



Bounded Rationality, Abstraction, and Hierarchical Decision-Making: An Information-Theoretic Optimality Principle

Tim Genewein^{1,2,3*}, Felix Leibfried^{1,2,3}, Jordi Grau-Moya^{1,2,3} and Daniel Alexander Braun^{1,2}

¹ Max Planck Institute for Intelligent Systems, Tübingen, Germany, ² Max Planck Institute for Biological Cybernetics, Tübingen, Germany, ³ Graduate Training Centre of Neuroscience, Tübingen, Germany

OPEN ACCESS

Edited by:

Joschka Boedecker,
University of Freiburg, Germany

Reviewed by:

Dimitrije Markovic,
Dresden University of Technology,
Germany

Sam Neymotin,
State University of New York
Downstate Medical Center, USA

*Correspondence:

Tim Genewein
tim.genewein@tuebingen.mpg.de

Specialty section:

This article was submitted to
Computational Intelligence,
a section of the
journal *Frontiers in Robotics and AI*

Received: 31 August 2015

Accepted: 23 October 2015

Published: 11 November 2015

Citation:

Genewein T, Leibfried F, Grau-Moya J and Braun DA (2015) Bounded Rationality, Abstraction, and Hierarchical Decision-Making: An Information-Theoretic Optimality Principle. *Front. Robot. AI* 2:27. doi: 10.3389/frobt.2015.00027

Abstraction and hierarchical information processing are hallmarks of human and animal intelligence underlying the unrivaled flexibility of behavior in biological systems. Achieving such flexibility in artificial systems is challenging, even with more and more computational power. Here, we investigate the hypothesis that abstraction and hierarchical information processing might in fact be the consequence of limitations in information-processing power. In particular, we study an information-theoretic framework of bounded rational decision-making that trades off utility maximization against information-processing costs. We apply the basic principle of this framework to perception-action systems with multiple information-processing nodes and derive bounded-optimal solutions. We show how the formation of abstractions and decision-making hierarchies depends on information-processing costs. We illustrate the theoretical ideas with example simulations and conclude by formalizing a mathematically unifying optimization principle that could potentially be extended to more complex systems.

Keywords: information theory, bounded rationality, computational rationality, rate-distortion, decision-making, hierarchical architecture, perception-action system, lossy compression

1. INTRODUCTION

A key characteristic of intelligent systems, both biological and artificial, is the ability to flexibly adapt behavior in order to interact with the environment in a way that is beneficial to the system. In biological systems, the ability to adapt affects the fitness of an organism and becomes key to survival not only of individual organisms but species as a whole. Both in the theoretical study of biological systems and in the design of artificial intelligent systems, the central goal is to understand adaptive behavior formally. A formal framework for tackling the problem of general adaptive systems is decision-theory, where behavior is conceptualized as a series of optimal decisions or actions that a system performs in order to respond to changes to the input of the system. An important idea, originating from the foundations of decision-theory, is the maximum expected utility (MEU) principle (Ramsey, 1931; Von Neumann and Morgenstern, 1944; Savage, 1954). Following MEU, an intelligent system is formalized as a decision-maker that chooses actions in order to maximize the desirability of the expected outcome of the action, where the desirability of an outcome is quantified by a utility function.

A fundamental problem of MEU is that the computation of an optimal action can easily exceed the computational capacity of a system. It is for example in general prohibitive trying to compute

an optimal chess move due to the large number of possibilities. One way to deal with such problems is to study optimal decision-making with information-processing constraints. Following the pioneering work of Simon (1955, 1972) on bounded rationality, decision-making with limited information-processing resources has been studied extensively in psychology (Gigerenzer and Todd, 1999; Camerer, 2003; Gigerenzer and Brighton, 2009), economics (McKelvey and Palfrey, 1995; Rubinstein, 1998; Kahneman, 2003; Parkes and Wellman, 2015), political science (Jones, 2003), industrial organization (Spiegler, 2011), cognitive science (Howes et al., 2009; Janssen et al., 2011), computer science, and artificial intelligence research (Horvitz, 1988; Lipman, 1995; Russell, 1995; Russell and Subramanian, 1995; Russell and Norvig, 2002; Lewis et al., 2014). Conceptually, the approaches differ widely ranging from heuristics (Tversky and Kahneman, 1974; Gigerenzer and Todd, 1999; Gigerenzer and Brighton, 2009; Burns et al., 2013) to approximate statistical inference schemes (Levy et al., 2009; Vul et al., 2009, 2014; Sanborn et al., 2010; Tenenbaum et al., 2011; Fox and Roberts, 2012; Lieder et al., 2012).

In this study, we use an information-theoretic model of bounded rational decision-making (Braun et al., 2011; Ortega and Braun, 2012, 2013; Braun and Ortega, 2014; Ortega and Braun, 2014; Ortega et al., 2014) that has precursors in the economic literature (McKelvey and Palfrey, 1995; Mattsson and Weibull, 2002; Sims, 2003, 2005, 2006, 2010; Wolpert, 2006) and that is closely related to recent advances in the information theory of perception-action systems (Todorov, 2007, 2009; Still, 2009; Friston, 2010; Peters et al., 2010; Tishby and Polani, 2011; Daniel et al., 2012, 2013; Kappen et al., 2012; Rawlik et al., 2012; Rubin et al., 2012; Neymotin et al., 2013; Tkačik and Bialek, 2014; Palmer et al., 2015). The basis of this approach is formalized by a free energy principle that trades off expected utility, and the cost of computation that is required to adapt the system accordingly in order to achieve high utility. Here, we consider an extension of this framework to systems with multiple information-processing nodes and in particular discuss the formation of information-processing hierarchies, where different levels in the hierarchy represent different levels of abstraction. The basic intuition is that information-processing nodes with little computational resources can adapt only a little for different inputs and are therefore forced to treat different inputs in the same or a similar way, that is the system has to abstract (Genewein and Braun, 2013). Importantly, abstractions arising in decision-making hierarchies are a core feature of intelligence (Kemp et al., 2007; Braun et al., 2010a,b; Gershman and Niv, 2010; Tenenbaum et al., 2011) and constitute the basis for flexible behavior.

The paper is structured as follows. In Section 2, we recapitulate the information-theoretic framework for decision-making and show its fundamental connection to a well-known trade-off in information theory (the rate-distortion problem for lossy compression). In Section 3, we show how the extension of the basic trade-off principle leads to a theoretically grounded design principle that describes how perception is shaped by action. In Section 4, we apply the basic trade-off between expected utility and computational cost to a two-level hierarchy and show how this leads to emergent, bounded-optimal hierarchical decision-making systems. In Section 5, we present

a mathematically unifying formulation that provides a starting point for generalizing the principles presented in this paper to more complex architectures.

2. BOUNDED RATIONAL DECISION-MAKING

2.1. A Free Energy Principle for Bounded Rationality

In a decision-making task with context, an actor or agent is presented with a world-state w and is then faced with finding an optimal action a_w^* out of a set of actions \mathcal{A} in order to maximize the utility $U(w, a)$:

$$a_w^* = \arg \max_a U(w, a). \quad (1)$$

If the cardinality of the action-set is large, the search for the single best action can become computationally very costly. For an agent with limited computational resources that has to react within a certain time-limit, the search problem can potentially become infeasible. In contrast, biological agents, such as animals and humans, are constantly confronted with picking an action out of a very large set of possible actions. For instance, when planning a movement trajectory for grasping a certain object with a biological arm with many degrees of freedom, the number of possible trajectories is infinite. Yet, humans are able to quickly find a trajectory that is not necessarily optimal but good enough. The paradigm of picking a good enough solution that is actually computable has been termed *bounded rational* acting (Simon, 1955, 1972; Horvitz, 1988; Horvitz et al., 1989; Horvitz and Zilberstein, 2001). Note that bounded rational policies are in general stochastic and thus expressed as a probability distribution over actions given a world-state $p(a|w)$.

We follow the work of Ortega and Braun (2013), where the authors present a mathematical framework for bounded rational decision-making that takes into account computational limitations. Formally, an agent's initial behavior (or search strategy through action-space) is described by a prior distribution $p_0(a)$. The agent transforms its behavior to a posterior $p(a|w)$ in order to maximize expected utility $\sum_a p(a|w)U(w, a)$ under this posterior policy. The computational cost of this transformation is measured by the KL-divergence between prior and posterior and is upper-bounded in case of a bounded rational actor. Decision-making with limited computational resources can then be formalized with the following constrained optimization problem:

$$\begin{aligned} p^*(a|w) &= \arg \max_{p(a|w)} \sum_a p(a|w)U(w, a) \\ \text{s.t. } D_{\text{KL}}(p(a|w)||p_0(a)) &\leq K. \end{aligned} \quad (2)$$

This principle models bounded rational actors that initially follow a prior policy $p_0(a)$ and then use information about the world-state w to adapt their behavior to $p(a|w)$ in a way that optimally trades off the expected gain in utility against the transformation costs for adapting from $p_0(a)$ to $p(a|w)$. The constrained optimization problem in equation (2) can be rewritten as an unconstrained

variational problem using the method of Lagrange multipliers:

$$p^*(a|w) = \arg \max_{p(a|w)} \underbrace{\sum_a p(a|w)U(w, a)}_{\mathbb{E}_{p(a|w)}[U(w, a)]} - \frac{1}{\beta} \underbrace{\sum_a p(a|w) \log \frac{p(a|w)}{p_0(a)}}_{D_{KL}(p(a|w)||p_0(a))}, \quad (3)$$

where β is known as the *inverse temperature*. The inverse temperature acts as a conversion-factor, translating the amount of information imposed by the transformation (usually measured in nats or bits) into a cost with the same units as the expected utility (utils). The distribution $p^*(a|w)$ that maximizes the variational principle is given by

$$p^*(a|w) = \frac{1}{Z(w)} p_0(a) e^{\beta U(w, a)}, \quad (4)$$

with *partition sum* $Z(w) = \sum_a p_0(a) e^{\beta U(w, a)}$. Evaluating equation (3) with the maximizing distribution $p^*(a|w)$ yields the *free energy difference*

$$\begin{aligned} \Delta F(w) &= \max_{p(a|w)} \mathbb{E}_{p(a|w)}[U(w, a)] - \frac{1}{\beta} D_{KL}(p(a|w)||p_0(a)) \\ &= \frac{1}{\beta} \log Z(w), \end{aligned} \quad (5)$$

which is well known in thermodynamics and quantifies the energy of a system that can be converted to work. $\Delta F(w)$ is composed of the expected utility under the posterior policy $p^*(a|w)$ minus information processing cost that is required for computing the

posterior policy measured as the Kullback-Leibler (KL) divergence between the posterior $p^*(a|w)$ and the prior $p_0(a)$.

The inverse temperature β governs the influence of the transformation cost and thus the boundedness of the actor which determines the maximally allowed deviation of the final behavior $p^*(a|w)$ from the initial behavior $p_0(a)$. A perfectly rational actor that maximizes its utility can be recovered as the limit case $\beta \rightarrow \infty$ where transformation cost is ignored. This case is identical to equation (1) and simply reflects maximum utility action selection, which is the foundation of most modern decision-making frameworks. Note that the optimal policy $p^*(a|w)$ in this case collapses to a delta over the best action $p^*(a|w) = \delta_{aa^*}$. In contrast, $\beta \rightarrow 0$ corresponds to an actor that has infinite transformation cost or no computational resources and thus sticks with its prior policy $p_0(a)$. An illustrative example is given in **Figure 1**.

Interestingly, the free energy principle for bounded rational acting can also be used for inference problems. In particular if the utility is chosen as a log-likelihood function $U(w, a) = \log q(w|a)$ and the inverse temperature β is set to one, Bayes' rule is recovered as the optimal bounded rational solution [by plugging into equation (4)]:

$$p^*(a|w) = \frac{p_0(a)q(w|a)}{\sum_a p_0(a)q(w|a)}.$$

Importantly, the inverse temperature β can also be interpreted in terms of computational or sample complexity (Braun and Ortega, 2014; Ortega and Braun, 2014; Ortega et al., 2014).

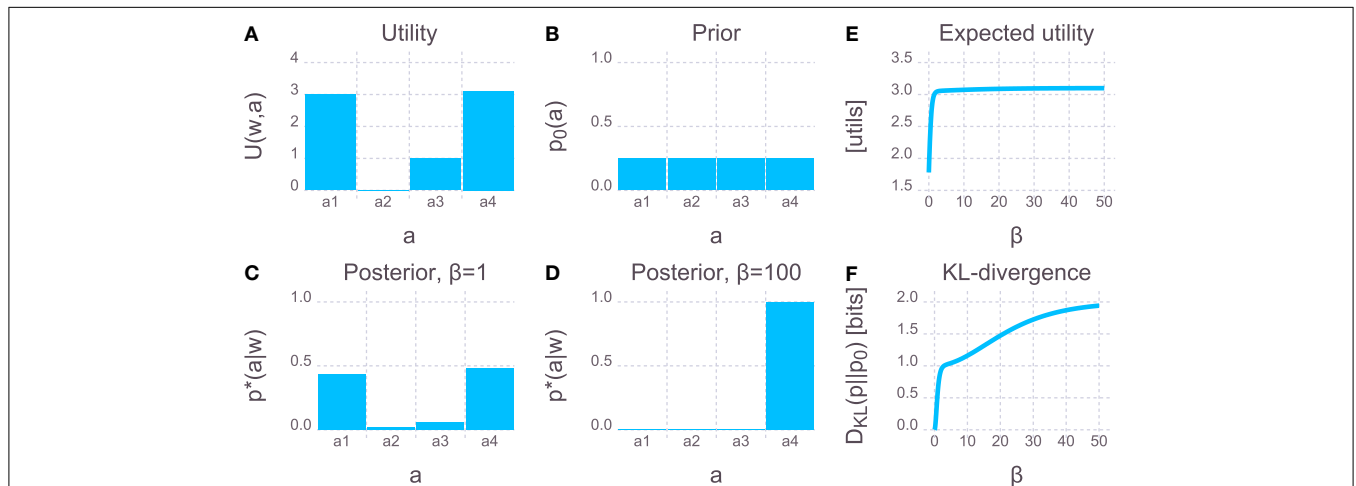


FIGURE 1 | Bounded rationality and the free energy principle. Imagine an actor that has to grasp a particular cup w from a table. There are four options a_1 to a_4 to perform the movement and the utility $U(w, a)$ shown in **(A)** measures the performance of each option. There are two actions a_1 and a_4 that lead to a successful grasp without spilling, and a_4 is minimally better. Action a_3 leads to a successful grasp but spills half the cup, and a_2 represents an unsuccessful grasp. **(B)** Prior distribution over actions $p_0(a)$: no preference for a particular action. **(C)** Posterior $p^*(a|w)$ [equation (4)] for an actor with limited computational capacity. Due to the computational limits, the posterior cannot deviate from the prior arbitrarily far, otherwise the KL-divergence constraint would be violated. The computational resources are mostly spent on increasing the chance of picking one of the two successful options and decreasing the chance of picking a_2 or a_3 . The agent is almost indifferent between the two successful options a_1 and a_4 . **(D)** Posterior for an actor with large computational resources. Even though a_4 is only slightly better than a_1 , the agent is almost unbounded and can deviate a lot from the prior. This solution is already close to the deterministic maximum expected utility solution and incurs a large KL-divergence from prior to posterior. **(E)** Expected utility $\mathbb{E}_{p^*(a|w)}[U(w, a)]$ as a function of the inverse temperature β . Initially, allowing for more computational resources leads to a rapid increase in expected utility. However, this trend quickly flattens out into a regime where small increases in expected utility imply large increases in β . **(F)** KL-divergence $D_{KL}(p||p_0)$ [bits] as a function of the inverse temperature β . In order to achieve an expected utility of $\approx 95\%$ of the maximum utility roughly 1 bit suffices [leading to a posterior similar to **(C)**]. Further increasing the performance by 5% requires twice the computational capacity of 2 bits [leading to a posterior similar to **(D)**]. A bounded rational agent that performs reasonably well could thus be designed at half the cost (in terms of computational capacity) compared to a fully rational maximum expected utility agent. An interactive version of this plot where β can be freely changed is provided in the Supplementary Jupyter Notebook “1-FreeEnergyForBoundedRationalDecisionMaking.”

The basic idea is that in order to make a decision, the bounded rational decision-maker needs to generate a sample from the posterior $p^*(a|w)$. Assuming that the decision-maker can draw samples from the prior $p_0(a)$, samples from the posterior $p^*(a|w)$ can be generated by rejecting any samples from $p_0(a)$ until one sample is accepted as a sample of $p^*(a|w)$ according to the acceptance rule $u \leq \exp(\beta(U(w, a) - T(w)))$, where u is drawn from the uniform distribution over the unit interval $[0;1]$ and $T(w)$ is the aspiration level or acceptance target value with $T(w) \geq \max_a U(w, a)$. This is known as rejection sampling (Neal, 2003; Bishop, 2006). The efficiency of the rejection sampling process depends on how many samples are needed on average from $p_0(a)$ to obtain one sample from $p^*(a|w)$. This average number of samples $\overline{\#Samples}(w)$ is given by the mean of a geometric distribution

$$\begin{aligned} \overline{\#Samples}(w) &= \frac{1}{\sum_a p_0(a) \exp(\beta(U(w, a) - T(w)))} \\ &= \frac{\exp(\beta T(w))}{Z(w)}, \end{aligned} \tag{6}$$

where the partition sum $Z(w)$ is defined as in equation (4). The average number of samples increases exponentially with increasing resource parameter β when $T(w) > \max_a U(w, a)$. It is also noteworthy that the exponential of the Kullback-Leibler divergence provides a lower bound for the required number of samples that is $\overline{\#Samples}(w) \geq \exp(D_{KL}(p^*(a|w)||p_0(a)))$ (see Section 6 in the Supplementary Methods for a derivation). Accordingly, a decision-maker with high β can manage high sampling complexity, whereas a decision-maker with low β can only process a few samples.

2.2. From Free Energy to Rate-Distortion: The Optimal Prior

In the free energy principle in equation (3), the prior $p_0(a)$ is assumed to be given. A very interesting question is which prior distribution $p_0(a)$ maximizes the free energy difference $\Delta F(w)$ for all world-states w on average (assuming that $p(w)$ is given). To formalize this question, we extend the variational principle in equation (3) by taking the expectation over w and the arg max over $p_0(a)$

$$\begin{aligned} &\arg \max_{p_0(a)} \sum_w p(w) \\ &\times \left[\arg \max_{p(a|w)} \mathbb{E}_{p(a|w)}[U(w, a)] - \frac{1}{\beta} D_{KL}(p(a|w)||p_0(a)) \right]. \end{aligned}$$

The inner arg max-operator over $p(a|w)$ and the expectation over w can be swapped because the variation is not over $p(w)$. With the KL-term expanded this leads to

$$\begin{aligned} &\arg \max_{p_0(a), p(a|w)} \sum_{w,a} p(w, a) U(w, a) \\ &- \frac{1}{\beta} \sum_w p(w) \sum_a p(a|w) \log \frac{p(a|w)}{p_0(a)}. \end{aligned}$$

The solution to the arg max over $p_0(a)$ is given by $p_0^*(a) = \sum_w p(w)p(a|w) = p(a)$. [see Section 2.1.1 in Tishby et al. (1999)

or Csiszár and Tuszáný (1984)]. Plugging in the marginal $p(a)$ as the optimal prior $p_0^*(a)$ yields the following variational principle for bounded rational decision-making

$$\begin{aligned} &\arg \max_{p(a|w)} \underbrace{\sum_{w,a} p(w, a) U(w, a)}_{\mathbb{E}_{p(a|w)}[U(w,a)]} - \frac{1}{\beta} \underbrace{\sum_w p(w) D_{KL}(p(a|w)||p(a))}_{I(W;A)} \\ &= \arg \max_{p(a|w)} J_{RD}(p(a|w)), \end{aligned} \tag{7}$$

where $I(W; A)$ is the *mutual information* between actions A and world-states W . The mutual information $I(W; A)$ is a measure of the reduction in uncertainty about the action a after having observed w or vice versa since the mutual information is symmetric

$$I(W; A) = H(W) - H(W|A) = H(A) - H(A|W) = I(A; W),$$

where $H(L) = -\sum_l p(l) \log p(l)$ is the Shannon entropy of random variable L .

The exact same variational problem can also be obtained as the Langragian for maximizing expected utility with an upper bound on the mutual information

$$p^*(a|w) = \arg \max_{p(a|w)} \sum_{w,a} p(w, a) U(w, a) \quad \text{s.t. } I(W; A) \leq R \tag{8}$$

or in the dual point of view, as minimizing the mutual information between actions and world-states with a lower bound on the expected utility. Thus, the problem in equation (7) is equivalent to the problem formulation in rate-distortion theory (Shannon, 1948; Cover and Thomas, 1991; Tishby et al., 1999; Yeung, 2008), the information-theoretic framework for lossy compression. It deals with the problem that a stream of information must be transmitted over a channel that does not have sufficient capacity to transmit all incoming information – therefore some of the incoming information must be discarded. In rate-distortion theory, the distortion $d(w, a)$ quantifies the recovery error of the output symbol a with respect to the input symbol w . Distortion corresponds to a negative utility which thus leads to an arg min instead of an arg max and a positive sign for the mutual information term in the optimization problem. In this case, a maximum expected utility decision-maker would minimize the expected distortion which is typically achieved by a one-to-one mapping between w and a , which implies that the compression is not lossy. From this, it becomes obvious why MEU decision-making might be problematic: if the MEU decision-maker requires a rate of information processing that is above channel capacity, it simply cannot be realized with the given system.

The solution that extremizes the variational problem of equation (7) is given by the self-consistent equations [see Tishby et al. (1999)]

$$p^*(a|w) = \frac{1}{Z(w)} p(a) e^{\beta U(w,a)}, \tag{9}$$

$$p(a) = \sum_w p(w) p^*(a|w), \tag{10}$$

with *partition sum* $Z(w) = \sum_a p(a) e^{\beta U(w,a)}$.

In the limit case $\beta \rightarrow \infty$ where transformation costs are ignored, $p^*(a|w) = \delta_{aa^*}$ is the perfectly rational policy for each value of w independent of any of the other policies and $p(a)$ becomes a mixture of these solutions. Importantly, due to the low price of information processing $\frac{1}{\beta}$, high values of the mutual information term in equation (7) will not lead to a penalization, which means that actions a can be very informative about the world-state w . The behavior of an actor with infinite computational resources will thus in general be very world-state-specific.

In the case where $\beta \rightarrow 0$ the mutual information between actions and world-states is minimized to $I(W; A) = 0$, leading to $p^*(a|w) = p(a) \forall w$, the maximal abstraction where all w elicit the same response. Within this limitation, the actor will, however, emit actions that maximize the expected utility $\sum_{w,a} p(w)p(a)U(w, a)$ using the same policy for all world-states.

For values of the rationality parameter β in between these limit cases, that is $0 < \beta < \infty$, the bounded rational actor trades off world-state-specific actions that lead to a higher expected utility for particular world-states (at the cost of an increased information processing rate), against more robust or abstract actions that yield a “good” expected utility for many world-states (which allows for a decreased information processing rate).

Note that the solution for the conditional distribution $p^*(a|w)$ in the rate-distortion problem [equation (9)] is the same as the solution in the free energy case of the previous section [equation (4)], except that the prior $p_0(a)$ is now defined as the marginal distribution $p_0(a) = p(a)$ [see equation (10)]. This particular prior distribution minimizes the average relative entropy between $p(a|w)$ and $p(a)$ which is the mutual information between actions and world-states $I(W; A)$.

An alternative interpretation is that the decision-maker is a channel that transmits information from w to a according to $p(a|w)$. The channel has a limited capacity, which could arise from the agent not having a “brain” that is powerful enough, but a limited channel capacity could also arise from noise that is induced into the channel, i.e., an agent with noisy sensors or actuators. For a large capacity, the transmission is not severely influenced and the best action for a particular world-state can be chosen. For smaller capacities, however, some information must be discarded and robust (or abstract) actions that are “good” under a number of world-states must be chosen. This is possible by lowering β until the required rate $I(W; A)$ does no longer exceed the channel capacity. The notion that a decision-maker can be considered as an information processing channel is not new and goes back to the cybernetics movement (Ashby, 1956; Wiener, 1961). Other recent applications of rate-distortion theory to decision-making problems can be found for example in Sims (2003, 2006) and Tishby and Polani (2011).

2.3. Computing the Self-Consistent Solution

The self-consistent solutions that maximize the variational principle in equation (7) can be computed by starting with an initial distribution $p_{\text{init}}(a)$ and then iterating equations (9) and (10) in an alternating fashion. This procedure is well known in the rate-distortion framework as a Blahut-Arimoto-type algorithm (Arimoto, 1972; Blahut, 1972; Yeung, 2008). The iteration is

guaranteed to converge to a unique maximum [see Section 2.1.1 in Tishby et al. (1999) and Csizsár and Tusnády (1984) and Cover and Thomas (1991)]. Note that $p_{\text{init}}(a)$ has to have the same support as $p(a)$. Implemented in a straightforward manner, the Blahut-Arimoto iterations can become computationally costly since the iterations involve evaluating the utility function for every action-world-state-pair (w, a) and computing the normalization constant $Z(w)$. In case of continuous-valued random variables, closed-form analytic solutions exist only for special cases. Extending the sampling approach presented at the end of Section 2.1 could be one potential alleviation. A proof-of-concept implementation of the extended sampling scheme is provided in the Supplementary Jupyter Notebook “S1-SampleBasedBlahutArimoto.”

2.4. Emergence of Abstractions

The rate-distortion objective for decision-making [equation (7)] penalizes high information processing demand measured in terms of the mutual information between actions and world-states $I(W; A)$. A large mutual information arises when actions are very informative about the world-state which is the case when a particular action is mostly chosen under a particular world-state and is rarely chosen otherwise. Policies $p(a|w)$ with many world-state-specific actions are thus more demanding in terms of informational cost and might not be affordable by an agent with limited computational capacity. In order to keep informational costs low while at the same time optimizing expected utility, actions that yield a “good” expected utility for many different world-states must be favored. This leads to abstractions in the sense that the agent does not discriminate between different world-states out of a subset of all world-states, but rather responds with the same policy for the entire subset. Importantly, these abstractions are driven by the agent-environment structure encoded through the utility function $U(w, a)$. Limits in computational resources thus lead to abstractions where different world-states are treated as if they were the same.

To illustrate the influence of different degrees of computational limits and the resulting emergence of abstractions we constructed the following example. The goal is to design a recommender system that observes an item bought w and then recommends another item a . In this example the system can either recommend another concrete item or the best-selling item of a certain category or the best-selling item of a super-category which subsumes several categories (see Table 1). An illustration of the example is shown in Figure 2A. The possible items bought are shown on the x-axis and possible recommendations are shown on the y-axis. The super-categories and categories as well as the corresponding bought items can be seen in Table 1 where each bought item also indicates the corresponding concrete item that scores highest when recommended.

The utility of each (w, a) -pair is color-coded in blue in Figure 2A. For each possible world-state there is one concrete item that can be recommended that will (deterministically) yield the highest possible utility of 3 utils. Further, each bought item belongs to a category and recommending the best-selling item of the corresponding category leads to a utility of 2.2 utils. Finally, recommending the best-selling item of the corresponding super-category yields a utility of 1.6 utils. For each world-state there is

TABLE 1 | Recommender system example.

Super-category	Category	Bought item	Best recommended item	
Electric devices and electronics	Computers	Laptop	Laptop sleeve	
		Monitor	Monitor cable	
		Game pad	Video game	
	Small appliances	Coffee machine	Coffee capsules	
		Vacuum cleaner	Vacuum cleaner bags	
		Electric toothbrush	Brush heads	
Food and cooking	Fruit	Grapes	Cheese	
		Strawberries	Cream	
		Limes	Cane sugar	
	Baking	Pancake mix	Maple syrup	
		Baking soda	Vinegar	
		Baker's yeast	Flour	
		Muffin cups	Flour and chocolate chips	

The system observes an item bought w and can then recommend another item to buy a . For each bought item w , there is one other concrete item a that yields the maximum utility when recommended (indicated in the last column of the table). Additionally, each bought item belongs to a category and a less specific super-category. Recommending the best-selling item of the corresponding category or super-category yields sub-optimal but non-zero utility values. A depiction of the utility function $U(w, a)$ is shown in Figure 2A.

one specific action that leads to the highest possible utility but zero utility for all other world-states. At the same time there exist more abstract actions that are sub-optimal but still “good” for a set of world-states. See the legend of Figure 2 for more details on the example.

Figure 2B shows the result $p^*(a|w)$ obtained through Blahut-Arimoto iterations [equations (9) and (10)] for $\beta = 1.3$. For each world-state (on the x-axis) the probability over all actions (y-axis) corresponds to one column in the plot and is color-coded in red. For this particular value of β the agent cannot afford to pick the specific actions for most of the world-states (except for the last three world-states) in order to stay within the limit on the maximum allowed rate. Rather, the agent recommends the best-selling items of the corresponding category which allows for a lower rate by having identical policies (i.e., columns in the plot) for sets of world-states. The optimal policies thus lead to abstractions, where several different world-states elicit identical responses of the agent. Importantly, the abstractions are not induced because some stimuli are more similar than others under some utility-free measure and they are also not the result of a *post hoc* aggregation or clustering scheme. Rather, the abstractions are shaped by the utility function and appear as a consequence of bounded rational decision-making in the given task.

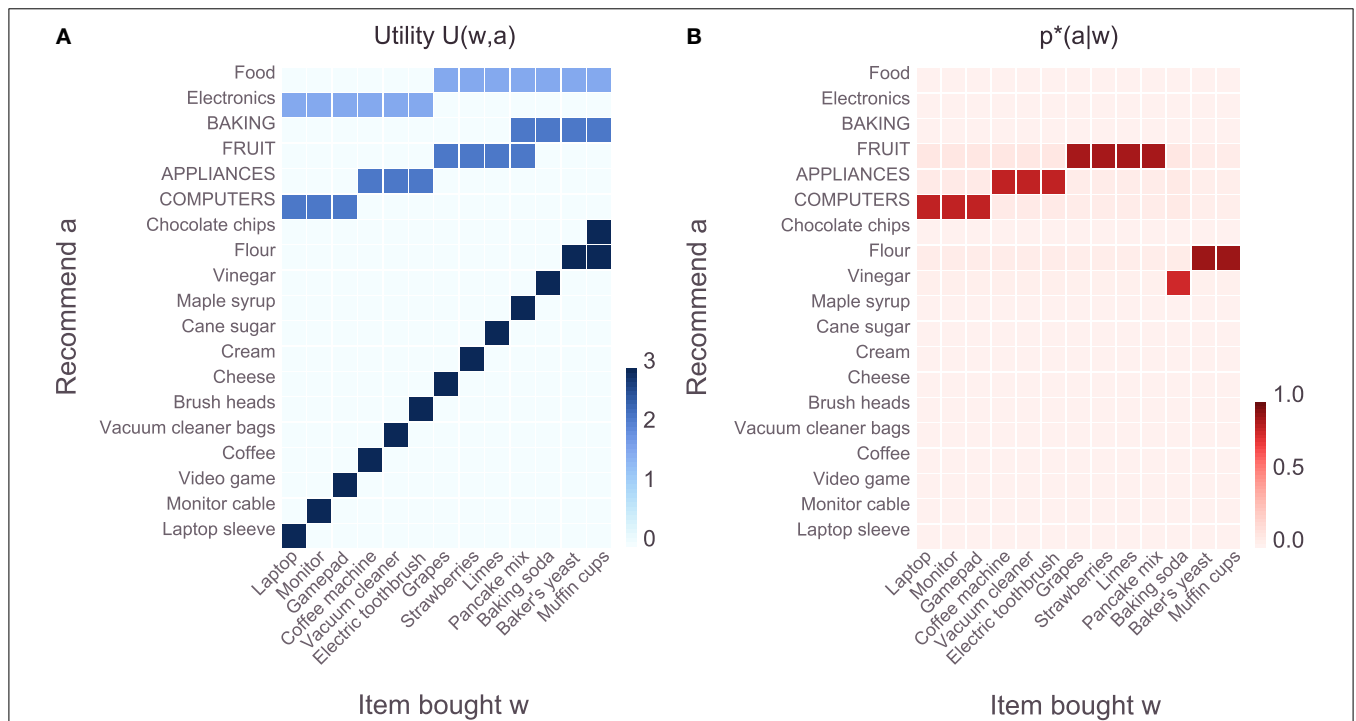


FIGURE 2 | Task setup and solution $p^*(a|w)$ for $\beta = 1.3$. (A) Utility function $U(w, a)$ for the recommender system task. The recommender system observes an item w bought by a customer and recommends another item a to buy to the customer. For each item bought there is another concrete item that has a high chance of being bought by the customer. Therefore, recommending the correct concrete item leads to the maximum utility of 3. However, each item also belongs to a category (indicated by capital letters) and recommending the best-selling item of the corresponding category leads to a utility of 2.2. Finally, each item also belongs to a super-category (either “food” or “electronics”) and recommending the best-selling item of the corresponding super-category leads to a utility of 1.6. There is one item (muffin cups) where two concrete items can be recommended and both yield maximum utility. Additionally there is one item (pancake mix) where the recommendation of the best-selling item of both categories “fruit” and “baking” yields the same utility. (B) Solution $p^*(a|w)$ [equation (9)] for $\beta = 1.3$. Due to the low β , the computational resources of the recommender system are quite limited and it cannot recommend the highest scoring items (except in the last three columns). Instead, it saves computational effort by applying the same policy to multiple items.

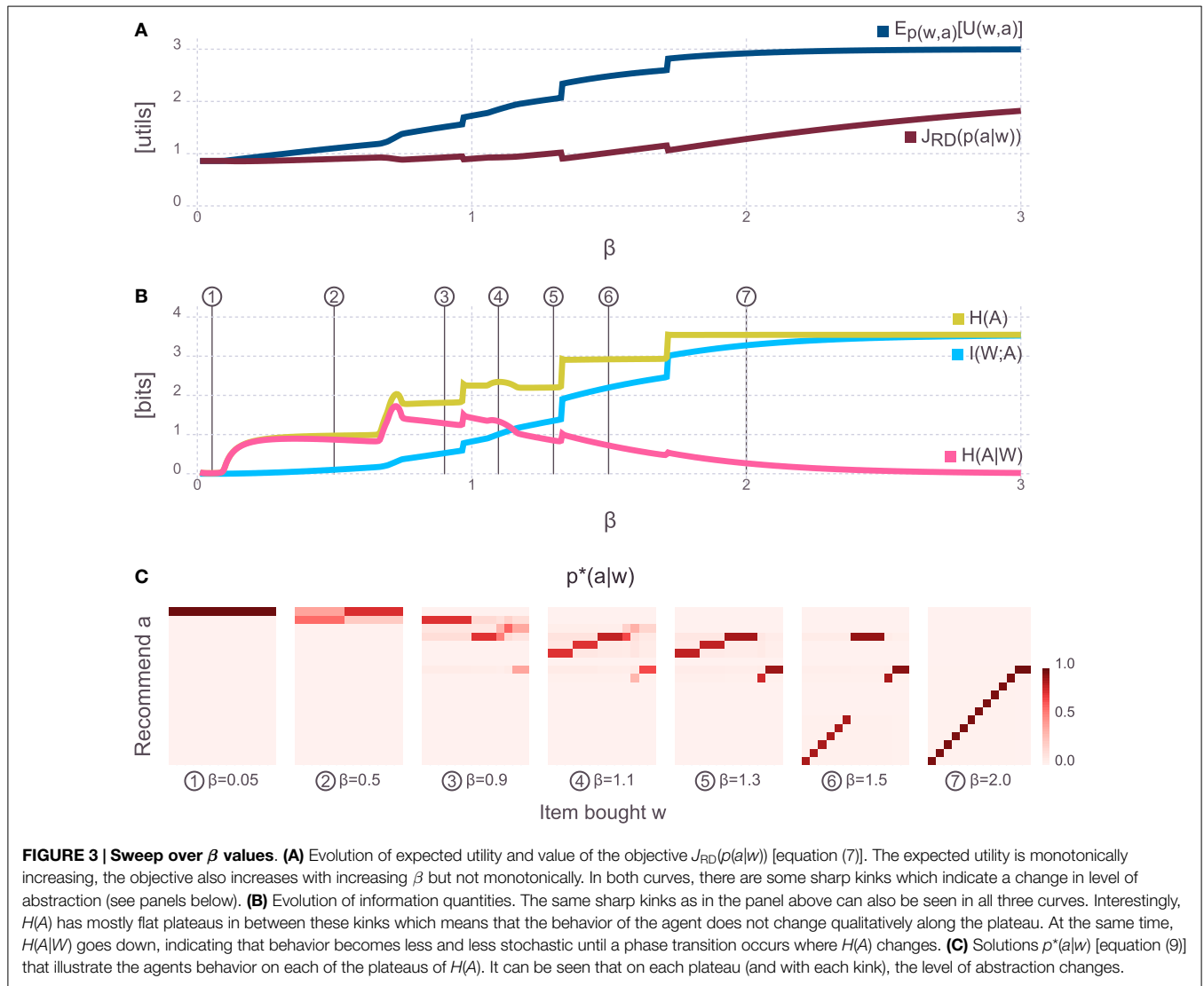


Figure 3A shows the expected utility $E_{p(w,a)}[U(w, a)]$ and the rate-distortion objective $J_{RD}(p(a|w))$ as a function of the inverse temperature β . The plot shows that by increasing β the expected utility increases monotonically, whereas the objective $J_{RD}(p(a|w))$ also shows a trend to increase but not monotonically. Interestingly, there are a few sharp transitions at the same points in both curves. The same steep transitions are also found in **Figure 3B**, which shows the mutual information and its decomposition into the entropic terms $I(W, A) = H(A) - H(A|W)$ as a function of β . The line corresponding to the entropy over actions $H(A)$ shows flat plateaus in between these phase transitions. **Figure 3C** illustrates solutions $p^*(a|w)$ for β values corresponding to points on each of the plateaus (labels for bought and recommended items have been omitted for visual compactness but are identical to the plot in **Figure 2B**). Surprisingly, most of the solutions correspond to different levels of abstraction – from fully abstract for $\beta \rightarrow 0$, then going through several levels of abstraction and getting more and more specific up to the case $\beta \rightarrow \infty$ where the conditional entropy $H(A|W)$ goes to zero implying that the

conditionals $p^*(a|w)$ become deterministic and identical to the maximum expected utility solutions. Within a plateau of $H(A)$, the entropy over actions does not change but the conditional entropy $H(A|W)$ tends to decrease with increasing β . This means that qualitatively the behavior along a plateau does not change in the sense that across all world-states the same subset of actions is used. However, the stochasticity within this subset of actions decreases with increasing β (until at some point a phase-transition occurs). Changing the temperature leads to a natural emergence of different levels of abstraction – levels that emerge from the agent-environment interaction structure described by the utility function. Each level of abstraction corresponds to one plateau in $H(A)$.

In general, abstractions are formed by reducing the information content of an entity until it only contains relevant information. For a discrete random variable $w \in \mathcal{W}$, this translates into forming a partitioning over the space \mathcal{W} where “similar” elements are grouped into the same subset of \mathcal{W} and become indistinguishable within the subset. In physics, changing the granularity of a

partitioning to a coarser level is known as *coarse-graining* which reduces the resolution of the space \mathcal{W} in a non-uniform manner. Here, the partitioning emerges in $p^*(a|w)$ as a *soft-partitioning* (see Still and Crutchfield, 2007), where “similar” world-states w get mapped to an action a (or a subset of actions) and essentially become indistinguishable. Readers are encouraged to interactively explore the example in the Supplementary Jupyter Notebook “2-RateDistortionForDecisionMaking.”

In analogy to rate-distortion theory where the rate-distortion function serves as an information-theoretic characterization of a system, one can define the *rate-utility function*

$$U(R) = \max_{p(a|w): I(W;A) \leq R} \mathbb{E}_{p(w,a)}[U(w, a)]. \quad (11)$$

where the expected utility is a function of the information processing rate $I(W; A)$. If the decision-maker is conceptualized as a communication channel between world-states and actions, the rate $I(W; A)$ defines the minimally required capacity of that channel. The rate-utility function thus specifies the minimum required capacity for computing actions given a certain expected utility target, or analogously the maximally achievable expected utility given a certain information processing capacity. The rate-utility curve is obtained by varying the inverse temperature β (corresponding to different values of R) and plotting the expected utility as a function of the rate. The resulting plot is shown in **Figure 4**, where the solid line denotes the rate-utility curve and the shaded region corresponds to systems that are theoretically infeasible and cannot be achieved regardless of the implementation. Systems in the white region are sub-optimal, meaning that they could either achieve the same performance with a lower rate or given their limits on computational capacity they could theoretically achieve higher performance. This curve is interesting for both designing systems as well as characterizing the degree of sub-optimality of given systems.

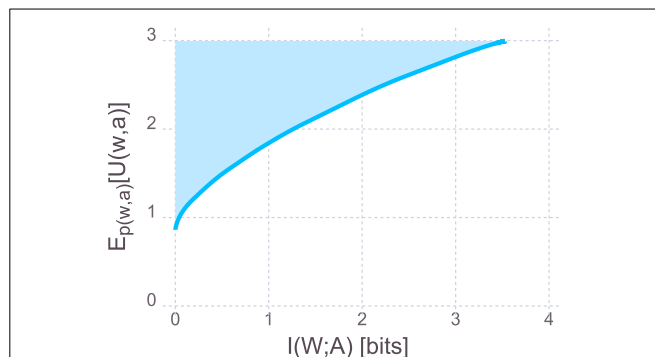


FIGURE 4 | Rate-utility curve. Analogously to the rate-distortion curve in rate-distortion theory, the rate-utility curve shows the minimally required information processing rate to achieve a certain level of expected utility or dually, the maximally achievable expected utility, given a certain rate. Systems that optimally trade-off expected utility against cost of computation lie exactly on the curve. Systems in the shaded region are theoretically impossible and cannot be realized. Systems that lie in the white region of the figure are sub-optimal in the sense that they could achieve a higher expected utility given their computational resources or they could achieve the same expected utility with lower resources.

3. SERIAL INFORMATION-PROCESSING HIERARCHIES

In this section, we apply the rate-distortion principle for decision-making to a serial perception-action system. We design two stages: a perceptual stage $p(x|w)$ that maps world-states w to observations x and an action stage $p(a|x)$ that maps observations x to actions a . Note that the world-state w does not necessarily have to be considered as a latent variable but could in general also be an observation from a previous processing stage. The action stage implements a bounded rational decision-maker (similar to the one presented in the previous section) that optimally trades off expected utility against cost of computation [see equation (7)]. Classically, the perceptual stage might be designed to represent w as faithfully as possible, given the computational limitations of the perceptual stage. Here, we show that trading off expected utility against the cost of information processing on *both* the perceptual and the action stage leads to bounded-optimal perception that does not necessarily represent w as faithfully as possible but rather extracts the most relevant information about w such that the action stage can work most efficiently. As a result, bounded-optimal perception will be tightly coupled to the action stage and will be shaped by the utility function as well as the computational capacity of the action channel.

3.1. Optimal Perception is Shaped by Action

To model a perceptual channel we extend the model from Section 2.2 as follows: The agent is no longer capable of fully observing the state of the world W but using its sensors it is capable to form a percept X as $p(x|w)$ which then allows for adaptation of behavior according to $p(a|x)$. The three random variables for world-state, percept, and action form a serial chain of channels, one channel from world-states to percepts expressed by $p(x|w)$ and another channel from percepts to actions expressed by $p(a|x)$ which implies the following conditional independence

$$p(w, x, a) = p(w)p(x|w)p(a|x),$$

that is also expressed by the graphical model $W \rightarrow X \rightarrow A$. We assume that $p(w)$ is given and the utility function depends on the world-state and the action $U(w, a)$. Note that mathematically, the results are identical for $U(w, x, a)$, but in this paper we consider the utility independent of the internal percept x .

Classically, inference and decision-making are separated – for instance, by first performing Bayesian inference over the state of the world w using the observation x and then choosing an action a according to the maximum expected utility principle. The MEU action-selection principle can be replaced by a bounded rational model for decision-making that takes into account the computational cost of transforming a (optimal) prior behavior $p_0(a)$ to a posterior behavior $p(a|x)$ as shown in Section 2.

Bayesian inference :
$$p(w|x) = \frac{p(w)p(x|w)}{\sum_w p(w)p(x|w)} \quad (12)$$

Bounded rational decision :
$$p^*(a|x) = \frac{1}{Z(x)} p(a) e^{\beta_2 U(x,a)} \quad (13)$$

where $U(x, a) = \sum_w p(w|x)U(w, a)$ is the expectation of the utility under the Bayesian posterior over w given x . Note that the bounded rational decision-maker in equation (13) is identical to the rate-distortion decision-maker introduced in Section 2 that minimizes the trade-off given by equation (7) by implementing equation (9). It includes the MEU solution as a special case for $\beta_2 \rightarrow \infty$. Here, the inverse temperature is denoted by β_2 (instead of β as in the previous section) for notational reasons that ensure consistency with later results of this section.

In equation (12), the choice of the likelihood model $p(x|w)$ remains unspecified and the question is where does it come from? In general, it is chosen by the designer of a system and the choice is often driven by bandwidth or memory constraints. In purely descriptive scenarios, the likelihood model is determined by the sensory setup of a given system and $p(x|w)$ is obtained by fitting it to data of the real system. In the following, we present a particular choice of $p(x|w)$ that is fundamentally grounded on the principle that any transformation of behavior or beliefs is costly (which is identical to the assumption of limited-rate information processing channels) and this cost should be traded off against gains in expected utility. Remarkably, equations (12) and (13) drop out naturally from the principle.

Given the graphical model: $W \rightarrow X \rightarrow A$, we consider an information processing channel between W and X and another one between X and A and introduce different rate-limits on these channels, i.e., the information processing price on the perceptual level $\frac{1}{\beta_1}$ can be different from the price of information processing on the action level $\frac{1}{\beta_2}$. Formally, we set up the following variational problem:

$$\begin{aligned} \arg \max_{p(x|w), p(a|x)} \mathbf{E}_{p(w,x,a)}[U(w, a)] - \frac{1}{\beta_1} I(W; X) - \frac{1}{\beta_2} I(X; A) \\ = \arg \max_{p(x|w), p(a|x)} J_{\text{ser}}(p(x|w), p(a|x)). \end{aligned} \quad (14)$$

Similar to the rate-distortion case, the solution is given by the following set of four self-consistent equations:

$$p^*(x|w) = \frac{1}{Z(w)} p(x) \exp(\beta_1 \Delta F_{\text{ser}(w,x)}) \quad (15)$$

$$p(x) = \sum_w p(w) p^*(x|w) \quad (16)$$

$$p^*(a|x) = \frac{1}{Z(x)} p(a) \exp\left(\beta_2 \sum_w p(w|x) U(w, a)\right) \quad (17)$$

$$p(a) = \sum_{w,x} p(w) p^*(x|w) p^*(a|x), \quad (18)$$

where $Z(w)$ and $Z(x)$ denote the corresponding normalization constants or partition sums. The conditional probability $p(w|x)$ is given by Bayes' rule $p(w|x) = \frac{p(w)p^*(x|w)}{p(x)}$ and $\Delta F_{\text{ser}(w,x)}$ is the free energy difference of the action stage:

$$\Delta F_{\text{ser}(w,x)} := \mathbf{E}_{p^*(a|x)}[U(w, a)] - \frac{1}{\beta_2} D_{\text{KL}}(p^*(a|x) || p(a)), \quad (19)$$

see also equation (5). More details on the derivation of the solution equations can be found in the Supplementary Methods Section 2.

The bounded-optimal perceptual model is given by equation (15). It follows the typical structure of a bounded rational solution consisting of a prior times the exponential of the utility multiplied by the inverse temperature. Compare equation (9) to see that the downstream free-energy trade-off $\Delta F_{\text{ser}(w,x)}$ now plays the role of the utility function for the perceptual model. The distribution $p^*(x|w)$ thus optimizes the downstream free-energy difference in a bounded rational fashion, that is taking into account the computational resources of the perceptual channel. Therefore, the optimal percept becomes tightly coupled to the agent-environment interaction structure as described by the utility function or in other words: the optimal percept is shaped by the embodiment of the agent and, importantly, is not simply a maximally faithful representation of W through X given the limited rate of the perceptual channel. A second interesting observation is that the action stage given by equation (17) turns out to be a bounded rational decision-maker using the Bayesian posterior $p(w|x)$ for inferring the true world-state w given the observation x . This is identical to equation (13) (using the optimal prior $p(a) = \sum_{w,x} p(w) p^*(x|w) p^*(a|x)$), even though the latter was explicitly modeled by first performing Bayesian inference over the world-state w given the percept x [equation (12)] and then performing bounded rational decision-making [equation (13)], whereas the same principle drops out naturally in equation (17) as a result of optimizing equation (14).

3.2. Illustrative Example

In this section, we design a hand-crafted perceptual model $p_\lambda(x|w)$ with precision-parameter λ , that drives a subsequent bounded rational decision-maker that maps an observation x to a distribution over actions $p(a|x)$ in order to maximize expected utility while not exceeding a constraint on the rate of the action channel. The latter is implemented by following equation (13) and setting β_2 according to the limit on the rate $I(X; A)$. We compare the bounded rational actor with hand-crafted perception against a bounded-optimal actor that maximizes equation (14) by implementing the four corresponding self-consistent equations (15)–(18). Importantly, the perceptual model $p^*(x|w)$ of the bounded-optimal actor maximizes the downstream free-energy trade-off of the action stage $\Delta F_{\text{ser}(w,x)}$ which leads to a tight coupling between perception and action that is not present in the hand-crafted model of perception. The action stage is identical in both models and given by equation (17).

We designed the following example where the actor is an animal in a predator-prey scenario. The actor has sensors to detect the size of other animals it encounters. In this simplified scenario, animals can only belong to one of three size-groups and their size correlates with their hearing-abilities:

- Small animals (insects): either 2, 3, or 4 size-units cannot hear very well.
- Medium-sized animals (rodents): either 6, 7, or 8 size-units can hear quite well.
- Large animals (cats of prey): either 10, 11, or 12 size-units can hear quite well.

The actor has a sensor for detecting the size of an animal, however, depending on the capacity of the perceptual channel this sensor will either be more or less noisy. To survive, the actor can

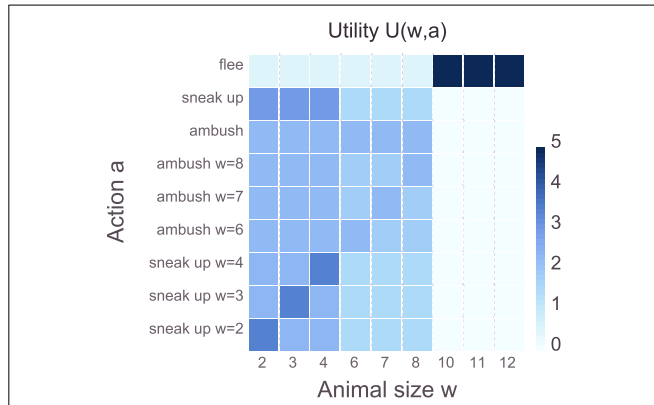


FIGURE 5 | Predator-prey utility function $U(w, a)$. The actor encounters different groups of animals – small, medium-sized, and large – and should hunt small and medium-sized animals for food and flee from large animals. Animals within a group have a similar size w : small animals $\in \{2,3,4\}$ size-units, medium-sized animals $\in \{6,7,8\}$ size-units, and large animals $\in \{10,11,12\}$ size-units. For hunting – denoted by the action a – the actor can either ambush the other animal or alternatively sneak up to the other animal and strike once close enough. In general, sneaking up works well on small animals as they cannot hear very well, whereas medium-sized animals can hear quite well which decreases the chance of a successful sneak-up. The ambush works equally well on both groups of prey. Large animals should never be engaged, but the actor should flee in order to avoid falling prey to them. In addition to the generic hunting patterns that work equally well for all members of a group, for small animals there are specific sneak-up patterns that have an increased chance of success if applied to the correct animal but a decreased chance of success when applied to the wrong animal. For medium-sized animals, there are specific ambush behaviors that have the same chance of success as the generic ambush behavior when applied to the correct animal but a decreased chance of success when applied to the wrong animal. Since the specific ambush behaviors for the medium-sized animals have no advantage over the generic ambush behavior but require more information processing, any bounded rational actor should never choose the specific ambush behaviors. This somewhat artificial construction is meant to illustrate that bounded rational actors do not add unnecessary complexity to their behavior. See main text for more details on the example.

hunt animals from both the small and the medium-sized group for food. On the other hand, it can fall prey to animals of the large group. The actor has three basic actions:

- Ambush: steadily wait for the other animal to get close and then strike.
- Sneak-up: slowly move closer to the animal and then strike.
- Flee: quickly move away from the other animal.

The advantage of the ambush is that it is silent, however, the risk is that the animal might not move toward the position of the ambush – it works equally well on animals from the small and medium-sized group. The sneak-up is not silent but does not rely on the other animal coincidentally getting closer – it works better than the ambush for small-sized animals but the opposite is true for medium-sized animals. If the actor encounters a large animal the only sensible action is to flee in order to avoid falling prey to the large animal. Besides these generic actions, the actor also has a repertoire of more specific hunting patterns – see **Figure 5** which shows the full details of the utility function for the predator-prey scenario. The exact numeric values are found in the Supplementary Jupyter Notebook “3-SerialHierarchy.”

The hand-crafted model of perception is specified by $p_\lambda(x|w)$, where the observed size x corresponds to the actual size of the animal w corrupted by noise. The precision-parameter λ governs the noise-level and thus the quality of the perceptual channel which can be measured with $I(W; X)$. In particular, the observation o is a discretized noisy version of w with precision λ :

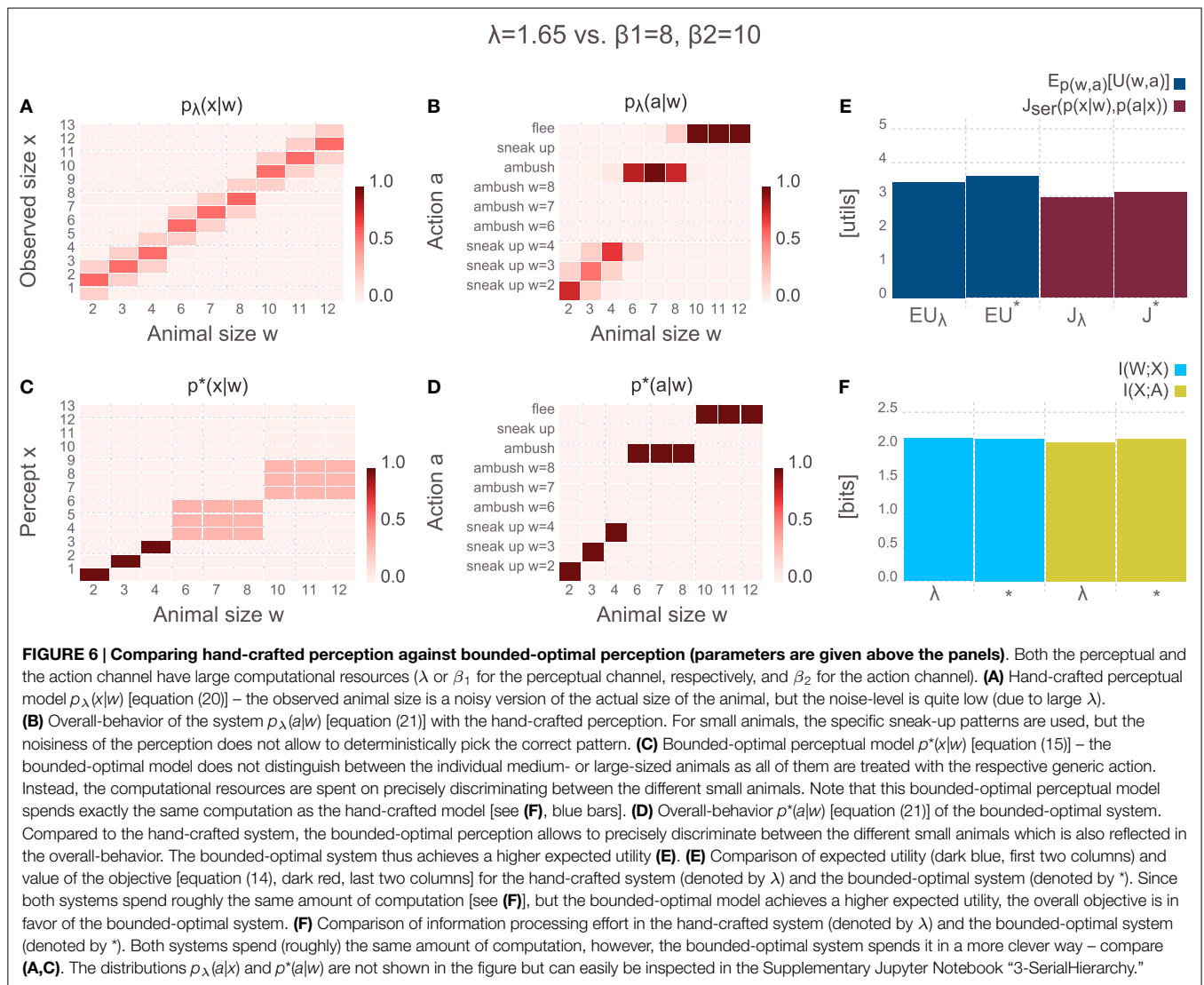
$$x|w, \lambda \sim \text{round}(\mathcal{N}_{\text{trunc}}(w, 1/\lambda)), \quad (20)$$

where the set of world-states is given by all possible animal sizes $w \in \mathcal{W} = \{2,3,4,6,7,8,10,11,12\}$ and the set of possible observations is given by $x \in \mathcal{X} = \{1,2,3, \dots, 11,12,13\}$. To avoid a boundary-bias due to the limited interval \mathcal{X} we reject and re-sample all values of x that would fall outside of \mathcal{X} . For $\lambda \rightarrow \infty$, the perceptual channel is very precise, and there is no uncertainty about the true value of w after observing x . However, such a channel incurs a large computational effort as the mutual information $I(W; X)$ is maximal in this case. If the perceptual channel has a smaller capacity than required to uniquely map each w to an x , the rate must be reduced by lowering the precision λ . Medium precision will mostly lead to within-group confusion whereas low precision will also lead to across-group confusion and corresponds to perceptual channels with a very low rate $I(W; X)$.

The results in **Figure 6** show solutions when having large computational resources on both the perception and action channel. As the figure clearly shows, the hand-crafted model $p_\lambda(x|w)$ looks quite different from the bounded-optimal solution $p^*(x|w)$, even though the rate on the perceptual channel is identical in both cases (given by the mutual information $I(W; X) \approx 2$ bits). The difference is that the bounded-optimal percept spends the two bits mainly on discriminating between specific animals of the small group and on discriminating between medium-sized and large animals. It does not discriminate between specific sizes within the latter two groups. This makes sense, as there is no gain in utility by applying any specific actions to specific animals in the medium- or large-sized group. **Figure 6** also shows the overall-behavior from the point of view of an external observer $p(a|w)$, which is computed as follows

$$p_\lambda(a|w) = \sum_x p_\lambda(x|w)p_\lambda(a|x) \quad \text{and} \\ p^*(a|w) = \sum_x p^*(x|w)p^*(a|x) \quad (21)$$

The overall-behavior in the bounded-optimal case is more deterministic, leading to a higher expected utility in the bounded-optimal case. The distributions $p_\lambda(a|x)$ and $p^*(a|w)$ are not shown in the figure but can easily be inspected in the Supplementary Jupyter Notebook “3-SerialHierarchy.” If the price of information processing on the perceptual channel in the hand-crafted model is the same as in the bounded-optimal model (given by β_1), then the overall objective $J_{\text{ser}}(p(x|w), p(a|x))$ is larger for the bounded-optimal case compared to the hand-crafted case, implying that the bounded optimal actor achieves a better trade-off between expected utility and computational cost. The crucial insight of this example is that the optimal percept depends on the utility function, where in this particular case it does for instance make no sense to waste computational resources on discriminating

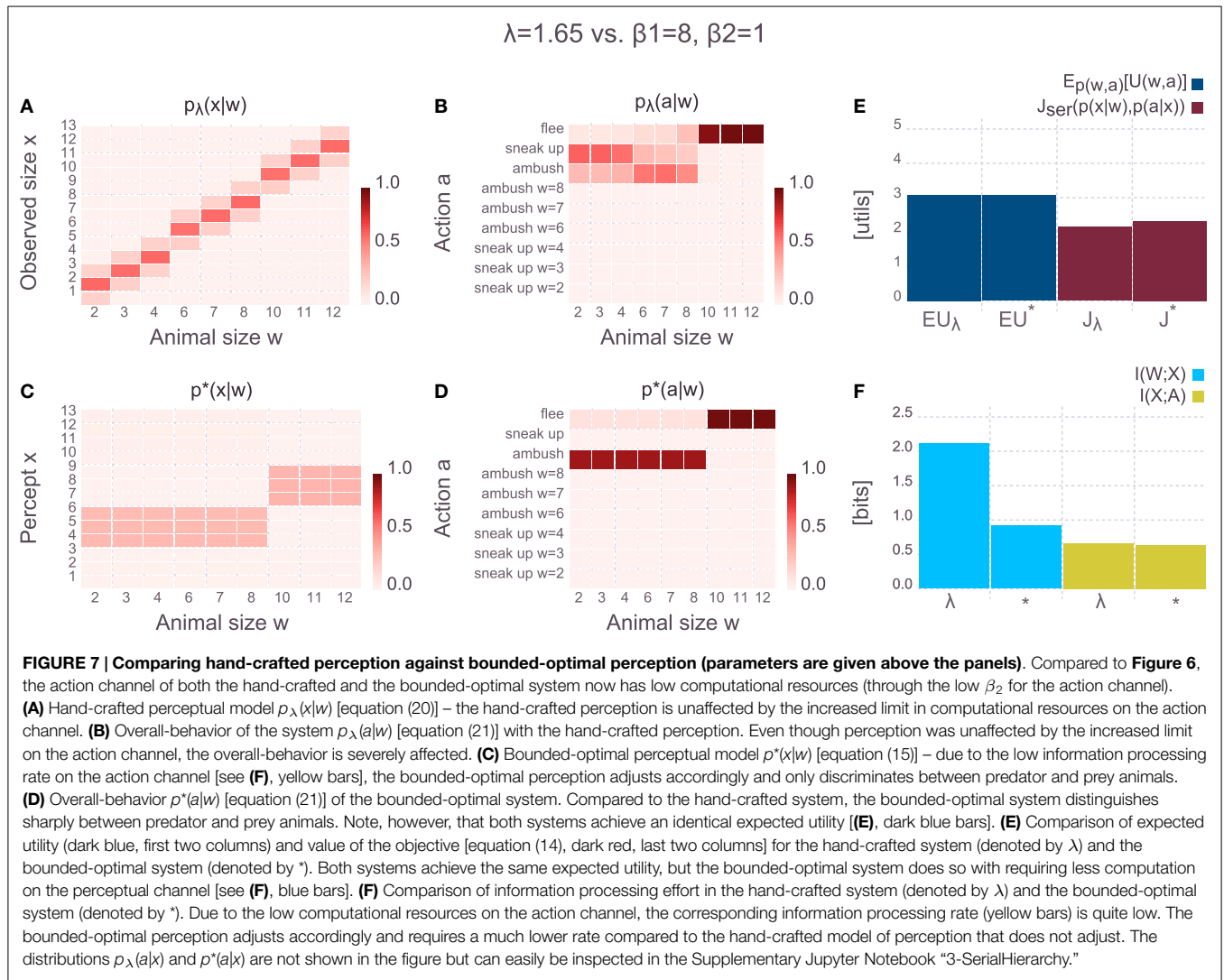


between the specific animals of the large group because the optimal response (flee with certainty) is identical to all of them. In the Supplementary Jupyter Notebook “3-SerialHierarchy” the utility function can easily be switched while keeping all other parameters identical in order to observe how the bounded-optimal percept changes accordingly. Note that the bounded-optimal behavior $p^*(a|w)$ shown in **Figure 6D** yields the highest possible expected utility in this task setup – there is no behavior that would lead to a higher expected utility (though there are other solutions that lead to the same expected utility).

The bounded-optimal percept depends not only on the utility function but also on the behavioral richness of the actor which is governed by the rate on the action channel $I(X; A)$. In **Figure 7** we show the results of the same setup as in **Figure 6** with the only change being the significantly increased price for information processing in the action stage (as specified by $\beta_2 = 1$ bit per util whereas it used to be $\beta_2 = 10$ bits per util in the previous figure). The hand-crafted perceptual model is unaffected by this change of the action stage, but the bounded-optimal model of perception has changed compared to the previous figure and now reflects

the limited behavioral richness. As shown in $p^*(a|w)$ in **Figure 7**, the actor is no longer capable of applying different actions to animals of the small group and animals of the medium-sized group. Accordingly, the bounded-optimal percept does not waste computational resources for discriminating between small and medium-sized animals since the downstream policy is identical for both groups of animals. In terms of expected utility, both the hand-crafted model as well as the bounded-optimal decision-maker score equally at ≈ 3 utils. However, the bounded-optimal model does so by using lower computational resources and thus scoring better on the overall trade-off $J_{ser}(p(x|w), p(a|x))$.

In **Figure 8** we again use large resources on the action channel $\beta_2 = 10$ (as in the first example in **Figure 6**), but now the resources on the perceptual channel are limited by setting $\beta_1 = 1$ (compared to $\beta_1 = 8$ in the first case). Accordingly, the precision of the hand-crafted perceptual model is tuned to $\lambda = 0.4$ (compared to $\lambda = 1.65$ in the first case) such that it has the same rate $I(W; X)$ as the bounded-optimal model. By comparing the two panels for $p_\lambda(x|w)$ and $p^*(x|w)$, it can clearly be seen that the bounded-optimal perceptual model now spends its scarce

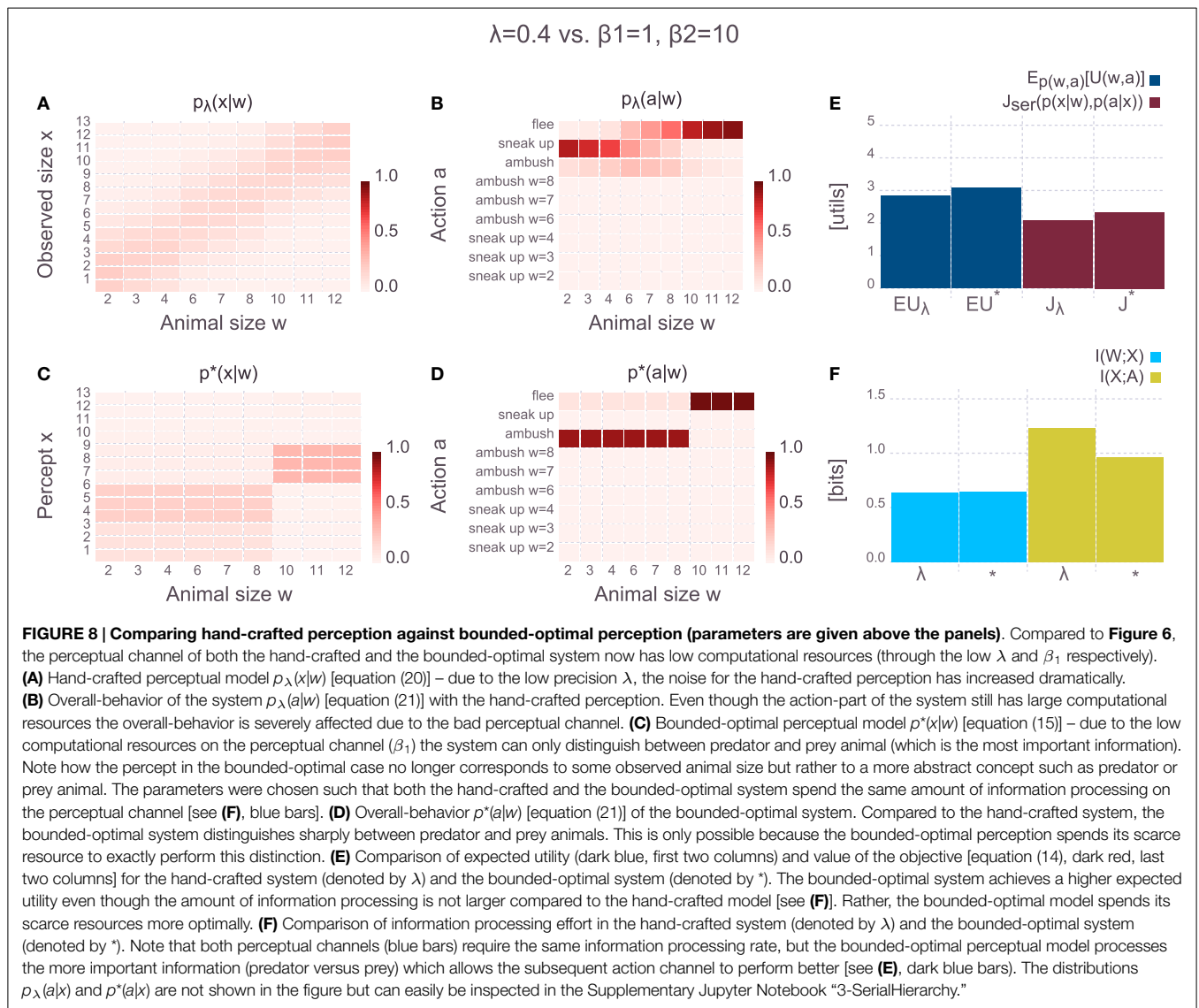


resources to reliably discriminate between large animals and all other animals. The overall behavioral policies $p(a|w)$ reflect the limited perceptual capacity in both cases, however, the bounded-optimal case scores a higher expected utility of ≈ 3 utils compared to the hand-crafted case. The overall objective $J_{ser}(p(x|w), p(a|x))$ is also higher for the bounded-optimal model, indicating that this model should be preferred because it finds a better trade-off between expected utility and information processing cost.

Note that in all three examples the optimal percept $p^*(x|w)$ often leads to a uniform mapping of an exclusive subset of world-states w to the same set of percepts x . Importantly, these percepts do not directly correspond to an observed animal size as in the case of the hand-crafted model of perception. Rather, the optimal percepts often encode more abstract concepts such as medium- or large-sized animal (as in **Figure 6**) or predator and prey animal (as in **Figures 7** and **8**). In a sense, abstractions similar to the ones shown in the recommender system example in the previous section (**Figure 3**) emerge in the predator-prey example as well but now they also manifest themselves in the form of abstract percepts. Crucially, the abstract percepts allow

for more efficient information processing further downstream in the decision-making part of the system. The formation of these abstract percepts is driven by the embodiment of the agent and reflects certain aspects of the utility function of the agent. For instance, unlike the actor in **Figure 6**, the actors in **Figures 7** and **8** would not “understand” the concept of medium-sized animals as it is of no use to them: with their very limited resources it is most important for them to have the two perceptual concepts of predator and prey. Note that the cardinality of X in the bounded-optimal model of perception is fixed in all examples in order to allow for easy comparison against the hand-crafted model, but it could be reduced further without any consequences (up to a certain point) – this can be explored in the Supplementary Jupyter Notebook “3-SerialHierarchy.”

The solutions shown in this section were obtained by iterating the self-consistent equations until numerical convergence. Since there is no convergence-proof, it cannot be fully ruled out that the solutions are sub-optimal with respect to the objective. However, the point of the simulation results shown here is to allow for easier interpretation of the theoretical results and highlight



certain aspects of the theoretical findings. We discuss this issue in Section 5.2.

4. PARALLEL INFORMATION-PROCESSING HIERARCHIES

Rational decision-making requires searching through a set of alternatives a and picking the option with the highest expected utility. Bounded rational decision-making replaces the “hard maximum” operation with a soft selection mechanism where the first action that satisfies a certain level of expected utility is picked. A parallel hierarchical architecture allows for a prior partitioning of the search space which reduces the effective size of the search space and thus speeds up the search process. For instance, consider a medical system that consists of general practitioners and specialist doctors. The general doctor can restrict the search space for a particular ailment of a patient by determining which specialist the patient should see. The specialist doctor in turn can determine the exact disease. This leads to a two-level decision-making hierarchy

consisting of a high-level partitioning that allows for making a subsequent low-level decision with reduced (search) effort. In statistics, the partitioning that is induced by the high-level decision is often referred to as a *model* and is commonly expressed as a probability distribution over the search space $p(a|m)$ (where m indicates the model) which also allows for a soft-partitioning. The advantage of hierarchical architectures is that the computation that leads to the high-level reduction of the search space can be stored in the model (or in a set of parameters in case of a parametric model). This computation can later be re-used by using the correct model (or set of parameters) in order to perform the low-level computation more efficiently. Interestingly, it should be most economic to put the most re-usable, and thus more abstract, information into the models $p(a|m)$ which leads to a hierarchy of abstractions. However, in order to make sure that the correct model is used, another deliberation process $p(m|w)$ is required (where w indicates the observed stimulus or data). Another problem is how to choose the partitioning to be most effective. In this section, we address both problems from a bounded rational point

of view. We show that the bounded optimal solution $p^*(a|m)$ trades off the computational cost for choosing a model m against the reduction in computational cost for the low-level decision.

To keep the notation consistent across all sections of the paper we denote the model m in the rest of the paper with the variable x . This is in contrast to Section 3, where x played the role of a percept. The advantage of this notation is that it allows to easily see similarities and differences of the information terms and solution equations of the different cases. In particular, in Section 5 we present a unifying case that includes the serial and parallel case as special cases – by keeping the notation consistent this can easily be seen.

4.1. Optimal Partitioning of the Search Space

Constructing a two-level decision-making hierarchy requires the following three components: high-level models $p(a|x)$, a model selection mechanism $p(x|w)$ and a low-level decision maker $p(a|w, x)$ (w denotes the observed world-state, x indicates a particular model and a is an action). The first two distributions are free to be chosen by the designer of the system, for $p(a|w, x)$ a maximum expected utility decision-maker is the optimal choice if computational costs are neglected. Here, we take computational cost into account and replace the MEU decision-maker with a bounded rational decision-maker that includes MEU as a special case ($\beta_3 \rightarrow \infty$) – the bounded rational decision-maker optimizes equation (7) by implementing equation (9). In the following we show how all parts of the hierarchical architecture:

1. Selection of model (or expert): $p(x|w)$ (22)
2. Prior knowledge of model (or expert): $p(a|x)$ (23)
3. Bounded rational decision of model (or expert):

$$p^*(a|w, x) = \frac{1}{Z(w, x)} p(a|x) e^{\beta_3 U(w, a)} \quad (24)$$

emerge from optimally trading off computational cost against gains in utility. Importantly, $p(a|x)$ plays the role of a prior distribution for the bounded rational decision-maker and reflects the high-level partitioning of the search space.

The optimization principle that leads to the bounded-optimal hierarchy trades off expected utility against the computational cost of model selection $I(W; X)$ and the cost of the low-level decision using the model as a prior $I(W; A|X)$:

$$\begin{aligned} \arg \max_{p(x|w), p(a|w, x)} \mathbf{E}_{p(w, x, a)} [U(w, a)] - \frac{1}{\beta_1} I(W; X) - \frac{1}{\beta_3} I(W; A|X) \\ = \arg \max_{p(x|w), p(a|w, x)} J_{\text{par}}(p(x|w), p(a|w, x)). \end{aligned} \quad (25)$$

The set of self-consistent solutions is given by

$$p^*(x|w) = \frac{1}{Z(w)} p(x) \exp(\beta_1 \Delta F_{\text{par}(w, x)}) \quad (26)$$

$$p(x) = \sum_w p(w) p^*(x|w) \quad (27)$$

$$p^*(a|w, x) = \frac{1}{Z(w, x)} p^*(a|x) \exp(\beta_3 U(w, a)) \quad (28)$$

$$p^*(a|x) = \sum_w p(w|x) p^*(a|w, x), \quad (29)$$

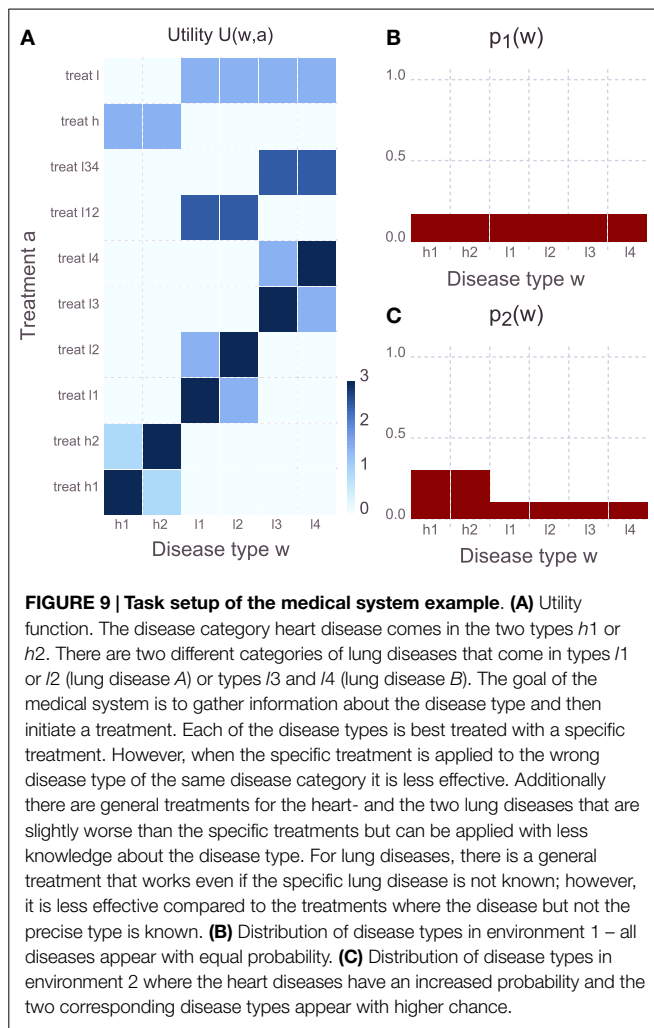
where $Z(w)$ and $Z(w, x)$ denote the corresponding normalization constants or partition sums. $p(w|x)$ is given by Bayes' rule $p(w|x) = \frac{p(w)p^*(x|w)}{p(x)}$ and $\Delta F_{\text{par}(w, x)}$ is the free energy difference of the low-level stage:

$$\Delta F_{\text{par}(w, x)} := \mathbf{E}_{p^*(a|w, x)} [U(w, a)] - \frac{1}{\beta_3} D_{\text{KL}}(p^*(a|w, x) || p^*(a|x)), \quad (30)$$

see equation (5). More details on the derivation of the solution equations can be found in the Supplementary Methods Section 3. By comparing the solution equations (26)–(29) with equations (22)–(24) the hierarchical structure of the bounded-optimal solution can be seen clearly. The bounded-optimal model selector in equation (26) maximizes the downstream free-energy trade-off $\Delta F_{\text{par}(w, x)}$ in a bounded rational fashion and is similar to the optimal perceptual model of the serial case [equation (15)]. This means that the optimal model selection mechanism is shaped by the utility function as well as the computational process on the low-level stage of the hierarchy (governed by β_3) but also by the computational cost of model selection (governed by β_1). The optimal low-level decision-maker given by equation (28) turns out to be exactly a bounded rational decision-maker with $p(a|x)$ as a prior – identical to the low-level decision-maker that was motivated in equation (24). Importantly, the bounded-optimal solution provides a principled way of designing the models $p(a|x)$ [see equation (29)]. According to the equation, the optimal model $p^*(a|x)$ is given by a Bayesian mixture over optimal solutions $p^*(a|w, x)$ where w is known. The Bayesian mixture turns out to be the optimal compressor of actions for unknown w under the belief $p(w|x)$.

4.2. Illustrative Example

To illustrate the formation of bounded-optimal models, we designed the following example: in a simplified environment, only three diseases can occur – a heart disease or one of two possible lung diseases. Each of the diseases comes in two possible types (e.g., type 1 or type 2 diabetes). Depending on how much information is available on the symptoms of a patient, diseases can be treated according to the specific type (which is most effective) or with respect to the disease category (which is less effective but requires less information). See **Figure 9** for a plot of the utility function and a detailed description of the example. The goal is to design a medical analysis hierarchy that initiates the best possible treatment, given its limitations. The hierarchy consists of an automated medical system that can cheaply take standard measurements to partially assess a patient's disease category. Additionally, the patient is then sent to a specialist who can manually perform more elaborate measurements if necessary to further narrow down the patient's precise disease type and recommend a treatment. The automated system should be designed in a way that minimizes the additional measurements required by the specialists. More formally, the automated system delivers a first diagnosis x given the patient's precise disease type w according to $p(x|w)$. The first diagnosis narrows down the possible treatments a according to a model $p(a|x)$. For each x , a specialist can further reduce uncertainty about the correct treatment by performing more measurements $p(a|w, x)$. We compare the optimal design of the automated system and the corresponding optimal



treatment recommendations to the specialist $p^*(a|x)$ according to equation (29) in two different environments: one, where all disease types occur with equal probability (Figure 9B) versus two, where heart diseases occur with increased chance (Figure 9C). For this example, the number of different high-level diagnoses X is set to $|\mathcal{X}| = 3$ which also means that there can be three different treatment recommendations $p(a|x)$. Since in the example the total budget for performing measurements is quite low (reflected by β_1, β_3 both being quite low), the whole system (automated plus specialists) can in general not gather enough information about the symptoms to treat every disease type with the correct specific treatment. Rather, the low budget has to be spent on gathering the most important information.

Figure 10 shows bounded-optimal hierarchies for the medical system in both environments. The top row in Figure 10 shows the optimal hierarchy for the environment where all diseases appear with equal probability: the automated system $p^*(x|w)$ (see Figure 10A) distinguishes between a heart disease, lung disease A and lung disease B, which means that there is one treatment recommendation for heart diseases and one treatment recommendation for each of the two possible lung diseases respectively (see the three columns of $p^*(a|x)$ in Figure 10B). Since the general

treatment for the heart disease works less effective than the general treatments for the two lung diseases, the (very limited) budget of the specialists is completely spent on finding the correct specific heart treatment. Both lung diseases are treated with their respective general treatments since the two lung specialists have no budget for additional measurements. Since the automated system already distinguishes between the two lung diseases, it can narrow down the possible treatments to a delta over the correct general treatment, thus requiring no additional measurements by the lung specialists (shown by the two columns in $p^*(a|x)$ that have a delta over the treatment).

The bottom row in Figure 10 shows the optimal hierarchy for the environment where heart diseases appear with higher probability. In this case it is optimal to redesign the automated system to distinguish between the two types of the heart disease $h1, h2$, and lung diseases in general (see $p^*(x|w)$ in Figure 10D of the figure). This means that there are now treatment recommendations $p^*(a|x)$ for $h1$ and $h2$ that do not require any more measurements by the specialists (shown by the delta over a treatment in the first two columns of $p^*(a|x)$ in Figure 10E) and there is another treatment recommendation for lung diseases. The corresponding specialist can use the limited budget to perform additional measurements to distinguish between the two categories of lung disease (but not between the four possible types as this would require more measurements than the budget allows). The example illustrates how the bounded-optimal decision-making hierarchy is shaped by the environment and emerges from optimizing the trade-off between expected utility and overall information processing cost. Readers can interactively explore the example in the Supplementary Jupyter Notebook “4-ParallelHierarchy” – in particular by changing the information processing costs of the specialists β_3 or changing the number of specialists by increasing or decreasing the cardinality of X .

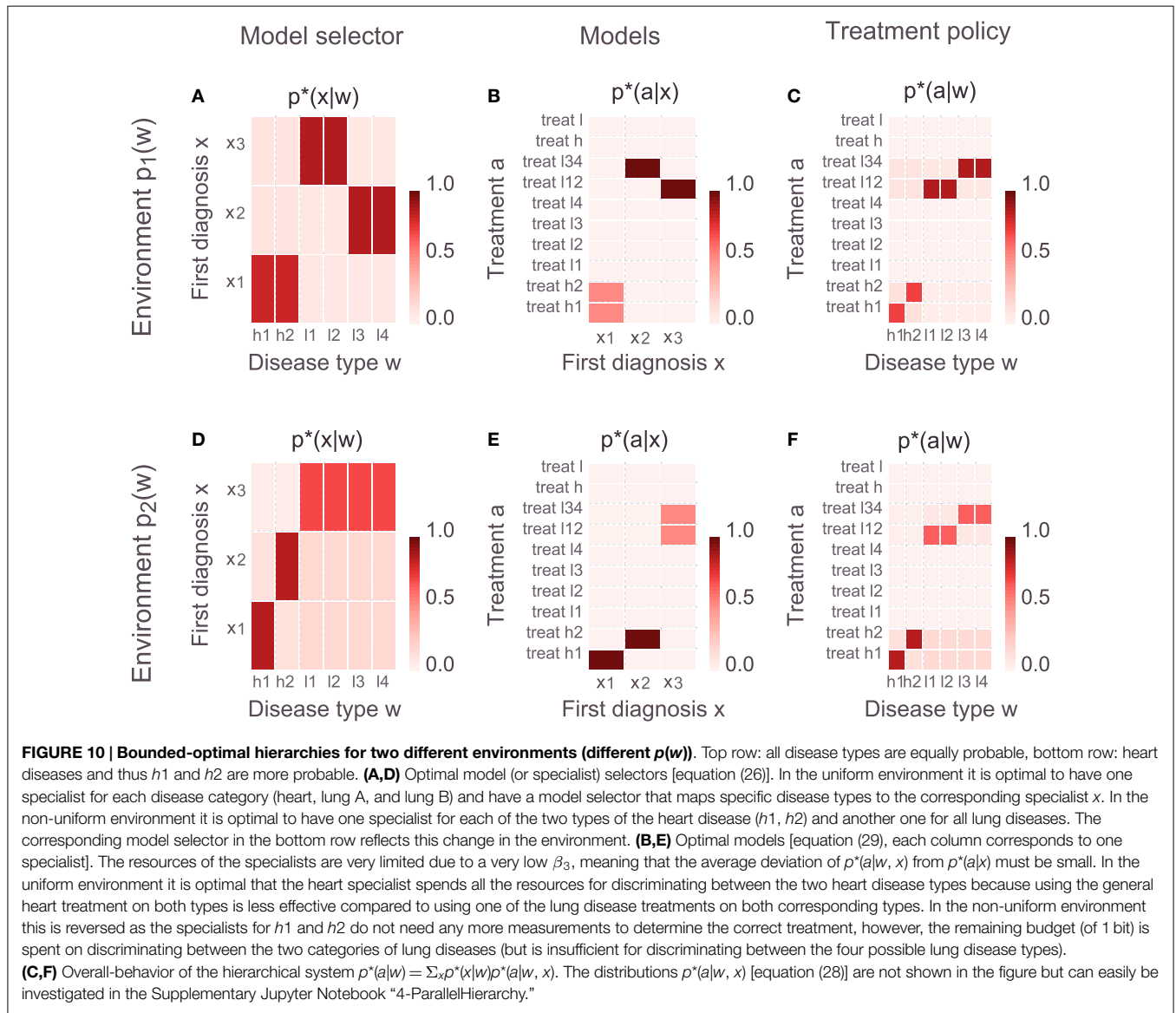
The solutions shown in this section were obtained by iterating the self-consistent equations until numerical convergence. Since there is no convergence-proof, it cannot be fully ruled out that the solutions are sub-optimal with respect to the objective. However, the point of the simulation results shown here is to allow for easier interpretation of the theoretical results and highlight certain aspects of the theoretical findings. We discuss this issue in Section 5.2.

4.3. Comparing Parallel and Serial Information Processing

In order to achieve a certain expected utility, a certain overall rate $I(W; A)$ is needed. In the one-step rate-distortion case (Section 2) the channel from w to a must have a capacity larger or equal to that rate. In the serial case (Section 3) there is a channel from w to x and another channel from x to a . Both serial channels must at least have a capacity of $I(W; A)$ in order to achieve the same overall rate, as the following inequality always holds for the serial case

$$I(W; A) \leq \min \{I(W; X), I(X; A)\}.$$

In contrast, the parallel architecture allows for computing a certain overall rate $I(W; A)$ using channels with a lower capacity



because the contribution in reducing uncertainty about a on each level of the hierarchy splits up as follows:

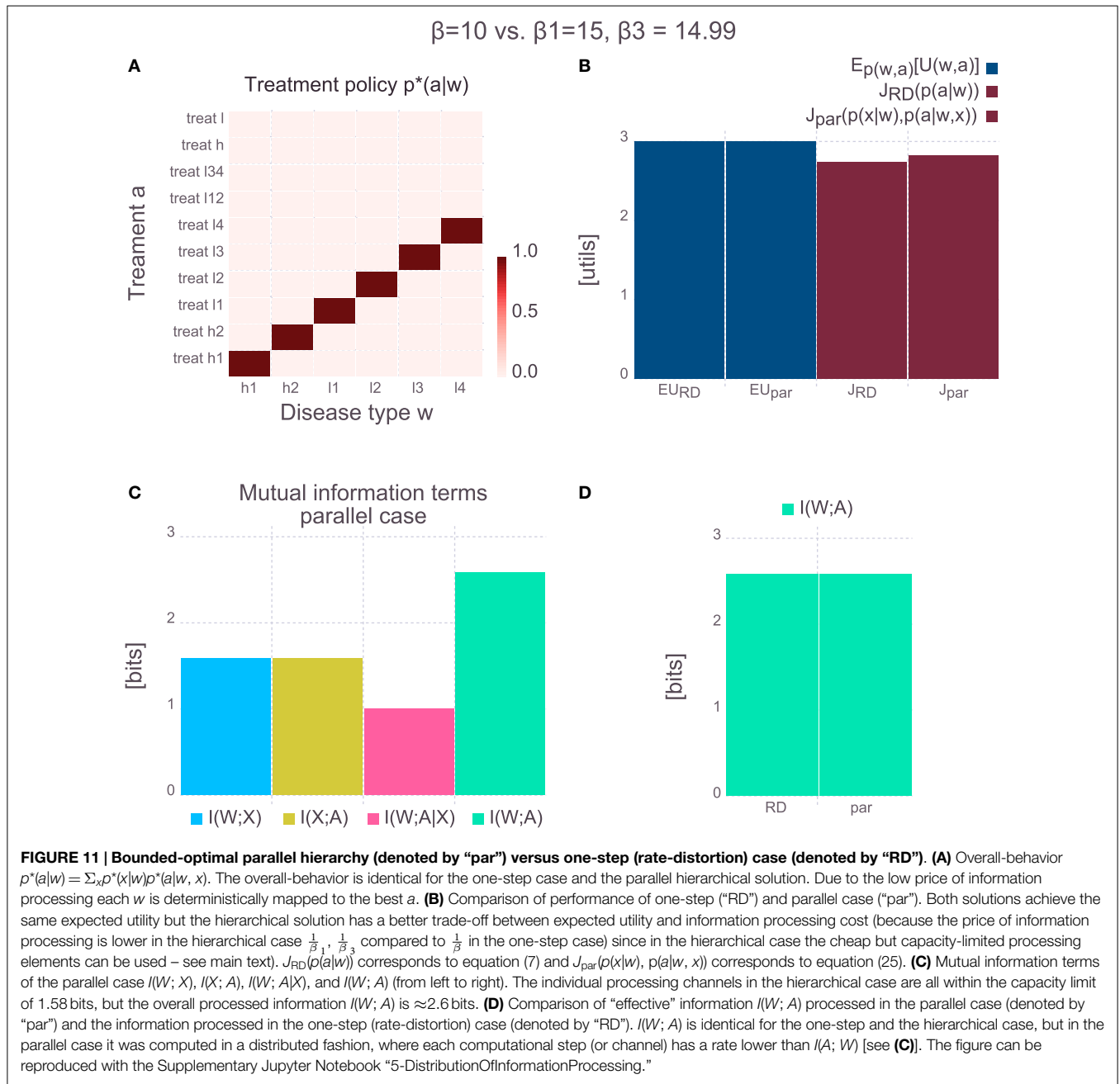
$$\underbrace{I(W, X; A)}_{\text{total reduction}} = \underbrace{I(X; A)}_{\text{high-level}} + \underbrace{I(W; A|X)}_{\text{low-level}}$$

which implies $I(W, X; A) \geq I(X; A)$. In particular, if the low-level step contributes information then $I(W; A|X) > 0$ and the previous inequality becomes strict. The same argument also holds when considering $I(W; A)$ (see Section 5.1).

In many scenarios the maximum capacity of a single processing element is limited and it is desirable to spread the total processing load on several elements that require a lower capacity. For instance, there could be technical reasons why processing elements with 5 bits of capacity can easily be manufactured but processing elements with a capacity of 10 bits cannot be manufactured or are disproportionately more costly to produce. In the one-step case and the serial case the only way to stay below a

certain capacity limit is by tuning β until the required rate is below the capacity – however, in both cases this also decreases the overall rate $I(W; A)$. In the parallel hierarchical case several building blocks with a limited capacity can be used to produce an overall rate $I(W; A)$ larger than the capacity of each processing block.

Splitting of information processing load onto several processing blocks is illustrated in **Figure 11**, where the one-step and parallel hierarchical solutions to the medical example are compared. In this example, the price of information processing in the one-step case is quite low ($\beta = 10$ bits per util) such that the corresponding solution leads to a deterministic mapping of each w to the best a (see **Figure 11A**). Doing so requires $I(W; A) \approx 2.6$ bits (see **Figure 11B**). Now assume for the sake of this example that processing elements, where information processing cost is reduced (15 bits per util), could be manufactured, but the maximum capacity of these elements is 1.58 bits. In the one-step case these processing elements can only be used if it is acceptable to reduce the rate $I(W; A)$ to 1.58 bits (by tuning β) which



would imply a lower expected utility. However, in the parallel hierarchical case the new processing elements can be used (see **Figure 11C**) which leads to a reduced price of information processing ($\beta_1 = 15, \beta_3 = 14.999$). In conjunction the new processing elements process the same effective information $I(W; A)$ (see **Figure 11D**) and achieve the same expected utility as the one-step case (see **Figure 11B**). However, since the price for information processing is lower on the more limited elements, the overall trade-off between expected utility and information processing cost is in favor of the parallel hierarchical architecture. Note that in this example it is important that the cardinality of X is limited (in this case $|\mathcal{X}| = 3$) and $\beta_1 > \beta_3$. We discuss this in the next paragraphs.

In the parallel hierarchical case, there are two possible pathways from w to a :

$$\begin{aligned} \text{Two-stage serial pathway} \quad & I(W; X) \rightarrow I(X; A) \\ \text{Parallel pathway} \quad & I(W; A|X) \end{aligned}$$

Note that $I(X; A)$ does not appear in the objective [equation (25)]; however, it is crucial for distributing information processing on both levels of the hierarchy (see more analysis of the medical system example in **Figure 11**). $I(X; A)$ measures the average adaptation effort for going from $p(a)$ to $p(a|x)$. In the parallel hierarchical case it is a measure of how much the different models $p(a|x)$ narrow down the search space compared to the average

$p(a) = \sum_x p(x)p(a|x)$. If all models are equal $p(a|x) = p(a) \forall x$ the mutual information $I(X; A)$ is zero. Note, however, that a large $I(X; A)$ is rendered useless by a low $I(W; X)$ and vice versa – if the model selector is very bad, even the best models are not useful and vice versa.

Since there is no cost for having a large rate $I(X; A)$, the overall throughput of the serial pathway is effectively governed by β_1 as it affects the rate $I(W; X)$. Similarly, β_3 governs the rate on the parallel pathway $I(W; A|X)$. As a result, whenever one of the two inverse temperatures β_1 and β_3 is larger than the other, it becomes more economic to shift all the information processing to the cheaper pathway (either serial or parallel) thus rendering the other pathway obsolete. The only scenario where it can be advantageous to use both pathways (and distribute computation) is when the cheaper pathway has insufficient capacity and the more expensive pathway is used to take on additional computational load that cannot be handled by the cheap pathway alone. Effectively, this translates into the constraint that the serial pathway must be cheaper $\beta_1 > \beta_3$ and additionally the serial pathway must be limited in its capacity by limiting the cardinality $|\mathcal{X}|$ (see Supplementary Methods Section 5 for a detailed discussion).

Note the important difference between changing the cardinality of X which governs the channel capacity of the serial pathway (that is the maximally possible rates $I(W; X)$, $I(X; A)$) but has no influence on the price of information processing and changing β_1 which governs the price of processing $I(W; X)$ and hence affects the actual rate on the serial pathway but has no effect on the capacity of the channels of the serial pathway.

$$I(W; X) = H(X) - H(X|W) \leq \mathbf{H}(\mathbf{X}) \quad (31)$$

$$I(X; A) = H(A) - H(A|X) = H(X) - H(X|A) \leq \mathbf{H}(\mathbf{X}) \quad (32)$$

$$I(W; A|X) = H(A|X) - H(A|W, X) \leq H(A|X) \quad (33)$$

$H(X)$ is an upper bound for both $I(W; X)$ but also $I(X; A)$ and the upper bound of $H(X)$ itself is a function of $|\mathcal{X}|$. Note that since there is no cost associated with $I(X; A)$ it is generally desirable to maximize $I(X; A)$ at least such that $I(X; A) \geq I(W; X)$. To do so $H(A|X)$ must be pushed toward zero [equation (32)] – however, this simultaneously pushes the upper bound for $I(W; A|X)$ toward zero [equation (33)]. In case of a sufficiently limited $H(X)$ (through a low $|\mathcal{X}|$), $I(X; A)$ cannot be fully maximized, therefore leaving a non-zero upper bound for $I(W; A|X)$.

In the example shown in **Figure 11** information processing is performed on both the serial pathway ($I(W; X)$ and $I(X; A)$) but also on the parallel pathway ($I(W; A|X)$) because the constraints for distribution of information processing are fulfilled: $\beta_1 > \beta_3$ and the capacity (that is the maximum rate $I(W; X)$ and $I(X; A)$) of the serial pathway is limited by the (low) cardinality $|\mathcal{X}| = 3$. The cardinality of X for the example can easily be changed in the Supplementary Jupyter Notebook “4-ParallelHierarchy” – if it is for instance increased to $|\mathcal{X}| = 6$ while keeping all other parameters the same, the whole information processing load will be entirely on the serial pathway and $I(W; A|X) = 0$. Alternatively to limiting the cardinality of X , a cost for $I(X; A)$ could be introduced to limit the computational resources for computing $p(a|x)$ from $p(a)$. This is explored in Section 5.

5. TOWARD MORE GENERAL ARCHITECTURES

In the serial case in Section 3, information processing cost arises from adapting $p(x)$ to $p(x|w)$ and $p(a)$ to $p(a|x)$, and the average informational effort is measured by $I(W; X)$ and $I(X; A)$. In the parallel hierarchical case in Section 4 the two information processing terms considered are $I(W; X)$ and $I(W; A|X)$, where the latter measures the average informational effort for adapting from $p(a|x)$ to $p(a|w, x)$. In this section, we present a mathematically unifying case that considers all three mutual information terms and includes the serial and the parallel case as special cases. This unifying formulation might also be a starting point for generalizing toward more than three random variables as the corresponding objective function could easily be extended to include more variables.

The general case uses the same factorization of the three variables W, X, A as the parallel case: $p(w, x, a) = p(w)p(x|w)p(a|w, x)$. Given this factorization, the KL-divergence between the joint $p(w, x, a)$ and the product of all three marginals, also known as the total correlation $C(W, X, A)$, leads to:

$$C(W, X, A) = D_{\text{KL}}(p(w, x, a) || p(w)p(x)p(a)) \\ = H(W) + H(X) + H(A) - H(W, X, A) \quad (34)$$

$$= \sum_{w,x,a} p(w, x, a) \log \frac{p(w, x, a)}{p(w)p(x)p(a)}$$

$$= \sum_{w,x} p(w, x) \log \frac{p(x|w)}{p(x)}$$

$$+ \sum_{w,a,x} p(w, x, a) \log \frac{p(a|w, x)}{p(a)}$$

$$= I(W; X) + I(W, X; A) \quad (35)$$

$$= I(W; X) + I(X; A) + I(W; A|X). \quad (36)$$

The total correlation (Watanabe, 1960), also called multivariate constraint (Garner, 1962) or multiinformation (Studený and Vejnarová, 1998), is the sum of the three information processing terms considered in the serial and parallel case. The general objective is formed by assigning different prices to each of the terms and trading off the resulting information processing cost against the expected utility:

$$\arg \max_{p(x|w), p(a|w, x)} \mathbf{E}_{p(w, x, a)} [U(w, a)] - \frac{1}{\beta_1} I(W; X) \\ - \frac{1}{\beta_2} I(X; A) - \frac{1}{\beta_3} I(W; A|X) \\ = \arg \max_{p(x|w), p(a|w, x)} J_{\text{gen}}(p(x|w), p(a|w, x)). \quad (37)$$

Identical to the parallel hierarchical case, the general case has two information processing pathways that allow for splitting up the total computational load: a serial pathway consisting of the two stages $I(W; X)$ and $I(X; A)$ and a parallel pathway $I(W; A|X)$. If any of the pathways is cheaper than the other one, it is more economical to shift all the computation to the cheaper pathway. However, the capacity of the serial pathway can be limited, for

example by reducing the cardinality of X . In such a case the parallel pathway can take on additional computational load, leading to a parallel hierarchical information processing architecture.

The solution to the general objective is given by the following set of five self-consistent equations (the detailed derivation of the solutions is included in the Supplementary Methods Section 1):

$$p^*(x|w) = \frac{1}{Z(w)} p(x) \exp \left(\beta_1 \Delta F_{\text{gen}}(w, x) - \left(\frac{\beta_1}{\beta_3} - \frac{\beta_1}{\beta_2} \right) D_{\text{KL}}(p^*(a|w, x) || p^*(a|x)) \right) \quad (38)$$

$$p(x) = \sum_w p(w) p^*(x|w) \quad (39)$$

$$p^*(a|w, x) = \frac{1}{Z(w, x)} p^*(a|x) \exp \left(\beta_3 U(w, a) - \frac{\beta_3}{\beta_2} \log \frac{p^*(a|x)}{p(a)} \right) \quad (40)$$

$$p^*(a|x) = \sum_w p(w|x) p^*(a|w, x) \quad (41)$$

$$p(a) = \sum_{w,x} p(w) p^*(x|w) p^*(a|w, x), \quad (42)$$

where $Z(w)$ and $Z(w, x)$ denote the corresponding normalization constants or partition sums. The conditional distribution $p(w|x)$ is given by Bayes' rule $p(w|x) = \frac{p(w)p^*(x|w)}{p(x)}$ and $\Delta F_{\text{gen}}(w, x)$ is the free energy difference

$$\Delta F_{\text{gen}}(w, x) := \mathbf{E}_{p^*(a|w,x)}[U(w, a)] - \frac{1}{\beta_2} D_{\text{KL}}(p^*(a|w, x) || p(a)).$$

For $\beta_3 < \beta_2$ the KL-term in equation (38) has a positive sign, implying that the KL-divergence is a utility instead of a cost which makes sense if computation on $I(W; A|X)$ is cheaper than computation on $I(A|X)$. For $\beta_3 > \beta_2$ the KL-term gets a negative sign, implying that the KL-divergence is a cost, as a result of computation on $I(W; A|X)$ being more expensive than computation on $I(A|X)$.

Equation 38 can also be rewritten as (see Supplementary Methods Section 1.2):

$$p^*(x|w) = \frac{1}{Z(w)} p(x) \exp \left(\beta_1 \Delta F_{\text{par}}(w, x) - \frac{\beta_1}{\beta_2} \sum_a p^*(a|w, x) \log \frac{p^*(a|x)}{p(a)} \right) \quad (43)$$

where $\Delta F_{\text{par}}(w, x)$ is the same free energy difference as in the parallel case

$$\Delta F_{\text{par}}(w, x) := \mathbf{E}_{p^*(a|w,x)}[U(w, a)] - \frac{1}{\beta_3} D_{\text{KL}}(p^*(a|w, x) || p^*(a|x)) \quad (44)$$

see equation (30).

Comparing the objective in equation (37) with the objective of the parallel case in equation (25), it can be seen that by setting $\beta_2 \rightarrow \infty$ the two objective functions become equal and the implicit assumption that in the parallel case there is no cost for going from $p(a)$ to $p(a|x)$ (as the latter is considered a prior) is made explicit. The solution equations of the general case also collapse

TABLE 2 | Recovery of special cases from the general, unifying case by specific settings of the inverse temperatures.

Case	β_1	β_2	β_3	(inverse) price per transformation
General	β_1	β_2	β_3	$\beta_1: p(x) \rightarrow p(x w)$ $\beta_2: p(a) \rightarrow p(a x)$ $\beta_3: p(a x) \rightarrow p(a w, x)$
Total correlation	β	β	β	$\beta: p(x) \rightarrow p(x w)$ $\beta: p(a) \rightarrow p(a w, x)$
Degenerate TC	β_1	β	β	$\beta_1: p(x) \rightarrow p(x w)$ $\beta: p(a) \rightarrow p(a w, x)$
Serial	β_1	β_2	$\rightarrow 0$	$\beta_1: p(x) \rightarrow p(x w)$ $\beta_2: p(a) \rightarrow p(a x)$ $p(a w, x) = p(a x) \forall w$ $I(W; A X) = 0$
Parallel	β_1	$\rightarrow \infty$	β_3	$\beta_1: p(x) \rightarrow p(x w)$ $\beta_3: p(a x) \rightarrow p(a w, x)$
Joint (x,a)	β	$\rightarrow \infty$	β	$\beta: p(x, a) \rightarrow p(x, a w)$

The table shows how to set the inverse temperatures in the general case to recover particular special cases. The last column shows for all cases which probability-transformations are considered as computational effort and the corresponding (inverse) price. The case "degenerate total correlation" is not described in the main paper, but is outlined in the Supplementary Methods Section 4 – it could be relevant in a two-dimensional decision-making scenario, that is when x is considered one dimension of the decision and a is considered the other dimension. This implies that the utility function also depends on $x: U(w, x, a)$. Similarly, the case "joint (x,a)" is only described in the Supplementary Methods Section 5 and describes how the one-step (rate-distortion) case is related to the general case.

to the solutions of the parallel case by letting $\beta_2 \rightarrow \infty$: compare equations (43) and (40) against equations (26) and (28). The general case thus also allows for designing more realistic hierarchical cases where there is a small cost for switching models $p^*(a|x)$ that arises, for instance, from loading a certain set of parameters or switching to a particular sampler or reading the model from memory. Similarly, the serial case can be recovered by $\beta_3 \rightarrow 0$. The special cases of the general objective are summarized in **Table 2**.

5.1. Effective Information Throughput $I(W; A)$

The amount of information processing that effectively contributes toward achieving a high expected utility is measured by $I(W; A)$ which does not directly appear in the objective of the general case (nor the serial and parallel case). However, the effective information throughput of the system is given by

$$I(W; A) = I(W; X; A) + I(W; A|X) \quad (45)$$

$$= I(W; X) + I(X; A) - I(X; W, A) + I(W; A|X) \quad (46)$$

$$= C(W, X, A) - I(X; W, A) \quad (47)$$

where $I(W; X; A)$ denotes the multivariate mutual information (MMI; Yeung, 1991). The first equation above is obtained by re-ordering the definition of the MMI $I(K; L; M) = I(K; M) - I(K; M|L)$. Note that in the serial hierarchical case $I(W; A|X) = 0$ always holds. The equations above also show how the total correlation $C(W, X, A)$ and the MMI $I(W; X; A)$ are related.

The multivariate mutual information is upper-bounded by $I(W; X; A) \leq \min\{I(W; X), I(W; A), I(X; A)\}$ (see (Yeung, 1991)). Using the bound in equation (45) leads to an upper bound for the effective information throughput:

$$I(W; A) \leq \min \{I(W; X), I(X; A)\} + I(W; A|X) \quad (48)$$

Equation 48 shows how information processing in the general case can be distributed between a two-stage serial pathway (consisting of $I(W; X)$ and $I(X; A)$) and a parallel pathway ($I(W; A|X)$). The general case forms a parallel hierarchy similar to Section 4, but it allows to associate a cost with $I(X; A)$ (which is a measure of how costly it is to switch models). Importantly the discussion on splitting up information processing between both levels of the parallel hierarchy as in Section 4.3 also holds for the general case.

5.2. Iterating the Self-Consistent Equations

For the simulation results shown in this paper the corresponding set of self-consistent equations was iterated until convergence (by checking that the total change in probability distributions between two iteration steps is below a certain threshold – see code underlying the Supplementary Notebooks for details). This is inspired by the Blahut-Arimoto scheme that is proven to converge to the global maximum in the rate-distortion case (Csiszar, 1974; Cover and Thomas, 1991) (Section 2). Unfortunately there is no such proof for iterating the sets of self-consistent equations of the general, serial or parallel case. It is not clear whether the optimization problems are still convex and have a global solution, nor is it clear that iterating the self-consistent equations would converge toward these global solutions. A convexity and convergence analysis is certainly among the most important steps for future investigations of the principles presented here. At this point, we can only report empirical observations and interested readers are encouraged to explore convergence behavior using the Supplementary Jupyter Notebooks (which include plots that show convergence behavior across iterations) but also the underlying code (published in the Supplementary Material).

6. DISCUSSION AND CONCLUSION

The overarching principle behind this paper is the consistent application of the trade-off of gains in expected utility against the computational cost that these gains require. Here, computational cost is defined as the average effort of computational adaptation (measured by the mutual information) multiplied by the price of information processing. This definition is motivated by first principles (Mattsson and Weibull, 2002; Ortega and Braun, 2010; Ortega and Braun, 2011) and is grounded in a thermodynamic framework for decision-making (Ortega and Braun, 2013). Mathematically, the basic principle is identical to the principle behind rate-distortion theory, the information-theoretic framework for lossy compression (Genewein and Braun, 2013; Still, 2014). This connection is no coincidence as bounded rational decision-making can be cast as a lossy compression problem in lossy compression the goal is to transmit the most relevant information (given the limited channel capacity) in order to minimize a distortion-function. In bounded rational decision-making the goal is to process the most relevant information in order maximize a utility function, given the limitations on information processing. In Section 2, we have shown how different levels of behavioral abstraction can be induced by different computational limitations. The authors in (van Dijk and Polani, 2013) use the *Relevant Information* method, which is a particular application

of rate-distortion theory and find a very similar emergence of “natural abstractions” and “ritualized behavior” when studying goal-directed behavior in the MDP case. We have shown how the basic principle can be extended to more complex cases and that analytic solutions can be obtained for these cases. Importantly, the solutions allow for interesting interpretations, highlighting how the same fundamental trade-off can lead to systems that elegantly solve more complex problems. For instance, when designing a perception-action system, the perceptual part of the system can easily be understood as a lossy compressor, but the corresponding distortion-function is not intuitively clear. We have shown in Section 3 how the extended lossy compression principle leads to a well-defined distortion-function for the perceptual part of the system that optimizes the downstream trade-off between expected utility and computational cost. In a similar fashion we have shown in Section 4 how the problem of designing bounded-optimal decision-making hierarchies is fundamentally equal to designing a distributed lossy compressor (that is spread over both levels of the hierarchy).

In the serial hierarchy in Section 3, we compared a perceptual channel that performs Bayesian inference against a bounded-optimal perceptual channel that optimizes the downstream free energy trade-off. We found that the difference between both models of perception was that in one case the likelihood model $p(x|w)$ was unspecified (Bayesian inference) whereas in the other case it was well defined (bounded-optimal solution). Perception is often conceptualized as (Bayesian) inference, however, given our findings there is a subtle but important difference. In our model of a perception-action system, the goal of the perceptual model $p(x|w)$ is to extract the most relevant information from w for choosing an action according to $p(a|x)$, given the computational limitations of the system. In plain inference, the goal is to predict w from x very well and the likelihood model is thus chosen to maximize predictive power. In many cases the two objectives coincide as achieving a large expected utility often requires precise knowledge about w . However, this must not always be the case and in particular for systems where computational limitations play a large role, the (limited) computational resources can often be spent more economically which allows for a higher expected utility at the cost of not being able to predict w from x that well. An interesting machine-learning application of the serial principle could be the design of optimal features for classification.

In Section 4, we showed how parallel bounded-optimal decision-making hierarchies can emerge from solving the trade-off between utility and cost of computation. We found that the condition for parallel hierarchies to being optimal solutions was that the price for model selection is lower than the price for processing information on the low level of the hierarchy ($\beta_1 > \beta_3$). At the same time, the upper level of the hierarchy must be limited in capacity (for instance through the cardinality $|\mathcal{X}|$). Intuitively this makes sense and fits with the general observation that often hardware that allows for cheap information processing is itself quite expensive to build (low signal to noise ratio, etc.). Therefore the amount of hardware that allows for cheap information processing is likely to be quite limited. It remains an open question whether this is a fundamental constraint for hierarchies being optimal solutions or whether there are other arguments in

favor of hierarchical architectures. In changing environments, for example, the overall change required to adapt a system is smaller for a hierarchical system, compared to a flat system, because the more abstract levels of the hierarchy might require little or no change at all. It could also be that the upper levels of a hierarchical model based on our principle contain more transferable knowledge that can be applied to novel but similar tasks. Changing the task corresponds to changing the utility-function, which requires a non-equilibrium analysis (Grau-Moya and Braun, 2013) that we leave for future investigation.

In our simulations, we initialize $p(x|w)$ and either $p(a|w, x)$ (parallel hierarchies) or $p(a|x)$ (serial hierarchies) and iterate the equations until (numerical) convergence. We found sometimes that the solutions can be sensitive to the initialization. This hints at the problem being non-convex or the iteration-scheme being prone to get stuck in local optima or plateaus. In particular, we find that in the serial hierarchy with low cost of computation, a sparse, diagonal-like initialization of $p(x|w)$ works much better than a random initialization. For the parallel hierarchies, we found that a random or sparse initialization of $p(x|w)$ combined with a uniform initialization of $p(a|w, x)$ works most reliably. Additionally we found that in the hierarchical case if β_3 is slightly larger than β_1 the iterations converge to sub-optimal solutions where both pathways are used instead of shifting all the computation to the parallel pathway. The toy simulations presented here are illustrative examples only and numerically efficient implementations of the iteration-schemes are beyond the scope of the current paper. These problems might be addressed by other solution schemes like sampling-based or parametric model-based solutions. Nevertheless, these other solution schemes (that potentially do not even require the sets of analytical solutions) can benefit from the interpretations given by the analytic solution equations in this paper.

The ability to form abstractions is thought of as a hallmark of intelligence, both in cognitive tasks and in basic sensorimotor behaviors (Kemp et al., 2007; Braun et al., 2010a,b; Gershman and Niv, 2010; Tenenbaum et al., 2011; Genewein and Braun, 2012). Traditionally, the formation of abstractions is conceptualized as being computationally costly because particular entities have to be grouped together by neglecting irrelevant information. Recently, abstractions that arise from sensory evolution and hierarchical behaviors have been studied from an information-theoretic perspective (Salge and Polani, 2009; Van Dijk et al., 2011). Here, we study abstractions in the process of decision-making, where “similar” situations elicit the same behavior when partially ignoring the current situational context. Extending our principle to hierarchies with more than two levels might provide novel points of view on the formation of hierarchies in biological systems, such as the early visual system (DiCarlo et al., 2012). One fundamental prediction, based on our current work is that the formation of abstractions and concepts should be heavily shaped by the agent-environment structure (the utility function). Following the work of (Simon, 1972) decision-making with limited information-processing resources has been studied extensively in psychology, economics, political science, industrial organization, computer science, and artificial intelligence research. In

this paper, we use an information-theoretic model of decision-making under resource constraints (McKelvey and Palfrey, 1995; Kappen, 2005; Wolpert, 2006; Todorov, 2009; Peters et al., 2010; Theodorou et al., 2010; Rubin et al., 2012). In particular, Braun et al. (2011) and Ortega and Braun (2011, 2012, 2013) present a framework in which gain in expected utility is traded off against the adaptation cost of changing from an initial behavior to a posterior behavior. The variational problem that arises due to this trade-off has the same mathematical form as the minimization of a *free energy difference* functional in thermodynamics. Here, we discuss the close connection between the thermodynamic decision-making framework (Ortega and Braun, 2013) and rate-distortion theory which is an information-theoretic framework for lossy compression. The problem in lossy compression is essentially the problem of separating structure from noise and is thus highly related to finding abstractions (Tishby et al., 1999; Still and Crutchfield, 2007; Still et al., 2010). In the context of decision-making the rate-distortion framework can be applied by conceptualizing the decision-maker as a channel from observations to actions *with limited capacity*, which is known in economics as the framework of “Rational Inattention” (Sims, 2003).

The rate-distortion principle and all the extended principles presented in this paper measure computational cost with the mutual information which is an abstract measure that quantifies the average KL-divergence. The mutual information measures the actual transformation of probabilities and thus provides a lower bound for any possible implementation. In fact, different implementations could perform the same transformation more or less efficiently which should reflect in the price of information processing but not the amount of information processed. The advantage of using a generic measure is that the principle is universal and can be applied to any system. The downside of this is that it cannot be directly used to analyze specific implementations. In practice it can be hard to determine how difficult or “costly” it is to implement a certain transformation of probability distributions. Rather, the price for information processing is often set implicitly, for instance by certain computation-time constraints or by constraining the number of samples, etc. When applying the principle to a specific implementation it might be required to derive a novel, specific solution scheme for the corresponding optimization problem. In Leibfried and Braun (2015), for instance, the authors apply the rate-distortion principle for decision-making to a spiking neuron model by deriving a gradient-based update rule for tuning the parameters of the model (the weights of the neuron). In their case, the price of information processing β appears directly in the parameter update equations which leads to an interesting regularizer for the (online) parameter update rule.

The fundamental trade-off between large expected utility and low computational cost appears in many domains such as machine learning, AI, economics, computational biology or neuroscience, and many solutions, such as heuristics, sampling-based approaches, and model-based approximation schemes, exist (Gershman et al., 2015; Jordan and Mitchell, 2015; Parkes and Wellman, 2015). One of the exciting prospects of such an approach

is that it might provide a common ground for research-questions from artificial intelligence and neuroscience, thus partially unifying the two fields that share common origins but have drifted apart over the last decades (Gershman et al., 2015). The main contribution of this paper is to advance a principled mathematical framework that formalizes the problem objective such that the trade-off between large expected utility and low computational cost and its solutions can be addressed in both a qualitative but also quantitative way. The main finding is that the consistent application of the principle beyond simple one-stage information processing systems leads to non-trivial solutions that address questions like optimal likelihood model design or the design of optimal decision-making hierarchies. Since the mathematics can easily be extended to more variables while the underlying principle remains the same, we believe that the formulation presented in this paper is a good candidate for a general underlying objective that is also applicable to biological organisms and evolutionary processes. We find the principle an interesting starting point for solving timely problems in machine learning, robotics, and AI but also for providing an interesting novel angle for research in computational neuroscience and biology. The principle also provides a promising basis for the design and analysis of guided self-organizing systems as most of the inner structure of systems following our principle is emergent (and thus self-organized) but ultimately aimed at solving particular tasks (through the utility function).

AUTHORS CONTRIBUTION

TG and DB conceived the project, TG and JG performed simulations, TG and FL did analysis and derivations, and TG, FL, JG, and DB wrote the paper.

FUNDING

This study was supported by the DFG, Emmy Noether grant BR4164/1-1.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at <http://journal.frontiersin.org/article/10.3389/frobt.2015.00027>

A Supplementary Methods provides detailed steps to derive the solution to the general case (Section 5) and how to rewrite the solution equations of the general case. Additionally it outlines how to derive the solutions for the serial and the parallel case. It also provides the set of self-consistent equations for the “degenerate total correlation” and “total correlation” case that drop out mathematically from the general case but are not used in this paper (see **Table 2**). The Supplementary Methods provides details to the discussion on the different information processing pathways of the parallel case (Section 4.3). Finally, it contains the proof for the inequality based on equation (6).

The simulations underlying the results presented in this paper are published as supplementary material using Jupyter (<http://jupyter.org/>) notebooks. The notebooks are considered part of the results of this paper and readers are encouraged

to use the notebooks to interactively explore the examples and concepts presented here. The underlying code is written in Julia (Bezanson et al., 2014) and uses the Gadfly package (<http://gadflyjl.org/>) for visualization. At the time of writing, the notebooks can be run with a local installation of Jupyter and Julia or without any installation in a web-browser through the JuliaBox project (<https://www.juliabox.org/>). The notebooks and code at the time of publication are provided in a supplementary.zip file but also under (Genewein, 2015). The notebooks and the code behind the notebooks as well as information on different methods to run the notebooks will be kept up-to-date in the accompanying GitHub repository: <https://github.com/tgenewein/BoundedRationalityAbstractionAndHierarchicalDecisionMaking>. If compatibility issues with future Julia versions are encountered, please refer to the GitHub repository and feel free to submit an issue. A readme-file on how to run the notebooks (with or without installation) is also provided in the supplementary data as well as in the GitHub repository.

The following notebooks are provided:

- “1-FreeEnergyForBoundedRationalDecisionMaking”: Illustrates the results of Section 1 and reproduces **Figure 1**.
- “2-RateDistortionForDecisionMaking”: Illustrates the results of Section 4 (the recommender system example) and reproduces **Figures 2–4**. The notebook can be used as a general template for setting up any of the examples presented in the paper and solving it using Blahut-Arimoto.
- “S1-SampleBasedBlahutArimoto”: A simple proof-of-concept implementation of sample-based Blahut-Arimoto iterations. Due to space-constraints, this part has been omitted from the paper, but interested readers can find a short theoretical part on the sampling approach in the notebook. Additionally, the notebook shows an implementation of the sampling scheme and applies it to a toy example.
- “3-SerialHierarchy”: Illustrates the comparison between hand-crafted perception and bounded-optimal perception in the serial case (Section 3) using the predator-prey example. The notebook reproduces **Figures 5–8**. The notebook allows to easily modify the parameters (e.g., inverse temperatures) of the example or to switch to a different utility function. It can also be used to see how the parallel or general case solution for the predator-prey example would look like.
- “4-ParallelHierarchy”: Illustrates the emergence of bounded-optimal hierarchies in two different environments of the medical system example as presented in Section 4 and reproduces **Figures 9 and 10**. The notebook can be used to easily explore the different information processing pathways in the parallel case but also to compare any two cases against each other (because it compares two general case solutions and they can be tuned to all of the special cases).
- “5-DistributionOfInformationProcessing”: Compares the parallel hierarchical solution to the medical example to the one-step (rate-distortion) case as shown in **Figure 11**. Since it implements the parallel case through the general case, it also allows to compare any other case to the one-step solution.

REFERENCES

- Arimoto, S. (1972). An algorithm for computing the capacity of arbitrary discrete memoryless channels. *IEEE Trans. Inf. Theory* 18, 14–20. doi:10.1109/TIT.1972.1054753
- Ashby, W. R. (1956). *An Introduction to Cybernetics*. London: Chapman & Hall.
- Bezanson, J., Edelman, A., Karpinski, S., and Shah, V. B. (2014). Julia: a fresh approach to numerical computing. *arXiv preprint arXiv:1411.1607*.
- Bishop, C. M. (2006). “Sampling methods,” in *Pattern Recognition and Machine Learning, Number 4 in Information Science and Statistics*, Chap. 11 (New York: Springer).
- Blahut, R. (1972). Computation of channel capacity and rate-distortion functions. *IEEE Trans. Inf. Theory* 18, 460–473. doi:10.1109/TIT.1972.1054855
- Braun, D. A., Mehring, C., and Wolpert, D. M. (2010a). Structure learning in action. *Behav. Brain Res.* 206, 157–165. doi:10.1016/j.bbr.2009.08.031
- Braun, D. A., Waldert, S., Aertsen, A., Wolpert, D. M., and Mehring, C. (2010b). Structure learning in a sensorimotor association task. *PLoS ONE* 5:e8973. doi:10.1371/journal.pone.0008973
- Braun, D. A., and Ortega, P. A. (2014). Information-theoretic bounded rationality and epsilon-optimality. *Entropy* 16, 4662–4676. doi:10.3390/e16084662
- Braun, D. A., Ortega, P. A., Theodorou, E., and Schaal, S. (2011). “Path integral control and bounded rationality,” in *IEEE Symposium on Adaptive Dynamic Programming and Reinforcement Learning* (Piscataway: IEEE), 202–209.
- Burns, E., Ruml, W., and Do, M. B. (2013). Heuristic search when time matters. *J. Artif. Intell. Res.* 47, 697–740. doi:10.1613/jair.4047
- Camerer, C. (2003). *Behavioral Game Theory: Experiments in Strategic Interaction*. Princeton, NY: Princeton University Press.
- Cover, T. M., and Thomas, J. A. (1991). *Elements of Information Theory*. Hoboken: John Wiley & Sons.
- Csiszar, I. (1974). On the computation of rate-distortion functions. *IEEE Trans. Inf. Theory* 20, 122–124. doi:10.1109/TIT.1974.1055146
- Csiszár, I., and Tusnády, G. (1984). Information geometry and alternating minimization procedures. *Stat. Decis.* 1, 205–237.
- Daniel, C., Neumann, G., and Peters, J. (2012). “Hierarchical relative entropy policy search,” in *International Conference on Artificial Intelligence and Statistics*. La Palma.
- Daniel, C., Neumann, G., and Peters, J. (2013). “Autonomous reinforcement learning with hierarchical REPS,” in *International Joint Conference on Neural Networks*. Dallas.
- DiCarlo, J. J., Zoccolan, D., and Rust, N. C. (2012). How does the brain solve visual object recognition? *Neuron* 73, 415–434. doi:10.1016/j.neuron.2012.01.010
- Fox, C. W., and Roberts, S. J. (2012). A tutorial on variational Bayesian inference. *Artif. Intell. Rev.* 38, 85–95. doi:10.1007/s10462-011-9236-8
- Friston, K. (2010). The free-energy principle: a unified brain theory? *Nat. Rev. Neurosci.* 11, 127–138. doi:10.1038/nrn2787
- Garner, W. R. (1962). *Uncertainty and Structure as Psychological Concepts*. New York: Wiley.
- Genewein, T. (2015). Bounded rationality, abstraction and hierarchical decision-making: an information-theoretic optimality principle: supplementary code (v1.1.0). *Zenodo*. doi:10.5281/zenodo.32410
- Genewein, T., and Braun, D. A. (2012). A sensorimotor paradigm for Bayesian model selection. *Front. Hum. Neurosci.* 6:291. doi:10.3389/fnhum.2012.00291
- Genewein, T., and Braun, D. A. (2013). Abstraction in decision-makers with limited information processing capabilities. *arXiv preprint arXiv:1312.4353*.
- Gershman, S. J., Horvitz, E. J., and Tenenbaum, J. B. (2015). Computational rationality: a converging paradigm for intelligence in brains, minds, and machines. *Science* 349, 273–278. doi:10.1126/science.aac6076
- Gershman, S. J., and Niv, Y. (2010). Learning latent structure: carving nature at its joints. *Curr. Opin. Neurobiol.* 20, 251–256. doi:10.1016/j.conb.2010.02.008
- Gigerenzer, G., and Brighton, H. (2009). Homo heuristicus: why biased minds make better inferences. *Top. Cogn. Sci.* 1, 107–143. doi:10.1111/j.1756-8765.2008.01006.x
- Gigerenzer, G., and Todd, P. M. (1999). *Simple Heuristics That Make Us Smart*. Oxford: Oxford University Press.
- Grau-Moya, J., and Braun, D. A. (2013). Bounded rational decision-making in changing environments. *arXiv preprint arXiv:1312.6726*.
- Horvitz, E. (1988). “Reasoning under varying and uncertain resource constraints,” in *AAAI*, Vol. 88 (Palo Alto: AAAI), 111–116.
- Horvitz, E., and Zilberstein, S. (2001). Computational tradeoffs under bounded resources. *Artif. Intell.* 126, 1–4. doi:10.1016/S0004-3702(01)00051-0
- Horvitz, E. J., Cooper, G. F., and Heckerman, D. E. (1989). “Reflection and action under scarce resources: theoretical principles and empirical study,” in *Proceedings of the 11th International Joint Conference on Artificial Intelligence*, Vol. 2 (Detroit: Morgan Kaufmann Publishers, Inc.), 1121–1127.
- Howes, A., Lewis, R. L., and Vera, A. (2009). Rational adaptation under task and processing constraints: implications for testing theories of cognition and action. *Psychol. Rev.* 116, 717–751. doi:10.1037/a0017187
- Janssen, C. P., Brumby, D. P., Dowell, J., Chater, N., and Howes, A. (2011). Identifying optimum performance trade-offs using a cognitively bounded rational analysis model of discretionary task interleaving. *Top. Cogn. Sci.* 3, 123–139. doi:10.1111/j.1756-8765.2010.01125.x
- Jones, B. D. (2003). Bounded rationality and political science: lessons from public administration and public policy. *J. Public Adm. Res. Theory* 13, 395–412. doi:10.1093/jopart/mug028
- Jordan, M., and Mitchell, T. (2015). Machine learning: trends, perspectives, and prospects. *Science* 349, 255–260. doi:10.1126/science.aaa8415
- Kahneman, D. (2003). Maps of bounded rationality: psychology for behavioral economics. *Am. Econ. Rev.* 93, 1449–1475. doi:10.1257/000282803322655392
- Kappen, H. J. (2005). Linear theory for control of nonlinear stochastic systems. *Phys. Rev. Lett.* 95, 200–201. doi:10.1103/PhysRevLett.95.200201
- Kappen, H. J., Gómez, V., and Opper, M. (2012). Optimal control as a graphical model inference problem. *Mach. Learn.* 87, 159–182. doi:10.1007/s10994-012-5278-7
- Kemp, C., Perfors, A., and Tenenbaum, J. B. (2007). Learning overhypotheses with hierarchical Bayesian models. *Dev. Sci.* 10, 307–321. doi:10.1111/j.1467-7687.2007.00585.x
- Leibfried, F., and Braun, D. A. (2015). A reward-maximizing spiking neuron as a bounded rational decision maker. *Neural Comput.* 27, 1686–1720. doi:10.1162/NECO_a_00758
- Levy, R. P., Reali, F., and Griffiths, T. L. (2009). “Modeling the effects of memory on human online sentence processing with particle filters,” in *Advances in Neural Information Processing Systems* (Vancouver: NIPS), 937–944.
- Lewis, R. L., Howes, A., and Singh, S. (2014). Computational rationality: linking mechanism and behavior through bounded utility maximization. *Top. Cogn. Sci.* 6, 279–311. doi:10.1111/tops.12086
- Lieder, F., Griffiths, T., and Goodman, N. (2012). “Burn-in, bias, and the rationality of anchoring,” in *Advances in Neural Information Processing Systems* (Lake Tahoe: NIPS), 2690–2798.
- Lipman, B. (1995). Information processing and bounded rationality: a survey. *Can. J. Econ.* 28, 42–67. doi:10.2307/136022
- Mattsson, L. G., and Weibull, J. W. (2002). Probabilistic choice and procedurally bounded rationality. *Games Econ. Behav.* 41, 61–78. doi:10.1016/S0899-8256(02)00014-3
- McKelvey, R. D., and Palfrey, T. R. (1995). Quantal response equilibria for normal form games. *Games Econ. Behav.* 10, 6–38. doi:10.1006/game.1995.1023
- Neal, R. M. (2003). Slice sampling. *Ann. Stat.* 31, 705–767. doi:10.1214/aos/1056562461
- Neymotin, S. A., Chadderdon, G. L., Kerr, C. C., Francis, J. T., and Lytton, W. W. (2013). Reinforcement learning of two-joint virtual arm reaching in a computer model of sensorimotor cortex. *Neural Comput.* 25, 3263–3293. doi:10.1162/NECO_a_00521
- Ortega, P., and Braun, D. (2010). “A conversion between utility and information,” in *Third Conference on Artificial General Intelligence (AGI 2010)* (Lugano: Atlantis Press), 115–120.
- Ortega, P. A., and Braun, D. A. (2014). Generalized Thompson sampling for sequential decision-making and causal inference. *Complex Adapt. Syst. Model.* 2, 269–274. doi:10.1186/2194-3206-2-2
- Ortega, P. A., Braun, D. A., and Tishby, N. (2014). “Monte Carlo methods for exact & efficient solution of the generalized optimality equations,” in *Proceedings of IEEE International Conference on Robotics and Automation*. Hong Kong.
- Ortega, P. A., and Braun, D. A. (2011). “Information, utility and bounded rationality,” in *Proceedings of the 4th International Conference on Artificial General Intelligence* (Mountain View: Springer-Verlag), 269–274.
- Ortega, P. A., and Braun, D. A. (2012). “Free energy and the generalized optimality equations for sequential decision making,” in *Journal of Machine Learning Research: Workshop and Conference Proceedings* (Edinburgh: JMLR W&C Proceedings), 1–10.

- Ortega, P. A., and Braun, D. A. (2013). Thermodynamics as a theory of decision-making with information-processing costs. *Proc. R. Soc. A Math. Phys. Eng. Sci.* 469, 2153.
- Palmer, S. E., Marre, O., Berry, M. J., and Bialek, W. (2015). Predictive information in a sensory population. *Proc. Natl. Acad. Sci. U.S.A.* 112(22), 6908–6913. doi:10.1073/pnas.1506855112
- Parkes, D. C., and Wellman, M. P. (2015). Economic reasoning and artificial intelligence. *Science* 349, 267–272. doi:10.1126/science.aaa8403
- Peters, J., Mülling, K., and Altun, Y. (2010). “Relative entropy policy search,” in AAAI. Atlanta.
- Ramsey, F. P. (1931). “Truth and probability,” in *The Foundations of Mathematics and Other Logical Essays*, ed. R. B. Braithwaite (New York, NY: Harcourt, Brace and Co), 156–198.
- Rawlik, K., Toussaint, M., and Vijayakumar, S. (2012). “On stochastic optimal control and reinforcement learning by approximate inference,” in *Proceedings Robotics: Science and Systems*. Sydney.
- Rubin, J., Shamir, O., and Tishby, N. (2012). “Trading value and information in mdps,” in *Decision Making with Imperfect Decision Makers* (Springer), 57–74.
- Rubinstein, A. (1998). *Modeling Bounded Rationality*. Cambridge: MIT Press.
- Russell, S. (1995). “Rationality and intelligence,” in *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence*, ed. C. Mellish (San Francisco, CA: Morgan Kaufmann), 950–957.
- Russell, S. J., and Norvig, P. (2002). *Artificial Intelligence: A Modern Approach*. Upper Saddle River: Prentice Hall.
- Russell, S. J., and Subramanian, D. (1995). Provably bounded-optimal agents. *J. Artif. Intell. Res.* 2, 575–609.
- Salge, C., and Polani, D. (2009). Information-driven organization of visual receptive fields. *Adv. Complex Syst.* 12, 311–326. doi:10.1142/S0219525909002234
- Sanborn, A. N., Griffiths, T. L., and Navarro, D. J. (2010). Rational approximations to rational models: alternative algorithms for category learning. *Psychol. Rev.* 117, 1144. doi:10.1037/a0020511
- Savage, L. J. (1954). *The Foundations of Statistics*. New York: Wiley.
- Shannon, C. E. (1948). A mathematical theory of communication. *Bell Syst. Tech. J.* 27, 379–423, 623–656. doi:10.1002/j.1538-7305.1948.tb00917.x
- Simon, H. A. (1955). A behavioral model of rational choice. *Q. J. Econ.* 69, 99–118. doi:10.2307/1884852
- Simon, H. A. (1972). Theories of bounded rationality. *Decis. Organ.* 1, 161–176.
- Sims, C. A. (2003). Implications of rational inattention. *J. Monet. Econ.* 50, 665–690. doi:10.1016/S0304-3932(03)00029-1
- Sims, C. A. (2005). “Rational inattention: a research agenda,” in *Deutsche Bundesbank Spring Conference, Number 4*. Berlin.
- Sims, C. A. (2006). Rational inattention: beyond the linear-quadratic case. *Am. Econ. Rev.* 96, 158–163. doi:10.1257/00028280677212431
- Sims, C. A. (2010). “Rational inattention and monetary economics,” in *Handbook of Monetary Economics*, Vol. 3, Chap. 4 (Elsevier), 155–181.
- Spiegler, R. (2011). *Bounded Rationality and Industrial Organization*. Oxford: Oxford University Press.
- Still, S. (2009). Information-theoretic approach to interactive learning. *Europhys. Lett.* 85, 28005. doi:10.1209/0295-5075/85/28005
- Still, S. (2014). “Lossy is lazy,” in *Workshop on Information Theoretic Methods in Science and Engineering* (Helsinki: University of Helsinki), 17–21.
- Still, S., and Crutchfield, J. P. (2007). Structure or noise? *arXiv preprint arXiv:0708.0654*.
- Still, S., Crutchfield, J. P., and Ellison, C. J. (2010). Optimal causal inference: estimating stored information and approximating causal architecture. *Chaos* 20, 037111. doi:10.1063/1.3489885
- Studený, M., and Vejnarová, J. (1998). “The multiinformation function as a tool for measuring stochastic dependence,” in *Learning in Graphical Models* (New York: Springer), 261–297.
- Tenenbaum, J. B., Kemp, C., Griffiths, T. L., and Goodman, N. D. (2011). How to grow a mind: statistics, structure, and abstraction. *Science* 331, 1279–1285. doi:10.1126/science.1192788
- Theodorou, E., Buchli, J., and Schaal, S. (2010). A generalized path integral control approach to reinforcement learning. *J. Mach. Learn. Res.* 11, 3137–3181.
- Tishby, N., Pereira, F. C., and Bialek, W. (1999). “The information bottleneck method,” in *The 37th Annual Allerton Conference on Communication, Control, and Computing*.
- Tishby, N., and Polani, D. (2011). “Information theory of decisions and actions,” in *Perception-Action Cycle*, Chap. 19 (New York: Springer), 601–636.
- Tkačik, G., and Bialek, W. (2014). Information processing in living systems. *arXiv preprint arXiv:1412.8752*.
- Todorov, E. (2007). “Linearly-solvable Markov decision problems,” in *Advances in Neural Information Processing Systems* (Vancouver: NIPS), 1369–1376.
- Todorov, E. (2009). Efficient computation of optimal actions. *Proc. Natl. Acad. Sci. U.S.A.* 106, 11478–11483. doi:10.1073/pnas.0710743106
- Tversky, A., and Kahneman, D. (1974). Judgment under uncertainty: heuristics and biases. *Science* 185, 1124–1131. doi:10.1126/science.185.4157.1124
- van Dijk, S. G., and Polani, D. (2013). Informational constraints-driven organization in goal-directed behavior. *Adv. Complex Syst.* 16:1350016. doi:10.1142/S0219525913500161
- Van Dijk, S. G., Polani, D., and Nehaniv, C. L. (2011). “Hierarchical behaviours: getting the most bang for your bit,” in *Advances in Artificial Life: Darwin Meets von Neumann*, eds. R. Goebel, J. Siekmann, and W. Wahlster (New York: Springer), 342–349.
- Von Neumann, J., and Morgenstern, O. (1944). *Theory of Games and Economic Behavior*. Princeton: Princeton University Press.
- Vul, E., Alvarez, G., Tenenbaum, J. B., and Black, M. J. (2009). “Explaining human multiple object tracking as resource-constrained approximate inference in a dynamic probabilistic model,” in *Advances in Neural Information Processing Systems* (New York: Wiley), 1955–1963.
- Vul, E., Goodman, N., Griffiths, T. L., and Tenenbaum, J. B. (2014). One and done? Optimal decisions from very few samples. *Cogn. Sci.* 38, 599–637. doi:10.1111/cogs.12101
- Watanabe, S. (1960). Information theoretical analysis of multivariate correlation. *IBM J. Res. Dev.* 4, 66–82. doi:10.1147/rd.41.0066
- Wiener, N. (1961). *Cybernetics or Control and Communication in the Animal and the Machine*, Vol. 25. Cambridge: MIT press.
- Wolpert, D. H. (2006). “Information theory—the bridge connecting bounded rational game theory and statistical physics,” in *Complex Engineered Systems*, eds D. Braha, A. A. Minai, and Y. Bar-Yam (New York: Springer), 262–290.
- Yeung, R. W. (1991). A new outlook on Shannon’s information measures. *IEEE Trans. Inf. Theory* 37, 466–474. doi:10.1109/18.79902
- Yeung, R. W. (2008). *Information Theory and Network Coding*. New York: Springer.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 Genewein, Leibfried, Grau-Moya and Braun. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.