This is a repository copy of *Agent architecture for adaptive behaviours in autonomous driving*.

White Rose Research Online URL for this paper:
http://eprints.whiterose.ac.uk/163442/

Version: Published Version

Article:

Agent Architecture for Adaptive Behaviours in Autonomous Driving

MAURO DA LIO[1]**, (Member, IEEE), RICCARDO DONÀ**[1]**, GASTONE PIETRO ROSATI PAPINI**[1] **AND KEVIN GURNEY**[2]**.**

[1] Department of Industrial Engineering, University of Trento, Trento, Italy (e-mail: {mauro.dalio, riccardo.dona, gastone.rosatipapini}@unitn.it)
[2] Department of Psychology, University of Sheffield, UK (e-mail: k.gurney@sheffield.ac.uk)

Corresponding author: Mauro Da Lio (e-mail: mauro.dalio@unitn.it).

ABSTRACT Evolution has endowed animals with outstanding adaptive behaviours which are grounded in the organization of their sensorimotor system. This paper uses inspiration from these principles of organization in the design of an artificial agent for autonomous driving. After distilling the relevant principles from biology, their functional role in the implementation of an artificial system are explained. The resulting Agent, developed in an EU H2020 Research and Innovation Action, is used to concretely demonstrate the emergence of adaptive behaviour with a significant level of autonomy. Guidelines to adapt the same principled organization of the sensorimotor system to other agents for driving are also obtained. The demonstration of the system abilities is given with example scenarios and open access simulation tools. Prospective developments concerning learning via mental imagery are finally discussed.

INDEX TERMS Adaptive Behaviour, Affordance Competition Hypothesis, Autonomous Driving, Explainable Artificial Intelligence.

I. INTRODUCTION

THIS paper presents the architecture of an Agent for autonomous driving that was developed in an EU Horizon 2020 Research and Innovation Action (Grant 731593, Dreams4Cars). The organization of the Agent sensory-motor system was conceived with a twofold goal, both inspired by biological principles: *a)* to support learning in a "wake-sleep/dream" scheme, and *b)* to produce *adaptive behaviours* – in the ethological sense – when in use for vehicle operation (i.e., at the "wake" state) [1].

This paper focuses on the latter aspect, illustrating principles and good practices for designing a sensorimotor system. The focus includes ideas and expedients that can also be used with little effort with (most of) the path planning and control methods currently in use. Throughout, we use Open Access simulation tools, and prospective longer-term developments are discussed in the conclusions

Specifically, the paper invokes several biological ideas (see Section I-B2) which underpin adaptive behaviours in Nature, including: topographic organization of motor space, robust action selection and steering of agent behaviour via biasing of the action selection. Many of these come together in an overarching scheme – the Affordance Competition Hypothe-

sis [2] – which we use as a guiding framework.

When these principles are used to organize an artificial system, benefits similar to those seen in animals emerge such as safe, natural, adaptive and robust behaviours. A number of other desirable benefits, such as the ability to deal with a hierarchy of intentions, prioritizing safety vs legality, and explainable Artificial Intelligence, are also shown. Notwithstanding this, we do not claim that the Agent presented here is optimal; rather we emphasise a set of interrelated principles that can be adopted, together or separately, to enhance artificial behaviour.

The paper theory is substantiated with several working examples for unusual situations and/or scenarios with randomness in the environment and in the behaviour of the other agents. The examples and simulations tools are open access, as part of the Open Data produced by the Research and Innovation Action (Dreams4Cars) [3].

A. NOVELTY AND CONTRIBUTION

The novelty and contribution of this paper is a sensorimotor architecture based on principles that are successful in Nature. Among these, the topographic organization of encodings in motor space, behaviours that are emergent from the com-

petition between affordable actions, robust action selection (the MSPRT algorithm) and selection biasing as a means to prioritize behaviours (for safety/legal/comfort etc.). The paper describes, with examples, both the overall system and the individual principles to permit a modular adoption of these ideas.

As far as we are aware, this is the first time that an engineered system for self-driving cars has been based on the large scale architecture of the human brain. Thus, we invoke ideas contained in the affordance competition hypothesis [2] that sensory input, processed by visual and parietal cortical brain areas, provides action options or affordances for decision making by sub-cortical structures (the basal ganglia). The algorithms for decision making are based on those previously used to model the basal ganglia [4]. As such, our system brings together elements of sensory, motor and decision making competencies found in the brain.

B. RELATED WORK

1) Path planning and control

A large number of methods and variants for trajectory planning and control are known in the literature. Reviews may be found in [5]–[9]. In addition, machine learning approaches should also be mentioned, for example, those based on learning human driving, e.g., [10]–[12].

For the goals of this paper, only the aspects of this work that are relevant for the efficient organization of the sensorimotor system are reviewed here.

The first point to note is that almost all trajectory planning approaches work by first producing a number of candidate trajectories and then selecting the "best" one. To link with our biologically inspired agenda, this two stage approach may be couched in terms of an initial step of *action priming* (candidate trajectories being established) followed by an *action selection* (fixing one trajectory for execution). However, as will be clarified below, the way in which priming-selection is carried out can make a great difference in the effectiveness of an agent behaviour.

Concerning *trajectory generation* (action priming), two aspects that characterize the production of candidate trajectories are *completeness* and *computational efficiency*. Completeness means that the space of possible trajectories is entirely spanned, so that a safe trajectory is found if it exists. Many trajectory generation methods work by sampling either the physical (configuration) space or the control space. Too coarse a sampling may miss finding scarce evasive trajectories in critical situations. On the other hand, computational efficiency (software and hardware) limits the number of candidates that can be analysed at every iteration.

Highly dynamic situations (which often happen in critical conditions) also pose challenges. Ideally the planning update rate should be high and the planning latency should be low, so that fresh plans that respond to unpredictable environmental changes are promptly available, e.g., [13]. However, for the execution of a given trajectory, only predictive control schemes are surely compatible with continuous re-planning,

whereas other types of control (e.g., feedback and pursuit controllers) may not work because new plans may typically begin at the current vehicle state and thus cancel the instantaneous errors that is used in those control schemes.

Dealing with moving obstacles is also challenging: not only because obstacle trajectories must be predicted but, also, because they add one dimension (time) to the planning problem (a simple, but not complete, method that may be used, e.g., [14], is that paths are first planned without considering the obstacles and then the longitudinal dynamics is adapted to the obstacle movements).

Regarding *trajectory selection* (action selection), the choice of one trajectory in a pool of trajectories must meet multiple objectives: safety, compliance with traffic rules, travel time, comfort, energy efficiency etc. Often these objectives are combined into an unique cost function. However objectives may be better organised in a hierarchy of priorities. Thus, safety should have the highest priority, including priority over traffic rules (for example, a vehicle should be allowed crossing solid lane markings if that is the only way to avoid a collision). In many trade-off approaches, however, safety and compliance with traffic rules are considered at the same level as "hard" constraints; whereas only the others are considered separately as soft goals, e.g., [9].

Finally, the choice of the "optimal" manoeuvre is often carried out according to a "winner takes all" (WTA) method, i.e., by selecting the manoeuvre that maximizes the weighted optimally criterion. However this kind of selection may not be ideal in case of uncertainties and noise, where more robust action selection algorithms can be deployed to advantage.

2) Adaptive Behaviour

Discovering and selecting effective behaviours is critical for animal survival. Natural evolution has produced very efficient and highly effective methods for such adaptive behaviour and, by studying them, it may be possible to improve the robustness and autonomy of robot behaviour too.

One fundamental idea is the notion of an *affordance* [15], [16] which makes an intimate link between perception and action. In affordance theory, perception is not thought of as simply an elucidation of a set of abstract features describing the environment. Rather, the job of perception is to identify ways in which the animal (or agent) may interact with its environment by pursuing an effective course of action. For example, in the current context, a free space like a roadside parking spot (Fig. 4), offers and affordance for the action of driving into it and may, on occasion, elicit a possibly life saving behaviour in order to avoid a collision.

Cisek has articulated a neuroscientifically grounded version of affordances and actions in his Affordance Competition Hypothesis [2]. A computational neural model of affordance priming and selection, based on these ideas has also been developed [17]. In this model, potential actions are formed, simultaneously and in parallel, in pathways running from sensory to the motor cortices; anatomically this comprises a dorsal processing stream in the brain (dorsal is the

upper surface in primates). Potential actions are encoded as patterns of neural activity, with activation strength reflecting the value, or "salience" of the encoded action. Actions then compete for taking control of the agent, with their probability of success determined by these salience values. According to Cisek, adaptation to dynamically changing conditions occurs via "the continuous evaluation of alternative activities that may become available and continuous tradeoffs between choosing to persist in a given activity and switching to a different one" [18]. Interestingly for us, there are experimental studies that support the interpretation of driver behaviours in terms of selection between affordances [19], [20].

The emergence of adaptive behaviours from the priming-selection arrangement is also discussed in [21], which highlights how using a centralized selection mechanism encoding action salience with a common scale (instead of a distributed one such as proposed in [22]) realizes a common evaluation metric that permits seamless extension of motor abilities via learning of new action priming loops.

Concerning the sub-problem of the action-selection — which of the many possible actions is gated to the motor system— there also are several studies and biologically grounded computational models [23], [24]. Further, these neural models have been shown to be describable by decision making algorithms such as the the multiple-hypothesis sequential probability ratio test (MSPRT) [4]. Under certain conditions, the MSPRT allows optimal decision making with noisy and uncertain signals. It should therefore be unsurprising that the brain has recruited such algorithms to guide animal behaviour. The MSPRT is described in more detail in Section II-D.

Finally, action-selection via the competition of immediately available affordances may be "biased" by higher level influences, thereby offering the opportunity of steering agent behaviours towards long-term goals [25], and the exploration required for action discovery [26].

C. A CHALLENGING SITUATION

We introduce the architecture using examples of desired system competencies. Fig. 1 shows a motorway situation harbouring a danger in which objects might fall from the minivan.

A human driver would have the ability to predict this possible event, to mentally simulate possible object trajectories and to anticipate mitigation strategies (e.g., keep increased distance, prepare evasive actions, comfortably change the lane).

Let us assume an artificial driver is not yet capable of this level of prediction. However, we might request that — if an object actually falls onto the road— the agent has at least: 1) the ability of evaluating as many escape strategies as possible. Furthermore, the falling object, depending of its nature, might have irregular trajectory. So: 2) continuous quick adaptations of the current manoeuvre may also be necessary. Finally, it may happen that a collision-free trajectory is not strictly legal; for example if left and right lanes were

busy, it might happen that fitting in the middle between two lanes could avoid the accident. This means that safety must have the priority even, to some extent, over legality and that: 3) motor planning must be carried out to satisfy multiple hierarchical objectives where lower priority objectives can be given up if necessary.

II. AGENT ARCHITECTURE
A. FUNCTIONAL LOOPS

Fig. 2 shows the biological basis for the architecture of the Agent, adapted from [2][1]. Of course, we don't aim to model the brain architecture faithfully as a large scale neuronal network; rather, we use the the scheme in Fig. 2 to highlight a series of *functionalities* and their interrelationships, which may be modelled at a high level with other technologies (we use regular computer code with neural network modules).

We now describe this architecture in more detail, starting with the *action priming* stream (the solid red arrow) and the *action selection* loop (shown in solid green); the remaining processing streams will be discussed later.

In the human brain the *action priming* stream occurs in the dorsal regions of cortex, and comprises a pathway running form the sensory cortices (*a* in the figure) to the motor cortices *c*. Of course, the human sensorimotor system is more complex than the simple unidirectional data flow depicted by the arrow: other pathways are involved, the flow is not simply unidirectional (as indicated by the dashed red arrow in Fig. 2), and information is compressed and expanded by convergence and divergence in the neural pathways [27].

As noted earlier, we conceive of the action priming pathway as computing the *salience values* of candidate trajectories (the salience is obtained via learned perception-action *associations*, without evaluating the trajectories as an intermediate step; see Section II-C4).

Turning to action selection, in the human brain there is an *action selection loop*, at the heart of which is a sub-cortical brain system of interconnected nuclei called the basal ganglia [24]. The basal ganglia are evolutionary old and common to all vertebrates, reflecting the fundamental nature of the behavioural problem of action selection faced by all animals. As noted earlier, time-efficient decision-making equivalent to the biological solution can be implemented with the MSPRT algorithm (Section II-D1).

The basal ganglia are also a locus of learning – enabling the selection of new actions, and re-emphasising the importance of existing ones [26]. In this way they offer a mechanism for influencing decision making, effectively steering the agent behaviour for long-term rewards [25] (Section II-E).

B. ORGANIZATION OF THE ACTION SPACE: THE "MOTOR CORTEX"

The processing streams in the brain described above have the critical property of dealing with many actions and affordances in parallel. Further, the representations of the relevant

[1]For clarity, only the processing streams that are active during vehicle operation (point *b*, Section I) are shown.

FIGURE 1: Example dangerous motorway situation. Objects might fall from the minivan. A human driver has the ability to predict the possible event, mentally simulate possible object trajectories and elaborate mitigation strategies (keep increased distance while preparing evasive actions). With current technologies an autonomous vehicles will not be able of such cognition abilities (note it is not only a matter of perception). Hence, as a mitigation strategies, an automated vehicle should be able to react to a real falling object by quickly elaborating as many viable trajectories as possible (including trajectories not strictly legal such as squeezing between the lanes) in order to have the largest possible set of choices.

percepts and actions occurs in an ordered way with similar items, and features therein, being encoded in proximity to each other in the neural tissue. This notion of *topographic organization* is found across several brain structures [28]–[30], among which are the sensory and motor cortices referred to in Fig. 2, *a* and *c*, respectively. The use of topographic organisation has also found its way into abstract neural networks where, for example, the topographic organization of the visual cortex had been one inspiring idea of modern convolutional networks.

We adopt the notion of topographic organization of the motor cortex to the current context of driving, in the scheme given in Fig. 3 (bottom). Here, the salience of primed actions are arranged in a two-dimensional space corresponding to the instantaneous lateral and longitudinal control.

This arrangement carries a number of benefits that are not easily obtained when the pool of actions is not arranged in this way. To our knowledge, the first time topographic organization had been used for artificial "codrivers" was in the FP7 InteractIVe project ([31], Figure 7). The same organization has been used in FP7 AdaptIVe and improved in H2020 Dreams4Cars. We may, however, find a prelude of this idea in the organization of ALVINN [32] where an array of output neurons is used to topographically encode the steering angle control of an autonomous land vehicle. ALVINN is, also, relevant as an example of neuralised sensorimotor system different form those of Section I-B1 and more conceptually similar to this paper. Topographic organization and related action encoding has, also, relation to dynamical systems theory of behaviours [33], in particular for what concerns neural field dynamics and behaviour representation.

The notion of topographic organization is better clarified by means of the example in Fig. 3, which presents a situation with three legal-lane distinct affordances: *a*, remaining in the current lane, *b*, turning right and *c*, changing lane to the left. For each of these legal-level intentions, the agent could elaborate an infinite number of trajectories, as exemplified with the b_i for the right turn case (of course equally numerous trajectories also exist for lanes *a* and *c*, but they are simply not shown). There may also be non-legal, but physically feasible, intentions that basically correspond to using the entire road surface, e.g., entering the right road in the opposite lane.

The control space of a vehicle has two dimensions, corresponding to lateral and longitudinal control. So, in order to produce one trajectory, the agent must elaborate two functions of time: the longitudinal control $j(t)$ and the lateral control $r(t)$. For a given intention, not all trajectories have the same cost though. Some – namely the smoothest ones – are easier to produce and less prone to the risk of loss of control and out of lane/road deviations. One could, for example still take the right turn by abruptly steering to the opposite direction for a while and then recovering with a carefully controlled steering action to the right, for example implementing *b'*. However, this would be very difficult to execute (some could even be physically unfeasible).

For every trajectory γ, generated by a single choice $\{j(t), r(t)\}$, a scalar functional $V(\gamma)$ can be defined to represent the "value" of that particular trajectory (see also Section II-C).

To model different intentions, restrictions on the admissible trajectories γ can be set. For example, the trajectory for lane change intention must stay inside the current and
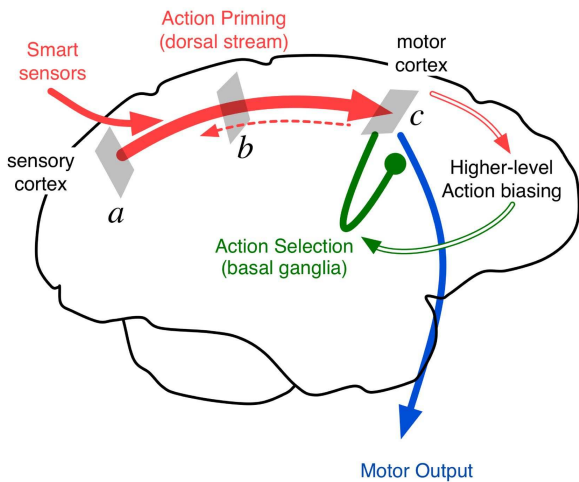
IEEE *Access*

FIGURE 2: The Agent architecture is adapted from [2]. It is made of a primary sensorimotor pathway that primes many candidate actions in parallel (red arrow). These potential actions are encoded, with their salience, topographically arranged in the motor space ("motor cortex"). The selection among the possible actions is carried out by means of a particular competition process that is robust against sensory and motor noise (green loop). An "action biasing" loop can steer action selection to implement hard constraints like traffic rules, as well as long-term action sequences. Once an action is selected, inverse models of the body dynamics are used to resolve the action into the low-level motor commands (blue arrow).

destination lane and terminate nearly aligned with the centre of the destination lane (e.g., Fig. 4, c, green area). Intentions are hence modelled with admissible sets g_i for γ (e.g., $\gamma \in g_i$ with g_i like a, b, c, d in Fig. 4).

Since the agent is primarily concerned with selecting the *current* control $\{j(0) = j_0, r(0) = r_0\}$ (in adaptive behaviour future controls can be modified later), a definition of salience as a means to express how good the choice of $\{j_0, r_0\}$ may be in relation to intention g_i, can be given as follows:

$$s_{g_i}(r_0, j_0) = \sup\left(V(\gamma)|r(0) = r_0, j(0) = j_0, \gamma \in g_i\right). \quad (1)$$

This means that the salience value of the instantaneous choice $\{j_0, r_0\}$ for intention g_i is that of the optimal γ among all the trajectories beginning with $\{j(0) = j_0, r(0) = r_0\}$ and belonging to the subset g_i, which models the intention.

It is not difficult to recognize the similarity with Reinforcement Learning, where $s_{g_i}(r_0, j_0)$ is the Q function estimating the future reward for choosing action $\{j_0, r_0\}$ at the current state. However, there are as many reward functions as the number of goals/intentions g_i, possibly organised hierarchically.

Our "motor cortex" therefore encodes control actions as being a two-dimensional array of discrete samples of the control space. It is therefore analogous to the neural structure in Fig. 2, c, and similar to the output neural array of ALVINN.

The value stored in the motor cortex array is $s_x(r_0, j_0)$, where x may be either one individual goal or the union of more (see next section). The discretization is typically not uniform: finer in the centre of the motor space, where minute precise control is desirable, and coarser at the edge. In Dreams4Cars this motor cortex array has size 41×41 (element $[21, 21]$ corresponds to the null action), which means that the agent at the lowest-level chooses among 1681 possible actions, organized in a hierarchy of intentions g_i.
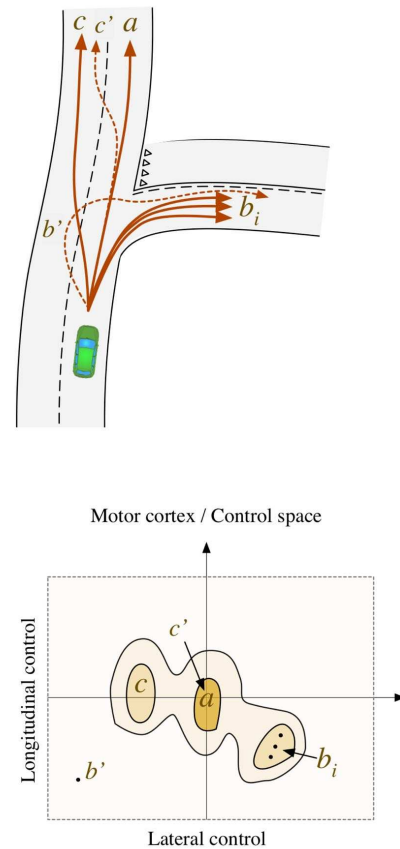

FIGURE 3: Topographic organization of the motor space. The salience of each available action is reported in the "motor cortex", which is the two-dimensional space representing the instantaneous lateral and longitudinal control. Hence, such map encodes the value of choosing one particular combination of lateral/longitudinal control at any time.

C. ACTION PRIMING

The instantiation of possible actions corresponds to the computation of the salience $s_{g_i}(r_0, j_0)$, $i = 1, \ldots, N$, where N is the number of affordances.

1) Modularity and parallelization

The process can be parallelized as shown in Fig. 4. Safety is the first concern (panel a): the vehicle must stay in the

road and, if necessary, drive over the lane markings or over other extra-lane room. The salience is computed for $\gamma \in a$ by means of (1), as shown in the small inset sketch of motor cortex to the right of the main panel (Fig. 4). Besides remaining on the road, the agent may have three legal intentions: b, lane following, c, lane change and d, stopping in the parking spot. For each of them, the salience can be computed in a similar manner ($\gamma \in b, c, d$), producing motor cortex activation patterns as shown in the central column of motor cortical sketches Fig. 4.
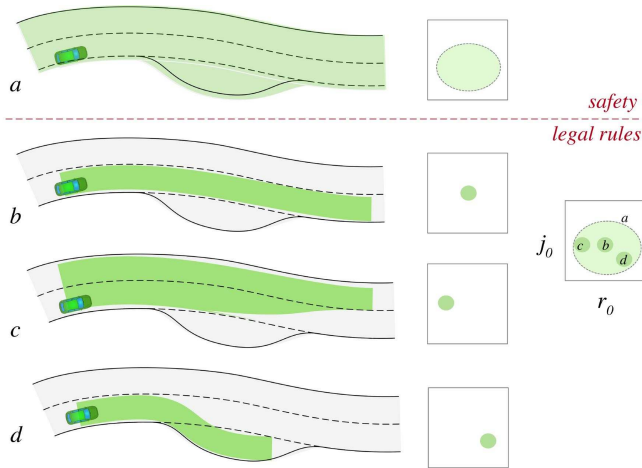

FIGURE 4: Modular parallel action priming. In a situation where there may be several alternative goals (possibly organised hierarchically), action priming can be parallelised, i.e. by overlapping the salience of each goal evaluated individually.

Note that, quite generally, intentions correspond to strips of possibly variable width (the simplest way to know the strips is if a digital map describing the road at the level of lanes is available). Hence, a module that computes the salience for a generic strip is sufficient for generating the individual activation of each intention. With modularity, the complexity of action priming is decomposed into developing simpler functions that prime individual goals, which can be verified in isolation.

A global salience function can then be obtained via aggregation of the individual ones, as shown in Fig. 4, right, or Fig. 3, bottom. One possible aggregating function is a weighted $\max$ operator, i.e.:

$$s(r_0, j_0) = \max(w_i s_{g_i}(r_0, j_0), g_i \in \textit{affordances}). \quad (2)$$

where weights w_i may serve to steer action selection as explained in the following.

2) Scalability

The encoding of action values with salience in the motor cortex implies scalability: new action possibilities would be enacted by new branches in Fig. 4 and appear as new active regions in the motor cortex. Encoded with the same salience scale, they would be immediately available for competition with the others and for selection [21].

3) Hierarchy of intentions

Legal intentions are assigned higher salience, symbolised by darker green tones, for example by using $w_b \approx w_c \approx w_d \approx 1$ and $w_a \ll 1$. In this way the agent will first seek to meet one action among b, c, d and only if no solution exists it will use one action in a. This means that only if no legal action is available the agent will resort on choosing a non-legal physically feasible action (remaining in the road) as a last resource.

4) Computation of the "motor cortex": declarative predictions

The computation of the salience by means of (1) can be carried out, in principle, with the trajectory planning methods mentioned in Section I-B1.

For this, (in principle) for every intention g_i the motor space $\{r_0, j_0\}$ must be sampled with sufficient density and, for each $\{r_0, j_0\}$, an optimal trajectory γ, maximizing functional $V(\gamma)$ that represents the optimality criteria such as those mentioned in Section I-B1, must be found. We are, in particular, interested in evaluating the maximum of $V(\gamma)$, which is the value of the choice $\{r_0, j_0\}$ for intention g_i, i.e., the salience $s_{g_i}(r_0, j_0)$[2]. However, one should note that this process implies that many optimality problems must be solved inline simultaneously, one for every discrete choice of $\{r_0, j_0\}$ and for every intention g_i yielding a corresponding large number of optimal trajectories that are used for computing their values; all together computing $s_{g_i}(r_0, j_0)$. Only one trajectory will be executed though. Evaluating so many trajectories in every detail for the purpose of extracting their values (function $s_{g_i}(r_0, j_0)$) is not very efficient, albeit there may be several means to accelerate and parallelize the process.

In the dorsal stream (Fig. 2), conversely, the salience of the affordable actions is not evaluated via detailed elaboration of all possible trajectories. Rather *associations* are learned that link perceived affordances to estimates of their value. These allow bypassing low-level detailed and computational demanding simulations (procedural simulations) to carry out faster and more abstract predictions (declarative predictions) [34]. One way to replicate this process, and accelerating the inline evaluation of $s_{g_i}(r_0, j_0)$, is training a functional approximation (e.g., a neural network) with examples generated offline by means of one trajectory planner as above. The neural network approximant will learn mapping the lane geometry to the activation pattern (salience). One early example of this was given in [35]. Another example of training neural

---

[2]Of course, for every intention g_i one might directly seek the intention-specific optimal trajectory without discretizing the motor space $\{r_0, j_0\}$. This would return the optimal control $\{r_0, j_0\}$ of each intention. However in this way the selection of one action (the optimal $\{r_0, j_0\}$) is entangled with action priming and the only remaining choice would be among the intentions g_i. We seek instead to completely separate action priming from action selection so that the optimal $\{r_0, j_0\}$ may be chosen with finer discretization and include obstacle inhibitions (see next sub-section). Following the competition between excitatory and inhibitory affordances, if necessary $\{r_0, j_0\}$ may depart from the local optimum of one single intention being attracted or repelled by others g_i, see example Section III-D.

network approximants may also be found in [36]. Since the generation of the training set is carried out offline, there are no real-time concerns and the number of training examples can be very large. At inference time the trained network will short-cut procedural computations quite quickly (if carefully crafted) and operating in parallel [36].
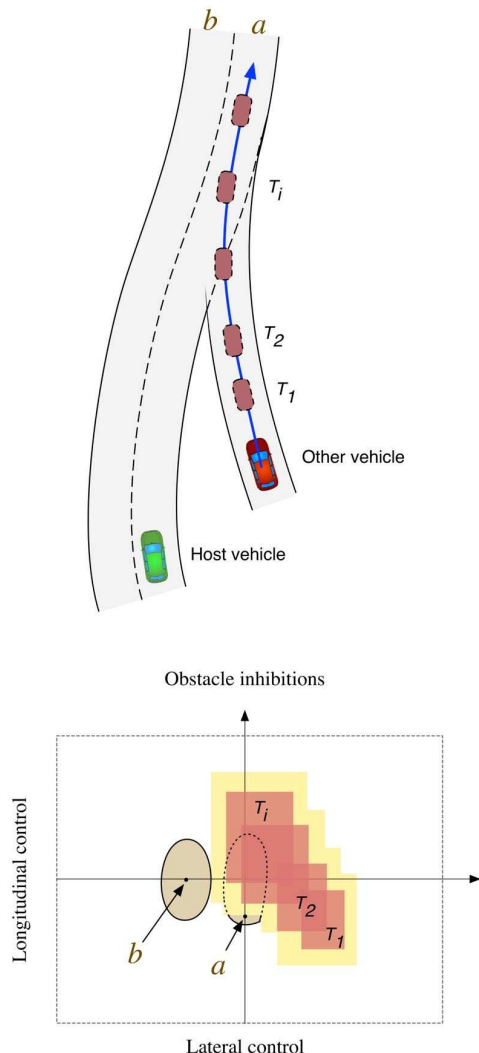

FIGURE 5: Modular parallel obstacle inhibitions. Inhibitions caused by obstacles have a second level of modularity. Besides the computation of individual obstacles in parallel, each obstacle can be resolved into a list of future space time locations which can be processed in parallel by a module that inhibits generic space-time regions.

5) Inhibitory circuits (obstacles)

Obstacles are treated as space-time locations to be avoided. The mapping between these space-time regions and the motor space is (with some adaptation) derived from the same functions used for mapping lane regions into humps of activities. The main difference is that the undesirable space-time locations are inhibited, essentially zeroing the salience for high collision probabilities (total inhibitions), and partially decreasing the salience where the probability of collision is secondary.

The computation of the inhibitions may be broken down into a further level of modularity: *a)* prediction of the obstacle trajectory (Fig. 5, top) and *b)* inhibition of space-time regions (bottom). Hence, in case of malfunctions one can diagnose whether the prediction of the obstacle trajectory was incorrect or whether the inhibitions were incorrectly computed [36]. The idea of separating desirable (mostly static) and undesirable (mostly dynamic) space-time regions via excitatory and inhibitory circuits is one way to solve the problem of trajectory planning with moving obstacles, which is otherwise very difficult to compute simultaneously and an often recognized hindrance for traditional trajectory planning.

In the example of Fig. 5 the choice between actions *a* and *b* depends on how much *a* is inhibited by the obstacle. The Agent will choose to change lane in response to a cut-in manoeuvre that requires significant speed reduction.

D. ACTION SELECTION

The action values stored in the motor cortex are readily available for action selection. The most obvious selection criterion is the "winner takes all" (WTA), i.e., choosing the action with the maximum instantaneous value.

1) Robust action selection: the MSPRT algorithm

In presence of noise the WTA criterion may not be the best option. An instantaneous snapshot of salience maps in the "motor cortex" may not reflect the distribution of mean values, derived by accumulating "evidence" over small periods of time, and which may offer the basis for a better decision. This problem of decision making using noisy evidence is a general problem in many domains and, one approach to its solution is supplied by the Multi-hypothesis Sequential Probability Ratio Test (MSPRT) [37], [38]. This choice of algorithm is supported the observation that the biological decision making mechanism – the basal ganglia – appears to have strong connections with the MSPRT [4].

The MSPRT, can be shown, under certain circumstances, to carry out time-optimal decision making with noise in the sense that it gives the shortest time to decision, given an acceptable error rate in making such decisions (to guarantee correct decision on every occasion would require evidence be accumulated indefinitely). We use an adaptation of the MSPRT (Algorithm 1), suitable for use with our action salience maps, and for online working with non-stationary inputs in a similar way to that described in [39].

The adapted MSPRT algorithm works by accumulating evidence for each action over time (`CurrentChannels` appended to `StoredChannels` in Algorithm 1), and finding the negative log likelihood that each channel is drawn from a distribution with a higher mean than the other channels (vector `NegLogLikelihoodChannels` in Algorithm 1).

Algorithm 1: MSPRT algorithm.

Data: `CurrentChannels`, `Threshold`, `Deadline`, `ForgetTime`
Result: Selected Action (`ArgMinLikelihood` and `MinLikelihood`)
Start with empty `StoredChannels`;
while *Decision not yet made* **do**
    `StoredChannels` ← **Append**(`CurrentChannels`) // store salience vectors;
    `AccumulatedChannels` ← **Total**(`StoredChannels`) //Total across list (temporal total);
    `NegLogLikelihoodChannels` ← − `AccumulatedChannels` + **Log**(**Total**(**Exp**(`StoredChannels`)));
    `MinLikelihood` ← **Min**(`NegLogLikelihoodChannels`);
    `ArgMinLikelihood` ← **ArgMin**(`NegLogLikelihoodChannels`);
    if *MinLikelihood < Threshold* **then**
        forget frames before `ForgetTime` in the `StoredChannels`;
        Return `ArgMinLikelihood` and `MinLikelihood` // selection before deadline;
    end
    if *Deadline is elapsed* **then**
        reset the `StoredChannels`;
        Return `ArgMinLikelihood` and `MinLikelihood` // selection after deadline;
    end
end

The algorithm may be implemented at the level of the aggregated motor cortex (Eq. (2)), in which case the competing channels (`CurrentChannels`) are the distinct values $s(r_0, j_0)$ of the salience array resulting from the discretization of the space r_0, j_0 (see also, [40]).

The algorithm can also be used at higher intentional levels: for example the weighted individual activation patterns of each intention ($w_i s_{g_i}(r_0, j_0)$) may be first summarised into scalar channels S_i by means of an appropriate aggregation operator. In this case the competition occurs among the intentions (the `CurrentChannels` are the S_i).

The competition between channels is enacted by the scalar term **Log**(**Total**(**Exp**(`StoredChannels`))), added to all `NegLogLikelihoodChannels` channels at each iteration. The **Total** operator works across the temporal dimension, i.e., by summing the `CurrentChannels` recorded in `StoredChannels` list. Once the log likelihood crosses the given `Threshold`, the action becomes selected. The `Threshold` has to be tuned such that some predetermined error rate is permitted. If the threshold is not passed before a given `Deadline`, the algorithm can be stopped by taking the most likely optimal choice accrued so far.

2) Trajectory instantiation

Once an action $\{r_0, j_0\}$ is selected, one can propagate the selection backwards in the dorsal stream finding (symbolised by the dashed red arrow in Fig. 3), for example, which object, and at which future time, is limiting the movement (for example Fig. 5); or which is the intended lane (Fig. 4). Then, (only) the trajectory to be actually used is computed with the necessary details and forwarded to the motor system (Fig. 4, blue arrow). This idea is also consistent with the architecture proposed by Meyer and Damasio [27], in particular where backwards signalling is foreseen.

3) Minimum commitment principle

The selection of one instantaneous action $\{r_0, j_0\}$, when propagated backwards in the dorsal stream, often identifies multiple goals that are compatible with $\{r_0, j_0\}$ (albeit one is the strongest). For example, in Fig. 3 the intention of lane keeping (*a*), and the possible intention of changing lane later (*c'*) map onto the same instantaneous control. Hence, with selection of the peak *a* in Fig. 3, bottom, the agent is also "keeping the door open" for *c'*. Choosing between *a* and *c'* does not require an immediate selection (at the level of $\{r_0, j_0\}$) and, with the choice of the instantaneous action the agent carries out only the minimum commitment possible: i.e., it chooses all the trajectories that share the same control with *a*, and excludes only *c* and *b*.

E. INTEGRATING TRAFFIC REGULATIONS VIA BIASING ACTION SELECTION

So far, the behaviours emerge from a proper architecture and the physical awareness of the environment. However, driving is also a matter of regulation (for example, one should not cross solid lane markings). The question of how to teach the traffic rules to an artificial driver may be solved, once again, with biological inspiration [25]. In particular, we exploit the idea that behavioural choice can be steered by biasing low level, motoric action selection with higher level goals (Fig. 2, "higher-level action biasing" loop). In this way, modules that implement rules can act on the agent by specifying desirable and undesirable space-time locations. The high level rules are used to bias individual intentions (e.g., multiplying the individual activations (after inhibitions) $s_{g_i}(r_0, j_0)$ by gains w_i before combining the aggregated motor cortex as in (2).

An example is given in Fig. 6, where the intention of remaining in lane may be artificially strengthened (green, peak *a*) whereas the possibility of turning right may be

artificially weakened (yellow, peak *b*). This way, all three possible actions are passed to the selection process, but *a* is recommended and *b* discouraged. If, for example, an obstacle were severely inhibiting the recommended lane keeping intention *a*, the Agent would resort to *c* and then *b* with this priority order. The biasing weights may be hand tuned with simulation (the process is not very critical) or, as we have recently shown, the weights may be learned within a low-dimensional Reinforcement Learning problem (e.g., [41]).
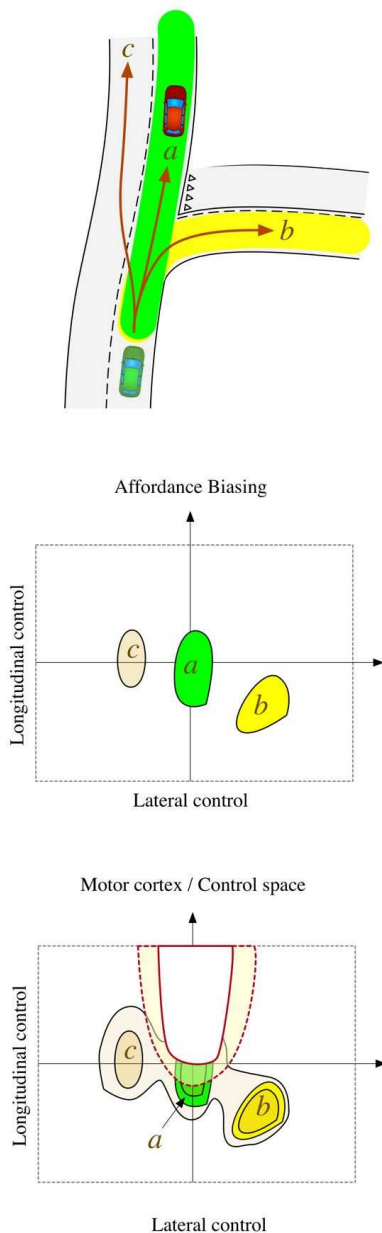

FIGURE 6: Action biasing principle. Action selection can be steered towards long-term goals by weighting the salience of individual intentions; for example increasing the salience of *a* and decreasing the salience of *b* (centre). Notably, the process of action biasing is safe because inhibited actions remain unselectable (bottom).

1) Bias versus lower-level veto

Biasing, as described above, can be used to program traffic rules by strongly recommending or discouraging particular actions. However, notice that biasing does not completely preclude an action, which may still be executed if its pre-bias salience is high enough; a situation which might occur in safety critical situations such as collision avoidance. This idea is developed further in Section III-C.

However, action biasing can only work in the space of safe and possible actions. That is, if part of the action space is inhibited then, whatever the biasing weights, no action may be performed in that sub-space. No completely inhibited action can ever be selected; the main sensorimotor loop will not implement recommendations that correspond to unsafe actions, a feature we call the "principle of lower-level veto". This relieves the need for testing safety of the higher-level biasing loops.

F. ADDITIONAL IMPLEMENTATION DETAILS

This paper has focused on sensorimotor principles in order to describe the whole picture. Many details that could not be fitted into the main narrative here can be found in the public deliverables of the Dreams4Cars Research Action [1].

III. OPEN ACCESS DEMONSTRATIONS

This section presents four different examples that, together, demonstrate the flexibility of the Agent architecture, which has been explained above from a principled/theoretical point of view.

A) The first example demonstrates complex adaptive and explainable behaviours emerging from the affordance competition principle; highlighting, in particular, the importance of the topographic organization of the motor cortex (Sections II-B to II-D).

B) The second example demonstrates robust action selection by comparing the commonly used Winner Takes All (WTA) selection criterion versus the Multi-hypothesis Sequential Probability Ratio Test algorithm (MSPRT) (Section II-D); in particular showing that the latter yields stable decisions at the cost of a minimal self-adapted increase in the decision time.

C) The third example demonstrates higher-level action biasing (Section II-E); in particular showing increased driving efficiency obtained via proactive steering of low-level action-selection.

D) The final example demonstrates hierarchical action-selection producing adaptation of the Agent to rapid unexpected events by, if necessary, forcing the traffic rules and choosing the lesser evil; highlighting, also, the importance of dense topographic organization of the motor space (Section II-B).

The examples (with the exception of the second, that includes real data experiments and a different simulation platform) may be found in the Open Data repository [3], where they can be reproduced or played. The environment

used for these is the OpenDS open source driving simulator[3] [42].

A. EMERGENT ADAPTIVE BEHAVIOURS

The example is given in Fig. 7, which presents the case of the Agent car (the black car) driving in a very wide lane, with two (red) vehicles —one is not visible yet in the camera view— closely following on both sides (time: 6.0 s). A standing (red) car is also present in the centre of the lane far ahead, which is also not visible yet.

The frames show a camera view (left) and the density plot of the motor cortex values (right). The dark blue circle is the selected action. The distinction between green and white is made to mark the longitudinal controls that comply with the speed limit. So, any choice in the green area does not violate the speed limit but, of course, the fastest option is on the boundary between green and white. For clarity the decrease of salience laterally, due to the lane limited width, is not shown. In the camera view, the pink line visualizes the instantaneous selected trajectory, which is generated following Section II-D2.

At time 6.0 s the inhibitions caused by the nearby cars (the right one is not visible yet in the camera view) appear in the motor cortex (red is total inhibition and yellow is partial inhibition). The Agent is travelling straight and hence its choice (blue circle) is not affected. Nonetheless one could say that the Agent is "aware" of the presence of the two vehicles because its motor space "reports" that some actions are no longer possible.

At time 11.5 s the far vehicle ahead is detected causing an inhibited region that overlaps with the previous current choice. At time 12.0 s the agent finds that it is possible to keep running at maximum speed by steering to the right. That means that the Agent "thinks" to pass on the right of the far obstacle and ahead of the vehicle that is following on the right.

This illustrates how the agent decisions can be explained. We, for example, know that if such an option was to be discarded from the beginning, a greater safety gap (yellow inhibition) should have been taught to the agent.

At time 13.3 s the option for passing on the right and ahead of the following vehicle is no longer safe enough (dark yellow), thus the agent opts for remaining in the centre of the lane and reducing the speed according to the distant obstacle. However, as the Agent reduces its speed a gap opens on the left side.

At time 16.9 s the agent makes the decision of passing on the left behind the left vehicle. Between time 16.9 s and 19.5 s we can see how this intention is maintained (no need for further revising it) and the manoeuvre that follows is exactly what could be expected. Eventually, after the overtake has been performed, at time 22.2 s the agent decides to return to the lane center.

Overall, the example shows the emergence of complex adaptive behaviours from basic principles, and from the way the sensorimotor architecture is organized (there were no rules programmed, as such). The agent decisions can always be clearly explained, and it is also clear what could be tuned for modifying the behaviours.

1) Relation to agents with programmed behaviours

The same situation was run using an agent with rule-based behaviours that could be transparently accessed. Here, the agent remained stuck behind the stopped vehicle, as if it were unable to make use of the unusual width of the lane; or as if the stopped vehicle were schematically blocking "the (entire) lane". Another shortcoming was found earlier in a scenario consisting of a pedestrian incorrectly crossing the road when the traffic light turns red for the pedestrian and green for the car, as shown in [43].

Nonetheless, when comparing agents with emergent behaviours – like this one – to agents with programmed behaviours – for example implemented with finite state machines – our examples cannot be generalised to say that the latter necessarily underperforms. First, the mechanism of the choice among different alternatives (albeit perhaps more schematic) also exists in more traditional designs as mentioned in Section I-B1, and so they also exhibit some degree of adaptive behaviour. Second, with programmed behaviours the outcome depends on which behaviours have been implemented. If some shortcoming is detected the program can be updated – for example new rules or states and transitions can be added.

Hence, we argue that, while in principle a given ability can be obtained with both approaches, the difference is in the effort needed for development. We argue that the development of agents that are programmed in detail is going to be more laborious, and requires us to identify every possible situation and identify how to operate therein. Conversely, and agent with emergent behaviors tends to be more robust, producing correct behaviours more often. Debugging is necessary also for the latter, but occurs more at the level of testing the implementation of principles like correctly computing inhibitions, correct biasing etc..

B. ROBUST ACTION SELECTION: WTA VS MSPRT

This scenario, shown in Fig. 8, is studied with IPG Car-Maker; an industry-standard simulation tool that was used in Dreams4Cars to create a virtual validated model of one real test vehicle (a Jeep Renegade). Since the overtake scenario was really tested on the Jeep Renegade (WTA only) and reproduced in simulation with CarMaker (both WTA and MSPRT), hence the example is presented with CarMaker instead of OpenDS data.

The scenario realizes one situation where two actions become equally salient, which is ideal to evaluate the effect of noise in action selection. Thus, re-entering from the left lane after an overtake manoeuvre may, at some point and conditions, form two almost equivalent choices (Fig. 8, point
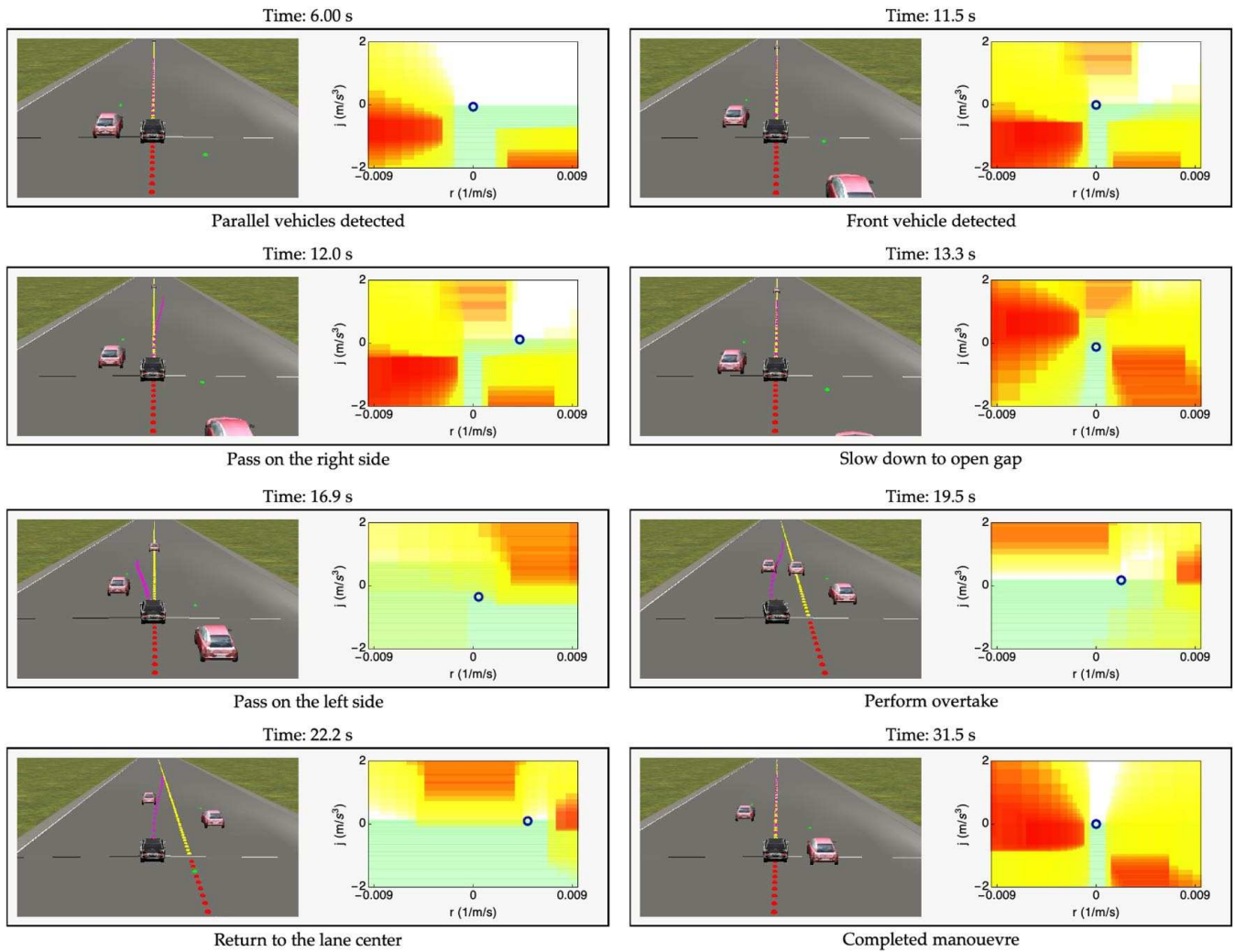
FIGURE 7: Example of emergent behaviours. The motor cortex activation is shown next to the camera view for different moments of the simulation. The inhibited (red/yellow for total/partial inhibition) and affordable actions (green) and the agent choices (blue circle) are easily explainable (see text for more comments), generating an, overall, complex behaviour that corresponds to what might be commonly expected.

2). The conditions are here explained: one way for returning to the right lane may be to bias the right lane choice (Fig. 8, *b*) according to Section II-E. The bias can be set in advance, such as, e.g, before point 2 in Fig. 8, or may even be let set forever. In both cases the lane change will occur only as soon as a collision free manoeuvre is available (i.e., at point 2) and not earlier (lower-level veto). Advance biasing is thus a way to induce right lane changes to occur as early as possible.

Fig. 9 compares the steering rate[4] r_0 for WTA and MSPRT. At point 1, travelled distance $86\,\mathrm{m}$, a stopped obstacle is detected for the first time $100\,\mathrm{m}$ ahead. Of the two options, stopping behind the obstacle (dashed arrow, Fig. 8) immediately has considerably less value than overtaking (solid line arrow). In this case the lane change is by far the winner action and no competition occurs in practice with the stop option. The

[4]The "steering rate" is defined here as the rate of change of curvature of the vehicle trajectory.
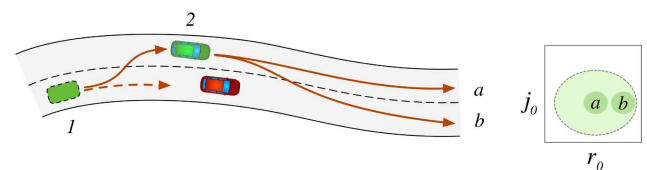

FIGURE 8: Returning to the right lane after an overtake. At point 2, actions *a* and *b* have the same salience. With noise, the WTA criterion may cause oscillations, which are not produced with MSPRT.

requested steering rate (that represents the selected action r_0) jumps immediately to approximately $0.02\,\mathrm{m}^{-1}\mathrm{s}^{-1}$ which corresponds to the lane change manoeuvre. The executed steering rate follows with a slight delay due to the steering actuator lags and to the vehicle dynamics lags (it is the
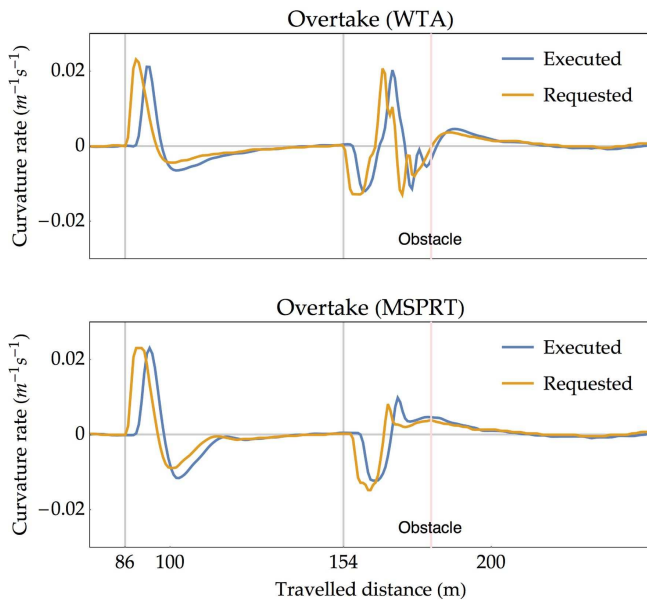
FIGURE 9: Returning to the right lane after overtake. WTA criterion causes oscillations, which are not seen with MSPRT (the oscillations of the steering rate control happening before the obstacle location for WTA).

requested steering rate that represents the selected action).

A different situation happens at point 2. When this point is approached the right lane affordance becomes gradually stronger until, at point 2, travelled distance $154\,\mathrm{m}$, the salience of the action b slightly tops the salience of a. The WTA selection (Fig. 9, "Requested" line on top) occurs as soon as b tops a and the execution follows with the usual delay. However, because of noise in the perception and the motor control, once the manoeuvre is initiated, there are oscillations in the obstacle inhibitions strength and fluctuations of the salience of b compared to a, which revert the decision as shown with the fluctuations in the requested and executed steering rate. These oscillations last until the obstacle is finally passed. The MSPRT selection is shown in (Fig. 9, bottom). In this case, since there is a competition between two almost equally strong affordances a and b, the MSPRT algorithm delays its decision until there is enough evidence that b is the winner action. The delay (the drop of the "Requested" line) is not severe: about $100\,\mathrm{ms}$–$150\,\mathrm{ms}$ as marked by the slight distance between grid line at $154\,\mathrm{m}$ and the drop of the orange line in the chart. On the other hand the decision is stable (the steering rate returns slight positive when the trajectory must be straightened to enter the destination lane). Note also that at point 1, when there was no doubt about the choice, MSPRT was almost as fast as the WTA. In this example MSPRT adapted the decision time.

C. (PROACTIVE) ACTION BIASING

For this demonstration, we considered a three lanes motorway scenario with stochastic traffic. The comparison takes place by running the same situations with two versions of

the agent: with and without biasing the selection of lanes. The objective is to show that, with proper biases, the agent develops more effective longer-term behaviours and, also, complies with the rule of using the right-most free lane, when available.

A total number of 100 simulations were performed, 50 with and 50 without the lane selection biasing. Every simulation corresponds to driving along a $5\,\mathrm{km}$ straight section of the motorway. The traffic is initialised by placing a random number of vehicles from a discrete uniform distribution between 30 and 70 vehicles. Each vehicle is placed on a random lane (left, centre, right) at a random distance drawn from a uniform distribution between $50\,\mathrm{m}$–$1750\,\mathrm{m}$ ahead of the agent's vehicle (avoiding traffic vehicles overlaps). The vehicles on the left lane are assigned a random velocity between $100\,\mathrm{km/h}$–$110\,\mathrm{km/h}$ (uniform distribution); on the centre lane between $80\,\mathrm{km/h}$–$90\,\mathrm{km/h}$; on the right lane between $50\,\mathrm{km/h}$–$70\,\mathrm{km/h}$. The ego vehicle target speed is set to $140\,\mathrm{km/h}$ such that it will overtake to run faster whenever the possibility occurs.

1) No biasing

In this version of the Agent, the salience of the affordable lanes is not biased (the weights w_i in Eq. (2) are the same for all the three lanes). The agent will remain in its lane until a leading vehicle begins inhibiting the free-flow longitudinal control (such as, e.g, Fig. 6, bottom). When the slow-down caused by this inhibition becomes large enough, a nearby lane, either left or right, will be selected if less inhibited (for this to happen the slow-down cost must exceed the lane change cost). Hence the Agent has to first enter a car-following penalising situation before feeling the need for a lane change.

2) Biasing lane selection

In this version, lane selection is biased according to the following logic:

- if the right lane is not slower than the Agent target speed, then the right lane is biased;
- else, if the current lane is slower than the target speed and the left lane is faster, then the left lane is given priority.

The speed of one lane is determined by the slowest vehicle travelling in the lane segment from the current position to a given distant horizon. This may include far vehicles that are not yet causing car-following. The logic is proactive, anticipating lane change before car-following conditions are reached. The speed is topped by the speed limit of the lane. If the lane is free, the lane's speed is the speed limit.

With the first criterion, the ego vehicle gradually moves to the rightmost lane if that does not reduce its travel time. With the second criterion, the ego vehicle moves to a left faster lane if that improves its travel time.

One important remark is that the evaluation of the criteria can be carried out on a simple schematic instantaneous

situation without needing precise evaluation of objects trajectories. The criteria, in fact, produce only priorities for the action selection. For example, if the criteria set the priority for a lane change, but there is a vehicle cutting in, the choice will be rejected because the obstacle will be inhibiting the lane salience.

Another remark is that these criteria, as formulated above, permit passing on the right. This may/may not be allowed depending on the particular road and highway code (preventing passing on the right would require additional criteria).

3) Results

Table 1 shows: *a*) the percentage of the time spent in the car-following condition, *b*) the mean time spent in one lane between lane changes and *c*) the average travelling velocity. Biasing yields a reduction of 32% (from 74.5% to 50.5%) of the car-following time, which is obtained with more frequent lane changes (from 31.9 s down to 17.9 s between lane changes) and an higher average speed (from 96.7 km/h to 109.5 km/h). In evaluating the latter, one must consider that the leftmost lane is populated with vehicles travelling at speeds between 100 km/h–110 km/h: the mean travelling speed is thus the average between car-following conditions and the occasionally faster free-flow conditions.

TABLE 1: Highway simulation scenario results.

case	mean time in car-follow (%)	mean time in one lane	mean velocity
"bias"	50.5	17.9 s	109.5 km/h
"no bias"	74.5	31.9 s	96.7 km/h

The vehicle velocity, per lane, is shown in Fig. 10 and Fig. 11, respectively for the "bias" and no "bias cases". So, in Fig. 10 the bottom sub-chart (red, right lane) shows the vehicle speed in the 50 simulations when the vehicle was in the right lane. One can notice a large number of cases where the speed drops from 140 km/h to 50 km/h–70 km/h. These are transitions from free flow to car-following, when a leading vehicle is approached. The car-following condition does not last much: as soon as it is possible, the car changes to the central lane (the speed profile continues in the middle lane chart, green). Often, the lane change happens already during the slow down phase, before entering the constant speed car-following condition. In only one case (at about 4500 m), the vehicle remained trapped in car-follow for some time. The middle lane chart shows the acceleration phases of the vehicle coming from the right lane (short traits below 50 km/h–70 km/h as well as further acceleration to 140 km/h when the lane is free. Notice that the vehicle never remains in car-following in the middle lane, because it has twice lane change opportunities (either to the left or to the right). The left lane chart shows accelerations from the vehicle coming from the middle lane, as well as transitions from full speed down to the 100 km/h–110 km/h traffic speed. In this lane the vehicle remains often trapped behind the fast traffic. The "no
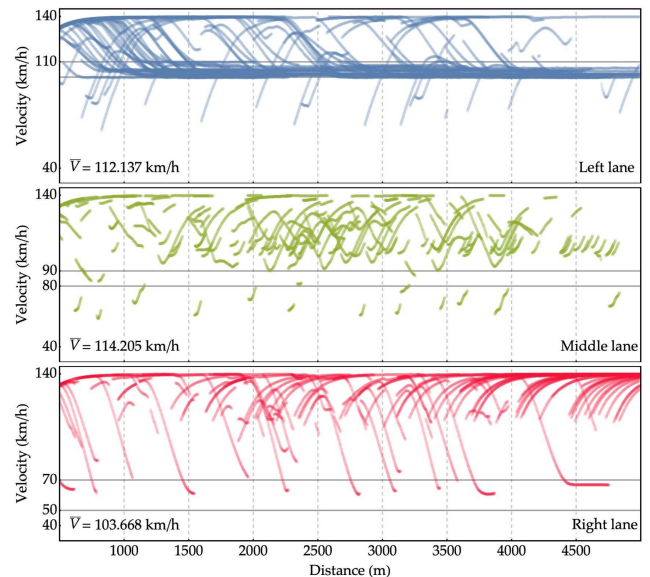

FIGURE 10: Lane selection with proactive bias ("bias" case). Velocity profiles of the 50 simulations. The speed of the vehicle is represented in the sub-chart of the lane where the vehicle is located. From top to bottom: the left, centre and right lanes are represented with blue, green and red respectively. The profiles on one lane are interrupted where the vehicle changes lane .
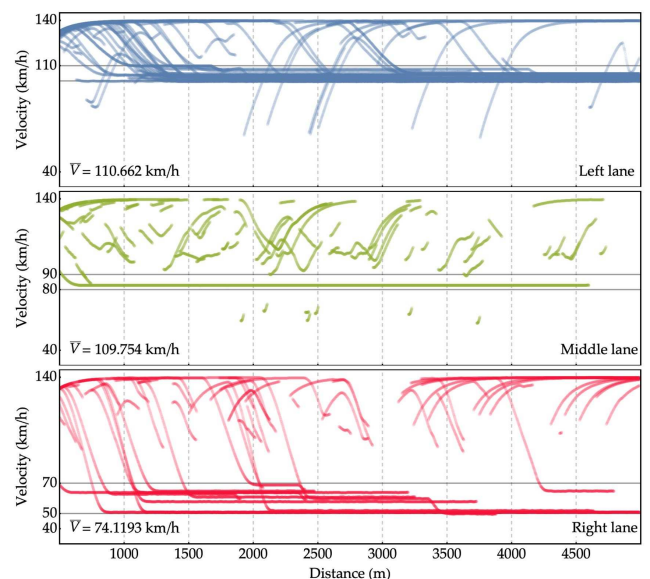

FIGURE 11: Lane selection without proactive bias ("no bias" case). Compared to the previous figure, the velocity profiles reveal that, without proactive lane change, the agent may remain frequently trapped behind slower vehicles in both the middle and the rightmost lane.

bias" case (Fig. 11) shows a less efficient behaviour that is

produced without the proactive promotion of lane changes. In particular, one can notice that the ego vehicle remains easily trapped in the traffic, both in the right and, also, middle lane. Furthermore, the lane changes from the right to the middle lane are less frequent in the slow-down phase. The mean travelling speed is annotated per lane. While there is no great improvement in the leftmost lane, a significant improvement is observed in the middle and especially the rightmost lane, which is conveniently used whenever it is free. The general conclusion is that proactively biasing action selection, by anticipating virtuous manoeuvres, increases efficiency.

D. HIERARCHICAL ACTION-SELECTION

This final example investigates the adaptation of the Agent to rapid unexpected events, if necessary responding by forcing the traffic rules and choosing the lesser evil. The study is inspired by the situation shown in Fig. 1. The ego car travels on a straight three-lane motorway following a vehicle that, unexpectedly drops an object (a traffic cone in the simulation, see schematic representation in Fig. 12). In the longitudinal direction the motion of the fallen object is uniformly decelerated. In the lateral direction it weaves from left to right with a linearly increasing amplitude, as defined by equation (3), which will be later commented.

The traffic is travelling at $60.0\,\mathrm{km/h}$, which is also the initial speed at which the cone is released. In order to have a schematic situation that helps interpretation, the spacing of the vehicles is uniform, but the relative position Δs of the next vehicle in both lanes may vary (Fig. 12).

Parametric simulations have been carried out by varying:

1) the traffic density, which may be either "high" (vehicle separation $1.8\,\mathrm{s}$, or $30\,\mathrm{m}$) or "low" (separation $3\,\mathrm{s}$, or $50\,\mathrm{m}$),
2) deceleration of the fallen object, $a_{\mathrm{cone}} \in \{-5.0, -4.5, -4.0, -3.5, -3.0, -2.5, -2.0\}\,\mathrm{m/s^2}$;
3) time headway between ego and leading vehicle $T_h \in \{1.25, 1.5, 1.75, 2.0, 2.25, 2.5\}\,\mathrm{s}$.

For each case, i.e. the combination of one traffic density (high/low), one value of a_{cone} and one value of T_h we performed 20 simulations, for a total of $20\times2\times7\times6 = 1680$ simulations.

Each simulation was carried out with partially random lateral cone movements, generated by (3) with phase ϕ_0 drawn from a uniform distribution $\mathcal{U}\left(-\pi/4, \pi/4\right)$. Hence, the lateral displacement of the cone $y_c(x)$ is given as a function of the longitudinal distance x travelled by the cone

$$y_c(x) = 0.05\,x\sin\left(\lambda x + \phi_0\right) + \mathcal{U}\left(-0.05, 0.05\right), \quad (3)$$

where $\lambda = 20\,\mathrm{m^{-1}}$ is the period of the oscillations whereas the amplitude increases linearly with the travelled distance x.

Also, while the traffic was regularly spaced, the relative position Δs of the next vehicle ahead (there are also vehicles behind) on both the left and right lanes was set randomly from a random uniform distribution $\mathcal{U}(0, L)$, with L being the spacing between the vehicles in the traffic ($L = 30\,\mathrm{m}$ for high density traffic and L = $50\,\mathrm{m}$ for low density traffic).
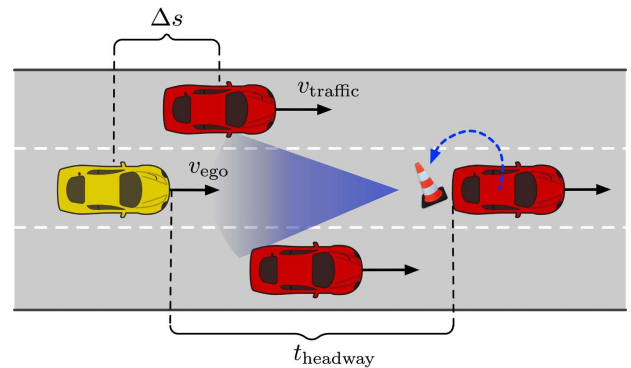
FIGURE 12: Schematic representation of the "US 101" scenario (Section III-D). The road cone falls from a leading car and follows a partially random trajectory. The ego car, yellow vehicle, attempts to either brakes or force change lane (including fitting in between lanes).

To model mechanical and actuation delays, the ego vehicle deceleration rate (longitudinal jerk) cannot exceed $-10\,\mathrm{m/s^3}$. This means that it takes $0.5\,\mathrm{s}$ to reach $-5\,\mathrm{m/s^2}$, which is the largest cone deceleration considered (in a real vehicle this figure may vary depending on the braking plant characteristics). Also, because the Agent re-planning loop runs at $20\,\mathrm{Hz}$ there is an additional $50\,\mathrm{ms}$ delay between perception and action request (this figure may also vary). Finally, we assume that the range sensor and perception system detecting the cone measure the distance and the velocity of the cone, but does not estimate the acceleration. Because of these limitations, and depending on the time headway at the moment of the object fall the collision might indeed be unavoidable.

Finally, it has to be said that, for this study, the mechanism for proactively changing lane, as described in Section III-C, is not active (otherwise the car would change to lanes as soon as they have farther obstacles and will not permanently remain in car-following, which is necessary for this study)

1) Results

With high-density traffic, there are no gaps in the adjacent lanes where the vehicle can fit while fully preserving the desired longitudinal safety distances. Yet, these gaps ($1.8\,\mathrm{s}$) are wide enough to fit inside without collision[5], especially if the car does not move entirely into the destination lane but stays on the lane border over the lane markings. Interestingly, the choice of such a "least evil" behaviour is generated by the play between partial and total inhibitions: with a dense "motor cortex" many intermediate actions exists among which some collision free choice, albeit not prefect, can be found.

This is shown in Fig. 13 which presents two instants of the same simulation. On top, the ego vehicle (the black car) follows the "dangerous" car before the object falls. When, sometimes later, the cone falls (time: $15.0\,\mathrm{s}$), the Agent

---

[5]Moving into a $1.8\,\mathrm{s}$ gap, the host vehicle is going to leave only a fraction of a second gap between the vehicle in front or behind.

initially adapts to the reducing speed of the obstacle until, at some point (time: $16.5\,\text{s}$) the action corresponding to squeezing between the current lane and the right lane change is chosen despite being partially inhibited by a nearer car behind (darker yellow area in Fig. 13 bottom). As shown in this example, the agent does not move completely into the right lane but shifts laterally only as much as needed to avoid the obstacle.
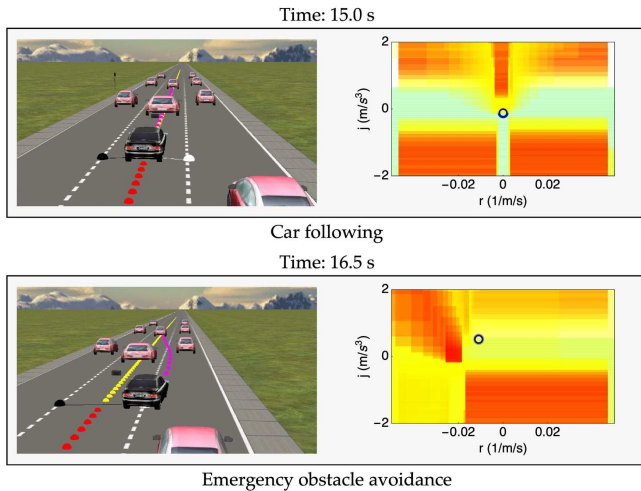

FIGURE 13: Example of emergency avoidance behaviour. The Agent chooses a risky action (bottom figure) to avoid the collision with the dropping obstacle.

Fig. 14 shows the trajectories of the ego vehicle for different combinations of cone deceleration and time headway. We distinguish the results in terms of three possible outcomes: *overtake*, in case the ego vehicle clears the cone by passing laterally and continues travelling along the road (blue lines); *stop*, if the ego vehicle stops without hitting the obstacle (green lines); *collision* if a collision occurs either in the attempt of passing or stopping (red lines). Similarly, Fig. 15 shows the trajectories for the high density traffic case.

Concerning trajectories that clear the obstacle (blue lines), one can notice that manoeuvres corresponding to travelling in between the lanes, such as the one shown in Fig. 13, are not very frequent. This is because the traffic reacts and it often happens that when the ego vehicle moves in front of one vehicle, this vehicle opens a greater gap which lets complete the lane change (this is revealed by changes in the trajectory shapes that follows adaptations of the traffic, especially at short time headways).

By counting the number of clear, stop and collision manoeuvres the probability of the three outcomes can be estimated (somewhat coarsely because there are only 20 simulations per case). These are shown in Fig. 16 and Fig. 17 respectively for the low and high density traffic cases.

If the low density traffic is considered (Fig. 16 and Fig. 14) one can notice the presence of 4 different clusters. For deceleration below $4\,\text{m/s}^2$ and time headway above $2\,\text{s}$ the Agent stops. Instead if the time headway is below $2\,\text{s}$ the
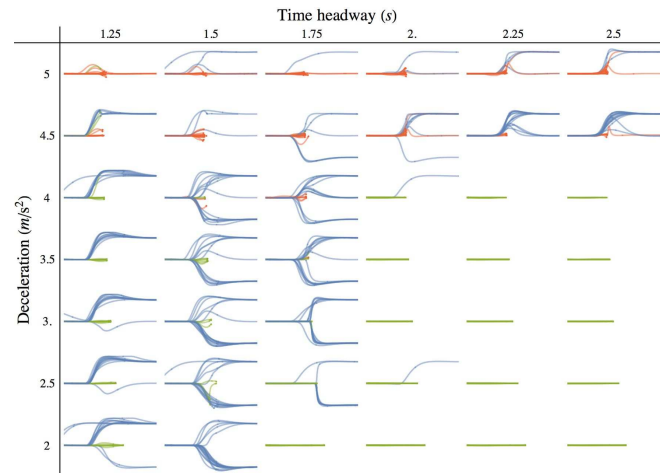

FIGURE 14: Low density traffic. Trajectories as function of time headway and cone deceleration with distinction for collision (red), stop (green) and clear (blue).
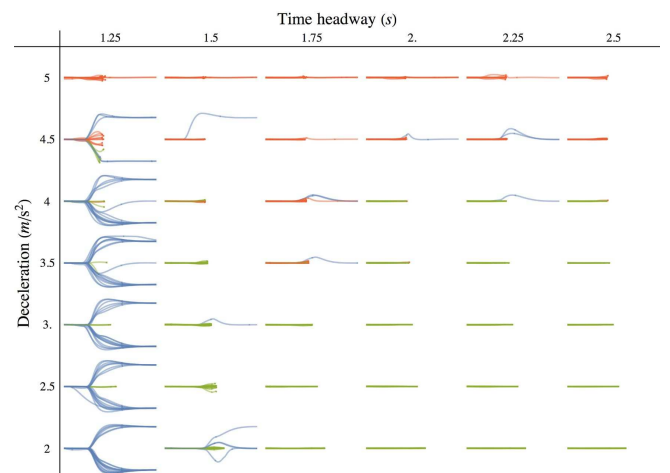

FIGURE 15: High density traffic. Trajectories as function of time headway and cone deceleration with distinction for collision (red), stop (green) and clear (blue).
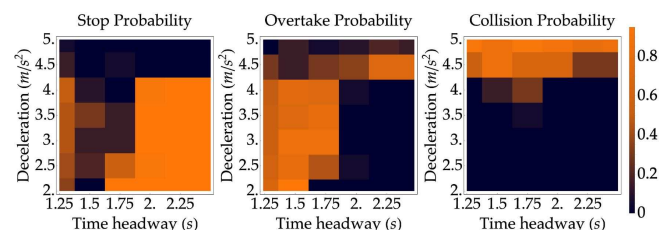

FIGURE 16: Low density traffic. Probability for collision, stop and clear manoeuvres. With high deceleration of the cone, the most probable outcome is collision. Otherwise, if the time headway is short, the agent tries to force a lane change (detailed discussion in the text).

Agent prefers to overtake. However, if the time headway is very small, $1.25\,\text{s}$ the Agent stops or overtake with similar probabilities. Which happens depends on the what is on the
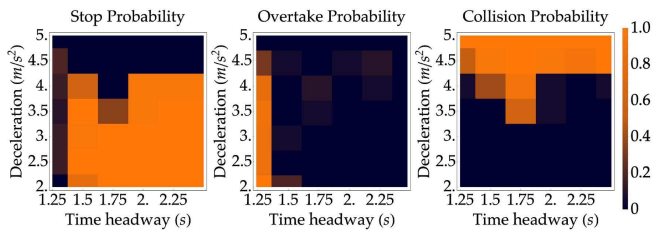
FIGURE 17: High density traffic. Probability for collision, stop and clear manoeuvres. With higher density of the traffic, forcing a lane change is less likely (detailed discussion in the text).

side: gaps or obstacles. For deceleration above $4.5\,\mathrm{m/s^2}$ a collision almost always occurs (with slight greater probability for shorter time hideaways). The collisions are due to the (assumed) limitation in the perception system that does not estimate the obstacle deceleration. Without knowing the deceleration, the average future velocity is systematical overestimated and this error is larger for larger deceleration until a collision may happen. The high density traffic is similar (Fig. 17 and Fig. 15) but in this case, below $4\,\mathrm{m/s^2}$ (where the car can stop) we see an expansion of the stop choice, because, of course, there is no completely free gap on both sides. When the time headway is very small ($1.25\,\mathrm{s}$) the Agent forces the lane change such as in Fig. 13. If the deceleration is above $4\,\mathrm{m/s^2}$ collisions occur, with higher probability and a similar pattern (worse for shorter time headway).

E. OTHER EXAMPLES

A project video with other examples is given in [44]. and we draw attention to several specific segments. Thus, at video times 0:37-1:29 there are four examples of emergent longitudinal behaviours (car following, stopped obstacle, suddenly braking leading car and pedestrian walking in the middle of the road), which are generated, similarly to those in Section III-A by the mechanism of Section II-C5.

At video times 1:29-1:53 there are two other examples of obstacles suddenly entering the road (a walking and a running pedestrian).

Proactive behaviour is shown at times 1:53-2:10. The example consists of the preventive adoption of a safe reduced speed in the proximity of a pedestrian that might enter the road (before the pedestrian starts moving). This high-level behaviour is further discussed in Section IV.

Action selection is presented between times 2:10 and 3:14. First, the effect of sensor and motor noise with the WTA selection criterion is shown. Then, better evidence-based MSPRT robust action selection is shown beginning at time 2:40.

Complex behaviours in a motorway scenario, generated by affordance competition with proactive biasing of action selection (similarly to Section III-C) is shown between times 3:53 and 4:51.

Finally, complex behaviours in a urban-like scenarios, which includes intersections, are shown beginning at time 4.52

IV. EXTENSIONS: LEARNING VIA MENTAL IMAGERY

While this paper dealt with emergent autonomy, the agent described here was conceived with the final goal of learning and optimizing behaviours via a process that might be called "mental simulation", which is inspired by the ability of humans to use mental imagery to explore future possibilities, including those in the dream state while sleeping [45]–[47].

This form of learning for artificial agents was the main goal of the Dreams4Cars project [1] (point *b* in the introduction), and was successfully achieved by building on the agent architecture described here. These results will be shown, in full, in future publications, but preliminary reports are available in the public project deliverables (such as D3.3, D7.3., that can be found in the project website), in [48]–[50] and in two communication project videos [44], [51].

In learning via mental simulation, the Agent does not interact with the plant/environment directly (which would be a form of Reinforcement Learning); instead, it first learns a model of the plant/environment and then interacts within that model. In this way, the agent can test actions that would be dangerous in the real world. Also the agent is focused on constructing its own predictive and control models, which can be tested and improved the next time that agent acts in the real world. Finally, a less evident, but actually very important point, is that, once a model of the environment/plant is available, there are ways to synthesize inverse models for control and behaviour that are more efficient than trial and error direct interaction with the learned models as shown next.

Learning with mental simulations requires additional loops that are not shown in Fig. 2. There are, in particular, two main loci/methods for learning.

One approach bootstraps the Agent sensorimotor system bottom up, starting with the learning of low-level forward models and, via progressive manipulation of these models, synthesizing low and higher-level forward and inverse models that may become particular instantiations of the motor output loop (motor control) and of the dorsal stream (new excitatory and inhibitory circuits representing the ability to detect/prime new/better affordances). Two examples of neural network controllers for predictive lateral control learned with this process are visible in the video [44] at, respectively, time 3:15 and time 3:29.

A second locus of learning is at the action selection and consists of learning the biases that produce longer terms rewards, while transversing in the short term states of little value (for example transversing a slow lane to get to a faster one could be an extension of the mechanism used in Section II-E). This can be regarded as a form of Reinforcement Learning (among the affordable actions). A slightly different variant of this method concerns the learning of behavioural parameters (for example the choice of a safe speed related to

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI
10.1109/ACCESS.2020.3007018, IEEE Access

IEEE *Access*

M. Da Lio *et al.*: Agent Architecture for Adaptive Behaviours in Autonomous Driving

a given context). This is also learnable quite efficiently with Reinforcement Learning. Two examples of this are given in the video [44] at, respectively, time 1:53 and time 5:04.

The project video [51] (simplified for communication purposes) gives an overall description of both the goals mentioned in the introduction, and clarifies how the agent architecture described in this paper functions with learning via mental imagery.

V. CONCLUSION

This paper described a sensorimotor architecture based on a few, biologically inspired principles, which is capable of producing adaptive autonomy while incorporating logical criteria. The position is supported by open accessible working examples. The same scenarios could also be used as benchmarks. Additional examples (not described in this paper) that extend the demonstration abilities may be found in the ZENODO repository [52]. This paper focused on the high level architectural principles and there are several ways to implement them; additional details for the Dreams4Cars implementation can be found in the public deliverables [1]. Explainability of the agent decision is, also, of prime importance. The topographic organization of the motor space is a key contributor to this feature because the competing actions can be identified, and the selected action can be back traced in the dorsal stream finding the intentions that generated the selection.
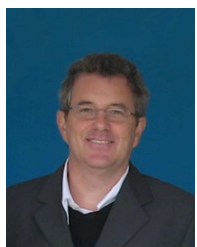
REFERENCES

[1] "Dreams4Cars - Dream-like simulation abilities for automated cars." [Online]. Available: https://www.dreams4cars.eu/en

[2] 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.

[3] M. Da Lio and R. Donà, "Dreams4Cars: Agent Architecture for Adaptive Behaviours in Autonomous Driving — supplementary material," May 2020. [Online]. Available: https://doi.org/10.5281/zenodo.3813250

[4] 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.

[5] C. Katrakazas, M. Quddus, W.-H. Chen, and L. Deka, "Real-time motion planning methods for autonomous on-road driving: State-of-the-art and future research directions," Transportation Research Part C: Emerging Technologies, vol. 60, pp. 416 – 442, 2015.

[6] B. Paden, M. Čáp, S. Z. Yong, D. Yershov, and E. Frazzoli, "A Survey of Motion Planning and Control Techniques for Self-Driving Urban Vehicles," IEEE Transactions on Intelligent Vehicles, vol. 1, no. 1, pp. 33–55, Mar. 2016.

[7] D. González, J. Pérez, V. Milanés, and F. Nashashibi, "A review of motion planning techniques for automated vehicles," IEEE Transactions on Intelligent Transportation Systems, vol. 17, no. 4, pp. 1135–1145, 2016.

[8] S. Pendleton, H. Andersen, X. Du, X. Shen, M. Meghjani, Y. Eng, D. Rus, and M. Ang, "Perception, Planning, Control, and Coordination for Autonomous Vehicles," Machines, vol. 5, no. 1, p. 6, Feb. 2017.

[9] L. Claussmann, M. Revilloud, D. Gruyer, and S. Glaser, "A Review of Motion Planning for Highway Autonomous Driving," IEEE Transactions on Intelligent Transportation Systems, pp. 1–23, 2019.

[10] C. Lu, J. Gong, C. Lv, X. Chen, D. Cao, and Y. Chen, "A Personalized Behavior Learning System for Human-Like Longitudinal Speed Control of Autonomous Vehicles," Sensors, vol. 19, no. 17, p. 3672, Aug. 2019.

[11] M. Bojarski, D. Del Testa, D. Dworakowski, B. Firner, B. Flepp, P. Goyal, L. D. Jackel, M. Monfort, U. Muller, J. Zhang et al., "End to end learning for self-driving cars," arXiv preprint arXiv:1604.07316, 2016.

[12] "Learning to Drive: Beyond Pure Imitation - Waymo - Medium." [Online]. Available: https://medium.com/waymo/learning-to-drive-beyond-pure-imitation-465499f8bcb2

[13] M. Otte and E. Frazzoli, "RRTX: Real-Time Motion Planning/Replanning for Environments with Unpredictable Obstacles," in Algorithmic Foundations of Robotics XI: Selected Contributions of the Eleventh International Workshop on the Algorithmic Foundations of Robotics, H. L. Akin, N. M. Amato, V. Isler, and A. F. van der Stappen, Eds. Cham: Springer International Publishing, 2015, pp. 461–478.

[14] A. Broggi, P. Cerri, S. Debattisti, M. C. Laghi, P. Medici, M. Panciroli, and A. Prioletti, "PROUD-Public road urban driverless test: Architecture and results," in 2014 IEEE Intelligent Vehicles Symposium Proceedings. MI, USA: IEEE, Jun. 2014, pp. 648–654.

[15] J. Gibson, "The theory of affordances," in The Ecological Approach To Visual Perception. Lawrence Erlbaum Associates, 1986, pp. 127–143.

[16] A. Chemero, "An outline of a theory of affordances," Ecological Psychology, vol. 15, no. 2, pp. 181–195, 2003.

[17] P. Cisek, "Integrated Neural Processes for Defining Potential Actions and Deciding between Them: A Computational Model," Journal of Neuroscience, vol. 26, no. 38, pp. 9761–9770, Sep. 2006.

[18] P. Cisek and J. F. Kalaska, "Neural mechanisms for interacting with a world full of action choices," Annual review of neuroscience, vol. 33, pp. 269–298, 2010.

[19] G. Marti, A. H. P. Morice, and G. Montagne, "Drivers' decision-making when attempting to cross an intersection results from choice between affordances," Frontiers in Human Neuroscience, vol. 8, Jan. 2015.

[20] M. Da Lio, A. Mazzalai, K. Gurney, and A. Saroldi, "Biologically guided driver modeling: The stop behavior of human car drivers," IEEE Transactions on Intelligent Transportation Systems, vol. 19, no. 8, pp. 2454–2469, 2018.

[21] T. J. Prescott, P. Redgrave, and K. Gurney, "Layered Control Architectures in Robots and Vertebrates," Adaptive Behavior, vol. 7, no. 1, pp. 99–127, Jan. 1999.

[22] R. Brooks, "A robust layered control system for a mobile robot," IEEE Journal of Robotics and Automation, vol. 2, no. 1, pp. 14–23, 1986.

[23] K. Gurney, T. J. Prescott, and P. Redgrave, "A computational model of action selection in the basal ganglia. II. Analysis and simulation of behaviour." Biological cybernetics, vol. 84, no. 6, pp. 411–23, Jun. 2001.

[24] 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.

[25] G. Pezzulo and P. Cisek, "Navigating the Affordance Landscape: Feedback Control as a Process Model of Behavior and Cognition," Trends in Cognitive Sciences, vol. 20, no. 6, pp. 414–424, Jun. 2016.

[26] R. Bolado-Gomez and K. Gurney, "A biologically plausible embodied model of action discovery," Frontiers in Neurorobotics, vol. 7, 2013.

[27] K. Meyer and A. Damasio, "Convergence and divergence in a neural architecture for recognition and memory," Trends in Neurosciences, vol. 32, no. 7, pp. 376–382, Jul. 2009.

[28] G. H. Patel, D. K. Michael, and L. H. Snyder, "Topographic organization in the brain: Searching for general principles," Trends in cognitive sciences, vol. 18, no. 7, pp. 351–363, Jul. 2014.

[29] J.-P. Thivierge and G. F. Marcus, "The topographic brain: from neural connectivity to cognition," Trends in Neurosciences, vol. 30, no. 6, pp. 251–259, Jun. 2007.

[30] M. A. Sommer and R. H. Wurtz, "Composition and topographic organization of signals sent from the frontal eye field to the superior colliculus," Journal of Neurophysiology, vol. 83, no. 4, pp. 1979–2001, 2000.

[31] 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 Vehicles, vol. 16, no. 1, pp. 244–263, 2015.

[32] D. A. Pomerleau, "Alvinn: An autonomous land vehicle in a neural network," in Advances in Neural Information Processing Systems 1, D. S. Touretzky, Ed. Morgan-Kaufmann, 1989, pp. 305–313.

[33] G. Schöner, M. Dose, and C. Engels, "Dynamics of behavior: Theory and applications for autonomous robot architectures," Robotics and Autonomous Systems, vol. 16, no. 2-4, pp. 213–245, Dec. 1995.

[34] K. L. Downing, "Predictive models in the brain," Connection Science, vol. 21, no. 1, pp. 39–74, Mar. 2009.

[35] M. Da Lio, A. Mazzalai, D. Windridge, S. Thill, H. Svensson, M. Yüksel, K. Gurney, A. Saroldi, L. Andreone, S. Anderson, and H.-J. Heich, "Exploiting dream-like simulation mechanisms to develop safer agents for automated driving: The 'Dreams4Cars' EU research and innovation action," in IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC, vol. 2018-March, 2018, pp. 1–6.

[36] A. Plebe, M. Da Lio, and D. Bortoluzzi, "On Reliable Neural Network Sensorimotor Control in Autonomous Vehicles," IEEE Transactions on Intelligent Transportation Systems, vol. 21, no. 2, pp. 711–722, Feb. 2020.

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

[38] 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.

[39] L. F. Nunes and K. Gurney, "Multi-alternative decision-making with non-stationary inputs," Open Science, vol. 3, no. 8, p. 160376, Aug. 2016.

[40] R. Donà, G. P. R. Papini, and G. Valenti, "MSPRT action selection model for bio-inspired autonomous driving and intention prediction," Macau, China, p. 4.

[41] G. P. Rosati Papini, M. Da Lio, R. Donà, and A. Plebe, "A Reinforcement Learning Approach for Enacting Cautious Behaviours in Automated Driving Agents: Safe Speed Choice in the Interaction with Distracted Pedestrians." IEEE Transactions on Intelligent Transportation Systems, vol. submitted.

[42] R. Math, A. Mahr, M. M. Moniri, and C. Müller, "Opends: A new open-source driving simulator for research," GMM-Fachbericht-AmE 2013, p. 2, 2013.

[43] "Dreams4cars Comparison With Other Autonomous Car Agents." [Online]. Available: https://www.youtube.com/watch?v=kOnEMwIB1gU

[44] "Explaining the Dreams4cars Autonomous Driving Capabilities." [Online]. Available: https://www.youtube.com/watch?v=ODuw9TNanE8

[45] C. T. Smith, J. B. Aubrey, and K. R. Peters, "Different Roles for REM and Stage 2 Sleep In Motor Learning: A Proposed Model," p. 24.

[46] H. Svensson and S. Thill, "Beyond bodily anticipation: Internal simulations in social interaction," Cognitive Systems Research, vol. 40, pp. 161–171, 2016.

[47] D. Windridge, H. Svensson, and S. Thill, "On the utility of dreaming: A general model for how learning in artificial agents can benefit from data hallucination," Adaptive Behavior, vol. 0, no. 0, p. 0, 0.

[48] A. Plebe, R. Donà, G. P. Papini Rosati, and M. Da Lio, "Mental Imagery for Intelligent Vehicles:," in Proceedings of the 5th International Conference on Vehicle Technology and Intelligent Transport Systems. Heraklion, Crete, Greece: SCITEPRESS - Science and Technology Publications, 2019, pp. 43–51.

[49] A. Plebe and M. Da Lio, "On the Road with 16 Neurons: Mental Imagery with Bio-inspired Deep Neural Networks," arXiv:2003.08745 [cs, stat], Mar. 2020, arXiv: 2003.08745.

[50] A. Plebe, G. P. R. Papini, R. Donà, and M. Da Lio, "Dreaming Mechanism for Training Bio-Inspired Driving Agents," in Intelligent Human Systems Integration 2019, W. Karwowski and T. Ahram, Eds. Cham: Springer International Publishing, 2019, vol. 903, pp. 429–434.

[51] "Explaining The Dreams4cars Autonomous Driving Project." [Online]. Available: https://www.youtube.com/watch?v=Bu5o8QzLETo

[52] M. Da Lio and R. Math, "Dreams4cars simulation tools and data," Jan. 2020. [Online]. Available: https://doi.org/10.5281/zenodo.3582054

IEEE *Access*

MAURO DA LIO received the Laurea degree in mechanical engineering from the University of Padova, Italy, in 1986. He is Full professor of mechanical systems with the University of Trento, Italy. His earlier research activity was on modelling, simulation and optimal control of mechanical multibody systems, in particular vehicle and spacecraft dynamics. More recently his focus shifted to the modelling of human sensory-motor control, in particular drivers and motor impaired people. Prior to his academic career, he worked for an offshore oil research company in underwater robotics (a EUREKA project). He was involved in several EU framework programme 6 and 7 projects (PReVENT, SAFERIDER, interactIVe, VERITAS, AdaptIVe, and No-Tremor). He is currently the coordinator of the EU Horizon 2020 Dreams4Cars Research and Innovation Action: a collaborative project in the Robotics domain which aims at increasing the cognition abilities of artificial driving agents by means of offline simulation mechanisms broadly inspired to the human dream state.

RICCARDO DONÀ is a PhD student working on the EU project Dreams4Cars at the University of Trento, Italy. He obtained the MSc in Mechatronics Engineering at the same university. His main scientific interest is the control of autonomous driving vehicles which constitutes the core topic of his PhD research project.

GASTONE PIETRO ROSATI PAPINI is a research fellow in advanced control systems applied to the field of autonomous driving and ADAS at Department of Industrial Engineering of the University of Trento and co-founder of Cheros s.r.l. a University of Pisa spinout company that deals with renewable energies and ICT solutions. Over 5 years experience in applied mechanics, in particular in the design of advanced control systems that also exploit machine learning techniques.

KEVIN GURNEY obtained a BSc in Mathematical Physics from the University of Sussex in 1977. He also has an MSc in Digital Systems and a PhD in Neural Networks (Brunel University). After his PhD, he held a postdoctoral position at Brunel in the Department of Human Sciences before moving to the University of Sheffield in 1995 where he is now Emeritus Professor of Computational Neuroscience. He has published in a wide variety of Journals including: Nature Reviews Neuroscience, IEEE Transactions on Neural Networks, and PLoS Biology. His previous research encompasses neural networks, vision, and the neuroscience of decision making and action selection.

• • •