An organismic inspired strategy for adaptive control

Alvaro Ovalle

Queen Mary University of London, London, UK
a.ovalle@qmul.ac.uk

Abstract

We focus on the connection between internal processes of regulation and their influence in the formation of behavioral patterns, and how they serve to provide a basis from where a biological organism derives a relation of value with its own activity. We touch on ongoing work exploring simulation modelling approaches to try to understand more about the issues required to overcome in order to capture the circularity involved in these processes, with particular attention to learning and sequential decision making formalisms of the perception-action loop, such as reinforcement learning and active inference. We describe an experimental setup where an agent learns to act in order to maintain the set of constraints that emerge and unfold as a consequence of the particular niche it inhabits.

Introduction

A striking feature exhibited by biological systems is their apparent intentional behavior. Organisms seem to act as to fulfill and obey drives in a way that confers their actions certain directedness, structure and significance. In contrast, if we compare to some recent and representative examples of reinforcement learning (RL) agents (Mnih et al., 2013, 2016; Pathak et al., 2017; Ha and Schmidhuber, 2018; Hafner et al., 2019) there is a general sense of detachment between them and their own behavior. For a successfully trained agent, for instance, its actions are intended to solve the task. But what is the task accomplishing for the agent? Unlike organisms, an RL agent is from the outset typically trained to achieve a specific competence. The agent acquires a behavior that is not a product of overcoming its constraints or satisfying its own needs or desires but those of the researcher. In this essay we briefly look at learning and sequential decision making frameworks through organismic lenses. That is, with the background that viability, continuity and survival provide for the emergence of primal behaviors (Weber and Varela, 2002; Di Paolo, 2003; Barandiaran et al., 2009; Man and Damasio, 2019). Thus the intention is not to focus on the merits these frameworks have for competence-based tasks in machine learning but to approach them as tools that can allow us a broader understanding of how to model autonomy and agent-centric intentionality.

Normativity, learning and decision making

The concept of normativity holds a prominent position in the study of autonomous behavior, due to its role in regulating action around a set of conditions from which to determine the level of success or failure of an activity pattern. In the standard formulation of RL, the extent to which the agent complies with those conditions depends on how adept the agent is at maximizing a reward. That is, the reward can be interpreted as a measure of preference between different choices, encouraging or discouraging behavioral policies. The situation, however, is not as straightforward as it may appear at a first glance. Rewards tend to be arbitrary without a clear criterion of how to specify them or what should be their magnitude. Since they are associated to specific events they are generally constrained to their domain or to a set of similar tasks. Moreover the situations that might trigger a particular reward could be different from those originally intended. In turn, for complex tasks rewards might have to be fine-tuned in order to obtain a desired behavior. But perhaps more problematically for our treatment, is that during this process of reward specification the agent remains disconnected from it. The norms or the goals are not grounded within the agent. It is not a process in which the agent participates to create or to internalize them, instead it presupposes the existence of a channel to which the agent is already attuned. Thus the process by which signals acquire a subjective significance that emerges from the particular historical contingencies is absent. A consequence that arises from this picture is that the internal dynamics of the agent are only in the service of the task because they lack a circular reference of their implication to the agent itself. It could be argued that the inclusion of *intrinsic motivations* [1] (Berlyne, 1960; Oudeyer and Kaplan, 2009) could in principle provide a mitigating component to some of these concerns. After all, in RL these are established along an epistemic dimension of the agent, for instance, computing state novelty (Bellemare et al., 2016; Burda et al., 2018), predic-

[1] Accounts of intrinsic motivation could be subjects of scrutiny as well considering that they accommodate an oracular worldview and operate under ambiguous notions of outcome separability.

tion error (Pathak et al., 2017) or uncertainty (Pathak et al., 2019). However their presence does not tend to be a product of a requirement associated to the structure of the agent but rather intended to alleviate deficiencies of the external reward, by providing support or by boosting the capacity to access certain regions of the state space in the service of the task. Thus we have a similar situation where goals do not emerge from following epistemic drives but rather the epistemic drives are there to assist in the discovery of those that have been exogenously imposed[2]. Closely related to probabilistic approaches in model-based RL, planning or control (Attias, 2003; Todorov, 2008; Kappen et al., 2012), active inference (AIF) offers a scheme that construes action and perception as complementary processes of inference (Friston, 2009). Conceptually it provides an alternative framework to model adaptive behavior that removes some of the ambiguity of the RL constructs by formulating the whole interaction in terms of beliefs. The basic premise is that the agent holds and updates a model of the world which provides a basis for its actions and predictions[3]. The agent tries to reduce the discrepancy between its predictions and sensory data either by changing its beliefs or by sampling regions of the space that might conform to those beliefs (Powers, 1973; Bell, 2007; Friston, 2009). A crucial aspect of this process in AIF is that the agent encodes its preferences in a target or reference distribution and thus it is possible to interpret the behavior *as if* it was trying to satisfy them. There are multiple ways to think about the role of the target distribution, one that is particularly relevant for the work presented here, is to conceive that it captures a series of constraints refined over evolutionary time-scales, and that if fulfilled, they may promote the continuity of an organism in an environment. However, as it has been noted in the literature (McGregor et al., 2015), AIF fails to provide satisfactory explanations for the normative conditions necessary for adaptive behavior as it does not directly address the specification of the target distribution.

A simple self-regulating agent

The previous sections have offered the context for our study in Ovalle and Lucas (2020), where we explore basic scenarios to try to understand how can an agent ground value in a context dependent and self-referential manner. We consider a simulated agent in the *Flappy Bird* environment (Lan, 2019) where it must navigate through narrow gaps between pipes. At each time step the agent can opt for flying upwards or to not perform any action in order to lower its position due to gravity. The agent *dies* if it goes above or below the edge of the frame or if it makes contact with the pipes. The agent measures its vertical position, velocity, the positions of the next two visible pipes and a quantity m that represents in abstract terms its state of life. Namely, whether it has registered an impact with the pipes or not. It is important to note that this and the other observations do not have an initial significance for the agent other than being available measurements. Although the dynamics offered by the environment are relatively simple, it provides a starting point to explore basic conditions in which an agent can learn self-preserving behavior. The agent cannot remain floating statically as it is subjected to gravity and the environment has a constant side-scrolling rate pushing it naturally towards the pipes. This permits us to establish some rudimentary comparisons to the conditions where natural organisms exist, as the agent is permanently challenged to maintain its continuity in an environment where the majority of the couplings lead to rapid non-existence. As we are not approaching the simulation from a task-competence perspective, the agent does not receive a reward at any point of the interaction. Instead we invoke the homeostatic interpretation of AIF which has also motivated other examples of agent-based simulations (McGregor et al., 2015; Baltieri and Buckley, 2019). From this perspective we can interpret that a system that exhibits homeostatic stability maintains low surprisal over a set of essential variables encoded by a reference distribution. For the simulations we do not assume previous knowledge or access to a prespecified reference distribution. Instead the agent dynamically defines and updates its own viability set from experience throughout its interaction with the environment. Concretely, the agent holds the sufficient statistics θ_t of its memory (i.e. queue) containing the last $t-k$ observations about its internal configuration m, which then uses to compute the surprisal generated by the new observation m_{t+1} during the next time step. We separately train two agents: a *model-free* agent that learns a surprisal minimizing policy by approximating state-action value functions through neural networks (NNs), and a second *model-based* agent that also learns via NNs the approximate transition function and predicted surprisal to simulate plans and act according to those it associates with the lowest *expected free energy* (i.e. surprisal and (negative) information gain) (Friston et al., 2015). The θ establishes a bridge between the conditions measured by the agent in the past that sustained its activity, and their involvement in providing a dynamic criteria that the agent uses to assess new (or predicted) measurements, as it continues to change on the basis of the conditions experienced by the agent. The self-reinforced sustaining behavior exhibited by the agents added new questions to the long list of open issues, such as the properties an environment has to possess in general to permit the presence of semi-stable configurations in the first place, how to establish more principled analogues of essential variables (Moerland et al., 2018) or the relations to recently proposed notions of semantic information (Kolchinsky and Wolpert, 2018).

[2]Conceived from biological considerations, *empowerment* (Klyubin et al., 2005) is a notable exception as it captures intuitions about agency and homeostatic behavior.

[3]We refer the reader to Buckley et al. (2017); Gottwald and Braun (2020) and Millidge et al. (2021) for thorough treatments.

References

Attias, H. (2003). Planning by Probabilistic Inference. In *Proc. of the 9th Int. Workshop on Artificial Intelligence and Statistics*.

Baltieri, M. and Buckley, C. L. (2019). The dark room problem in predictive processing and active inference, a legacy of cognitivism? In *ALIFE 2019: The 2019 Conference on Artificial Life*, pages 40–47. MIT Press.

Barandiaran, X. E., Di Paolo, E., and Rohde, M. (2009). Defining Agency: Individuality, Normativity, Asymmetry, and Spatio-temporality in Action. *Adaptive Behavior*, 17(5):367–386.

Bell, A. J. (2007). Towards a Cross-Level Theory of Neural Learning. *AIP Conference Proceedings*, 954(1):56–73.

Bellemare, M. G., Srinivasan, S., Ostrovski, G., Schaul, T., Saxton, D., and Munos, R. (2016). Unifying count-based exploration and intrinsic motivation. In *Proceedings of the 30th International Conference on Neural Information Processing Systems*, NIPS'16, pages 1479–1487, Red Hook, NY, USA. Curran Associates Inc.

Berlyne, D. E. (1960). *Conflict, Arousal, and Curiosity*. Conflict, Arousal, and Curiosity. McGraw-Hill Book Company, New York, NY, US.

Buckley, C. L., Kim, C. S., McGregor, S., and Seth, A. K. (2017). The free energy principle for action and perception: A mathematical review. *Journal of Mathematical Psychology*, 81:55–79.

Burda, Y., Edwards, H., Storkey, A., and Klimov, O. (2018). Exploration by random network distillation. In *International Conference on Learning Representations*.

Di Paolo, E. A. (2003). Organismically-inspired robotics : Homeostatic adaptation and teleology beyond the closed sensorimotor loop.

Friston, K. (2009). The free-energy principle: A rough guide to the brain? *Trends in Cognitive Sciences*, 13(7):293–301.

Friston, K., Rigoli, F., Ognibene, D., Mathys, C., Fitzgerald, T., and Pezzulo, G. (2015). Active inference and epistemic value. *Cognitive Neuroscience*, 6(4):187–214.

Gottwald, S. and Braun, D. A. (2020). The two kinds of free energy and the Bayesian revolution. *PLOS Computational Biology*, 16(12):e1008420.

Ha, D. and Schmidhuber, J. (2018). Recurrent World Models Facilitate Policy Evolution. In Bengio, S., Wallach, H., Larochelle, H., Grauman, K., Cesa-Bianchi, N., and Garnett, R., editors, *Advances in Neural Information Processing Systems 31*, pages 2450–2462. Curran Associates, Inc.

Hafner, D., Lillicrap, T., Fischer, I., Villegas, R., Ha, D., Lee, H., and Davidson, J. (2019). Learning Latent Dynamics for Planning from Pixels. In *International Conference on Machine Learning*, pages 2555–2565. PMLR.

Kappen, B., Gomez, V., and Opper, M. (2012). Optimal control as a graphical model inference problem. *Mach Learn*, 87(2):159–182.

Klyubin, A. S., Polani, D., and Nehaniv, C. L. (2005). Empowerment: A universal agent-centric measure of control. In *2005 IEEE Congress on Evolutionary Computation*, volume 1, pages 128–135 Vol.1.

Kolchinsky, A. and Wolpert, D. H. (2018). Semantic information, autonomous agency, and nonequilibrium statistical physics. *Interface Focus*, 8(6):20180041.

Lan, Q. (2019). Gym Compatible Games for Reinforcement Learning. *Github Repository*, https://github.com/qlan3/gym-games.

Man, K. and Damasio, A. (2019). Homeostatically Motivated Intelligence for Feeling Machines. In *AAAI Spring Symposium: Towards Conscious AI Systems*.

McGregor, S., Baltieri, M., and Buckley, C. L. (2015). A Minimal Active Inference Agent. *arXiv:1503.04187 [cs]*.

Millidge, B., Tschantz, A., and Buckley, C. L. (2021). Whence the Expected Free Energy? *Neural Computation*, 33(2):447–482.

Mnih, V., Badia, A. P., Mirza, M., Graves, A., Lillicrap, T., Harley, T., Silver, D., and Kavukcuoglu, K. (2016). Asynchronous Methods for Deep Reinforcement Learning. In *International Conference on Machine Learning*, pages 1928–1937. PMLR.

Mnih, V., Kavukcuoglu, K., Silver, D., Graves, A., Antonoglou, I., Wierstra, D., and Riedmiller, M. (2013). Playing Atari with Deep Reinforcement Learning. *arXiv:1312.5602 [cs]*.

Moerland, T. M., Broekens, J., and Jonker, C. M. (2018). Emotion in reinforcement learning agents and robots: A survey. *Mach Learn*, 107(2):443–480.

Oudeyer, P.-Y. and Kaplan, F. (2009). What is intrinsic motivation? A typology of computational approaches. *Front. Neurorobot.*, 1.

Ovalle, A. and Lucas, S. M. (2020). Modulation of Viability Signals for Self-regulatory Control. In Verbelen, T., Lanillos, P., Buckley, C. L., and De Boom, C., editors, *Active Inference*, Communications in Computer and Information Science, pages 101–113, Cham. Springer International Publishing.

Pathak, D., Agrawal, P., Efros, A. A., and Darrell, T. (2017). Curiosity-driven Exploration by Self-supervised Prediction. In *International Conference on Machine Learning*, pages 2778–2787. PMLR.

Pathak, D., Gandhi, D., and Gupta, A. (2019). Self-Supervised Exploration via Disagreement. In *International Conference on Machine Learning*, pages 5062–5071. PMLR.

Powers, W. T. (1973). *Behavior: The Control of Perception*. Behavior: The Control of Perception. Aldine, Oxford, England.

Todorov, E. (2008). General duality between optimal control and estimation. In *2008 47th IEEE Conference on Decision and Control*, pages 4286–4292.

Weber, A. and Varela, F. J. (2002). Life after Kant: Natural purposes and the autopoietic foundations of biological individuality. *Phenomenology and the Cognitive Sciences*, 1(2):97–125.