Supplementary: EMOTE - An Explainable architecture for Modelling the Other Through Empathy

1 Experiment Hyperparameters

Tables 3, 4, 5 and 6 list the experimental settings for all games. When tuning δ, it was found that it was made easier (less sensitive) when a constant (> 1) was multiplied with Loss term 2. δ can be tuned without the multiplier to achieve the same performance (as with the multiplier) but it is more sensitive and as such requires finer resolution tuning. When this multiplier was set in the range $[4, 8]$, δ values above 0.75 had the best performance.

Table 3: Assistive 1

parameter	value	comment
Episodes	2000	
No. Trials	10	
State View window	5 x 5	
Batch Size	16	
γ	0.9	
optimiser	Adam	
learning rate	1e-4	Q_{learn} DQN
learning rate	1e-5	Sypmpathy
learning rate	1e-5	E-Feature
learning rate	1e-5	E-Image
exploration initial	1	
exploration minimum	0.1	
exploration decay	0.99	
target network update		every 2 episodes
E-Feature δ	0.75	
E-Feature Loss 1 weight	1	
E-Feature Loss 2 weight	4	
E-Image δ	1.0	
E-Image Loss 1 weight	1	
E-Image Loss 2 weight	4	

Table 4: Assistive 2

parameter	value	comment
Episodes	2000	
No. Trials	10	
State View window	5 x 5	
Batch Size	16	
γ	0.9	
optimiser	Adam	
learning rate	1e-4	Q_{learn} DQN
learning rate	1e-3	Sypmpathy
learning rate	1e-5	E-Feature
learning rate	1e-5	E-Image
exploration initial	1	
exploration minimum	0.1	
exploration decay	0.99	
target network update		every 2 episodes
E-Feature δ	0.75	
E-Feature Loss 1 weight	1	
E-Feature Loss 2 weight	4	
E-Image δ	1.0	
E-Image Loss 1 weight	1	
E-Image Loss 2 weight	4	

Table 5: Adversarial 1

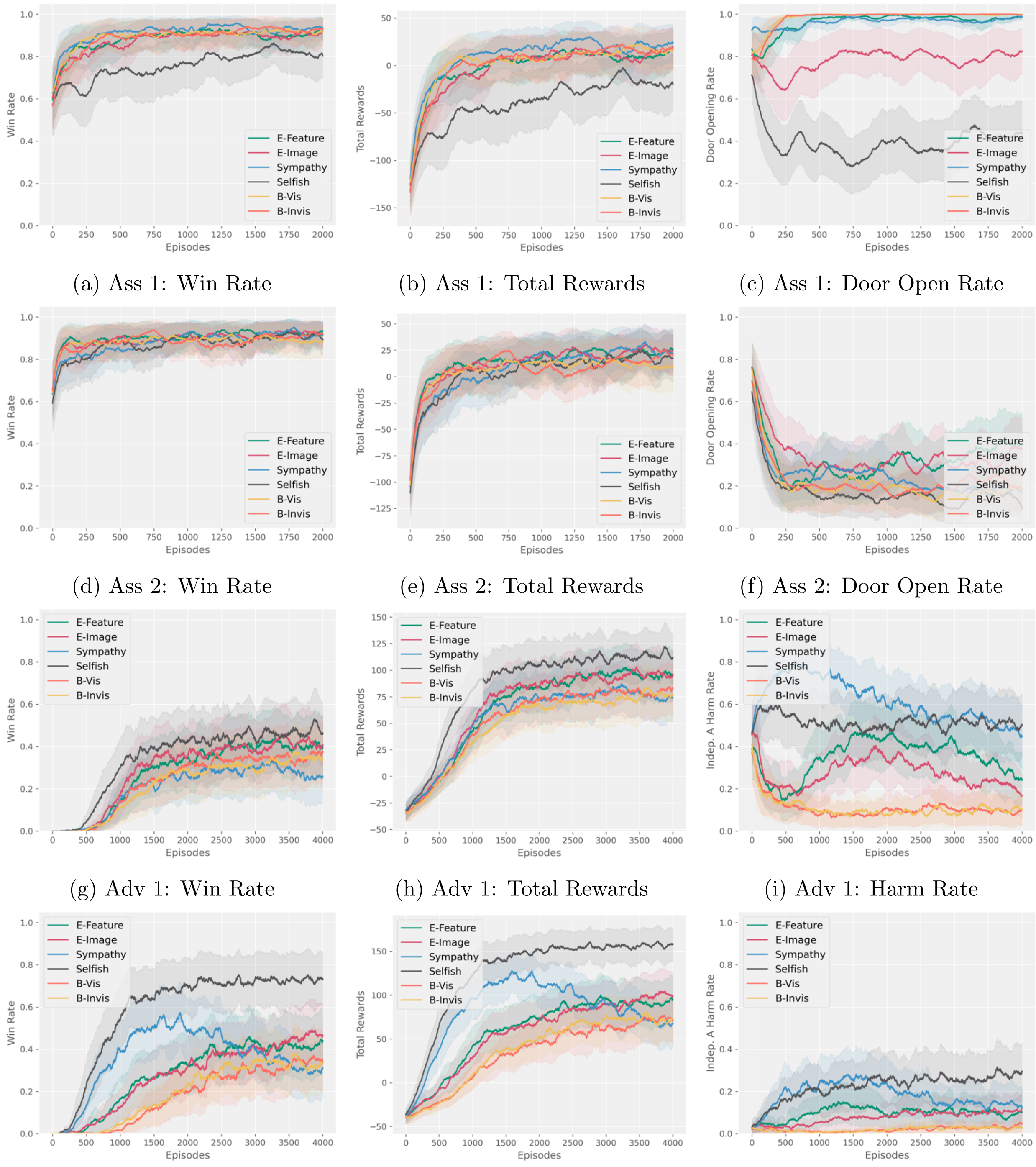
parameter	value	comment
Episodes	4000	
No. Trials	10	
State View window	5 x 5	
Batch Size	16	
γ	0.9	
optimiser	Adam	
learning rate	1e-4	Q_{learn} DQN
learning rate	1e-4	Sypmpathy
learning rate	1e-3	E-Feature
learning rate	1e-4	E-Image
exploration initial	1	
exploration minimum	0.1	
exploration decay	0.998	
target network update		every 2 episodes
E-Feature δ	0.95	
E-Feature Loss 1 weight	1	
E-Feature Loss 2 weight	4	
E-Image δ	0.95	
E-Image Loss 1 weight	1	
E-Image Loss 2 weight	8	

Table 6: Adversarial 2

parameter	value	comment
Episodes	4000	
No. Trials	10	
State View window	5 x 5	
Batch Size	16	
γ	0.9	
optimiser	Adam	
learning rate	1e-4	Q_{learn} DQN
learning rate	1e-4	Sypmpathy
learning rate	1e-3	E-Feature
learning rate	1e-4	E-Image
exploration initial	1	
exploration minimum	0.1	
exploration decay	0.998	
target network update		every 2 episodes
E-Feature δ	0.95	
E-Feature Loss 1 weight	1	
E-Feature Loss 2 weight	4	
E-Image δ	0.95	
E-Image Loss 1 weight	1	
E-Image Loss 2 weight	8	

2 Performance

Figure 6 plots the win rate, total rewards of the learning agent, and whether the door was opened (Assistive games) or the independent agent was harmed by the learning agent (Adversarial games) over the training period.



(a) Ass 1: Win Rate (b) Ass 1: Total Rewards (c) Ass 1: Door Open Rate

(d) Ass 2: Win Rate (e) Ass 2: Total Rewards (f) Ass 2: Door Open Rate

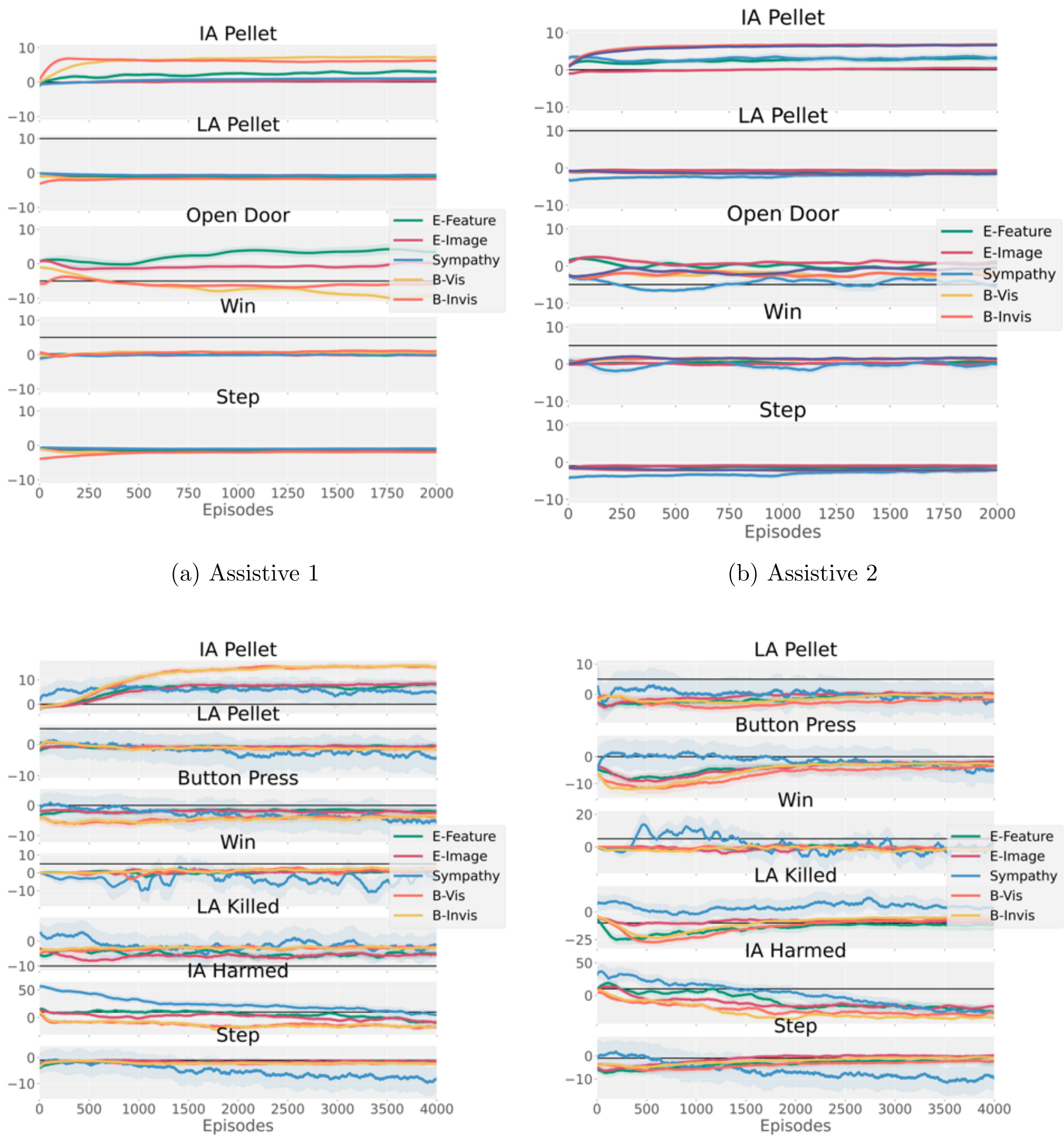
(g) Adv 1: Win Rate (h) Adv 1: Total Rewards (i) Adv 1: Harm Rate

(j) Adv 2: Win Rate (k) Adv 2: Total Rewards (l) Adv 2: Harm Rate

Figure 6: Performance in each game. Assistive 1 (a-c), Assistive 2 (d-f), Adversarial 1 (g-i) and Adversarial 2 (j-l)

3 IRL Trends

Figure 7 shows the trend of the inferred rewards of the independent agent ($\hat{R}_{indep}$) over the training period under the various runs. The weights of the Sympathy rewards have been scaled to have a $l1$ norm equal to that of the learning agent.



(a) Assistive 1

(b) Assistive 2

(c) Adversarial 1

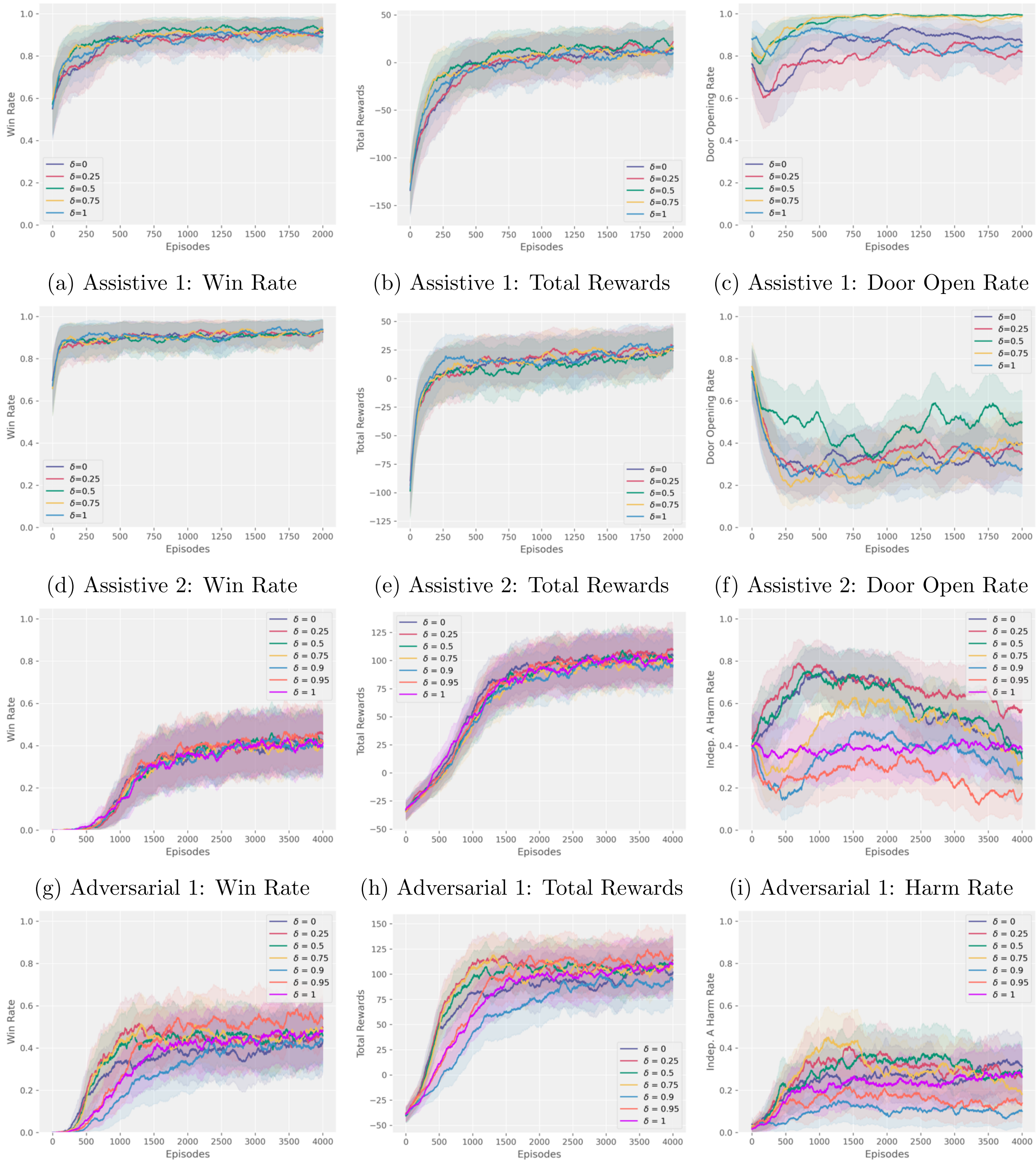
(d) Adversarial 2

Figure 7: Trend of $\hat{R}_{indep}$ over training. IA - Independent Agent, LA - Learning Agent

4 δ Sensitivity

4.1 δ Performance

Figures 8 and 9 show the impact of altering the hyperparameter δ on the win rate, total rewards, and either door open rate or rate of harm of the independent agent by the learning agent for each of the four games under both a E-Feature and E-Image Imagination Networks, respectively.



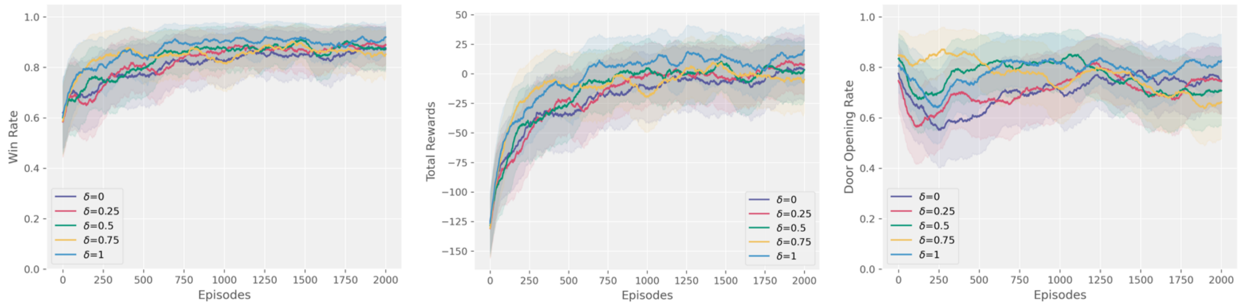
(a) Assistive 1: Win Rate (b) Assistive 1: Total Rewards (c) Assistive 1: Door Open Rate

(d) Assistive 2: Win Rate (e) Assistive 2: Total Rewards (f) Assistive 2: Door Open Rate

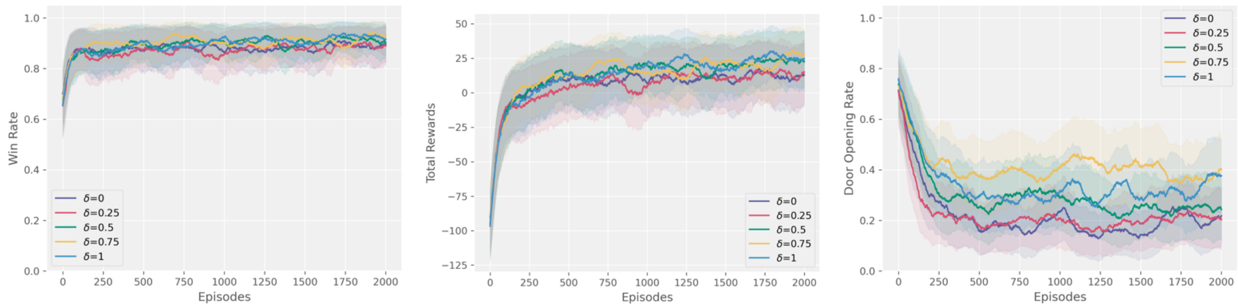
(g) Adversarial 1: Win Rate (h) Adversarial 1: Total Rewards (i) Adversarial 1: Harm Rate

(j) Adversarial 2: Win Rate (k) Adversarial 2: Total Rewards (l) Adversarial 2: Harm Rate

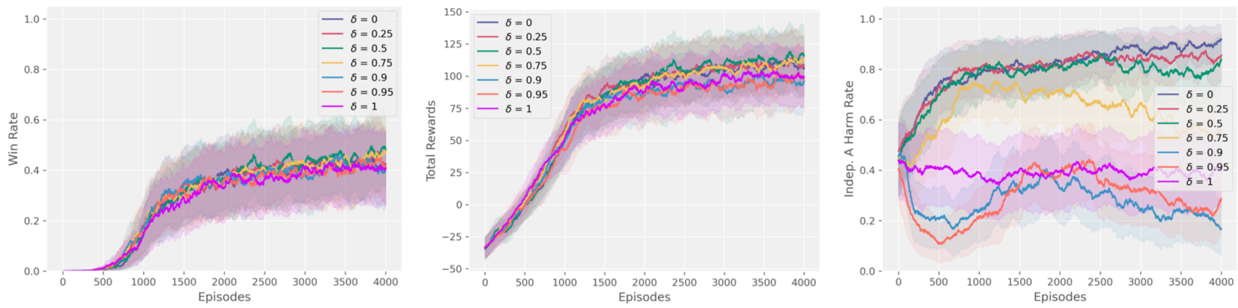
Figure 8: E-Feature: Impact of varying δ on learning agent's performance. Assistive 1 (a-c), Assistive 2 (d-f), Adversarial 1 (g-i) and Adversarial 2 (j-l)



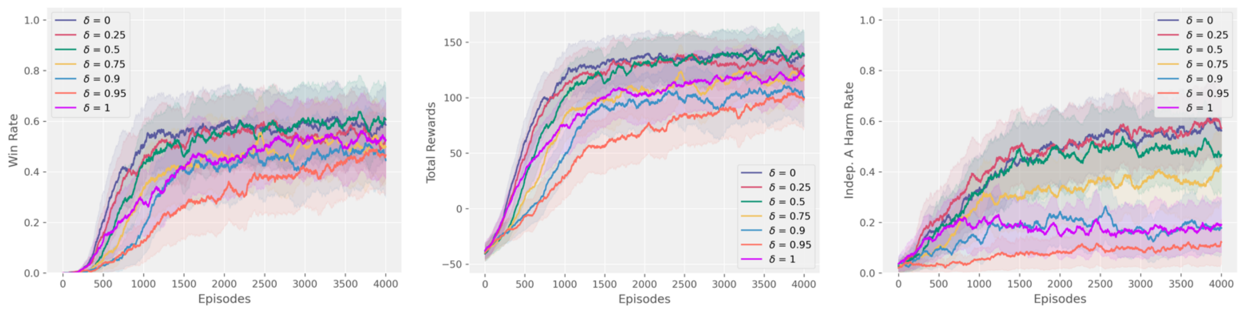
(a) Assistive 1: Win Rate (b) Assistive 1: Total Rewards (c) Assistive 1: Door Open Rate



(d) Assistive 2: Win Rate (e) Assistive 2: Total Rewards (f) Assistive 2: Door Open Rate



(g) Adversarial 1: Win Rate (h) Adversarial 1: Total Rewards (i) Adversarial 1: Harm Rate

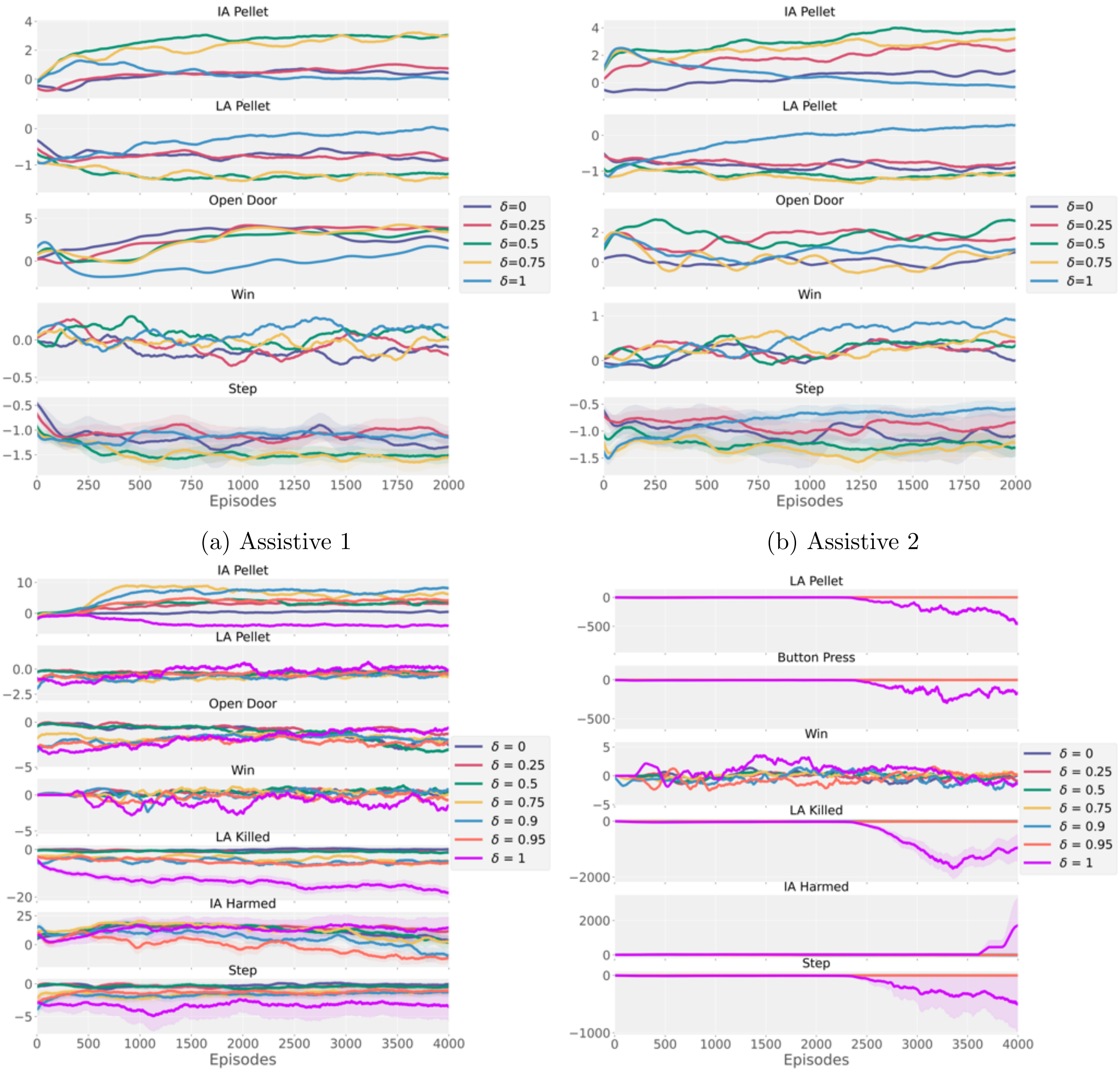


(j) Adversarial 2: Win Rate (k) Adversarial 2: Total Rewards (l) Adversarial 2: Harm Rate

Figure 9: E-Image: Impact of varying δ on learning agent's performance. Assistive 1 (a-c), Assistive 2 (d-f), Adversarial 1 (g-i) and Adversarial 2 (j-l)

4.2 δ IRL Trends

Figures 10 and 11 illustrate the impact to the resulting $\hat{R}_{indep}$ trends by altering δ on both the E-Feature and E-Image Imagination Networks, respectively.



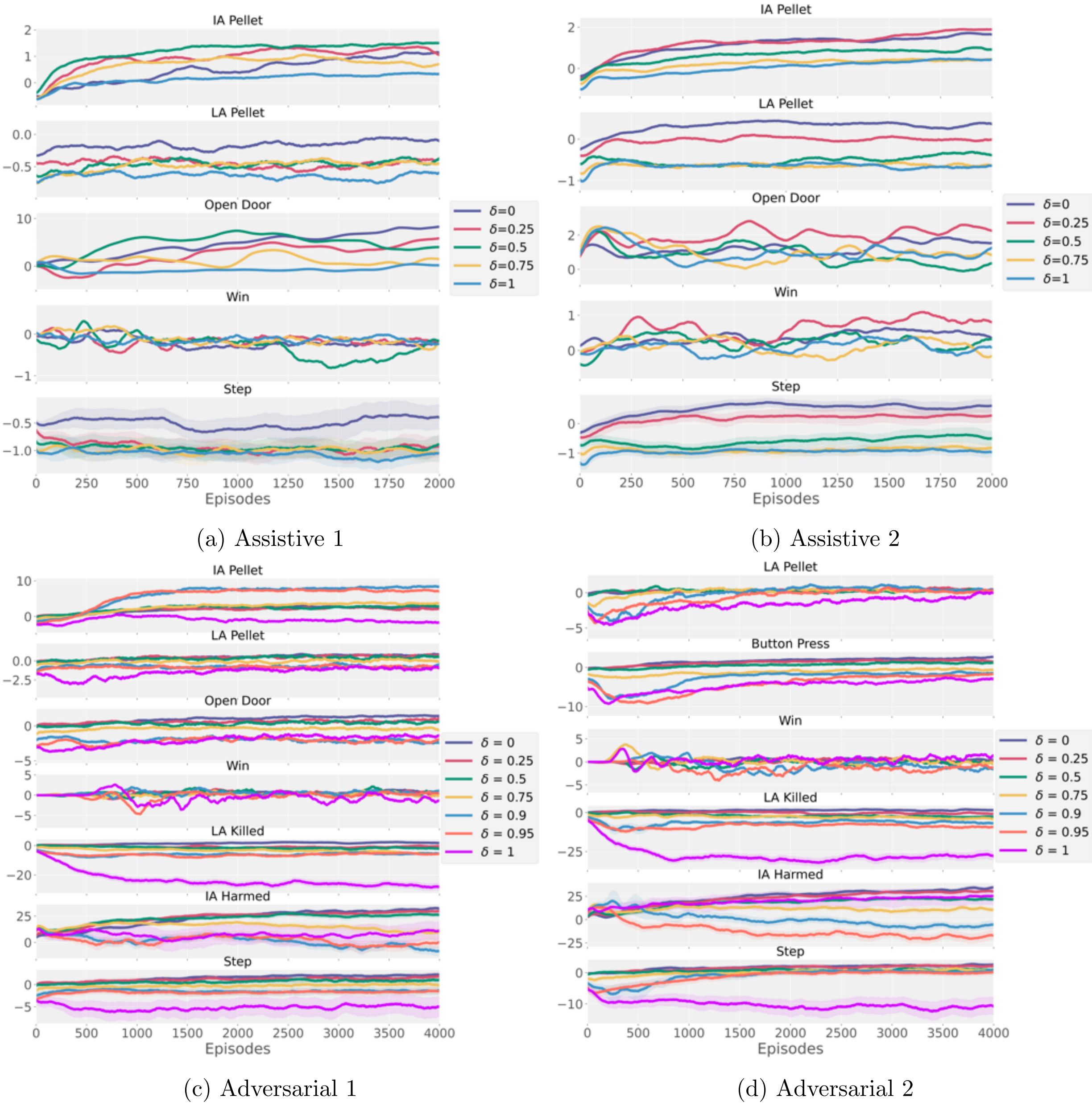
(a) Assistive 1

(b) Assistive 2

(c) Adversarial 1

(d) Adversarial 2

Figure 10: E-Feature: Trend of the estimated reward of the independent agent $\hat{R}_{indep}$ as δ is varied. IA - Independent Agent, LA - Learning Agent



(a) Assistive 1

(b) Assistive 2

(c) Adversarial 1

(d) Adversarial 2

Figure 11: E-Image: Trend of the estimated reward of the independent agent $\hat{R}_{indep}$ as δ is varied. IA - Independent Agent, LA - Learning Agent

5 Empathetic States

5.1 Generating Empathetic States s_e

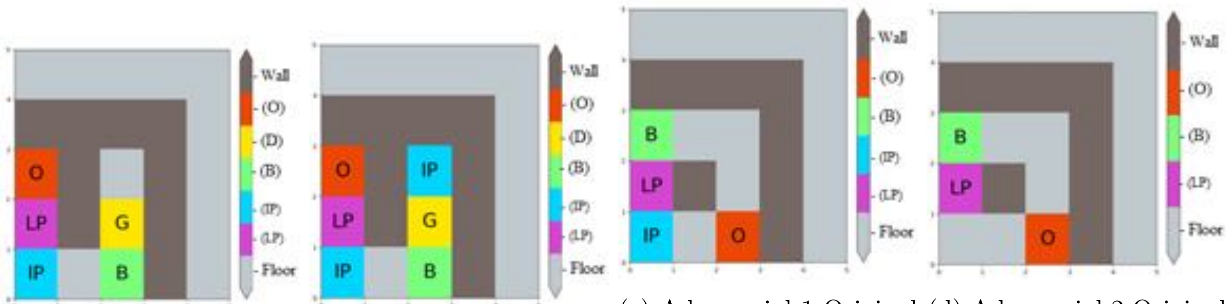
For the original states s_i input into the Imagination Networks shown in Col 1 of Figure 5, Table 7 and Table 8 list the corresponding value of each of the shown state features for the Assistive and Adversarial games, respectively. In order to generate the presented empathetic state representations, we manually specified (based on the ground truth values of each feature) ranges for each feature, which was then used to reconstruct the state features. For example, for a cell with value 0.82, this corresponded to "Other Agent", and was thus rendered in orange in the reconstructed state.

Table 7: Assistive Games: Value of each state feature

Feature	s_i value	s_e range
Floor	0	0.00 - 0.05
Learning Agent Pellet (LP)	0.13	0.05 - 0.25
Indep Agent Pellet (IP)	0.38	0.25 - 0.40
Button (B)	0.5	0.40 - 0.57
Door (D)	0.65	0.57 - 0.80
Other Agent (O)	0.85	0.80 - 0.90
Wall	1	0.90 - 1.00

Table 8: Adversarial Games: Value of each state feature

Feature	s_i value	Adv 1 s_e range	Adv 2 s_e range
Floor	0	0.00 - 0.05	0.00 - 0.05
Learning Agent Pellet (LP)	0.13	0.05 - 0.25	0.05 - 032
Indep Agent Pellet (IP)	0.38	0.25 - 0.40	-
Button (B)	0.5	0.40 - 0.57	0.32 - 0.57
Other Agent (O)	0.85	0.57 - 0.90	0.57 - 0.90
Wall	1	0.90 - 1.00	0.90 - 1.00

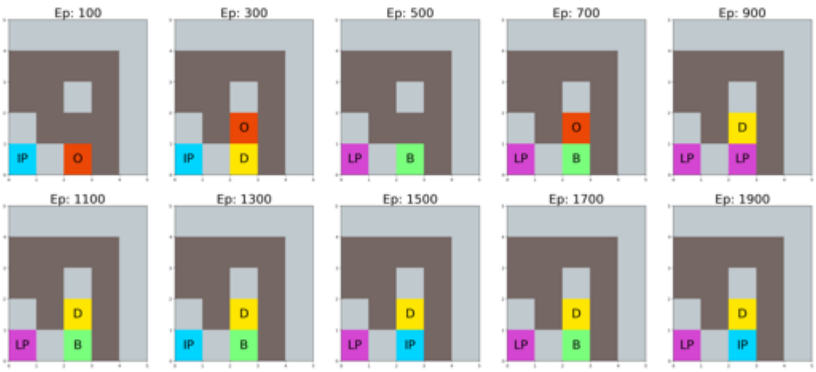


(a) Assistive 1 Original s_i (b) Assistive 2 Original s_i (c) Adversarial 1 Original s_i (d) Adversarial 2 Original s_i

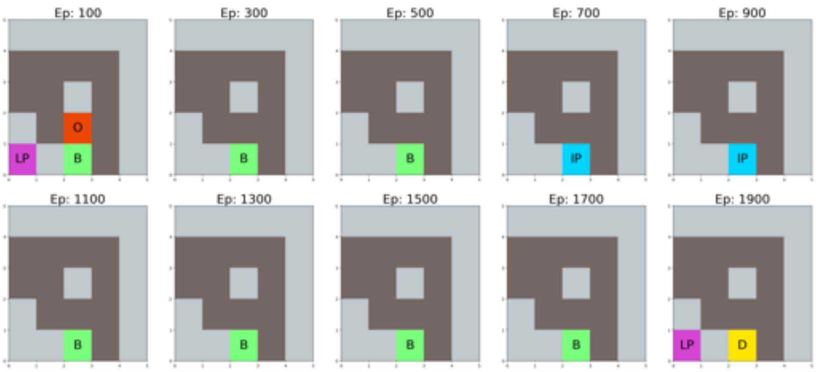
Figure 12: Original States s_i prior to transformation by the Imagination Networks.

5.2 Evolution of Empathetic State over time (during training)

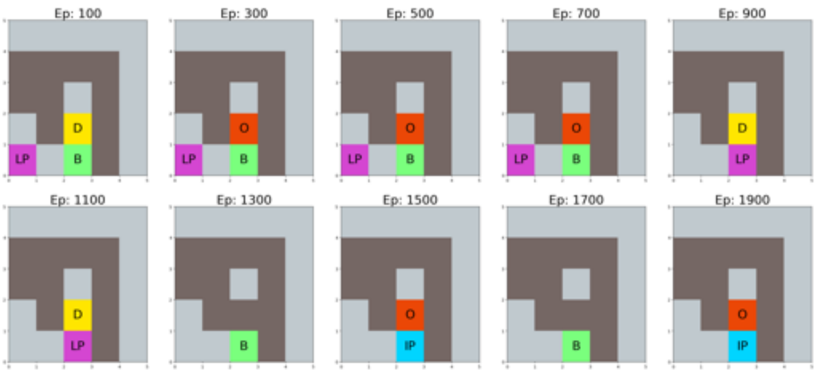
For each of the final empathetic states s_e shown in Figure 5, the figures below show the evolution of the empathetic state during training. These are for both the E-Feature and E-Image Imagination Network generated s_e.



(a) Assistive 1 Example 1: E-Feature s_e over time

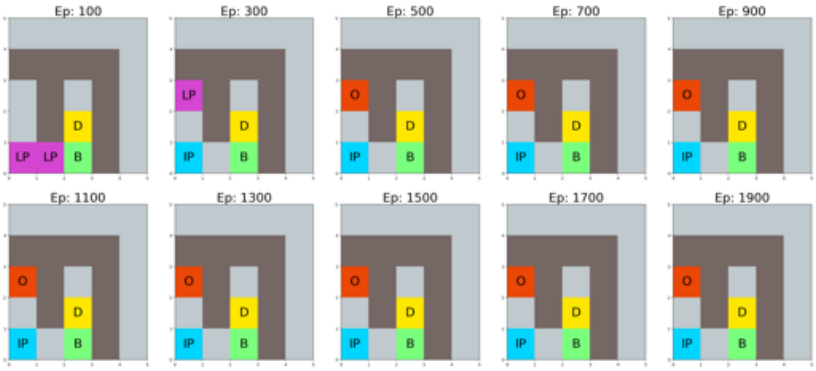


(b) Assistive 1 Example 2: E-Feature s_e over time

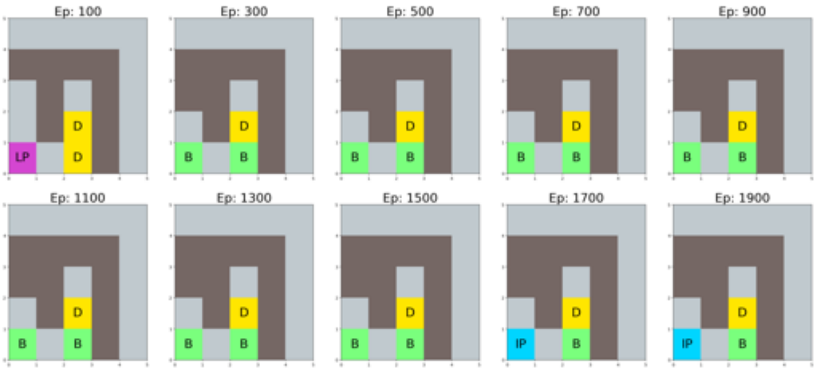


(c) Assistive 1 Example 3: E-Feature s_e over time

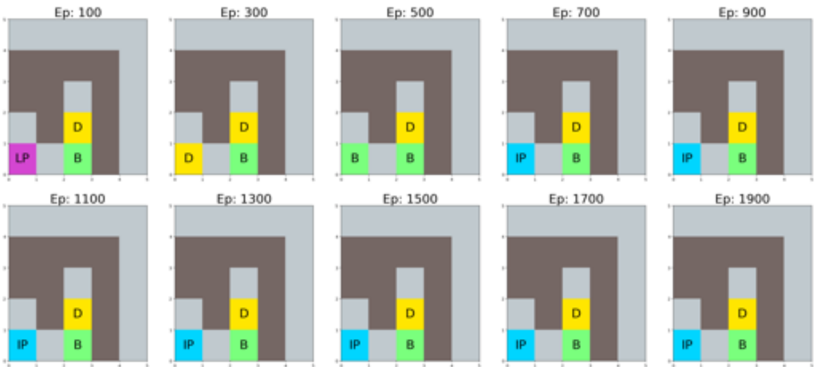
Figure 13: Assistive 1: Examples of the generated Empathetic State s_e over the training period under the E-Feature Imagination Networks for the final states shown in Figure 5.



(a) Assistive 1 Example 1: E-Image s_e over time

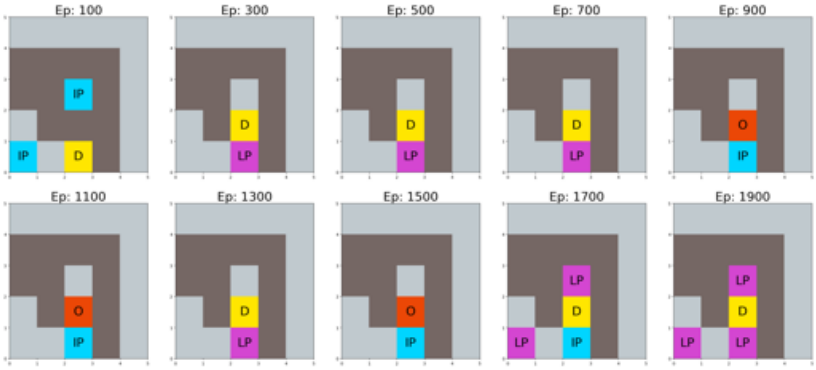


(b) Assistive 1 Example 2: E-Image s_e over time

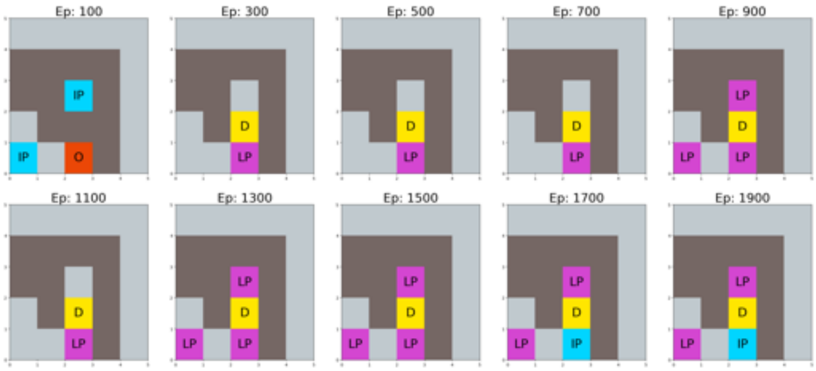


(c) Assistive 1 Example 3: E-Image s_e over time

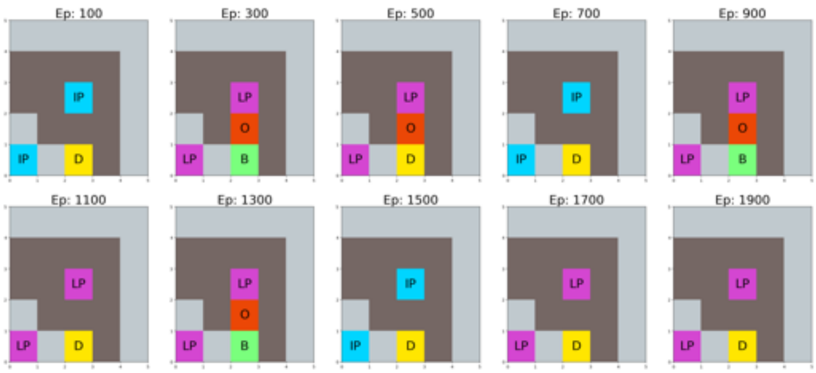
Figure 14: Assistive 1: Examples of the generated Empathetic State s_e over the training period under the E-Image Imagination Networks for the final states shown in Figure 5.



(a) Assistive 2 Example 1: E-Feature s_e over time

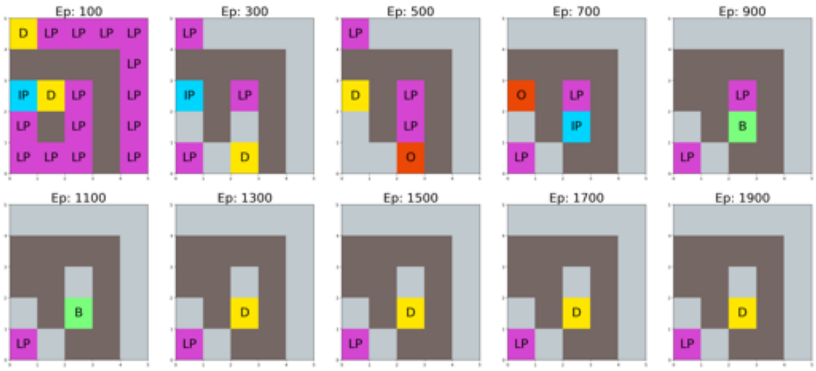


(b) Assistive 2 Example 2: E-Feature s_e over time

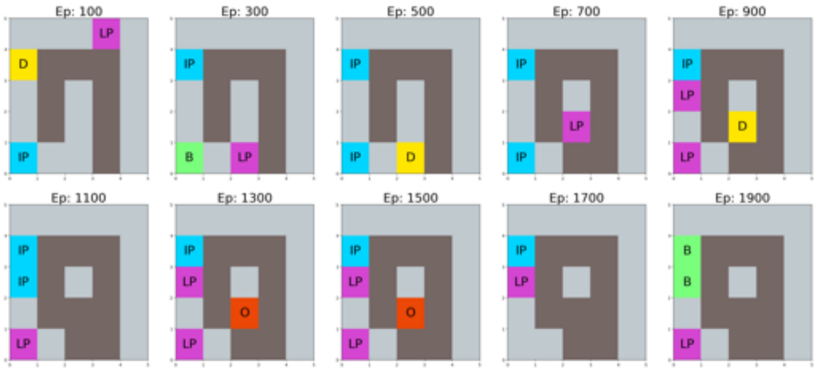


(c) Assistive 2 Example 3: E-Feature s_e over time

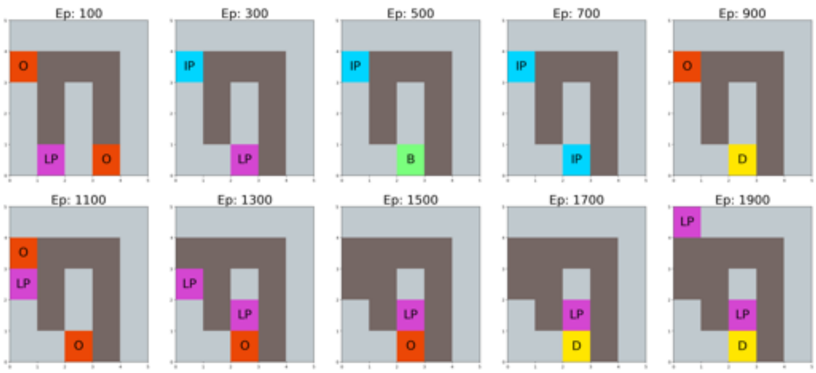
Figure 15: Assistive 2: Examples of the generated Empathetic State s_e over the training period under E-Feature Imagination Networks for the final states shown in Figure 5.



(a) Assistive 2 Example 1: E-Image s_e over time

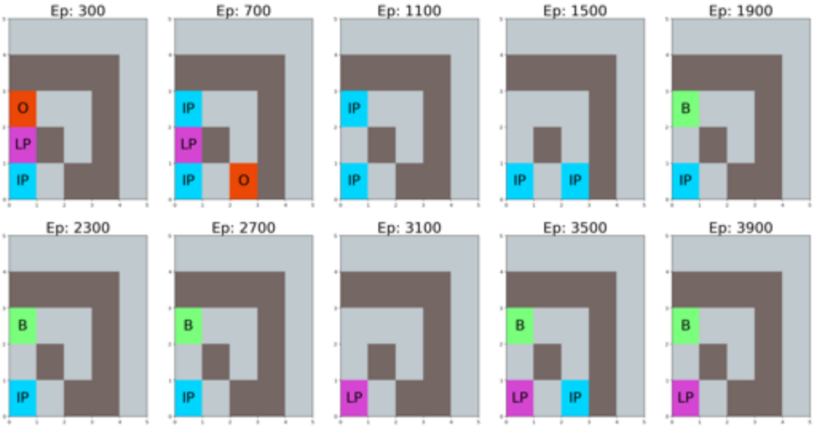


(b) Assistive 2 Example 2: E-Image s_e over time

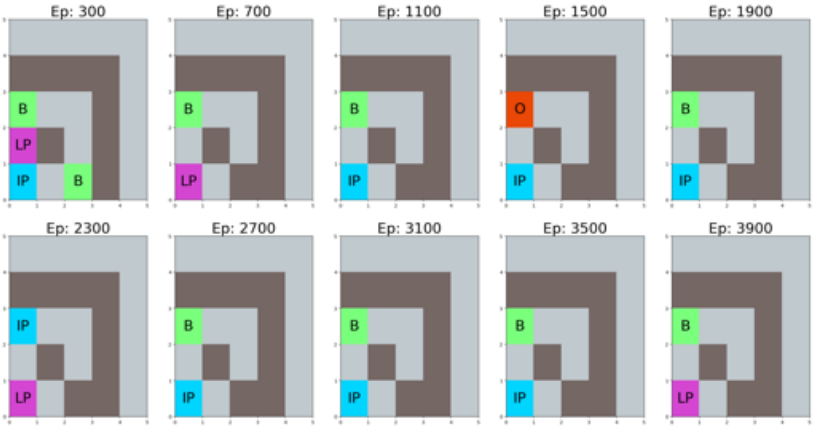


(c) Assistive 2 Example 3: E-Image s_e over time

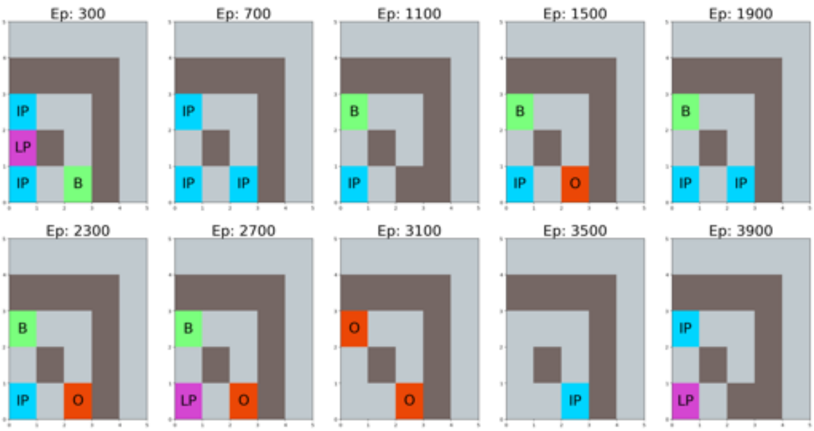
Figure 16: Assistive 2: Examples of the generated Empathetic State s_e over the training period under E-Image Imagination Networks for the final states shown in Figure 5.



(a) Adversarial 1 Example 1: E-Feature s_e over time

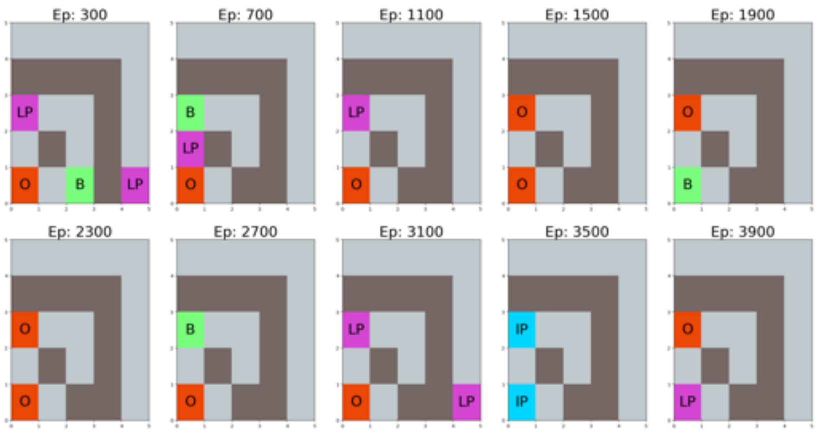


(b) Adversarial 1 Example 2: E-Feature s_e over time

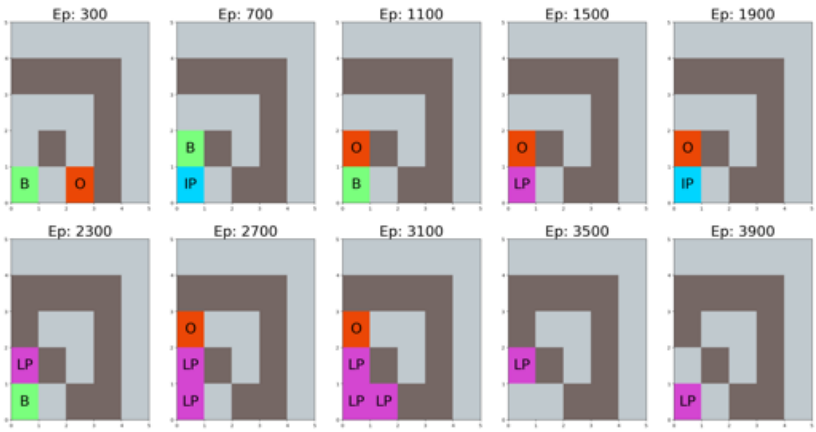


(c) Adversarial 1 Example 3: E-Image s_e over time

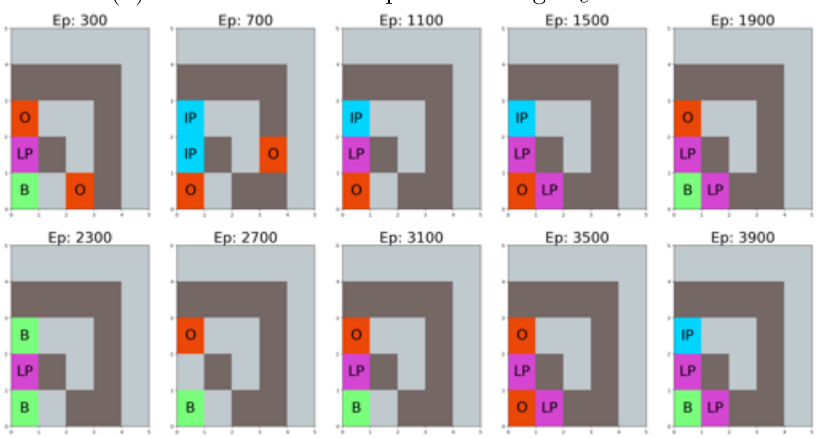
Figure 17: Adversarial 1: Examples of the generated Empathetic State s_e over the training period under E-Feature Imagination Networks for the final states shown in Figure 5.



(a) Adversarial 1 Example 1: E-Image s_e over time

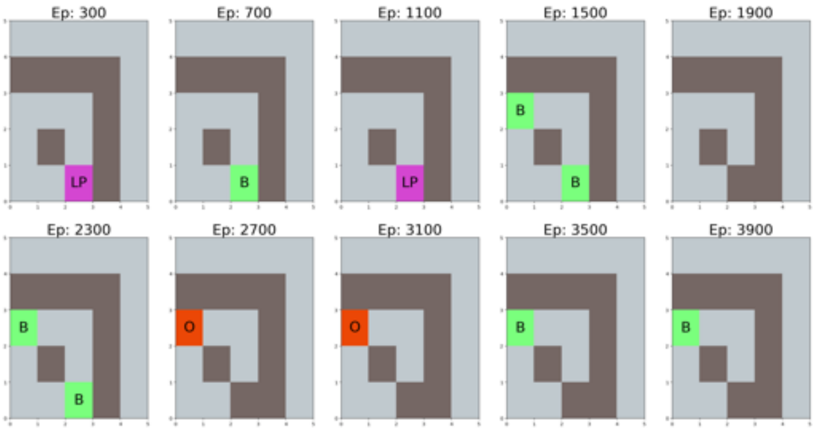


(b) Adversarial 1 Example 2: E-Image s_e over time

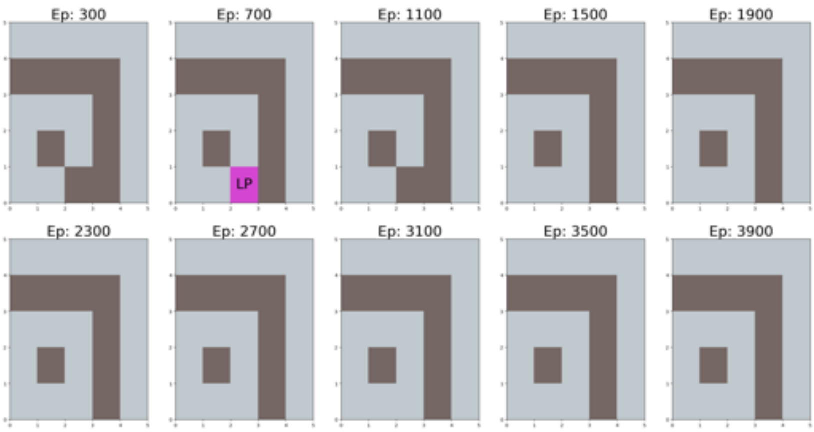


(c) Adversarial 1 Example 3: E-Image s_e over time

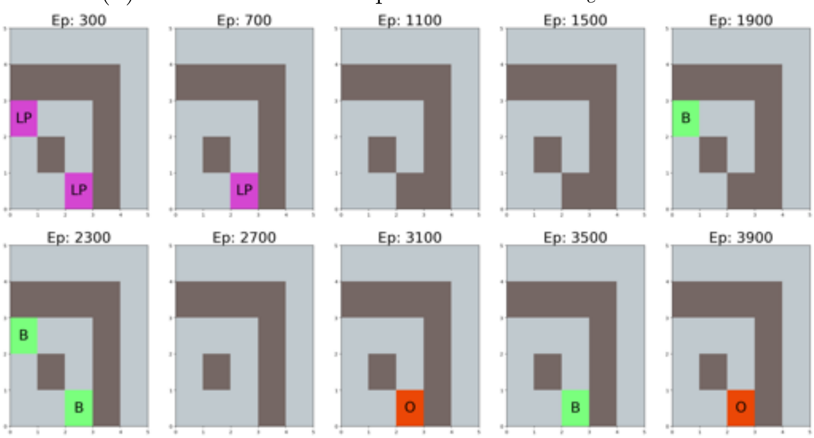
Figure 18: Adversarial 1: Examples of the generated Empathetic State s_e over the training period under E-Image Imagination Networks for the final states shown in Figure 5.



(a) Adversarial 2 Example 1: E-Feature s_e over time

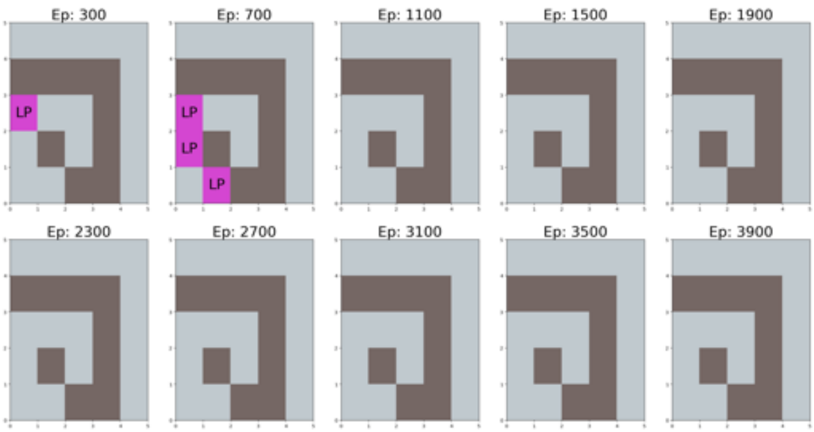


(b) Adversarial 2 Example 2: E-Feature s_e over time

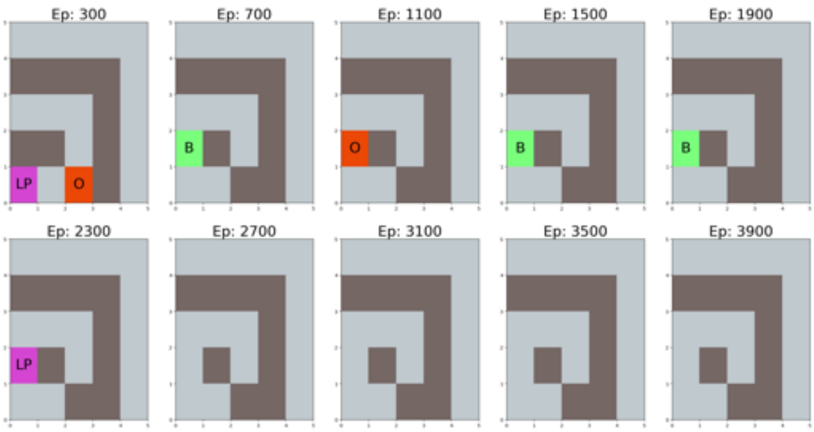


(c) Adversarial 2 Example 3: E-Image s_e over time

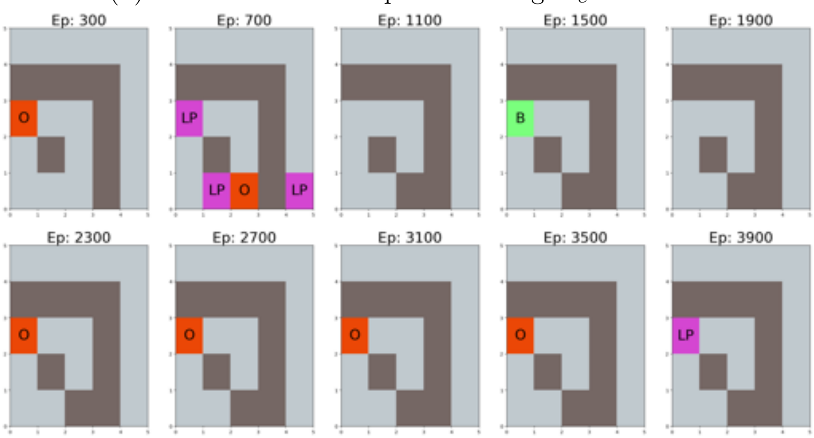
Figure 19: Adversarial 2: Examples of the generated Empathetic State s_e over the training period under E-Feature Imagination Networks for the final states shown in Figure 5.



(a) Adversarial 2 Example 1: E-Image s_e over time



(b) Adversarial 2 Example 2: E-Image s_e over time

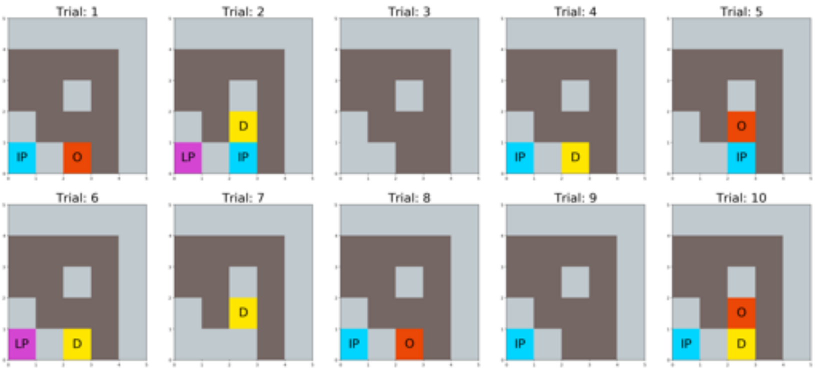


(c) Adversarial 2 Example 3: E-Image s_e over time

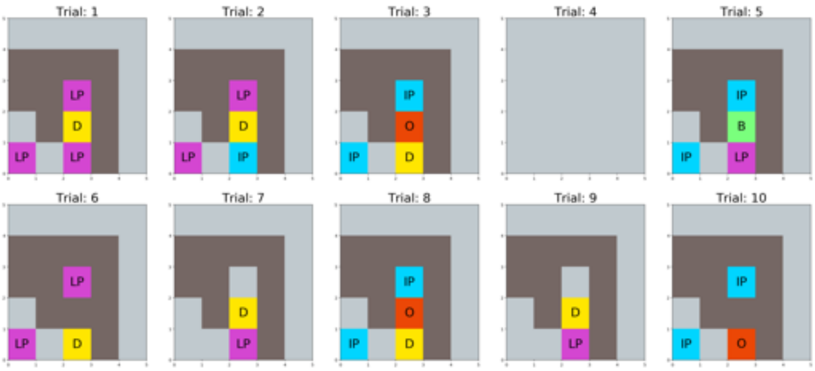
Figure 20: Adversarial 2: Examples of the generated Empathetic State s_e over the training period under E-Image Imagination Networks for the final states shown in Figure 5.

5.3 Final Empathetic States s_e from random trials

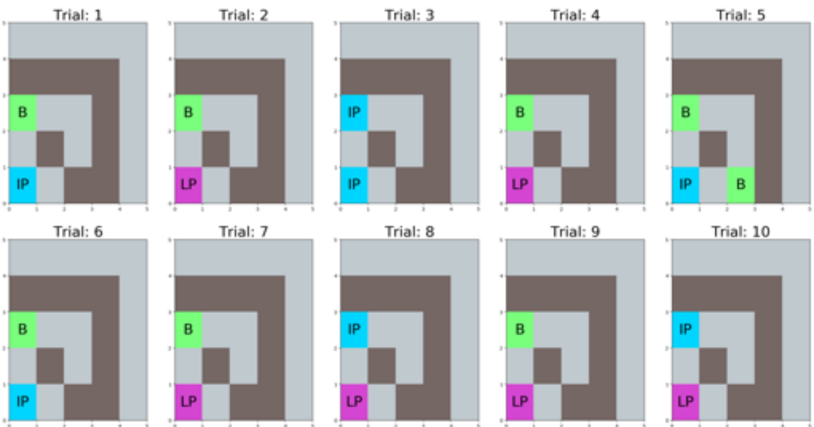
Figures 21 and 22 show the final empathetic state for all the randomly initialised trials of each game (10 of each). Results from both E-Feature and E-Image Imagination Networks are shown.



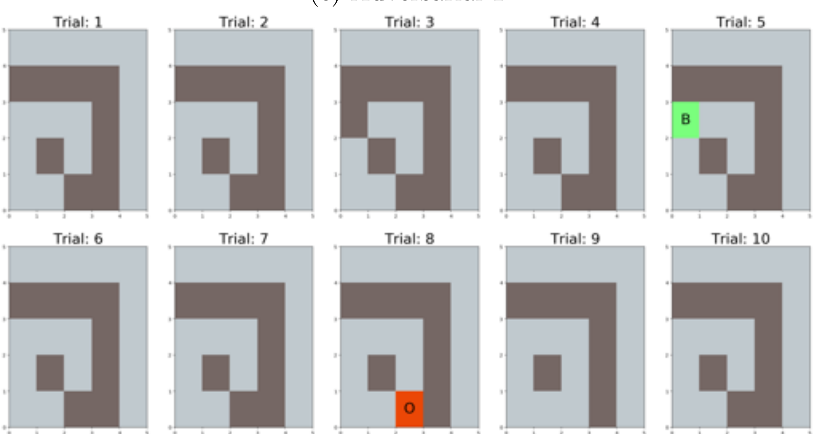
(a) Assistive 1



(b) Assistive 2

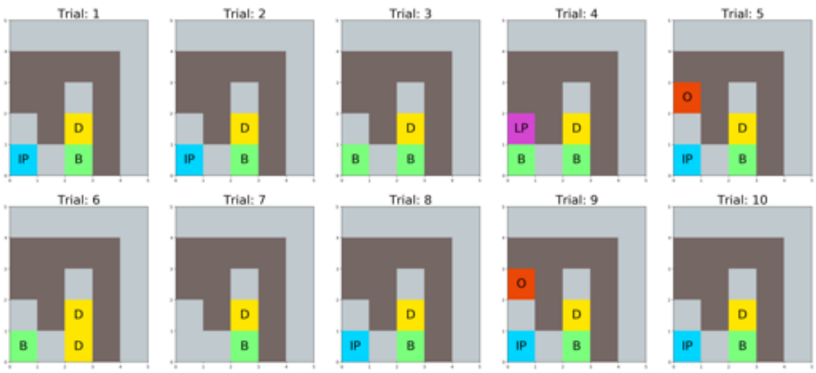


(c) Adversarial 1

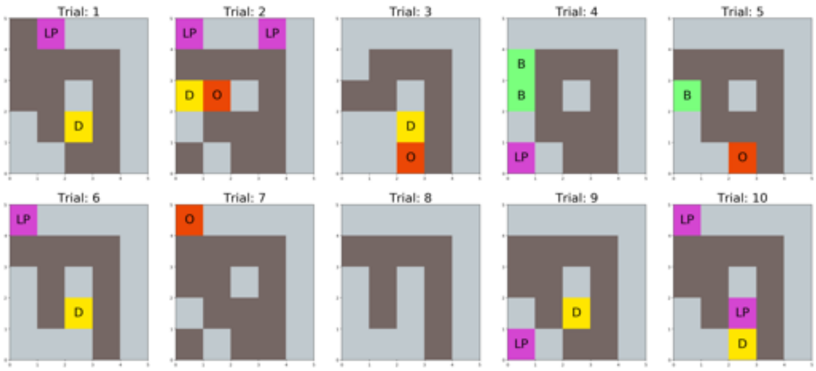


(d) Adversarial 2

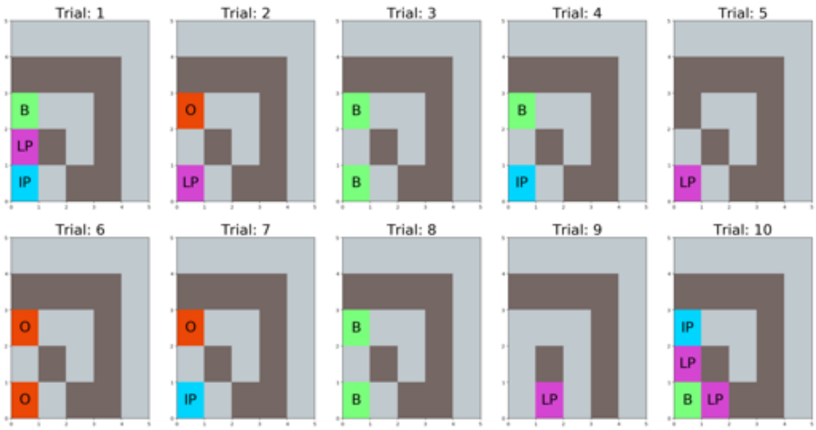
Figure 21: Final Empathetic state s_e produced via E-Feature Imagination Network for 10 random trials of each game.



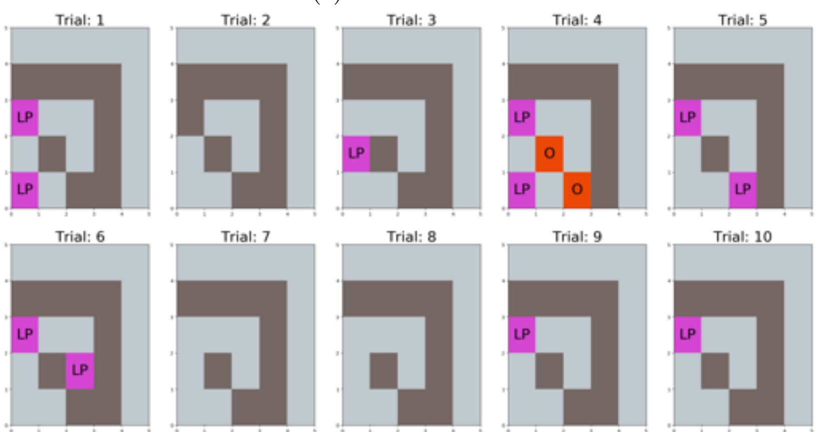
(a) Assistive 1



(b) Assistive 2



(c) Adversarial 1

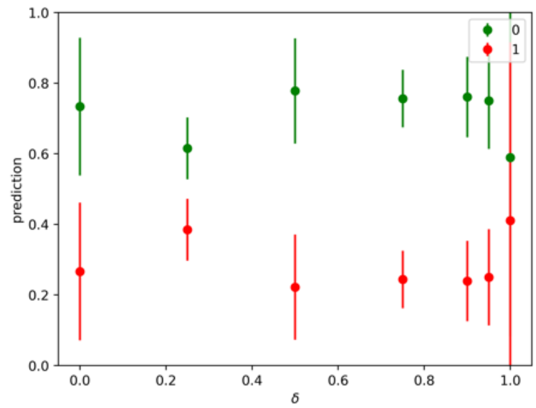


(d) Adversarial 2

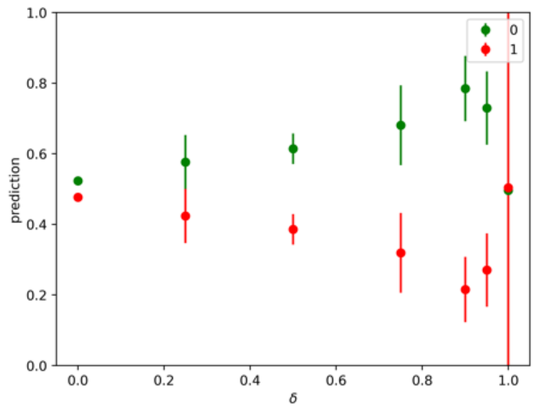
Figure 22: Final Empathetic state s_e produced via E-Image Imagination Network for 10 random trials of each game.

5.4 Adversarial Modelled Button

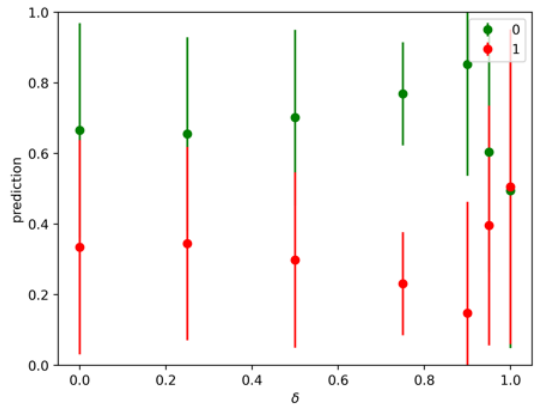
For the Adversarial games the button status was also modelled as part of the Imagination Model. As inputs this model took in the status of the button (either a 0 or a 1) and output a predicted empathetic button status. Figure 23 shows the predictions of this model after training for both the E-Feature and E-Image Imagination models for various δ values.



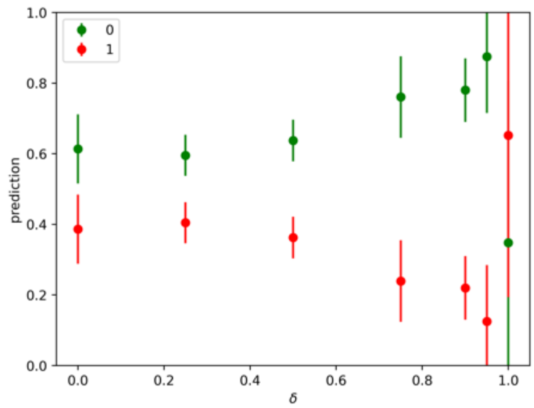
(a) Adversarial 1: E-Feature



(b) Adversarial 1: E-Image



(c) Adversarial 2: E-Feature



(d) Adversarial 2: E-Image

Figure 23: Predicted value of button status model at the end of training for various δ values. Adversarial 1 (a-b) and Adversarial 2 (c - d).