# Value-based decision making for human-machine robot control

Paul B. Reverdy

## I. Introduction

Humans negotiate and manipulate their environments with a facility that far surpasses the abilities of even the best current robots. This facility is likely the fruit of a complex set of systems that allow humans to process noisy stimuli about the world, form internal representations of the patterns contained in these stimuli, and use the representations to make decisions that are expressed through motor actions.

This perception-action loop is analogous to the one used to design robots, although the physical (i.e., biological) hardware used by humans is different in kind to that used by robots. The biological hardware has significantly less computational power than that of robots, so human capabilities are likely underpinned by approximations that enable high performance at low computational cost. Understanding these approximations is of great use to robotics, and this paper argues that the framework of value-based decision making is a fruitful one to pursue this understanding.

## II. Value-based decision making is optimization

Value-based decision making [1] is a paradigm for studying decision-making behavior. In the paradigm, subjects (who may be animal, human, or algorithmic) are given tasks where they are shown stimuli and have to react by performing certain actions, whereupon they receive rewards. The utility of this paradigm is due to the fact that the subject's optimal strategy can be written in terms of an optimization problem, namely, that of maximizing total rewards or reward rate. With such an optimal (often termed "normative") strategy defined, cognitive scientists can then empirically study deviations from optimal behavior and explain them in terms of heuristics and biases [2].

The properties of the value-based decision-making paradigm that make it useful for studying behavior also make it useful for robotics. It is natural to formulate robotics and control tasks as optimization problems; this is the basis of optimal control. Solving optimal control problems can be computationally costly, so understanding the heuristics and biases that permit biological systems to cheaply find approximate solutions is of great value to robotics. Conversely, when these biases create predictable patterns of suboptimal behavior, it is of interest to design computational aids to help humans improve their performance.

The author is with the Department of Aerospace and Mechanical Engineering, University of Arizona, Tucson, AZ 85721 USA. preverdy@email.arizona.edu. This work has been supported in part by startup funds provided by the University of Arizona, as well as grant number FA8650-15-D-1845 subcontract 669737-6 from the Air Force Research Laboratory.

A generic value-based decision-making task has the following structure: a subject is presented with a set of possible actions $\mathcal{A}$ which may be finite or infinite. At each of a sequence of times $t \in \mathbb{N}$, the subject is presented with a stimulus $s_t$ which may be visual, auditory, tactile, etc. The subject then selects an action $a_t \in \mathcal{A}$ and receives a reward $r_t \in \mathbb{R}$. Both the stimulus and the reward may be stochastic in the sense that they are corrupted by random noise. The subject's goal is to pick actions that maximize expected rewards:

$$\max_{\{a_t | \mathcal{F}_t\}_{t\in\mathbb{N}}} \sum_{t\in\mathbb{N}} \mathbb{E}\left[r_t\right], \tag{1}$$

where $\mathcal{F}_t$ represents the information available to the subject at time $t$. This paradigm is analogous to a Markov Decision Process (MDP) in obvious ways.

The normatively-optimal strategy to solve (1) can, in principle, be computed by exploiting the analogy to MDPs and employing an appropriate MDP solution algorithm, such as value iteration, Q-learning, etc. This provides a baseline for comparison with observed human behavior.

In solving control problems, one often employs the separation principle and designs a controller based on 1) a state-feedback control law, and 2) an estimator which produces estimates of the state that can be fed into the control law. Behavioral scientists build models that separate similarly. This introduces structure in two places: 1) heuristics, analogous to control laws, and 2) Bayesian statistics, which are used to build representations and estimators.

The structures of heuristics and priors are representations of task-relevant information that can be shared between humans and robots. By understanding the structures of heuristics that guide human decision making and the priors that inform these heuristics, we can develop algorithms that naturally interface between humans and machines. For example, a robotic system could learn from a human with high performance by fitting a model to the human's observed behavior and then using those model parameters in an algorithm with similar structure.

## III. Motor action selection

Several prototypical tasks from cognitive science are particularly relevant to our goal of using value-based decision making in robotics. In this section we introduce focus on one, called *motor action selection*, and in the following section we show how we have begun to connect it to robotics.

Situations where evidence must continuously be integrated and used to select among alternatives are at the heart of the so-called affordance competition hypothesis. An affordance

is an opportunity for action defined by the environment around an animal or robot. The affordance competition hypothesis states that, in animals, the processes of action selection and movement planning operate simultaneously and in an integrated manner [3].

The cognitive science literature provides a body of evidence for the affordance competition hypothesis [3], [4]. Often this evidence takes the form of brain structures which interweave responsibility for perception and motion planning and execution. In the context of robotics, an analogous sensorimotor system would similarly integrate processing of perceptual stimuli and motion planning. In the next section I present the details of such a sensorimotor system, called *motivation dynamics*, which I am actively developing.

## IV. Motivation dynamics

In ongoing work, I am pursuing a framework for value-based sensorimotor systems that implements a form of affordance competition and naturally interfaces with low-level signal processing models of the type used to study perceptual decision making. I call this framework *motivation dynamics*.

The motivation dynamics framework, introduced in [5], can be thought of as a convex relaxation of hybrid dynamical systems in the following sense. Like a hybrid dynamical system, a motivation dynamics system with continuous state $x$ has a finite set of low-level controllers $F_a(x), a \in \mathcal{A} = \{1, \ldots, N\}$ called *modes*. A hybrid dynamical system follows one mode $a_t$ at time $t$, where $a$ is the state variable of a discrete finite automaton. The continuous dynamics are then $\dot{x} = F_{a_t}(x)$, and various guard functions control the transitions of the automaton. Instead of the mode variable $a \in \mathcal{A}$, the motivation dynamics system maintains a *motivation* state $m \in \Delta^N = \{x \in \mathbb{R}^{N+1} | x_i \geq 0, \sum_i x_i = 1\}$, where $\Delta^N$ is the $N$-simplex. The vertices of $\Delta^N$ are analogous to the modes $a \in \mathcal{A}$, while other elements of $\Delta^N$ consist of convex combinations of the vertices. The dynamics of $x$ under motivation dynamics are given by $\dot{x} = \sum_{i=1}^{N} m_i F_i(x)$, which can be thought of a convex relaxation of the continuous dynamics of the hybrid dynamical system.

In place of the discrete finite automaton that selects modes in a hybrid system, motivation dynamics framework [5] uses a bio-inspired dynamical system from [6] which implements a value-based decision-making model. Specifically, the motivation dynamics framework associates a value state $v_i > 0$ to each mode $i \in \mathcal{A}$ ($v_i$ may have its own dynamics and depend on the environment or external stimuli) and the motivation dynamics $\dot{m} = f_m(m, v)$ is such that the motivation state $m$ will tend towards a point that puts most weight on the highest-value mode. By tightly coupling valuation, action (i.e., mode) selection, and physical dynamics, the motivation dynamics framework implements a form of affordance competition as discussed in [3].

The value state $v \in \mathbb{R}^N_+$ in the motivation dynamics provides a natural interface between the motivation dynamics framework and low-level perceptual decision models. For example, the likelihood or log-likelihood of a given category of stimulus can be used as the value input associated with the mode that should be triggered when that stimulus is detected. Such a connection was suggested in [7], which studied the dynamics $\dot{m} = f_m(m, v)$ in their original biological context. In recent work [8], we show that using log-likelihoods as value states permits a robot to smoothly select correct actions in response to noisy stimuli.

## References


[1] A. Rangel, C. Camerer, and P. R. Montague, "A framework for studying the neurobiology of value-based decision making," *Nature Reviews Neuroscience*, vol. 9, no. 7, pp. 545–556, Jun 2008. [Online]. Available: http://dx.doi.org/10.1038/nrn2357

[2] A. Tversky and D. Kahneman, "Judgment under uncertainty: Heuristics and biases," *science*, vol. 185, no. 4157, pp. 1124–1131, 1974.

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

[4] 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.

[5] P. B. Reverdy and D. E. Koditschek, "A dynamical system for prioritizing and coordinating motivations," *SIAM Journal on Applied Dynamical Systems*, vol. 17, no. 2, pp. 1683–1715, 2018.

[6] T. D. Seeley, P. K. Visscher, T. Schlegel, P. M. Hogan, N. R. Franks, and J. A. Marshall, "Stop signals provide cross inhibition in collective decision-making by honeybee swarms," *Science*, vol. 335, no. 6064, pp. 108–111, 2012.

[7] D. Pais, P. M. Hogan, T. Schlegel, N. R. Franks, N. E. Leonard, and J. A. Marshall, "A mechanism for value-sensitive decision-making," *PloS one*, vol. 8, no. 9, p. e73216, 2013.

[8] P. B. Reverdy, V. Vasilopoulos, and D. E. Koditschek, "Motivation dynamics for autonomous composition of navigation tasks," in *In preparation*, 2019.

[9] H. Kress-Gazit, G. E. Fainekos, and G. J. Pappas, "Temporal-logic-based reactive mission and motion planning," *IEEE Transactions on Robotics*, vol. 25, no. 6, pp. 1370–1381, 2009.

[10] J. Liu, N. Ozay, U. Topcu, and R. M. Murray, "Synthesis of reactive switching protocols from temporal logic specifications," *IEEE Transactions on Automatic Control*, vol. 58, no. 7, pp. 1771–1785, 2013.

[11] R. L. Keeney and H. Raiffa, *Decisions with multiple objectives: preferences and value trade-offs*. Cambridge University Press, 1993.