Supplementary Material: Improving Robot Motor Learning with Negatively Valenced Reinforcement Signals

Nicolás Navarro-Guerrero [1,*], **Robert Lowe** [2,3] **and Stefan Wermter** [1]

*Correspondence:
Nicolás Navarro-Guerrero
nicolas.navarro.guerrero@gmail.com

1 HYPERPARAMETER OPTIMIZATION

At a macroscopic level, i.e. at the evolutionary level, all solutions are comparable both in terms of convergence speed as well as in the quality of the best solution.

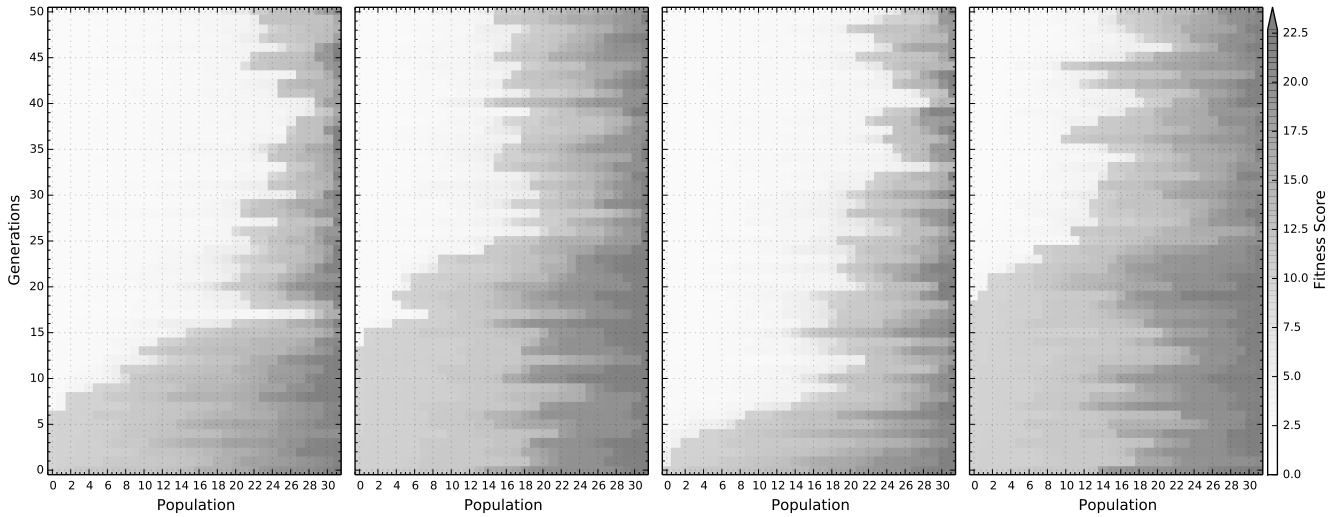

Figure S1. Fitness distribution over generations for all four conditions. From left to right, *reward only*, *reward+punishment*, *reward+nociception*, and *reward+punishment+nociception*. The fitness is directly computed from the total distance to the target, thus the lower the value the better.

2 RESULTS FOR THE BEST 4 HYPERPARAMETER SETS FOR EACH CONDITION

2.1 Positioning error

Figure S2 shows the average change of the positioning error during learning for the four best hyperparameter sets on the validation set. The average is over 10 different random network initializations.

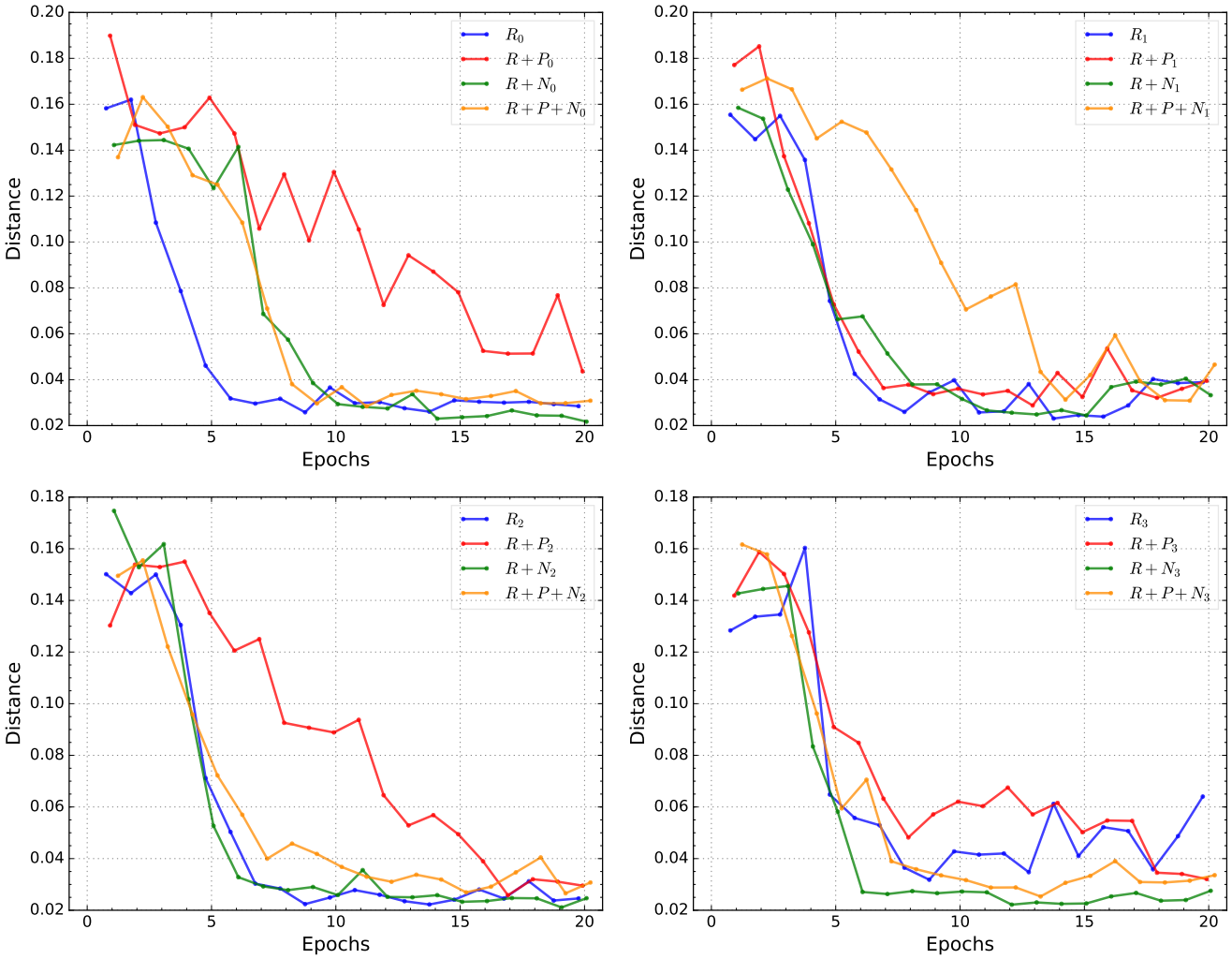

Figure S2. Mean positioning error for 10 runs of the best hyperparameters for each condition. Organised from left to right and top to bottom are the results for the best hyperparameter set and the fourth best hyperparameter set for each condition.

Figure S4 shows the average values for positioning error after 20 epochs for all four conditions and the best hyperparameter sets.

Figure S6 shows the 3 best runs versus the 3 worst runs in terms of positioning error for the best hyperparameters of each condition. All runs are sorted from smallest to largest positioning error. The distribution of the positioning error for all samples in the validation set is plotted. The blue (left side for each condition) shows the distribution of the 3 best runs whereas the red (right side for each condition) shows the distribution of the 3 worst runs. It seems that the distribution of

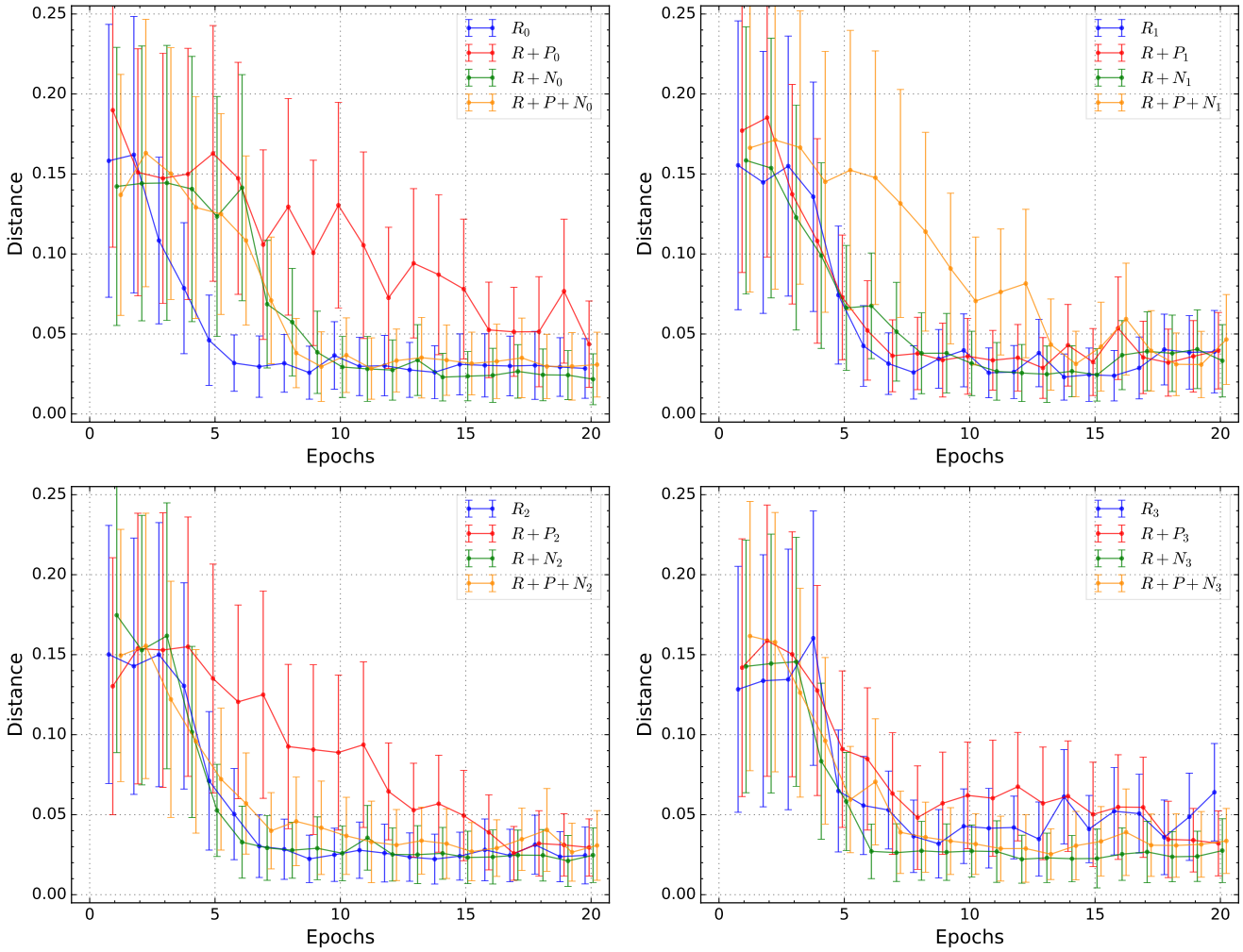

Figure S3. Mean positioning error for 10 runs of the best hyperparameters for each condition. Organised from left to right and top to bottom are the results for the best hyperparameter set and the fourth best hyperparameter set for each condition. The same as Figure S2 but this time including standard deviation.

the positioning error for all conditions and for all hyperparameter set is considerably larger for the worst initializations. This could be used as a test for discarding networks initialization that might be not so favourable early on in training.

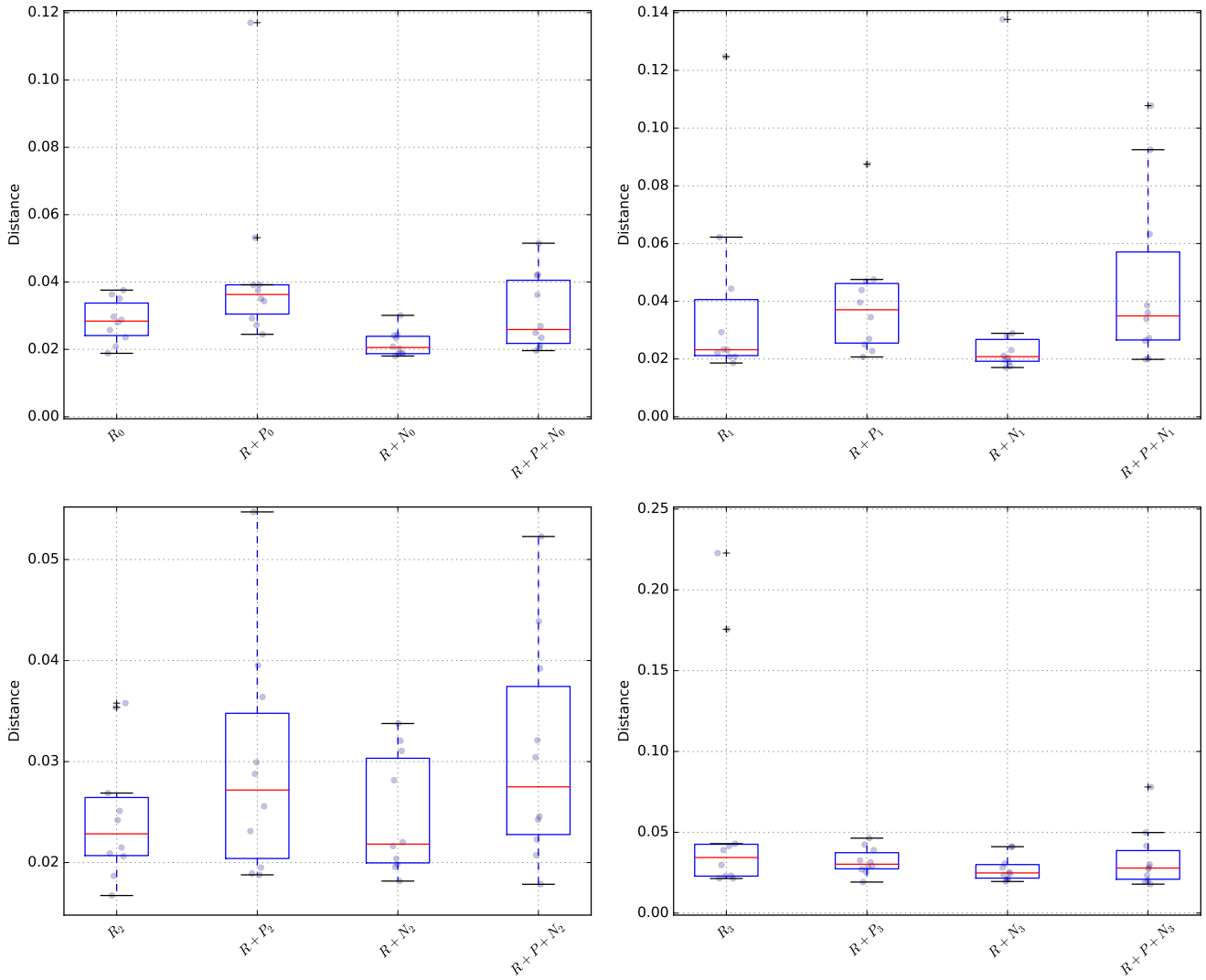

Figure S4. Mean positioning error for 10 runs of the best hyperparameters for each condition.

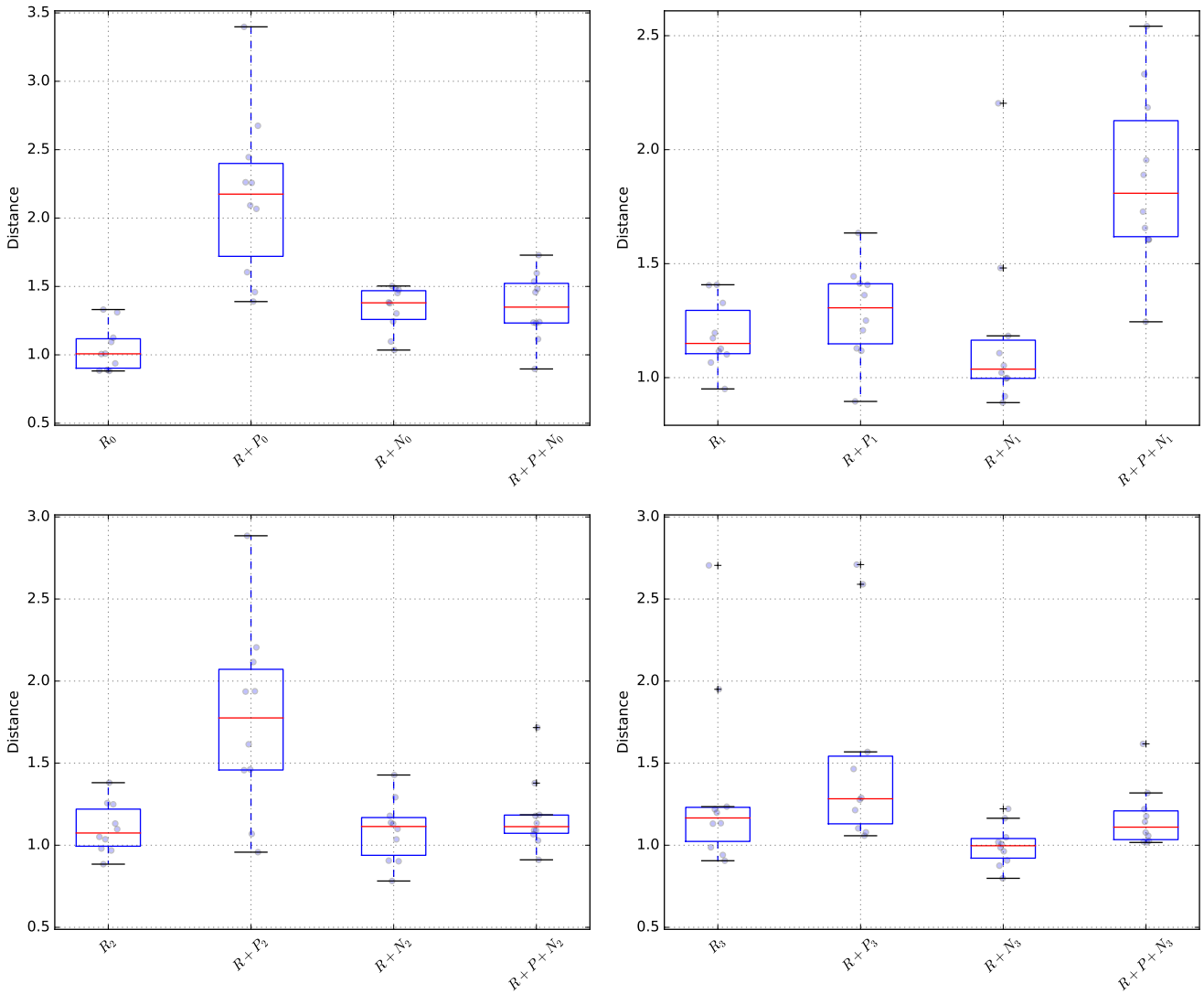

Figure S5. Mean convergence speed for 10 runs of the best hyperparameters for each condition.

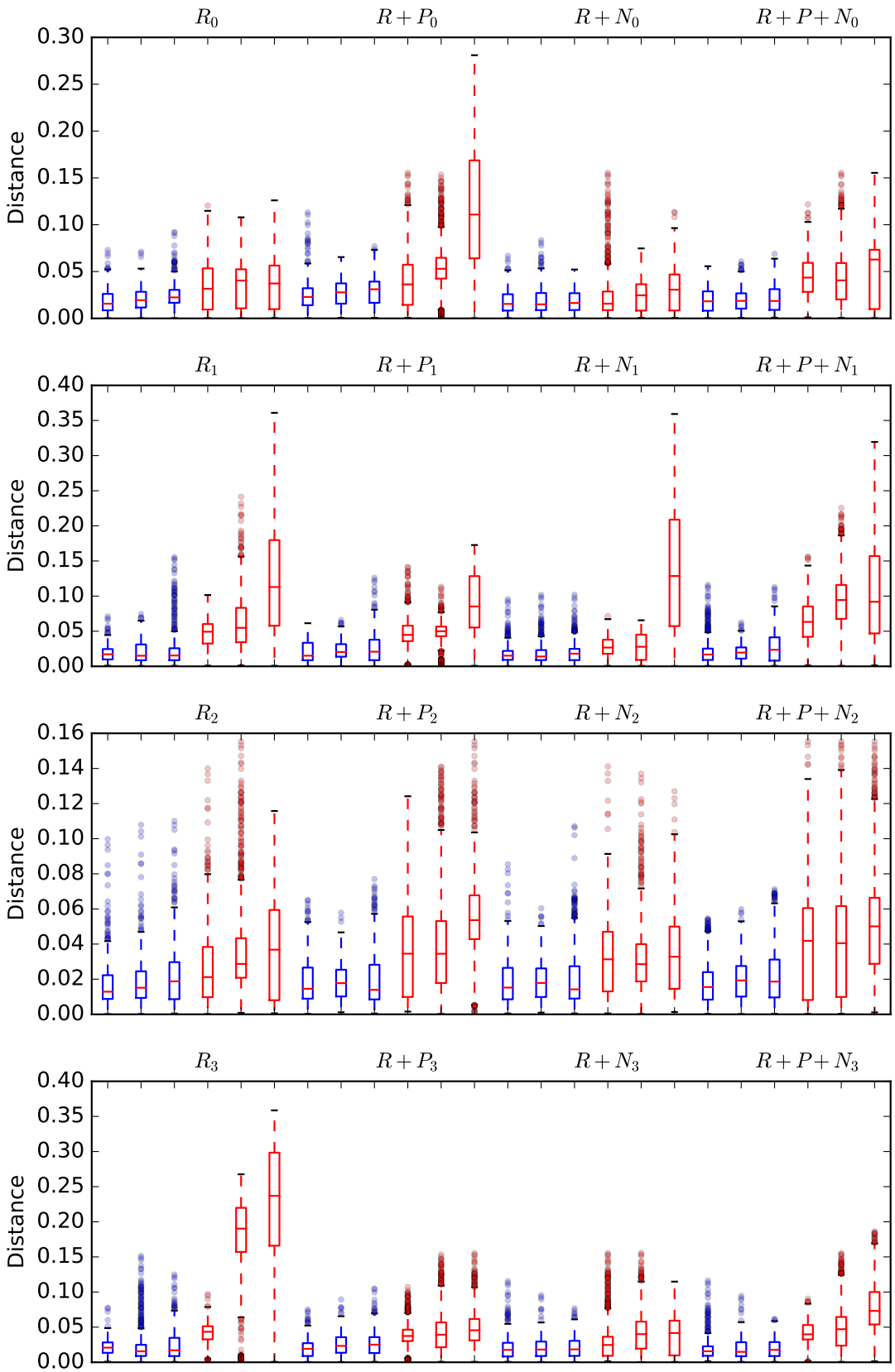

Figure S6. Performance distribution of all samples in the validation set respect to positioning error. Blue show the distribution of the 3 best runs of a condition best hyperparameters and red shows the distribution of the 3 worst runs.

2.2 Perceived Nociception

Figure S7 shows the average change of the perceived nociception (potential for damage) during learning for the four best hyperparameter sets on the validation set. The average is over 10 different random network initializations.

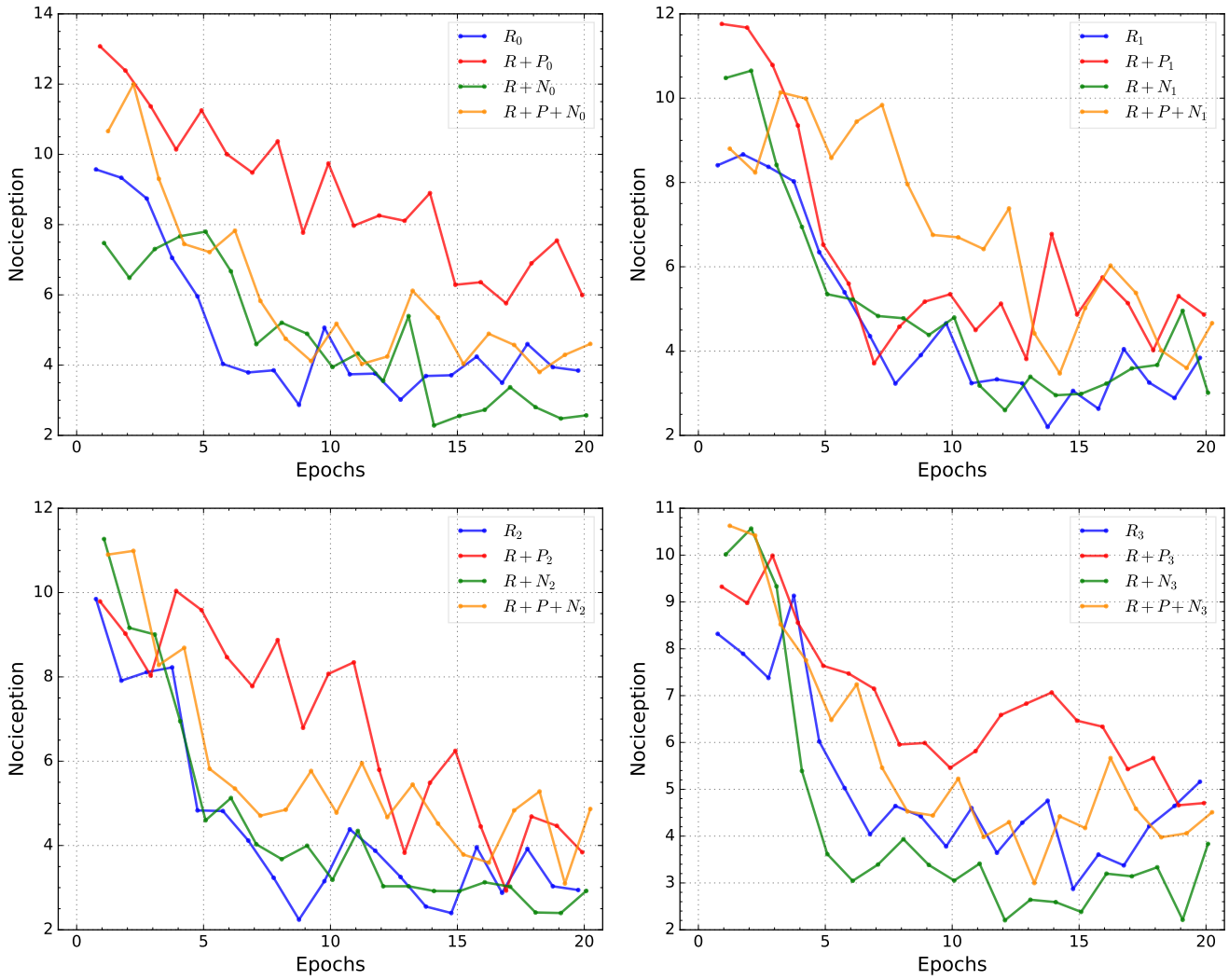

Figure S7. Mean potential for damage for 10 runs of the best hyperparameters for each condition. Organised from left to right and top to bottom are the results for the best hyperparameter set and the fourth best hyperparameter set for each condition.

Figure S9 shows the average values for the potential for damage after 20 epochs for all four conditions and the best hyperparameter sets.

Figure S11 shows the 3 best runs versus the 3 worst runs in terms of potential for damage for the best hyperparameters of each condition. All runs are sorted from smallest to largest potential for damage. The distribution of the potential for damage for all samples in the validation data set is plotted. The blue (left side for each condition) shows the distribution of the 3 best runs whereas the red (right side for each condition) shows the distribution of the 3 worst runs. Similarly as seen for

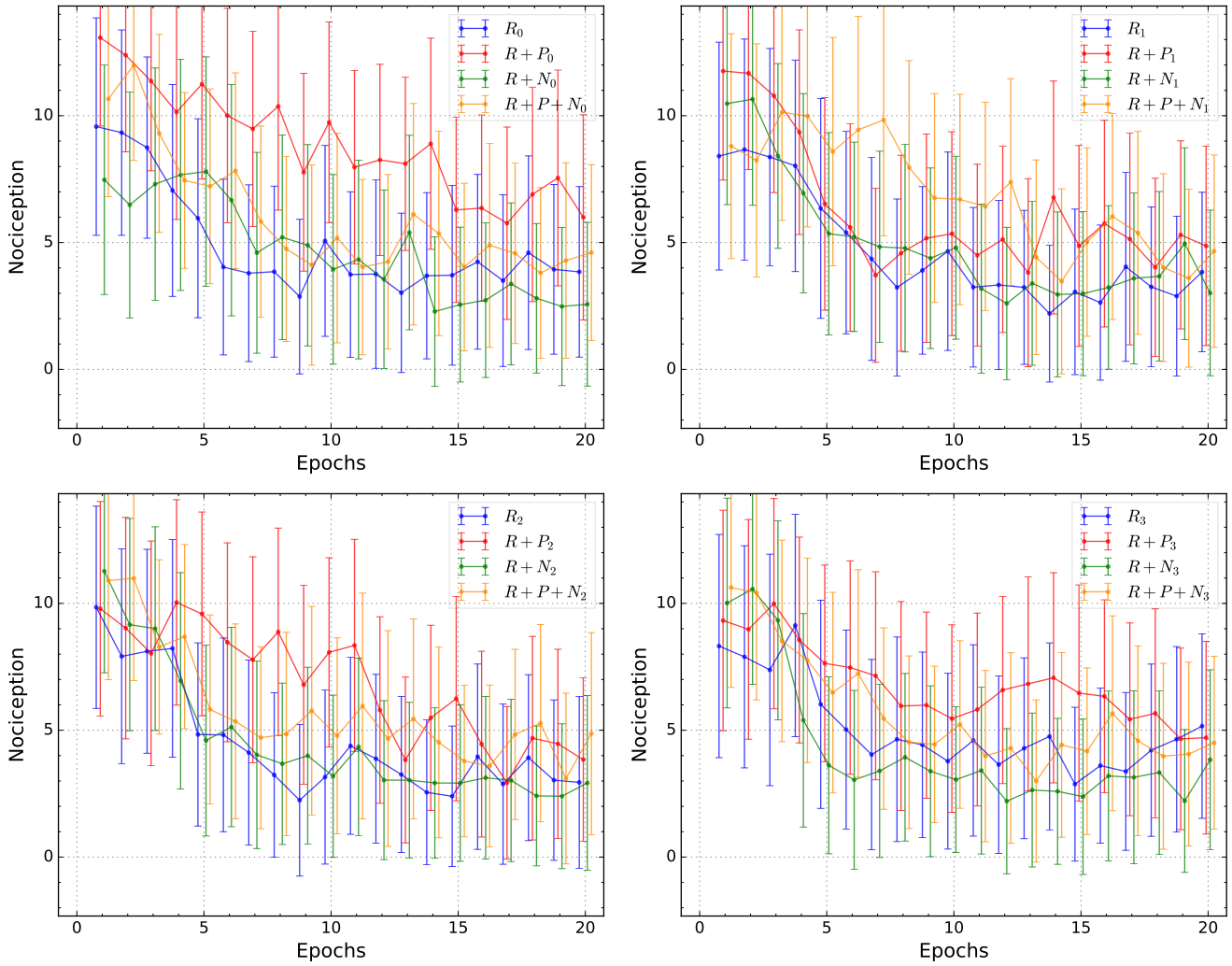

Figure S8. Mean potential for damage for 10 runs of the best hyperparameters for each condition. Organised from left to right and top to bottom are the results for the best hyperparameter set and the fourth best hyperparameter set for each condition. The same as Figure S7 but this time including standard deviation.

the metric positioning error, the distribution of the potential for damage within the validation set could be used as a test to determine what network initializations could be more favourable.

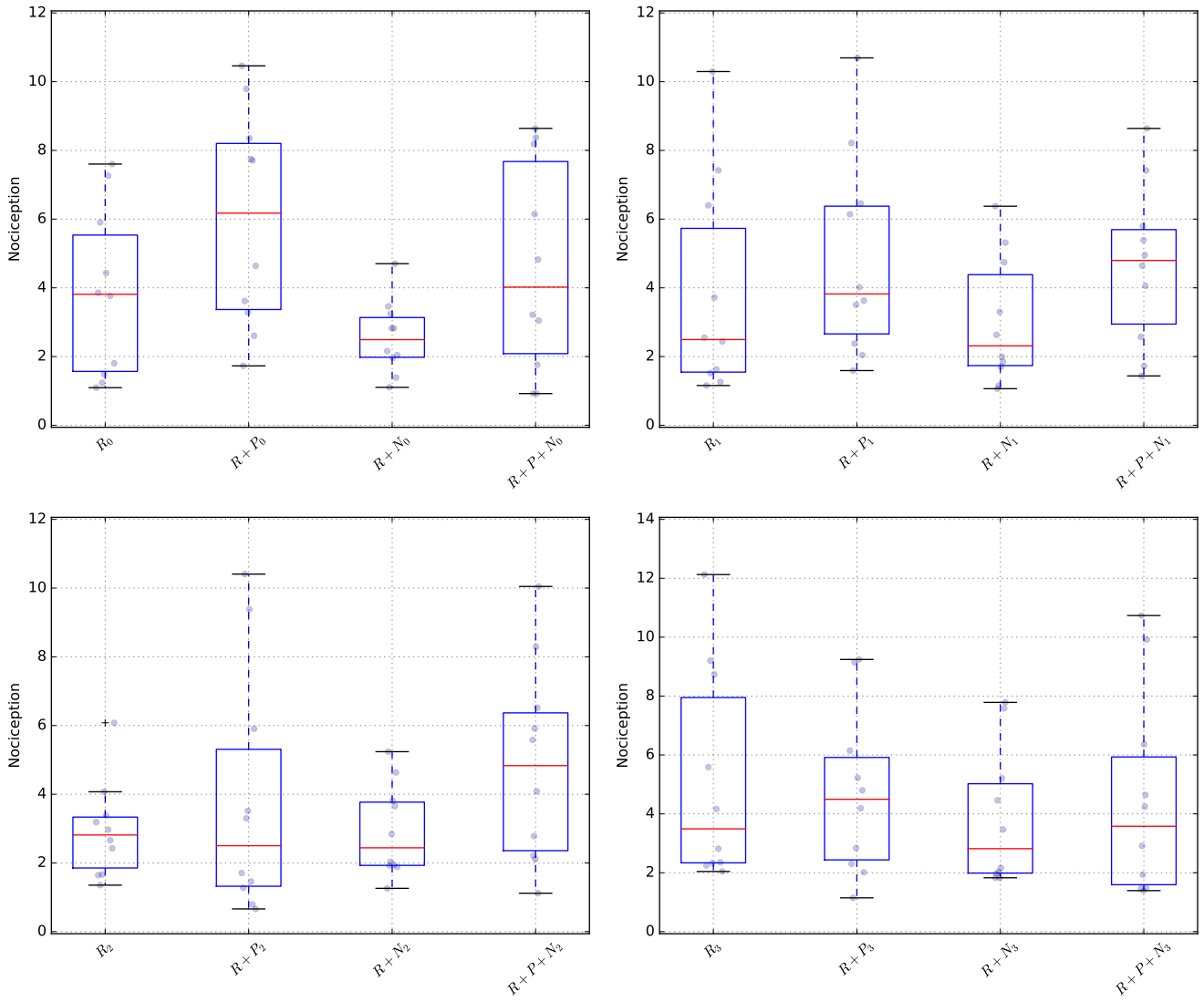

Figure S9. Mean potential for damage for 10 runs of the best hyperparameters for each condition.

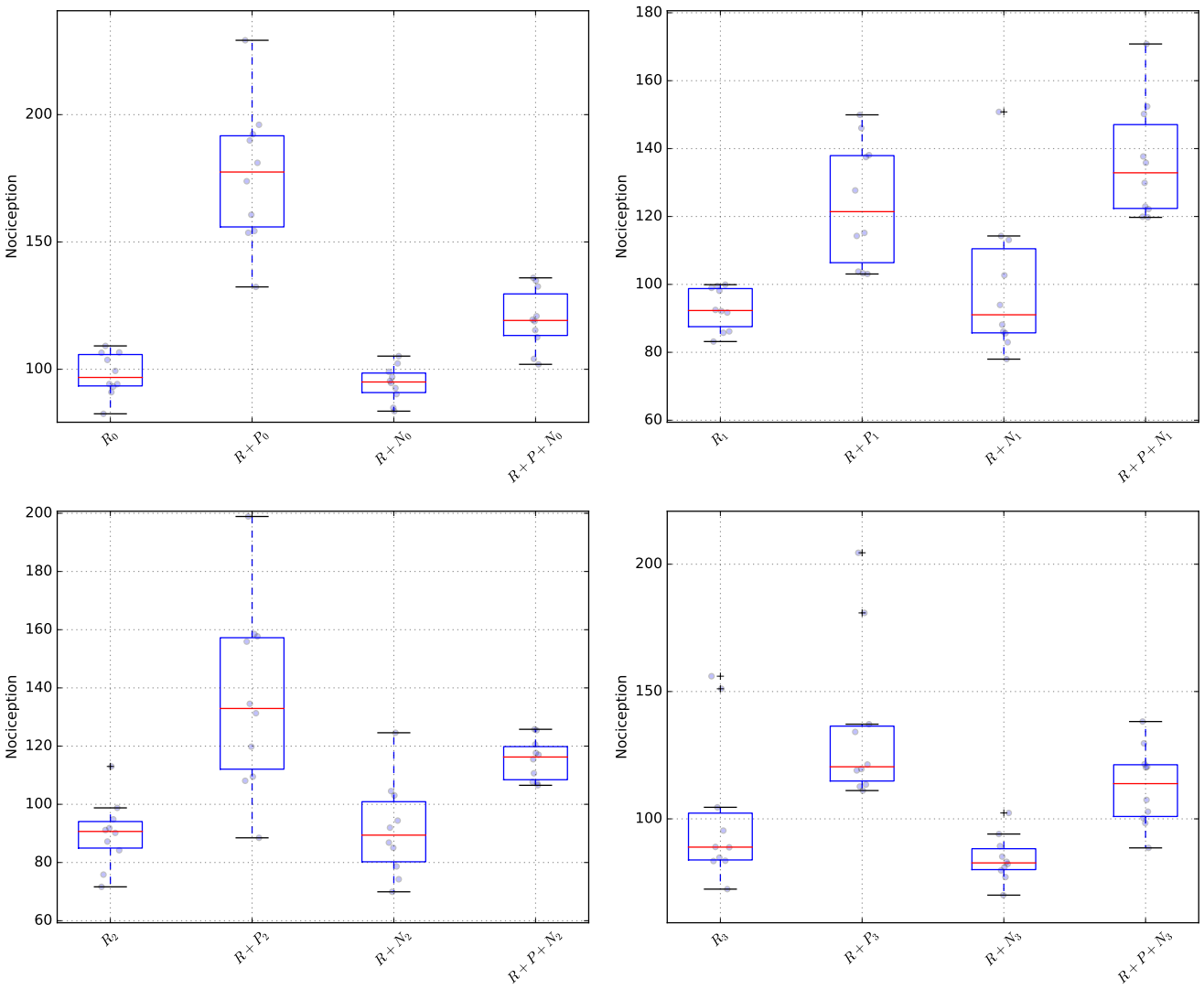

Figure S10. Mean cumulative absolute perceived nociception during learning for 10 runs of the best hyperparameters for each condition.

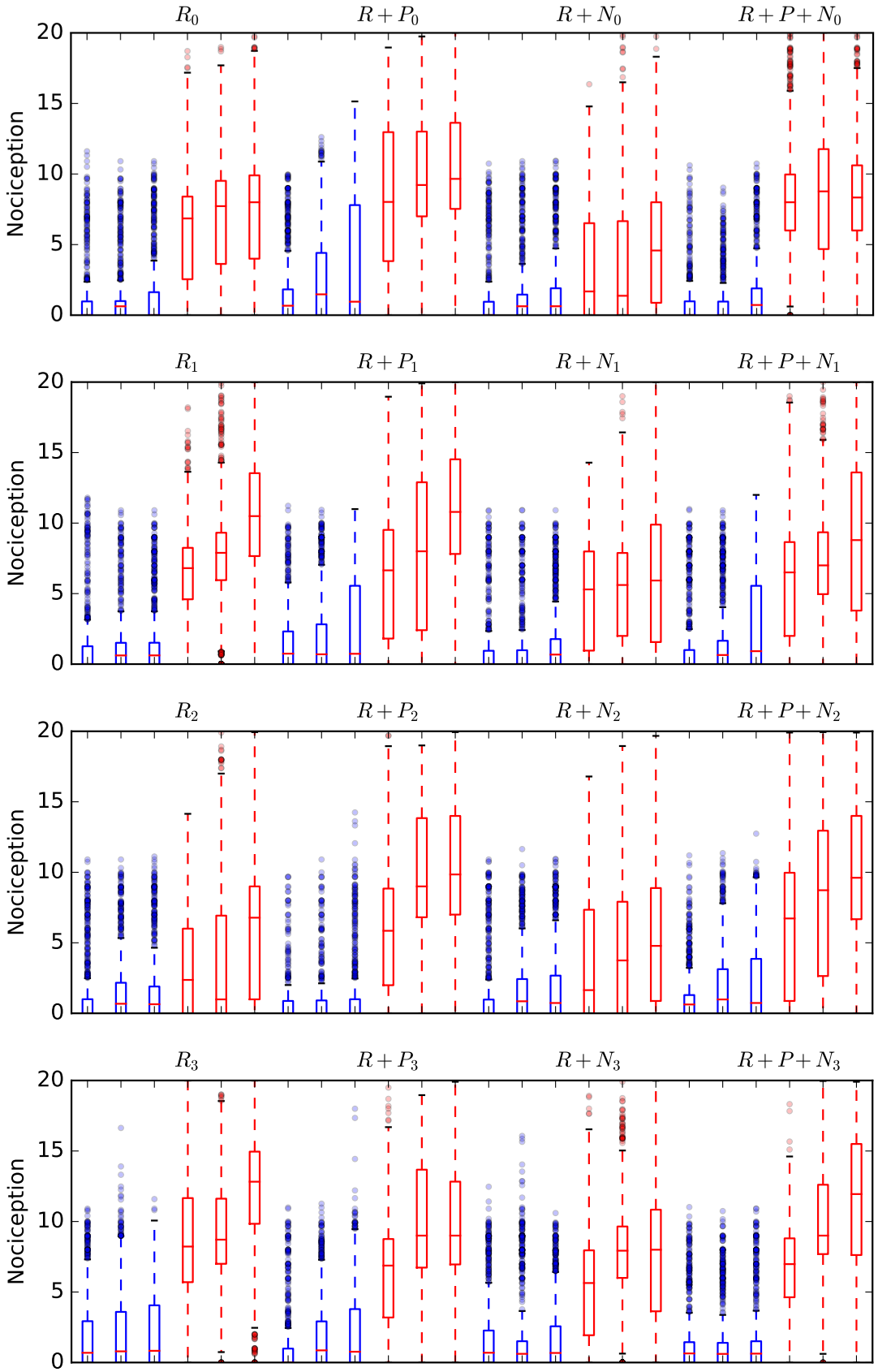

Figure S11. Performance distribution of all samples in the validation set respect to potential for damage. Blue show the distribution of the 3 best runs of a condition best hyperparameters and red shows the distribution of the 3 worst runs.

2.3 Positioning speed

Figure S12 shows the average change of the positioning speed during learning for the four best hyperparameter sets on the validation set. The average is over 10 different random network initializations.

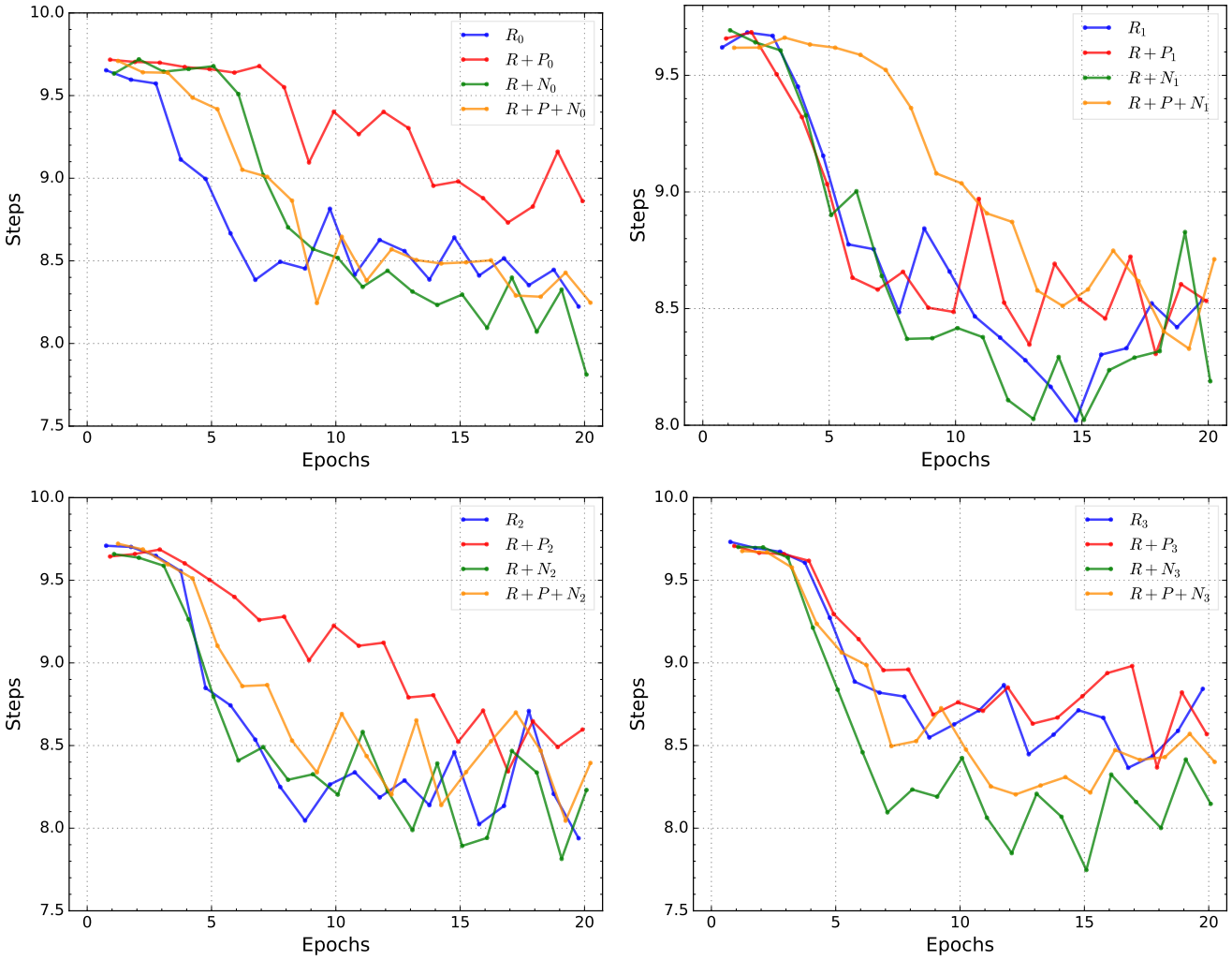

Figure S12. Mean positioning speed for 10 runs of the best hyperparameters for each condition. Organised from left to right and top to bottom are the results for the best hyperparameter set and the fourth best hyperparameter set for each condition.

Figure S14 shows the average values for positioning speed after 20 epochs for all four conditions and the best hyperparameter sets.

Figure S16 shows the 3 best runs versus the 3 worst runs in terms of positioning speed for the best hyperparameters of each condition. All runs are sorted from smallest to largest potential for damage. The distribution of the positioning speed for all samples in the validation data set is plotted. The blue (left side for each condition) shows the distribution of the 3 best runs whereas the red (right side for each condition) shows the distribution of the 3 worst runs. Similarly as seen for the metric positioning error and potential for damage, the distribution of the positioning speed within

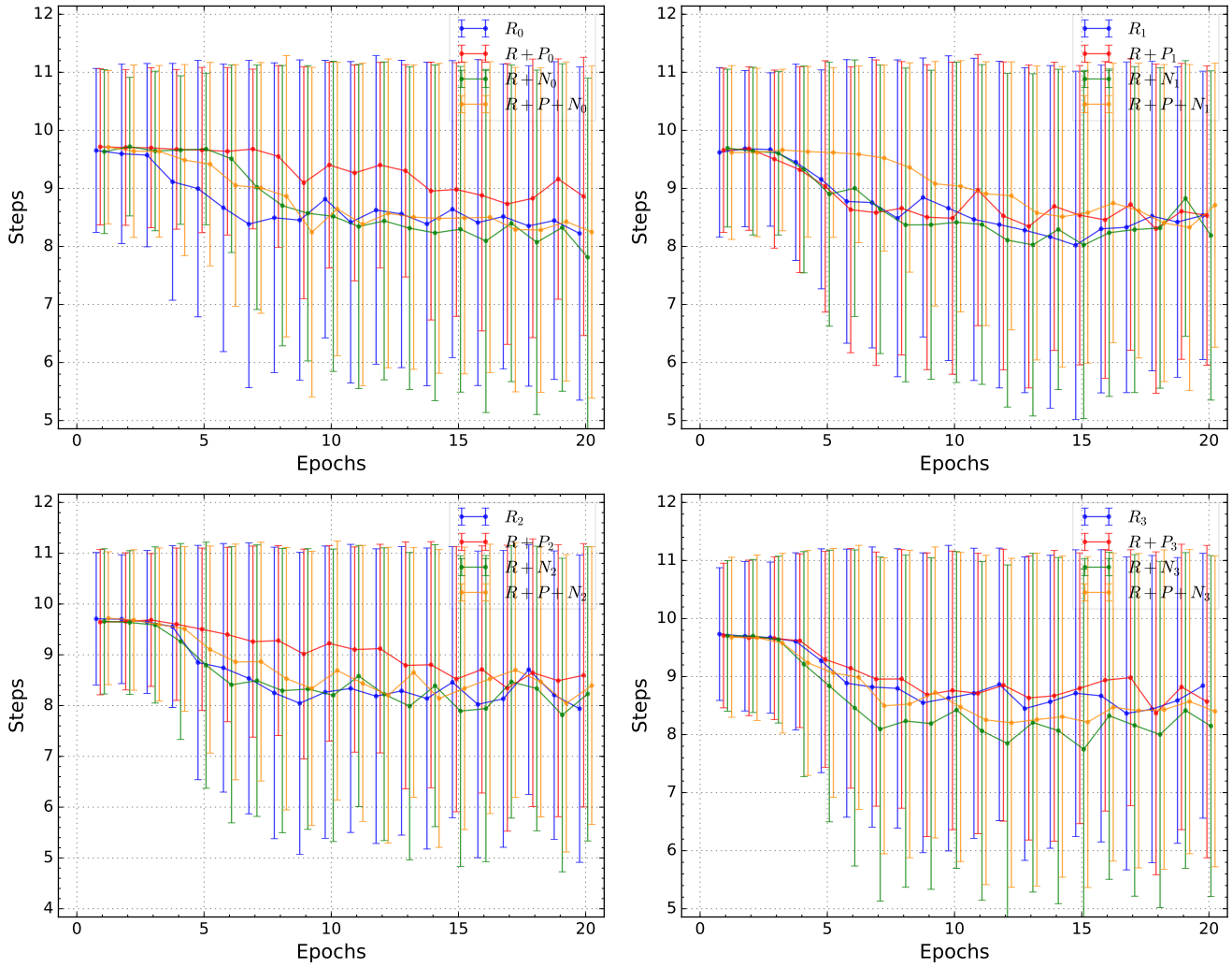

Figure S13. Mean positioning speed for 10 runs of the best hyperparameters for each condition. Organised from left to right and top to bottom are the results for the best hyperparameter set and the fourth best hyperparameter set for each condition. The same as Figure S12 but this time including standard deviation.

the validation set could be used as a test to determine what network initializations could be more favourable.

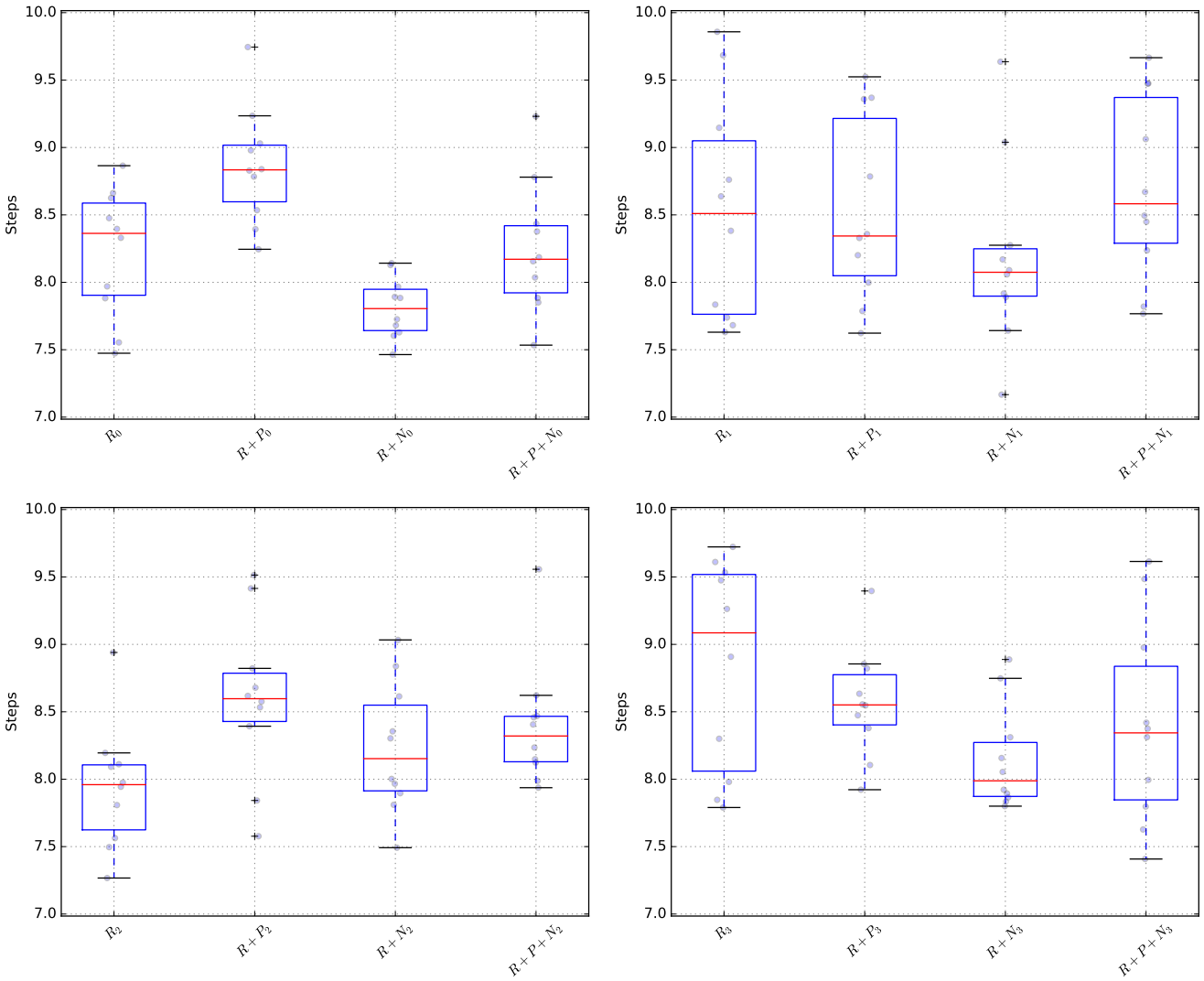

Figure S14. Mean positioning speed for 10 runs of the best hyperparameters for each condition.

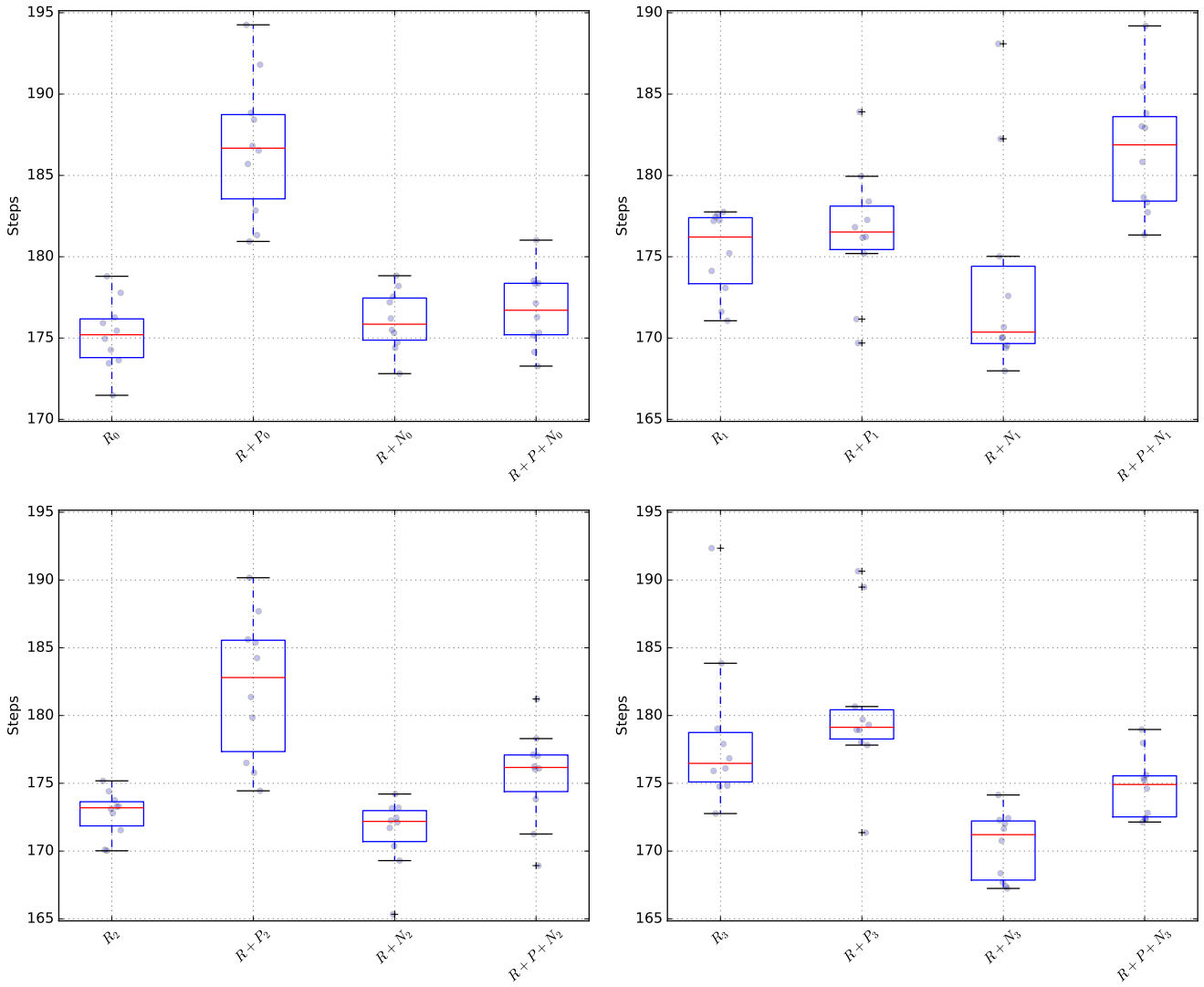

Figure S15. Mean cumulative number of steps needed during learning for 10 runs of the best hyperparameters for each condition.

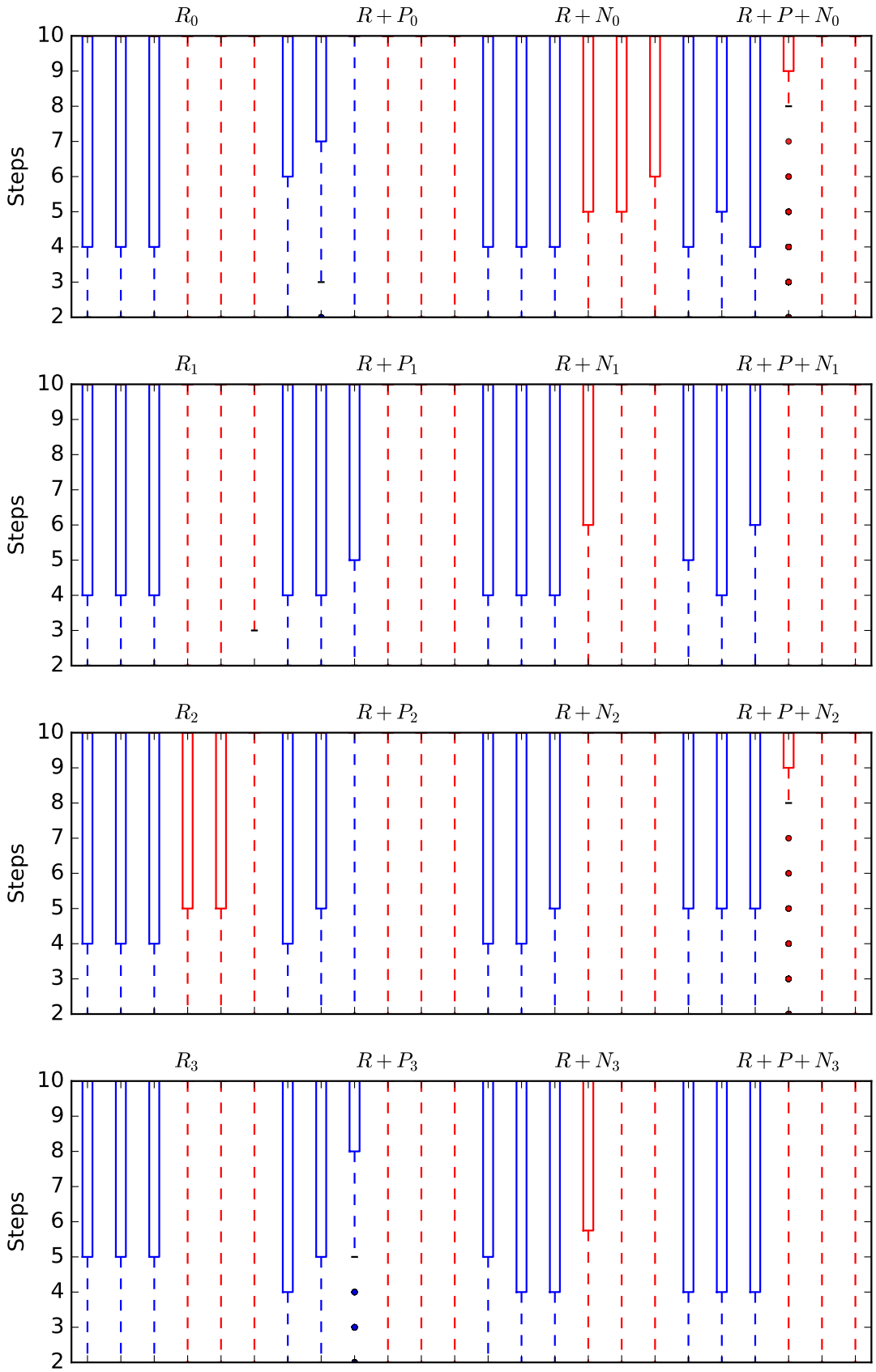

Figure S16. Performance distribution of all samples in the validation set respect to positioning speed. Blue show the distribution of the 3 best runs of a condition best hyperparameters and red shows the distribution of the 3 worst runs.