Neuroscience and Biobehavioral Reviews 68 (2016) 862–879

Contents lists available at ScienceDirect

Neuroscience and Biobehavioral Reviews

journal homepage: www.elsevier.com/locate/neubiorev

Active inference and learning

Karl Friston[a,*], Thomas FitzGerald[a,b], Francesco Rigoli[a], Philipp Schwartenbeck[a,b,c,d], John O'Doherty[e], Giovanni Pezzulo[f]

[a] *The Wellcome Trust Centre for Neuroimaging, UCL, 12 Queen Square, London, United Kingdom*
[b] *Max-Planck—UCL Centre for Computational Psychiatry and Ageing Research, London, United Kingdom*
[c] *Centre for Neurocognitive Research, University of Salzburg, Salzburg, Austria*
[d] *Neuroscience Institute, Christian-Doppler-Klinik, Paracelsus Medical University Salzburg, Salzburg, Austria*
[e] *Caltech Brain Imaging Center, California Institute of Technology, Pasadena, USA*
[f] *Institute of Cognitive Sciences and Technologies, National Research Council, Rome, Italy*

ARTICLE INFO

Article history:
Received 5 March 2016
Received in revised form 15 June 2016
Accepted 17 June 2016
Available online 29 June 2016

Keywords:
Active inference
Habit learning
Bayesian inference
Goal-directed
Free energy
Information gain
Bayesian surprise
Epistemic value
Exploration
Exploitation

ABSTRACT

This paper offers an active inference account of choice behaviour and learning. It focuses on the distinction between goal-directed and habitual behaviour and how they contextualise each other. We show that habits emerge naturally (and autodidactically) from *sequential policy* optimisation when agents are equipped with *state-action policies*. In active inference, behaviour has explorative (epistemic) and exploitative (pragmatic) aspects that are sensitive to ambiguity and risk respectively, where epistemic (ambiguity-resolving) behaviour enables pragmatic (reward-seeking) behaviour and the subsequent emergence of habits. Although goal-directed and habitual policies are usually associated with *model-based* and *model-free* schemes, we find the more important distinction is between *belief-free* and *belief-based* schemes. The underlying (variational) belief updating provides a comprehensive (if metaphorical) process theory for several phenomena, including the transfer of dopamine responses, reversal learning, habit formation and devaluation. Finally, we show that active inference reduces to a classical (Bellman) scheme, in the absence of ambiguity.

Contents

[*] Corresponding author at: The Wellcome Trust Centre for Neuroimaging, Institute of Neurology, 12 Queen Square, London WC1N 3BG, United Kingdom.
E-mail addresses: k.friston@ucl.ac.uk (K. Friston), thomas.fitzgerald@ucl.ac.uk (T. FitzGerald), f.rigoli@ucl.ac.uk (F. Rigoli), philipp.schwartenbeck.12@ucl.ac.uk (P. Schwartenbeck), jdoherty@hss.caltech.edu (J. O'Doherty), giovanni.pezzulo@istc.cnr.it (G. Pezzulo).

http://dx.doi.org/10.1016/j.neubiorev.2016.06.022

1. Introduction

There are many perspectives on the distinction between goal-directed and habitual behaviour (Balleine and Dickinson, 1998; Yin and Knowlton, 2006; Keramati et al., 2011; Dezfouli and Balleine, 2013; Dolan and Dayan, 2013; Pezzulo et al., 2013). One popular view rests upon model-based and model-free learning (Daw et al., 2005, 2011). In model-free approaches, the value of a state (e.g., being in a particular location) is learned through trial and error, while actions are chosen to maximise the value of the next state (e.g. being at a rewarded location). In contrast, model-based schemes compute a value-function of states under a model of behavioural contingencies (Gläscher et al., 2010). In this paper, we consider a related distinction; namely, the distinction between policies that rest upon beliefs about states and those that do not. In other words, we consider the distinction between choices that depend upon a (free energy) functional of beliefs about states, as opposed to a (value) function of states.

Selecting actions based upon the value of states only works when the states are known. In other words, a value function is only useful if there is no ambiguity about the states to which the value function is applied. Here, we consider the more general problem of behaving under ambiguity (Pearson et al., 2014). Ambiguity is characterized by an uncertain mapping between hidden states and outcomes (e.g., states that are partially observed) – and generally calls for policy selection or decisions under uncertainty; e.g. (Alagoz et al., 2010; Ravindran, 2013). In this setting, optimal behaviour depends upon beliefs about states, as opposed to states *per se*. This means that choices necessarily rest on inference, where optimal choices must first resolve ambiguity. We will see that this resolution, through epistemic behaviour, is an emergent property of (active) inference under prior preferences or goals. These preferences are simply outcomes that an agent or phenotype expects to encounter (Friston et al., 2015). So, can habits be learned in an ambiguous world? In this paper, we show that epistemic habits emerge naturally from observing the consequences of (one's own) goal-directed behaviour. This follows from the fact that ambiguity can be resolved, unambiguously, by epistemic actions.

To illustrate the distinction between belief-based and belief-free policies, consider the following examples: a predator (e.g., an owl) has to locate a prey (e.g., a field mouse). In this instance, the best goal-directed behaviour would be to move to a vantage point (e.g., overhead) to resolve ambiguity about the prey's location. The corresponding belief-free policy would be to fly straight to the prey, from any position, and consume it. Clearly, this belief-free approach will only work if the prey reveals its location unambiguously (and the owl knows exactly where it is). A similar example could be a predator waiting for the return of its prey to a waterhole. In this instance, the choice of whether to wait depends on the time elapsed since the prey last watered. The common aspect of these examples is that the belief state of the agent determines the optimal behaviour. In the first example, this involves soliciting cues from the environment that resolve ambiguity about the context (e.g., location of a prey). In the second, optimal behaviour depends upon beliefs about the past (i.e., memory). In both instances, a value-function of the states of the world cannot specify behaviour, because behaviour depends on beliefs or knowledge (i.e., *belief states* as opposed to states of the world).

Usually, in Markov decision processes (MDP), belief-based problems call for an augmented state-space that covers the belief or information states of an agent (Averbeck, 2015) – known as a belief MDP (Oliehoek et al., 2005). Although this is an elegant solution to optimising policies under uncertainty about (partially observed) states, the composition of belief states can become computationally intractable; not least because belief MDPs are defined over a continuous belief state-space (Cooper, 1988; Duff, 2002; Bonet and Geffner, 2014). Active inference offers a simpler approach by absorbing any value-function into a single functional of beliefs. This functional is variational free energy that scores the surprise or uncertainty associated with a belief, in light of observed (or expected) outcomes. This means that acting to minimise free energy resolves ambiguity and realises unsurprising or preferred outcomes. We will see that this single objective function can be unpacked in a number of ways that fit comfortably with established formulations of optimal choice behaviour and foraging.

In summary, schemes that optimise state-action mappings – via a value-function of states – could be considered as habitual, whereas goal-directed behaviour is quintessentially belief-based. This begs the question as to whether habits can emerge under belief-based schemes like active inference. In other words, can habits be learned by simply observing one's own goal-directed behaviour? We show this is the case; moreover, habit formation is an inevitable consequence of equipping agents with the hypothesis that habits are sufficient to attain goals. We illustrate these points, using formal (information theoretic) arguments and simulations. These simulations are based upon a generic (variational) belief update scheme that shows several behaviours reminiscent of real neuronal and behavioural responses. We highlight some of these behaviours in an effort to establish the construct validity of active inference.

This paper comprises four sections. The first provides a description of active inference, which combines our earlier formulations of planning as inference (Friston et al., 2014) with Bayesian model averaging (FitzGerald et al., 2014) and learning (FitzGerald et al., 2015a, 2015b). Importantly, action (i.e. policy selection), perception (i.e., state estimation) and learning (i.e., reinforcement learning) all minimise the same quantity; namely, variational free energy. In this formulation, habits are learned under the assumption (or hypothesis) there is an optimal mapping from one state to the next, that is not context or time-sensitive.[1] Our key interest was to see if habit-learning emerges as a Bayes-optimal *habitisation* of goal-directed behaviour, when circumstances permit. This follows a general line of thinking, where habits are effectively learned as the invariant aspects of goal-directed behaviour (Dezfouli and Balleine, 2013; Pezzulo et al., 2013, 2014, 2015). It also speaks to the *arbitration* between goal-directed and habitual policies (Lee et al., 2014). The second section considers variational belief updating from the perspective of standard approaches to policy optimisation based on the Bellman optimality principle. In brief, we will look at dynamic programming schemes for Markovian decision processes that are cast in terms of value-functions – and how the ensuing value (or policy) iteration schemes can be understood in terms of active inference.

[1] Here, we mean context insensitive in the sense of Thrailkill and Bouton (2015). In other words, context refers to outcome contingencies; not the paradigmatic context.