LETTER **Communicated by Karl Friston**

Systems of Bounded Rational Agents with Information-Theoretic Constraints

Sebastian Gottwald
sebastian.gottwald@uni-ulm.de
Daniel A. Braun
daniel.braun@uni-ulm.de
Institute of Neural Information Processing, Faculty of Engineering, Computer Science and Psychology, University of Ulm, Ulm, Baden-Württemberg, 89081 Germany

Specialization and hierarchical organization are important features of efficient collaboration in economical, artificial, and biological systems. Here, we investigate the hypothesis that both features can be explained by the fact that each entity of such a system is limited in a certain way. We propose an information-theoretic approach based on a free energy principle in order to computationally analyze systems of bounded rational agents that deal with such limitations optimally. We find that specialization allows a focus on fewer tasks, thus leading to a more efficient execution, but in turn, it requires coordination in hierarchical structures of specialized experts and coordinating units. Our results suggest that hierarchical architectures of specialized units at lower levels that are coordinated by units at higher levels are optimal, given that each unit's information-processing capability is limited and conforms to constraints on complexity costs.

1 Introduction

The question of how to combine a given set of individual entities in order to perform a certain task efficiently is a long-lasting question shared by many disciplines, including economics, neuroscience, and computer science. Although the explicit nature of a single individuum might differ between these fields—for example, an employee of a company, a neuron in a human brain, or a computer or processor as part of a cluster—they have one important feature in common that usually prevents them from functioning isolated by themselves: they are all limited. In fact, this was the driving idea that inspired Herbert A. Simon's early work on decision making within economic organizations (Simon, 1943, 1955), which earned him a Nobel prize in 1978. He suggested that a scientific behavioral grounding of economics should be based on bounded rationality, which has remained an active research topic (Russell & Subramanian, 1995; Lipman, 1995; Aumann,

Neural Computation **31, 440–476** (2019)
doi:10.1162/neco_a_01153

1997; Kaelbling, Littman, & Cassandra, 1998; DeCanio and Watkins, 1998; Gigerenzer & Selten, 2001; Jones, 2003; Sims, 2003; Burns, Ruml, & Do, 2013; Ortega & Braun, 2013; Acerbi, Vijayakumar, & Wolpert, 2014; Gershman, Horvitz, & Tenenbaum, 2015). Subsequent studies in management theory have been built on Simon's basic observation, because "if individual managers had unlimited access to information that they could process costlessly and instantaneously, there would be no role for organizations employing multiple managers" (Geanakoplos & Milgrom, 1991). In neuroscience and biology, similar concepts have been used to explore the evolution of specialization and modularity in nature (Kashtan & Alon, 2005; Wagner, Pavlicev, & Cheverud, 2007). In modern computer science, the terms *parallel computing* and *distributed computing* denote two separate fields that share the concept of decentralized computing (Radner, 1993)—the combination of multiple processing units in order to decrease the time of computationally expensive calculations.

Despite their success, there are also shortcomings of most approaches to the organization of decision-making units based on bounded rationality. As DeCanio and Watkins (1998) point out, existing agent-based methods (including their own) are not using an overreaching optimization principle but are tailored to the specific types of calculations the agents are capable of, and therefore lack in generality. Moreover, it is usually imposed as a separate assumption that there are two types of units, specialized operational units and coordinating nonoperational units, which was expressed by (Knight, 1921) as "workers do, and managers figure out what to do."

Here, we use a free energy optimization principle in order to study systems of bounded rational agents, extending the work in Ortega and Braun (2011, 2013), Genewein and Braun (2013) and Genewein, Leibfried, Grau-Moya, and Braun (2015) on decision making, hierarchical information processing, and abstraction in intelligent systems with limited information-processing capacity, that has precursors in the economic and game-theoretic literature (McKelvey & Palfrey, 1995; Ochs, 1995; Mattsson & Weibull, 2002; Wolpert, 2006; Spiegler, 2011; Howes, Lewis, & Vera, 2009; Todorov, 2009; Still, 2009; Tishby & Polani, 2011; Kappen, Gómez, & Opper, 2012; Vul, Goodman, Griffiths, & Tenenbaum, 2014; Lewis, Howes, & Singh, 2014). Note that the free energy optimization principle of information-theoretic bounded rationality is connected to the free energy principle used in variational Bayes and active inference (Friston, Levin, Sengupta, & Pezzulo, 2015; Friston, Rigoli et al., 2015; Friston, Lin, Frith, & Pezzulo, 2017; Friston, Parr, & de Vries, 2017), but has a conceptually distinct interpretation and some formal differences (see section 6.3 for a detailed comparison).

By generalizing the ideas in Genewein and Braun (2013) and Genewein et al. (2015) on two-step information processing to an arbitrary number of steps, we arrive at a general free energy principle that can be used to study systems of bounded rational agents. The advantages of our approach can be summarized as follows:

1. There is a unifying free energy principle that allows for a multiscale problem formulation for an arbitrary number of agents distributed among the steps of general multistep processes (see sections 3.3 and 4.2).
2. The computational nature of the optimization principle allows explicitly calculating and comparing optimal performances of different agent architectures for a given set of objectives and resource constraints (see section 5).
3. The information-theoretic description implies the existence of the two types of units already mentioned, nonoperational units (selector nodes) that coordinate the activities of operational units. Depending on their individual resource constraints, the free energy principle assigns each unit to a region of specialization that is part of an optimal partitioning of the underlying decision space (see section 4.3).

In particular, we find that for a wide range of objectives and resource limitations (see sections 5.3 and 5.3), hierarchical systems with specialized experts at lower levels and coordinating units at higher levels generally outperform other structures.

2 Preliminaries

This section serves as an introduction to the terminology required for our framework presented in sections 3 and 4.

2.1 Notation. We use curly letters (e.g., $\mathcal{W}$, $\mathcal{X}$, $\mathcal{A}$) to denote sets of finite cardinality—in particular, the underlying spaces of the corresponding random variables (e.g., W, A, X)—whereas the values of these random variables are denoted by lowercase letters: $w \in \mathcal{W}$, $a \in \mathcal{A}$, and $x \in \mathcal{X}$, respectively. We denote the space of probability distributions on a given set $\mathcal{X}$ by $\mathbb{P}_{\mathcal{X}}$. Given a probability distribution $p \in \mathbb{P}_{\mathcal{X}}$, the expectation of a function $f : \mathcal{X} \to \mathbb{R}$ is denoted by $\langle f \rangle_p := \sum_x p(x) f(x)$. If the underlying probability measure is clear without ambiguity, we just write $\langle f \rangle$.

For a function g with multiple arguments (e.g., for $g : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$, $(x, y) \mapsto g(x, y)$), we denote the function $\mathcal{X} \to \mathbb{R}$, $x \mapsto g(x, y)$ for fixed $y \in \mathcal{Y}$ by $g(\,\cdot\,, y)$ (partial application), that is, the dot indicates the variable of the new function. Similarly, for fixed $y \in \mathcal{Y}$, we denote a conditional probability distribution on $\mathcal{X}$ with values $p(x|y)$ by $p(\,\cdot\,|y)$. This notation shows the dependencies clearly without giving up the original function names and thus allows writing more complicated expressions in a concise form. For example, if F is a functional defined on functions of one variable, such as $F[f] := \sum_x f(x)$ for all functions $f : \mathcal{X} \to \mathbb{R}$, then evaluating F on the function g in its first variable while keeping the second variable fixed, is simply denoted by $F[g(\,\cdot\,, y)]$. Here, the dot indicates on which argument of g the functional F is acting, and at the same time it records that the resulting value

(which equals $\sum_x g(x, y)$ in the case of the example) does not depend on a particular x but on the fixed y.

2.2 Decision Making. Here, we consider (multitask) decision making as the process of observing a world state $w \in \mathcal{W}$, sampled from a given distribution $\rho \in \mathbb{P}_{\mathcal{W}}$, and choosing a corresponding action $a \in \mathcal{A}$ drawn from a posterior policy $P(\cdot\,|w) \in \mathbb{P}_{\mathcal{A}}$. Assuming that the joint distribution of W and A is given by $p(a, w) := \rho(w)P(a|w)$, then P is the conditional probability distribution of A given W. Unless stated otherwise, the capital letter P always denotes a posterior, while the lowercase letter p denotes the joint distribution or a marginal of the joint (i.e., a dependent variable).

A decision-making unit is called an *agent*. An agent is rational if its posterior policy P maximizes the expected utility,

$$\langle U \rangle = \sum_{w \in \mathcal{W}} \rho(w) \sum_{a \in \mathcal{A}} P(a|w)\, U(a, w), \tag{2.1}$$

for a given utility function $U : \mathcal{W} \times \mathcal{A} \to \mathbb{R}$. Note that the utility U may itself represent an expected utility over consequences in the sense of von Neumann and Morgenstern (1944), where W would serve as a context variable for different tasks. The posterior P can be seen as a state-action policy that selects the best action $a \in \mathcal{A}$ with respect to a utility function U given the state $w \in \mathcal{W}$ of the world.

2.3 Bounded Rational Agents. In the information-theoretic model of bounded rationality (Ortega & Braun, 2011, 2013; Genewein et al., 2015), an agent is bounded rational if its posterior P maximizes equation 2.1, subject to the constraint

$$\left\langle \mathrm{D}_{\mathrm{KL}}(P\|q) \right\rangle = \sum_{w \in \mathcal{W}} \rho(w)\, \mathrm{D}_{\mathrm{KL}}(P(\cdot\,|w)\|q) \leqslant D_0, \tag{2.2}$$

for a given bound $D_0 > 0$ and a prior policy $q \in \mathbb{P}_{\mathcal{A}}$. Here, $\mathrm{D}_{\mathrm{KL}}(p\|q)$ denotes the Kullback-Leibler (KL) divergence between two distributions $p, q \in \mathbb{P}_{\mathcal{Y}}$ on a set $\mathcal{Y}$, defined by $\mathrm{D}_{\mathrm{KL}}(p\|q) := \sum_{y \in \mathcal{Y}} p(y) \log(p(y)/q(y))$. Note that for $\mathrm{D}_{\mathrm{KL}}(p\|q)$ to be well defined, p must be absolutely continuous with respect to q, so that $q(y) = 0$ implies $p(y) = 0$. When p or q are conditional probabilities, we treat $\mathrm{D}_{\mathrm{KL}}(p\|q)$ as a function of the additional variables.

Given a world state w, the information processing consists of transforming a prior q to a world-state specific posterior distribution $P(\cdot\,|w)$. Since $\mathrm{D}_{\mathrm{KL}}(P(\cdot\,|w)\|q)$ measures by how much $P(\cdot\,|w)$ diverges from q, the upper bound D_0 in equation 2.2 characterizes the limitation of the agent's average information-processing capability: If D_0 is close to zero, the posterior must be close to the prior for all world states, which means that A contains

only little information about W, whereas if D_0 is large, the posterior is allowed to deviate from the prior by larger amounts and therefore A contains more information about W. We use the KL divergence as a proxy for any resource measure, as any resource must be monotone in processed information, which is measured by the KL divergence between prior and posterior.

Technically, maximizing expected utility under constraint 2.2 is the same as minimizing expected complexity cost under the constraint of a minimal expected performance, where complexity is given by the expected KL divergence between prior and posterior and performance by expected utility. Minimizing complexity means minimizing the number of bits required to generate the actions.

2.4 Free Energy Principle. By the variational method of Lagrange multipliers, the above constrained optimization problem is equivalent to the unconstrained problem,

$$\max_P \left(\langle U \rangle - \frac{1}{\beta} \left\langle \mathrm{D}_{\mathrm{KL}}(P \| q) \right\rangle \right), \tag{2.3}$$

where $\beta > 0$ is chosen such that the constraint 2.2 is satisfied. In the literature on information-theoretic bounded rationality (Ortega & Braun, 2011, 2013), the objective in equation 2.3 is known as the *free energy* $\mathcal{F}$ of the corresponding decision-making process. In this form, the optimal posterior can be explicitly derived by determining the zeros of the functional derivative of $\mathcal{F}$ with respect to P, yielding the Boltzmann-Gibbs distribution,

$$P(a|w) = \frac{1}{Z(w)} q(a)\, e^{\beta U(a,w)}, \quad Z(w) := \sum_{a \in \mathcal{A}} q(a)\, e^{\beta U(a,w)}. \tag{2.4}$$

Note how the Lagrange multiplier β (also known as inverse temperature) interpolates between an agent with zero processing capability that always acts according to its prior policy ($\beta = 0$) and a perfectly rational agent ($\beta \to \infty$). Note that plugging equation 2.4 back into the free energy equation 2.3, gives

$$\max_P \mathcal{F}[P] = \frac{1}{\beta} \langle \log Z \rangle. \tag{2.5}$$

2.5 Optimal Prior. The performance of a given bounded rational agent crucially depends on the choice of the prior policy q. Depending on D_0 and the explicit form of the utility function, it can be advantageous to a priori prefer certain actions over others. Therefore, optimal bounded rational decision making includes optimizing the prior in equation 2.3. In contrast to equation 2.3, the modified optimization problem,

$$\max_{P,q} \left(\langle U \rangle - \frac{1}{\beta} \langle \mathrm{D}_{\mathrm{KL}}(P \| q) \rangle \right), \tag{2.6}$$

does not have a closed-form solution. However, since the objective is convex in (P, q), a unique solution can be obtained iteratively by alternating between fixing one and optimizing the other variable (Csiszár & Tusnády, 1984), resulting in a Blahut-Arimoto-type algorithm (Arimoto, 1972; Blahut, 1972) that consists of alternating the equations

$$\begin{cases} P(a|w) = \dfrac{1}{Z(w)}\, q(a)\, e^{\beta U(a,w)}, \\ q(a) = p(a) = \displaystyle\sum_{w} \rho(w) P(a|w), \end{cases} \tag{2.7}$$

with $Z(w)$ given by equation 2.4. In particular, the optimal prior policy is the marginal p of the joint distribution of W and A. In this case, the average Kullback-Leibler divergence between prior and posterior coincides with the mutual information between W and A,

$$I(W; A) = \sum_{w \in \mathcal{W}} \sum_{a \in \mathcal{A}} p(w, a) \log \frac{p(w, a)}{\rho(w) p(a)} = \langle \mathrm{D}_{\mathrm{KL}}(P, p) \rangle.$$

It follows that the modified optimization principle, equation 2.6, is equivalent to

$$\max_{P} \left(\langle U \rangle - \frac{1}{\beta} I(W; A) \right). \tag{2.8}$$

Due to its equivalence to rate distortion theory (Shannon, 1959) (with a negative distortion measure given by the utility function), equation 2.8 is denoted as the rate distortion case of bounded rationality in Genewein and Braun (2013).

2.6 Multistep and Multiagent Systems. When multiple random variables are involved in a decision-making process, such a process constitutes a multistep system (see section 3). Consider the case of a prior over $\mathcal{A}$ that is conditioned on an additional random variable X with values $x \in \mathcal{X}$, that is, $q(\,\cdot\,|x) \in \mathbb{P}_{\mathcal{A}}$ for all $x \in \mathcal{X}$. Remember that we introduced a bounded rational agent as a decision-making unit that, after observing a world state w, transforms a single prior policy over a choice space $\mathcal{A}$ to a posterior policy $P(\,\cdot\,|w) \in \mathbb{P}_{\mathcal{A}}$. Therefore, in the case of a conditional prior, the collection of prior policies $\{q(\,\cdot\,|x)\}_{x \in \mathcal{X}}$ can be considered as a collection or ensemble of agents, or a multiagent system, where for a given $x \in \mathcal{X}$, the prior $q(\,\cdot\,|x)$ is transformed to a posterior $P(\,\cdot\,|x, w) \in \mathbb{P}_{\mathcal{A}}$ by exactly one agent. Note that a

single agent deciding about both, X and A, would be modeled by a prior of the form $q(x, a)$ with $x \in \mathcal{X}$ and $a \in \mathcal{A}$, instead.

Hence, in order to combine multiple bounded rational agents, we are first splitting the full decision-making process into multiple steps by introducing additional intermediate random variables (see section 3), which then will be used to assign one or more agents to each of these steps (see section 4). In this view, we can regard a multiagent decision-making system as performing a sequence of successive decision steps until an ultimate action is selected.

3 Multistep Bounded Rational Decision Making

3.1 Decision Nodes. Let W and A denote the random variables describing the full decision-making process for a given utility function $U : \mathcal{W} \times \mathcal{A} \to \mathbb{R}$, as described in section 2. In order to separate the full process into $N > 1$ steps, we introduce internal random variables $X_1, \ldots, X_{N-1}$, which represent the outputs of additional intermediate bounded rational decision-making steps. For each k, let $\mathcal{X}_k$ denote the target space and $x_k \in \mathcal{X}_k$ a particular value of X_k. We call a random variable that is part of a multistep decision-making system a *(decision) node*. For simplicity, we assume that all intermediate random variables are discrete (just like W and A).

Here, we are treating feedforward architectures originating at $X_0 := W$ and terminating in $X_N := A$. This allows labeling the variables $\{X_k\}_{k=0}^N$ according to the information flow, so that X_j potentially can only obtain information about X_i if $i < j$. The canonical factorization,

$$p(w, x_1, \ldots, x_{N-1}, a) = \rho(w)\, p(x_1|w)\, p(x_2|x_1, w) \cdots p(a|x_{N-1}, \ldots, x_1, w),$$

of the joint probability distribution of $\{X_k\}_{k=0}^N$ therefore consists of the posterior policies of each decision node.

3.2 Two Types of Nodes: Inputs and Prior Selectors. A specific multistep architecture is characterized by specifying the explicit dependencies on the preceding variables for each node's prior and posterior or, better, the missing dependencies. For example, in a given multistep system, the posterior of the node X_3 might depend explicitly on the outputs of X_1 and X_2 but not on W, so that $P(x_3|x_2, x_1, w) = P(x_3|x_2, x_1)$. If its prior has the form $q(x_3|x_1)$, then X_3 has to process the output of X_2. Moreover, in this case, the actual prior policy $q(\,\cdot\,|x_1) \in \mathbb{P}_{\mathcal{X}_3}$ that is used by X_3 for decision making is selected by X_1 (see Figure 1).

In general, the inputs $X_i, \ldots, X_j$ that have to be processed by a particular node X_k are given by the variables in the posterior that are missing from the prior, and if its prior q is conditioned on the outputs of $X_l, \ldots, X_m$, then these

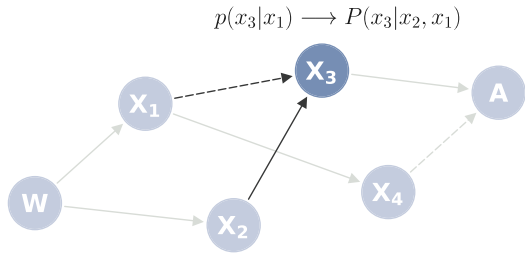

Figure 1: Example of a processing node that is part of a multistep architecture with $N = 5$, visualized as a directed graph. Here, X_3 processes the output of X_2 by transforming a prior policy $p(x_3|x_1)$ to a posterior policy $P(x_3|x_2, x_1)$. The prior of X_3 being conditioned on the output of X_1 (indicated by the dashed arrow) means that X_1 determines which of the prior policies $\{p(\cdot\,|x_1)\}_{x_1 \in \mathcal{X}_1}$ is used by X_3 to process a given output of X_2.

nodes select which of the prior policies $\{q(\cdot\,|x_l, \ldots, x_m)\}_{x_l \in \mathcal{X}_l, \ldots, x_m \in \mathcal{X}_m} \subset \mathbb{P}_{\mathcal{X}_k}$ is used by X_k for decision making, that is, for the transformation

$$q(x_k|x_l, \ldots, x_m) \longrightarrow P(x_k|x_l, \ldots, x_m, x_i, \ldots, x_j).$$

We denote the collection of input nodes of X_k by X^k_{in} ($:= \{X_i, \ldots, X_j\}$) and the prior selecting nodes of X_k by X^k_{sel} ($:= \{X_l, \ldots, X_m\}$). The joint distribution of $X_0, \ldots, X_N$ is then given by

$$p(x_0, \ldots, x_N) = \rho(w)\, P_1\big(x_1\big|x^1_{sel}, x^1_{in}\big) \cdots P_N\big(x_N\big|x^N_{sel}, x^N_{in}\big) \tag{3.1}$$

for all $x_k \in \mathcal{X}_k$ and $x^k_{sel} \in \mathcal{X}^k_{sel}$, $x^k_{in} \in \mathcal{X}^k_{in}$ $(k = 1, \ldots, N)$.

Specifying the sets X^k_{sel} and X^k_{in} of selectors and inputs for each node in the system then uniquely characterizes a particular multistep decision-making system. Note that we always have $(X^1_{sel}, X^1_{in}) = (\{\}, \{X_0\})$.

Decompositions of the form 3.1 are often visualized by directed acyclic graphs, so-called DAGs (see Bishop, 2006). Here, in addition to the decomposition of the joint in terms of posteriors, we have added the information about the prior dependencies in terms of dashed arrows, as shown in Figure 1.

3.3 Multistep Free Energy Principle. If P_k and q_k denote the posterior and prior of the kth node of an N-step decision process, then the free energy principle takes the form

$$\sup_{P_1, q_1, \ldots, P_N, q_N} \left(\langle U \rangle - \sum_{k=1}^{N} \frac{1}{\beta_k} \left\langle \mathrm{D}_{\mathrm{KL}}(P_k \| q_k) \right\rangle \right), \tag{3.2}$$

where, in addition to the expectation over inputs, the average of $\mathrm{D}_{\mathrm{KL}}(P_k \| q_k)$ now also includes the expectation with respect to X_{sel}:

$$\big\langle \mathrm{D}_{\mathrm{KL}}(P\|q) \big\rangle = \sum_{x_{sel}, x_{in}} p(x_{sel}, x_{in})\, \mathrm{D}_{\mathrm{KL}}\big(P(\,\cdot\,|x_{sel}, x_{in}) \| q(\,\cdot\,|x_{sel})\big).$$

Since the prior policies appear only in the KL divergences and, moreover, there is exactly one KL divergence per prior, it follows, as in section 2.5, that for each $k = 1, \ldots, N$, the optimal prior is the marginal given for all $x_k \in \mathcal{X}_k$ by

$$q_k(x_k|x_{sel}) = p_k(x_k|x_{sel}) := \frac{1}{p(x_{sel})} \sum_{\{x_0,\ldots,x_N\}\setminus(\{x_k\}\cup x_{sel})} p(x_0, \ldots, x_N), \tag{3.3}$$

whenever $X^k_{sel} = x_{sel}$. Hence, the free energy principle can be simplified to

$$\sup_{P_1,\ldots,P_N} \left(\langle U \rangle - \sum_{k=1}^{N} \frac{1}{\beta_k} I\big(X^k_{in}; X_k \big| X^k_{sel}\big) \right), \tag{3.4}$$

where $I(X; Y|Z)$ denotes the conditional mutual information of two random variables X, Y given a third random variable Z.

By optimizing equation 3.4 alternatingly, that is, optimizing one posterior at a time while keeping the others fixed, we obtain for each $k = 1, \ldots, N$,

$$P_k\big(x_k|x_{sel}, x_{in}\big) = \frac{p_k\big(x_k|x_{sel}\big)}{Z_k(x_{sel}, x_{in})} \exp\Big[\beta_k\, \mathcal{F}_k[P_1, \ldots, P_N](x_k, x_{sel}, x_{in})\Big], \tag{3.5}$$

whenever $X^k_{sel} = x_{sel}$ and $X^k_{in} = x_{in}$. Here, $Z_k(x_{sel}, x_{in})$ denotes the normalization constant and $\mathcal{F}_k[P_1, \ldots, P_N]$ denotes the (effective) utility function on which the decision making in X_k is based on. More precisely, given $\tilde{X} = (X_k, X^k_{sel}, X^k_{in})$, it is the free energy of the subsequent nodes in the system, that is, for any value of $\tilde{x} := (x_k, x_{sel}, x_{in})$ we obtain for $\mathcal{F}_k := \mathcal{F}_k[P_1, \ldots, P_N]$,

$$\mathcal{F}_k(\tilde{x}) = \frac{1}{p(\tilde{x})} \sum_{\{x_0,\ldots,x_N\}\setminus\tilde{x}} p(x_0, \ldots, x_N)\, \mathcal{F}_{k,\mathrm{loc}}(x_0, \ldots, x_N), \tag{3.6}$$

where

$$\mathcal{F}_{k,\mathrm{loc}}(x_0, \ldots, x_N) := U(x_0, x_N) - \sum_{i>k} \frac{1}{\beta_i} \log \frac{P_i(x_i|x^i_{sel}, x^i_{in})}{p_i(x_i|x^i_{sel})}.$$

Here, x^i_{in} and x^i_{sel} are collections of values of the random variables in X^i_{in} and X^i_{sel}, respectively. The final Blahut-Arimito-type algorithm consists of iterating equations 3.5, 3.3, and 3.6 for each $k = 1, \ldots, N$ until convergence is achieved. Note that since each optimization step is convex (marginal convexity), convergence is guaranteed but generally not unique (Jain & Kar, 2017), so that depending on the initialization, one might end up in a local optimum.

3.4 Example: Two-Step Information Processing. The cases of serial and parallel information processing studied in Genewein and Braun (2013) are special cases of the multistep decision-making systems introduced above. Both cases are two-step processes ($N = 2$) involving the variables $X_0 = W$, $X_1 = X$, and $X_2 = A$. The serial case is characterized by $(X^2_{sel}, X^2_{in}) = (\{\}, \{X_1\})$ and the parallel case by $(X^2_{sel}, X^2_{in}) = (\{X_1\}, \{X_0\})$. There is a third possible combination for $N = 2$, given by $(X^2_{sel}, X^2_{in}) = (\{\}, \{X_0, X_1\})$. However, it can be shown that this case is equivalent to the (one-step) rate distoration case from section 2, because if A has direct world state access, then any extra input to the final node $A = X_2$ that is not a prior selector contains redundant information.

4 Systems of Bounded Rational Agents

4.1 From Multistep to Multiagent Systems. As explained in section 2.6, a single random variable X_k that is part of an N-step decision-making system can represent a single agent or a collection of multiple agents, depending on the cardinality of $\mathcal{X}^k_{sel}$, that is, whether X_k has multiple priors selected by the nodes in X^k_{sel}. Therefore, an N-step bounded rational decision-making system with $N > 1$ represents a bounded rational multiagent system (of depth N).

For a given $k \in \{1, \ldots, N\}$, each value $x \in \mathcal{X}^k_{sel}$ of X^k_{sel} corresponds to exactly one agent in X_k. During decision making, the agents that belong to the nodes in X^k_{sel} are choosing which of the $|\mathcal{X}^k_{sel}|$ agents in X_k will receive a given input x_{in} (see section 4.4 for a detailed example). This decision is based on how well the selected agent x will perform on the input x_{in} by transforming its prior policy $p_k(\cdot|x)$ into a posterior policy $P_k(\cdot|x, x_{in})$, subject to the constraint

$$\langle \mathrm{D}_{\mathrm{KL}}(P_k||p_k)\rangle(x) := \sum_{x_{in}} p(x_{in}|x)\, \mathrm{D}_{\mathrm{KL}}\big(P_k(\cdot|x, x_{in})\|\, p_k(\cdot|x)\big) \leqslant D_x, \qquad (4.1)$$

where $D_x > 0$ is a given bound on the agent's information-processing capability. Similar to multistep systems, this choice is based on the performance measured by the free energy of the subsequent agents.

4.2 Multiagent Free Energy Principle. In contrast to multistep decision making, the information-processing bounds are allowed to be functions of the agents instead of just the nodes, resulting in an extra Lagrange multiplier for each agent in the free energy principle, equation 3.12. As in equation 3.4, optimizing over the priors yields the simplified free energy principle

$$\sup_{P_1,\ldots,P_N} \left(\langle U \rangle - \sum_{k=1}^{N} \sum_{x_{sel}^k \in \mathcal{X}_{sel}^k} \frac{p(x_{sel}^k)}{\beta_k(x_{sel}^k)} I\big(X_{in}^k; X_k \big| X_{sel}^k = x_{sel}^k\big) \right), \tag{4.2}$$

which can be solved iteratively as explained in the previous section, the only difference being that the Lagrange parameters β_k now depend on x_{sel}^k. Hence, for the posterior of an agent that belongs to node k, we have

$$P_k\big(x_k|x_{sel}, x_{in}\big) = \frac{p_k\big(x_k|x_{sel}\big)}{Z_k(x_{sel}, x_{in})} \exp\big[\beta_k(x_{sel}^k)\, \mathcal{F}_k(x_k, x_{sel}, x_{in})\big], \tag{4.3}$$

where $\beta_k(x_{sel}^k)$ is chosen such that constraint 4.1 is fulfilled for all $x \in x_{sel}^k$, and $\mathcal{F}_k$ is given by equation 3.6 except that now, we have

$$\mathcal{F}_{k,loc}(x_0, \ldots, x_N) := U(x_0, x_N) - \sum_{i>k} \frac{1}{\beta_i(x_{sel}^i)} \log \frac{P_i(x_i|x_{sel}^i, x_{in}^i)}{p_i(x_i|x_{sel}^i)}. \tag{4.4}$$

The resulting Blahut-Arimoto-type algorithm is summarized in algorithm 1.

4.3 Specialization. Although a given multiagent architecture predetermines the underlying set of choices for each agent, only a small part of such a set might be used by a given agent in the optimized system. For example, all agents in the final step potentially can perform any action $a \in \mathcal{A}$ (see Figure 2 and the example in section 4.4). However, depending on their indiviual information-processing capabilities, the optimization over the agents' priors can result in a (soft) partitioning of the full action space $\mathcal{A}$ into multiple chunks, where each of these chunks is given by the support of the prior of a given agent x, $\mathrm{supp}(p(\,\cdot\,|x)) \subset \mathcal{A}$. Note that the resulting partitioning is not necessarily disjoint, since agents might still be sharing a number of actions, depending on their available information-processing resources. If the processing capability is low compared to the number of possible actions in the full space and if there are enough agents at the same level, then this partitioning allows each agent to focus on a smaller number of options to choose from, provided that the coordinating agents have enough resources to decide among the partitions reliably.

Algorithm 1: Blahut-Arimito-Type Algorithm for Equation 4.2.

procedure GETMULTIAGENTSOLUTION($U, \rho, \{(X^k_{sel}, X^k_{in})\}_{k=1}^N, \beta, \epsilon$)

  initialize $p(x_0, \ldots, x_N)$ $\forall x_0, \ldots, x_N$,

  repeat

    $p_0 \leftarrow p$

    for $k = 1, \ldots, N$ **do**

      $\mathcal{F}_k(x_k, x_{sel}, x_{in}) \leftarrow$ (equations 3.6, 4.4) $\forall x_k, x_{sel}, x_{in}$ ▷ *calc. effective utility*

      $P_k(x_k|x_{sel}, x_{in}) \leftarrow$ (equation 4.3) $\forall x_k, x_{sel}, x_{in}$ ▷ *update posterior*

      $p(x_k|x_{sel}) \leftarrow$ (equation 3.3) $\forall x_k, x_{sel}$ ▷ *update prior*

      $p(x_0, \ldots, x_N) \leftarrow$ (equation 3.1) $\forall x_0, \ldots, x_N$ ▷ *update joint*

    end for

    $error \leftarrow \mathrm{dist}(p, p_0)$

  until $error < \epsilon$

  return $P_1, \ldots, P_N$

end procedure

Therefore, the amount of prior adaptation of an agent—by how much its optimal prior p deviates from a uniform prior p_0 over all accessible choices, which is measured by the KL divergence $\mathrm{D}_{\mathrm{KL}}(p\|p_0)$—determines its degree of specialization. More precisely, we define the specialization of an agent with prior p and choice space $\mathcal{X}$ by

$$\mathcal{S}[p] := \frac{\mathrm{D}_{\mathrm{KL}}(p\|p_0)}{\log|\mathcal{X}|} = 1 - \frac{H[p]}{\log|\mathcal{X}|}, \tag{4.5}$$

where $H[p] := -\sum_x p(x)\log p(x)$ denotes the Shannon entropy of p. By normalizing with $\log|\mathcal{X}|$, we obtain a quantity between 0 and 1, since $0 \leqslant H(p) \leqslant \log|\mathcal{X}|$. Here, $\mathcal{S}[p] = 0$ corresponds to $H[p] = \log|\mathcal{X}|$, which means that the agent is completely unspecialized, whereas $\mathcal{S}[p] = 1$ corresponds to $H[p] = 0$, which implies that p has support on a single option $x^* \in \mathcal{X}$, meaning that the agent deterministically performs the same action always and therefore is fully specialized.

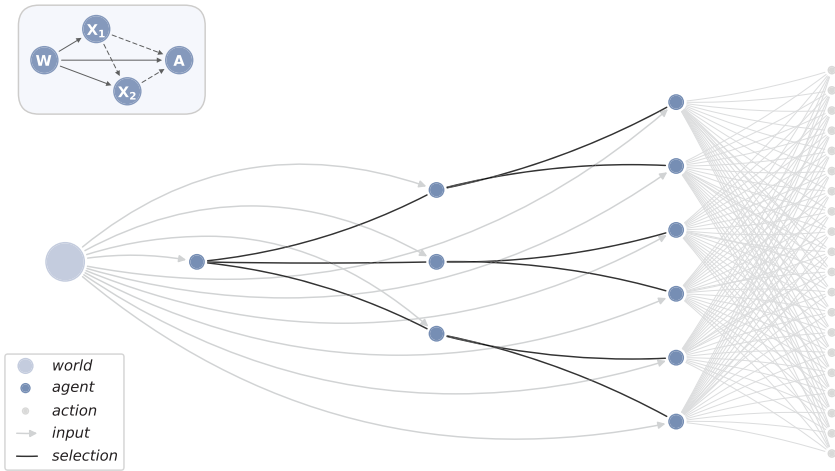

Figure 2: Example of a hierarchical architecture of 10 agents that are combined using the three-step decision-making system ($N = 3$) shown in the upper-left corner (see section 4.4 for details). Here, every node—and therefore every agent—has access to the world states (big circle). X_1 consists of one agent that decides about which of the $|\mathcal{X}_1| = 3$ agents in X_2 obtains a given world state as input. The selected agent in X_2 selects which of the $|\mathcal{X}_2| = 2$ agents out of the $|\mathcal{X}_1| \cdot |\mathcal{X}_2| = 6$ agents in A that are connected to it, obtains the world state to perform the final decision about an action $a \in \mathcal{A}$ (gray circles on the right). In our notation introduced below, this architecture is labeled by $(1, 4)_{[1,3,(3,2)]}$ (see section 5.1).

4.4 Example: Hierarchical Multiagent System with Three Levels. Consider the example of an architecture of 10 agents shown in Figure 2 that are combined via the three-step decision-making system given by

$$(X^2_{sel}, X^2_{in}) = (\{X_1\}, \{W\}), \quad (X^3_{sel}, X^3_{in}) = (\{X_1, X_2\}, \{W\}), \tag{4.6}$$

as visualized in the upper-left corner of Figure 2. The number of agents in each node is given by the cardinality of the target space of the selecting node(s) (or equals one if there are no selectors). Hence, X_1 consists of one agent, X_2 consists of $|\mathcal{X}_1|$ agents, and A consists of $|\mathcal{X}_1| \cdot |\mathcal{X}_2|$ agents. For example, if we have $|\mathcal{X}_1| = 3$ and $|\mathcal{X}_2| = 2$, as in Figure 2, then this results in a hierarchy of one, three, and six agents.

The joint probability of the system characterized by equation 4.6 is given by

$$p(w, x_1, x_2, a) = p(w)P_1(x_1|w)P_2(x_2|x_1, w)P_3(a|x_2, x_1, w)$$

and the free energy by

$$\mathcal{F}[P_1, P_2, P_3] = \sum_{w,x_1,x_2,a} p(w, x_1, x_2, a)\Big[U(a, w) - \frac{1}{\beta_1} \log \frac{P_1(x_1|w)}{p_1(x_1)} - \frac{1}{\beta_2(x_1)} \log \frac{P_2(x_2|x_1, w)}{p_2(x_2|x_1)} - \frac{1}{\beta_3(x_1, x_2)} \log \frac{P_3(a|x_2, x_1, w)}{p_3(a|x_2, x_1)}\Big],$$

where the priors p_1, p_2, and p_3 are given by the marginals (see equation 3.3):

$$p_1(x_1) = \sum_w \rho(w) P(x_1|w),$$
$$p_2(x_2|x_1) = \sum_w p(w|x_1) P_2(x_2|x_1, w),$$
$$p_3(a|x_2, x_1) = \sum_w p(w|x_1, x_2) P_3(a|x_2, x_1, w).$$

By equation 3.5, the posteriors that iteratively solve the free energy principle are

$$P_1(x_1|w) = \frac{p_1(x_1)}{Z(w)} \exp\big[\beta_1 \mathcal{F}_1(w, x_1)\big],$$
$$P_2(x_2|x_1, w) = \frac{p_2(x_2|x_1)}{Z(w, x_1)} \exp\big[\beta_2(x_1) \mathcal{F}_2(w, x_1, x_2)\big],$$
$$P_3(a|x_2, x_1, w) = \frac{p_3(a|x_2, x_1)}{Z(w, x_1, x_2)} \exp\big[\beta_3(x_1, x_2) U(a, w)\big],$$

where, by equations 3.6 and 4.4,

$$\mathcal{F}_1(w, x_1) := \sum_{x_2,a} p(x_2, a|x_1, w)\Big[U(a, w) - \frac{1}{\beta_2(x_1)} \log \frac{P_2(x_2|x_1, w)}{p_2(x_2|x_1)} - \frac{1}{\beta_3(x_1, x_2)} \log \frac{P_3(a|x_2, x_1, w)}{p_3(a|x_2, x_1)}\Big],$$
$$\mathcal{F}_2(w, x_1, x_2) := \sum_a P_3(a|x_2, x_1, w)\Big[U(a, w) - \frac{1}{\beta_3(x_1, x_2)} \log \frac{P_3(a|x_2, x_1, w)}{p_3(a|x_2, x_1)}\Big].$$

Given a world state $w \in \mathcal{W}$, the agent in X_1 decides which of the three agents in X_2 obtains w as an input. This narrows down the possible choices for the selected agent in X_2 to two out of the six agents in A. The selected agent performs the final decision by choosing an action $a \in \mathcal{A}$. Depending on its degree of specialization, which is a result of his own and the

coordinating agents' resources, this agent will choose his action from a certain subset of the full space $\mathcal{A}$.

5 Optimal Architectures

Here, we show how the framework we have described can be used to determine optimal architectures of bounded rational agents. Summarizing the assumptions made in the derivations, the multiagent systems that we analyze must fulfill the following requirements:

i. The information flow is feedforward. An agent in X_k can obtain information directly from another agent that belongs to X_m only if $m < k$.
ii. Intermediate agents cannot be end points of the decision-making process. The information flow always starts with the processing of W and always ends with a decision $a \in A$.
iii. A single agent is not allowed to have multiple prior policies. Agents are the smallest decision-making unit, in the sense that they transform a prior to a posterior policy over a set of actions in one step.

The performance of the resulting architectures is measured with respect to the expected utility they are able to achieve under a given set of resource constraints. To this end, we need to specify (1) the objective for the full decision-making process, (2) the number N of decision-making steps in the system, (3) the maximal number n of agents to be distributed among the nodes, and (4) the individual resource constraints $\{D_1, \ldots, D_n\}$ of those agents. We illustrate these specifications with a toy example in section 5.2 by showcasing and explicitly explaining the differences in performance of several architectures. Moreover, we provide a broad performance comparison in section 5.3, where we systematically vary a set of objective functions and resource constraints in order to determine which architectural features most affect the overall performance. For simplicity, in all simulations, we are limiting ourselves to architectures with $N \leqslant 3$ nodes and $n \leqslant 10$ agents. In the following section, we start by describing how we characterize the architectures conforming to the three requirements.

5.1 Characterization of Architectures

5.1.1 Type. In order to be able to reference the architectures resulting from *i–iii*, we label an N-step decision-making process with $N > 1$ by a tuple $(i_1, \ldots, i_{N-1})$ of length $N - 1$ which we call the *type* of the architecture, where i_{N-2} characterizes the relation between the first N variables $X_0, \ldots, X_{N-1}$, and i_{N-1} determines how these variables are connected to X_N. Note that the explicit mapping between an index and the corresponding relation of random variables is arbitrary.

For example, for $N \leqslant 3$, we define the types shown in Figure 3, where $i \in \{0, 1, 2\}$ and $j \in \{0, \ldots, 5\}$ represent the following relations:

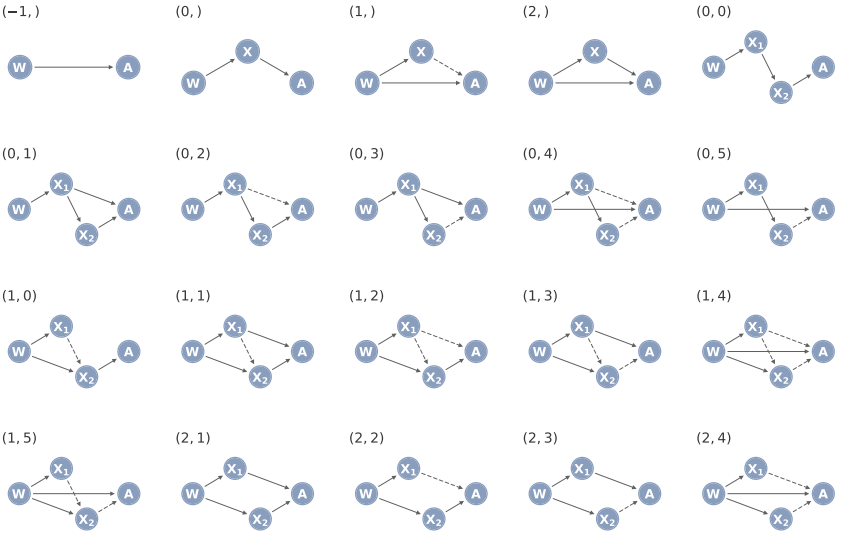

Figure 3: Overview of the resulting architectures for $N \leqslant 3$, each of them labeled by its type.

$$
\begin{aligned}
i = 0: &\ (X^2_{sel}, X^2_{in}) = (\{\}, \{X_1\}),\\
i = 1: &\ (X^2_{sel}, X^2_{in}) = (\{X_1\}, \{W\}),\\
i = 2: &\ (X^2_{sel}, X^2_{in}) = (\{\}, \{W\}),\\
j = 0: &\ (X^3_{sel}, X^3_{in}) = (\{\}, \{X_2\}),\\
j = 1: &\ (X^3_{sel}, X^3_{in}) = (\{\}, \{X_1, X_2\}),\\
j = 2: &\ (X^3_{sel}, X^3_{in}) = (\{X_1\}, \{X_2\}),\\
j = 3: &\ (X^3_{sel}, X^3_{in}) = (\{X_2\}, \{X_1\}),\\
j = 4: &\ (X^3_{sel}, X^3_{in}) = (\{X_1, X_2\}, \{W\}),\\
j = 5: &\ (X^3_{sel}, X^3_{in}) = (\{X_2\}, \{W\}).
\end{aligned}
$$

For example, the architecture shown in Figure 2 has the type $(1, 4)$. Correspondingly, the two-step cases are labeled by $(i,)$ for $i \in \{0, 1, 2\}$, and the one-step rate distoration case by $(-1,)$. Note that not every combination of $i \in \{0, 1, 2\}$ and $j \in \{0, \ldots, 5\}$ describes a unique system; for example, $(2, 3)$ is equivalent to $(2, 2)$ when replacing X_1 by X_2. Moreover, as already mentioned, $(2,)$ is equivalent to $(-1,)$, and similarly, $(0, 1)$ is equivalent to $(0,)$.

5.1.2 Shape. After the number of nodes has been fixed, the remaining property that characterizes a given architecture is the number of agents per

node. For most architectures, there are multiple possibilities to distribute a given number of agents among the nodes, even when neglecting individual differences in resource constraints. We call such a distribution a shape, denoted by $[n_1, n_2, \ldots]$, where n_k denotes the number of agents in node k. Note that not all architectures will be able to use the full number of available agents, most immanently the one-step rate distortion case (one agent), or the two-step serial case (two agents). For these systems, we always use the agents with the highest available resources in our simulations.

For example, for $N \leqslant 3$, the resulting shapes for a maximum of $n = 10$ agents are as follows:

- [1] for $(-1,)$, [1, 1] for (0,), and [1, 9] for (1,)
- [1, 1, 1] for (0, 0) and (2, 1)
- [1, 1, 8] for (0, 2), (0, 3), (0, 5), (2, 2)
- [1, 1, (2, 4)] and [1, 1, (4, 2)] for (0, 4) and (2, 4)
- [1, 8, 1] for (1, 0) and (1, 1)
- [1, 4, 4] for (1, 2)
- [1, 2, 7], [1, 3, 6], [1, 4, 5], [1, 5, 4], [1, 6, 3], [1, 7, 2] for (1, 3) and (1, 5)
- [1, 2, (2, 3)] and [1, 3, (3, 2)] for (1, 4),

where a tuple inside the shape means that two different nodes are deciding about the agents in that spot; for example $[1, 1, (2, 4)]$ means that there are eight agents in the last node, labeled by the values $(x_1, x_2) \in \mathcal{X}_1 \times \mathcal{X}_2$ with $|\mathcal{X}_1| = 2$ and $|\mathcal{X}_2| = 4$. In Figure 4, we visualize one example architecture for each of the above three-step shapes, except for the shapes of type $(1, 4)$ of which one example is shown in Figure 2.

Together, the type $(i, \ldots)$ and shape $[n_1, \ldots]$ uniquely characterize a given multiagent architecture, denoted by $(i, \ldots)_{[n_1, \ldots]}$.

5.2 Example: Call Center. Consider the operation of a company's call center as a decision-making process, where customer calls (world states) must be answered with an appropriate response (action) in order to achieve high customer satisfaction (utility). The utility function shown on the left of Figure 5 can be viewed as a simplistic model for a real-world call center of a big company such as a communication service provider. In this simplification, there are 24 possible customer calls that belong to three separate topics—for example, questions related to telephone, Internet, or television—which can be further subdivided into two subcategories—for example, consisting of questions concerning the contract or problems with the hardware. (See the Figure 5 caption for the explicit utility values.)

Handling all possible phone calls perfectly by always choosing the corresponding response with maximum utility requires $\log_2(24) \approx 4.6$ bits (see Figure 5). However, in practice, a single agent is usually not capable of knowing the optimal answers to every type of question. For our example, this means that the call center has access only to agents with information

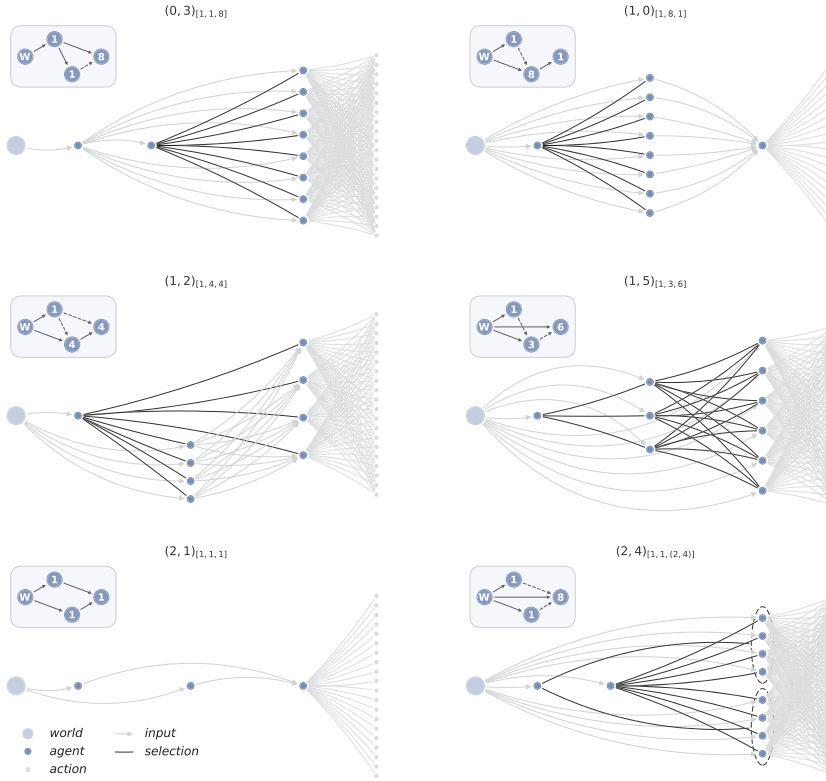

Figure 4: Visualization of exemplary three-step multiagent architectures specified by their types and shapes.

processing capability less than 4.6 bits. It is then required to organize the agents in a way so that each agent has to deal with only a fraction of the customer calls. This is often realized by first passing the phone call through several filters in order to forward it to a specialized agent. Arranging these selector or filter units in a strict hierarchy then corresponds to architectures of the form of $(1, 4)$ or $(1, 5)$ (see below for a comparison of these two), where at each stage, a single operator selects how a call is forwarded. In contrast, architectures of the form of $(2, 4)$ allow for multiple independent filters working in parallel—for example, realized by multiple trained neural networks, where each is responsible for a particular feature of the call (say, one node deciding about the language of the call and another node deciding about the topic). In the following, we do not discriminate between human and artificial decision makers, since both can qualify equally well as information-processing units.

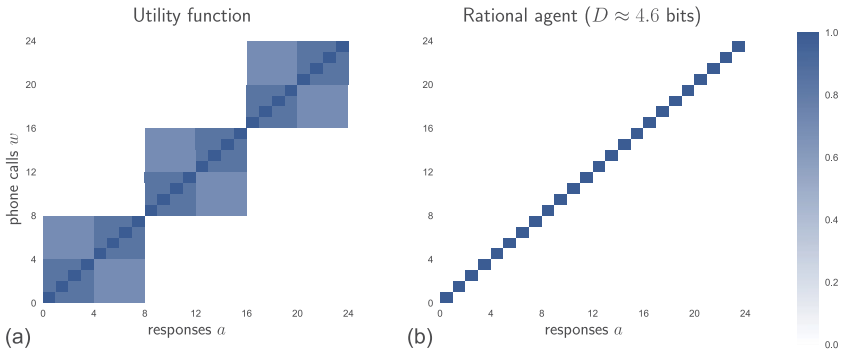

Figure 5: (a) Utility function for example 5.2. The set of phone calls W is partitioned into three separate regions, corresponding to three different topics about which customers might have complaints or questions. Each of these can be divided into two subcategories of four customer calls each. For each phone call, there is exactly one answer that achieves the best result ($U = 1$). Moreover, the responses that belong to one subcategory of calls are also suitable for the other calls in that particular subcategory, albeit slightly less effective ($U = 0.85$) than the optimal answers. Similarly, the responses that belong to the same topic of calls are still a lot better ($U = 0.7$) than responses to other topics ($U = 0$). (b) Posterior policy of a single agent with an information bound of 4.6 bits.

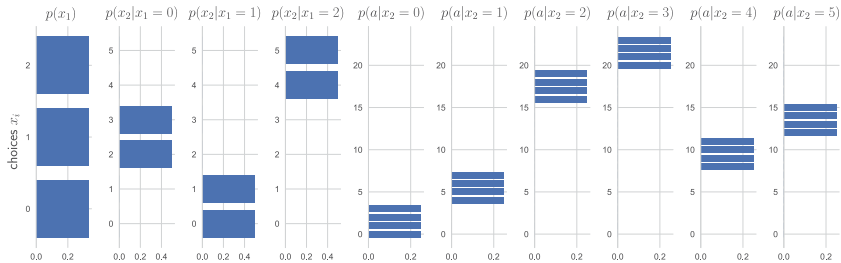

Figure 6: Optimal prior policies for each agent of the architecture $(1, 5)_{[1,3,6]}$ with an information bound of $(D_1, D_2, \ldots, D_{10}) = (1.6, 0.1, \ldots, 0.1)$.

Assume that there are $n = 10$ bounded rational agents available. Considering the given utility function, the architectures $(1, 4)_{[1,3,(3,2)]}$ (shown in Figure 2) and $(1, 5)_{[1,3,6]}$ (shown in Figure 4) might be obvious choices as they represent the hierarchical structure of the utility function. With an information bound of 1.6 ($\approx \log_2(3)$) bits for the first agent and 0.1 bits for the rest, the optimal prior policies for $(1, 5)_{[1,3,6]}$ obtained by our free energy principle are shown in Figure 6. We can see that for this architecture, the choice x_1 of the agent at the first step corresponds to the general topic of the phone call, the decisions x_2 of the three agents at the second stage

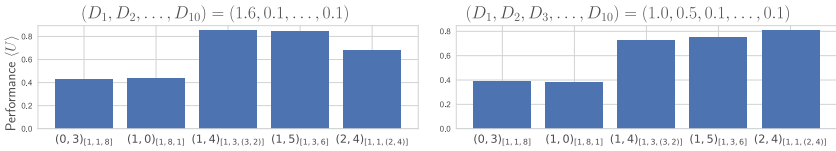

Figure 7: Performance comparison under two different information bounds.

correspond to the subcategory on which one of the six agents at the final stage is specialized to, who then makes the decision about the final response a by picking one of the four actions in the support of its prior.

We can see in Figure 7 on the left that a hierarchical structure as in $(1, 5)_{[1,3,6]}$ or $(1, 4)_{[1,3,(3,2)]}$ is indeed superior when comparing with the architecture $(2, 4)_{[1,1,(2,4)]}$, because there is no good selector for the second filter. We have also added two architectures to the comparison that have a bottleneck of the information flow at either end of the decision-making process, $(0, 3)_{[1,1,8]}$ and $(1, 0)_{[1,8,1]}$ (see Figure 4 for a visualization), which are performing considerably worse than the others: in $(0, 3)_{[1,1,8]}$, the first agent is the only one who has direct contact to the customer and passes the filtered information on to everybody else, whereas in $(1, 0)_{[1,8,1]}$, the customer talks to multiple agents; however, they cannot make any decisions and pass on the information to a final decision node who has to select from all possible options. Interestingly, as can be seen on the right side of Figure 7, when changing the resource bounds such that the first agent has only $D_1 = 1$ bits instead of 1.6 and the second agent has $D_2 = 0.5$ bits instead of 0.1, then the strictly hierarchical architectures $(1, 5)_{[1,3,6]}$ and $(1, 4)_{[1,3,(3,2)]}$ are outperformed by the architecture $(2, 4)_{[1,1,(2,4)]}$, because their first agent is not able to perfectly distinguish among the three topics anymore. This is an ideal situation for $(2, 4)_{[1,1,(2,4)]}$, since here, the total information processing for filtering the phone calls is split efficiently between the first two agents in the system.

Note that $(1, 4)$ and $(1, 5)$ do not necessarily perform identically (as can be seen on the right in Figure 7), even though the structure of the utility function might suggest that it is ideal for $(1, 5)_{[1,3,6]}$ to always have the optimal priors shown in Figure 6. However, this crucially depends on the given information-processing bounds. In Figure 8, we illustrate the difference between the two types in more detail by showing the processed information that can actually be achieved per agent in the respective architecture for an information bound of $D = (0.4, 2.6, 2.6, 2.6, 0.4, \ldots, 0.4)$. When the first agent in the hierarchy has low capacity, then the rigid structure of $(1, 4)$ is penalized because the agents at the second stage cannot compensate the errors of the first agent, regardless of their capacity. In contrast, for $(1, 5)$, the connection between the second stage and the executing stage can be changed freely, which leads to ignoring the first agent and letting the three

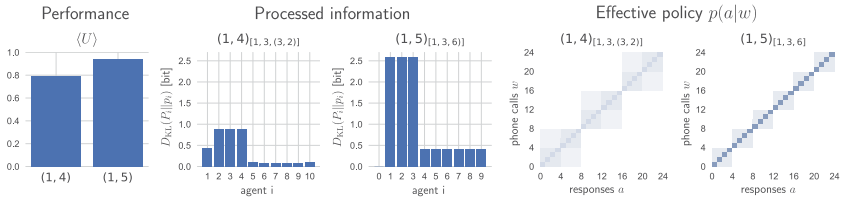

Figure 8: Demonstration of the difference between the two architectures $(1,4)_{[1,3,(3,2)]}$ and $(1,5)_{[1,3,6]}$ for an information bound of $D = (0.4, 2.6, 2.6, 2.6, 0.4, \ldots, 0.4)$.

agents in the second stage determine the distribution of phone calls completely. In this sense, $(1,5)$ is more robust to errors in the first filter than $(1,4)$.

5.3 Systematic Performance Comparison. In this section, we move away from an explicit toy example to a broad performance comparison of all architectures for $N \leqslant 3$, averaged over multiple types of utility functions and a large number of resource constraints (as defined below). In section 6.1, this is supplemented with an analysis of the architectural features that best explain the performances.

5.3.1 Objectives. We compare all possible architectures for 12 different utility functions, $\{U_k\}_{k=1}^{12}$, defined on a world and action space of $|\mathcal{W}| = |\mathcal{A}| = 20$ elements, and we assume the same cardinality for the range of all hidden variables. Note that the cardinality of the target set $\mathcal{X}$ for selector nodes $X \in X_{sel}$ is given by the number of agents it decides about. In particular, we consider three kinds of utility functions (one-to-one, many-to-one, one-to-many) that we vary in a 2×2 paradigm, where the first dimension is the number of maximum utility peaks (single, multiple) and the second dimension is the range of utility values (binary, multivalued). The utility functions are visualized in Figure 9, where the three kinds of functions correspond to the three rows of the plot. A one-to-one scenario applies to a needle-in-a-haystack situation where each world state affords only a unique action and, vice versa, each optimal action allows uniquely identifying the world state, for example, an absolute identification task. A many-to-one scenario allows for abstractions in the world states, for example, in categorization when multiple instances are judged to belong to the same class (e.g., vegetables are boiled; fruit is eaten raw). A one-to-many scenario allows for abstractions in the action space—for example, in hierarchical motor control when a grasp action can be performed in many different ways.

5.3.2 Resource Limitations. We are considering three schemes of resource constraints:

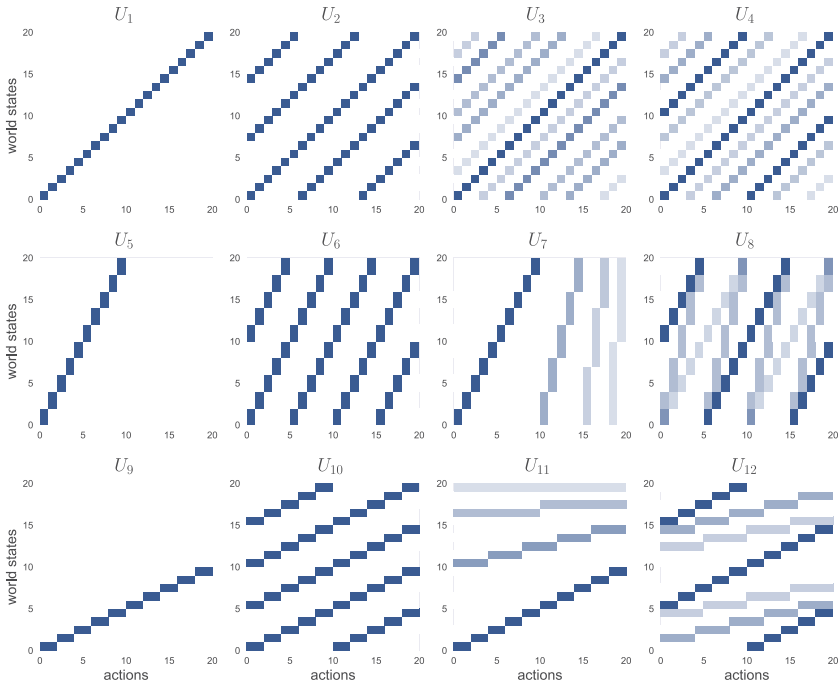

Figure 9: Utility functions on which the performances are measured.

1. Same constraints for all agents
2. Same constraints for all agents but one, which has a higher limit than the other agents
3. Same constraints for all but two agents, which can have a different limit and have higher limits than all the other agents.

For the first constraint, we compare 20 sets of constraints $\{D_0, D_1, \ldots\}$ with D_i equally spaced in the range between 0 and 3 bits; for the second, we compare 39 sets in the same range but the high resource agent having 1, 2 and 3 bits; and for the third, we allow 89 sets with similar constraints than in the second constraint but additional combinations for the second high-resource agent.

5.3.3 Simulation Results. The performance of an architecture is given by its expected utility with respect to a given objective and a given information bound as defined above. In Figure 10, we show which of the architectures won at least one condition, together with the proportion of conditions won by each of these architectures. We can see that $(2, 4)_{[1,1,(2,4)]}$ overall outperforms all the other systems (see Figure 4 for a visualization). When all

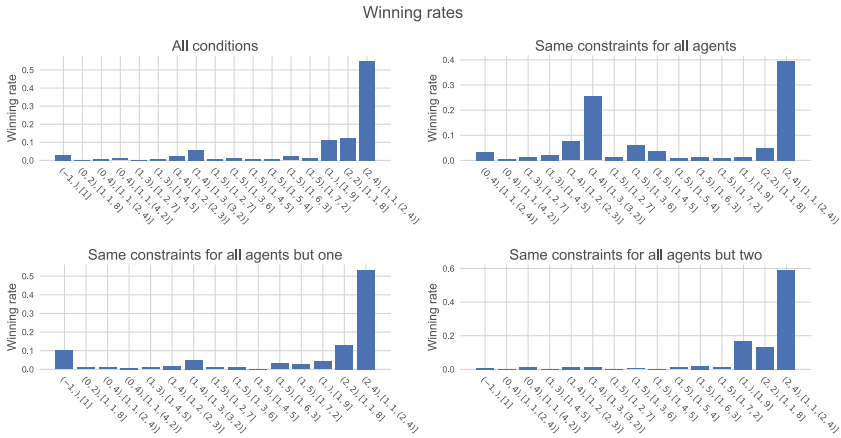

Figure 10: Proportion of conditions where the given architectures had the highest performance, for all conditions, and separately for each of the three different schemes of resource constraints.

agents have the same resource constraints, the architecture $(1, 4)_{[1,3,(3,2)]}$ is a strong second winner; however, this is not the case if one or two agents have more resources than the rest. It is not surprising that in these situations, the parallel case with one high-resource agent distributing the work among the low-resource agents, and even the case of a single agent that does everything by himself, are both performing well.

A closer look on the achieved expected utilities, however, shows several architectures that are almost equally well performing for many conditions. In order to increase comparability between the different utility functions, we measure performance in terms of a relative score, which, for a given utility function and resource constraint, is given by the architectures' expected utility divided by the maximum expected utility of all architectures. The score averaged over all conditions is shown for each architecture in Figure 11 in the top row. We can see that the best architectures are pretty close to each other. As expected, the architecture that won the most conditions also has the highest overall performance; however, there are multiple architectures that are very close. The top three architectures are

$$(2, 4)_{[1,1,(2,4)]},\ (1, 5)_{[1,3,6]},\ (1, 4)_{[1,3,(3,2)]}, \tag{5.1}$$

which have already been visualized (see Figures 2 and 4).

A better understanding of their performances under different resource constraints can be gathered from the remaining graphs in Figure 11. In the second row, we can see that the top three overall architectures also perform best for almost all utility functions when averaged over the information

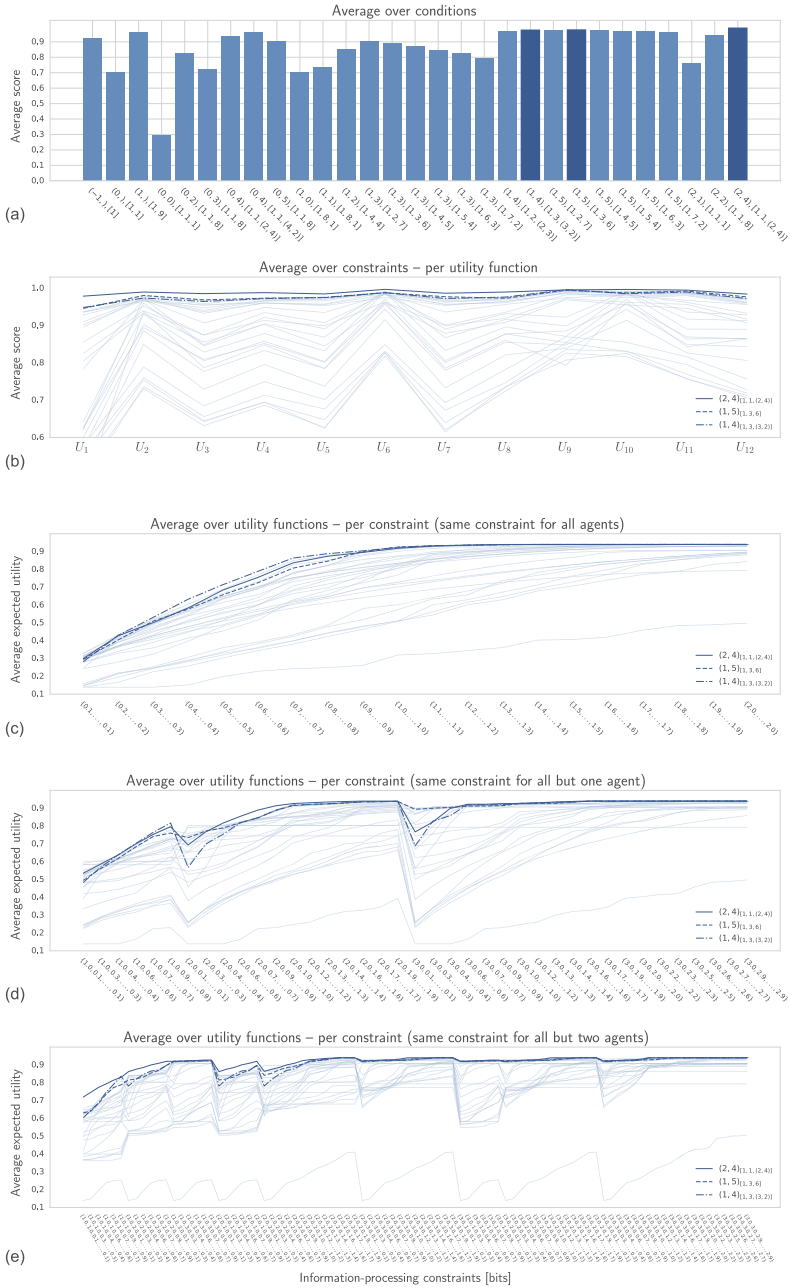

Figure 11: Architecture performances averaged over all conditions (a), averaged over all information bounds for each utility function (b), and averaged over all utility functions for each information bound (c–e).

bounds. The last three graphs in Figure 11 show the expected utility of each architecture averaged over all utility functions for each information bound. We can see how the expected utility increases with higher information bounds for some architectures more than for others. The top three architecures perform differently for most of the bounds, with spans of bounds where each of them clearly outperforms the others.

6 Discussion

6.1 Analysis of the Simulations. Plenty of factors influence the performance of each of the given architectures. Here, we attempt to unfold the features that determine their performances in the clearest way. To this end, we compare the architectures with respect to the following quantities:

Average specialization of operational agents: The specialization (see equation 4.5) averaged over all agents in the final stage of the architecture
Hierarchical: Boolean value that specifies whether an architecture is hierarchical or not, meaning that consecutive nodes are occupied by an increasing amount of agents
Agents with direct w-access: The number of agents with direct world state access
Operational agents with direct w-access: The number of agents in the last node of the architecture
Number of w-bottlenecks: The total number of nodes that are missing direct access to the world state

As can be seen from Figure 12, we found that these architectural features explain the differences in performance quite well. More precisely, the architectures can be roughly grouped into different categories, indicated by slightly different color saturations in Figure 12). The poorest-performing group consists of architectures that have between one and two w-bottlenecks, and therefore have only a few agents with direct w-access; in particular, none of their operational agents has direct w-access. Moreover, in this group, most architectures are not hierarchical at all, and their operational agents have low specialization, with two exceptions that both have two w-bottlenecks.

The architectures with medium performance have maximally one w-bottleneck, and many of them are hierarchical. Here, systems that have operational units with high specialization are missing direct w-access, and the systems that have operational units with direct w-access have low specialization.

All architectures in the top group have many agents with direct world-state access, and they have no w-bottlenecks. Interestingly, the best six architectures are all strictly hierarchical. Moreover, the order of performance

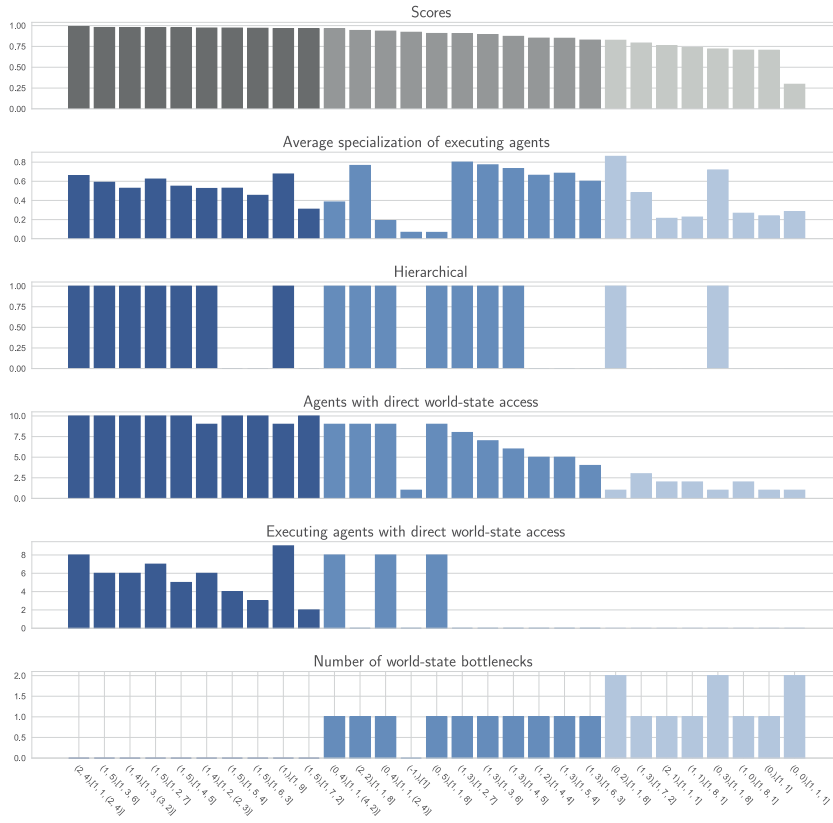

Figure 12: Proposed features to explain the architectures' performances (see section 6.1).

is almost in direct accordance with the average specialization of the operational agents.

Overall we can say that it is best to have as many operational units as needed to discriminate the actions well, as long as the coordinating agents have enough resources to discriminate among them properly. The architecture $(1, 4)_{[1,1,(2,4)]}$ has eight operational agents managed by two coordinating units, which need maximally two bits (for choosing among four agents) and one bit (for choosing among two agents) in order to perform well. Both of the other top three architectures, $(1, 5)_{[1,3,6]}$ and $(1, 4)_{[1,3,(3,2)]}$, have six operational agents, managed by three coordinating units, so that each of them needs maximally one bit. But compared to $(1, 4)_{[1,1,(2,4)]}$, there are fewer agents to spare for the operational stage. Hence, if the operational

units have low resources, it is always a trade-off between the number of operational units and the resources of the coordinating ones.

Another way to see why the architecture $(1, 4)_{[1,1,(2,4)]}$ overall outperforms all the other high-ranked systems might be its lower average choice-per-agent ratio—the average number of options for the decision of each agent in the system. In $(1, 4)_{[1,1,(2,4)]}$, the second agent also directly observes the world state; moreover, the choice space of eight agents at the operational stage is split into two and four choices. Therefore, there are only $\frac{2+4+20}{10} = 2.6$ choices per agent on average, whereas for $(1, 5)_{[1,3,6]}$ and $(1, 4)_{[1,3,(3,2)]}$, there are $\frac{3+6+20}{10} = 2.9$.

6.2 Limitations of Our Analysis. The analysis we have presented provides only a rough explanation of the differences in performance. Which architecture is optimal depends a lot on the actual information bounds of each agent. In all of our conditions, we assumed that most agents have the same processing capabilities, which is why there is a certain bias toward architectures that perform well under this assumption (low variance in choice-per-agent ratio across the agents).

Due to the large number of Lagrange parameters in the free energy principle (see equation 4.2), the data generation was done by running the Blahut-Arimoto-type algorithm for 10,000 different combinations of parameters for each of the architectures, for each type of the three different types of resource limitations in section 5.3, and for each of the utility functions defined in section 5.3. For a given information bound, the corresponding parameters were determined by looking for the points with the highest free energy that still respect the bound. A better approach would be to enhance the global parameter search by a more finely grained local search. Another possibility is to use an evolutionary algorithm, where each population is given by multiple sets of parameters and the information constraints are built in by a method similar to Chehouri, Younes, Perron, and Ilinca (2016). This works well but requires significantly more time to process.

Since the Blahut-Arimoto type of algorithm is not guaranteed to converge to a global maximum, the resulting values for the expected utility and mutual information for a given set of parameters can depend on the initialization of the algorithm. In practice, this variation is small enough so that it influences the average performance over multiple conditions by only a negligable amount. However, direct comparisons of architectures for a given information bound and utility function should be repeated multiple times to make sure that the results are stable.

6.3 Relation to Variational Bayes and Active Inference. Above, we determined the architectures that achieve the highest expected utility under a given resource constraint. These constraints are fulfilled by tuning the Lagrange multipliers in the free energy principle. If the Lagrange multipliers

themselves are fixed, for instance, as exchange rates between information and utility (Ortega & Braun, 2010), or inverse temperatures in thermodynamics (Ortega & Braun, 2013), then the free energy itself would be an appropriate performance measure. This is done, for example in Bayesian model selection, which is also known as structure learning and represents an important problem in Bayesian inference and machine learning. The Bayesian approach for evaluating different Bayesian network structures, in order to find the relation of a given set of hidden variables that best explains a data set $\mathcal{D}$, consists in comparing the marginal likelihood or evidence $p(\mathcal{D}|S)$ of the structures S (Friedman & Koller, 2003). This can be seen to be analogous to a performance comparison of different decision-making architectures measured by the free energy. In the simple case of one observable Y and one hidden variable X, we have

$$p(y|S) = \sum_{x \in \mathcal{X}} p(x|S)\, p(y|x, S) \qquad \forall y \in \mathcal{Y},$$

where the likelihood $p(y|x, S)$ is assumed to be known. Given a prior $p(x|S)$ and, for simplicity, a single observed data point $y \in Y$, the posterior distribution of X can be inferred by using Bayes' rule:

$$p(x|y, S) = \frac{p(x|S)\, p(y|x, S)}{p(y|S)} \qquad \forall x \in \mathcal{X}. \tag{6.1}$$

As Ortega and Braun (2013) have noted, when comparing equation 6.1 with the Boltzmann equation, 2.4, we can see that equation 6.1 is equivalent to the posterior P of a bounded rational decision maker with choice space $\mathcal{X}$, prior policy $p(x|S)$, Lagrange parameter $\beta = 1$, and utility function given by $U(x) := \log p(y|x, S)$. Since the marginal likelihood $p(y|S)$ is the normalization constant in equation 6.1, it follows immediately from equation 2.5 that $\log p(y|S)$ is the optimal free energy $\mathcal{F}_{var}[P = p(\cdot\,|y, S)]$ of this decision maker, where

$$\begin{aligned}
\mathcal{F}_{var}[P] &:= \sum_x P(x) \log p(y|x, S) - \sum_x P(x) \log \frac{P(x)}{p(x|S)} \\
&= \sum_x P(x) \log \frac{p(x|y, S)}{P(x)} + \log p(y|S) \\
&= \sum_x P(x) \log \frac{p(x, y|S)}{P(x)}.
\end{aligned} \tag{6.2}$$

In Bayesian statistics, $\mathcal{F}_{var}$ is known as the variational free energy, and the given decomposition is often referred to in terms of the difference between accuracy (expected log likelihood) and complexity (KL divergence

between prior and posterior). It is used in the variational characterization of Bayes' rule, that is, the approximation of the exact Bayesian posterior $p(\cdot|y, S)$ given by equation 6.1 in terms of a simpler—for example, a parameterized—distribution q by minimizing the KL divergence between q and $p(\cdot|y, S)$. Since $D_{KL}(q\|p(\cdot|y, S)) = -\mathcal{F}_{var}[q] + \log p(y, S)$, this is equivalent to the maximization of $\mathcal{F}_{var}$.

The same is true for multiple hidden variables. For example, let S be the three-step architecture of type $(1, 4)$ from section 4.4 with $W = Y$ and hidden variables X_1, X_2, and $X_3 = A$. Setting $\beta_1 = \beta_2 = \beta_3 = 1$ and $U(a, x_1, x_2, y) = \log p(y|a, x_1, x_2, S)$, we obtain

$$\mathcal{F}_2(y, x_1, x_2) = \log p(y|x_1, x_2, S), \quad \mathcal{F}_1(y, x_1) = \log p(y|x_1, S)$$

and

$$Z(y, x_1, x_2) = p(y|x_1, x_2, S), \quad Z(y, x_1) = p(y|x_1, S), \quad Z(y) = p(y|S).$$

Note that although so far, we always assumed that the utility function depends on only the world states and actions, the equations in sections 3, 4, and 4.4 are also valid in the general case of U depending on all the variables in the system. The total free energy for a given $y \in \mathcal{Y}$ then takes the form

$$\sum_{x_1,x_2} p(x_1, x_2|y, S) \left(\log p(y|x_1, x_2, S) - \log \frac{p(x_2|x_1, y, S)}{p(x_2|x_1, S)} - \log \frac{p(x_1|y, S)}{p(x_1|S)} \right)$$

$$= \sum_{x_1} p(x_1|y, S) \left(\log p(y|x_1, S) - \log \frac{p(x_1|y, S)}{p(x_1|S)} \right) = \log p(y|S).$$

Hence, also in this case, the logarithm of the marginal likelihood is given by the free energy of the corresponding decision-making system. Choosing the multistep architecture with the highest free energy is then analogous to Bayesian model selection with the marginal likelihood or Bayesian model evidence as performance measure.

Another interesting interpretation of equation 6.2 is that here, the hidden variable X can be thought of as an action causing observed outcomes y. This is close to the framework of active inference (Friston, Rigoli et al., 2015; Friston, Parr et al., 2017), where actions directly cause transitions of hidden states, which generate outcomes that are observed by the actor. More precisely, there the real-world process generating observable outcomes is distinguished from an internal generative model describing the beliefs about the external generative process (e.g., a Markov decision process). Observations are generated from transitions of hidden states, which depend on the decision maker's actions. Decision making is given by the optimization of a variational free energy analogous to equation 6.2, where the log likelihood

is given by the generative model, which describes beliefs about the hidden and control states of the generative process. This way, utilities are absorbed into a (desired) prior (Ortega & Braun, 2015). There are several differences to our approach. First, the structure of the free energy principle of bounded rationality originates from the maximization of a given predefined external utility function under information constraints, whereas the free energy principle of active inference aims to minimize surprise or Bayesian model evidence, effectively minimizing the divergence between approximate and true posterior. Second, in active inference, utility is transformed into preferences in terms of prior beliefs, while in bounded rationality, prior policies over actions can be part of the optimization process, which results in specialization and abstraction. In constrast, active inference compounds utilities and priors into a single desired prior, which is fixed and does not allow separately optimizing utility and action priors.

7 Conclusion

In this work, we have presented an information-theoretic framework to study systems of decision-making units with limited information-processing capabilities. It is based on an overreaching free energy optimization principle that, on the one hand, allows computing the optimal performances of explicit architectures and, on the other hand, produces optimal partitions of the involved choice spaces into regions of specialization. In order to combine a given set of bounded rational agents, the full decision-making process is split into multiple decision steps by introducing intermediate decision variables, and then a given set of agents is distributed among these variables. We have argued that this leads to two types of agents, nonoperational units that distribute the work among subordinates and operational units that are doing the actual work in the sense of choosing a particular action that either serves as an input for another agent in the system or represents the final decision of the full process. This "vertical" specialization is enhanced by optimizing over the agents' prior policies, which leads to an optimal soft partitioning of the underlying choice space of each step in the system, resulting in a "horizontal" specialization as well.

In order to illustrate the proposed framework, we have simulated and analyzed the performances under a number of different resource constraints and tasks for all possible three-step architectures whose information flow starts by observing a given world state and ends with the selection of a final decision. Although the relative architecture performances depend crucially on the explict information-processing constraints, the overall best-performing architectures tend to be hierarchical systems of nonoperational manager units at higher hierarchical levels and operational worker units at the lowest level.

Our approach is based on earlier work on information-theoretic bounded rationality (Ortega & Braun, 2011, 2013; Genewein & Braun, 2013; Genewein

et al., 2015). In particular, the N-step decision-making systems introduced in section 3 generalize the two-step processes studied in Genewein and Braun (2013) and Genewein et al. (2015). According to Simon (1979), there are three different bounded rational procedures that can transform intractable into tractable decision problems: (1) Looking for satisfactory choices instead of optimal ones, (2) replacing global goals with tangible subgoals, and (3) dividing the decision-making task among many specialists. From this point of view, the decision-making process of a single agent, given by the one-step case of information-theoretic bounded rationality (Ortega & Braun, 2011, 2013) described in section 2, corresponds to the first procedure, while the bounded rational multistep and multiagent decision-making processes introduced in sections 3 and 4 can be attributed to the second and third procedures.

The main advantage of a purely information-theoretic treatment is its universality. To our knowledge, this work is the first systematic theory-guided approach to the organization of agents with limited resources in the generality of information theory. In other approaches, more specific methods, tailored to each particular focus of study, are used instead. In particular, bounded rationality has usually a very specific meaning, often being implemented by simply restricting the cardinality of the choice space. For example, in management theory, the well-known results by Graicunas from the 1930s (Graicunas, 1933) suggest that managers must have a limited span of control in order to be efficient. By counting the number of possible relationships between managers and their subordinates, he concludes that there is an explicit upper bound of five or six subordinates. Of course, there are many cases of successful companies today that disagree with Graicunas's claim; for example, Apple's CEO has 17 managers reporting directly to him. However, current management experts think that the optimal number is somewhere between 5 and 12. The idea of restricting the cardinality of the space of decision making is also studied for operational agents. For example, Camacho and Persky (1988) explore the hierarchical organization of specialized producers with a focus on production. Although their treatment is more abstract and more general than many preceding studies, their take on bounded rationality is very explicit and based on the assumption that the number of elementary parts that form a product, as well as the number of possibilities of each part, are larger than a single individual can handle. Similarly, in most game-theoretic approaches that are based on automaton theory (Neyman, 1985; Abreu & Rubinstein, 1988; Hernández & Solan, 2016), the boundedness of an agent's rationality is expressed by a bound on the number of states of the automaton. Most of these non-information-theoretic treatments consider cases when there is a hard upper bound on the number of options, but they usually lack a probabilistic description of the behavior in cases when the number of options is larger than the given bound.

The work by Geanakoplos and Milgrom (1991) uses "information" to describe the limited attention of managers in a firm. But here, we use the

term more informally, and not in the classical information-theoretical sense. However, one of their results suggests that "firms with more prior information about parameters . . . will employ less able managers, or give their managers wider spans of control" (Geanakoplos & Milgrom, 1991, p. 207). This observation is in line with information-theoretic bounded rationality, since by optimizing over priors in the free energy principle, the required processing information is decreased compared to the case of nonoptimal priors, so that less able agents can perform a given task or, similarly, an agent with a higher information bound can have a larger choice space.

In neuroscience, the variational Bayes approach explained in section 6.3 has been proposed as a theoretical framework to understand brain function in terms of active inference (Friston 2009, 2010; Friston, Levin et al., 2015; Friston, Rigoli et al., 2015; Friston, Lin et al., 2017; Friston, Parr et al., 2017), where perception is modeled as variational Bayesian inference over hidden causes of observations. There, a processing node (usually a neuron) is limited in the sense that it can only linearly combine a set of input signals into a single output signal. Decision making is modeled by approximating Bayes' rule in terms of these basic operations and then tuning the weights of the resulting linear transformations in order to optimize the free energy (see equation 6.2). Hence, there, the free energy serves as a tool to computationally simplify Bayesian inference on the neuronal level, whereas our free energy principle is a tool to computationally trade off expected utility and processing costs, providing an abstract probabilistic description of the best possible choices when the information-processing capability is limited.

In the general setting of approximate Bayesian inference, there are many interesting algorithms and belief update schemes—for example, belief propagation in terms of message passing on factor graphs (see Yedidia, Freeman, & Weiss, 2005). These algorithms make use of the notion of the Markov boundary (minimal Markov blanket) of a node X, which consists of the nodes that share a common factor with X (so-called neighbors). Conditioned on its Markov boundary, a given random variable is independent of all other variables in the system, which allows approximating marginal probabilities in terms of local messages between neighbors. These approximations are generally exact only on tree-like factor graphs without loops (Mézard & Montanari, 2009, theorem 14.1). This raises the interesting question of whether such algorithms could also be applied to our setting. First, it should be noted that variational Bayesian inference constitutes only a subclass of problems that can be expressed by utility optimization with information constraints. In this subclass, all random variables have to appear either in utility functions (they have to be given as log likelihoods) or in marginal distributions that are kept fixed—see, for example, the definition of the utility in the inference example above where $U(a, x_1, x_2, y) = \log p(y|a, x_1, x_2, S)$ compared to the utility functions of the form $U(\omega, a)$ used throughout the letter that leave all intermediate random variables $X_1, \ldots, X_{N-1}$ unspecified. Second, while it may be possible to

exploit the notion of Markov blankets by recursively computing free energies among the nodes in a similar fashion to message passing, there can also be contributions from outside the Markov boundary—for example, when the action node has to take an expectation over possible world states that lie outside the Markov boundary. Finally, it may be interesting to study whether message-passing algorithms can be extended to deal with our general problem setting and at least to approximately generate the same kind of solutions as Blahut-Arimoto, even though in general, we do not have tree-structured graphs.

There are plenty of other possible extensions of the basic framework introduced in this work. Marschak and Reichelstein (1998) study multiagent systems in terms of communication cost minimization, while ignoring the actual decision-making process. One could combine our model with the information bottleneck method (Tishby, Pereira, & Bialek, 1999) and explicitly include communication costs in order to study more general agent architectures—in particular, systems with nondirected information flow. Moreover, we have seen in our simulations that specialization of operational agents is an important feature shared among all of the best-performing architectures. In the biological literature, specialization is often paired with modularity. For example Kashtan and Alon (2005) and Wagner et al. (2007) show that modular networks are an evolutionary consequence of modularly varying goals. Similarly, it would be interesting to study the effects of changing environments on specialization, abstraction, and optimal network architectures of systems of bounded rational agents.

Appendix: Proof of Equation 3.5

The free energy functional $\mathcal{F}$ that is optimized in the free energy principle, equation 3.4, is given by

$$\mathcal{F}[P_1, \ldots, P_N] = \sum_x p(x)\, F_{0,\mathrm{loc}}(x),$$

where $x := (x_0, \ldots, x_N)$, and for all $k \in \{0, \ldots, n\}$

$$p(x) = \rho(x_0)\, P_1\big(x_1 \big| x^1_{sel}, x^1_{in}\big) \cdots P_N\big(x_N \big| x^N_{sel}, x^N_{in}\big),$$

$$\mathcal{F}_{k,\mathrm{loc}}(x) = U(x_0, x_N) - \sum_{i>k} \frac{1}{\beta_i} \log \frac{P_i(x_i | x^i_{sel}, x^i_{in})}{p_i(x_i | x^i_{sel})}.$$

By writing

$$\mathcal{F}_{0,\mathrm{loc}}(x) = \mathcal{F}_{k,\mathrm{loc}}(x) \; - \; \frac{1}{\beta_k} \log \frac{P_k(x_k | x^k_{sel}, x^k_{in})}{p_k(x_k | x^k_{sel})} \; - \; \mathcal{R}_k(x_{<k}),$$

where $x_{<k} := (x_0, \ldots, x_{k-1})$, and

$$\mathcal{R}_k(x_{<k}) := \sum_{i<k} \frac{1}{\beta_i} \log \frac{P_i(x_i|x^i_{sel}, x^i_{in})}{p_i(x_i|x^i_{sel})},$$

we obtain for any $k \in \{0, \ldots, n\}$,

$$\begin{aligned}\mathcal{F}[P_1, \ldots, P_N] = \sum_{x^k_{sel}, x^k_{in}} p(x^k_{sel}, x^k_{in}) \Bigg[& \sum_{x_k} P_k(x_k|x^k_{sel}, x^k_{in}) \sum_{\tilde{x}^c} p(\tilde{x}^c|\tilde{x}) \mathcal{F}_{k,\mathrm{loc}}(x) \\ & - \frac{1}{\beta_k} \mathrm{D}_{\mathrm{KL}}\big(P_k(\,\cdot\,|x^k_{sel}, x^k_{in}) \,\big\|\, p_k(x_k|x^k_{sel})\big) \Bigg] \\ & - \sum_{x_{<k}} p(x_{<k})\, \mathcal{R}_k(x_{<k})\end{aligned}$$

with $\tilde{x} = (x_k, x^k_{sel}, x^k_{in})$ and $\tilde{x}^c := (x_0, \ldots, x_N) \setminus \tilde{x}$. In this form, we can see that optimizing for P_k yields the Boltzmann distribution, equation 3.5, with respect to the effective utility $\mathcal{F}_k(\tilde{x}) = \sum_{\tilde{x}^c} p(\tilde{x}^c|\tilde{x}) \mathcal{F}_{k,\mathrm{loc}}(x)$ as defined in equation 3.6.

Acknowledgments

This study was funded by the European Research Council (ERC-StG-2015-ERC Starting Grant, Project ID: 678082, "BRISC: Bounded Rationality in Sensorimotor Coordination").

References

Abreu, D., & Rubinstein, A. (1988). The structure of Nash equilibrium in repeated games with finite automata. *Econometrica*, *56*, 1259–1281.

Acerbi, L., Vijayakumar, S., & Wolpert, D. M. (2014). On the origins of suboptimality in human probabilistic inference. *PLoS Computational Biology*, *10*(6), 1–23.

Arimoto, S. (1972). An algorithm for computing the capacity of arbitrary discrete memoryless channels. *IEEE Transactions on Information Theory*, *18*, 14–20.

Aumann, R. J. (1997). Rationality and bounded rationality. *Games and Economic Behavior*, *21*(1), 2–14.

Bishop, C. M. (2006). *Pattern recognition and machine learning*. New York: Springer-Verlag.

Blahut, R. E. (1972). Computation of channel capacity and rate-distortion functions. *IEEE Transactions on Information Theory*, *18*(4), 460–473.

Burns, E., Ruml, W., & Do, M. B. (2013). Heuristic search when time matters. *Journal of Artificial Intelligence Research*, *47*(1), 697–740.

Camacho, A., & Persky, J. J. (1988). The internal organization of complex teams: Bounded rationality and the logic of hierarchies. *Journal of Economic Behavior and Organization*, *9*(4), 367–380.

Chehouri, A., Younes, R., Perron, J., & Ilinca, A. (2016). A constraint-handling technique for genetic algorithms using a violation factor. *Journal of Computer Sciences*, *12*(7), 350–362.

Csiszár, I., & Tusnády, G. (1984). Information geometry and alternating minimization procedures. *Statistics and Decisions*, *1*, (Suppl.), 205–237.

DeCanio, S. J., & Watkins, W. E. (1998). Information processing and organizational structure. *Journal of Economic Behavior and Organization*, *36*(3), 275–294.

Friedman, N., & Koller, D. (2003). Being Bayesian about network structure: A Bayesian approach to structure discovery in bayesian networks. *Machine Learning*, *50*(1), 95–125.

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

Friston, K. J. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, *11*, 127–138.

Friston, K. J., Levin, M., Sengupta, B., & Pezzulo, G. (2015). Knowing one's place: A free-energy approach to pattern regulation. *Journal of the Royal Society Interface*, *12*, 20141383.

Friston, K. J., Lin, M., Frith, C., & Pezzulo, G. (2017). Active inference, curiosity and insight. *Neural Computation*, *29*(10), 2633–2683.

Friston, K. J., Parr, T., & de Vries, B. (2017). The graphical brain: Belief propagation and active inference. *Network Neuroscience*, *1*(4), 381–414.

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

Geanakoplos, J., & Milgrom, P. (1991). A theory of hierarchies based on limited managerial attention. *Journal of the Japanese and International Economies*, *5*(3), 205–225.

Genewein, T., & Braun, D. A. (2013). Abstraction in decision-makers with limited information processing capabilities. NIPS Workshop on Planning with Information Constraints. arXiv:1312.4353

Genewein, T., Leibfried, F., Grau-Moya, J., & Braun, D. A. (2015). Bounded rationality, abstraction, and hierarchical decision-making: An information-theoretic optimality principle. *Frontiers in Robotics and AI*, *2*, 27.

Gershman, S. J., Horvitz, E. J., & Tenenbaum, J. B. (2015). Computational rationality: A converging paradigm for intelligence in brains, minds, and machines. *Science*, *349*(6245), 273–278.

Gigerenzer, G., & Selten, R. (2001). *Bounded rationality: The adaptive toolbox*. Cambridge, MA: MIT Press.

Graicunas, V. A. (1933). Relationship in organization. *Bulletin of the International Management Institute*, *7*, 39–42.

Hernández, P., & Solan, E. (2016). Bounded computational capacity equilibrium. *Journal of Economic Theory*, *163*, 342–364.

Howes, A., Lewis, R. L., & Vera, A. (2009). Rational adaptation under task and processing constraints: Implications for testing theories of cognition and action. *Psychological Review*, *116*(4), 717–751.

Jain, P., & Kar, P. (2017). *Non-convex optimization for machine learning*. Boston: Now Publishers.

Jones, B. D. (2003). Bounded rationality and political science: Lessons from public administration and public policy. *Journal of Public Administration Research and Theory*, *13*(4), 395–412.

Kaelbling, L. P., Littman, M. L., & Cassandra, A. R. (1998). Planning and acting in partially observable stochastic domains. *Artificial Intelligence*, *101*(1), 99–134.

Kappen, H. J., Gómez, V., & Opper, M. (2012). Optimal control as a graphical model inference problem. *Machine Learning*, *87*(2), 159–182.

Kashtan, N., & Alon, U. (2005). Spontaneous evolution of modularity and network motifs. *Proceedings of the National Academy of Sciences*, *102*(39), 13773–13778.

Knight, F. (1921). *Risk, uncertainty and profit*. Cambridge, MA: Houghton Mifflin.

Lewis, R. L., Howes, A., & Singh, S. (2014). Computational rationality: Linking mechanism and behavior through bounded utility maximization. *Topics in Cognitive Science*, *6*(2), 279–311.

Lipman, B. L. (1995). Information processing and bounded rationality: A survey. *Canadian Journal of Economics/Revue canadienne d'Economique*, *28*, 42–67.

Marschak, T., & Reichelstein, S. (1998). Network mechanisms, informational efficiency, and hierarchies. *Journal of Economic Theory*, *79*(1), 106–141.

Mattsson, L.-G., & Weibull, J. W. (2002). Probabilistic choice and procedurally bounded rationality. *Games and Economic Behavior*, *41*(1), 61–78.

McKelvey, R. D., & Palfrey, T. R. (1995). Quantal response equilibria for normal form games. *Games and Economic Behavior*, *10*(1), 6–38.

Mézard, M., & Montanari, A. (2009). *Information, physics, and computation*. Oxford: Oxford University Press.

Neyman, A. (1985). Bounded complexity justifies cooperation in the finitely repeated prisoners' dilemma. *Economics Letters*, *19*(3), 227–229.

Ochs, J. (1995). Games with unique, mixed strategy equilibria: An experimental study. *Games and Economic Behavior*, *10*(1), 202–217.

Ortega, P. A., & Braun, D. A. (2010). A conversion between utility and information. In *Proceedings of the Third Conference on Artificial General Intelligence* (pp. 115–120). www.istcs.org/

Ortega, P. A., & Braun, D. A. (2011). *Information, utility and Bounded Rationality*. *Lecture Notes in Computer Science: Vol. 6830. Artificial General Intelligence* (pp. 269–274). Berlin: Springer-Verlag.

Ortega, P. A., & Braun, D. A. (2013). Thermodynamics as a theory of decision-making with information-processing costs. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, *469*, 20120683.

Ortega, P. A., & Braun, D. A. (2015). What is epistemic value in free energy models of learning and acting? A bounded rationality perspective. *Cognitive Neuroscience*, *6*(4), 215–216. PMID:25990838.

Radner, R. (1993). The organization of decentralized information processing. *Econometrica*, *61*(5), 1109–1146.

Russell, S. J., & Subramanian, D. (1995). Provably bounded-optimal agents. *Journal of Artificial Intelligence Research*, *2*(1), 575–609.

Shannon, C. E. (1959). Coding theorems for a discrete source with a fidelity criterion. *IRE International Convention Record*, *7*, 142–163.

Simon, H. A. (1943). *A theory of administrative decision*. PhD diss., University of Chicago.

Simon, H. A. (1955). A behavioral model of rational choice. *Quarterly Journal of Economics*, *69*(1), 99–118.

Simon, H. A. (1979). Rational decision making in business organizations. *American Economic Review*, *69*(4), 493–513.

Sims, C. A. (2003). Implications of rational inattention. *Journal of Monetary Economics*, *50*(3), 665–690.

Spiegler, R. (2011). *Bounded rationality and industrial organization*. Oxford: Oxford University Press.

Still, S. (2009). Information-theoretic approach to interactive learning. *Europhysics Letters*, *85*(2), 28005.

Tishby, N., Pereira, F. C., & Bialek, W. (1999). The information bottleneck method. In *Proceedings of the 37th Allerton Conference on Communication, Control, and Computing* (pp. 368–377). Monticello, IL: University of Illinois.

Tishby, N., & Polani, D. (2011). Information theory of decisions and actions. In V. Cutsuidis, A. Hussain, & J. Taylor (Eds.), *Perception-action cycle: Models, architectures, and hardware*. Berlin: Springer.

Todorov, E. (2009). Efficient computation of optimal actions. *Proceedings of the National Academy of Sciences*, *106*(28), 11478–11483.

von Neumann, J., & Morgenstern, O. (1944). *Theory of games and economic behavior*. Princeton: Princeton University Press.

Vul, E., Goodman, N., Griffiths, T. L., & Tenenbaum, J. B. (2014). One and done? Optimal decisions from very few samples. *Cognitive Science*, *38*(4), 599–637.

Wagner, G. P., Pavlicev, M., & Cheverud, J. M. (2007). The road to modularity. *Nature Reviews Genetics*, *8*, 921.

Wolpert, D. H. (2006). *Information theory: The bridge connecting bounded rational game theory and statistical physics*. Berlin: Springer.

Yedidia, J. S., Freeman, W. T., & Weiss, Y. (2005). Constructing free-energy approximations and generalized belief propagation algorithms. *IEEE Transactions on Information Theory*, *51*(7), 2282–2312.

Received April 30, 2018; accepted September 28, 2018.