Bayes optimality of human perception, action and learning
Beierholm, Ulrik

DOI:
10.1007/978-3-319-12084-3_10

Citation for published version (Harvard):

General rights
When referring to this publication, please cite the published version. Copyright and associated moral rights for publications accessible in the public portal are retained by the authors and/or other copyright owners. It is a condition of accessing this publication that users abide by the legal requirements associated with these rights.

- You may freely distribute the URL that is used to identify this publication.
- Users may download and print one copy of the publication from the public portal for the purpose of private study or non-commercial research.
- If a Creative Commons licence is associated with this publication, please consult the terms and conditions cited therein.
- Unless otherwise stated, you may not further distribute the material nor use it for the purposes of commercial gain.

Take down policy
If you believe that this document infringes copyright please contact UBIRA@lists.bham.ac.uk providing details and we will remove access to the work immediately and investigate.
Bayes optimality of human perception, action and learning: behavioural and neural evidence

Ulrik R. Beierholm

University of Birmingham, Birmingham, B15 2TT, UK
U.Beierholm@bham.ac.uk,
WWW home page: http://beierholm.net/

Abstract. The primary role of any biological nervous system (including the human) is to process incoming information in a way that allows motor choices to be made that increases the subjective utility of the organism. Or put slightly differently, "to make sure good things happen". There are a number of ways that such a process can be done, but one possible hypothesis is that the human nervous system has been optimized to maximize the use of available resources, thus approximating optimal computations. In the following I will discuss the possibility of the nervous system performing such computations in perception, action and learning, and the behavioural and neural evidence supporting such ideas.

Keywords: optimality, Bayesian Inference, reinforcement learning, behaviour, fMRI, neural recordings

1 Introduction

Evolution is a continuous process forcing biological organisms to constantly evolve or face extinction (red queen hypothesis [37]). Any biological function that requires metabolic energy expenditure has to provide an important role to be worthwhile. As such it is probably fair to assume that the nervous system of an animal plays an important role for its ecological fitness. If it is so important what is the role that it performs, or rather the roles as it is unlikely to simply perform a single task? While the nervous system undoubtedly plays a large part in regulating physiological factors such as hormones and metabolism, in the following I will primarily focus on the elements required for motor planning.

According to this argument (as well described by e.g. Daniel Wolpert [41, 22]) the primary role of the brain is not to 'think', write poems or contemplate modern existence. Instead the goal is to make decisions about how to change the physical environment through motor control, the only means through which the brain can influence its environment. The brain thus needs to collect information about the environment, make a motor decision and await to see the outcome of its choice in order to learn to improve its choices for the future.

The process may be imagined more clearly if we contemplate a fictive organism, e.g. a hungry frog on a lily pad. (fig. 1). The animal needs to consume

1 http://www.ted.com/talks/daniel_wolpert_the_real_reason_for_brains.html—
calories (e.g. flies for the sake of this argument) and needs to make motor choices that maximize the number of flies caught while minimizing the effort exerted.

Fig. 1. Optimal behaviour: A frog on a lily pad has to infer the location of the flies, in order to perform a motor action to catch the fly. Based on the outcome of the action it can potentially learn about properties of its perceptual and motor system as well as the environment it is in (e.g. different lilly pads might be preferable).

Given sensory input (e.g. visual and auditory stimuli) the animal first needs to infer the location of the insect (we are here disregarding whether or not frogs actually utilize sound for prey localization). Based on some estimate of this (see below) the animal can decide on a motor response that gives an optimal chance of catching the fly. Based on how well the motor response worked the animal learns about its perceptual system, motor control and the external environment.

In the follow the theoretical aspects of these processes, inference, decision making and learning, will be described in further details based on achieving the optimal behaviour for the animal.

2 Inference

Inference about states of the world becomes necessary due to stochastic properties and/or uncertainty about the states. If we have perfect knowledge about the state of the world in the past, present and future, no inference is necessary.

However the world does not present itself to us directly as von Helmholtz [19] and others have surmised. Instead we have to combine different sources of information in order to create a coherent interpretation of the world.
2.1 Bayesian inference

Given such stochasticity and the need for inference, the most natural approach is to use Bayesian statistical inference which combines previous (prior) information with current estimates (based on a likelihood function) to provide an updated (posterior) distribution of possible states.

2.2 The simple case

Imagine for now that the frog has visual information, $V$ about the location of the insect $X$. Given a visual cue $V$ which is related to the true location $X$ through a known stochastic process, $P(V|X)$, (see figure 2) and a prior distribution of possible estimates $P(X)$ the best way to combine these sources of information is through Bayes’ theorem:

$$P(X|V) = \frac{P(V|X)P(X)}{P(V)} \quad (1)$$

where $P(V) = \sum P(V|X)P(X)$ functions as a normalizing term.

The new distribution of the state of the world is merely given as the normalised product of the the prior knowledge $P(X)$ and the distribution for the new cue $P(V|X)$.

The location of the insect is now represented as a probability distribution based on the frogs visual estimate (and the uncertainty in that estimate), and its expectation about its location. This can be continuously updated (with a discounting over time if the insect is moving) and can for example also be used to estimate the velocity of the insect (and thus predict future locations).

![Fig. 2. Graphical models: a) Simple model with one cause, $S$, and one cue, $X$. b) Model with one source and two cues, $X_V$ and $X_A$. c) Causal inference model where the probability of either model is compared.](image-url)
When the distributions are Normally distributed the calculations get further simplified since estimates based on the posterior distribution can easily be shown to be a linear combination of estimates of the prior and likelihood distributions with a weight specified by the covariances of the distributions.

2.3 Two cues

A special case that has seen a lot of interest is the one of combining sources of information. Given two sources of information, e.g. audio $A$ and visual $V$ cues about a hidden property $X$, the information can again be combined in a similar fashion:

$$P(X|V, A) = \frac{P(V|X)P(A|X)P(X)}{P(V, A)} \quad (2)$$

Again, the new distribution of the estimate is given by the normalised product of the likelihood function for each of the two cues and the prior distribution.

From the viewpoint of our frog, if it has access to two sources of information about the location of the insect, it would be potentially beneficial to combine this information in a way that decreases the variance of the estimate of the location. This is exactly what the Bayesian cue integration does.

2.4 Causality

However, critically in order to perform such computations as described above, the causal structure underlying the cues needs to be known. E.g. it only makes sense to combine information from an audio and a visual cue if it is known that they refer to the same object. For large discrepancies in estimate (e.g. in space or time) from two cues automatic cue combination indeed becomes problematic. Hence it becomes necessary to estimate how likely it is that two cues are indeed from the same source, a process referred to as causal inference.

One of the simplest such cases (and yet most relevant) is the inference of whether two cues indeed were generated by one ($C = 1$) or two sources ($C = 2$).

$$P(C|V, A) = \frac{P(V, A|C)P(C)}{P(V, A)} \quad (3)$$

This allows inference about the probability of each of the causal structures given the cues, , $P(C = 1|A, V)$ and , $P(C = 2|A, V)$, allowing the estimation of a underlying source to be done through e.g. model averaging

$$\hat{X} = P(C = 1|V, A)\hat{X}_{C=1} + P(C = 2|V, A)\hat{X}_{C=2} \quad (4)$$

where $\hat{X}_{C=1}$ and $\hat{X}_{C=2}$ are the estimates of $X$ according to each of the two causal models.

This is equivalent to comparing the probability of competing hypotheses and doing averaging over them based on this value.
For the frog, if there are multiple insects nearby, it would be detrimental to combine cues from separate sources (e.g. audio from insect $\alpha$ and visual from insect $\beta$). Instead the causal relationship has to be estimated, i.e. the perceptual cues have to be attributed to the correct insects before any integration is performed.

2.5 Behavioural and Neural evidence

While this may be optimal approach to combine information, does the human or animal brains actually use this? Behaviourally this has been studied in a large number of cases. For the single cue and its relationship with prior expectations e.g. in the way human subjects combine prior expectation of the speed of an object with the visually perceived speed [39, 31].

For cue combination the Bayesian integration has been examined in a large number of human behavioural studies, showing optimal cue combination for e.g. audio-visual [1], visual-haptic [10] and slant and texture cues [21], i.e. for both within- and cross-modal stimuli.

For the estimation of causality it is perhaps somewhat surprising that the human brain seems to indeed be doing a calculation like this seemingly fast and effortless [22, 30]

While there has been overwhelming support for such models in human behavioural experiments (see also [33]), the neural evidence is only now being uncovered.

A recent study [5] examined the neural recordings of awake ferret during development and found that spontaneous activity resembled what you would expect to record for a Bayesian system without visual input, e.g. just representing the prior expectations.

Intricate neural models of Bayesian inference have also been suggested using either neural networks [8], single neurons [25, 3] as well as by single synapses [26]. While at least one study found neural activity in accordance with such models, [13], there is certainly potential for future experiment (see Fetsch et al. for a recent review[12]).

3 Choice and movement

Statistical inference provides a probability distribution over the possible values of the unknown variable of interest. But critically it does not specify what you should do with that information. For that an optimal model requires you to take into account your subjective expected utility (SEU) of the options ([4, 28]).

3.1 Utility function

SEU describes the expectations that you have about possible outcomes and the subjective value that you assign to them. Obviously you will want make decisions that maximise the probability of a positive outcome for yourself:
\[ \hat{O} = \arg \max E < U(X, O) >= \arg \max \int U(X, O)P(X|A,V) dX \tag{5} \]

The utility function can encapsulate phenomena such as rewards vs penalty trade off or risk aversion [28]. For perceptual problems it is often assumed that the utility of a choice decreases as the squared deviation from the true value \( U(X, O) \sim (X - O)^2 \). When the posterior probability of the unknown variable \( X \) is Normal distributed (as in several of the examples above) this leads to the optimal choice \( O \sim \hat{x} \sim \int XP(X|A,V) dX \) although other utility functions will lead to other optimal responses ([23]).

For our pensive frog it has to decide not just which insect to try to catch, but calculate the best location to aim for given the movement of the insect. There may be biophysical costs involved making the motor response (muscle movement is not free) as well as opportunity costs if a better possibility of capture will be arriving soon. It is not easy being green...

### 3.2 Motor control

The issue becomes a little more complicated when we take into account that not only do we not have perfect knowledge about the state of the world \( X \) (and thus have to infer a probability distribution over \( X, P(X|A,V) \)), we are also restricted in how we influence the world, not being able to directly do so in a way we wish. This is encapsulated by the problem of motor control; a specific motor signal ('point to the red dot on the screen') will lead to a distribution of outcome locations when repeated [34], \( P(O|m) \). Furthermore the reaction time will not be perfect, hence variability in lag in response also has to be considered. An optimal decision thus also has to take such variability into account

\[ m = \arg \max \int \int U(X, O)P(O|m)P(X|A,V) dXdO \tag{6} \]

where \( O \) refers to the eventual outcome, e.g. a location on a screen, while \( m \) is the motor command submitted. (The problem is actually more complex than this [32], but this will suffice for the current discussion.)

The frog needs to have a representation of its own motor uncertainty, i.e. what is the locations that its tongue can reach, what is the variability etc. A fast moving insect may become less attractive if the speed is too large once the variability of the motor response is taken into account.

### 3.3 Behavioural and Neural evidence

There is currently more behavioural than neurophysiological evidence for the brain utilizing optimal decision making and motor control, one exception being the study by Wunderlich et al. that showed activity in prefrontal cortex that was consistent with optimal use of reward cues.
Behaviourally there are however a large number of studies. Whiteley and Sahani [40] gave subjects a perceptual task where subjects had to integrate the potential payoffs when deciding on different options, showing that perceptual judgements can be influenced by the subjective utility.

In a series of experiments Trommershauser and colleagues studied the ability of human subjects to account for their own movement variability (motor noise), showing near optimal performance on simple tasks [34, 35, 43] but sub-optimal performance as the complexity increases.

4 Learning

Optimal behaviour also requires adapting to the environment through learning how the organism interacts with the environment as well as about the environment itself.

4.1 Bayesian learning

Any parameters in the proposed mechanisms above have to be learnt and adapted to the statistics of the environment. Priors and likelihoods have to reflect respectively the statistics of the environment, and the variability in the perceptual system (e.g. visual).

In our frog example the animal has to learn e.g. how reliable its visual system is relative to the auditory system or the variability of its motor system. However it will also learn about the environment itself, e.g. which lilly pad allows it to have a better chance of catching flies?

The obvious way to learn all of these variables is to utilize Bayesian inference for the parameters themselves. Bayesian models are excellent for this, due to the way new information is integrated, i.e. the posterior of a variable at time point $t$ is

$$P(X|V_{1:t}) = \frac{P(V_t|X)P(X|V_{1:t-1})}{P(V_{1:t})} \quad (7)$$

where the posterior at time $t - 1$ becomes the prior for the update at time $t$. However very little is currently known about how the nervous system would achieve this, from a theoretical as well as experimental viewpoint. It would be advantageous if the updating of the parameters could happen ‘online’ i.e. while the animal is interacting with the environment, although ‘off-line’ updating, e.g. while sleeping, may also be possible for some variables [20].

4.2 Reward learning - Model based

A special case of learning that has gathered a lot of attention, partly due to its link with the literature on Pavlovian and Operant conditioning, is that of reward learning. Compared to the full sensory input available to an organism, only utilising a reward as a single signal for learning seems somewhat impoverished
in the available information. However, the goal of the organism as argued above
should be to optimise the potential future rewards from the environment, making
it undoubtedly the most important single sensory signal to the organism.

In reward learning the total expected future rewards due to taking action $a_t$
while in state $s_t$ are calculated over a fixed time horizon or with a discount $\gamma$ of
future rewards

$$Q(s_t, a_t) = \sum_{t}^{\infty} \gamma^t E < r_t > = E < r_t > + \gamma E < r_{t+1} > + ...$$  \hspace{1cm} (8)

$$= E < r_t > + \gamma \sum_{s_{t+1}} P(s_{t+1}|s_t, a_t) \max_{a_{t+1}} Q(s_{t+1}, a_{t+1})$$  \hspace{1cm} (9)

where $E < r_t >$ is the expected reward at time step $t$ (substituting for the ex-
pected utility proposed above). Optimal behaviour (often referred to as a policy
$\hat{\pi}$) is to choose the action that maximises future rewards, $\arg \max_{a_t} Q(s_t, a_t)$.

Optimal learning can still rely on Bayesian mechanisms, e.g. to try to estimate
the probabilities for transitioning between states or receiving rewards. Such
approaches are often referred to as model-based reinforcement learning due to
their reliance on the establishment of a statistical model of the environment.

For our frog it now has to worry about not just the expected instantaneous
outcome, but also what will happen at later times. It may for example be ad-
vantageous to move to a different lilly pad to increase future expected rewards,
despite the instantaneous cost in terms of movement and opportunity cost.

4.3 Reward learning - Model free

An alternative approach to the model-based RL does not rely on a model of
the environment, instead approximating the expected reward $Q(s_t, a_t)$ by the
learnt values from previous experiences. The goal is thus still to maximize the
total expected future rewards (or utility) but this is achieved by approximating
future rewards with past rewards. One approach updates $Q(s_t, a_t)$ after each
trial according to:

$$Q(s_t, a_t) \rightarrow Q(s_t, a_t) + \alpha (r_t + \gamma \max_{a'} Q(s_{t+1}, a') - Q(s_t, a_t))$$  \hspace{1cm} (10)

This method is referred to as Q-learning [38], and while it is an approximation
to the optimal model it can be shown (given certain assumptions) to converge
to the true values of $Q(s_t, a_t)$.

In the model-free reinforcement learning there is no explicit model of the
environment, instead there is an assumption that the world is stationary and
thus that past rewards are a predictor of future rewards. This will of course lead
to deviations from optimal behaviour in situations where the environment does
indeed vary [7].
For our frog, moving to the lilly pad that has previously had most flies around it may be a generally good choice, but if things change in the lake (e.g. rise in water level) then this may no longer be a good choice, indicating how relying on past performance can be dangerous.

4.4 Behavioural and Neural evidence

Behaviourally optimal updating of parameters has been studied in both perceptual [14] and motor control systems [15]. Furthermore a few imaging studies have examined Bayesian learning of parameters, e.g. Hampton et al. found that neural activity in prefrontal cortex during a reversal learning task corresponded more closely to a model-based or Bayesian learner than a model-free reinforcement learning model [18].

The relationship between model based and model-free learning and their relative behavioural and neural evidence has been reviewed in a few papers (e.g. [7, 9]) but the the largest amount of evidence of near-optimal learning probably relates to the study of model free learning.

The discovery that the phasic firing of dopaminergic mid-brain neurons are seemingly encoding the reward prediction error of the model-free system [29] has led to a large number of studies relating reinforcement learning to both behaviour and neural recordings (see [9] for a recent review). The idea that the brain is encoding aspects of subjective utility through mechanisms similar to reinforcement learning is today no longer a controversial claim [27].

5 Discussion

In the sections above I have presented the different elements that goes into a model of optimal behaviour, inference, decision making and learning, and presented some of the evidence that the human (or animal) nervous system performs each of these computations.

5.1 Optimal Behaviour

However in a realistic environment an organism is faced with performing all of these within a task. In our toy example our pondering frog needs to infer the location of multiple flies, decide on the most valuable target and plan for the motor command that takes into account both perceptual and motor variability. After the movement it has to assess if it needs to update parameters for its visual estimation model, its motor precision as well as whether it in fact needs to choose a different target or move to a different more lucrative location.

Obviously experimentally incorporating all of these elements into a single task becomes complex too study (although for an approximation see [44] ). Optimisation of behaviour requires a number of computational steps as outlined above, computations that can only be performed by the nervous system and which constitute one of its primary functions. Each step of this process may be
performed in a separate neural population, or the nervous system may have a
different way of splitting the task that is more amenable to the type of computa-
tions in neural hardware.

5.2 Limitations
While I have debated the evidence for optimal processing in the nervous system,
there is certainly also evidence against such an idea. Some of the most well
known examples are from Kahneman & Tverskys studies [36] which highlighted
the non-optimal heuristic methods and biases typically used by subjects when
faced with written questions related to inference or decision making. Certain
criticism has been levelled at this approach (e.g. [16]) while others have tried to
bridge the gap between the performance on perceptual-motor tasks and these
cognitive tasks (e.g. [42]).

The concept of integrating the subjective utility into the decision making
process has also been questioned in human cognition, especially in behavioural
economics where researchers such as Allais [2] have shown clear inconsistencies
incompatible with the use of a single subjective utility function.

The complexity of the task faced by an organism should not be underesti-
mated. In the real world that a frog is operating in multiple targets, distractors,
irrelevant clutter etc. require the animal to perform much more intricate com-
putations than the ones outlined above. Clearly approximations will have to be
made. As researchers in this field we have multiple avenues to proceed. One
approach is to continually rely on new developments in theoretical models and
ideas, expanding the complexity of the tasks that can be processed. A different
approach is to abandon the optimal (or near optimal) scheme and instead find a
process that is ‘good-enough’. According to this idea, the performance of a real
organism can be just as good when using simplified models, similar to heuristics
[17]. The debate amongst researchers on these issues is still ongoing, regarding
the level of approximation needed or heuristics employed. However even less neu-
rophysiological evidence currently exists for such models, than for the optimal
models described above.

An obvious omission from the analysis of behaviour above is the issue of social
decision making. In our example the frog is a solitary creature who does not have
to worry about competitive or collaborative influences from e.g. members of the
same species. For some animals this seems fair, but humans are extremely social
creatures who are daily interacting with up to hundreds of other humans. Due
to space constraint I have not debated this very complex issue, one which the
economic field of behavioural game theory [6, 24] is dedicated to.

5.3 Conclusion
Evolution on this planet (earth) has shaped the form and behaviour of every
species for at least 2 billion years, forcing each species to adapt to the dynamic
environment they are placed in. Given this process and the evolutionary pressure
faced, it would be surprising if the organisms had not been trying to improve
their fitness by approximating optimal processing in the neural systems, optimising inference about properties of the environment and the choices and motor commands that can improve the utility of the organism. In the previous sections I have attempted to draft a rough sketch of the processes involved in such an optimization, and what we currently know about how the nervous system might be approximating optimal behaviour. I have only scratched the surface of such a comprehensive topic, a topic which has generated large discussions of the feasibility of the process itself, but I hope that in describing the current state of knowledge I have at least convinced the reader of the importance of considering the task of the organism when trying to understand the nervous system.

References


