INTERNSHIP PROGRESS

Modeling curriculum learning

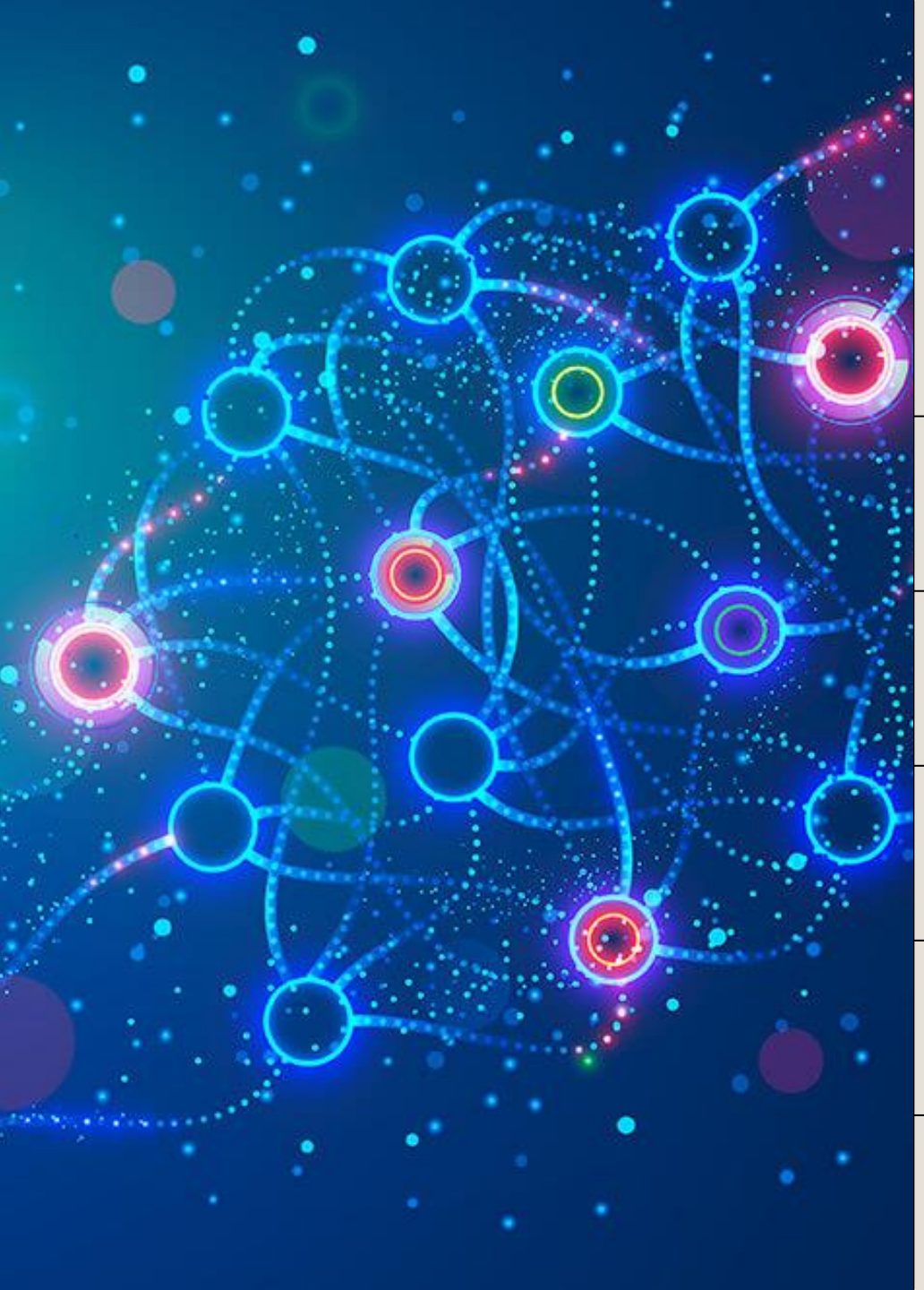


TABLE OF CONTENTS

WEEKLY OBJECTIVES

	FEBRUARY	MARCH	APRIL	MAY
W1	/	Level 2: accuracy RL Level 2: learning progress RL	Level 3: integrate all LVL2 and manually set weights	Finishing touches Written report
W2	/	Unify level 1 Compare different versions of the model	Level 3: create the policy gradient model	Written report
W3	Level 1: implement tasks and neural network	Prepare the presentation	Level 3: explore more options (chaining effect? Between-task learning?)	(exams)
W4	Finish 3 models for Level 1 Level 2: accuracy RL learning	**LAB PRESENTATION** Integrate all suggestions	Level 3: adjust and compare model performances	(exams)

Past week

Upcoming week

LAST WEEK'S OBJECTIVES

1 IMPLEMENT OFFLINE LEARNING

For LP signed/unsigned

COMPARE DIFFERENT WINDOWS

For LP signed/unsigned

2

CREATE A SIMPLE P.OLICY GRADIENT

3

LOOK INTO POLICY GRADIENTS

Read the article

4

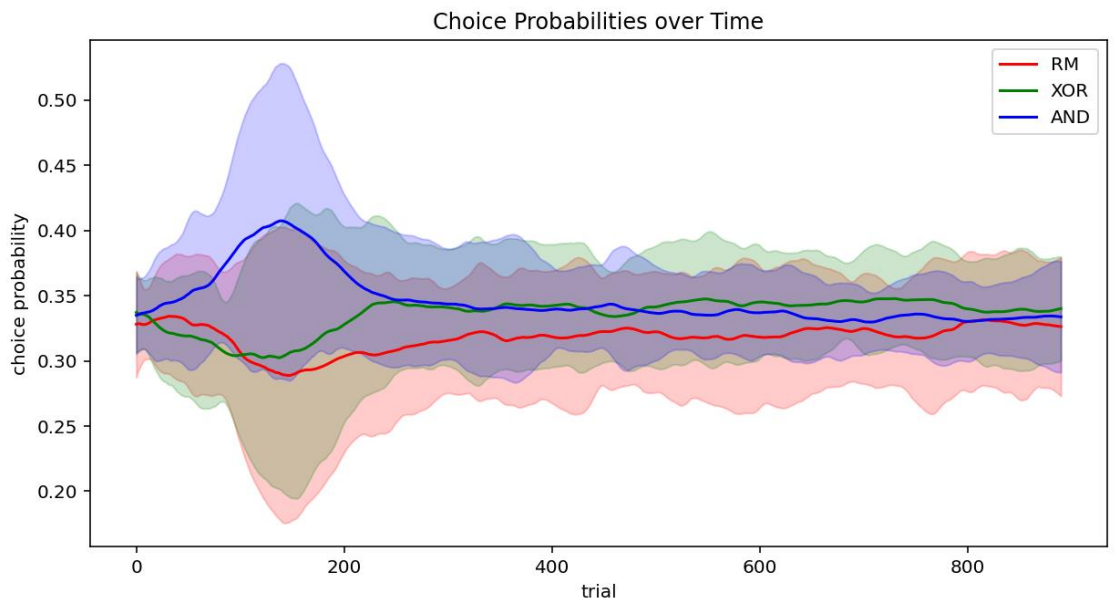
SUMMARISE FINDINGS

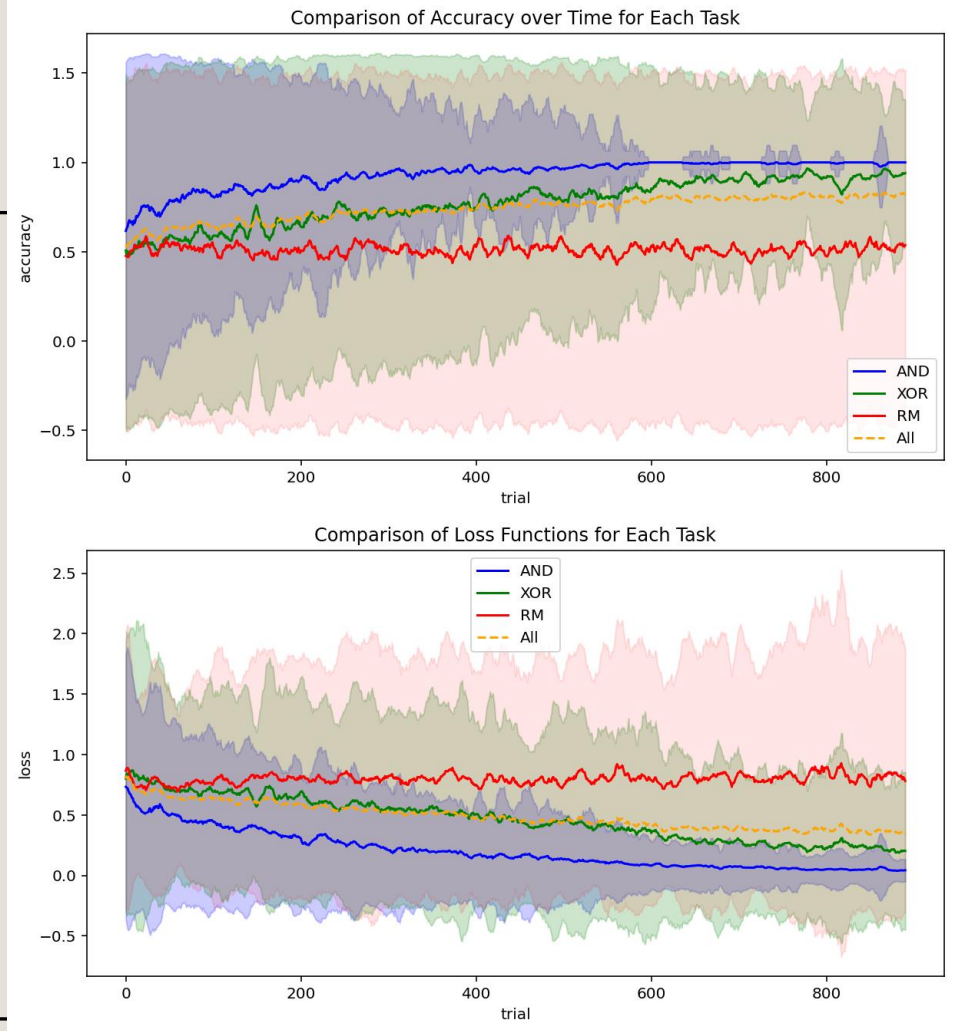
Write a short summary of findings, questions and plan for next week.

5

Comparing different window for signed LP

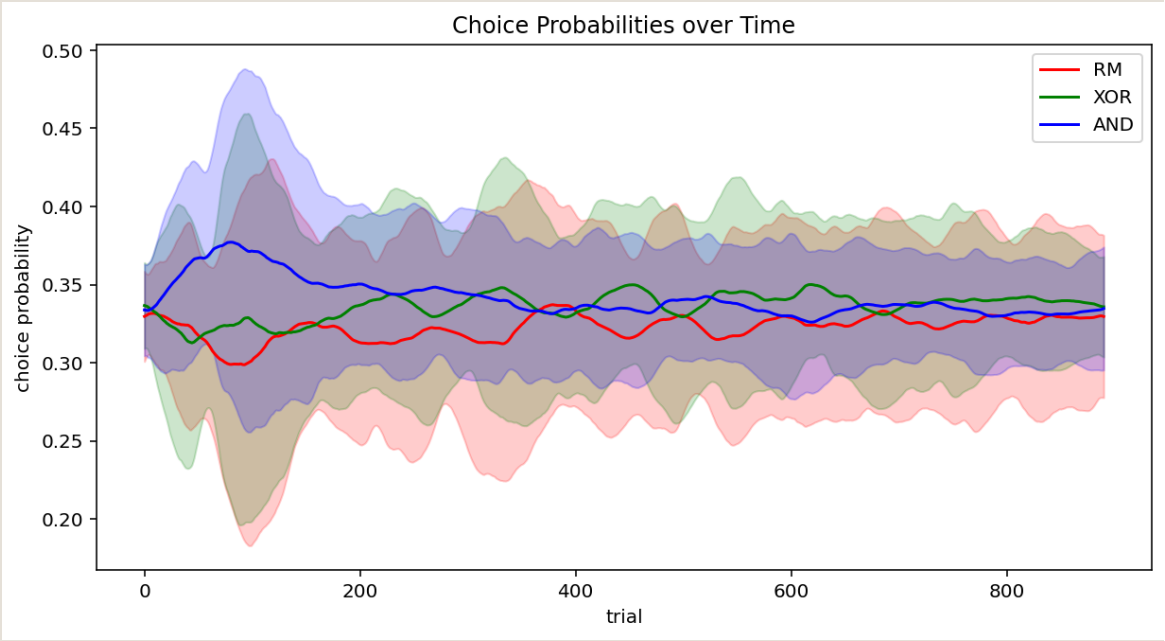
Signed LP model (window = 60)

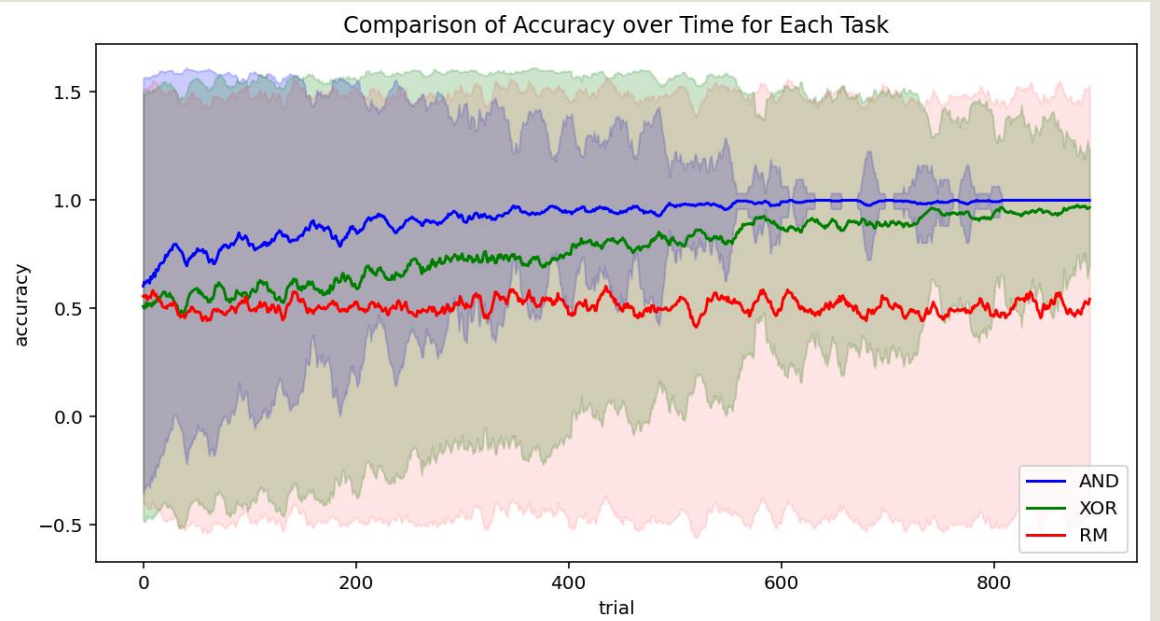


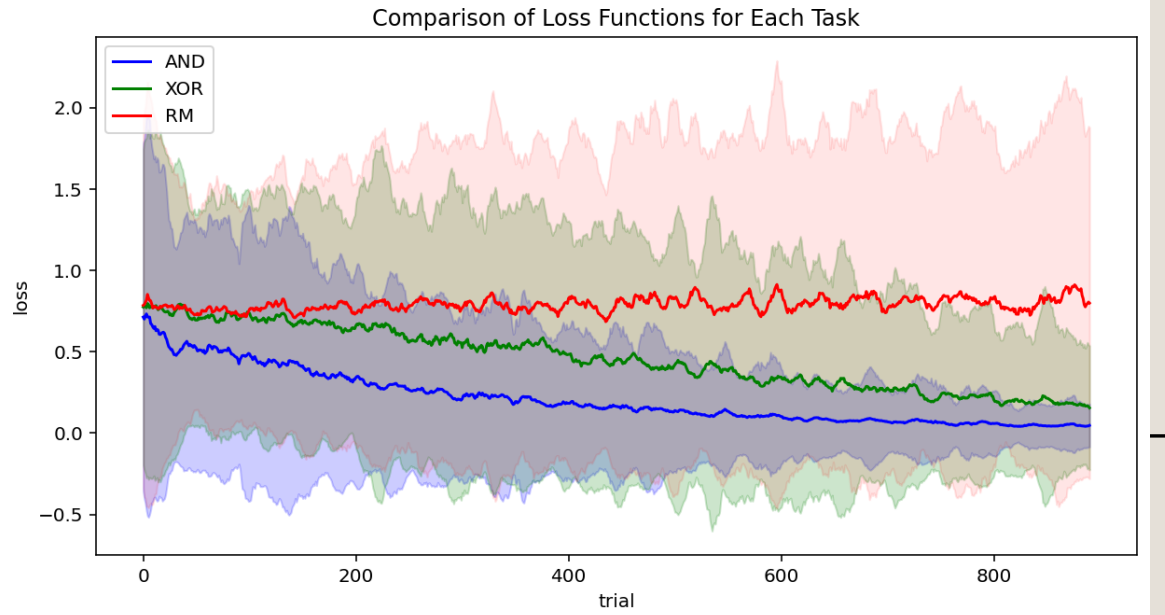


What the models look like

Signed LP model (window = 40)

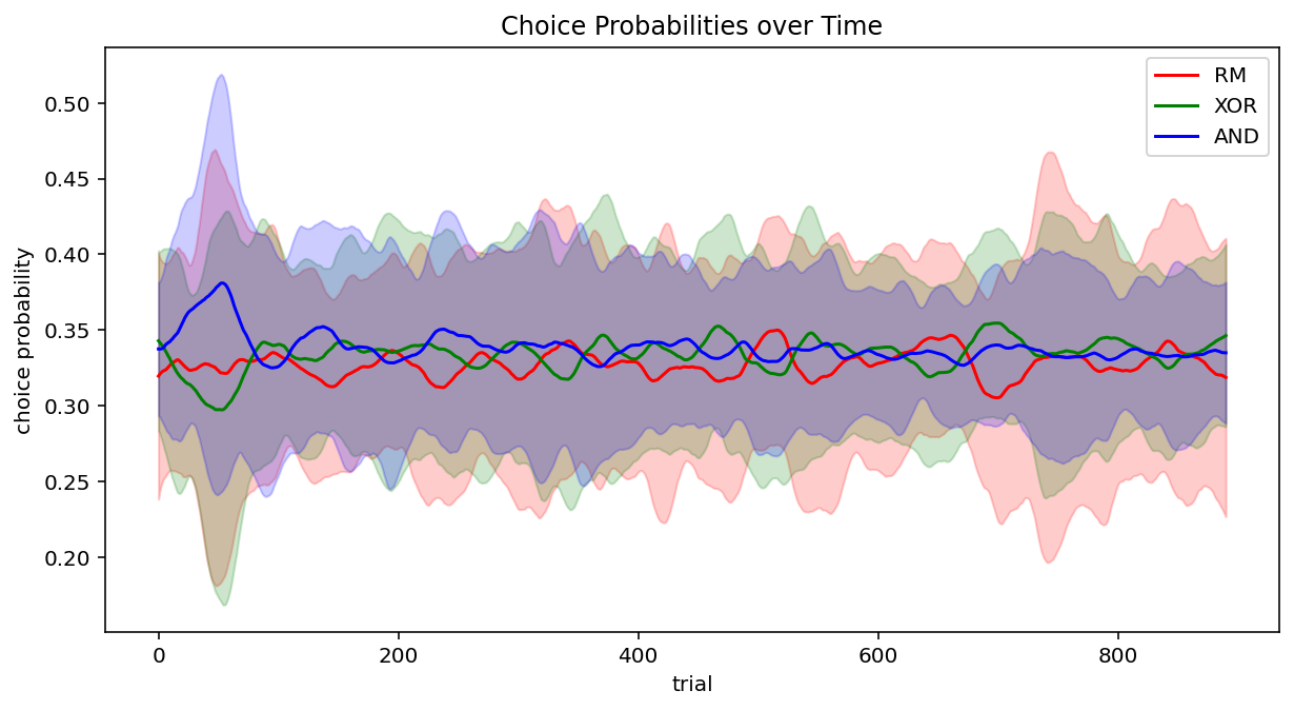


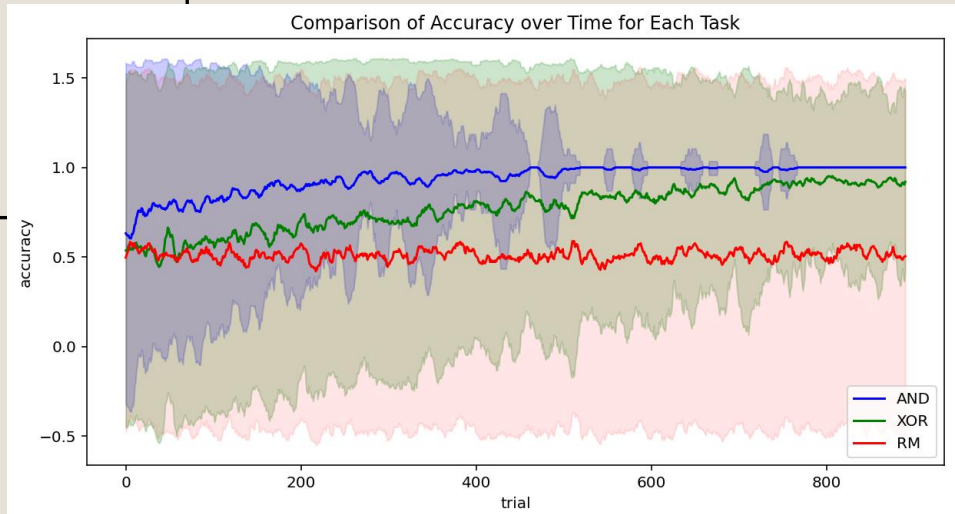


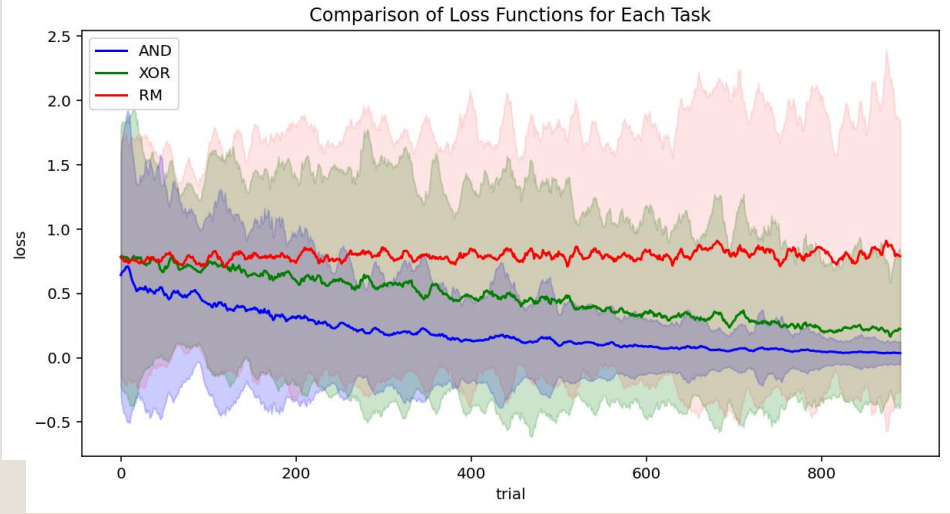


What the models look like

Signed LP model (window = 20)



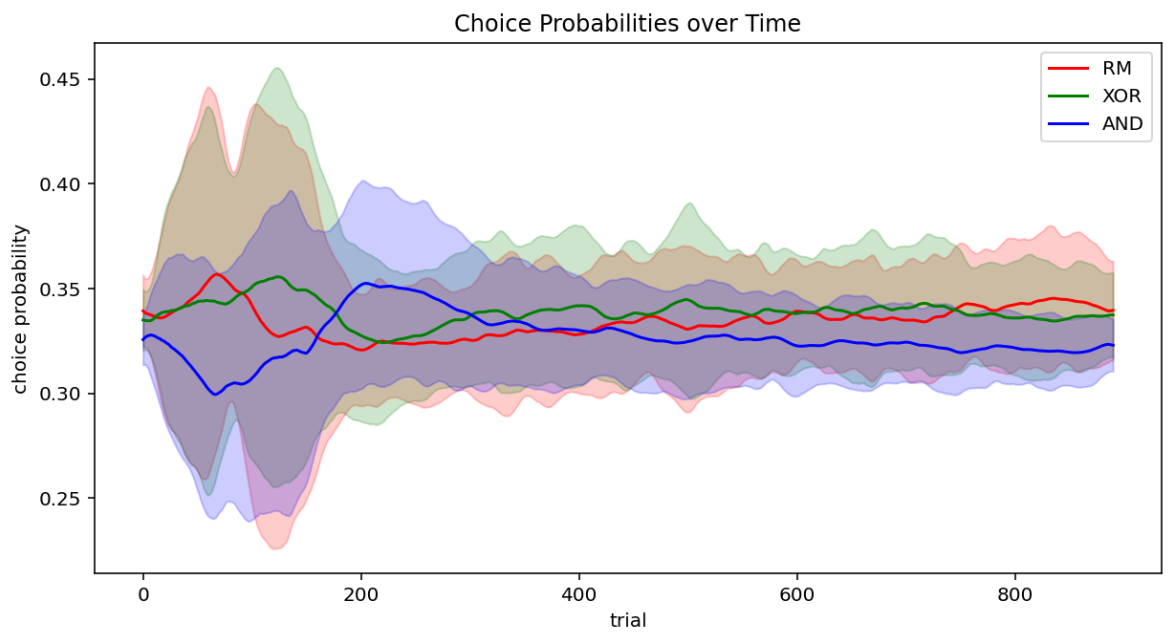


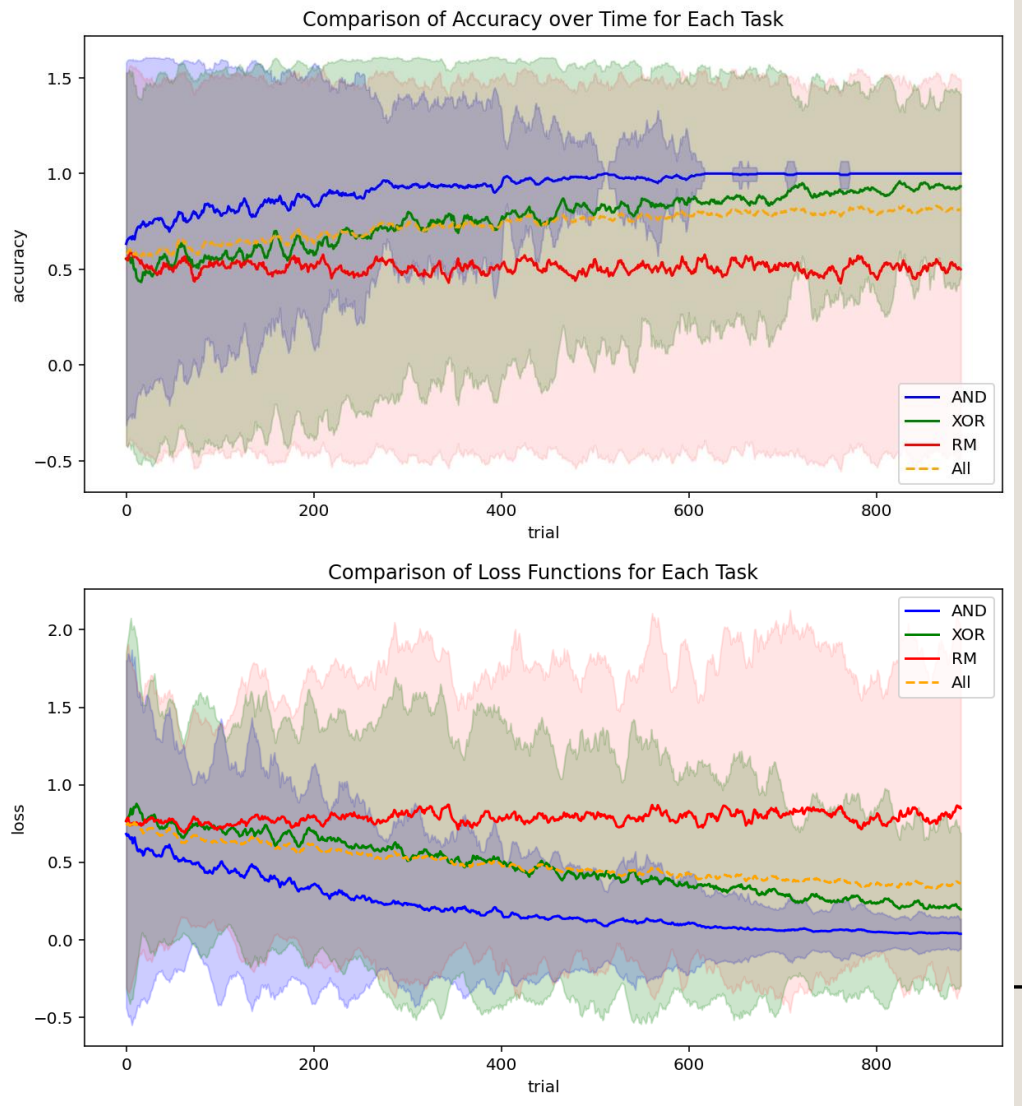


- The preference for the easy task appears earlier
- But also ends earlier (even before the task is learned)
- Still no clear preference for the hard task

Comparing different window for unsigned LP

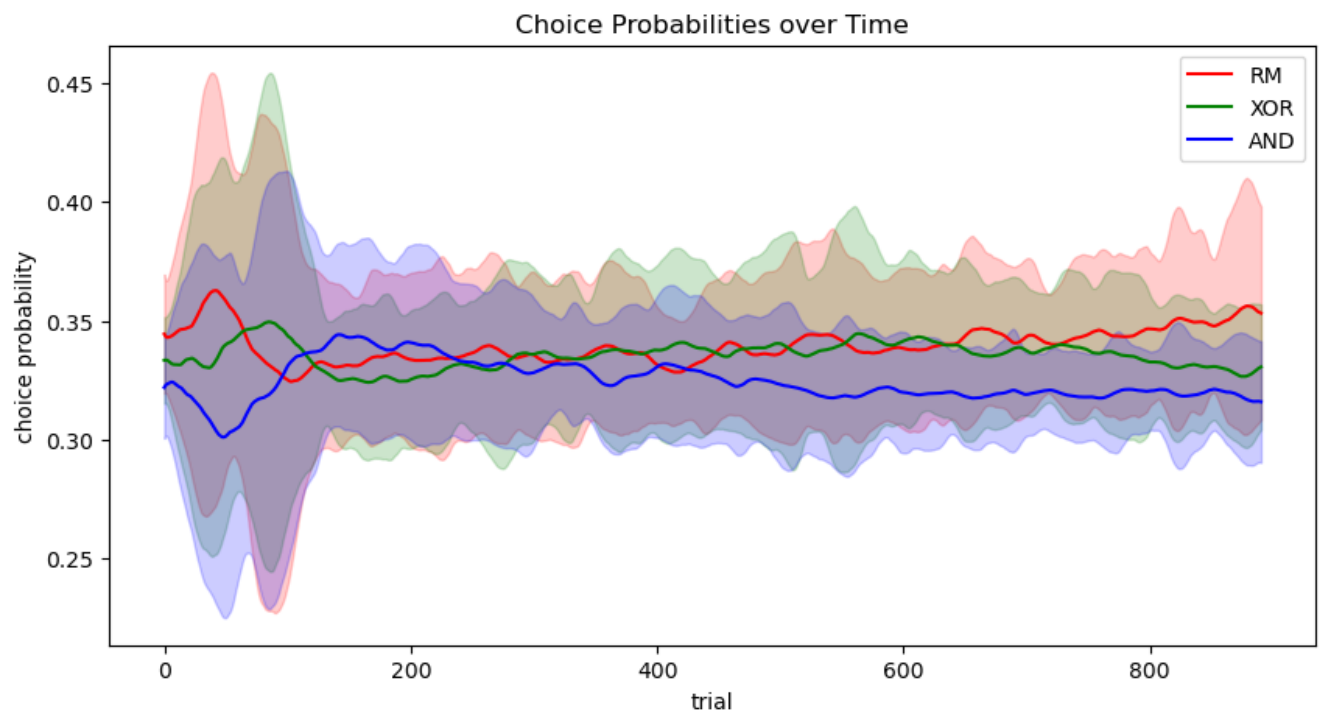
Unsigned LP model (window = 60)

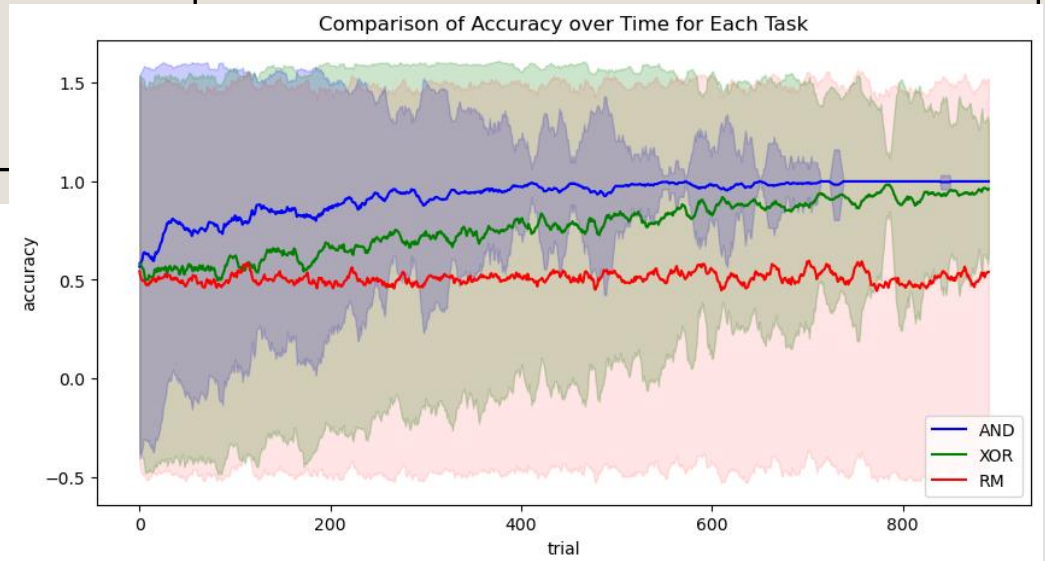


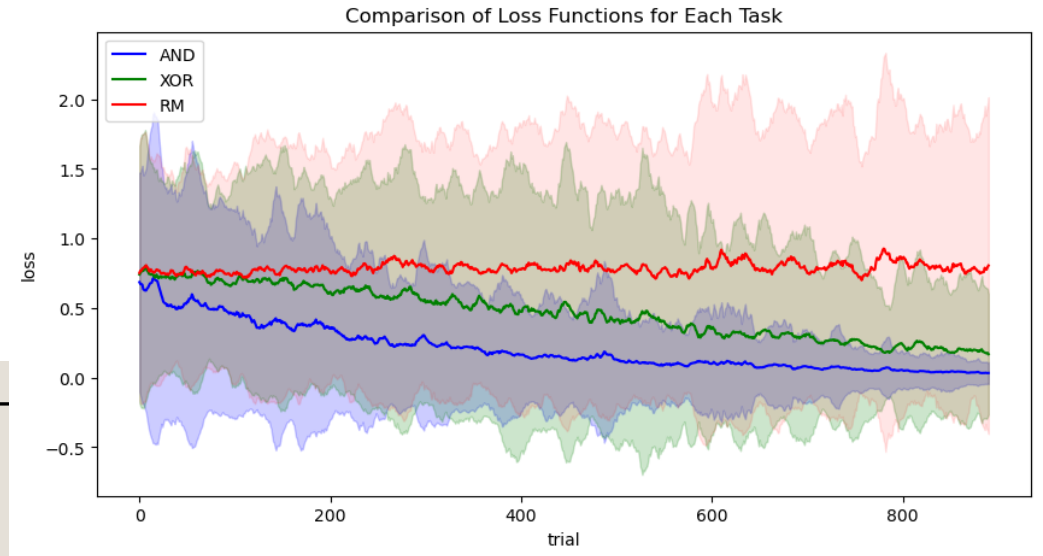


What the models look like

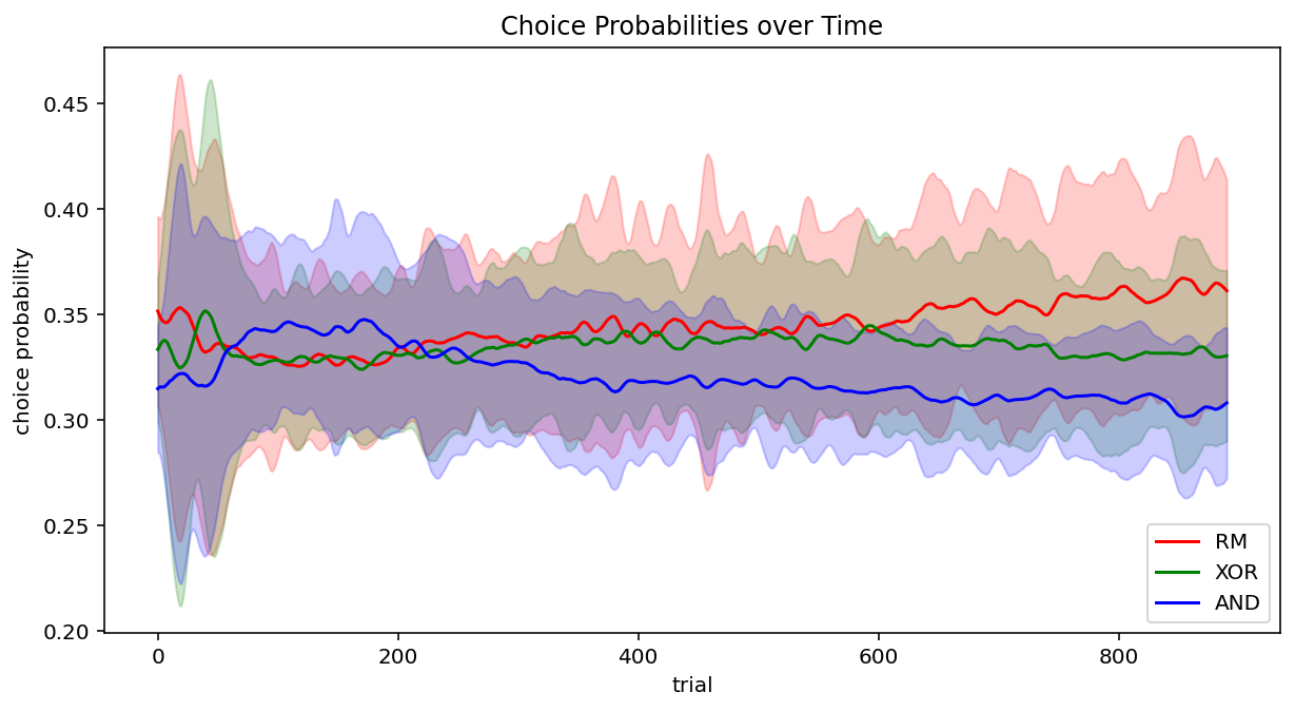
Unsigned LP model (window = 40)

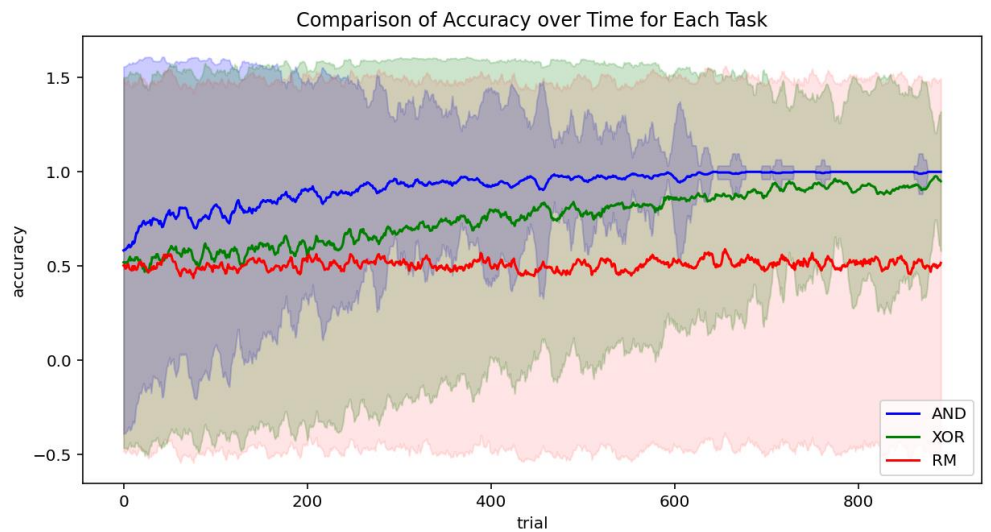


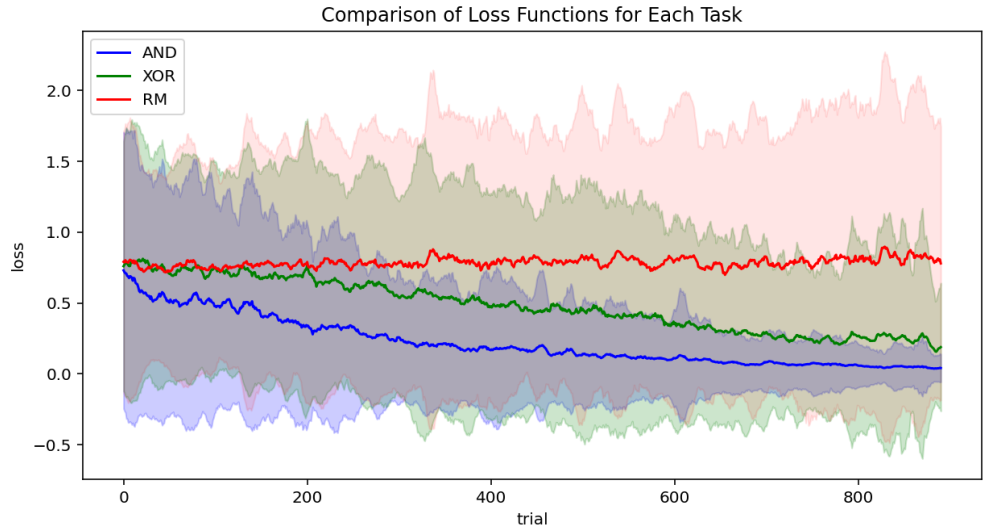




Unsigned LP model (window = 20)



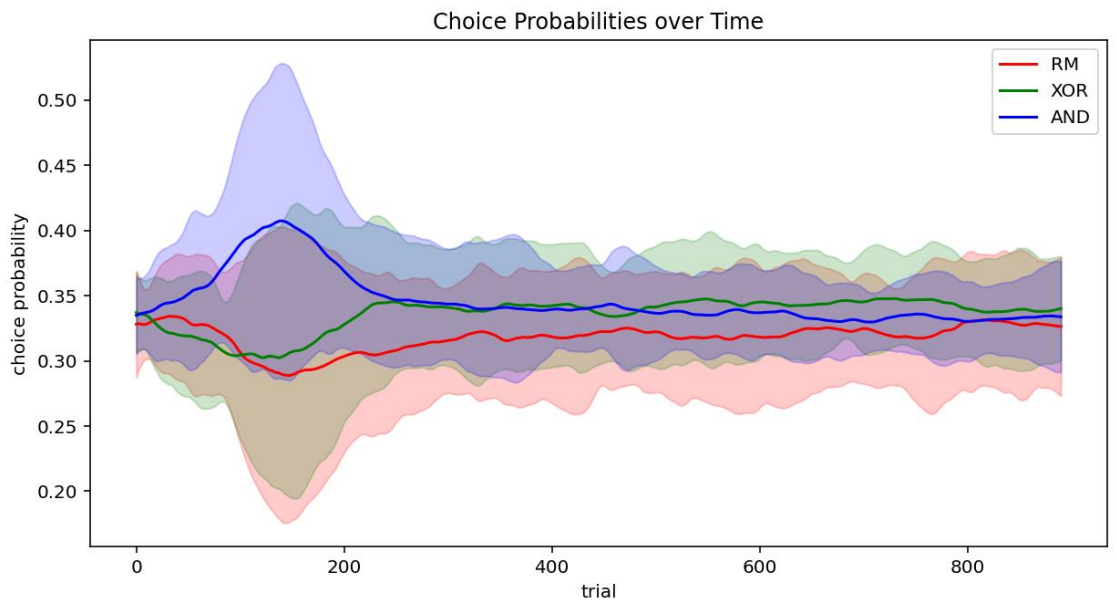


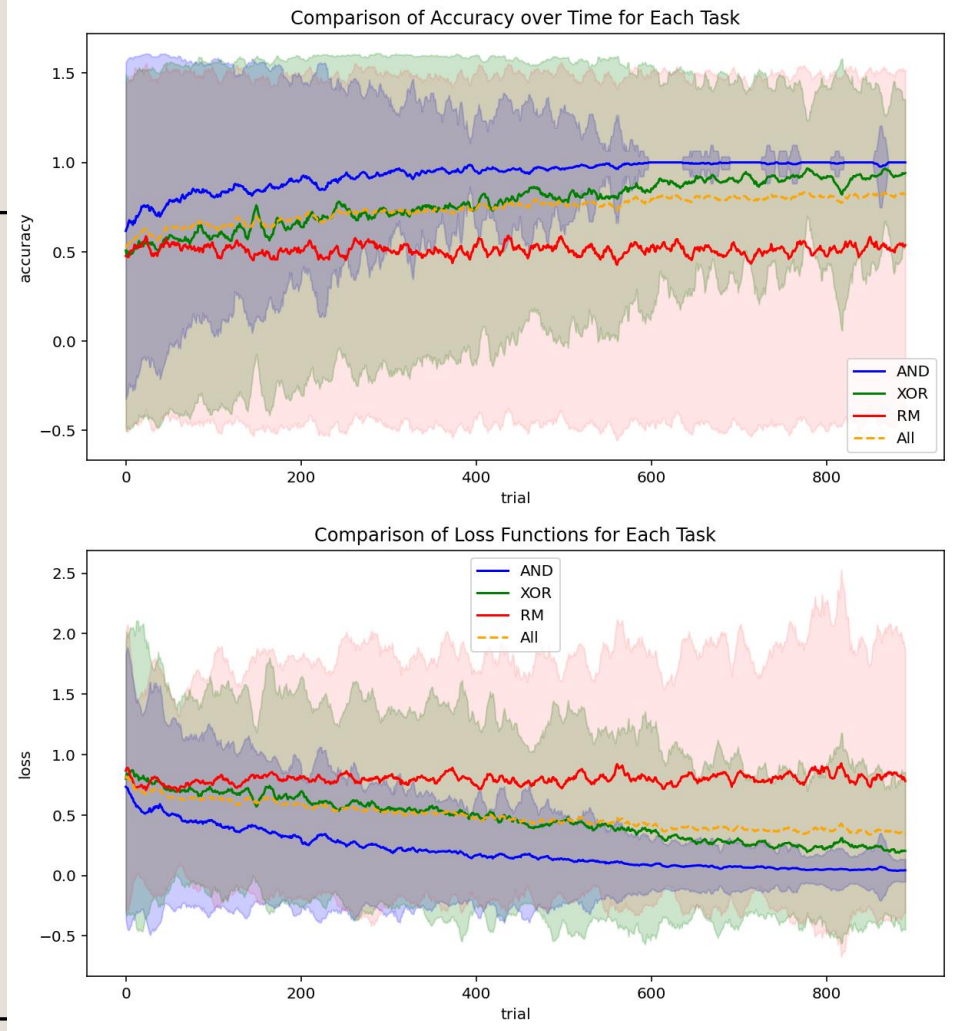


(note: no difference in performance)

Offline/online learning for un/signed LP

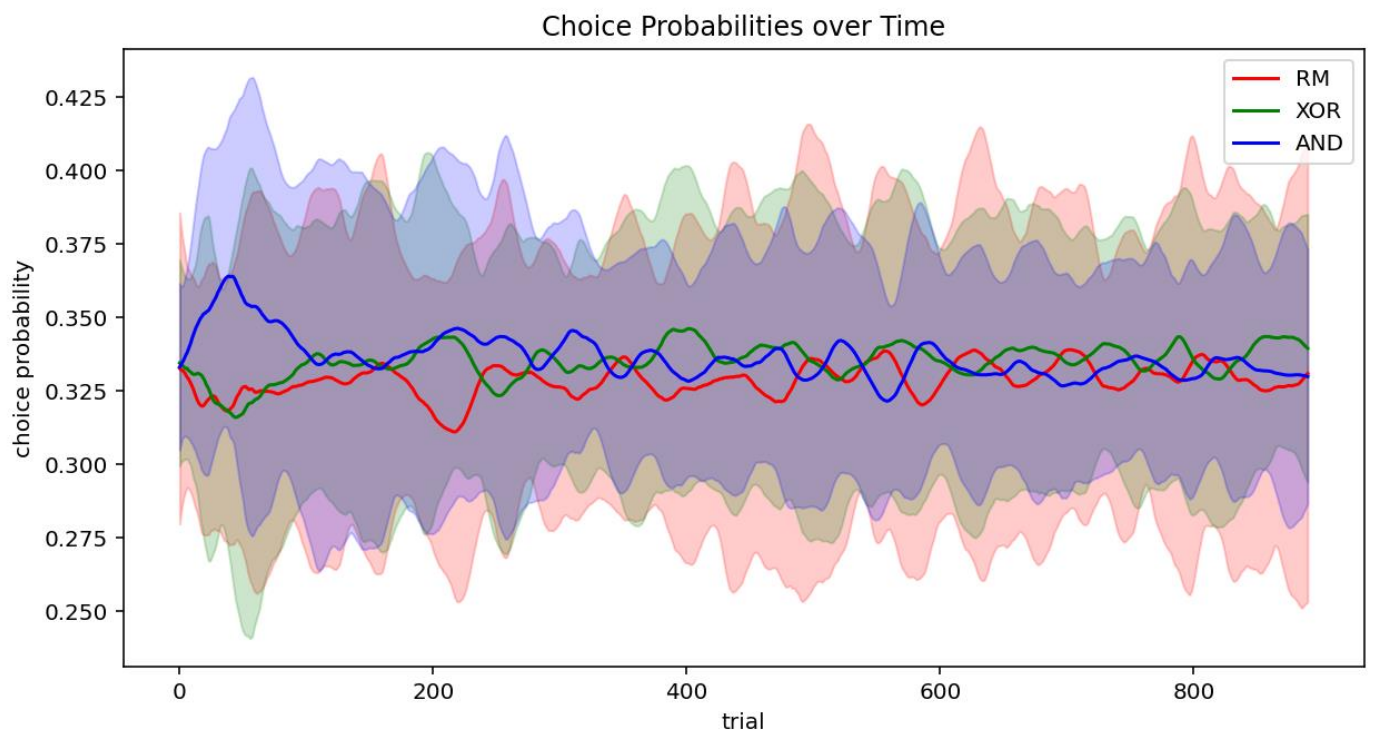
Signed LP model (window = 60, online learning)

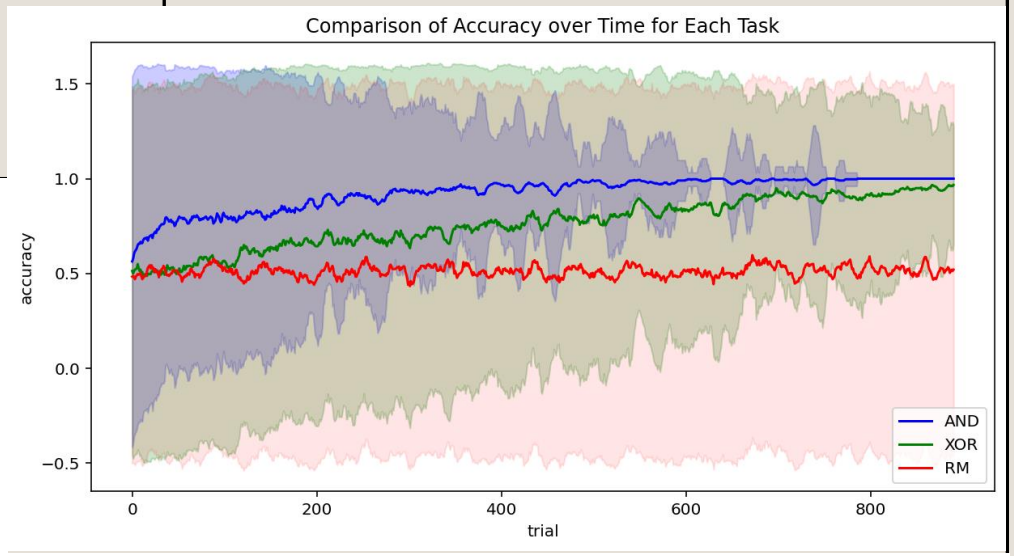


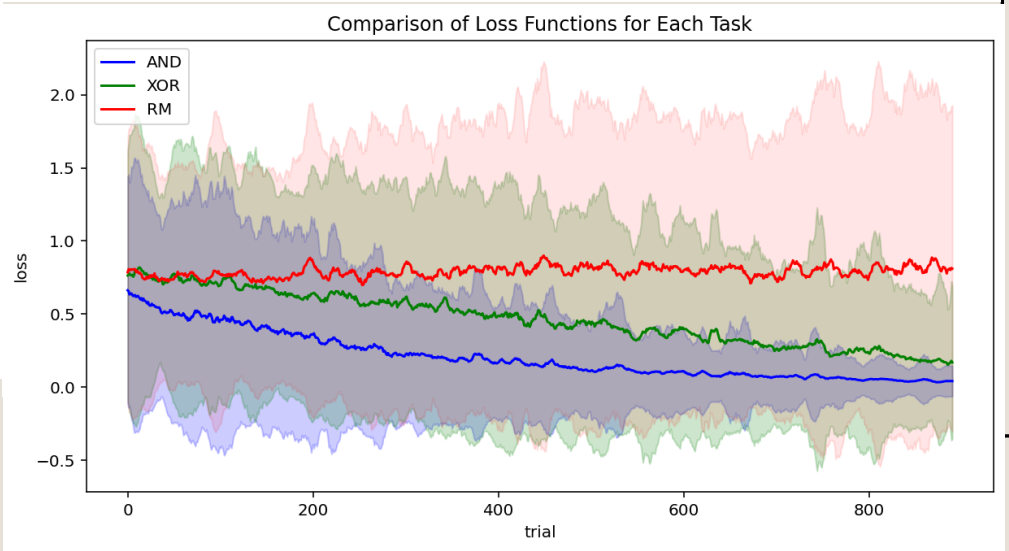


What the models look like

Signed LP model (window = 60, offline learning)

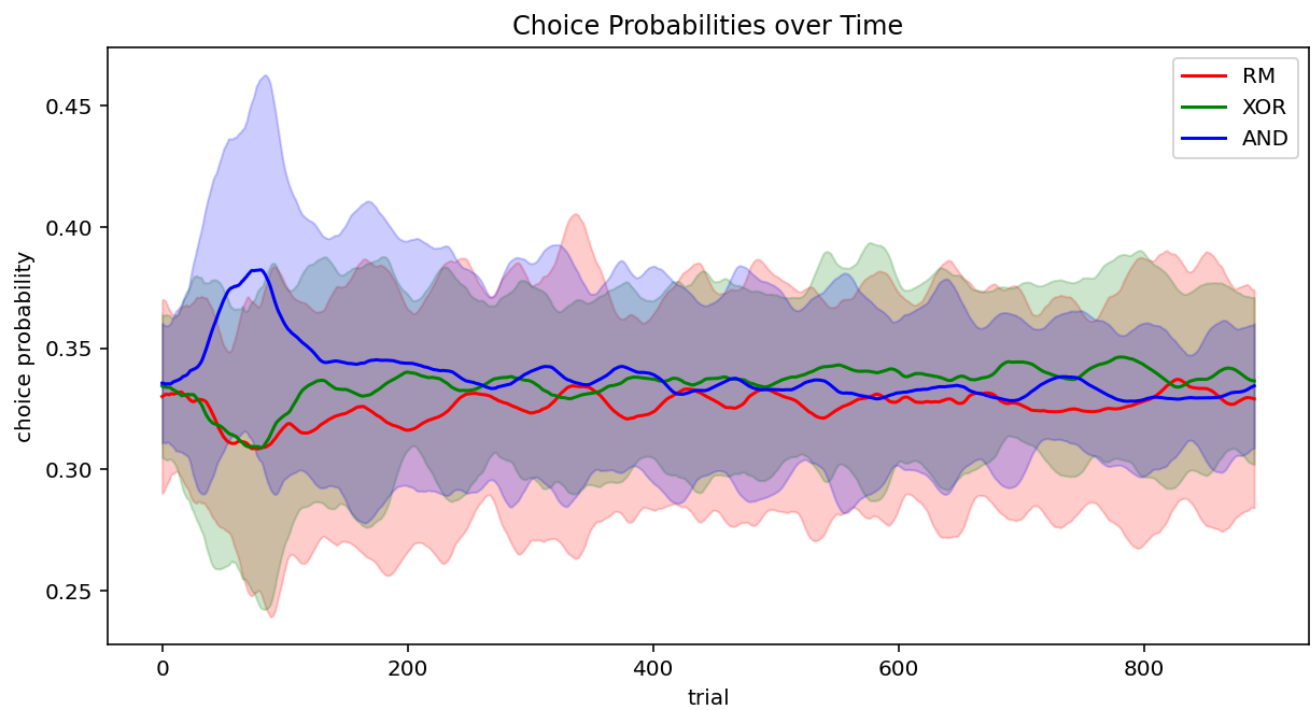


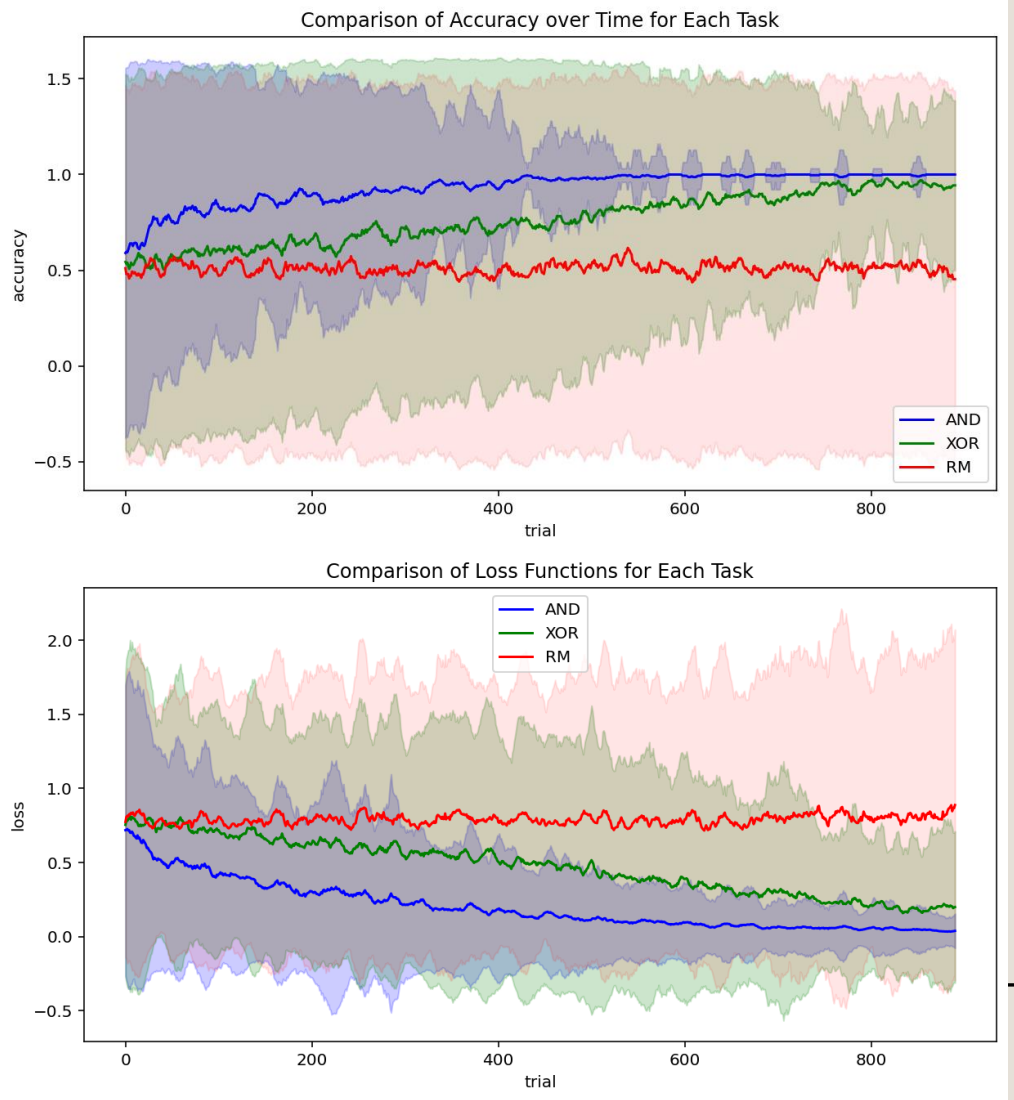




- Changes in weights are less likely to be relevant or significant for the task
- learning over a smaller time frame (by ~3).

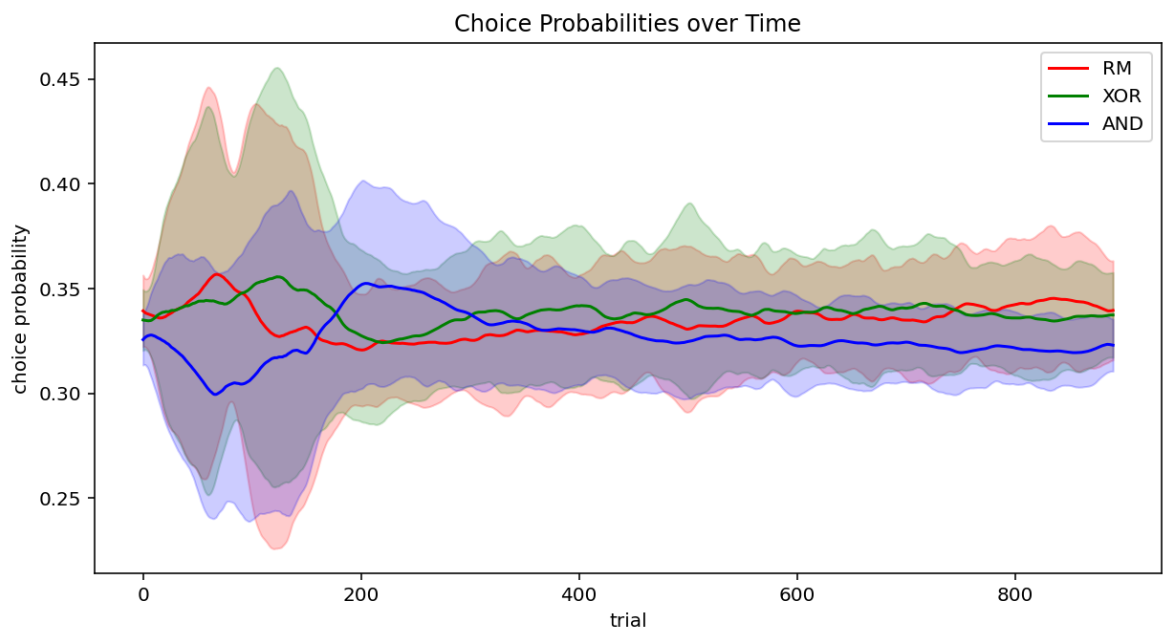
Signed LP model (window = 90, offline learning)

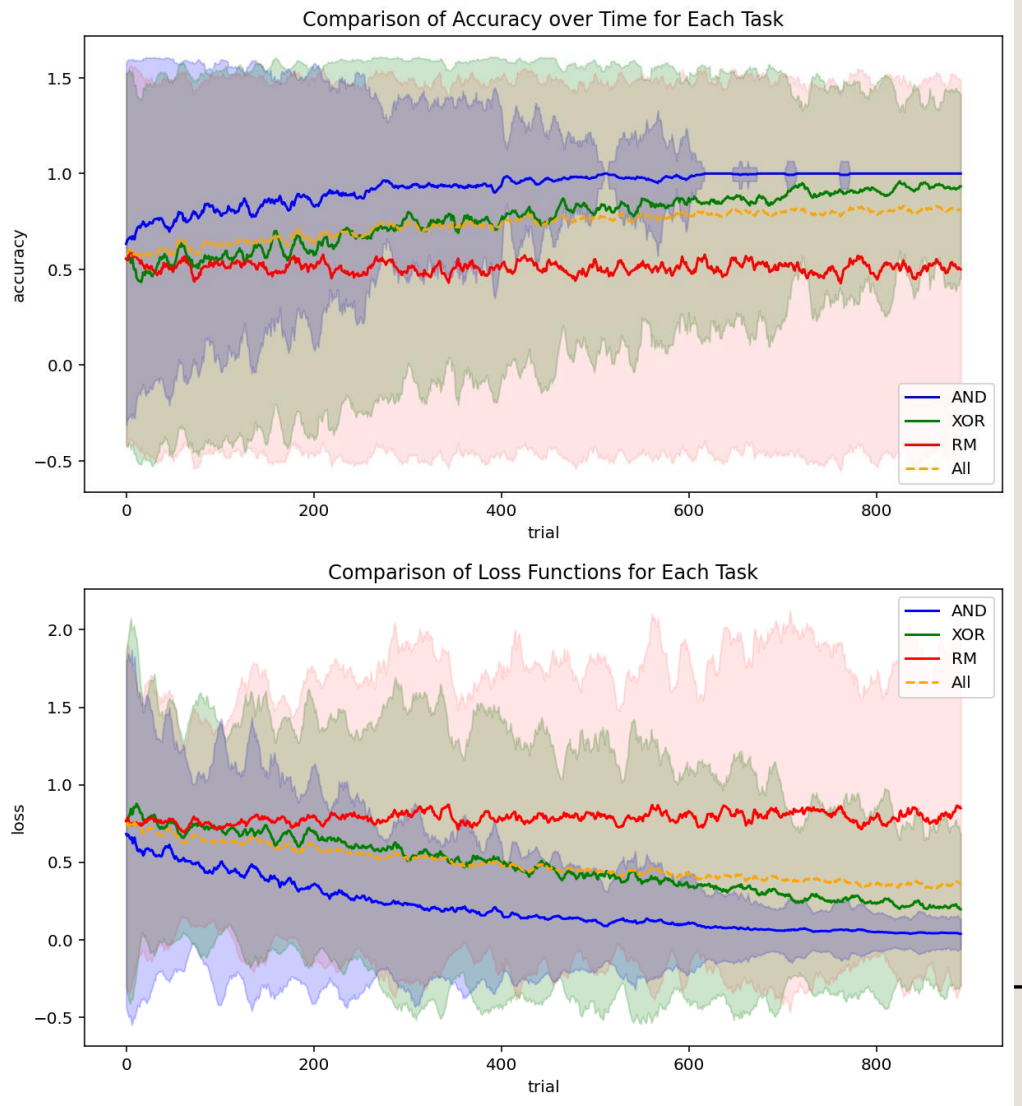




What the models look like

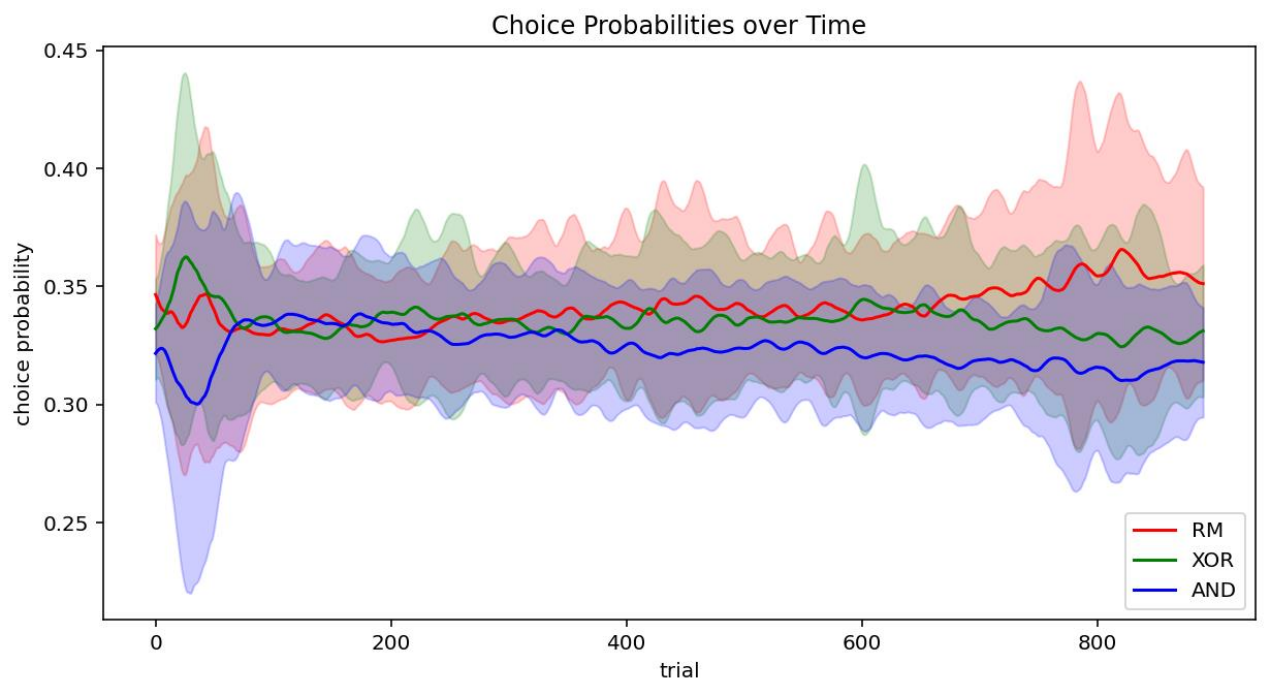
Unsigned LP model (window = 60, online learning)



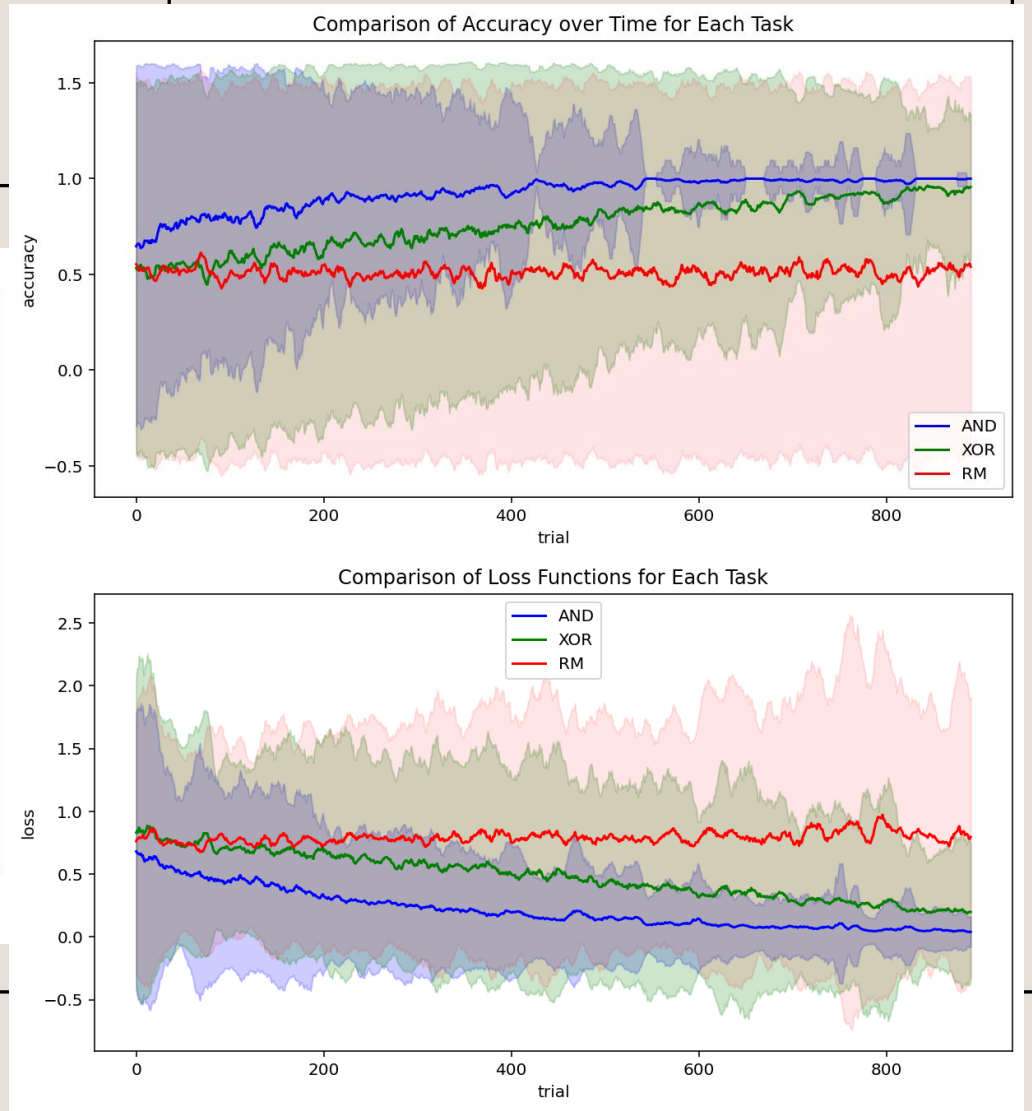


What the models look like

Unsigned LP model (window = 60, offline learning)



Note: no difference in performance with iid

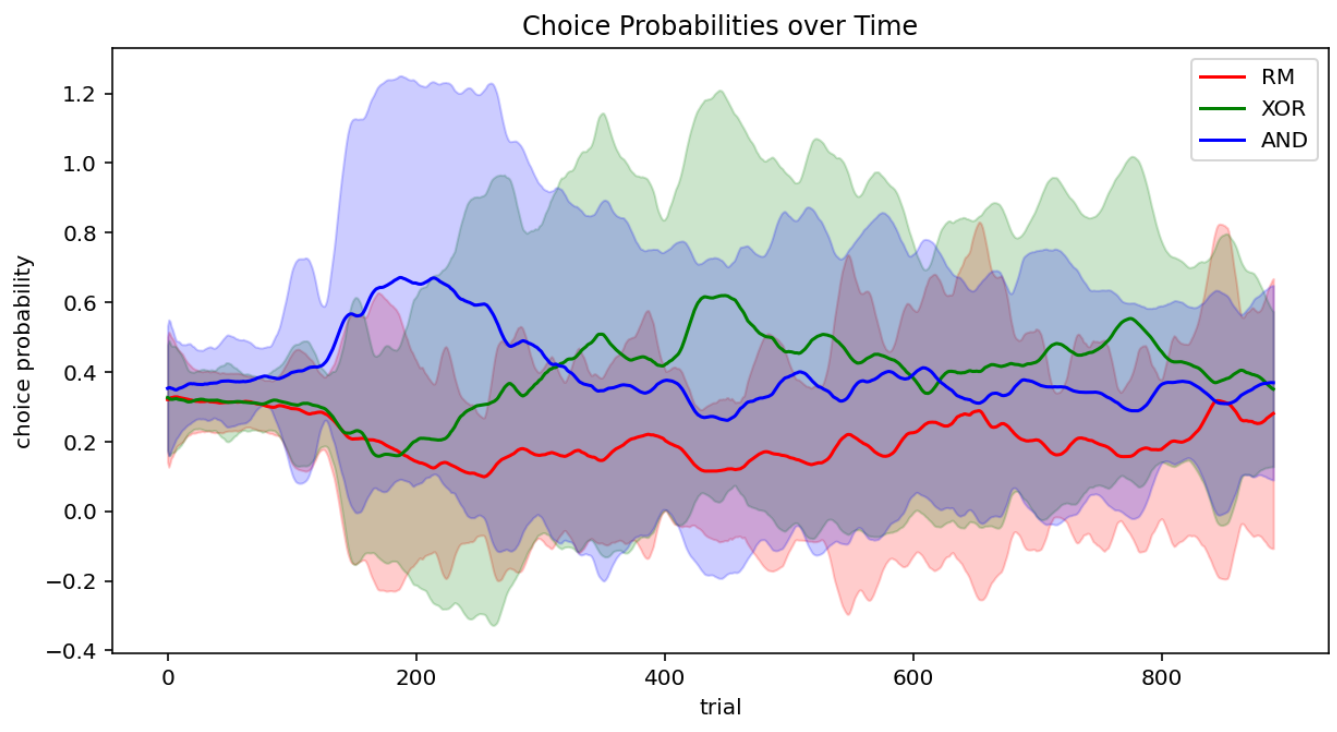


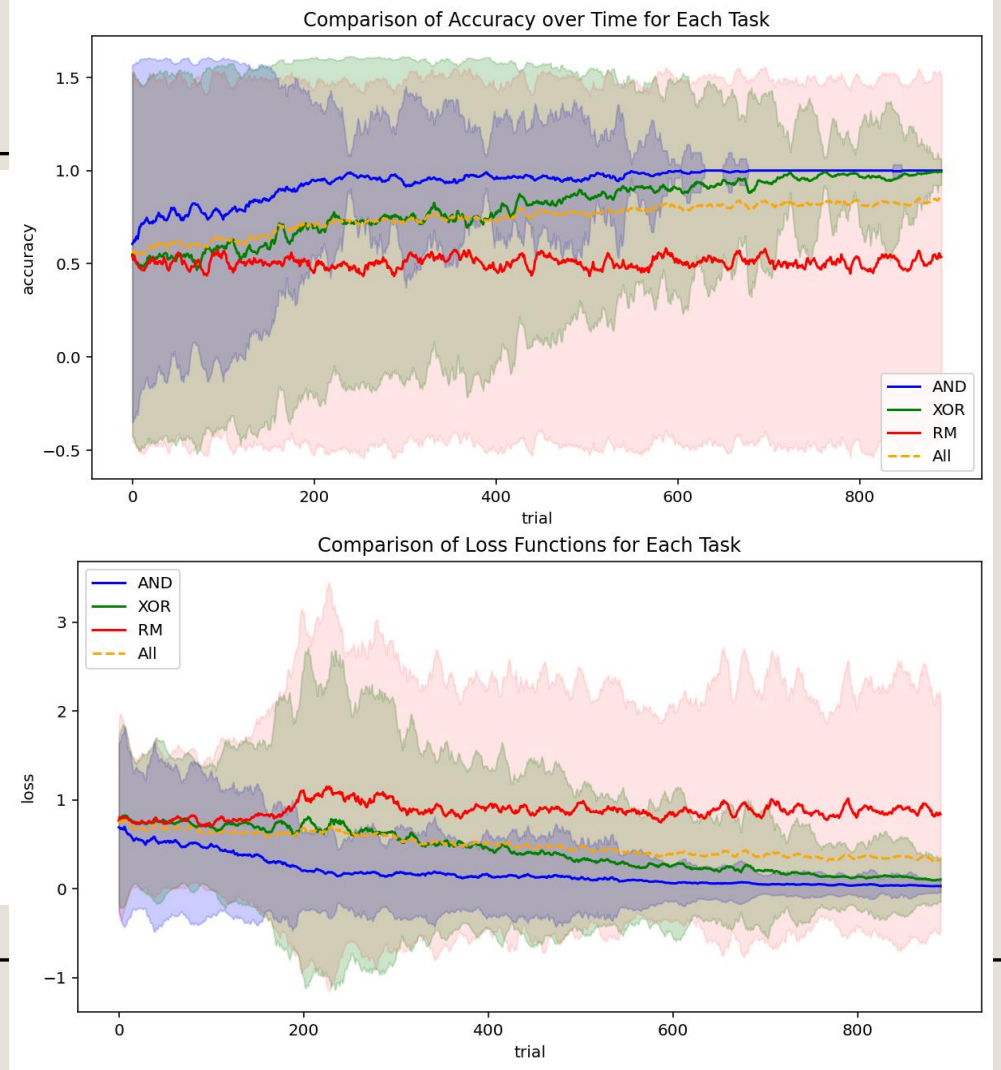
Ideas to fix LP signed/unsigned

- Different mathematic formula to increase sensitivity? (log, exp etc?)
- Perhaps due to how subtil the changes are? → the other criteria help make up for it

Performance of combined criteria compared to iid
+ 2 different windows

Combined criteria, manual weights (window = 60)

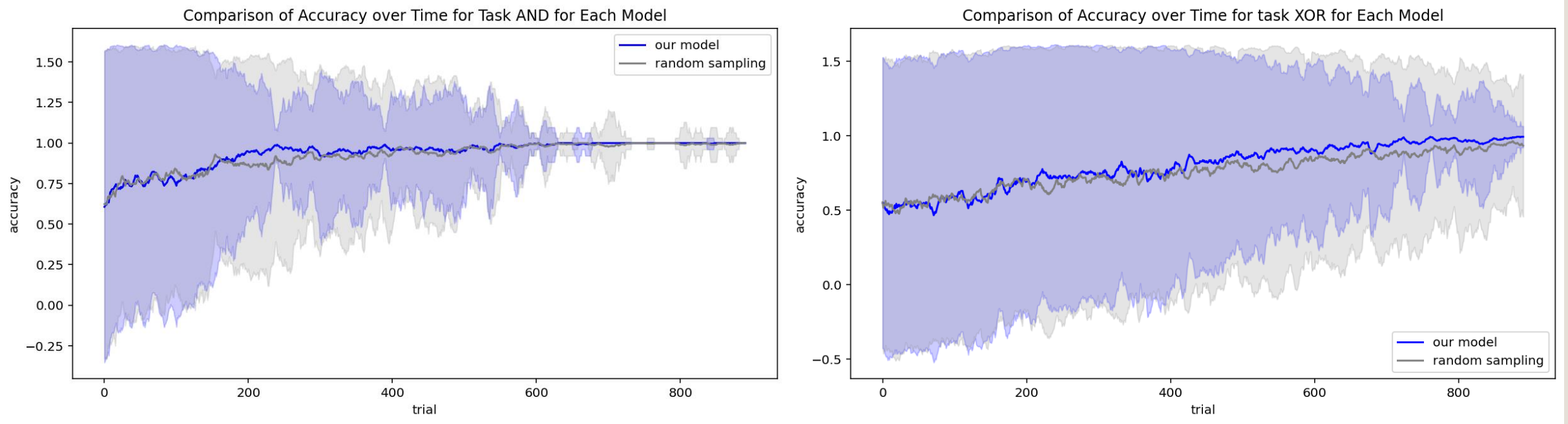




What the models look like

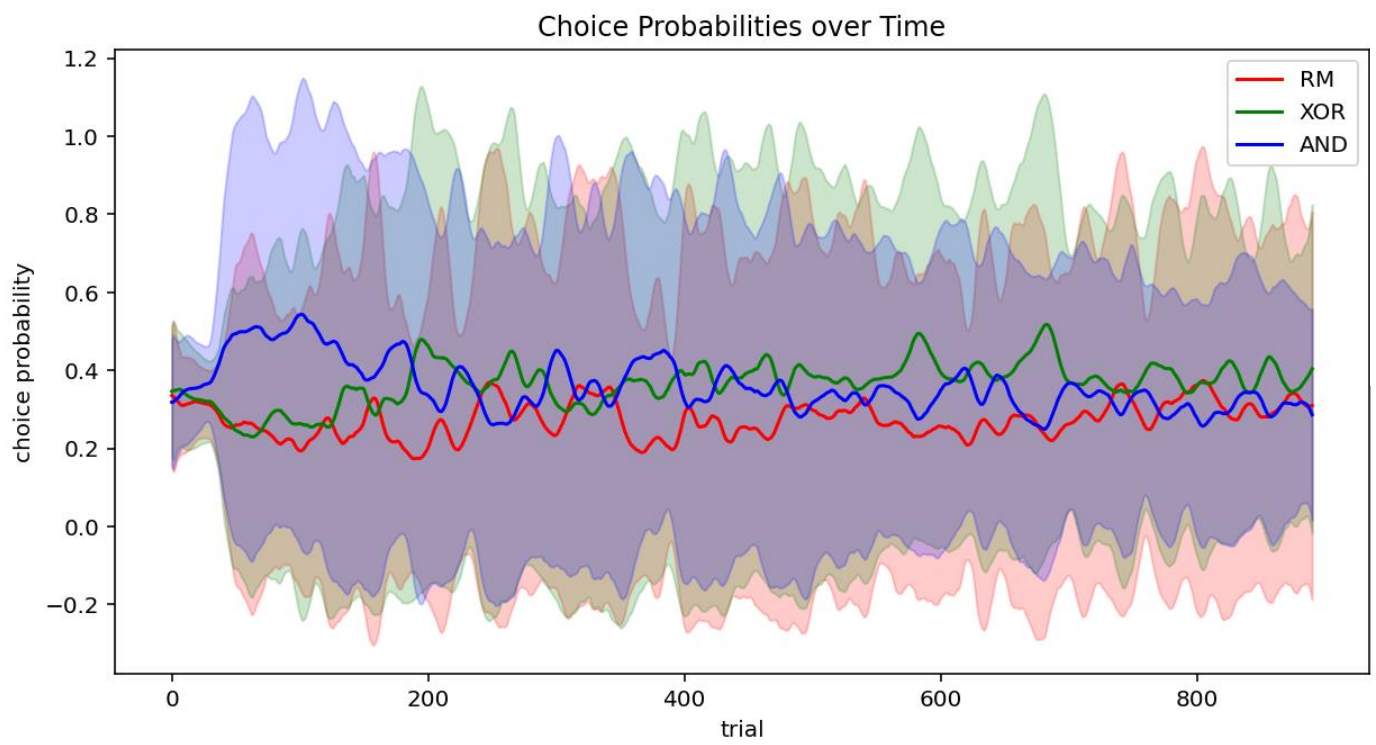
```
weight = {
'LP_signed': 10,
'LP_unsigned': 10,
'acc': 1,
'novelty': 1
} #manually set for now
```

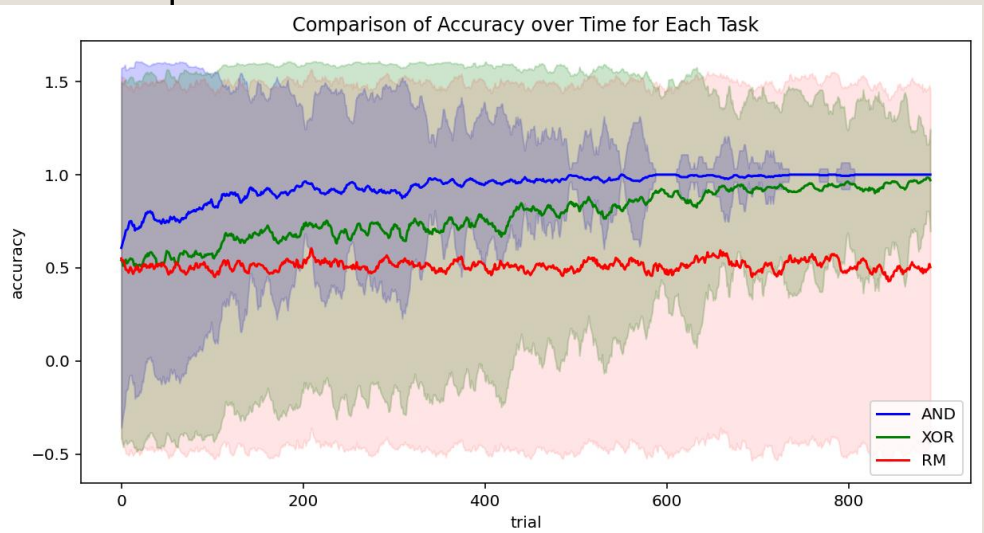
Combined criteria, manual weights (window = 60)

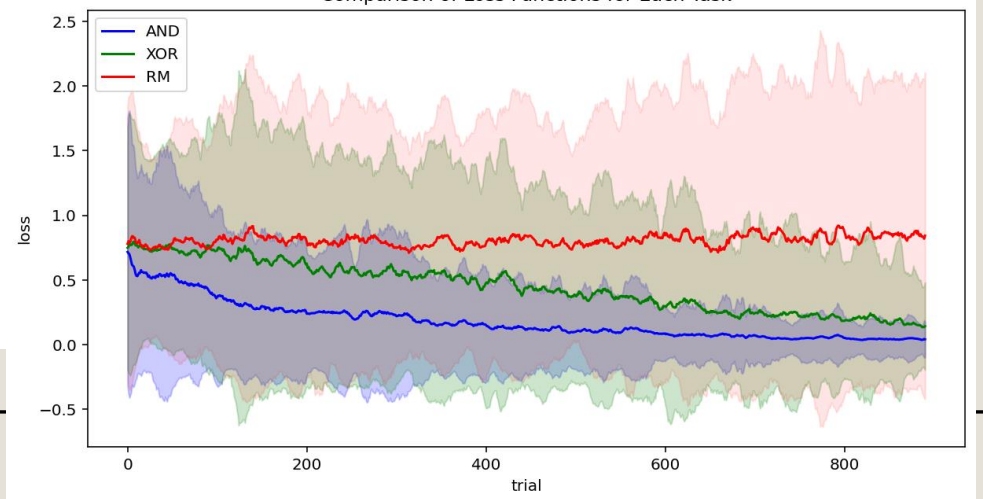


What the models look like

Combined criteria, manual weights (window = 20)



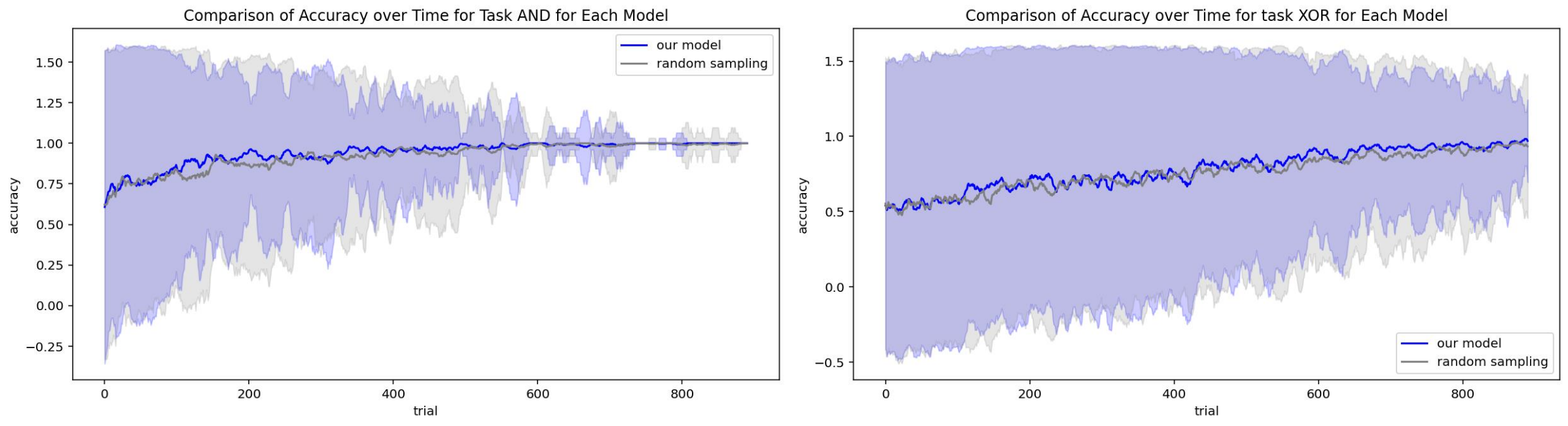




What the models look like

```
weight = {
'LP_signed': 10,
'LP_unsigned': 10,
'acc': 1,
'novelty': 1
} #manually set for now
```

Combined criteria, manual weights (window = 20)



What the models look like

QUESTIONS

QUESTION 1	▪ For policy gradient, use any packages?
QUESTION 2	▪ REINFORCE (with/out baseline) or actor critics?
NOTES	▪ …

NEXT WEEK'S OBJECTIVES

1 MAKE A SIMPLE POLICY GRADIENT MODEL

APPLY POLICY GRADIENT

To the complete model

2

KEEP IMPROVING LP

By trying new mathematics options

4

TRY NEW CRITERIA

Entropy

3

SUMMARISE FINDINGS

Write a short summary of findings, questions and plan for next week.

5