Entropy **2014**, *16*, 4662-4676; doi:10.3390/e16084662

OPEN ACCESS

entropy

ISSN 1099-4300

www.mdpi.com/journal/entropy

Article

Information-Theoretic Bounded Rationality and ϵ-Optimality

Daniel A. Braun [1,*] and Pedro A. Ortega [2,*]

[1] Max Planck Institute for Biological Cybernetics, Max Planck Institute for Intelligent Systems, Spemannstrasse 38, Tübingen 72076, Germany

[2] GRASP Laboratory, Electrical and Systems Engineering Department, University of Pennsylvania, Philadelphia, PA 19104, USA

* Authors to whom correspondence should be addressed;
E-Mails: daniel.braun@tuebingen.mpg.de (D.A.B.); ope@seas.upenn.edu (P.A.O.).

Received: 19 July 2014; in revised form: 11 August 2014 / Accepted: 15 August 2014 / Published: 21 August 2014

Abstract: Bounded rationality concerns the study of decision makers with limited information processing resources. Previously, the free energy difference functional has been suggested to model bounded rational decision making, as it provides a natural trade-off between an energy or utility function that is to be optimized and information processing costs that are measured by entropic search costs. The main question of this article is how the information-theoretic free energy model relates to simple ϵ-optimality models of bounded rational decision making, where the decision maker is satisfied with any action in an ϵ-neighborhood of the optimal utility. We find that the stochastic policies that optimize the free energy trade-off comply with the notion of ϵ-optimality. Moreover, this optimality criterion even holds when the environment is adversarial. We conclude that the study of bounded rationality based on ϵ-optimality criteria that abstract away from the particulars of the information processing constraints is compatible with the information-theoretic free energy model of bounded rationality.

Keywords: bounded rationality; ϵ-optimality; probabilistic choice; ambiguity

1. Introduction

Decision making under uncertainty is studied by means of optimal actor models in a broad spectrum of sciences with remarkably different historical roots, like economics, artificial intelligence research, biology, sociology, and even fields, like legal studies, ethics and philosophy [1–3]. Usually, when we talk about decision making, we imagine a human mind (for example, a chess player) that ponders a variety of possible options for action, deliberates about their potential outcomes and finally picks one of these actions for execution; namely, the one that is expected to have the most beneficial consequences. Recently, the same paradigm has also been extended to model sensorimotor integration and control [4–6], where consequences of actions can be anticipated by implicit learning processes. Crucially, however, in either case, classic decision-theoretic models [3,7] ignore the details of the underlying cognitive or implicit processes preceding a decision by simply assuming that these processes optimize a performance criterion. This ignorance is both boon and bane, as, on the one hand, it allows the statement of many general results that do not depend on the details of the decision making process, but on the other hand, the often unrealistic assumption of perfect optimization limits the applicability of classic decision theory.

Classic decision theory rests on two conceptual pillars: the notion of probability and the notion of utility. Their intertwined occurrence may be best understood on the basis of the concept of lotteries. A lottery is defined as a set of N different outcomes $o_j \in \mathcal{O}$ each of which can occur with a respective probability $P(o_j)$ where $j = 1, \dots, N$. We can imagine a lottery as a roulette wheel or a gamble where we obtain a prize o_j with probability $P(o_j)$ that has a subjective utility $U(o_j)$ for the decision maker. The compound value of the lottery can then be determined by the expected utility $\mathbf{E}[U] = \sum_j P(o_j)U(o_j)$, which is commonly used as the standard performance criterion in decision making. The concept of expected utility was first axiomatized by Neumann and Morgenstern [8]. In their axiomatic system, Neumann and Morgenstern [8] define a binary preference relation $\succ$ over the set of probability distributions $\mathcal{P}$ defined over the set of outcomes $\mathcal{O}$. If (and only if) this binary relation satisfies the axioms of completeness, transitivity, continuity and independence, then there exists a function $U : \mathcal{O} \mapsto \mathbb{R}$, such that:

$$P \succ P' \quad \Longleftrightarrow \quad \sum_j P(o_j)U(o_j) > \sum_j P'(o_j)U(o_j),$$

where $P, P' \in \mathcal{P}$. This utility function U is unique up to a positive affine transform.

When designing optimal actors, most designers use the Neumann and Morgenstern [8] conception of probability and utility; see for example Russell and Norvig [2]. Such optimal actors are typically equipped with a probabilistic model of the world $P(o_j|a_i)$, where $a_i \in \mathcal{A}$ is an action that leads to consequence o_j with probability $P(o_j|a_i)$. The decision maker can assess the expected utility of each action as $\mathbf{E}[U|a_i] = \sum_j P(o_j|a_i)U(o_j)$. Thus, the probabilistic model of the world defines a set of M different lotteries indexed by a_i, where $i = 1, \dots, M$. The decision maker can compare the expected utilities of all the lotteries and choose the one with the highest expected utility, such that:

$$a_{max} = \arg\max_i \mathbf{E}[U|a_i]. \tag{1}$$

However, there are at least two important assumptions. First, the decision maker requires an accurate probability model. Second, the decision maker requires enough computational resources to find the best lottery. What happens if one of the two assumptions is violated? This question has spurred research on bounded rationality where decision makers have limited knowledge and bounded computational resources.

The modern study of bounded rationality began with Herbert Simon [9–11] and has since been continued in economics [12–14], game theory [15–17], industrial organization [18] and political science [19], but also in psychology [20,21], cognitive science [22–24], computer science and artificial intelligence research [25–27]. One of the fundamental questions faced by bounded rationality models is whether they should attend to the actual physical or cognitive processes underlying decision making or whether it is also possible to gain a more general understanding of bounded rational decision making by abstracting away from the details of the actual decision making process. While the first approach is taken, for example, by the new field of neuroeconomics relating decision making processes to anatomical structures [28,29], one of the simplest approaches in the second tradition is the concept of ϵ-optimality [30], where the decision maker does not search for a single best action a_{max}, but for any action from a set of permissible actions $\mathcal{A}^{\epsilon}$ whose expected utility deviates at most by $\epsilon > 0$ from the optimal expected utility of a_{max}, such that:

$$\mathcal{A}^{\epsilon} = \{a_i \in \mathcal{A} : \mathbf{E}[U|a_i] \geq \mathbf{E}[U|a_{max}] - \epsilon\} \,. \tag{2}$$

The main question of this article is how to relate this simple model of bounded rationality to the information-theoretic bounded rationality model discussed in Ortega and Braun [31–34] that we recapitulate in the next section.

2. Methods

Most models of decision making ignore information processing costs and assume that the decision maker can simply handpick the action that yields the highest (expected) utility. Presupposing that there is a unique maximum, this would correspond to a deterministic strategy as in Equation (1). In general, however, a decision maker with limited information processing capabilities might be unable to handpick the best option with certainty. Such a bounded rational strategy must therefore be described by a probability distribution $P(a_i)$ reflecting this uncertainty. Information-theoretic models of bounded rational decision making quantify the cost of information-processing by entropic measures of information [15–17,31–35] and are closely related to softmax-choice rules that have been extensively studied in the psychological and econometric literature, but also in the literature on reinforcement learning and game theory [36–42]. In [31–34], Ortega and Braun discuss an information-theoretic model of bounded rational decision making where information processing costs are quantified by the relative entropy with the idea that information processing costs can then be measured with respect to changes in the choice strategy $P(a_i)$.

Let us assume that the initial strategy of the decision maker can be described by a probability distribution $P_0(a_i)$. This could include the uniform distribution over a_i as a special case, if the decision maker has no prior preferences between different actions. Next, this decision maker is exposed to a utility function $V(a_i)$, which includes the case of $V(a_i) = \mathbf{E}[U|a_i]$, implying that the decision maker does not have to compute the expectation values, but the expectation values are simply given. Ideally, the decision maker will arrive at the new distribution $P(a_i) = \delta_{a_i,a_{max}}$. The underlying computation can be imagined as a search process that reduces the uncertainty over the action by $D_{KL}[P||P_0] = \sum_i P(a_i) \log\left[P(a_i)/P_0(a_i)\right]$. In general, such a search is costly, and the decision maker might not be able to afford such a stark reduction in uncertainty. Assuming a price $1/\alpha$ for $1 bit$ of information gain, we can then design a bounded optimal decision maker that trades off gains in utility resulting from changes in $P(a_i)$ against the search costs that these changes imply, such that, overall, the decision maker optimizes a free energy difference in utility gains and information costs:

$$\Delta F[\tilde{P}] = \left\{ \sum_i \tilde{P}(a_i)V(a_i) - \frac{1}{\alpha}\sum_i \tilde{P}(a_i)\log\frac{\tilde{P}(a_i)}{P_0(a_i)} \right\}, \tag{3}$$

where the maximizing distribution $P = \arg\max_{\tilde{P}} \Delta F[\tilde{P}]$ is the equilibrium distribution:

$$P(a_i) = \frac{1}{Z_\alpha} P_0(a_i) e^{\alpha V(a_i)}, \qquad \text{where } Z_\alpha = \sum_i P_0(a_i) e^{\alpha V(a_i)}, \tag{4}$$

and represents the choice probabilities after deliberation. Note that the free energy difference $\Delta F[\tilde{P}]$ can be expressed as $\Delta F[\tilde{P}] = F_1[\tilde{P}] - F_0$, with the free energies:

$$\begin{aligned} F_1[\tilde{P}] &= \sum_i \tilde{P}(a_i)\Phi_1(a_i) - \frac{1}{\alpha}\sum_i \tilde{P}(a_i)\log\tilde{P}(a_i) \\ F_0 &= \sum_i P_0(a_i)\Phi_0(a_i) - \frac{1}{\alpha}\sum_i P_0(a_i)\log P_0(a_i), \end{aligned}$$

where $P_0(a_i) = \exp\left(\alpha(\Phi_0(a_i) - F_0)\right)$ and $V(a_i) = \Phi_1(a_i) - \Phi_0(a_i)$. Hence, the utility function $V(a_i)$ expresses changes in value Φ, that are gains or losses with respect to the status quo. In the case of inference, the utility function is given by a negative log-likelihood and measures informational surprise. The temperature parameter corresponds then to a precision parameter in exponential family distributions. Casting the problem of acting as an inference problem has been previously discussed in [43–48]. The certainty-equivalent value V_{CE} under strategy P can be determined from the same variational principle:

$$\begin{aligned} V_{CE} &= \max_{\tilde{P}} \left\{ \sum_i \tilde{P}(a_i)V(a_i) - \frac{1}{\alpha}\sum_i \tilde{P}(a_i)\log\frac{\tilde{P}(a_i)}{P_0(a_i)} \right\} \\ &= \frac{1}{\alpha}\log\left(\sum_i P_0(a_i)e^{\alpha V(a_i)} \right) = \frac{1}{\alpha}\log Z_\alpha. \end{aligned} \tag{5}$$

For the two different limits of α, the value and the equilibrium distribution take the asymptotic forms:

$$\begin{array}{llll} \alpha \to +\infty & \frac{1}{\alpha}\log Z_\alpha = \max\limits_i V(a_i) & P(a_i) = \delta_{a_i,a_{max}} & \text{(perfectly rational)} \\ \alpha \to 0 & \frac{1}{\alpha}\log Z_\alpha = \sum\limits_i P_0(a_i)V(a_i) & P(a_i) = P_0(a_i) & \text{(irrational)} \end{array}$$

It can be seen that a perfectly rational agent with $\alpha \to \infty$ is able to handpick the optimal action, which is a deterministic policy in the case of a unique optimum, whereas finitely rational agents have stochastic policies with a non-zero probability of picking a sub-optimal action.

In the case that $V(a_i)$ are not simply given, the decision maker has to compute the expectation values herself from the prior $P_0(o_j|a_i)$ and the utility $U(o_j)$, such that search costs have to be considered both for a_i and o_j. The variational problem can then be formulated as a nested expression [32,34,49]:

$$\arg\max_{\tilde{P}} \sum_i \tilde{P}(a_i)\left[-\frac{1}{\alpha}\log\frac{\tilde{P}(a_i)}{P_0(a_i)} + \sum_j \tilde{P}(o_j|a_i)\left[U(o_j) - \frac{1}{\beta}\log\frac{\tilde{P}(o_j|a_i)}{P_0(o_j|a_i)} \right]\right]. \tag{6}$$

If we assume that the estimation of the expected utilities $V(a_i)$ is much cheaper than the calculation of the optimal action, then the price $1/\beta$ should be much lower than $1/\alpha$, such that $\alpha \gg \beta$, implying that we can simply obtain samples from $P_0(o_j|a_i)$ for our computation of the expectation, but that it is much more difficult to compute a_i, because we cannot simply rely on our prior $P_0(a_i)$. The two-part solution to the nested variational problem is given by:

$$P(o_j|a_i) = \frac{1}{Z_\beta(a_i)} P_0(o_j|a_i) \exp\left(\beta U(o_j)\right) \tag{7}$$

with the normalization constant: $Z_\beta(a_i) = \sum_j P_0(o_j|a_i) \exp\left(\beta U(o_j)\right)$ and:

$$P(a_i) \quad = \quad \frac{1}{Z_{\alpha\beta}} P_0(a_i) \exp\left(\frac{\alpha}{\beta} \log Z_\beta(a_i)\right) \tag{8}$$

with the normalization constant: $Z_{\alpha\beta} = \sum_i P_0(a_i) \exp\left(\frac{\alpha}{\beta} \log Z_\beta(a_i)\right)$. The perfectly rational decision maker is obtained in the limit $\alpha \to \infty$ and $\beta \to 0$, that is:

$$\begin{aligned} P(o_j|a_i) &= P_0(o_j|a_i) \\ P(a_i) &= \delta_{a_i, a_{max}}. \end{aligned} \tag{9}$$

The computational complexity of the information-theoretic model of bounded rational decision making can also be interpreted in terms of a sampling complexity [50,51]. In particular, Equation (4) can be interpreted under a rejection sampling scheme where we want to obtain samples from $P(a_i)$, but we are only able to sample from the distribution $P_0(a_i)$. In this scheme, we generate a sample $a_i \sim P_0(a_i)$ and then accept the sample if:

$$u \leq \frac{e^{\alpha V(a_i)}}{e^{\alpha T}}, \tag{10}$$

where u is drawn from the uniform $\mathcal{U}[0;1]$ and T is the acceptance target value with $T \geq \max_i V(a_i)$. Otherwise, the sample is rejected. The efficiency of the sampling process depends on how many samples we will need on average from P_0 to obtain one sample from P. This average number of samples from P_0 needed for one sample of P is given by the mean of a geometric distribution:

$$\overline{\sharp Samples} \;=\; \frac{1}{\sum_i P_0(a_i) \frac{e^{\alpha V(a_i)}}{e^{\alpha T}}} = \frac{e^{\alpha T}}{Z_\alpha}. \tag{11}$$

It is important to note that the average number of samples increases exponentially with increasing the rationality parameter, such that:

$$\frac{e^{\alpha T}}{Z_\alpha} \xrightarrow{\alpha \to \infty} \frac{e^{\alpha(T - V(a_{max}))}}{P_0(a_{max}))},$$

where $a_{max} = \arg\max V(x)$ and $T > \max_i U(a_i)$.

This interpretation in terms of sampling complexity can also be extended to Equation (6), where the decision maker has to estimate the expected utilities from samples. In line with Equation (8), we should accept a sample $a_i \sim P_0(a_i)$ if it fulfils the criterion:

$$u \leq \frac{e^{\alpha \frac{1}{\beta} \log Z_\beta(a_i)}}{e^{\alpha T}} = \left[\frac{Z_\beta(a_i)}{e^{\beta T}}\right]^{\frac{\alpha}{\beta}}, \tag{12}$$

where $u \sim \mathcal{U}[0;1]$ and $T \geq \frac{1}{\beta}\log Z_\beta(a_i)$. From Equation (11), we know that the ratio $Z_\beta(a_i)/e^{\beta T}$ can be interpreted as an acceptance probability; in this case, the acceptance probability of $\theta \sim P_0(\theta)$. Thus, in order to accept one sample from x, we need to accept $\frac{\alpha}{\beta}$ consecutive samples of θ, with acceptance criterion:

$$u \leq \frac{e^{\beta U(x,\theta)}}{e^{\beta T}} \tag{13}$$

with $u \sim \mathcal{U}[0;1]$ and T as set above.

3. Results

Here, we investigate the question of how close a bounded rational decision maker gets to the optimal (expected) utility achieved by the perfectly rational decision maker. Since we assume that the strategy of a bounded rational decision maker is inherently stochastic and can be described by a probability distribution according to Equation (4), we can only compare some statistical measure of the performance of the bounded rational decision maker to the performance of the perfectly rational decision maker. In the following, we will consider the expected performance.

Theorem 1 (ϵ-Optimality)**.** *Given a bounded rational decision maker with information cost $1/\alpha$ that optimizes (3), one can bound the expected performance of this decision maker from below within an ϵ-neighborhood of the optimal performance $V_{max} = \max_i \mathbf{E}[U|a_i]$ of the perfectly rational decision maker, such that:*

$$\sum_i P(a_i)V(a_i) \quad \geq \quad V_{max} - \underbrace{\left(-\frac{1}{\alpha}\log P_0(a_{max})\right)}_{=:\epsilon}.$$

Proof. The certainty-equivalent value V_{CE} under the bounded rational strategy $P(a_i)$ is given by:

$$\begin{aligned} V_{CE} &= \frac{1}{\alpha}\log\sum_i P_0(a_i)e^{\alpha V(a_i)} \\ &= \sum_i P(a_i)V(a_i) - \frac{1}{\alpha}\underbrace{\sum_i P(a_i)\log\frac{P(a_i)}{P_0(a_i)}}_{\geq 0}, \end{aligned}$$

where $P(a_i) = \frac{1}{Z}P_0(a_i)e^{\alpha V(a_i)}$. From the positiveness of the Kullback–Leibler divergence, it follows that:

$$\begin{aligned} \sum_i P(a_i)V(a_i) &\geq \frac{1}{\alpha}\log\sum_i P_0(a_i)e^{\alpha V(a_i)} \\ \Rightarrow \sum_i P(a_i)V(a_i) &\geq \frac{1}{\alpha}\log P_0(a_{max})e^{\alpha V_{max}} \\ \Rightarrow \sum_i P(a_i)V(a_i) &\geq V_{max} + \frac{1}{\alpha}\log P_0(a_{max}) \end{aligned}$$

□

As a corollary, we can conclude for the special case of uniform prior $P_0(a_i) = 1/M$ that the ϵ-bound is given by $\epsilon = 1/\alpha \log M$. Conversely, given an $\epsilon > 0$, there exists an $\bar{\alpha} = \frac{\log M}{\epsilon}$, such that for $\alpha \geq \bar{\alpha}$, any decision taken yields a utility within epsilon of the optimum.

In the case of (6), the bounded rational decision maker has to determine the expected utilities by sampling, and the above lower bound cannot be guaranteed anymore. Instead of the expected utilities $V(a_i) = \mathbf{E}[U|a_i]$, such a decision maker optimizes the "distorted" certainty-equivalent value:

$$\tilde{V}(a_i) = \frac{1}{\beta} \log Z_\beta(a_i) = \frac{1}{\beta} \log \sum_j P_0(o_j|a_i) e^{\beta U(o_j)},$$

with $Z_\beta(a_i)$ from Equation (7). Only for $\beta \to 0$, the expectation value $\tilde{V}(a_i) \to \mathbf{E}[U|a_i]$ is retained. Due to $\frac{1}{\beta} \log Z_\beta(a_i) \geq \mathbf{E}[U|a_i]$, such a decision maker with positive β will overestimate the certainty-equivalent value for sub-optimal actions a_i. For small $\beta \ll 1$, the certainty-equivalent value can be approximated by a Taylor expansion in β:

$$\frac{1}{\beta} \log \sum_j P_0(o_j|a_i) e^{\beta U(o_j)} \approx \mathbf{E}_{P_0(o_j|a_i)}[U(o_j)] + \frac{\beta}{2} \mathbb{VAR}_{P_0(o_j|a_i)}[U(o_j)] + O(\beta^2),$$

where $O(\beta^2)$ are higher-order cumulants that can be neglected. Due to Theorem 1, we have:

$$\sum_i P(a_i) \left[\frac{1}{\beta} \log \sum_j P_0(o_j|a_i) e^{\beta U(o_j)} \right] \geq V_{max} + \frac{1}{\alpha} \log P_0(a_{max}),$$

from which we can conclude for the limit $\beta \ll 1$ and $\alpha \gg \beta$ that:

$$\sum_i P(a_i) V(a_i) \geq V_{max} - \underbrace{\left(-\frac{1}{\alpha} \log P_0(a_{max}) + \frac{\beta}{2} \mathbf{E}_{P(a_i)} \left[\mathbb{VAR}_{P_0(o_j|a_i)}[U(o_j)] + O(\beta^2) \right] \right)}_{=:\epsilon}.$$

For such a bounded rational decision maker, the error bound is increased by higher order cumulants.

If all of the (expected) utilities $V(a_i)$ are very similar in magnitude, it requires a high rationality parameter α to differentiate between them. A tighter ϵ-bound in α can be given, if we assume that there is an interval $V(a_i) \in [V_{\min}; V_{\max}]$ and that all the utilities are discriminable by at least one "utile", such that for any choice a_i and a_k, we have $|V(a_i) - V(a_k)| \geq 1$, which is the case, for example, when utilities reflect rank.

Theorem 2 (ϵ-Optimality for rank utilities)**.** *Given a bounded rational decision maker with information cost $1/\alpha$ that optimizes Equation (3) and assuming a uniform prior $P_0(a_i) = 1/M$, bounded (expected) utilities $V(a_i) \in [V_{\min}; V_{\max}]$ for all i and $|V(a_i) - V(a_k)| \geq 1$ for every pair (i,k), one can bound the expected performance of this decision maker from below within an ϵ-neighborhood of the optimal performance $V_{max} = \max_i \mathbf{E}[U|a_i]$ of the perfectly rational decision maker, such that:*

$$\sum_i P(a_i) V(a_i) \geq V_{max} - \underbrace{\left(e^{-\alpha} \left(V_{max} - V_{min} \right) \right)}_{=:\epsilon}.$$

Proof. We express the choice probability $P(a_i)$ derived from Equation (4) under uniform prior $P_0(a_i) = 1/M$ as:

$$P(a_i) = \frac{e^{\alpha V(a_i)}}{\sum_k e^{\alpha V(a_k)}} = \frac{\left(\frac{1}{\delta}\right)^{V(a_i)}}{\sum_k \left(\frac{1}{\delta}\right)^{V(a_k)}},$$

where we have introduced the variable $\delta = \exp(-\alpha)$. We can then express the expected performance as:

$$\begin{aligned}\sum_i P(a_i)V(a_i) &= \frac{1}{\sum_k \left(\frac{1}{\delta}\right)^{V(a_k)}} \sum_i \left(\frac{1}{\delta}\right)^{V(a_i)} V(a_i)\\ &\geq \left(\frac{\left(\frac{1}{\delta}\right)^{V_{max}}}{\sum_k \left(\frac{1}{\delta}\right)^{V(a_k)}}\right) V_{max} + \left(1 - \frac{\left(\frac{1}{\delta}\right)^{V_{max}}}{\sum_k \left(\frac{1}{\delta}\right)^{V(a_k)}}\right) V_{min}\\ &\geq V_{max} - \left(1 - \frac{\left(\frac{1}{\delta}\right)^{V_{max}}}{\sum_k \left(\frac{1}{\delta}\right)^{V(a_k)}}\right) \big(V_{max} - V_{min}\big), \end{aligned} \tag{14}$$

where the inequality is obtained by taking out the largest summand and then finding a lower bound for the remaining terms. The second summand in the last equality can be further delimited as:

$$1 - \frac{\left(\frac{1}{\delta}\right)^{V_{max}}}{\sum_k \left(\frac{1}{\delta}\right)^{V(a_k)}} = 1 - \frac{1}{\sum_k \delta^{V_{max}-V(a_k)}} \leq \delta,$$

since we can limit $\sum_k \delta^{V_{max}-V(a_k)} \leq \sum_k \delta^k \leq \frac{1}{1-\delta}$ from $|V(a_i) - V(a_k)| \geq 1 \, \forall i, k$ and the limit properties of the geometric series. Therefore, we have:

$$\sum_i P(a_i)V(a_i) \geq V_{max} - \delta \left(V_{max} - V_{min}\right).$$

□

As a corollary, we can conclude in the case of minimal interval size $[V_{min}; V_{max}] = [V_{min}; V_{min} + M]$ that the performance bound is given by $\sum_i P(a_i)V(a_i) \geq V_{max} - e^{-\alpha}M$. Conversely, given an $\epsilon > 0$, there exists an $\bar{\alpha} = \log \frac{V_{max}-V_{min}}{\epsilon}$, such that for $\alpha \geq \bar{\alpha}$, any decision made yields a utility within epsilon of the optimum.

4. Adversarial Environments

So far, we have considered stochasticity in action selection to arise due to limited computational power, even in the absence of any uncertainty in the environment. Naturally, in this setting, stochastic choice yields less (expected) utility than deterministic choice of the best option, but the performance decrement can be bounded by ϵ. If, however, the environment is potentially adversarial, stochastic action selection can also be superior in terms of utility alone, since it does not allow the opponent to perfectly predict and thwart any deterministic action plan that the decision maker might have. In the following, we will discuss two different scenarios for decision making in adversarial environments, where the decision maker chooses between different actions $a_i \in \mathcal{A}$ with (expected) utility $V(a_i) = \mathbf{E}[U|a_i]$.

4.1. Unknown Action Set

In the first scenario, we assume that the decision maker starts by choosing a probability distribution $P(a_i)$ over actions $a_i \in \mathcal{A}$, and then, the environment chooses a subset $\mathcal{S} \in \mathcal{P}(\mathcal{A})\backslash\{\}$ of permissible actions, where $\mathcal{P}(\mathcal{A})$ denotes the powerset. All actions that are not part of the subset are eliminated. Finally, the action a_i is randomly determined from the set of permissible actions with their renormalized probabilities. The problem is to find the betting probability $P(a_i)$ such that we maximize our expected return; however, the expectation has to be taken over the unknown subset $\mathcal{S}$ capriciously chosen by the opponent. This models a decision maker, who has to choose a generic hedging strategy by allocating resources to different alternatives, but where the rules of the game are only fully revealed after the choice is made. Formally, we want to choose the probability $P(a_i)$, such that the conditional expectation $\mathbf{E}[V(a_i)|\mathcal{S}]$ is as large as possible. Unsurprisingly, we cannot provide a deterministic optimal solution $P(a_i) = \delta(a_i - a^*)$, since the environment could always eliminate a^*. However, if we allow ourselves an arbitrarily small, non-zero performance loss $\epsilon > 0$, then there is a way to assign probabilities $P(a_i)$, such that the conditional expectation is almost equal to the optimum, *i.e.*, to the highest utility in the subset chosen by the opponent. This is precisely the result of the following theorem.

Theorem 3 (ϵ-Optimality in adversarial environments)**.** *The expected utility achieved by a bounded rational decision maker that optimizes* (3) *lies within an ϵ-neighborhood of the optimal utility $V^{\mathcal{S}}_{max} = \max_{a_i \in \mathcal{S}} V(a_i)$ in $\mathcal{S}$ for any subset $\mathcal{S}$ of possible actions selected by nature, such that:*

$$\begin{aligned} \frac{1}{\sum_{a_k \in \mathcal{S}} P(a_k)} \sum_{a_i \in \mathcal{S}} P(a_i)V(a_i) &\geq V^{\mathcal{S}}_{max} - \left(-\frac{1}{\alpha}\log P_0(a^{\mathcal{S}}_{max})\right) \\ &=: V^{\mathcal{S}}_{max} - \epsilon. \end{aligned}$$

Proof.

$$\begin{aligned} \frac{1}{\sum_{a_k \in \mathcal{S}} P(a_k)} \sum_{a_i \in \mathcal{S}} P(a_i)V(a_i) &= \sum_{a_i \in \mathcal{S}} \frac{P_0(a_i)e^{\alpha V(a_i)}}{\sum_{a_k \in \mathcal{S}} P_0(a_k)e^{\alpha V(a_k)}} V(a_i) \\ &= \sum_{a_i \in \mathcal{S}} \frac{\frac{P_0(a_i)}{\sum_{a_l \in \mathcal{S}} P_0(a_l)} e^{\alpha V(a_i)}}{\sum_{a_k \in \mathcal{S}} \frac{P_0(a_k)}{\sum_{a_l \in \mathcal{S}} P_0(a_l)} e^{\alpha V(a_k)}} V(a) \end{aligned}$$

where $P(a_i) = \frac{1}{Z}P_0(a_i)e^{\alpha V(a_i)}$. We can then apply Theorem 1 to the expression in the last equality to find that:

$$\begin{aligned} \frac{1}{\sum_{a_k \in \mathcal{S}} P(a_k)} \sum_{a_i \in \mathcal{S}} P(a_i)V(a_i) &\geq V^{\mathcal{S}}_{max} + \frac{1}{\alpha}\log \frac{P_0(a^{\mathcal{S}}_{max})}{\sum_{a_k \in \mathcal{S}} P(a_k)} \\ &\geq V^{\mathcal{S}}_{max} + \frac{1}{\alpha}\log P_0(a^{\mathcal{S}}_{max}) \end{aligned}$$

where $a^{\mathcal{S}}_{max} = \arg\max_{a_i} V^{\mathcal{S}}(a_i)$. □

As a corollary, we obtain in the case $P_0(a_i) = \frac{1}{M}$ an ϵ-bound of $\epsilon = \frac{1}{\alpha}\log M$.

Similarly, Theorem 2 holds for any chosen subset $\mathcal{S}$, such that:

$$\frac{1}{\sum_{a_k \in \mathcal{S}} P(a_k)} \sum_{a_i \in \mathcal{S}} P(a_i)V(a_i) \geq V^{\mathcal{S}}_{max} - \underbrace{\left(e^{-\alpha}\left(V_{max} - V_{min}\right)\right)}_{=:\epsilon}.$$

4.2. Unknown Utility

In the second scenario of an adversarial environment, the agent chooses a distribution $P_0(a_i)$ and the environment subsequently chooses $V(a_i)$ in an arbitrary fashion, such that, in general, the choice of $V(a_i)$ may depend on $P_0(a_i)$. Once the $V(a_i)$ are revealed, the decision maker updates the choice strategy according to Equation (4). Importantly, the new distribution $P(a_i)$ is not used as a choice strategy to choose between the different $V(a_i)$ as in the previous theorems, but is only used in a later choice with new, yet unknown utilities. If we denote the trial number or time step by t and assume a trial-by-trial update:

$$P_{t+1}(a_i) = \frac{1}{Z_t} P_t(a_i) \exp(\alpha V_t(a_i)), \tag{15}$$

where the utilities $V_t(a_i)$ are bounded in each time step to lie within the unit interval, that is $V_t(a_i) \in [0;1]$, then the expected performance of the decision maker can be bounded from below by:

$$\sum_t \sum_i P_t(a_i) V_t(a_i) \geq \frac{\log(1+\epsilon)}{\epsilon} V_{max}^T - \frac{\log M}{\epsilon}, \tag{16}$$

where $\epsilon = \exp(\alpha) - 1$. This performance bound can be derived from a hedging analysis originally proposed by Freund and Shapire in a full information game where the decision maker learns about all possible utilities $V_t(a_i)$ in each time step [52,53]. In this case, the decision maker chooses between i different options with probability $p_i(t) = w_i(t)/\sum_j w_j(t)$, where the weights $w_i(t)$ are updated according to:

$$w_i(t+1) = w_i(t)\,(1+\epsilon)^{V_i(t)}$$

and where $V_i(t)$ is the utility of option i at time t. It is straightforward to see that a bounded rational decision maker following Equation (4) is hedging, when acting according to $P_t(a_i)$ before receiving feedback $V_i(t)$; that is, the bounded rational decision maker has a delay of one time step, as it is the distribution $P_{t+1}(a_i)$ that is bounded optimal for the utility $V_i(t)$ under the prior $P_t(a_i)$.

5. Discussion and Conclusion

Information-theoretic bounded rationality can be viewed as a prescriptive model of optimal decision making when the decision maker can only afford a certain amount of information processing. Information processing is formalized as a change in probability distribution from a prior distribution representing an *a priori* choice strategy to a posterior distribution over actions after information processing has taken place. Such changes in distributions can be measured by the relative entropy between prior and posterior distribution and be related to actual physical state changes in thermodynamic systems [34], where the concept of energy is analogous to the concept of utility and computational costs are analogous to entropic costs that reduce the system's capability to do work. This interpretation builds on previous work that has related computational and physical processes; see for example [54] for an overview. As discussed in the Methods, the cost of changing distributions can also be expressed in terms of complexity of sampling processes [50,51].

In this paper, we show that we can abstract away even further both from physical and computational processes when modeling bounded rational decision making with entropic information processing constraints. We show that the performance of information-theoretic bounded rational decision makers can be ϵ-bounded compared to the perfectly rational decision maker and that, therefore, information-theoretic bounded rationality naturally implies ϵ-optimality. In this sense, bounded rational decision making is strictly inferior to perfect rationality, which selects deterministically the best action. This, however, changes in adversarial environments. We discuss two scenarios. In the first scenario, the opponent can eliminate any non-empty subset of actions from the choice set after the decision maker has specified her strategy. Here, bounded rationality allows defining an ϵ-optimal performance criterion under any subset. In the second scenario, the opponent can arbitrarily select utilities for each action, and the agent responds with the bounded rational strategy with respect to the previous utilities. This scenario is equivalent to hedging and also comes with performance bounds, but in contrast to the previous setting, these bounds do not correspond to ϵ-optimality, since the difference between optimal and actual utility also depends on a multiplicative factor.

The concept of ϵ-optimality has been previously discussed in the economic literature, in particular within the context of game theory and the solution concept of ϵ-equilibria [55,56]. In particular, Fudenberg and Levine [57] have investigated the concept of ϵ-universal consistency in games where players learn a smooth best response to another player from observations. They could show that learning with a softmax-decision rule performs within an ϵ-bound of the best response with known frequencies of the opponent's play. Importantly, the concept of ϵ-optimality extends the usual black box approach taken in perfect rationality models of economic decision making where the details of the reasoning process are ignored [30]. In ϵ-optimality models, the decision maker is assumed to make decisions that are (approximately) optimal; how these decisions are arrived at is largely ignored. The choice of the ϵ in such models is typically arbitrary. Here, we link the parameter ϵ quantitatively to the temperature parameter of information-theoretic bounded rationality, that is a Lagrange multiplier indicating the shadow price of changing the distribution representing the choice strategy.

Economic models of decision making are usually considered to be as if models. The fact that behavior is consistent with an optimality criterion does not imply that an actual optimization process causes this behavior. Similarly, we could consider the information-theoretic bounded rationality model as an as if model, where the decision maker behaves as if optimizing a trade-off between utility and information cost or as if optimizing utility under information processing constraints. In contrast, when engineering an optimal decision maker (for example, a planning algorithm in a robot), typically the utility function is provided by the engineer, and the action is selected by the system after an optimization process. Here, we can consider the information-theoretic bounded rationality model as an anytime search for the optimum that stops when resources run out. Most importantly, however, independent of whether one regards utility functions as causal for behavior or not, bounded rational decision making does not necessarily imply optimizing a constrained optimization problem that is more difficult to solve than the original unconstrained problem, but the decision maker can be regarded as optimizing utility until running out of resources, thereby implicitly optimizing the constrained problem.

Acknowledgments

This study was supported by the Deutsche Forschungsgemeinschaft, Emmy Noether Grant BR4164/1-1.

Author Contributions

Daniel A. Braun and Pedro A. Ortega conceived of and wrote the paper. Both authors have read and approved the final manuscript.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Gintis, H. A Framework for the Unification of the Behavioral Sciences. *Behav. Brain Sci.* **2006**, *30*, 1–61.
2. Russell, S.; Norvig, P. *Artificial Intelligence: A Modern Approach*, 1st ed.; Prentice-Hall: Englewood Cliffs, NJ, USA, 1995.
3. Kreps, D.M. *Notes on the Theory of Choice*; Westview Press: Boulder, CO, USA, 1988.
4. Trommershauser, J.; Maloney, L.T.; Landy, M.S. Decision making, movement planning and statistical decision theory. *Trends Cogn. Sci.* **2008**, *12*, 291–297.
5. Braun, D.A.; Nagengast, A.J.; Wolpert, D. Risk-sensitivity in sensorimotor control. *Front. Hum. Neurosci.* **2011**, *5*, doi:10.3389/fnhum.2011.00001
6. Wolpert, D.M.; Landy, M.S. Motor control is decision-making. *Curr. Opin. Neurobiol.* **2012**, *22*, 996–1003.
7. Fishburn, P. *The Foundations of Expected Utility*; D. Reidel Publishing: Dordrecht, The Netherlands, 1982.
8. Neumann, J.V.; Morgenstern, O. *Theory of Games and Economic Behavior*; Princeton University Press: Princeton, NJ, USA, 1944.
9. Simon, H.A. Rational choice and the structure of the environment. *Psychol. Rev.* **1956**, *63*, 129–138.
10. Simon, H. Theories of Bounded Rationality. In *Decision and Organization*; McGuire, C.B., Radner, R., Eds.; North Holland Pub. Co.: Amsterdam, The Netherlands, 1972; pp. 161–176.
11. Simon, H. *Models of Bounded Rationality*; MIT Press: Cambridge, MA, USA, 1984.
12. Aumann, R.J. Rationality and Bounded Rationality. *Games Econ. Behav.* **1997**, *21*, 2–14.
13. Rubinstein, A. *Modeling bounded rationality*; MIT Press: Cambridge, MA, USA, 1998.
14. Kahneman, D. Maps of Bounded Rationality: Psychology for Behavioral Economics. *Am. Econ. Rev.* **2003**, *93*, 1449–1475.
15. McKelvey, R.D.; Palfrey, T.R. Quantal Response Equilibria for Normal Form Games. *Games Econ. Behav.* **1995**, *10*, 6–38.
16. Mckelvey, R.; Palfrey, T.R. Quantal Response Equilibria for Extensive Form Games. *Exp. Econ.* **1998**, *1*, 9–41.
17. Wolpert, D.H. Information Theory—The Bridge Connecting Bounded Rational Game Theory and Statistical Physics. In *Complex Engineered Systems*; Braha, D., Minai, A.A., Bar-Yam, Y., Eds; Springer: Berlin/Heidelberg, Germany, 2006; pp. 262–290.

18. Spiegler, R. *Bounded Rationality and Industrial Organization*; Oxford University Press: Oxford, UK, 2011.
19. Jones, B.D. Bounded Rationality and Political Science: Lessons from Public Administration and Public Policy. *J. Public Adm. Res. Theory* **2003**, *13*, 395–412.
20. Gigerenzer, G.; Selten, R. Bounded rationality: The adaptive toolbox; MIT Press: Cambridge, MA, USA, 2001.
21. Camerer, C. *Behavioral Game Theory: Experiments in Strategic Interaction*; Princeton University Press: Princeton, NJ, USA, 2003.
22. Howes, A.; Lewis, R.; Vera, A. Rational adaptation under task and processing constraints: implications for testing theories of cognition and action. *Psychol. Rev.* **2009**, *116*, 717–751.
23. Janssen, C.P.; Brumby, D.P.; Dowell, J.; Chater, N.; Howes, A. Identifying Optimum Performance Trade-Offs Using a Cognitively Bounded Rational Analysis Model of Discretionary Task Interleaving. *Top. Cogn. Sci.* **2011**, *3*, 123–139.
24. Lewis, R.; Howes, A.; Singh, S. Computational rationality: Linking mechanism and behavior through bounded utility maximization. *Top. Cogn. Sci.* **2014**, in press.
25. Lipman, B. Information Processing and Bounded Rationality: A Survey. *Can. J. Econ.* **1995**, *28*, 42–67.
26. Russell, S. Rationality and Intelligence. In Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence, Montreal, Canada, 20–25 August 1995; Morgan Kaufmann: San Francisco, CA, USA, 1995; pp. 950–957.
27. Russell, S.; Subramanian, D. Provably bounded-optimal agents. *J. Artif. Intell. Res.* **1995**, *3*, 575–609.
28. Glimcher, P.; Fehr, E.; Camerer, C.; Poldrack, R. *Neuroeconomics: Decision Making and the Brain*; Elsevier Science: Amsterdam, The Netherlands, 2008.
29. Friston, K.; Schwartenbeck, P.; Fitzgerald, T.; Moutoussis, M.; Behrens, T.; Dolan, R.J. The anatomy of choice: Active inference and agency. *Front. Hum. Neurosci.* **2013**, *7*, doi:10.3389/fnhum.2013.00598
30. Dixon, H. Some thoughts on economic theory and artificial intelligence. In *Artificial Intelligence and Economic Analysis: Prospects and Problems*; Moss, S., Rae, J., Eds.; Edward Elgar Publishing: Cheltenham, UK, 1992; pp. 131–154.
31. Ortega, P.; Braun, D. A conversion between utility and information. In Proceedings of the Third Conference on Artificial General Intelligence, Lugano, Switzerland, 5–8 March 2010; Atlantis Press: Paris, France, 2010; pp. 115–120.
32. Ortega, P.A.; Braun, D.A. Information, utility and bounded rationality. In *Artificial General Intelligence*, Proceedings of the 4th International Conference on Artificial General Intelligence (AGI 2011), Mountain View, CA, USA, 3–6 August 2011; Schmidhuber, J., Thórisson, K.R., Looks, M., Eds.; Lecture Notes on Artificial Intelligence, Volume 6830; Springer: Berlin/Heidelberg, Germany, 2011; pp. 269–274.
33. Braun, D.A.; Ortega, P.A.; Theodorou, E.; Schaal, S. Path integral control and bounded rationality. In Proceedings of IEEE Symposium on Adaptive Dynamic Programming and Reinforcement Learning, Paris, France, 11–15 April 2011; pp. 202–209.

34. Ortega, P.A.; Braun, D.A. Thermodynamics as a theory of decision-making with information-processing costs. *Proc. R. Soc. A* **2013**, *469*, doi:10.1098/rspa.2012.0683.
35. Wolpert, D.; Harre, M.; Bertschinger, N.; Olbrich, E.; Jost, J. Hysteresis effects of changing parameters of noncooperative games. *Phys. Rev. E* **2012**, *85*, 036102.
36. Luce, R. *Individual choice behavior*; Wiley: Oxford, UK, 1959.
37. McFadden, D. Conditional logit analysis of qualitative choice behavior. In *Frontiers in Econometrics*; Zarembka, P., Ed.; Academic Press: New York, NY, USA, 1974; pp. 105–142.
38. Meginnis, J. A new class of symmetric utility rules for gambles, subjective marginal probability functions, and a generalized Bayesian rule. In *1976 Proceedings of the American Statistical Association, Business and Economic Statistics Section*; American Statistical Association: Washington, DC, USA, 1976; pp. 471–476.
39. Fudenberg, D.; Kreps, D. Learning mixed equilibria. *Games Econ. Behav.* **1993**, *5*, 320–367.
40. Sutton, R.; Barto, A. *Reinforcement Learning: An Introduction*; MIT Press: Cambridge, MA, USA, 1998.
41. Luce, R. *Utility of gains and losses: Measurement-theoretical and experimental approaches*; Erlbaum: Mahwah, NJ, USA, 2000.
42. Train, K. *Discrete Choice Methods with Simulation*, 2nd ed.; Cambridge University Press: Cambridge, UK, 2009.
43. Toussaint, M.; Harmeling, S.; Storkey, A. *Probabilistic inference for solving (PO)MDPs*; Technical Report; University of Edinburgh: Edinburgh, UK, 2006.
44. Ortega, P.A.; Braun, D.A. A minimum relative entropy principle for learning and acting. *J. Artif. Intell. Res.* **2010**, *38*, 475–511.
45. Friston, K. The free-energy principle: A unified brain theory? *Nat. Rev. Neurosci.* **2010**, *11*, 127–138.
46. Tishby, N.; Polani, D. Information Theory of Decisions and Actions. In *Perception-reason-action cycle: Models, algorithms and systems*; Vassilis, H.T., Ed.; Springer: Berlin, Germany, 2011.
47. Kappen, H.; Gómez, V.; Opper, M. Optimal control as a graphical model inference problem. *Mach. Learn.* **2012**, *1*, 1–11.
48. Vijayakumar, S.; Rawlik, K.; Toussaint, M. On Stochastic Optimal Control and Reinforcement Learning by Approximate Inference. In Proceedings of Robotics: Science and Systems, Sydney, Australia, 9–13 July 2012; MIT Press: Cambridge, MA, USA, 2013.
49. Ortega, P.A.; Braun, D.A. Free Energy and the Generalized Optimality Equations for Sequential Decision Making. In Proceedings of the Tenth European Workshop on Reinforcement Learning, Edinburgh, Scotland, 30 June–1 July 2012.
50. Ortega, P.A.; Braun, D.A. Generalized Thompson sampling for sequential decision-making and causal inference. *Complex Adap. Syst. Model.* **2014**, *5*, 269–274.
51. Ortega, P.A.; Braun, D.A.; Tishby, N. Monte Carlo Methods for Exact & Efficient Solution of the Generalized Optimality Equations. In Proceedings of IEEE International Conference on Robotics and Automation (ICRA), Hong Kong, China, 31 May–5 June 2014.

52. Auer, P.; Cesa-Bianchi, N.; Freund, Y.; Schapire, R.E. Gambling in a rigged casino: The adversarial multi-armed bandit problem. In Proceedings of IEEE 36th Annual Symposium on Foundations of Computer Science, Milwaukee, WI, USA, 23–25 October 1995; pp. 322–331.
53. Freund, Y.; Schapire, R.E. A Decision-theoretic Generalization of On-line Learning and an Application to Boosting. *J. Comput. Syst. Sci.* **1997**, *55*, 119–139.
54. Feynman, R.P. *The Feynman Lectures on Computation*; Addison-Wesley: Boston, MA, USA, 1996.
55. Fudenberg, D.; Levine, D. *The Theory of Learning in Games*; MIT Press: Cambridge, MA, USA, 1998.
56. Noam, N.; Roughgarden, T.; Éva, T.; Vazirani, V. *Algorithmic Game Theory*; Cambridge University Press: Cambridge, UK, 2007.
57. Fudenberg, D.; Levine, D.K. Consistency and cautious fictitious play. *J. Econ. Dyn. Control* **1995**, *19*, 1065–1089.