Free energy, empowerment, and predictive information compared

Martin Biehl[1], Christian Guckelsberger[2], Christoph Salge[3,4], Simón C. Smith[4], and Daniel Polani[4]

[1] *Araya Inc., Tokyo, Japan*
[2] *Goldsmiths, University of London, London, UK*
[3] *New York University, New York, USA*
[4] *University of Hertfordhire, Hatfield, UK*

Introduction The free energy principle/active inference (FEP, Friston et al. 2006, 2015), empowerment maximization (EM, Klyubin et al. 2005; Salge et al. 2014), and predictive information maximization (PIM, Ay et al. 2008, 2012; Martius et al. 2013) have all been proposed as information theoretical principles for driving autonomous behaviour. In this sense they are instances of intrinsic motivations (see Oudeyer and Kaplan, 2008; Schmidhuber, 2010; Barto et al., 2013; Santucci et al., 2013, for definition attempts and other approaches). In spite of the similar purpose of these principles, we are not aware of systematic comparisons between them. Since they are usually presented in differing formal frameworks, comparisons remain unnecessarily time-consuming and intransparent. Furthermore, simulation comparisons of the generated behaviours have not been undertaken either. Here we sketch our work on formulating the three principles within the same formalism. This will permit a direct and transparent comparison. A simulation analysis based on this unifying formalism is in work.

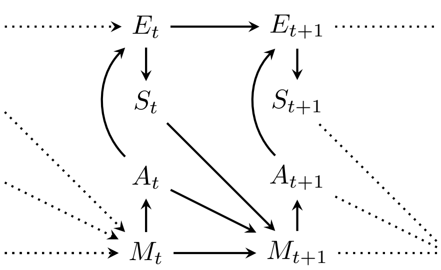

Figure 1: One PA-loop step.

Perception-action (PA-) loop We use the Bayesian network representation of the PA-loop as in Fig. 1. E_t, S_t, A_t, M_t are finite discrete random variables representing Environment, Sensor, Action, and agent Memory respectively. We assume an initial distribution over E_0, M_0 at time $t = 0$ and a final time T. For all times and all values of the involved random variables the environment and sensor dynamics are given via conditional probabilities $p(e_{t+1}|a_{t+1}, e_t)$ and $p(s_t|e_t)$. An agent is defined by specifying the values stored in M_t as well as their dynamics $p(m_{t+1}|s_t, a_t, m_t)$ and an action selection mechanism $p(a_t|m_t)$. We choose the same M_t and agent dynamics for all considered principles and only vary the action selection mechanisms. For clarity, we formally distinguish explicitly between transition probabilities constituting the PA-loop (denoted by the symbol p as above) and probabilistic models (denoted by symbols q, r) that are used to define and calculate the agents' dynamics and actions. Parameters of the agents' models are denoted using the Greek alphabet (e.g. Eq. (2)). We also indicate estimates, predictions, and "contemplated" values of random variables with a hat: an estimate of environment state e_t will be denoted $\hat{e}_t$ and a future action at $t + 1$ that is contemplated is denoted $\hat{a}_{t+1}$.

Sensor value prediction and agent dynamics Our setup so far corresponds to that of a partially observable Markov decision process (POMDP) without a given reward function. In place of the reward, the evaluation of actions is provided by the three principles. More precisely, all three can be formulated as evaluating predicted future consequences/sensor values of contemplated actions and then choosing an actual action accordingly. In order to allow a fair comparison, the predictions are generated by the same variational inference (VI) mechanism in each case. Note that due to differences in action choice, the sensor values each agent will encounter will differ and so will the quality of predictions. Assume a model $q(\hat{s}_{\succ t}|\hat{a}_{\succ t};\hat{\theta})$ of all future sensor value responses to future actions with parameters (possibly including latent variables, e.g. environment states $\hat{e}_t$) $\hat{\theta} \in \Delta_{\hat{\Theta}}$ and hyperprior $q(\hat{\theta};\hat{\alpha})$ with hyperparameters $\hat{\alpha} \in \Delta_{\hat{\alpha}}$. Then after a history $h_{\preceq t} = (s_{\preceq t}, a_{\preceq t})$ of all sensor values and actions until time t, full Bayesian inference (e.g. Bishop, 2011) gives us:

$$q(\hat{s}_{\succ t}|\hat{a}_{\succ t}, h_{\preceq t}) = \int q(\hat{s}_{\succ t}|\hat{a}_{\succ t};\hat{\theta})\; q(\hat{\theta}|h_{\preceq t};\hat{\alpha})\; q(\hat{\alpha}|h_{\preceq t})\; \mathrm{d}\hat{\theta}\; \mathrm{d}\hat{\alpha}. \tag{1}$$

Variational inference (e.g. Beal and Ghahramani, 2006) approximates the two history dependent terms by a "recognition" model $r(\hat{\theta};\phi_{t+1})$ where $\phi_{t+1} = \phi_{t+1}(h_{\preceq t})$. If the recognition model is chosen as a conjugate prior to $q(\hat{s}_{\succ t}|\hat{a}_{\succ t};\hat{\theta})$ the resulting integral

$$r(\hat{s}_{\succ t}|\hat{a}_{\succ t};\phi_t) := \int q(\hat{s}_{\succ t}|\hat{a}_{\succ t};\hat{\theta})\; r(\hat{\theta}|\phi_t)\; \mathrm{d}\hat{\theta} \quad (2)$$

becomes tractable (notationally the resulting model inherits the symbol r via "contamination"). In summary, we have $m_t = (h_{\preceq t-1}, \phi_t)$ and $p(m_{t+1}|s_t, a_t, m_t) = p(\phi_{t+1}|h_{\preceq t})p(h_{\preceq t}|a_t, s_t, h_{\preceq t-1})$ with the first factor obtained from variational inference and the second just by concatenation. Both factors are deterministic. We also choose a deterministic action selection with $p(a_t|m_t) = p(a_t|\phi_t) = \delta_{a_t^*(\phi_t)}(a_t)$ where $a_t^*(\phi_t)$ is defined for each principle below.

Free energy principle As mentioned in Friston et al. (2015, p.188), active inference results from "one straightforward imperative – to minimize surprise [...]". Free energy minimization itself is only invoked as a proxy for surprise minimization (see also Friston et al., 2012, Sec.3.1). We therefore formulate the FEP as the choice of actions that minimize expected surprise directly. For times t to $t+k$ surprise is defined as $-\log r(\hat{s}_t^{t+k}|\hat{a}_t^{t+k};\phi_t)$, so its expectation is an entropy. Since we only want to choose the next action, we can take into account longer action sequences by minimizing the expected surprise over distributions of subsequent actions:

$$a_t^*(\phi_t) = \arg\min_{\hat{a}_t} \min_{q(\hat{a}_{t+1}^{t+k})} \mathrm{H}(\hat{S}_t^{t+k}|\hat{A}_{t+1}^{t+k}, \hat{a}_t;\phi_t). \quad (3)$$

Empowerment maximization Empowerment is the channel capacity from a sequence of actions to the subsequent sensor value. Again we want to choose only the immediate action, and do this such that empowerment starting of the subsequent actions is maximized. Formally:

$$a_t^*(\phi_t) = \arg\max_{\hat{a}_t} \max_{q(\hat{a}_{t+1}^{t+k})} \mathrm{I}(\hat{A}_{t+1}^{t+k} : \hat{S}_{t+k}|\hat{a}_t;\phi_t). \quad (4)$$

Predictive information maximization Predictive information is the mutual information between past and future. Similar to the time-local predictive information of Martius et al. (2013), we choose the immediate action such that the predicted subsequent predictive information of the sensor values is maximized. For the subsequent actions we again choose a maximizing distribution (k now an even integer):

$$a_t^*(\phi_t) = \arg\max_{\hat{a}_t} \max_{q(\hat{a}_{t+1}^{t+k})} \mathrm{I}(\hat{S}_{t+1}^{t+k/2} : \hat{S}_{t+k/2+1}^{t+k}|\hat{a}_t;\phi_t). \quad (5)$$

Notes Variational inference involves the minimization of past surprise $-\log q(s_{\preceq t}|a_{\preceq t},\hat{\alpha})$ w.r.t. $\hat{\alpha}$ via the minimization of a "free energy". As action selection in the FEP usually also minimizes (expected) surprise via a free energy-like term, it is justifiable to package both together as a single "free energy principle". From a Bayesian perspective, however, the past surprise minimization (and in turn free energy minimization) is only an approximation and it would seem a stretch to elevate it to a principle. The reason for this elevation stems from the formal similarity to thermodynamic entropy and free energy (Friston et al., 2006).

In our formulation the three principles look strikingly similar. This is not only due to rigorous derivation but also to deliberate reformulation. In particular this concerns the use of optimization of $q(\hat{a}_{t+1}^{t+k})$ in the FEP and PIM which resembles the standard method for EM. We are investigating the assumptions necessary to justify these reformulations. Intuitively however, the expressions still capture the main ideas behind the principles.

It also remains to be seen whether and how any similarities will extend to the behaviour exhibited in simulations.

Bibliography

Ay, N., Bernigau, H., Der, R., and Prokopenko, M. (2012). Information-driven self-organization: the dynamical system approach to autonomous robot behavior. *Theory in Biosciences*, 131(3):161–179.

Ay, N., Bertschinger, N., Der, R., Güttler, F., and Olbrich, E. (2008). Predictive information and explorative behavior of autonomous robots. *The European Physical Journal B-Condensed Matter and Complex Systems*, 63(3):329–339.

Barto, A., Mirolli, M., and Baldassarre, G. (2013). Novelty or Surprise? *Frontiers in Psychology*, 4.

Beal, M. J. and Ghahramani, Z. (2006). Variational Bayesian learning of directed graphical models with hidden variables. *Bayesian Analysis*, 1(4):793–831.

Bishop, C. M. (2011). *Pattern Recognition and Machine Learning*. Springer, New York.

Friston, K., Kilner, J., and Harrison, L. (2006). A free energy principle for the brain. *Journal of Physiology-Paris*, 100(1):70 – 87.

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.

Friston, K., Samothrakis, S., and Montague, R. (2012). Active inference and agency: optimal control without cost functions. *Biological Cybernetics*, 106(8-9):523–541.

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

Martius, G., Der, R., and Ay, N. (2013). Information Driven Self-Organization of Complex Robotic Behaviors. *PLoS ONE*, 8(5):e63400.

Oudeyer, P.-Y. and Kaplan, F. (2008). How can we define intrinsic motivation? In *Proceedings of the 8th International Conference on Epigenetic Robotics: Modeling Cognitive Development in Robotic Systems, Lund University Cognitive Studies, Lund: LUCS, Brighton*. Lund University Cognitive Studies, Lund: LUCS, Brighton.

Salge, C., Glackin, C., and Polani, D. (2014). Changing the Environment Based on Empowerment as Intrinsic Motivation. *Entropy*, 16(5):2789–2819.

Santucci, V. G., Baldassarre, G., and Mirolli, M. (2013). Which is the best intrinsic motivation signal for learning multiple skills? *Frontiers in Neurorobotics*, 7.

Schmidhuber, J. (2010). Formal Theory of Creativity, Fun, and Intrinsic Motivation (1990 -2010). *IEEE Transactions on Autonomous Mental Development*, 2(3):230–247.