Supplementary Material: The effects on adaptive behaviour of negatively valenced signals in reinforcement learning

Nicolás Navarro-Guerrero
Universität Hamburg
Department of Informatics
Vogt-Koelln-Str. 30
22527 Hamburg, Germany
navarro@informatik.uni-hamburg.de

Robert Lowe
University of Gothenburg
Department of Applied IT
Forskningsgången 6
41296 Gothenburg, Sweden
robert.lowe@ait.gu.se

Stefan Wermter
Universität Hamburg
Department of Informatics
Vogt-Koelln-Str. 30
22527 Hamburg, Germany
wermter@informatik.uni-hamburg.de

A. *Positioning Error*

Table I shows the p-values and effect size in percentage of all 4 activation function of punishment and nociception in comparison to our baseline condition. It can be seen that in terms of performance after learning the smooth exponential function of punishment is the only one that leads to a significant improvement of approx. 40%. However, all tested activation functions for punishment significantly reduce the convergence speed by at least 15% and up to 70%. With respect to convergence speeds the linear activation function is the lesser evil but still slower than the baseline.

On the contrary, both exponential functions used as nociceptive inputs seem to converge faster than our baseline, albeit the effect size is modest. In terms of performance after learning, the abrupt exponential function used as nociceptive input is comparable to the performance of the smooth exponential function used as punishment without sacrificing convergence speed. However, non of the improvements observed for nociceptive inputs is statistically significant.

Overall, in both performance after learning as well as convergence speed the binary function for both punishment and nociception is often the lowest performing function but ironically is the most common in the literature.

B. *Perceived Nociception or Potential for Damage*

Table II shows the significance and effect size of the different punishment and nociception functions in comparison to the baseline. The smooth exponential function of punishment is the only condition that significantly reduces the potential for damage after learning, ca. 56%. However, similarly as what we observed for the task performance metric (positioning error) this comes at a cost of higher potential for damage during learning, ca. 42%. Other functions of punishment also increase the potential for damage during learning without having beneficial effect after learning. On the other hand, none of the activation functions used as nociceptive input lead to significant differences to our baseline.

TABLE I

COMPARISON AGAINST BASELINE. P-VALUES AND EFFECT SIZE FOR THE DIFFERENT ACTIVATION FUNCTIONS FOR BOTH THE AVERAGE POSITIONING ERROR AFTER LEARNING AND THE AVERAGE CUMULATIVE POSITIONING ERROR DURING LEARNING. POSITIVE PERCENTAGE (SIGNIFICANT: BLUE) IN THE EFFECT SIZE INDICATES THAT GROUP 2 IS BETTER THAN GROUP 1 AND NEGATIVE PERCENTAGE (SIGNIFICANT: RED) INDICATES THAT GROUP 1 IS BETTER THAN GROUP 2.

		Binary	Linear	$e \propto \sigma$	$e \propto 3\sigma$
Baseline compared to R+P	After	0.9873 6.67 %	0.8123 14.36 %	0.9821 7.18 %	0.0306 38.21 %
	Cum.	0.0173 −22.39 %	0.2038 −15.51 %	0.0000 −44.29 %	0.0000 −70.51 %
Baseline compared to R+N	After	0.6371 −21.79 %	0.8629 15.38 %	0.2689 31.28 %	0.8037 17.44 %
	Cum.	0.0000 −27.33 %	0.8237 −5.44 %	0.9985 1.48 %	0.8913 4.67 %
Baseline compared to R+P+N	After	0.9923 7.95 %	0.8456 18.72 %	0.9873 9.23 %	0.5106 −28.72 %
	Cum.	0.0000 −42.08 %	0.9571 4.54 %	0.0278 −19.10 %	0.0000 −63.24 %

C. *Positioning Speed*

Table III shows the significance and effect size of all tested functions of punishment and nociception with respect to their performance on positioning speed both after and during learning. All activation functions of punishment, except the smooth exponential, seem to reduce the positioning speed after learning. While all activation functions of punishment significantly reduce the positioning speed during learning, in other words, punishment leads to slower convergence speed.

On the contrary, almost all activation functions of nociception are comparable to our baseline in terms of positioning speed after learning. The abrupt exponential function of nociception seems to improve positioning speed after learning and significantly improves positioning speed during learning. While the binary activation function seems to reduce position speed.

TABLE II

Comparison against baseline. p-values and effect size for the different activation functions for both the average potential for damage after learning and the average cumulative potential for damage during learning. Positive percentage (significant: blue) in the effect size indicates that Group 2 is better than Group 1 and negative percentage (significant: red) indicates that Group 1 is better than Group 2.

		Binary	Linear	$e \propto \sigma$	$e \propto 3\sigma$
Baseline compared to *R*+*P*	After	0.6624	0.9958	0.5170	0.0009
		19.64 %	−5.47 %	−22.90 %	56.92 %
	Cum.	0.8582	0.0000	0.0000	0.0000
		5.13 %	−24.84 %	−49.86 %	−42.47 %
Baseline compared to *R*+*N*	After	0.9995	0.9696	0.5188	0.9703
		−2.99 %	−8.84 %	21.85 %	8.79 %
	Cum.	0.9982	0.2476	0.9306	0.2790
		−1.17 %	−7.89 %	3.10 %	7.64 %
Baseline compared to *R*+*P*+*N*	After	0.9991	0.9847	0.8131	0.9271
		4.21 %	8.74 %	−18.07 %	13.48 %
	Cum.	0.0000	0.9562	0.0000	0.0000
		−20.89 %	3.08 %	−26.82 %	−30.68 %

TABLE III

Comparison against baseline. p-values and effect size for the different activation functions for both the average positioning speed after learning and the average cumulative positioning speed during learning. Positive percentage (significant: blue) in the effect size indicates that Group 2 is better than Group 1 and negative percentage (significant: red) indicates that Group 1 is better than Group 2.

		Binary	Linear	$e \propto \sigma$	$e \propto 3\sigma$
Baseline compared to *R*+*P*	After	0.8127	0.8164	0.2267	0.6679
		−1.59 %	−1.58 %	−3.04 %	1.95 %
	Cum.	0.0001	0.0004	0.0000	0.0000
		−2.09 %	−1.97 %	−3.45 %	−4.90 %
Baseline compared to *R*+*N*	After	0.5139	0.9222	0.1492	0.9724
		−2.39 %	−1.25 %	3.47 %	−0.93 %
	Cum.	0.0002	0.7994	0.0071	0.7828
		−1.85 %	−0.48 %	1.45 %	0.49 %
Baseline compared to *R*+*P*+*N*	After	0.2348	0.8850	0.9951	0.0733
		−3.36 %	−1.51 %	−0.63 %	−4.20 %
	Cum.	0.0000	0.9036	0.1925	0.0000
		−3.51 %	0.41 %	−1.01 %	−3.97 %