Perceptual uncertainty modulates human reward-based learning

Rasmus Bruckner (r.bruckner@fu-berlin.de), Hauke R. Heekeren, Dirk Ostwald
Freie Universität Berlin, Habelschwerdter Allee 45
14195, Berlin, Germany

Abstract

To successfully interact with an everchanging world imbued with uncertainties, humans have to learn probabilistic state-action-reward contingencies. Here we investigate the computational principles that govern decision making and state-action-reward contingency learning under perceptual uncertainty. To this end, we designed an integrated perceptual and economic decision making learning task and acquired behavioural data from 52 human participants. To interpret the participants' choice data, we developed a set of seven artificial agent models that allow testing if humans consider or ignore perceptual uncertainty. Moreover, we apply these models to test if learning under perceptual uncertainty can be better described according to principles of Bayesian inference or a temporal-difference learning rule. Our results favor a Bayesian agent model that suggests that humans integrate their subjective perceptual uncertainty when learning probabilistic state-action-reward contingencies. Importantly, humans partly deviate from optimal Bayesian inference in that previous perceptual choices influence the regulation of learning at the cost of an underestimation of perceptual uncertainty. Together, this study provides a better understanding of the computational mechanisms of human state-action-reward contingency learning under perceptual uncertainty.

Keywords: Reinforcement learning; decision making; perceptual uncertainty; Bayesian inference

Introduction

Humans often have to learn state-action-reward contingencies under considerable perceptual uncertainty. For example, when learning which varieties of wild berries are edible, perceptual uncertainty about the type of berry can significantly degrade the correct credit assignment between the state (the type of berry), an action (picking a berry) and an experienced reward (an increase in blood glucose level). Previous work has extensively studied learning of state-action-reward contingencies in the absence of perceptual uncertainty (e.g., Glimcher and Fehr (2013)). Moreover, studies on perceptual decision making with asymmetric rewards for identifying correct and incorrect options indicate that humans combine perceptual uncertainty and information about reward to maximize gains (e.g., Whiteley and Sahani (2008)). However, it is currently unclear how humans consider perceptual uncertainty during state-action-reward contingency learning. We have previously reported on a novel behavioral task and preliminarily investigated in how far human learning and decision making conforms to a Bayes-optimal treatment of perceptual uncertainty (Ostwald, Bruckner, & Heekeren, 2018). Here, we present extensions of our agent-based computational modelling framework. Specifically, (1) to examine if the consideration of perceptual uncertainty can be better described according to principles of Bayesian inference or a temporal-difference (TD) learning rule, we included additional TD learning agents in our model space, (2) to formally integrate perceptual inference and perceptual decision making in our framework, we specified a perceptual decision making policy and, (3) to evaluate and estimate the computational models in the presence of participants' noisy observations that are not directly observable for the experimenter, we developed an additional experimental observation model. In the following, we briefly summarize the task to study state-action-reward contingency learning and decision making in humans and artificial agents ("The Gabor-bandit task"), then provide an overview about our computational framework ("Agent models"), and finally discuss the results of applying these agent models to human behavioural data ("Experimental results").

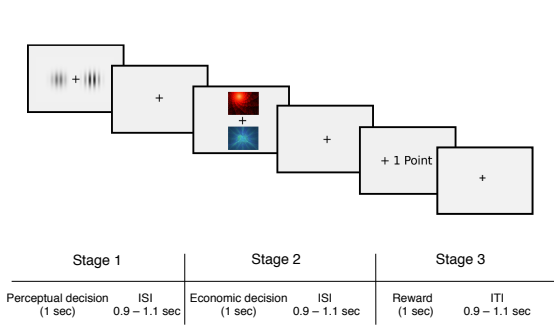
The Gabor-bandit task

As previously reported in Ostwald et al. (2018), the Gabor-bandit (GB) task is a novel state-action-reward contingency learning task which combines aspects of perceptual and economic decision making (Figure 1A). In brief, during each trial participants indicate a perceptual decision about which of two Gabor patches has the higher contrast (stage 1) and an economic decision about the fractal with the higher expected reward (stage 2), which is followed by reward feedback (stage 3). To induce varying levels of perceptual uncertainty, we manipulated the contrast difference between the patches on a trial-by-trial basis. The central feature of the GB task is the dependency of the fractal choice option reward probabilities on the relative display location of the high-contrast Gabor patch. For example, if on a given trial the high-contrast Gabor patch is displayed on the left side, then the blue fractal choice option is associated with a higher reward probability than the red fractal choice option. In contrast, if the high contrast Gabor patch is displayed on the right side, then the blue fractal choice option is associated with a lower reward probability than the red fractal choice option. In effect, over the course of each task block, participants face a credit-assignment problem regarding the contingency of the high-contrast Gabor patch location, the fractal choice options, and the received rewards.

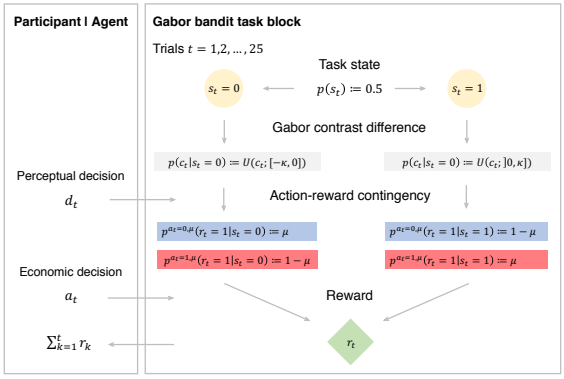
To render the GB task amenable to computational modelling, we first formulated a mathematical model of the task. In particular, we model a block of the GB task by the tuple

$$\left(T, S, C, R, D, A, p^{\phi}(s_t), p^{\kappa}(c_t|s_t), p^{a_t,\mu}(r_t|s_t)\right),$$

A



B



C

$T := 25$	Number of trials per block, which are indexed as t = 1,2, …, T
$S := \{0,1\}$	Set of task states, governing the action-reward contingencies in the task
$R := \{0,1\}$	Set of rewards
$D := \{0,1\}$	Set of perceptual decisions, where 0 and 1 represent the selection of the left and right patch, respectively
$A := \{0,1\}$	Set of economic decisions, where 0 and 1 represent the selection of the red and blue fractals, respectively
$p^{\phi} := B(s_t;\phi)$	Bernoulli state distribution with parameter $\phi := 0.5$
$p^{\kappa}(c_t\|s_t)$	State-conditional contrast difference-uniform distribution, defined by $p^{\kappa}(c_t\|s_t) := U(c_t;[-\kappa,0])^{1-s_t}U(c_t;]0,\kappa])^{s_t}$
$p^{a_t,\mu}(r_t\|s_t)$	Action-and parameter-dependent and state-conditional reward distribution, which represents the true state-action-reward contingencies of the task: $p^{a_t,\mu}(r_t\|s_t) := (B(r_t;\mu)^{1-s_t}B(r_t;1-\mu)^{s_t})^{1-a_t}(B(r_t;1-\mu)^{1-s_t}B(r_t;\mu)^{s_t})^{a_t}$

Figure 1: **Experimental task.** (A) Gabor-bandit task trial structure. (B) Illustration of the task structure. (C) Description of the task model.

according to the definitions in Figure 1C.

Agent models

To formalize the putative cognitive processes of human participants interacting with the GB task, we applied a set of seven neuroscience-inspired agent models, including a uniform random choice model (agent A7), which is not further described here.

Perceptual decision policy

For all agents, the perceptual model

$$\left(T,S,C,O,D,p^{\phi}(s_t),p^{\kappa}(c_t|s_t),p^{\sigma^2}(o_t|c_t)\right),$$

was applied to formalize perceptual decision making and perceptual uncertainty. Here

- $T,S,C,D,p^{\phi}(s_t),p^{\kappa}(c_t|s_t)$ are as for the task model,
- $O \in \mathbb{R}$ is a set of internal agent observations o_t that are assumed to result from the external Gabor patch contrast difference c_t under additive perceptual noise,
- $p^{\sigma^2}(o_t|c_t)$ is the agent's observation likelihood, which we defined by the conditional normal distribution $p^{\sigma^2}(o_t|c_t) := N(o_t;c_t,\sigma^2)$,

where σ^2 was a free parameter. In particular, we assume that each agent first infers its observation-conditional state distribution (belief state) according to $\pi_s := p^{\kappa,\sigma^2}(s_t = s|o_t)$. For the perceptual decision, we then assume that the agent aims at minimizing the 0-1 loss function

$$\ell : \{0,1\}^2 \to \{0,1\}, (d_t,s_t) \mapsto \ell(d_t,s_t) = \begin{cases} 1, d_t \neq s_t \\ 0, d_t = s_t. \end{cases} \quad (1)$$

Because the true state is unknown, the agent computes its expected loss according to

$$\mathbb{E}_{p^{\kappa,\sigma^2}(s_t|o_t)}(\ell(d_t,s_t)) = \sum_{s_t=0,1} \ell(d_t,s_t)p^{\kappa,\sigma^2}(s_t|o_t). \quad (2)$$

The agent then follows the Bayes-optimal decision policy to minimize the expected loss,

$$d_t^* = \underset{d_t \in D}{\operatorname{argmin}}\{\mathbb{E}_{p^{\kappa,\sigma^2}(s_t|o_t)}(\ell(d_t,s_t))\}, \quad (3)$$

that is, it takes the perceptual decision d_t^* that is more likely equal to the true but unknown state s_t.

Bayesian inference agents (A1-A3)

All Bayesian inference agents are represented by the tuple

$$\left(T,M,S,C,O,D,A,R,p(\mu),p^{\phi}(s_t),p^{\kappa}(c_t|s_t),p^{\sigma^2}(o_t|c_t)p^{a_t}(r_t|s_t,\mu)\right),$$

where

- the difference to the task model is that μ assumes the status of a random variable,
- $M := [0,1]$ is the outcome space of this random variable, which represents the agent's uncertainty about the state-action-reward contingency parameter on a given task block,
- $p(\mu)$ is the agent's task block-specific initial uncertainty about μ, corresponding to a uniform distribution over M.

Agent A1. During the economic decision phase, agent A1 chooses that action a_t^* for which

$$a_t^* = \underset{a_t \in A}{\operatorname{argmax}}\{\mathbb{E}_{p^{a_{1:t},a=0}(r_t|o_{1:t},r_{1:t-1})}(r_t)\}, \quad (4)$$

i.e., it chooses that fractal which promises the highest expected reward. In response to reward feedback, the agent then updates the distribution of the state-action-reward contingency parameter according to $p_t(\mu) := p^{a_{1:t}}(\mu|r_{1:t},o_{1:t})$, i.e.,

by optimally considering the observations and obtained rewards.

Agent A2. Agent A2 differs from A1 in that it ignores perceptual uncertainty. To model this categorical strategy, we adjusted the belief state to a binary state representation. In particular, we set

$$\pi_0 = \begin{cases} 0, d_t = 1 \\ 1, d_t = 0, \end{cases} \qquad \pi_1 = \begin{cases} 1, d_t = 1 \\ 0, d_t = 0. \end{cases} \tag{5}$$

Thus, the belief state of Agent A2 is entirely driven by the perceptual decision d_t. Consequently, action-dependent reward expectations do not probabilistically factor in the belief state. Similarly, the magnitude of the expected value update during learning is not modulated by perceptual uncertainty.

Agent A3. Agent A3 is a mixture model between A1 and A2, where the free parameter λ_b indicates the extent to which A3 behaves as A1 as opposed to A2. The action-dependent expected rewards are combined according to

$$E^{A3}_{p^{a_t}(r_t|o_{1:t},r_{1:t-1})}(r_t) = E^{A1}_{p^{a_t}(r_t|o_{1:t},r_{1:t-1})}(r_t)\lambda_b + E^{A2}_{p^{a_t}(r_t|o_{1:t},r_{1:t-1})}(r_t)(1-\lambda_b). \tag{6}$$

Similarly, the inferred task parameter is a linear combination of $p_t(\mu)$ of A1 and A2:

$$p^{A3}_t(\mu) = p^{A1}_t(\mu)\lambda_b + p^{A2}_t(\mu)(1-\lambda_b). \tag{7}$$

TD-learning agents (A4-A6)

In Ostwald et al. (2018), we investigated to which degree human learning and decision making in the GB task conforms to a Bayes-optimal treatment of perceptual uncertainty. However, numerous studies including behavioral, neural and animal findings suggest that TD algorithms provide a mechanistic explanation of state-action-reward contingency learning under full state observability (e.g. Glimcher and Fehr (2013)). More recent work in animals (e.g., Lak et al. (2018)) suggests that TD-learning algorithms that take perceptual uncertainty into account (Chrisman, 1992), may also provide an account for learning under perceptual uncertainty. Here we therefore additionally test various TD-learning agents that are represented by the tuple

$$\left(T, M, S, C, O, D, A, R, p^{\phi}(s_t), p^{\kappa}(c_t|s_t), p^{\sigma^2}(o_t|c_t)p^{a_t}(r_t|s_t,\mu)\right),$$

where in contrast the Bayesian agent models, the distribution over the contingency parameter $p(\mu)$ is not included. Instead, the TD-learning agents sequentially update state-action values. In particular, we consider a sequence of state-action-value functions

$$q_t : S \times A \to \mathbb{R}, (s,a) \mapsto q_t(s,a) \text{ for } t = 0,1,...,T, \tag{8}$$

where $q_t(s,a)$ denotes the value that is assigned to state s and action a at trial t. At the beginning of each block, we initialize q_0 for all state-action pairs to $q_0 = 0.5$, reflecting the agent's assumption that reward is equally likely for both actions.

Agent A4. A4 considers the belief state during economic decision making and learning. This agent chooses that action a^*_t, which maximizes the probability to obtain a reward according to

$$a^*_t = \underset{a \in A}{\operatorname{argmax}}\, Q_t(a), \tag{9}$$

where

$$Q_t(a) = \sum_{s=0,1} \pi_s q_t(s,a), \tag{10}$$

and where

$$q_t(s,a) = \begin{cases} q_t(s=0,a=0), & s=0 \wedge a=0 \\ 1-q_t(s=0,a=0), & s=0 \wedge a=1 \\ 1-q_t(s=0,a=0), & s=1 \wedge a=0 \\ q_t(s=0,a=0), & s=1 \wedge a=1. \end{cases} \tag{11}$$

During learning, $q_{t+1}(s=0,a=0)$ is then updated under consideration of the belief state according to

$$q_{t+1}(s=0,a=0) := q_t(s=0,a=0) + \alpha \begin{cases} \pi_0(\tilde{r}_t - q_t(s=0,a=0)), & \pi_0 \geq \pi_1 \\ \pi_1((1-\tilde{r}_t) - q_t(s=0,a=0))), & \pi_0 < \pi_1. \end{cases} \tag{12}$$

where $\tilde{r}_t := r_t + a_t(-1)^{2+r_t}$ accounts for the action-dependency of the reward probability and α is a free learning rate parameter.

Agent A5. Agent A5 uses a categorical economic decision making and TD-learning strategy. That is, like Agent A2, A5 represents a categorical belief state (eq. 5).

Agent A6. Agent A6 is a mixture model between A4 and A5. During economic decision making, Q-values are combined according to

$$Q^{A6}_t(a) = Q^{A4}_t(a)\lambda_r + Q^{A5}_t(a)(1-\lambda_r) \tag{13}$$

and during learning $q_{t+1}(s=0,a=0)$ according to

$$q^{A6}_{t+1}(s=0,a=0) = q^{A4}_{t+1}(s=0,a=0)\lambda_r + q^{A5}_{t+1}(s=0,a=0)(1-\lambda_r). \tag{14}$$

Experimental observation model

In tasks with perceptual uncertainty, the experimenter has no direct access to participants' observations because of internal sensory noise. An estimation of the agent models' parameters therefore requires an embedding of the agents into a statistical framework that accounts for the experimenters' uncertainty over participants' observations (Daunizeau et al., 2010). For agent models A1, A3, A4 and A6, we therefore integrated over participants' observation space conditional on the presented contrast differences. In particular, we computed the probability of a perceptual decision according to

$$p^{\sigma^2}(d_t|c_t) = \int_{-\infty}^{\infty} p(d_t|o_t)p^{\sigma^2}(o_t|c_t)do_t, \tag{15}$$

the probability of an economic decision according to

$$p^{\beta,\sigma^2}(a_t|c_{1:t},r_{1:t-1}) = \int_{-\infty}^{\infty} s_\beta(\mathbb{E}_{p^{a_{1:t-1},a_t=1,\kappa,\sigma^2}(r_t|o_t,c_{1:t-1},r_{1:t-1})}(r_t))p^{\sigma^2}(o_t|c_t)do_t \tag{16}$$

where

$$s_\beta : \mathbb{R}^2 \to \mathbb{R}^2, v \mapsto s_\beta(v), \text{where } p_j = \frac{\exp(\beta v_j)}{\sum_{k=1,2}\exp(\beta v_k)} \tag{17}$$

and the inferred μ parameter during learning according to

$$p^{\sigma^2}(\mu|r_t,c_t) := \int_{-\infty}^{\infty} p(\mu|r_t,o_t)p^{\sigma^2}(o_t|c_t)do_t. \tag{18}$$

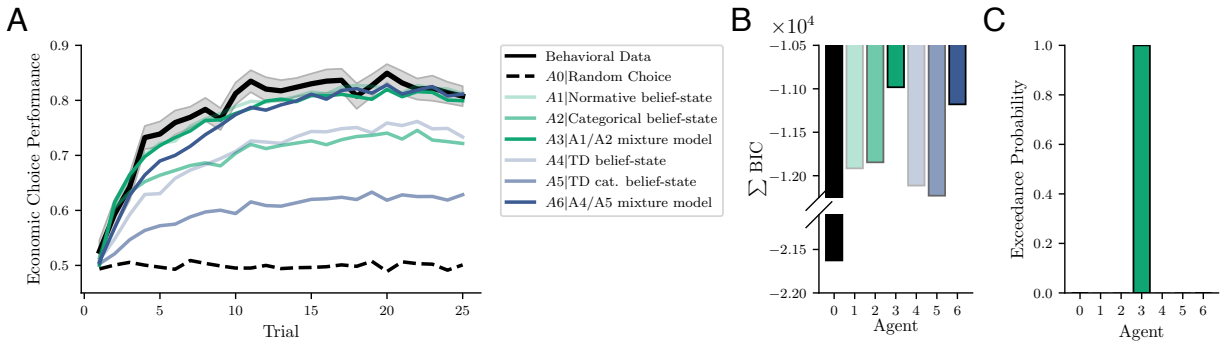


Figure 2: **Experimental results.** (A) Average proportion of economic choices associated with the higher expected reward of human participants and agent model simulations. The grey error bars depict the SEM of the human participant data. (B) Cumulative BIC scores for each agent model over participants. (C) Model exceedance probabilities.

In agents A2 and A5 this was not necessary because, as described above, we assumed that perceptual decisions directly indicate participants' categorical belief states.

Experimental results

Simulations. Figure 2A depicts participants' economic choice accuracies and the behavioral modeling results. As a measure of the agent model's face validity, we first compared the average performance achieved by human participants and simulated task-agent interactions. These simulations were conducted under similar conditions as in the experimental study with human participants (e.g., number of GB task blocks, trials, observed stimulus and rewarded frequencies, estimated parameters). The simulations suggest that, first, a mixture between the consideration of perceptual uncertainty and categorical influences of perceptual choices (A3, A6) may capture the human behavioral data well, second, that a purely categorical learning and decision making strategy does no accurately account for the data (A2, A5), and third, that a Bayes-optimal consideration of perceptual uncertainty (A1) describes the data better than a belief-state TD-learning model (A4).

Model comparison. To formally compare the agent models in light of participants' choice data, we evaluated the cumulative BIC scores over participants for each agent model (Figure 2B). Indeed, these indicate that model A3 explains the behavioral data best, which is followed by agent model A6. Moreover, assessing model plausibility using a random-effects Bayesian model selection procedure (Stephan, Penny, Daunizeau, Moran, & Friston, 2009) confirms this result by allocating a protected model exceedance probability (pEP) of more than 0.99 to agent A3 (Figure 2C).

Conclusion

Our results reveal that human participants take perceptual uncertainty during economic decision making and learning into account. Given our current model space, our results indicate that human learning and decision making under perceptual uncertainty can be better described within a Bayesian inference framework than within a TD framework that accounts for perceptual uncertainty. However, in contrast to optimal Bayesian inference, we found that perceptual choices can bias learning and decision making towards a categorical style. In summary, our work provides a mechanistic account of the modulation of reward-based learning by perceptual uncertainty and may form the basis for developing TD-learning agents that account for perceptual uncertainty in a manner comparable to Bayes-optimal inference algorithms.

References

Chrisman, L. (1992). Reinforcement learning with perceptual aliasing: The perceptual distinctions approach. *AAAI-92 Proceedings*, 183–188.

Daunizeau, J., den Ouden, H. E., Pessiglione, M., Kiebel, S. J., Stephan, K. E., & Friston, K. J. (2010). Observing the observer (I): Meta-bayesian models of learning and decision-making. *PLoS ONE*, *5*(12), e15554.

Glimcher, W., Paul, & Fehr, E. (2013). *Neuroeconomics: Decision Making and the Brain*. Academic Press.

Lak, A., Okun, M., Moss, M., Gurnani, H., Wells, M. J., Reddy, C. B., ... Carandini, M. (2018). Dopaminergic and frontal signals for decisions guided by sensory evidence and reward value. *bioRxiv*, 411413.

Ostwald, D., Bruckner, R., & Heekeren, H. (2018). Computational mechanisms of human state-action-reward contingency learning under perceptual uncertainty. In *Conference on cognitive computational neuroscience, https://doi.org/10.32470/ccn.2018.1078-0.*

Stephan, K. E., Penny, W. D., Daunizeau, J., Moran, R. J., & Friston, K. J. (2009). Bayesian model selection for group studies. *NeuroImage*, *46*(4), 1004–1017.

Whiteley, L., & Sahani, M. (2008). Implicit knowledge of visual uncertainty guides decisions with asymmetric outcomes. *Journal of vision*, *8*(3), 1–15.