MSPRT action selection model for bio-inspired autonomous driving and intention prediction

Riccardo Donà[1], Gastone Pietro Rosati Papini [1] and Giammarco Valenti[1]

Abstract— **This paper proposes the usage of a bio-inspired action selection mechanism, known as multi-hypothesis sequential probability ratio test (MSPRT), as a decision making tool in the field of autonomous driving. The focus is to investigate the capability of the MSPRT algorithm to effectively select the optimal action whenever the autonomous agent is required to drive the vehicle or, to infer the human driver intention when the agent is acting as an intention prediction mechanism. After a brief introduction to the agent, we present numerical simulations to demonstrate how simple action selection mechanisms may fail to deal with noisy measurements while the MSPRT provides the robustness needed for the agent implementation on the real vehicle.**

I. INTRODUCTION

Autonomous vehicles (AVs) require effective algorithms to perform robust decision making in the shortest time frame possible. Indeed, in a dynamic environment such as the one faced by the AVs, the capability of reacting promptly is a major factor in potentially avoiding collisions wand saving lives. The inherent complexity of the process is worsened by the presence of sensors' noise and uncertainties which affect the way the behavioural level selects the proper action. In the early days of autonomous driving, tactical/behavioral level planning typically relied on manually engineered state machines, this approach has been adopted by many competitors of the 2007 DARPA Grand Challenge (a.k.a. *urban challenge*) [1], [2]. Despite some participants actually managed to succeed, state machines inherently lack the capability of safely generalizing to unmodeled scenarios. More recent autonomous driving softwares are built on top of probabilistic approaches including Markov Decision Process [3] or machine learning-based techniques such as behaviour networks [4] or support vector machines [5]. A promising method is the adoption of reinforcement learning (RL) as a high level biasing mechanism for learning an optimal action selection policy [6] or oppositely, the exploitation of the inverse reinforcement learning (IRL) framework to learn the reward function from human data [7].

Conversely, the problem of action selection is not a peculiar feature of AVs, instead any agent (both artificial and biological) dealing with complex dynamical environments where multiple mutually exclusive behaviours are possible, shares similar dilemmas. Indeed there exists a huge amount of ethology literature investigating "behaviour switching" and "decision making" [8], the common jargon among cognitive scientists to refer to the action selection problem in robotics.

Several theories have been proposed in literature on how animals perform effective decision making [9]. For instance, in [10] the *affordance competition* concept underlines a parallel processing of multiple actions competing against each other until the selection of the winning behavior. Such a modeling framework is based on the definition of criteria for assessing the *worthiness* of the action and the *selection* process itself.

We exploit this concept of parallel competing actions in the context of the European Projects SafeStrip[1] and Dreams4Cars[2]. In particular, in SafeStrip we take advantage of the mirroring mechanism introduced in [11] to infer the human driver intended action in several dangerous scenarios, like in the proximity of a pedestrian crossing, in a road work zone or in an intersection. In the latter case a more complex mirroring is performed, taking into account the right of way rules and mirroring other vehicles. This is made through vehicle to vehicle and vehicle to infrastructure communication [12].

Such an inference process boils down to the selection among a set of longitudinal maneuvers, called *motor primitives*, of the one matching the driver intended action in terms of instantaneous jerk j_0. Each motor primitive has an optimality-based formulation characterized by an initial jerk associated with. By defining the jerk space as a one 1-dimensional grid we can explore a set of possible actions taking also into account infrastructure-based information.

In Dreams4Cars we utilize a similar optimality-based motor primitives approach for the synthesis of an autonomous driving agent called Co-driver [13]. In addition to the longitudinal manoeuvres, we also generate set of lateral manoeuvres by defining a 1-dimensional grid on instantaneous lateral jerk r_0. By combining the two grids we devise a 2-dimensional matrix where each entry is a pair of (j_0, r_0) which encodes a latent action. Each pair is then assign a merit via the definition of a scenario dependent *salience*.

Common to both the project there is the need to select the best action after the computations of the grids. The rest of this paper is devoted to demonstrate how we can perform such a task taking advantage of a biologically inspired action selection mechanism.

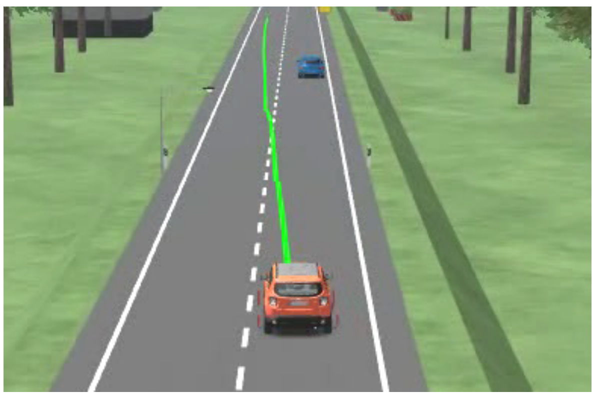
II. THE MOTOR CORTEX CONCEPT

In order to better clarify how the affordances competition process takes place, let us inspect an example simulation

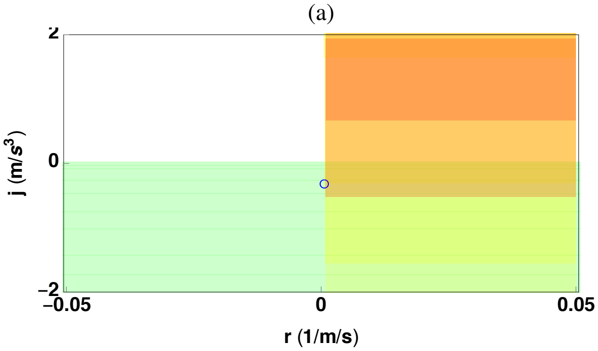
[1] Department of Industrial Engineering, University of Trento, 38123 Trento, Italy riccardo.dona@unitn.it

[1]https://www.safestrip.eu
[2]https://www.dreams4cars.eu

scenario as in Fig. 1. In the proposed situation the ego car, driven by the Co-driver agent, is travelling at high speed on a straight road when a slower vehicle is detected (Fig. 1a). This scenario translates into the control space representation shown in Fig. 1b. Physical space to control space transformation is performed via the analytical solution of a linearized vehicle kinematic plant optimal control similarly to [14], [13]. The green portion of the control space representation expresses the feasible control actions, *i.e.* the set of pairs $(j_0,\ r_0)$ which allow the ego car to stay within the solid lane markings. On the other side, the orange/yellow portion conveys the control inhibition caused by the presence of a slower leading vehicle. The solid orange region is associated to controls that lead to a collision while the yellow area encodes the potential danger in staying too close to the obstacle. Eventually the white region expresses the speed limit exceedance.



(a)



(b)

Fig. 1: Example simulation scenario bird-eye view (a) and corresponding control space representation (b).

The motor cortex corresponding to the action space in Fig. 1b can be computed by introducing some merit criterion. For the considered example scenario we model the merit as the maximum time at which, given the pair $(j_0,\ r_0)$, the vehicle will leave the road or collide with other road users. In other words we are trying to find which are the controls that allow the vehicle to navigate the longest without any further intervention during the execution. This idea is also known as *minimum intervention principle* [15]. Given the biological inspiration of the procedure, we refer to such a time as the *salience* of the action. By establishing the criterion above, we can compute an artificial *motor cortex* as in Fig. 2, where the salience is displayed along the z-axis of the 3D plot. It can be noticed how lateral controls close to zero have high merit values as, clearly, steering abruptly will drive the vehicle out of the road sooner than steering mildly while the orange region in Fig. 1b has a close to zero salience due to the inherent risk of collide in a short time frame.

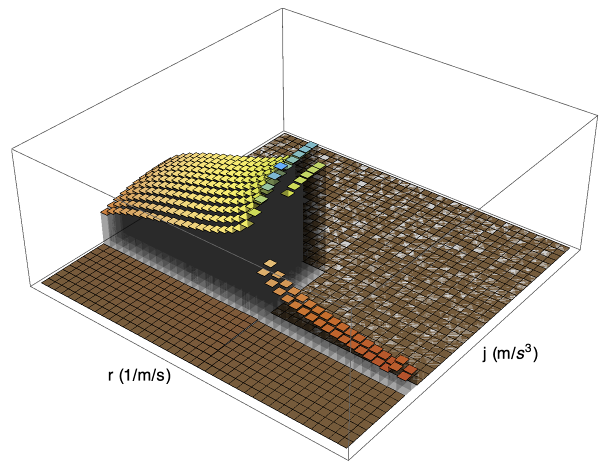


Fig. 2: Minimum intervention principle based motor cortex for the scenario in Fig. 1.

The motor cortex in Fig. 2 encodes the *affordance* concepts previously mentioned. Each of the action is in fact associated to a merit value and compete against the others for winning the selection process. The outcome of the "competition" is the optimal pair $(j_0^*,\ r_0^*)$ that will eventually guide the car for the next time-step.

In the inference-via-mirroring application, the merit assignment procedure is slightly modified to account for both the potential maneuvers and the one currently performed by the driver. After the computation of the scenario-based merit for each initial control, a bias function measuring the proximity of the driver maneuver to each action is applied to the motor cortex as shown in [16] for the longitudinal control only.

III. ACTION SELECTION

A. WTA algorithm

The most trivial approach to model the affordances contetition would be to simply choose the pair having the highest instantaneous salience. This selection mechanism is known as *winner takes all* (WTA) [9] and has proven to be fairly efficient in the simulation environment where there is no signal noise. On the other side, this action selection procedure is likely to choose sub-optimal action in the presence of noise such as when the agent is driving a real car. Furthermore, even in the simulation environment, this mechanism may give rise to hysteresis when two competing actions share similar salience values which can cause loss of vehicle control.

B. MSPRT algorithm

In order to overcome the problematics of the WTA procedure we propose here the introduction of the multi-hypotheses sequential probability ratio test (MSPRT) [17] decision making algorithm.

The key idea of the MSPRT algorithm is to accumulate *evidence* for each channel and then pick an action only when the integral reaches a predefined *threshold* level. This mechanism should guarantee more robustness to noisy decisions by trading off some responsiveness.

The MSPRT has been shown to be asymptotically time-optimal in a multi alternatives process [18]. More recently a link between the action selection process happening in the basal ganglia of the human brain and the MSPRT algorithm has been drawn [19].

The overall procedure for action-selection using the MSPRT algorithm is reported in Algorithm 1. First we append the set of observations at time-step $\mathcal{M}_t$ to the list of observations $\mathcal{M}_{\text{list}}$ which contains observations from $t - ws$ up to t where ws is the dimension of the averaging window. Then we compute the mean of observation along the time dimension in order to get an average of each channel evidence for the considered window. Next we compute the likelihood of each channel according to

$$L(t) = y(t) - \log \sum_{k=1}^{N} \exp\left(y_k(t)\right), \tag{1}$$

where $y(t)$ represents the vector of evidence at time-step t obtained via flattening the motor cortex. Then we compute the $\max(\exp(L))$ to investigate whether some of the channels reached a predefined threshold value. In the positive case we reset the moving average list by taking only a percentage λ of the current average value. Otherwise we continue to follow the previous action until eventually a new action will win the affordance race.

Algorithm 1: MSPRT algorithm

Result: Action log-likelihood
$\mathcal{M}_{\text{list}} \leftarrow \mathcal{M}_t$;
$\bar{\mathcal{M}} \leftarrow \text{mean}\{\mathcal{M}_{\text{list}}\}$;
compute likelihood L as in (1);
if $\max(\exp(L)) >$ *threshold* **then**
    take action;
    $\mathcal{M}_{\text{list}} = \lambda\, \bar{\mathcal{M}}$
else
    follow previous action;
end

Overall the behaviour of the MSPRT algorithm can be shaped by adjusting the hyper-parameters in Table I.

IV. SIMULATION COMPARISON

We compare the performances of the MSPRT against the WTA on simulated logged data. Firstly we let the agent drove on a simulated scenario with no noise affecting the measurements. According to this set-up we can perform optimal decision making using a simple WTA algorithm. We then select a 9-seconds long critical double lane change maneuver where the responsiveness of the action selection plays a fundamental role. Next, we re-execute the simulation offline, *i.e.* we take the logged motor cortex history, we apply some random noise on the channels and we re-execute the decision making algorithm only on the corrupted motor cortex. We then analyze again the performances of the WTA against MSPRT with respect to the ground-truth case obtained previously. The exact parameters used in the simulation above are reported in Table I.

TABLE I: Parameters of the MSPRT algorithm.

name	symbol	value	effect
threshold	th	0.0005	slows down the switch to a new channel
windows size	ws	8	average out noise, brings in more robustness
forgetting factor	λ	0.9	introduces a memory effect after the switching to a new channel

Fig. 3 reports the results of assessment as a function of the adimensional noise variance σ injected into the motor cortex. In case of limited noise figures, the WTA still outperforms MSPRT due to the worse transient performance of the latter. As soon as we introduce noise in the simulation, however, the advantages of the MSPRT start to be evident. In this case we chose a fairly conservative tuning for the MSPRT that will make it behave correctly even in the presence of high noise while the performance of the WTA drops in a more significant manner. Indeed by shrinking the threshold value and setting the λ to zero MSPRT will perform exactly like WTA.

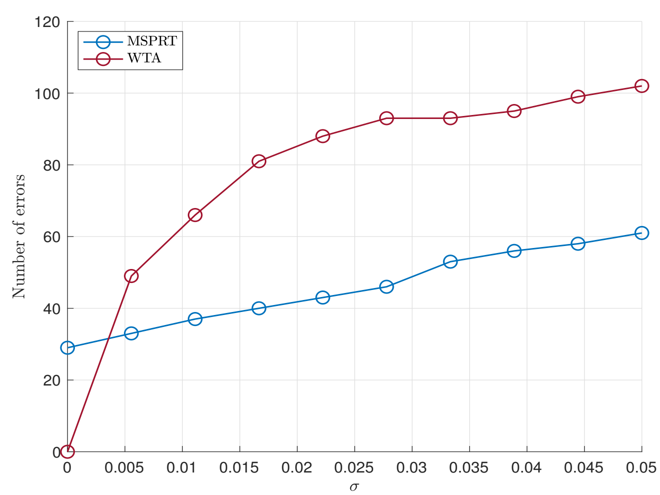


Fig. 3: MSPRT vs. WTA channels selection errors. Parameters of the simulation as in Table I.

Another valuable performance index is the number of switches, the lower the number of switches the more stable the behaviour of the agent. Fig. 4 shows the switching logic for the MSPRT and WTA for a selection of the data-set.

It is evident how the MSPRT not only picks the best action more effectively than WTA but also tends to stick with a sub-optimal action rather than continuously changing the channel which could lead to vehicle instability.

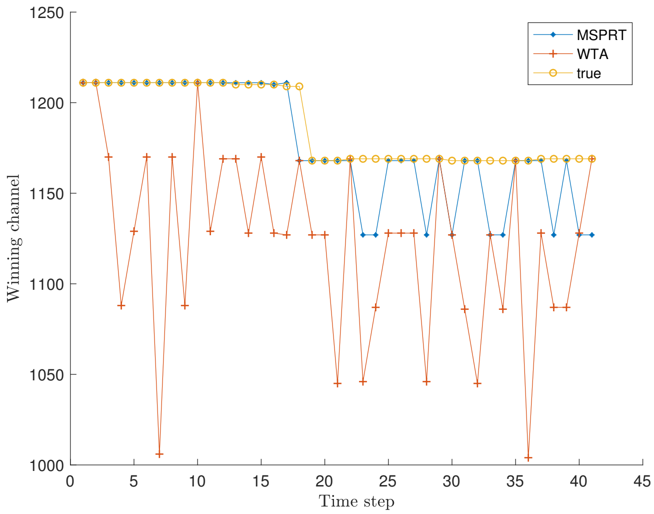


Fig. 4: MSPRT vs. WTA channels switching for $\sigma = 0.5$. Parameters of the simulation as in Table I.

V. CONCLUSIONS

We have shown that bio-inspired cognitive models can play a substantial role in the process of decision making in automated driving. In particular we demonstrated how the biologically inspired MSPRT algorithm can be adapted to both the inference and the action selection process given a suitable lower-lever architecture for the agent. The advantages of the proposed formulation lie in an improved robustness to noisy observations (Fig. 3) and a greater stability of the chosen action (Fig. 4) with respect to traditional action selection. Indeed the effectiveness of the MSPRT depends on the tuning of application dependent hyperparameters. The activation *threshold* shapes the sensitivity to the process noise: the lower the threshold the more responsive will be the action picking, the higher the threshold the more robust the selection. Similar considerations apply for the tuning of the forgetting factor λ. However, we proved via simulation that it is possible to find an effective trade off adjustment for the MSPRT such that the algorithm outperforms other techniques. In particular for the considered data-set and $\sigma = 0.5$ the MSPRT guarantees an error rate up to 40% inferior to the WTA algorithm. Further work will be devoted to the set-up of a "layered" action selection process where a lower layer will be in charge of merging the contribution of channels encoding the same action to make sure that the affordance competition takes place among statistically independent channels only in order to run the MSPRT more efficiently.

VI. FUNDING

This work is supported by the European Commission Grant 731593 (Dreams4Cars) and 723211 (SafeStrip).

REFERENCES

[1] T. Gindele, D. Jagszent, B. Pitzer, and R. Dillmann, "Design of the planner of team annieways autonomous vehicle used in the darpa urban challenge 2007," in *2008 IEEE Intelligent Vehicles Symposium*. IEEE, 2008, pp. 1131–1136.

[2] C. Urmson, J. Anhalt, D. Bagnell, C. Baker, R. Bittner, M. Clark, J. Dolan, D. Duggins, T. Galatali, C. Geyer *et al.*, "Autonomous driving in urban environments: Boss and the urban challenge," *Journal of Field Robotics*, vol. 25, no. 8, pp. 425–466, 2008.

[3] S. Brechtel, T. Gindele, and R. Dillmann, "Probabilistic MDP-behavior planning for cars," in *2011 14th International IEEE Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2011, pp. 1537–1542.

[4] J. Schroder, M. Hoffmann, M. Zollner, and R. Dillmann, "Behavior decision and path planning for cognitive vehicles using behavior networks," in *2007 IEEE Intelligent Vehicles Symposium*. IEEE, 2007, pp. 710–715.

[5] C. Vallon, Z. Ercan, A. Carvalho, and F. Borrelli, "A machine learning approach for personalized autonomous lane change initiation and control," in *2017 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 2017, pp. 1590–1595.

[6] M. Mukadam, A. Cosgun, A. Nakhaei, and K. Fujimura, "Tactical decision making for lane changing with deep reinforcement learning," 2017.

[7] D. Sadigh, S. Sastry, S. A. Seshia, and A. D. Dragan, "Planning for autonomous cars that leverage effects on human actions." in *Robotics: Science and Systems*, vol. 2. Ann Arbor, MI, USA, 2016.

[8] D. McFarland, *Problems of animal behaviour*. Longman Sc & Tech, 1989.

[9] P. Redgrave, T. J. Prescott, and K. Gurney, "The basal ganglia: a vertebrate solution to the selection problem?" *Neuroscience*, vol. 89, no. 4, pp. 1009–1023, 1999.

[10] P. Cisek, "Cortical mechanisms of action selection: the affordance competition hypothesis," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 362, no. 1485, pp. 1585–1599, 2007.

[11] M. Da Lio, A. Mazzalai, and M. Darin, "Cooperative Intersection Support System Based on Mirroring Mechanisms Enacted by Bio-Inspired Layered Control Architecture," *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 5, pp. 1415–1429, May 2018. [Online]. Available: https://ieeexplore.ieee.org/document/8011474/

[12] G. Valenti, D. Piscini, and F. Biral, "A cooperative intersection support application enabled by safe strip technology both for c-its equipped, non-equipped and autonomous vehicles," in *European conference on Intelligent Transportation Systems*, 2019.

[13] M. Da Lio, F. Biral, E. Bertolazzi, M. Galvani, P. Bosetti, D. Windridge, A. Saroldi, and F. Tango, "Artificial Co-Drivers as a Universal Enabling Technology for Future Intelligent Vehicles and Transportation Systems," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 1, pp. 244–263, 2015.

[14] M. Werling, S. Kammel, J. Ziegler, and L. Gröll, "Optimal trajectories for time-critical street scenarios using discretized terminal manifolds," *The International Journal of Robotics Research*, vol. 31, no. 3, pp. 346–359, 2012.

[15] E. Todorov and M. I. Jordan, "A minimal intervention principle for coordinated movement," in *Advances in neural information processing systems*, 2003, pp. 27–34.

[16] G. Valenti, L. De Pascali, and F. Biral, "Estimation of longitudinal speed profile of car drivers via bio-inspired mirroring mechanism," in *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2018, pp. 2140–2147.

[17] C. W. Baum and V. V. Veeravalli, "A sequential procedure for multihypothesis testing," *IEEE Transactions on Information Theory*, vol. 40, no. 6, 1994.

[18] V. Draglia, A. G. Tartakovsky, and V. V. Veeravalli, "Multihypothesis sequential probability ratio tests. i. asymptotic optimality," *IEEE Transactions on Information Theory*, vol. 45, no. 7, pp. 2448–2461, 1999.

[19] R. Bogacz and K. Gurney, "The basal ganglia and cortex implement optimal decision making between alternative actions," *Neural computation*, vol. 19, no. 2, pp. 442–477, 2007.