Evolving Neuromodulated Controllers in Variable Environments

Chloe M. Barnes*, Anikó Ekárt*, Kai Olav Ellefsen†, Kyrre Glette†‡, Peter R. Lewis§ and Jim Tørresen†‡

*School of Informatics and Digital Engineering, Aston University, Birmingham, UK

{c.barnes1, a.ekart}@aston.ac.uk

†Department of Informatics, University of Oslo, Oslo, Norway

{kaiolae, kyrrehg, jimtoer}@ifi.uio.no

‡RITMO, University of Oslo, Oslo, Norway

peter.lewis@ontariotechu.ca

‡Faculty of Business and IT, Ontario Tech University, Oshawa, Canada

Abstract—**Modern technical systems are increasingly composed of heterogeneous components that are situated in variable environments. In nature, organisms can temporarily adapt their behaviour to novel stimuli with *behavioural plasticity*; this can be achieved with *neuromodulation*, a biological process that modulates synaptic activity in the brain. We explore how neuromodulation affects goal-achievement in evolved neural controllers for artificial agents in variable environments. As variability can arise from the actions of others, we show that the benefit of plasticity can increase with variability, as agents can temporarily change their phenotype within their lifetime. Further, we show that cooperation can emerge between plastic agents that cannot perceive one another in highly variable environments.**

I. INTRODUCTION

Both humans and animals alike are often situated in dynamic environments prone to change. One way that organic creatures have adapted to survive in variable environments is with *behavioural plasticity*; temporarily altering behaviour depending on environmental stimuli helps to overcome the unknown [1], [2]. Plasticity can be achieved with neuromodulation, a biological process found within the brain whereby synaptic activity is regulated in response to changing stimuli [3], resulting in short- or long-term changes in behaviour.

Modern technical systems are growing in both size and complexity, and interactions between components are evermore prevalent [4]; these systems and their components are likely to encounter unknown situations, which must be dealt with appropriately for the safety and functionality of the systems as a whole, and the environments in which they operate [5].

In previous work [6], we showed that plasticity helps artificial agents with neural controllers to solve different complexities of task in single- or multi-agent environments. In this work however, we explore how plasticity can help agents to achieve goals when others may or may not exist. This differs to our previous work [6] as we look here at how agents evolve when the environment differs between and during generations (henceforth termed 'environmental variability'), rather than how agents solve different tasks. The study is conducted using the River Crossing Dilemma testbed [7], designed to study how agent evolve to achieve goals in arbitrarily complex tasks.

Komers [2] states that the degree of plasticity increases with variability in nature. The focus of this paper is thus to explore whether a relationship between variability and plasticity is also observed in artificial agents. Specifically, we observe how neuromodulation may help agents with no capacity to perceive others to evolve in increasingly variable environments.

II. BACKGROUND AND RELATED WORK

A. Behavioural Plasticity in Nature

Behavioural plasticity, or changing behaviour in response to novel stimuli, is important for survival in variable environments [1]. Plastic behaviour is shown to lower risk of extinction in birds, than those that cannot respond to change as effectively [8]. If an organism experiences environmental changes, 'activational plasticity' [1] can facilitate behavioural changes in a short-term, immediate, and reversible way [2], rather than long-term learning. In nature, this can be achieved with *neuromodulation* – a biological process found within the brain; strengthening or weakening synapses between neurons results in synaptic plasticity being regulated based on chemical modulatory signals, the pre- and post-synaptic activities, and the stimuli [3], [9]. This causes long-term learning if synaptic strengths are changed over time, or immediate behavioural changes if synaptic activity is changed temporarily.

B. Neuromodulation and Variability in Artificial Systems

Neuromodulation has been used in computer science to facilitate learning and behavioural changes in neural networks situated in variable environments [10]–[12]. As highlighted in our previous work [6], plasticity via neuromodulation occurring *within* a single neural network has received little attention. Often, neuromodulation is used to regulate long-term *learning* [10], rather than behaviour in the short-term. These approaches tend to employ a *separate* network or group of neurons to regulate learning, or network activity [11], [12], rather than within a single neural network – like chemical signals modulating synaptic activity in the brain.

Agents that learn or evolve in variable environments have been explored in the domains of reinforcement learning [13],
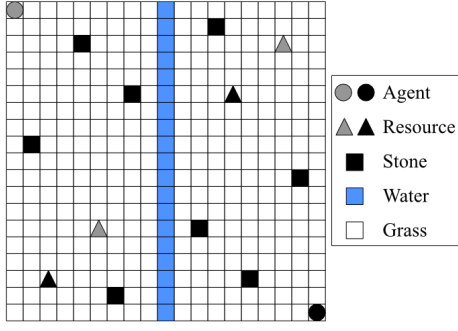
Fig. 1. The River Crossing Dilemma testbed [6]. Agents are allocated two Resources (black/grey), and can interact with all objects except other agents' Resources. Bridges can be built with two Stones to cross the river safely.
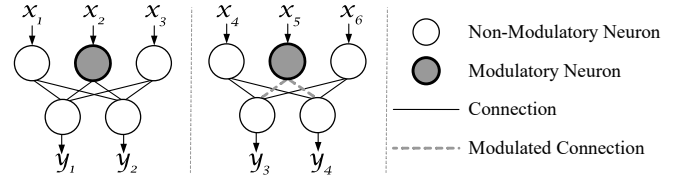

Fig. 2. Modulatory neurons modulate outgoing activity when incoming signals to the neuron are negative; when the incoming signal to x_5 is negative, signals of 0 are sent along each of its *own* outgoing connections (in this case to y_3 and y_4) which effectively 'turns off' outgoing signals. When the incoming signal to a modulatory neuron is positive, such as in x_2, activity propagates like in a standard non-modulatory neuron. Agents can thus express behavioural plasticity, as behaviour can temporarily change depending on the network inputs/stimuli – and importantly, without permanently modifying the network.

social learning [14] and robotics [15], to name a few. In nature, if full information about the environment, the actions of others, or even the presence of others is not known, evolution can be negatively affected [2]. This however is becoming increasingly commonplace in technical systems, as system components may interact unknowingly [4]. In reality, a wealth of knowledge about others is an expensive and often unrealistic luxury. In this study, we explore how agents with no knowledge of others evolve to achieve goals in variable environments. Environmental variability can arise from the actions of others, thus we explore how plasticity via neuromodulation may help agents to achieve goals when the presence of others can vary.

III. TESTBED AND AGENT DESIGN

This study is designed to explore how neuromodulation and variability affect agent evolution and goal-achievement.

A. The River Crossing Dilemma Testbed

The River Crossing Dilemma (RCD) testbed, introduced by Barnes et al. [7] is a 19×19 grid-world used to study social phenomena when agents have no *a priori* knowledge of others (Figure 1). This is one instance in a family of River Crossing testbeds, first introduced by Robinson et al. [16]. The RCD has a river of Water in the centre, with four Stones on either side; all other cells are Grass. An agent's goal is to collect its two allocated Resources from either side of the river, rewarding it with a highly positive fitness; stepping into the river gives a highly negative fitness. To achieve this goal, agents must learn how to cross the river safely by building a bridge with Stones.

B. Gamification of the RCD

The RCD is gamified with an increasing, *personal* cost for Stones placed in the river (e.g. first Stone costs 0.1, second costs 0.2, etc.). Bridges are built with two Stones as the river is two cells deep. Agents may cooperate to share the cost, exploit the other to avoid a cost, or build a bridge alone to endure the full cost. This cost means less incentive to discover bridge-building behaviour. An agent i's fitness p_i is calculated based on its own behaviour with Equation 1:

$$p_i \;=\; \frac{r_i}{N} \;-\; \left[\; \frac{C \times s_i}{2}\Big(1 + s_i\Big) \; \right] - f \qquad (1)$$

where r_i is how many Resources collected by agent i, s_i is how many Stones agent i has put in the river, and $f = 1$ if agent i drowns – or 0 if not. $N = 2$ and $C = 0.1$ are constants for how many Resources are allocated to each agent, and the cost of placing Stones in Water respectively. Fitness is 1.0 if exploiting another agent (no cost incurred), 0.7 if achieved independently, and 0.9 if cooperating to share the cost.

C. Agent Design

Agents make decisions using a two-tiered architecture introduced by Barnes et al. [6], and inspired by the original River Crossing task [16]. The first tier, the deliberative network, generates sub-goals based on the current state of the agent; this has three layers of eight, six and four hidden neurons, six input neurons, and three output neurons. Input neurons represent the agent's state: on Grass, Resource, Water, or Stone, carrying a Stone, and whether a Bridge is partially built (1 if true, 0 if false). Outputs are the sub-goals: what the agent is attracted to, will avoid, or will be neutral about (1, -1 and 0 respectively).

The second tier, the reactive network, uses the sub-goals from the previous tier to generate activity landscapes, enabling them to hill-climb towards these sub-goals. This network does not evolve, rather the activity of each neuron updates at each timestep. Details of how this works can be found in [6].

D. Operationalising Activity-Gating Neuromodulation

Neuromodulation occurs by gating activity *in* the deliberative network itself [6]; modulatory agents can thus temporarily change behaviour. Each neuron outputs the sum of each input signal multiplied by connection weight, passed through a $tanh$ activation function; this is then passed along the outgoing connections. In modulatory neurons however, if the incoming signal is negative, a signal of 0 is instead output along each of its *own* outgoing connections to the next layer. (Figure 2). One genotype (deliberative network) can thus represent many phenotypes depending on the network inputs (environmental stimuli), without changing the network. Neurons in the deliberative network may evolve to be non-modulatory or modulatory.

E. Evolutionary Algorithm

Experiments are conducted using the RCD testbed with the following setup, inspired by Barnes et al. [6]. A Steady State

Genetic Algorithm [17] is used to evolve a population of 25 randomly initialised agents in each experiment. At each generation, a tournament of three randomly selected agents are evaluated for 500 timesteps. The worst-performer in the tournament is replaced with the offspring of the best two. This offspring has a probability of $P_{one} = 0.95$ to inherit each chromosome (layer of weights in the deliberative network) from a random parent, otherwise single-point crossover is used; each weight w is then mutated by a random value from a Gaussian distribution with $\mu = w$ and $\sigma = 0.01$.

Modulatory agents are initialised with only non-modulatory neurons in the deliberative network, but evolve the types of hidden neurons (non-/modulatory). At each generation, the offspring inherits the deliberative network configuration from a random parent; further, there is a probability of $P_{mut} = 0.15$ that one randomly chosen neuron in one of the three hidden layers of the deliberative network will be mutated from non-modulatory to modulatory, or vice versa. Non-modulatory agents do not evolve the neurons in their deliberative network.

IV. Experimental Design

Each experiment is repeated 100 times; randomly initialised agents are evolved with or without neuromodulation for 500,000 generations in the RCD testbed. Agents do not 'learn' in their lifetime, only in the breeding process; an agent's genetic code (deliberative network weights and configuration) remains unchanged in each generation. Agents are evaluated on one, two, or four environments at every generation to observe how they evolve in variable environments. The experiments below are repeated twice, where agents evolve with either *coevolved*, or *random* partners at each generation. The actions of random agents are unpredictable and thus increase variability compared to a coevolved partner.

The first experiments explore how plasticity affects agent evolution in a shared environment with one other agent.

The second experiments explore how agents may retain goal-achieving behaviour when the environment can be shared or solitary; here, agents are evaluated on two consecutive environments at every generation (firstly alone, then with another). By evolving agents in both a single-agent and a multi-agent environment at each generation, they have the opportunity to evolve appropriate behaviours that are successful whenever they are alone or with a partner; this could lead to more generalised behaviour being observed.

The final experiments evaluate agents on four consecutive environments at each generation, exposing them to more variability. Agents exist alone in the first and third environments, and share the second and fourth environments with another agent. The partners in shared environments are different to one another, to increase variability further. Evaluating agents in two environments where they exist alone may seem redundant, however agent movement is stochastic so they would experience more variability in four compared to three environments.

An agent's fitness at each generation is the sum of fitnesses achieved in each individual environment it is evaluated on (Equation 1); the maximum fitness for agents evolving in one, two and four environments is thus 1.0, 1.7 and 3.4, since agents can only receive a maximum fitness of 0.7 when alone.

V. Results

A. Evolving with a Coevolved Partner

Here, the impact that sharing an environment with another agent can have on evolution is explored. As these agents coevolve in tandem, they face the least variability during evolution as partners become predictable. Figure 3a shows the mean best-in-population fitness of agents evolving with and without neuromodulation. Modulatory agents have a higher mean best-in-population fitness than non-modulatory agents throughout evolution. Further, 68% of non-modulatory agents receive a goal-achieving fitness at the end of evolution compared to 94% of modulatory agents, when they coevolve in one environment (Table I). This indicates that more agents are able to achieve their goal with neuromodulation than without.
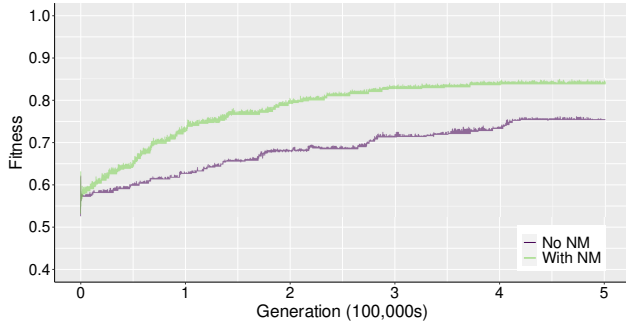
B. Evolving with Random Partners

These experiments explore how evolving with random partners can affect agent evolution. Each random partner can potentially affect the environment in an unpredictable way, resulting in more variable conditions. The mean best-in-population fitness of agents, with and without neuromodulation, is presented in Figure 3b. Only 60% of non-modulatory agents achieve their goal after evolving with random partners, compared to 86% of modulatory agents (Table I). Plasticity is thus observed to be beneficial to agents that experience more variability when evolving with unpredictable partners.

The predictability of another agent's actions can dramatically affect the evolution and fitness of agents. The line graph in Figure 3b indicates that the actions of random partners create high levels of variability. The fitness fluctuates often in each individual run, which means the overall mean fitness can vary by large amounts *between* generations. Despite this, the mean best-in-population fitness is higher earlier on in evolution when agents evolve with random partners, than when agents coevolve. Interestingly, this is not a result of more agents achieving their goal when evolving with random partners compared to coevolved partners – in fact, the opposite is true when looking at the data in Table I. There is however a shift from agents that tend to exploit the other or achieve their goals alone when coevolving, to those that cooperate with random partners. Exploitation is more risky when partners are random, as not all partners may perform behaviour that can be exploited; this would leave the agent unable to achieve its goal. Evolving with random partners seems to incentivise agents to evolve goal-achieving behaviour that involves either cooperation or exploitation. This is an interesting observation, as the agents have no capacity to perceive others.

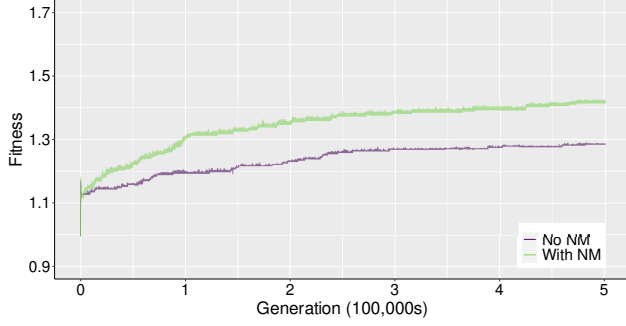### C. Evolving in Two Environments with a Coevolved Partner

Here, agents are evaluated on two RCD instances at each generation. Agents thus experience more variability than previous experiments due to evolving in both single- and multi-agent environments. The mean best-in-population fitness is
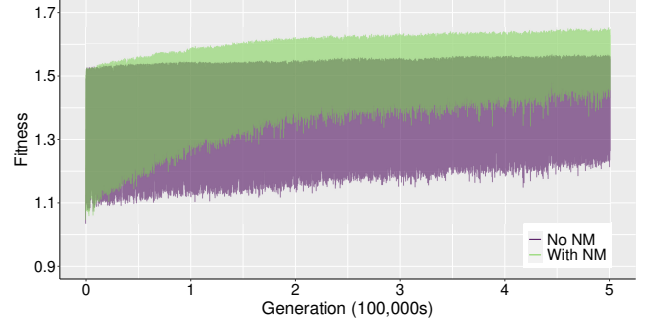
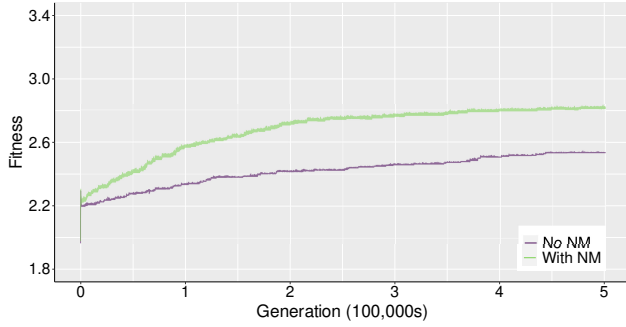(a) One Environment with a Coevolved Partner
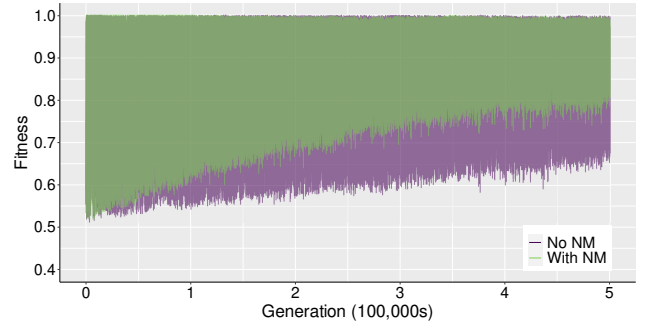
(b) Evolving in One Environment with Random Partners
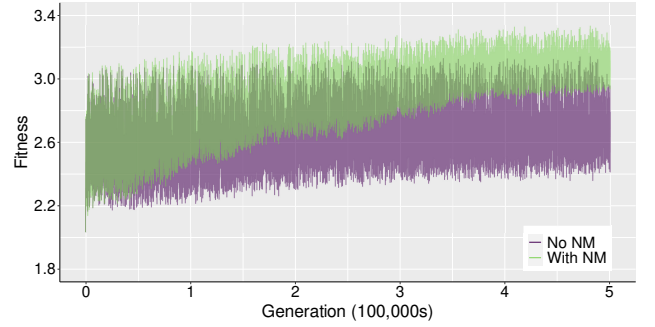

(c) Two Environments with a Coevolved Partner

(d) Two Environments with Random Partners

(e) Four Environments with Coevolved Partners

(f) Four Environments with Random Partners

Fig. 3. The mean best-in-population fitnesses of agents in the RCD evolving for 500,000 generations, with and without neuromodulation (NM). At each generation, agents are evaluated on: one ((a) and (b)), two ((c) and (d)), or four ((e) and (f)) environments, with coevolved ((a), (c) and (e)) or random ((b), (d) and (f)) partners. For (a) and (b), agents share one environment so the maximum fitness is 1.0. For (c) and (d), agents evolve alone, then in a shared environment; the maximum fitness is 1.7. For (e) and (f), agents evolve in an environment: alone, shared, alone, then shared; the maximum fitness is 3.4. A fitness of: 0.7 indicates the goal is achieved individually; 0.9 is cooperation; 1.0 is exploitation; ≥ 0.7 is a goal-achieving fitness (Equation 1). *Note: y-axis scales between Figures 3a and 3b, 3c and 3d, and 3e and 3f are comparable, but are not comparable otherwise as agents are evaluated on different numbers of environments and thus have different maximum fitnesses. Also note the data for (b), (d) and (f) are downsampled by a factor of 10.*

higher in modulatory agents than in non-modulatory agents (Figure 3c). Further, 37% of non-modulatory agents achieve their goal in both environments, increasing to 66% in modulatory agents (Table I). 37% of non-modulatory agents achieve their goal in the first environment when alone compared to 66% of modulatory agents; in the second, shared environment, this rises to 65% and 96% respectively. Agents are more successful in shared environments as they can capitalise on the actions of others to receive a higher fitness. Ultimately, this shows plasticity has a positive effect on the success rate of agents that both exist alone, and coevolve with another.

D. Evolving in Two Environments with Random Partners

Here, agents are evaluated firstly in an environment alone, then in an environment shared with random partners; the unpredictable actions of the partners increases variability. The line graph depicting the mean best-in-population is erratic as fitness fluctuates often during evolution (Figure 3d); this indicates that evolving successful behaviour both when alone and with random partners is hard, as each partner behaves differently. The mean best-in-population fitness is greater in modulatory agents than in non-modulatory agents. Further, only 28% of non-modulatory agents achieve their goal in both

Exp	NM	Env	Fitness (% of Runs)				Goals Achieved				
			0.7	0.9	1.0	≥ 0.7	0	1	2	3	4
1C	No	1	27	5	36	68	32	68	–	–	–
	Yes	1	44	0	50	94	6	94	–	–	–
1R	No	1	2	29	29	60	40	60	–	–	–
	Yes	1	3	59	24	86	14	86	–	–	–
2C	No	1	37	0	0	37	35	28	37	–	–
		2	37	2	26	65					
	Yes	1	66	0	0	66	4	30	66	–	–
		2	66	0	30	96					
2R	No	1	33	0	0	33	65	7	28	–	–
		2	5	25	0	30					
	Yes	1	77	0	0	77	23	10	67	–	–
		2	10	57	0	67					
3C	No	1	26	0	0	26	25	31	18	0	26
		2	26	3	33	62					
		3	26	0	0	26					
		4	26	3	28	57					
	Yes	1	66	0	0	66	2	8	24	0	66
		2	65	1	26	92					
		3	66	0	0	66					
		4	66	0	30	96					
3R	No	1	31	0	0	31	64	2	3	4	27
		2	7	24	0	31					
		3	32	0	0	32					
		4	21	13	0	34					
	Yes	1	86	0	0	86	13	1	1	6	79
		2	12	68	0	80					
		3	85	0	0	85					
		4	63	23	0	86					

environments after evolution, compared to 67% of modulatory agents (Table I). In the first environment, 33% of non-modulatory agents were successful compared to 77% of modulatory agents. In the second environment shared with random partners, only 30% of non-modulatory agents were successful compared to 67% of modulatory agents. More modulatory agents were successful in each individual environment than non-modulatory agents, and more agents were successful in the first environment than the second environment; this shows that evolving successful behaviour that generalises across multiple environments and partners is difficult.

E. Evolving in Four Environments with Coevolved Partners

Here, agents are evaluated on four environments at each generation. Figure 3e shows the mean best-in-population fitness over time is higher in modulatory agents than non-modulatory agents. 75% of non-modulatory agents were successful in at least one environment, however only 26% achieved their goal in all four environments (Table I). In

comparison, 98% of modulatory agents were successful in at least one environment, but only 66% were successful in all four. Modulatory agents can therefore be expected to achieve their goal in more environments than non-modulatory agents, and are also more likely to succeed in all four environments.

F. Evolving in Four Environments with Random Partners

Here, agents are evaluated on four environments at each generation. The variability arising from the unpredictable actions of random partners causes the mean best-in-population fitness to fluctuate often (Figure 3f). Only 36% of non-modulatory agents were successful in at least one environment, and 27% in all four; this is compared to 87% of modulatory agents and 79% respectively (Table I). Modulatory agents more likely to achieve their goal not only in each individual environment, but in all four. Agent fitness fluctuates more, but is higher overall, when agents face high variability caused by random partners (Figures 3e and 3f); this is also true when agents are evaluated on one, two, or four environments at each generation. This shows that the increased variability that comes with evolving with random partners helps agents to evolve behaviour that is useful in a range of environments.

G. Discussion and Further Analysis

Table I shows a qualitative shift in evolved goal-achieving behaviour between agents that evolve with random or coevolved partners. Specifically, exploitation is more prevalent when the actions of an unknown, coevolving partner causes the environment to become more predictable. However, when variability arises through evolving with random partners, cooperation is favoured. Agents evolving random partners are shown to increase their fitness whenever possible, by cooperating with other agents unintentionally. This shows that agents in highly variable environments are likely to evolve behaviour that enables them to achieve higher fitnesses, without affecting their ability to succeed when conditions inevitably change.

Statistical tests were conducted to compare the best-in-population fitnesses achieved after evolution by modulatory and non-modulatory agents. A Shapiro-Wilk test for normality was conducted for each experiment, since it is powerful for a range of distributions [18]. These results were significant as $p < 0.05$ for each distribution, indicating non-normality. Consequently, one two-tailed and two one-tailed Wilcoxon Signed Rank statistical tests were conducted for each experiment to ascertain whether plasticity has any effect on agent fitness (Table II). The median best-in-population agent fitness after evolution of non-modulatory agents is significantly lower than modulatory agents in each experiment ($p < 0.05$, $m_n < m_m$). Behavioural plasticity thus positively affects the fitness that agents receive – especially when they are evaluated on increasing numbers of environments.

The effect size estimate r can measure the magnitude of such an effect, and relationship between two variables (Table II); this is calculated with $r = \frac{Z}{\sqrt{N}}$, where Z is the z-score and $N = 100$ for the number of observations. As r is

Exp	Statistical Test Alternative Hypothesis			r
	$m_n \neq m_m$	$m_n < m_m$	$m_n > m_m$	
1C	1.594×10^{-2} *	7.970×10^{-3} *	0.9922	-0.2413
1R	1.871×10^{-3} *	9.355×10^{-4} *	0.9991	-0.3112
2C	2.331×10^{-6} *	1.165×10^{-6} *	1	-0.4725
2R	1.248×10^{-6} *	6.239×10^{-7} *	1	-0.4851
3C	4.243×10^{-9} *	2.121×10^{-9} *	1	-0.5877
3R	1.951×10^{-11}*	9.757×10^{-12}*	1	-0.6712

negative, modulatory agents receive higher fitnesses than non-modulatory agents; this effect is stronger when agents evolve with random partners, and evaluated on more environments, as variability increases. The effect that plasticity has on agent fitness increases with variability; $r = -0.2413$ and $r = -0.6712$ in the least and most variable environments respectively. This shows a correlation between plasticity and variability, where the benefit of neuromodulation increases with variability. This benefit is felt more strongly by agents evolving with random partners, as the actions of these partners increase variability.

We show that regulating behaviour is beneficial to agents in variable environments, as agents can change their phenotype (behaviour) without affecting their genotype (deliberative network); behavioural changes are short-lived.

VI. CONCLUSION

In nature, behavioural plasticity is one example of how animals can survive in variable environments [1]. In this paper, we explore the effect of plasticity on agent evolution and goal-achievement in the River Crossing Dilemma testbed, when agents cannot perceive others. Here, variability can arise from the actions or presence of other agents. It is harder for agents to maintain goal-achieving behaviour when other agents are unpredictable. Whilst evolving in highly variable environments is challenging, we show that plasticity can increase agent fitness across the study. Further, cooperation is more prevalent when agents experience more variability, and exploitation with less variability. These findings are important considering that technical systems are increasing in size and complexity, and thus variability. We show that the benefit of plasticity increases with variability – even when other agents are unknown.

This study demonstrates the importance of considering how systems may overcome unforeseen interactions or situations. Technical systems comprise many interrelated components; a limitation of this study is that the abstracted agent-based model studies how only a maximum of two agents interact. Future studies will explore how many agents achieve goals in highly variable environments, and how the consequences of increased unintended interactions can be mitigated. We use neuromodulation here to increase plasticity in agents, meaning they can temporarily change their sub-goals if network activity is modulated; agents may, for example, become attracted to Stones – crucial for achieving the goal – without explicitly learning the importance of doing so. These behavioural switches will be explored qualitatively in the future.

REFERENCES

[1] E. C. Snell-Rood, "An overview of the evolutionary causes and consequences of behavioural plasticity," *Animal Behaviour*, 2013.

[2] P. E. Komers, "Behavioural plasticity in variable environments," *Canadian Journal of Zoology*, vol. 75, no. 2, pp. 161–169, 1997.

[3] L. F. Abbott, "Modulation of function and gated learning in a network memory," *Proc. of the National Academy of Sciences of the United States of America*, vol. 87, no. 23, pp. 9241–9245, 1990.

[4] J. Hähner, U. Brinkschulte, P. Lukowicz, S. Mostaghim, B. Sick, and S. Tomforde, "Runtime self-integration as key challenge for mastering interwoven systems," in *Proc. of the 28th Intl. Conf. on Architecture of Computing Systems (ARCS)*. VDE, 2015, pp. 1–8.

[5] A. Burger, D. W. King, and G. Schiele, "Reconfigurable embedded devices using reinforcement learning to develop action-policies," in *Proc. of the 1st IEEE Intl. Conf. on Autonomic Computing and Self-Organizing Systems (ACSOS)*, 2020, pp. 232–241.

[6] C. M. Barnes, A. Ekárt, K. O. Ellefsen, K. Glette, P. R. Lewis, and J. Tørresen, "Coevolutionary learning of neuromodulated controllers for multi-stage and gamified tasks," in *Proc. of the IEEE 1st Intl. Conf. on Autonomic Computing and Self-Organizing Systems (ACSOS)*. IEEE, 2020, pp. 129–138.

[7] C. M. Barnes, A. Ekárt, and P. R. Lewis, "Social action in socially situated agents," in *Proc. of the IEEE 13th Intl. Conf. on Self-Adaptive and Self-Organizing Systems (SASO)*. IEEE, 2019, pp. 97–106.

[8] S. Ducatez, D. Sol, F. Sayol, and L. Lefebvre, "Behavioural plasticity is associated with reduced extinction risk in birds," *Nature Ecology & Evolution*, vol. 4, no. 6, pp. 788–793, 2020.

[9] L. F. Abbott and S. B. Nelson, "Synaptic plasticity: Taming the beast," *Nature Neuroscience*, vol. 3, no. 11, pp. 1178–1183, 2000.

[10] K. O. Ellefsen, J. B. Mouret, and J. Clune, "Neural modularity helps organisms evolve to learn new skills without forgetting old skills," *PLoS Computational Biology*, vol. 11, no. 4, pp. 1–24, 04 2015.

[11] S. Beaulieu, L. Frati, T. Miconi, J. Lehman, K. O. Stanley, J. Clune, and N. Cheney, "Learning to continually learn," in *Proc. of the 24th European Conf. on Artificial Intelligence (ECAI)*. IOS Press, 2020, pp. 992–1001.

[12] N. Vecoven, D. Ernst, A. Wehenkel, and G. Drion, "Introducing neuromodulation in deep neural networks to learn adaptive behaviours," *PLOS ONE*, vol. 15, no. 1, pp. 1–13, 01 2020.

[13] K. K. Ndousse, D. Eck, S. Levine, and N. Jaques, "Emergent social learning via multi-agent reinforcement learning," in *Proc. of the 38th Intl. Conf. on Machine Learning*, ser. Proc. of Machine Learning Research, M. Meila and T. Zhang, Eds., vol. 139. PMLR, 7 2021, pp. 7991–8004.

[14] B. P. Jolley, J. M. Borg, and A. Channon, "Analysis of social learning strategies when discovering and maintaining behaviours inaccessible to incremental genetic evolution," in *Lecture Notes in Computer Science*, vol. 9825 LNCS, 2016, pp. 293–304.

[15] J. T. Carvalho and S. Nolfi, "Behavioural plasticity in evolving robots," *Theory in Biosciences*, vol. 135, no. 4, pp. 201–216, 2016.

[16] E. Robinson, T. Ellis, and A. Channon, "Neuroevolution of agents capable of reactive and deliberative behaviours in novel and dynamic environments," in *Advances in Artificial Life*. Springer, 2007, pp. 1–10.

[17] G. Syswerda, "Uniform crossover in genetic algorithms," in *Proc. of the 3rd Intl. Conf. on Genetic Algorithms*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1989, pp. 2–9.

[18] B. W. Yap and C. H. Sim, "Comparisons of various types of normality tests," *Journal of Statistical Computation and Simulation*, vol. 81, no. 12, pp. 2141–2155, 2011.