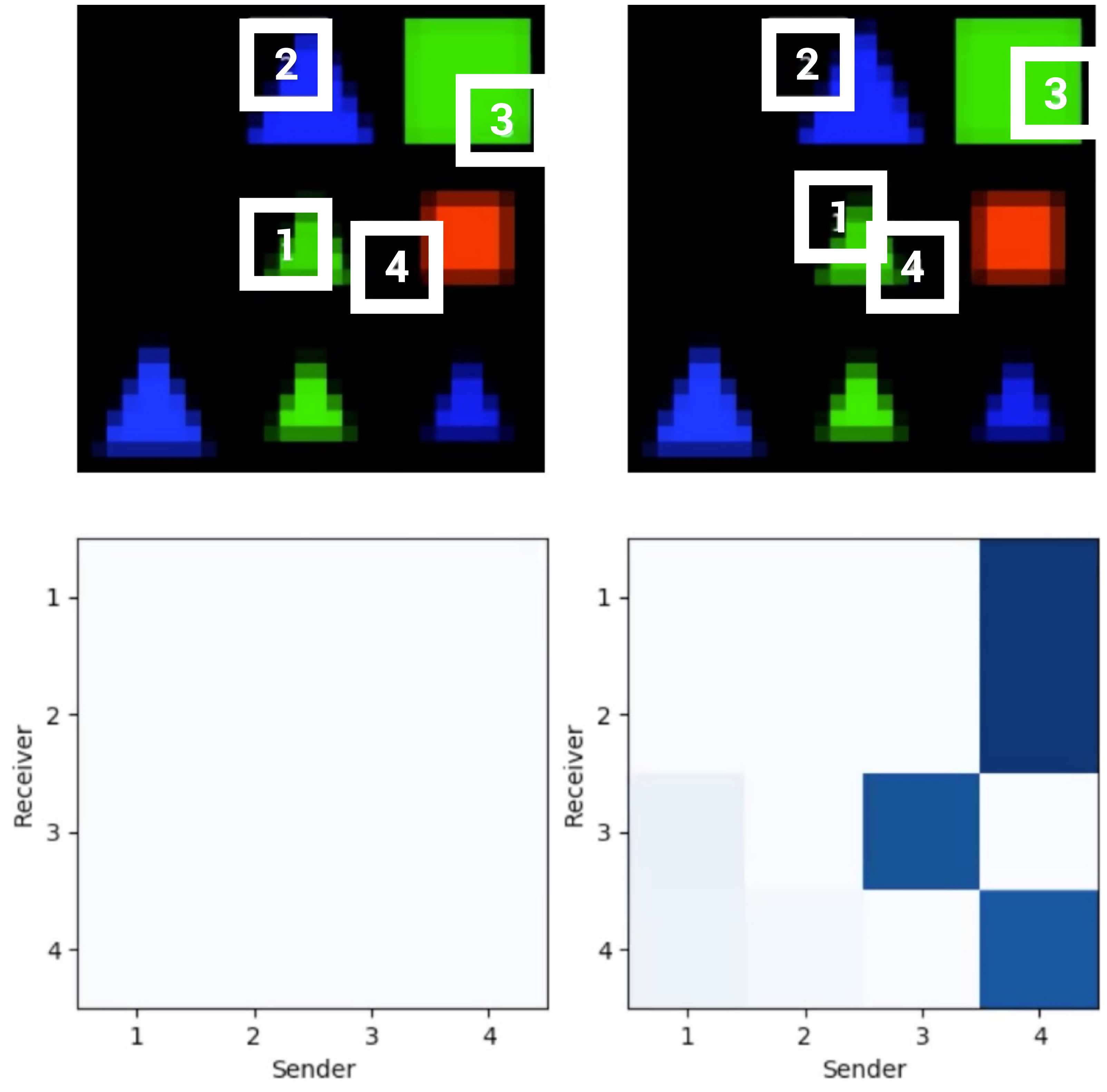




t = 1

Results: SHAPES

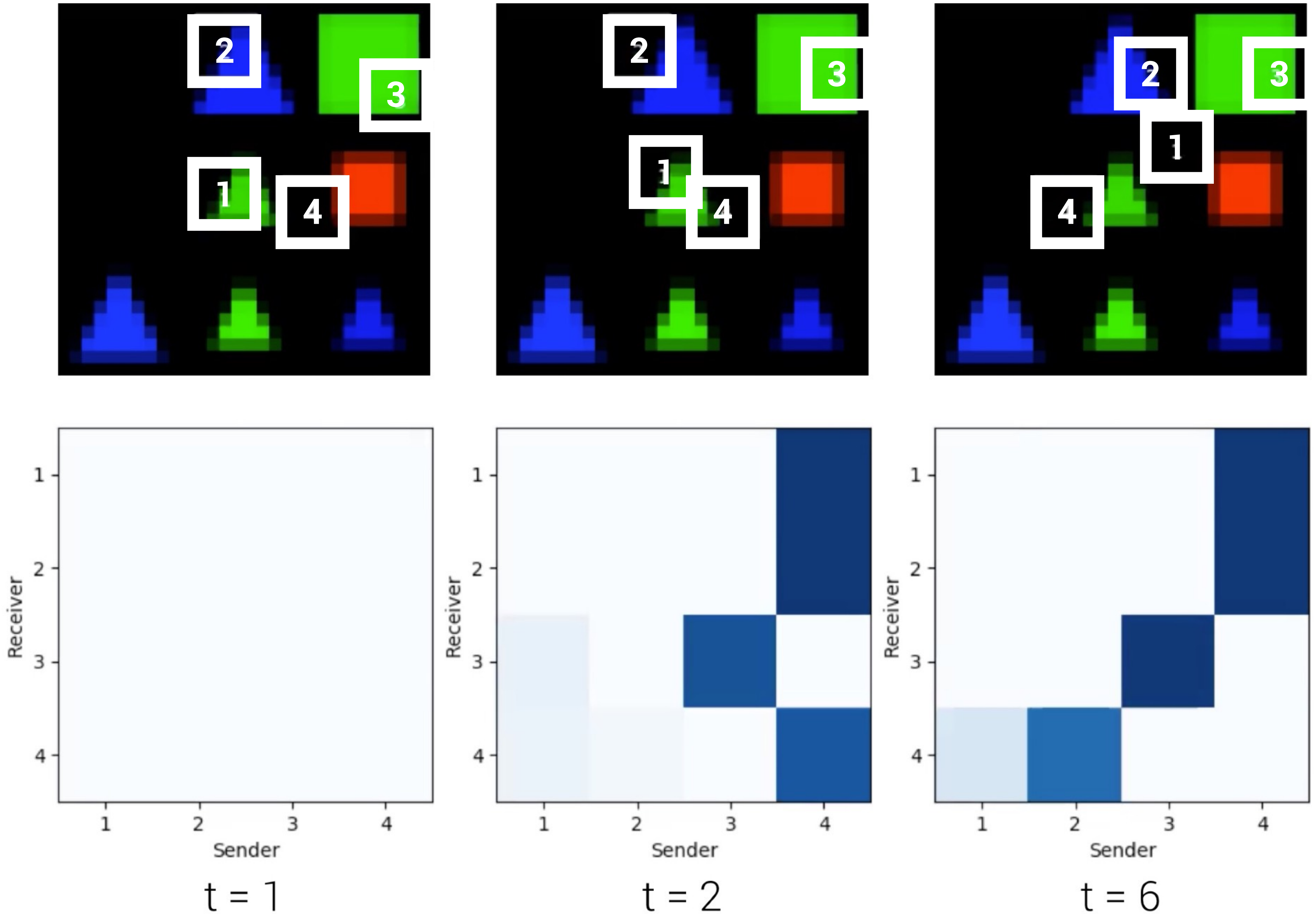
4 agents in a gridworld with partial observations
looking for **red**, **red**, **green**, **blue** respectively



t = 1   t = 2

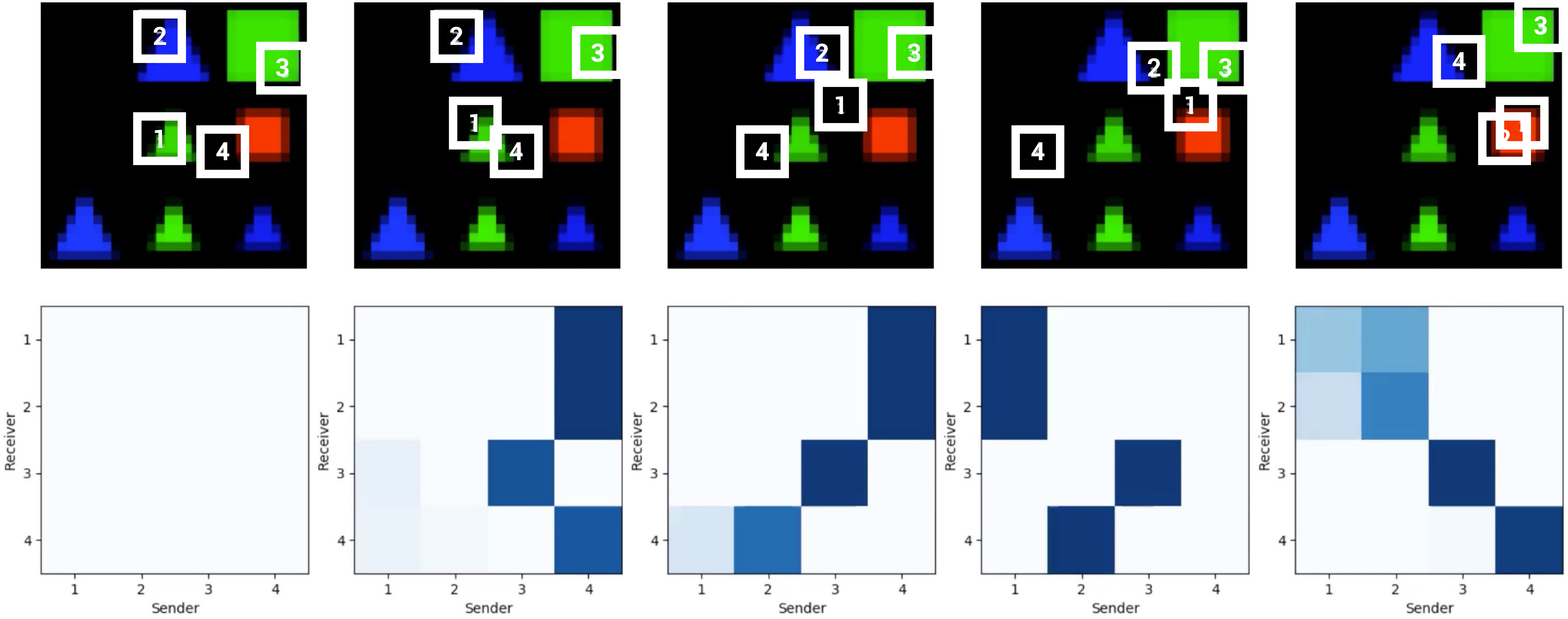
Results: SHAPES

4 agents in a gridworld with partial observations
looking for **red**, **red**, **green**, **blue** respectively



Results: SHAPES

4 agents in a gridworld with partial observations
looking for **red**, **red**, **green**, **blue** respectively



t = 1 t = 2 t = 6 t = 8 t = 21

Results: SHAPES

	30 × 30, 4 agents, find[red]	50 × 50, 4 agents, find[red]	50 × 50, 4 agents, find[red,red,green,blue]
No communication	95.3±2.8%	83.6±3.3%	69.1±4.6%
No attention	**99.7**±0.8%	**89.5**±1.4%	82.4±2.1%
TarMAC	**99.8**±0.9%	**89.5**±1.7%	**85.8**±2.5%

Table 2: Success rates on 3 different settings of cooperative navigation in the SHAPES environment.

Results: SHAPES

	30×30, 4 agents, find[red]	50×50, 4 agents, find[red]	50×50, 4 agents, find[red,red,green,blue]
No communication	$95.3 \pm 2.8\%$	$83.6 \pm 3.3\%$	$69.1 \pm 4.6\%$
No attention	$\mathbf{99.7} \pm 0.8\%$	$\mathbf{89.5} \pm 1.4\%$	$82.4 \pm 2.1\%$
TarMAC	$\mathbf{99.8} \pm 0.9\%$	$\mathbf{89.5} \pm 1.7\%$	$\mathbf{85.8} \pm 2.5\%$

Table 2: Success rates on 3 different settings of cooperative navigation in the SHAPES environment.

Benefits of communication and attention increase with task complexity.

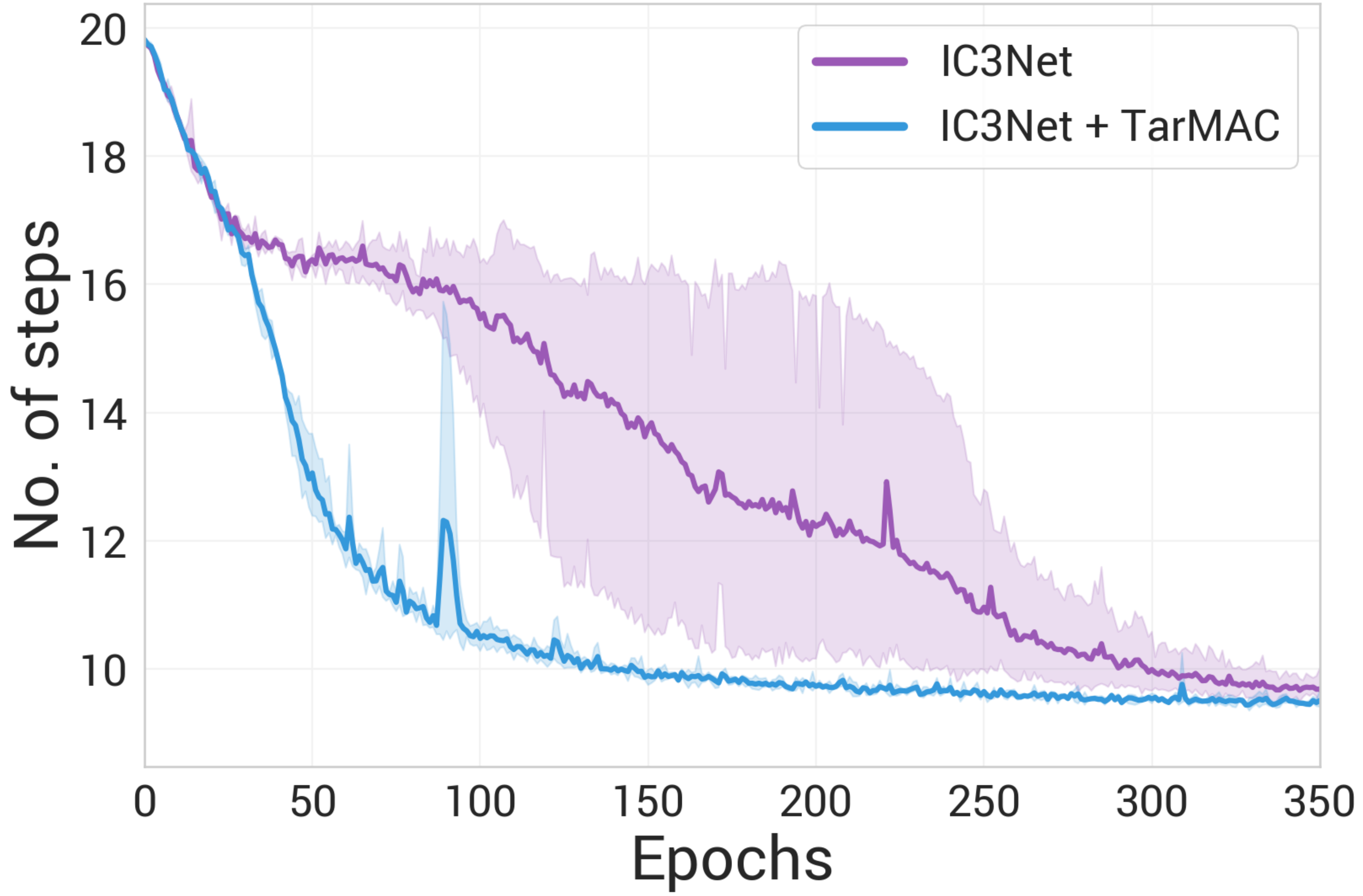
Results: Traffic Junction

	Easy	Hard
No communication	84.9±4.3%	74.1±3.9%
CommNet (Sukhbaatar et al., 2016)	99.7±0.1%	78.9±3.4%
TarMAC 1-round	**99.9**±0.1%	84.6±3.2%
TarMAC 2-round	**99.9**±0.1%	**97.1**±1.6%

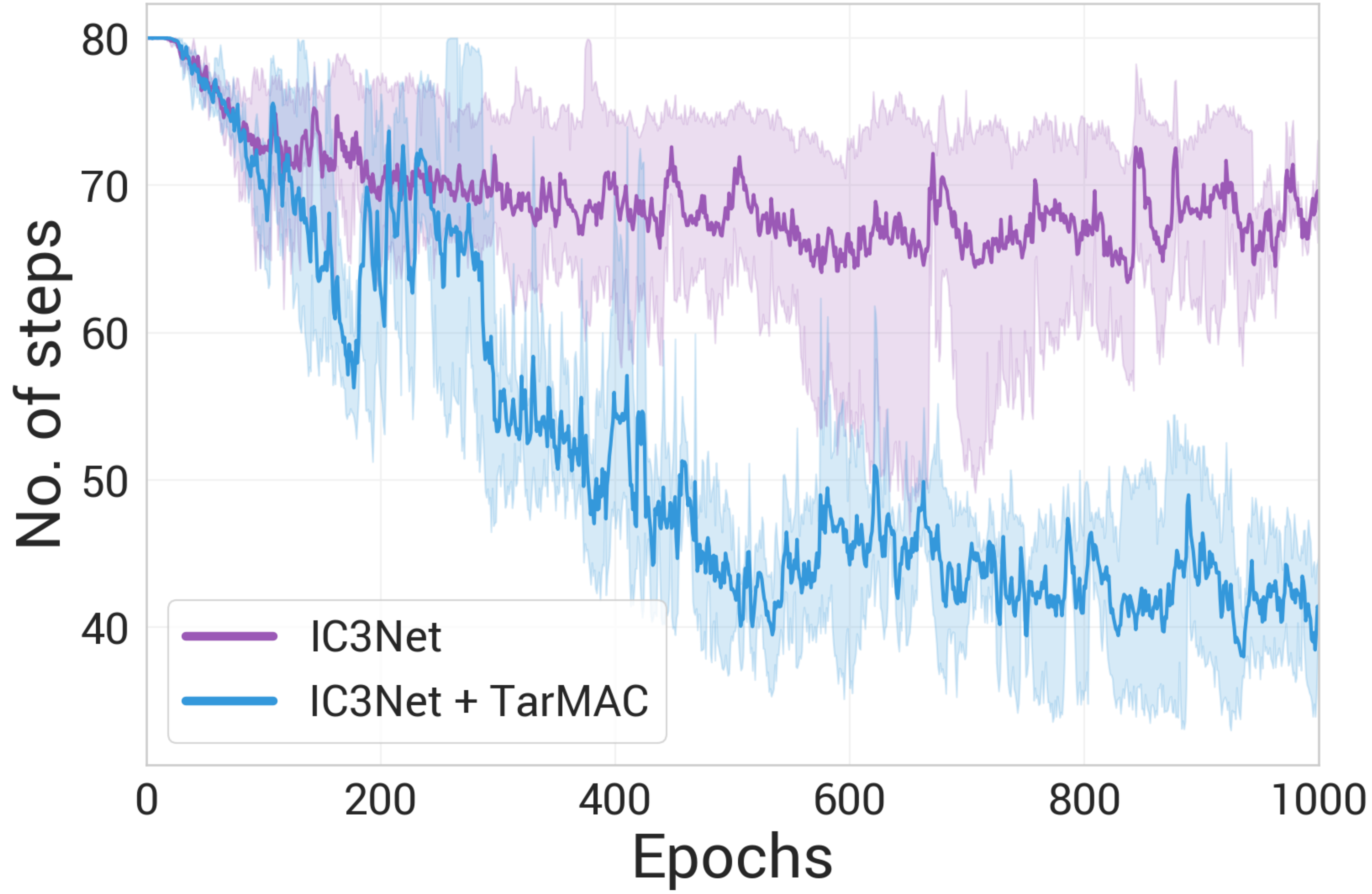
Table 3: Success rates on traffic junction. Our targeted 2-round communication architecture gets a success rate of 97.1±1.6% on the 'hard' variant, significantly outperforming Sukhbaatar et al. (2016). Note that 1- and 2-round refer to the number of rounds of communication between actions (Equation 4).

Benefits of communication and attention increase with task complexity.

Results: Extension to competitive tasks — Predator-Prey



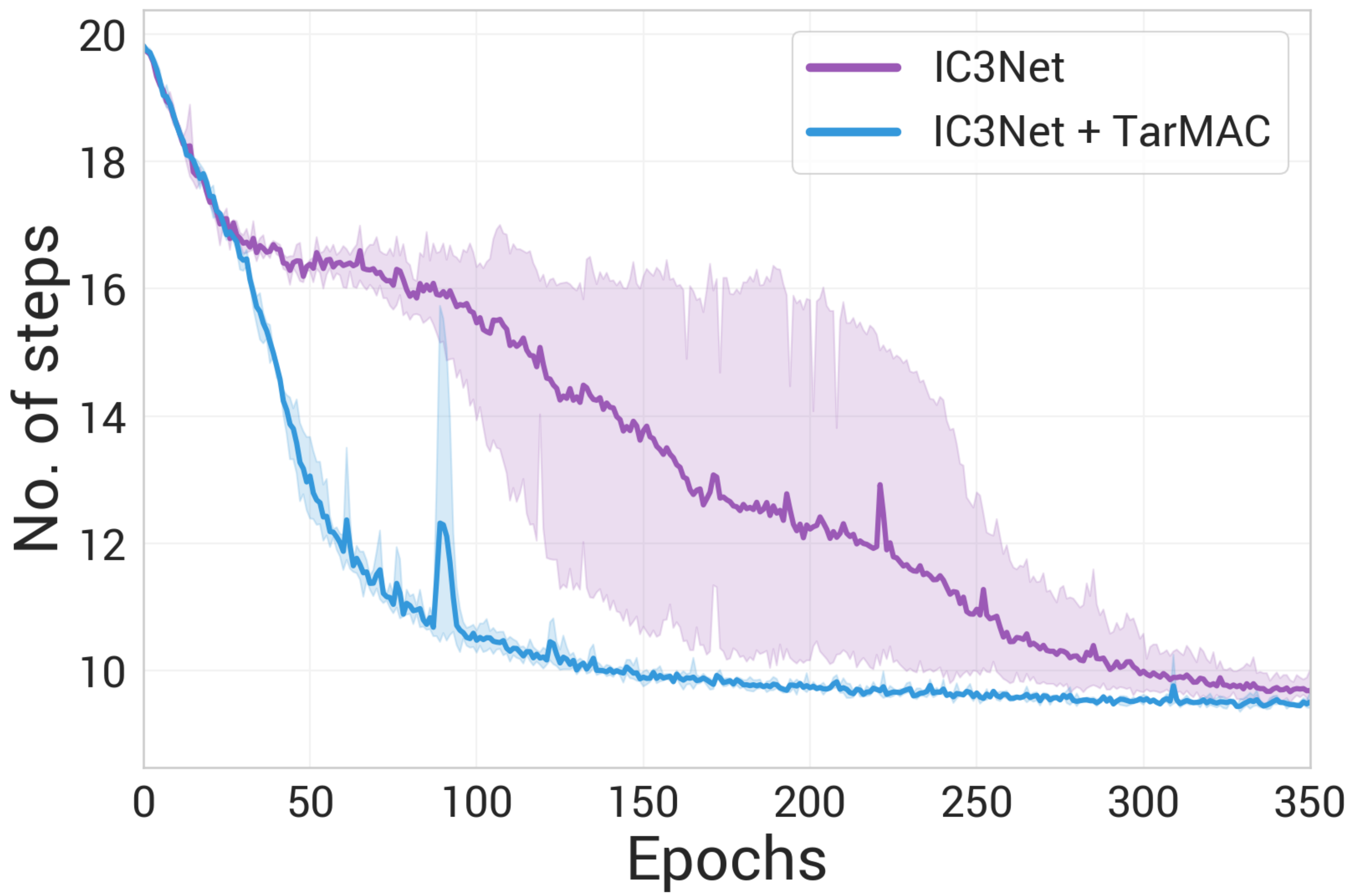
(a) 3 agents, 5 × 5 grid, vision=0, max steps=20



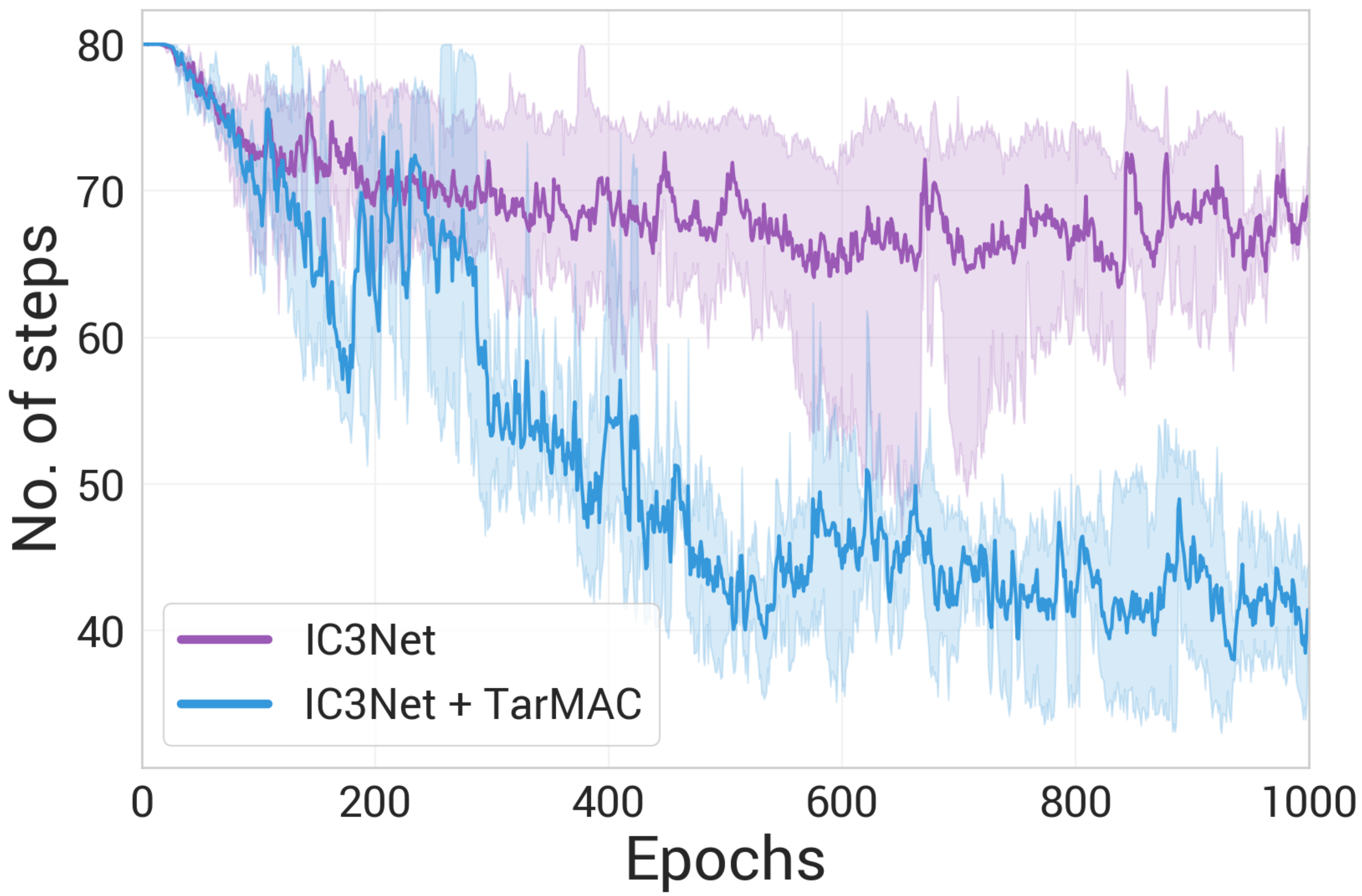
(b) 10 agents, 20 × 20 grid, vision=1, max steps=80

No. of steps taken by predators to reach prey as training progresses.

Results: Extension to competitive tasks — Predator-Prey



(a) 3 agents, 5 × 5 grid, vision=0, max steps=20



(b) 10 agents, 20 × 20 grid, vision=1, max steps=80

No. of steps taken by predators to reach prey as training progresses.

When combined with prior approaches for competitive environments, TarMAC leads to better sample efficiencies

Summary

- We introduced TarMAC, a simple architecture that allows for targeted multi-agent communication.

Summary

- We introduced TarMAC, a simple architecture that allows for targeted multi-agent communication.

- Targeting is learnt via sender-receiver soft attention.

Summary

- We introduced TarMAC, a simple architecture that allows for targeted multi-agent communication.
- Targeting is learnt via sender-receiver soft attention.
- Evaluation on multiple environments shows that TarMAC leads to intuitive communication attention and better performance.

Summary

- We introduced TarMAC, a simple architecture that allows for targeted multi-agent communication.

- Targeting is learnt via sender-receiver soft attention.

- Evaluation on multiple environments shows that TarMAC leads to intuitive communication attention and better performance.

- For more details, come to our poster
  - Pacific Ballroom #57 (06:30 PM to 09:00 PM)