Incentivising Pragmatism

Advanced methods for multi-agent communication

Mortimer von Chappuis

11 October 2023

Setup

Environment

- Self play
- partial observability
- fully cooperative

Symmetries

- Valuefunction
- Policy
- Belief update

Three levels of incentives

descriptive

$$argmax_m I(m^{self}, b^{self})$$

epistemic

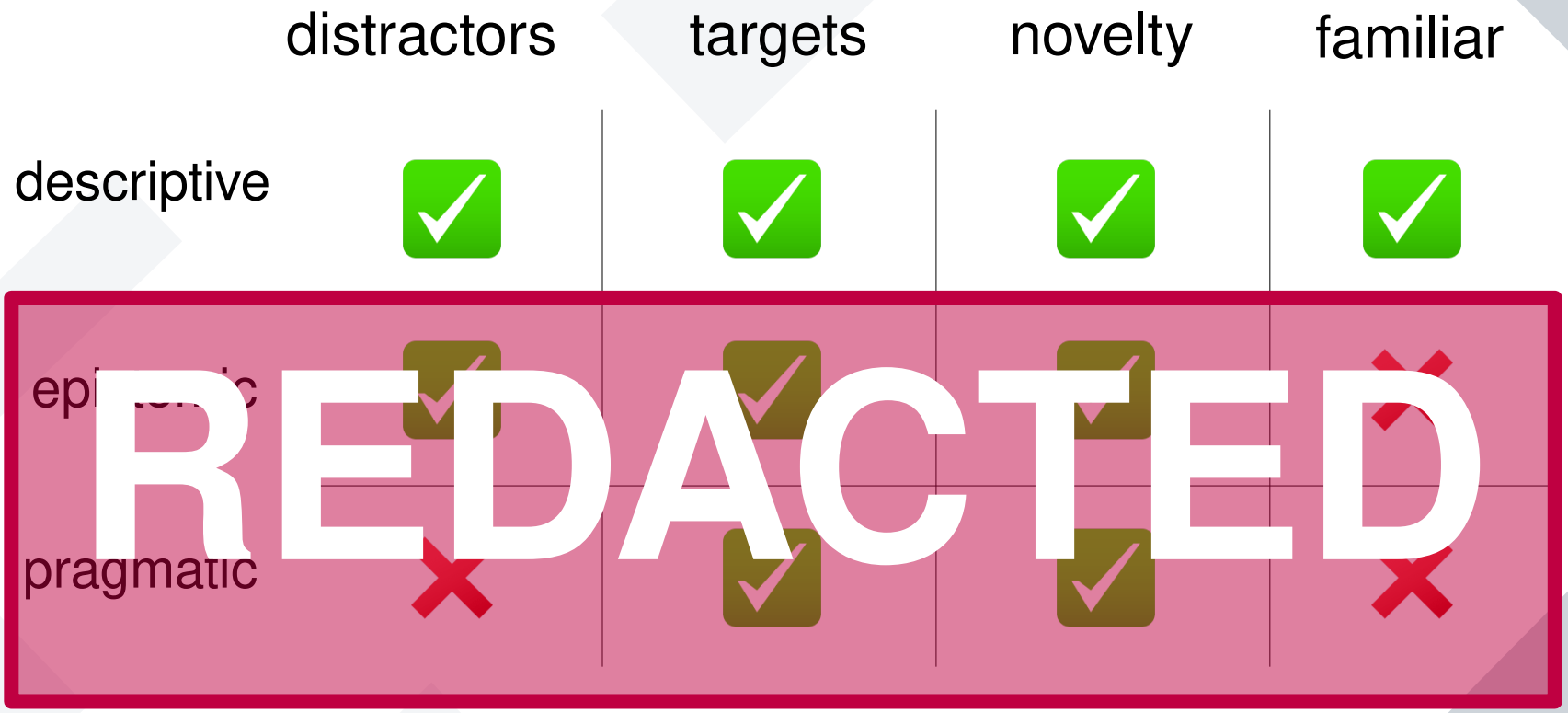
$$argmax_m I(m^{self}, b^{other})$$

pragmatic

$$argmax_m Q(b^{other}, u^{other} | m^{self}) - Q(b^{other}, u^{other})$$

reward shaping $r_t + \alpha \hat{r}_t - \beta |m_t|$

Hypothesis


distractors
targets
novelty
familiar
descriptive
epistemic
pragmatic
REDACTED

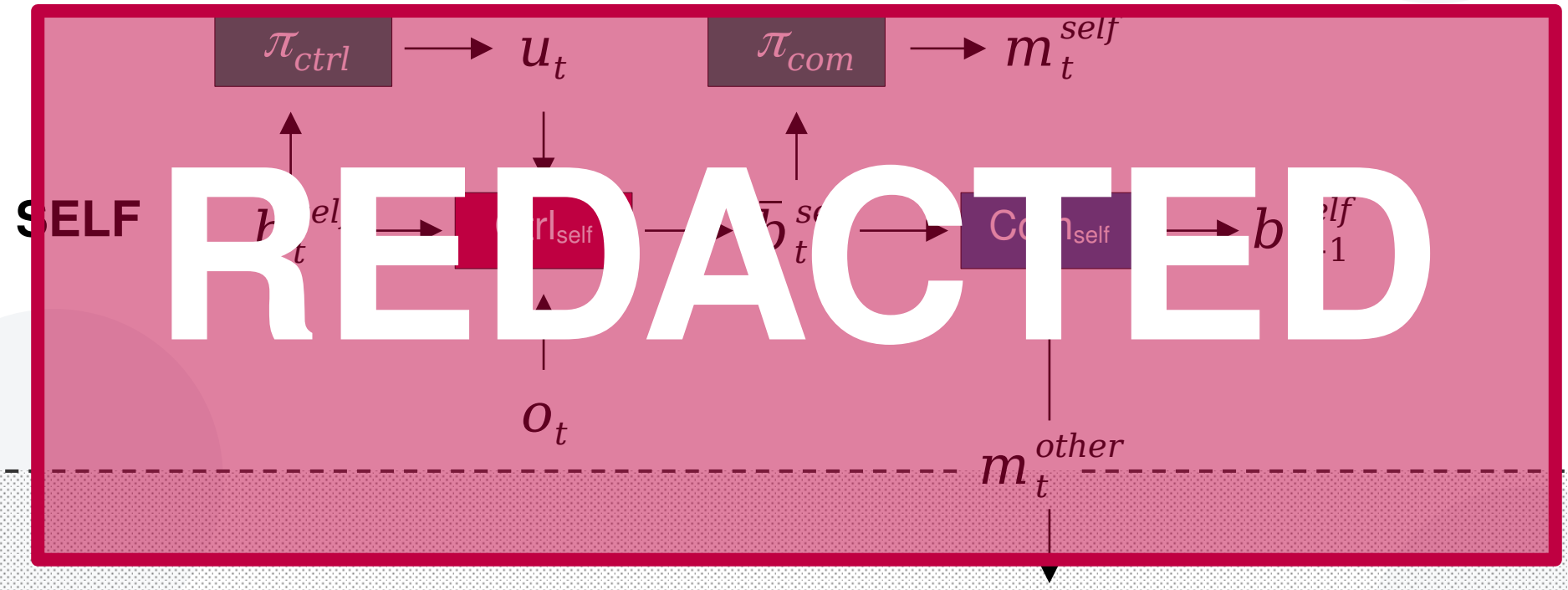
Mutual Information

$$I(b,m)=\sum_{b\in B,m\in M} p(b,m)[\log(p(m|b))-\log(p(m))]$$

$\pi(m|b)$ policy

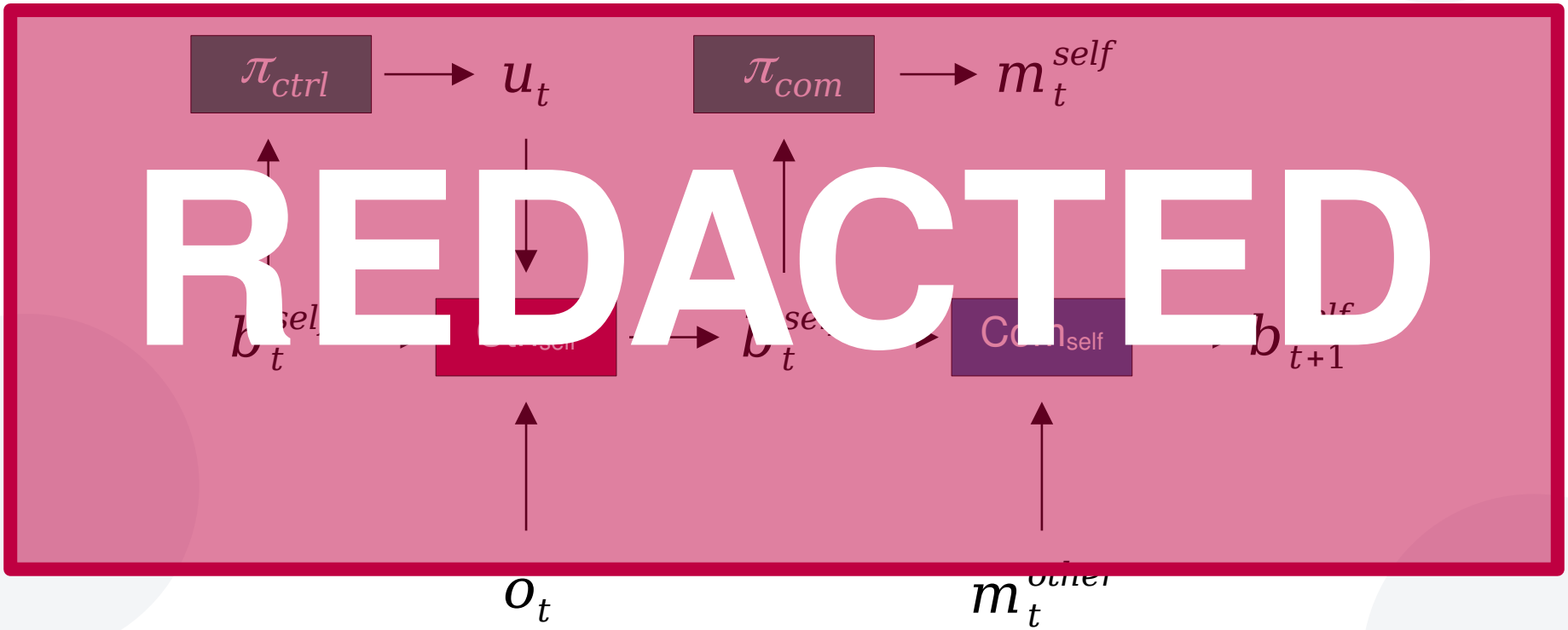
$f(m)$ language model

Architecture



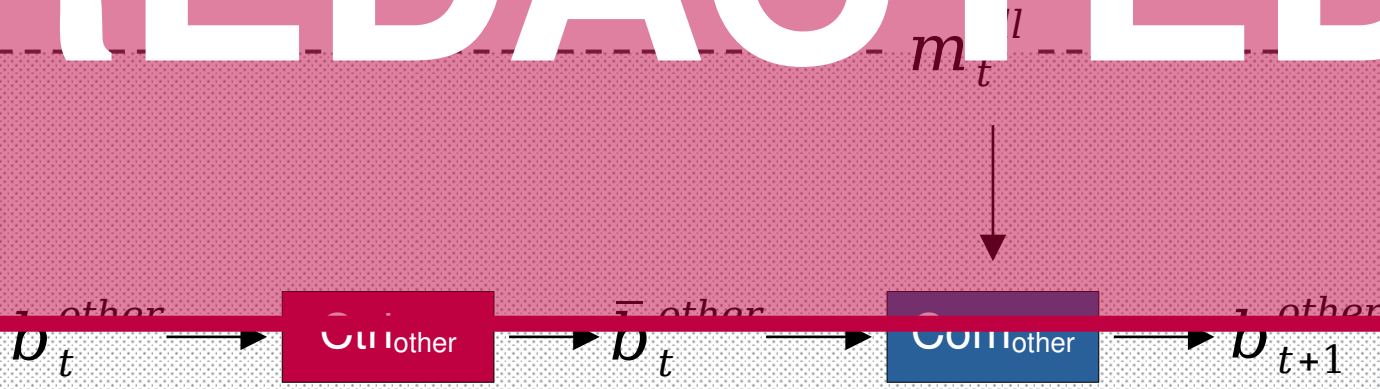
REDACTED

Architecture — Self Belief

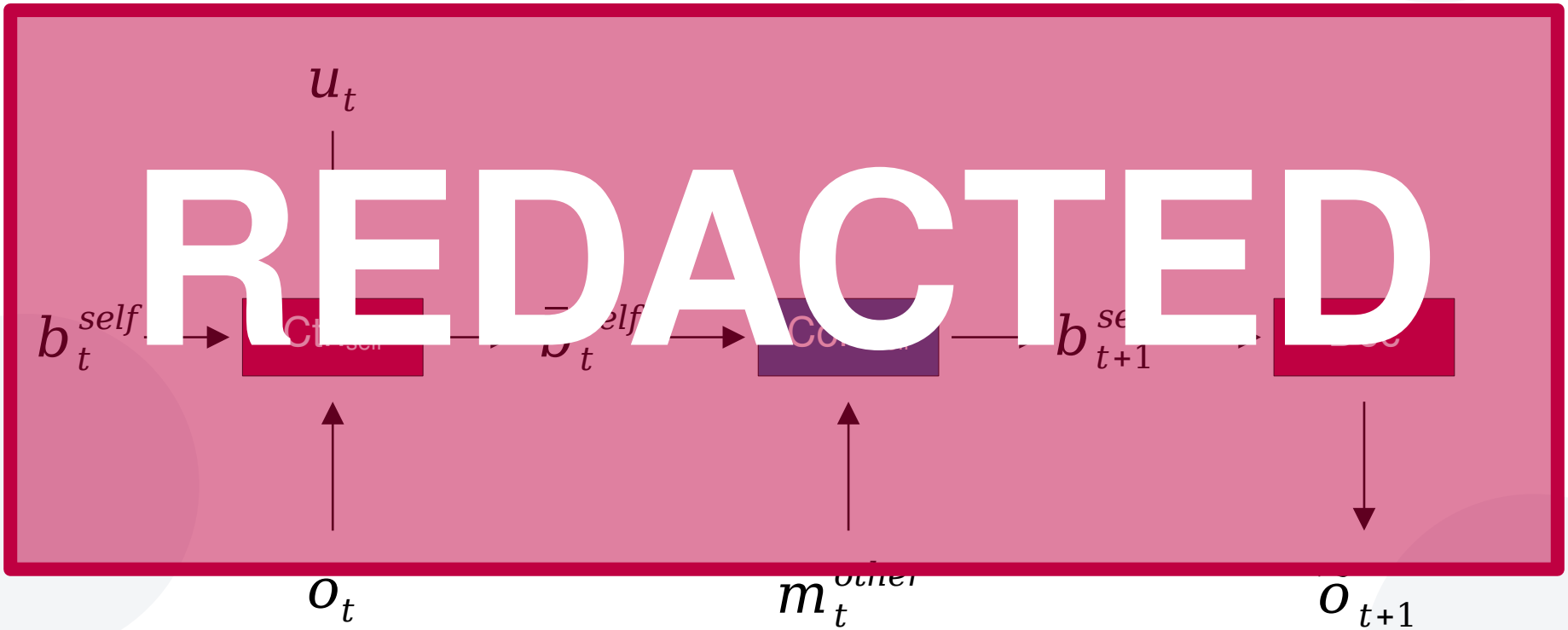


Architecture — Other Belief

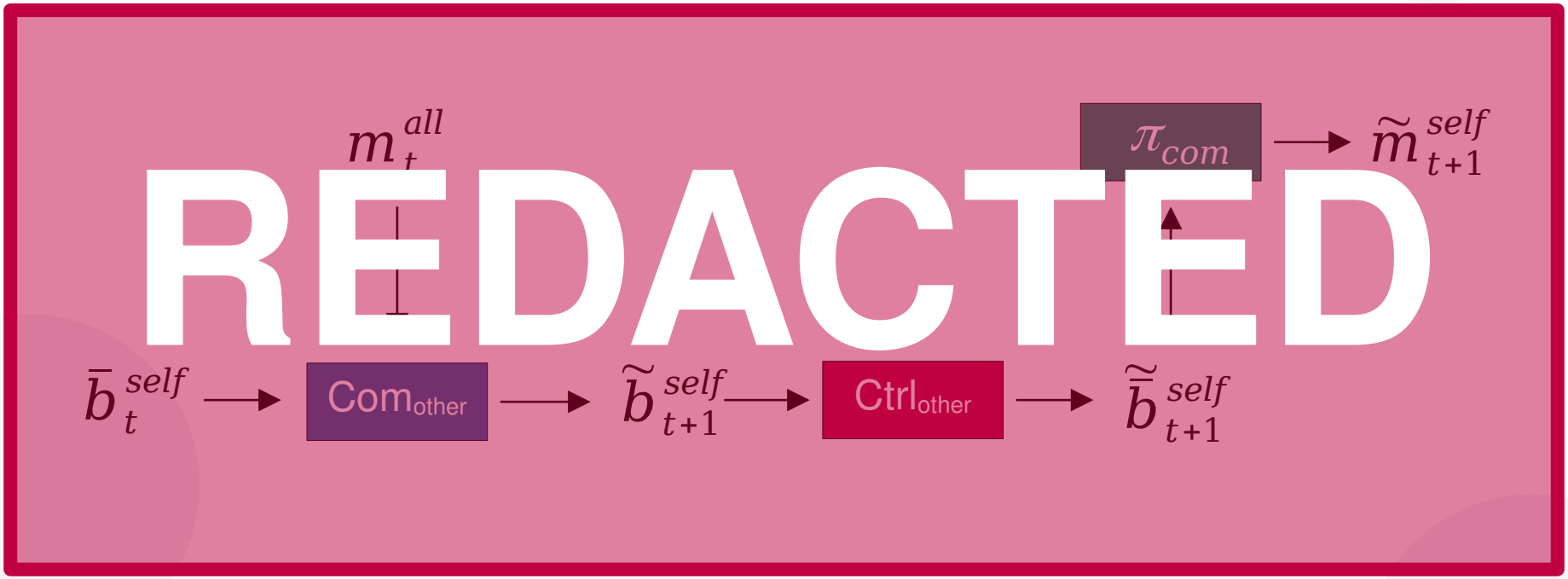
REDACTED



Training — Self



Training — Other



Summary

Networks

- actor policy (π_{com}, π_{ctrl})
- critic (Q function or Advantage)
- language model
- ctrl (self, other)
- com (self, other)
- dec

Training

- RL (actor, critic)
- language model
- belief self ($Ctrl_{self}$, Com_{self}, Dec)
- belief other ($Ctrl_{other}$, Com_{other})

Thanks for your attention! — Questions?

Influences

- Cheap Talk Discovery and Utilization in Multi-Agent Reinforcemenr Learning
- Learning Attentional Communication for Multi-Agent Cooperation
- Mastering Diverse Domains through World Models