2013 IEEE/RSJ International Conference on
Intelligent Robots and Systems (IROS)
November 3-7, 2013. Tokyo, Japan

# Active Bayesian perception and reinforcement learning

Nathan F. Lepora, Uriel Martinez-Hernandez, Giovanni Pezzulo, Tony J. Prescott

*Abstract*— **In a series of papers, we have formalized an *active Bayesian perception* approach for robotics based on recent progress in understanding animal perception. However, an issue for applied robot perception is how to tune this method to a task, using: (i) a belief threshold that adjusts the speed-accuracy tradeoff; and (ii) an active control strategy for relocating the sensor *e.g.* to a preset fixation point. Here we propose that these two variables should be learnt by reinforcement from a reward signal evaluating the decision outcome. We test this claim with a biomimetic fingertip that senses surface curvature under uncertainty about contact location. Appropriate formulation of the problem allows use of multi-armed bandit methods to optimize the threshold and fixation point of the active perception. In consequence, the system learns to balance speed versus accuracy and sets the fixation point to optimize both quantities. Although we consider one example in robot touch, we expect that the underlying principles have general applicability.**

## I. INTRODUCTION

A main principle underlying animal perception is the accumulation of evidence for multiple perceptual alternatives until reaching a preset belief threshold that triggers a decision [1], [2], formally related to sequential analysis methods for optimal decision making [3]. In a series of papers [4]–[10], we have formalized a *Bayesian perception* approach for robotics based on this understanding of animal perception. Our formalism extends naturally to active perception, by moving the sensor with a control strategy based on evidence received during decision making. Benefits of active Bayesian perception include: (i) robust perception in unstructured environments [8]; (ii) an order-of-magnitude improvement in acuity over passive methods [9]; and (iii) a general framework for Simultaneous Object Localization and IDentification (SOLID), or 'where' and 'what' [9], [10].

This work examines a key issue for applying active Bayesian perception to practical scenarios: how to choose the parameters for the optimal decision making and active perception strategy. Thus far, the belief threshold has been treated as a free parameter that adjusts the balance between mean errors and decision times (*e.g.* [7, Fig. 5]); furthermore, the active control strategy was hand-tuned to fixate to a region with good perceptual acuity [8]–[10]. Here we propose that these free parameters should be learnt by reinforcement from a reward signal evaluating the decision outcome, and demonstrate this method on a task in robot touch.

Past work on reinforcement learning and active perception has been confined to active vision, and was motivated initially by the *perceptual aliasing* problem for agents with limited sensory information [11]–[13]. Later studies shifted emphasis to optimizing perception, such as learning good viewpoints [14]–[16]. Just one paper has considered active (not reinforcement) learning to optimize active touch [17]. There has also been interest in applying reinforcement learning to visual attention [18]–[20]. We know of no work on learning an optimal decision making threshold and active control strategy, by reinforcement or otherwise.

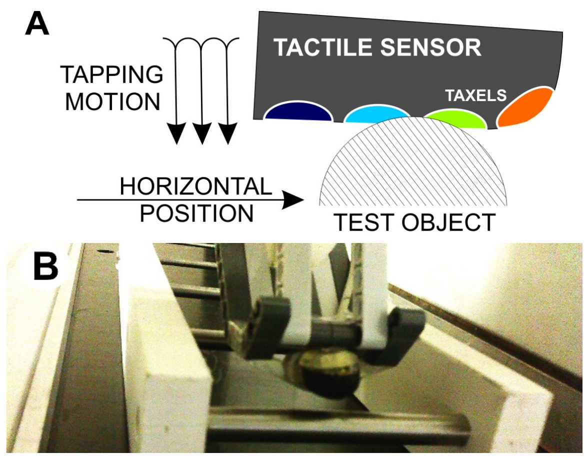


Fig. 1. Experimental setup. (A) Schematic of tactile sensor tapping against a cylindrical test object: the fingertip taps down and then back up again to press its pressure-sensitive taxels (colored) against the test object; each tap is then followed by a horizontal move. (B) Forward view of the experiment showing the fingertip mounted on the arm of the Cartesian robot.

Our proposal for active Bayesian perception and reinforcement learning is tested with a simple but illustrative task of perceiving object curvature via tapping movements of a biomimetic fingertip with unknown contact location (Fig. 1). We demonstrate first that active perception with fixation point control can give robust and accurate perception, but the decision time and acuity depend strongly on the fixation point and belief threshold. Next, we introduce a reward function of the decision outcome, which for illustration is a linear Bayes risk of decision time and error. Interpreting each active perception strategy (parameterized by the threshold and fixation point) as an action, then allows use of multi-armed bandit methods to balance exploitation and exploration of the most rewarding strategies [21]. In consequence, the appropriate decision threshold is learnt to balance the risk of making mistakes versus the risk of reacting too slowly, while the fixation point is tuned to optimize both quantities.

Although we consider one example in robot touch, we expect that the underlying principles are sufficiently general to be applicable across a range of other percepts and modalities.

This work was supported by EU Framework projects EFAA (ICT-270490) and GOAL-LEADERS (ICT-270108), and also by CONACyT (UMH).

NL, UMH and TP are with the SCentRo, University of Sheffield, UK. Email: {n.lepora, uriel.martinez, t.j.prescott}@sheffield.ac.uk

GP is with the ILC, Pisa and ISTC, Roma, Consiglio Nazionale delle Ricerche (CNR), Italy. Email: giovanni.pezzulo@cnr.it

978-1-4673-6357-0/13/$31.00 ©2013 IEEE 

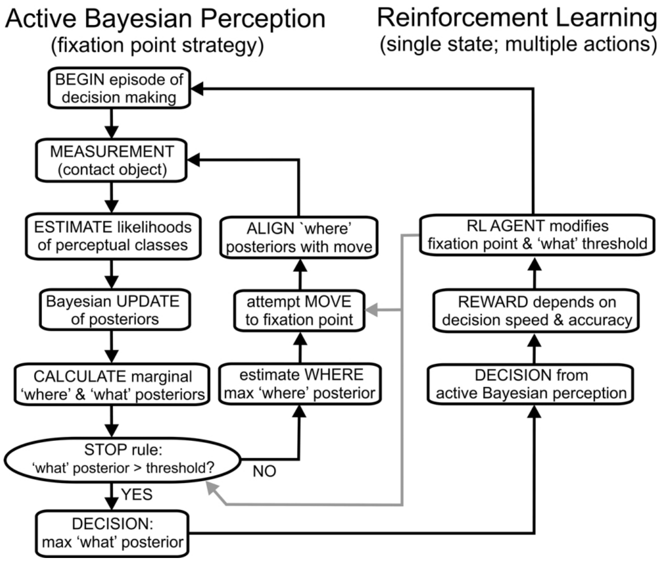


Fig. 2. Algorithm for active Bayesian perception with reinforcement learning. Active Bayesian perception (left) has a recursive Bayesian update of the posterior beliefs, marginalized over 'what' identity and 'where' location, while also actively controlling sensor location according to those beliefs; decision termination is at sufficient 'what' belief. When the sensor moves, the 'where' component of the beliefs are re-aligned with the new location. Reinforcement learning (right) modifies the belief threshold and active control strategy based on rewards derived from the decisions.

## II. METHODS

### A. Active Bayesian perception with reinforcement learning

Our algorithm for active perception is based on including a sensorimotor feedback loop in an existing method of Bayesian perception, whereby beliefs are recursively updated during perception while relocating the sensor based on those beliefs [9], [10]. Following sequential analysis methods for optimal decision making, the belief threshold to complete the decision is a free parameter that adjusts the speed-accuracy tradeoff [3]. In active perception, a control strategy relocates the sensor, here to attain a preset fixation point that is another free parameter. This study applies reinforcement learning to set these two free parameters according to a reward function of the speed and accuracy of the decision outcome.

Reinforcement learning is concerned with how agents should take actions to maximize a cumulative reward that asses the outcome of those actions. In doing so, the agent should balance exploration of new information against exploitation based on current knowledge. Multi-armed bandit problems consider an agent sequentially selecting one of multiple actions, and have well-known algorithms for action selection, *e.g.* [21]. Interpreting the choices of decision threshold and fixation point as potential actions, we can thus apply these bandit methods to optimize active perception.

Because the active perception part of these methods has been presented in other work [9], [10], we give a brief summary of active Bayesian perception and refer to previous work for more details. Our methods are framed in a general notation for any simultaneous object localization and identification task, with $N_{\rm loc}$ 'where' location classes $x_l$ and $N_{\rm id}$ 'what' identity classes $w_i$ comprising $N = N_{\rm loc}N_{\rm id}$ joint classes $c_n = (x_l, w_i)$. Each contact gives a multi-dimensional time series of sensor values $z = \{s_k(j)\}$ over time samples $j \in [1, N_{\rm samples}]$ and sensor channels $k \in [1, N_{\rm channels}]$. The $t$th contact in a sequence is denoted by $z_t$ with $z_{1:t-1} = \{z_1, \cdots, z_{t-1}\}$ its contact history.

*Measurement model and likelihood estimation:* The likelihoods of all perceptual classes are found using a measurement model of the contact data, by applying a histogram method to training examples of each perceptual class [4], [5]. First, the sensor values $s$ for channel $k$ are binned into 100 intervals; then, given a test tap $z$, the log likelihood is given by the mean log sample distribution for that tap

$$\log P(z|c_n) = \sum_{k=1}^{N_{\rm channels}} \sum_{j=1}^{N_{\rm samples}} \frac{\log P(b_k(j)|c_n, k)}{N_{\rm samples}N_{\rm channels}}, \quad (1)$$

where $b_k(j)$ is the bin occupied by sample $s_k(j)$.

*Bayesian update:* Bayes' rule is used after each successive test contact $z_t$ to recursively update the posterior beliefs $P(c_n|z_{1:t})$ for the perceptual classes with the estimated likelihoods $P(z_t|c_n)$ of that contact data

$$P(c_n|z_{1:t}) = \frac{P(z_t|c_n)P(c_n|z_{1:t-1})}{\sum_{n=1}^{N} P(z_t|c_n)P(c_n|z_{1:t-1})}, \quad (2)$$

from background information $P(c_n|z_{1:t-1})$ initialized from uniform priors $P(c_n|z_0) := P(c_n) = \frac{1}{N}$.

*Marginal 'where' and 'what' posteriors:* Because each class $c_n = (x_l, w_i)$ has a 'where' location $x_l$ and 'what' identity $w_i$ component, the beliefs for just location or identity

$$P(x_l|z_{1:t}) = \sum_{i=1}^{N_{\rm id}} P(x_l, w_i|z_{1:t}), \quad (3)$$

$$P(w_i|z_{1:t}) = \sum_{l=1}^{N_{\rm loc}} P(x_l, w_i|z_{1:t}), \quad (4)$$

are found from marginalizing the joint 'where-what' beliefs over all identity classes $w_i$ or location classes $x_l$ respectively.

*Final decision on the 'what' posteriors:* The Bayesian update stops when the marginal 'what' identity belief passes a threshold, giving a final decision

$$\text{if any } P(w_i|z_{1:t}) > \theta_{\rm id} \text{ then } w_{\rm id} = \arg\max_{w_i} P(w_i|z_{1:t}). \quad (5)$$

This belief threshold $\theta_{\rm id}$ is a free parameter that adjusts the balance between decision speed and accuracy.

*Active control strategy*: Here we consider a 'fixation point' control strategy that relocates the sensor to a point $x_{\rm fix}$ assuming its present location $x_{\rm loc}$ is the most probable:

$$x_{\rm sensor} \leftarrow x_{\rm sensor} + \Delta(x_{\rm loc}), \quad \Delta(x_{\rm loc}) = x_{\rm fix} - x_{\rm loc}, \quad (6)$$

$$x_{\rm loc} = \arg\max_{x_l} P(x_l|z_{1:t}). \quad (7)$$

This fixation point $x_{\rm fix}$ is a free parameter that adjusts the set-point of the active perception (see *e.g.* Fig. 4).

*Align 'where' posteriors*: In applying the control strategy, the 'where' location beliefs should be kept aligned with the sensor by shifting the 'where-what' beliefs upon each move

$$P(x_l, w_i|z_{1:t}) \leftarrow P(x_l - \Delta(x_{\rm loc}), w_i|z_{1:t}), \quad (8)$$

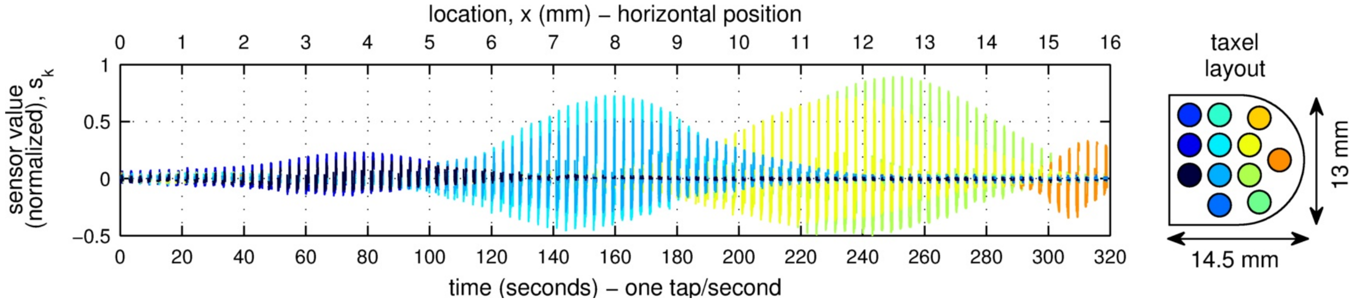


Fig. 3. Tactile dataset (for test rod of diameter 4 mm). Entire dataset, with 320 taps over horizontal positions spanning 16 mm. Taps are every 0.05 mm horizontal displacement with 1 second from each tap displayed. The taxel layout with color-code is shown on the right. As the fingertip moved across its horizontal range, the taxels were activated initially at its base (dark blue), then its middle (light blue) and finally its tip (green/yellow).

recalculating the beliefs outside the original range by assuming they are uniform and the shifted beliefs sum to unity.

*Reinforcement learning*: The active perception strategy is defined by two free parameters, the decision threshold $\theta_{\rm id}$ and fixation point $x_{\rm fix}$, to be learnt by reinforcement. Each learning trial $i$ is a perceptual decision with decision time $T_i$ (number of taps) and error $e_i$ (difference $|w_{\rm id} - w_{\rm test}|$ between identity percept and test object, measured here in mm diameter). Then the ensuing scalar reward signal $r(T, e)$ is taken here as the negative Bayes risk [3]

$$r_i = -\alpha T_i - \beta e_i, \tag{9}$$

where $\alpha$, $\beta$ are positive coefficients that parameterize the riskiness of increasing decision times and errors. Note that only the relative value $\alpha/\beta$ is important, because we aim to learn the optimal speed-accuracy tradeoff.

Standard techniques from reinforcement learning can be used to learn the active perception strategy that maximizes reward. If each strategy $(\theta_{\rm id}, x_{\rm fix})$ is considered an action, then the problem is equivalent to a multi-armed bandit. Discretizing the decision threshold $\theta_{\rm id} \in \{\theta(1), \cdots, \theta(N_\theta)\}$ and noting the $N_{\rm loc}$ 'where' classes are already discrete, allows the use of standard methods for balancing reward exploration versus exploitation (see *e.g.* [21, ch. 2]). Here we consider $N_{\rm loc} = 16$ locations (see *e.g.* Fig. 5) and $N_\theta = 13$ thresholds, giving 208 distinct actions. We use a standard algorithm that keeps a running reward average $Q = \langle r \rangle$ for each action $a = (\theta(d), x_l)$ from an incremental update

$$Q_a \leftarrow Q_a + \frac{1}{n_a + 1}\left(r_i - Q_a\right), \tag{10}$$

on trial $i$ with action $a$ chosen and $n_a$ the number of trials that this action has been chosen up to now. Exploration is achieved with initially optimistic $Q_a$ values (100 in the units of Figs 6,7). Exploitation is via a greedy policy that at each trial chooses the action with maximal $Q_a$.

### B. Tactile data collection

The tactile sensors were those used in previous studies of Bayesian perception [4]–[10]: they consist of an inner support wrapped with a flexible printed circuit board containing $N_{\rm channels} = 12$ conductive patches for the touch sensor 'taxels' [22]. These are coated with non-conductive foam and conductive silicone layers that together comprise a capacitive touch sensor that detects pressure by compression. Data were collected at 8 bit resolution and 50 cycles/sec then normalized and high-pass filtered before analysis.

For precise and exhaustive data collection, the tactile sensor was mounted on a Cartesian robot able to move the sensor in a highly controlled manner in a horizontal/vertical plane onto various test stimuli ($\sim$20 $\mu$m accuracy) [23]. The fingertip was mounted at an angle appropriate for contacting axially symmetric shapes such as cylinders aligned perpendicular to the plane of movement (Fig. 1). $N_{\rm id} = 5$ smooth steel rods with diameters 4,6,8,10,12 mm were used as test objects, mounted with their centers offset to align their closest point to the fingertip in the direction of tapping.

Touch data were collected while the fingertip tapped vertically onto and off each test object, followed by a horizontal move $\Delta x = 0.05$ mm across the closest face of the object (Fig. 1A). A horizontal $x$-range of 16 mm was used, giving 320 taps for each of the $N_{\rm id} = 5$ objects, or 1600 taps in total. From each tap of the fingertip against the object, a 1 sec time series of pressure readings ($N_{\rm samples} = 50$) was extracted for all $N_{\rm channels} = 12$ taxels (Fig. 3). All data were collected twice to give distinct training and test sets.

For analysis, the data were separated into $N_{\rm loc} = 16$ distinct location classes, by collecting groups of 20 taps each spanning 1 mm of the 16 mm $x$-range (Fig. 3) In total, there were thus $N = N_{\rm loc} N_{id} = 80$ distinct perceptual classes. These were used to set up a 'virtual environment' in which our methods could be compared off-line on identical data. A Monte Carlo validation ensured good statistics, by averaging perceptual acuities over many test runs with taps drawn randomly from the perceptual classes (typically 10000 runs per data point in results). Perceptual acuities $e_{\rm loc}$, $e_{\rm id}$ were quantified using the mean absolute error (MAE) between the actual $x_{\rm test}$, $w_{\rm test}$ and classified values $x_{\rm loc}$, $w_{\rm id}$ of object location and identity over the test runs.

## III. RESULTS

### A. Active Bayesian perception

Previous work has compared active and passive Bayesian perception methods for simultaneous object localization and identification on this and related datasets [7]–[10]. Active perception can control changes in location of the sensor during the decision making process, whereas for passive

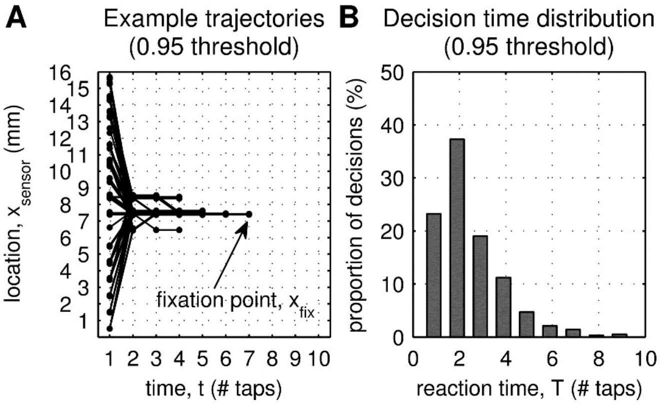


Fig. 4. Active perception with fixation point control strategy. (A) Trajectories converge on the fixation point independent of starting position. (B) Decision times have a positively skewed distribution with mean $\sim$3 taps.

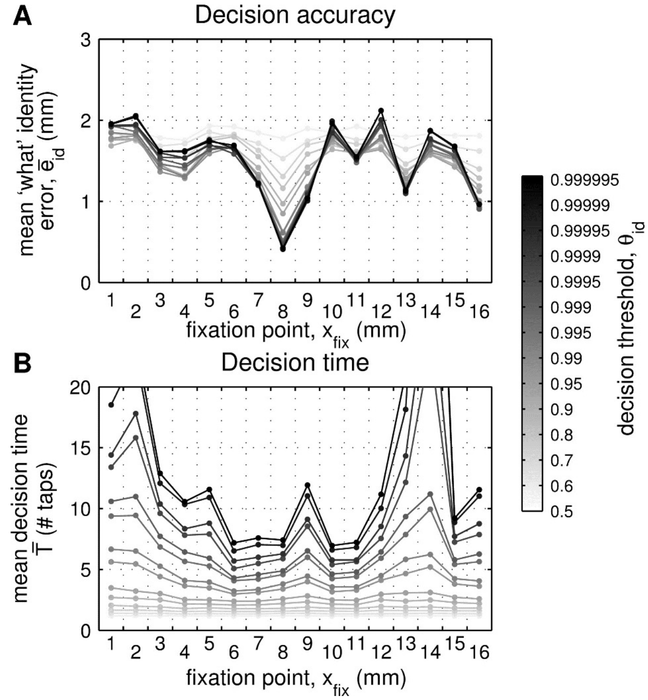


Fig. 5. Dependence of active perception on the belief threshold and fixation point. The mean identification error of rod diameter is shown in (A) and the mean decision time in panel (B), plotted against fixation point $x_{\rm fix}$. The gray-scale denotes the belief threshold. Each data point corresponds to 10000 decision trials. Perceptual performance improves in the center of the sensor location range and at greater belief thresholds.

perception the location is fixed at where the sensor initially contacted the object. We found that active perception gave far more accurate perception in situations of uncertain object location and identity than passive perception [8]–[10].

Active Bayesian perception is here applied to a 'where' and 'what' perceptual task of identifying rod location $x_{\rm loc}$ (horizontal position) and identity $w_{\rm id}$ (diameter). Results are generated with a Monte Carlo procedure using test data as a virtual environment (Sec. II-B), with an active control strategy that tries to re-locate the sensor to a preset fixation point (example trajectories in Fig. 4A).

For the present set of $N_{\rm id} = 5$ rods (4-12 mm diameter) over $N_{\rm loc} = 16$ location classes (each of 1 mm range), the active perception oriented the sensor to the fixation location within a few taps independent of starting placement (Fig. 4A; example fixation $x_{\rm fix} = 8$ mm and threshold $\theta_{\rm id} = 0.95$). The decision times to reach belief threshold had a positively skewed distribution (Fig. 4B) reminiscent of those from behavioral/psychological experiments with humans and animals [24]. The 'where' and 'what' decisions for perceiving object location and identity were measured by the mean perceptual errors $\bar{e}_{\rm loc}$, $\bar{e}_{\rm id}$ over all initial contact locations. The perceptual acuities ranged over $\bar{e}_{\rm loc} \sim 0.5$-1 mm for location and $\bar{e}_{\rm id} \sim 0.1$-1 mm for rod identity, depending on the belief threshold and fixation point.

An aspect of these results is that the localization acuity is far finer than the taxel resolution ($\sim$4 mm spacing). In previous work we have emphasized that this tactile hyperacuity [7] is a consequence of both the Bayesian perception method and the tactile sensors being designed with broad ($\sim$8 mm) but sensitive, overlapping receptive fields (Fig. 3).

### *B. Perception depends on belief threshold and active control*

The decision accuracy and decision times for active Bayesian perception depended strongly on both the belief threshold and fixation point (Fig. 5; threshold $\theta_{\rm id}$ indicated by gray shade of plot, fixation point $x_{\rm fix}$ on abscissa). Raising the belief threshold (darker gray plots), requires more evidence to make a decision, which results in improved 'what' identity perception of rod diameter and also delayed decision times. The choice of fixation point is also important for perception, with the central region of the horizontal range giving the best perceptual acuity and briefest decision times. This dependence on fixation point is due to the physical properties (morphology) of the tactile sensor coupled with shape and dynamics of the perceived object: central contacts of the fingertip activate more taxels and have improved reliability, in contrast to glancing contacts at its base or tip (Fig. 3). In consequence, errors improved from $\sim$2 mm for fixation at the base or tip, down to $\lesssim$1 mm at the center (Fig. 5; belief thresholds $\theta_{\rm id} \gtrsim$0.95).

Fig. 5 reveals that an active perception strategy with central fixation point gives the finest perceptual acuity and quickest decision times. However, the plots in Fig. 5 were obtained by 'brute force' over millions of validation trials. This raises the question of how to optimize active perception in practice over a manageably small number of decisions.

### *C. Reinforcement learning can optimize active perception*

The main theme of this paper is that the parameters controlling active perception (the decision threshold $\theta_{\rm id}$ and fixation point $x_{\rm fix}$) should be learnt by reinforcement using a reward function that evaluates the decision outcome (Fig. 2).

For simplicity, we use an example reward function given by (minus) the linear Bayes risk of decision time and absolute error of rod identity (Eq. 9). Although the proposed approach should be independent of reward function, we use the Bayes risk to give a simple example that can be interpreted as minimizing the relative risks of taking too long

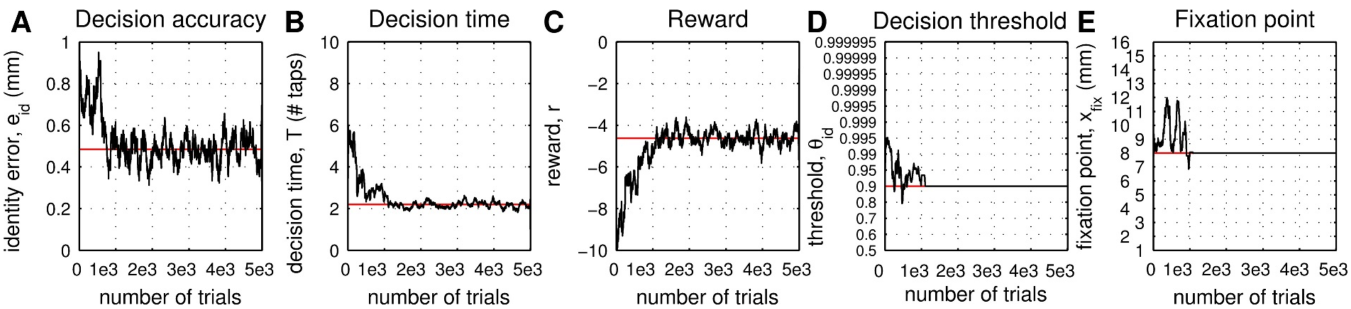


Fig. 6. Example run of reinforcement learning to optimize the active perception strategy. Change in decision error (A) and decision time (B) as the belief threshold (D) and fixation point (E) are learnt to optimize mean reward (C). Target values from brute-force optimization of the reward function are shown in red. All plots are smoothed over 100 trials. Results are for risk parameter $\alpha/\beta = 0.2$. The active perception rapidly converges to the optimal strategy.

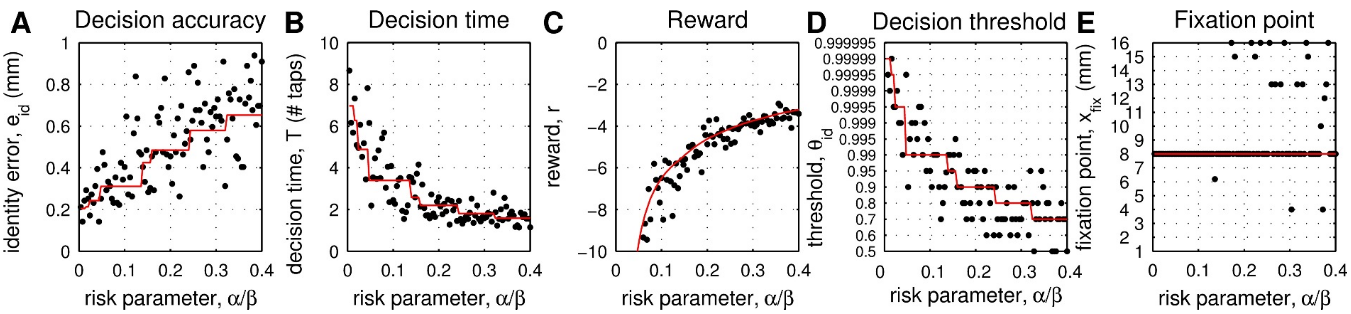


Fig. 7. Dependence of optimal active perception strategy on Bayes risk parameter. The final error (A), decision time (B), reward (C), decision threshold (D) and fixation point (E) are shown after 5000 reinforcement learning trials. The risk parameter (100 values between 0-0.4) described the relative reward benefits of improving speed versus accuracy. Results are similar to those from brute-force optimization ($\sim 10^7$ trials) of the reward function (red plots).

to reach a decision versus making errors. Then the resulting speed-accuracy tradeoff depends only upon the ratio $\alpha/\beta$ of the two coefficients in the Bayes risk, with a smaller ratio placing more risk on the decision error and a larger ratio on the decision time. Maximizing reward minimizes this risk.

In this work, each combination $(\theta_{\rm id}, x_{\rm fix})$ of decision threshold and fixation point defines a distinct active perception strategy, with the threshold taking one of the $N_\theta = 13$ discrete values $\theta(d)$ shown in Fig. 5 and the fixation point one of the $N_{\rm loc} = 16$ location classes. If the optimal strategy is to be learnt by reinforcement over many trials, each active perception strategy may be considered a distinct action. The overall situation therefore reduces to a standard multi-armed bandit problem. In consequence, the optimal active perception strategy can be learnt efficiently using standard methods for balancing exploration versus exploitation (*e.g.* those from [21, ch. 2]). In practice, all such methods that we tried converged well for appropriate learning parameters, hence we simplify our explanation by considering only a greedy method with incrementally updated reward estimates from optimistic initial values (Eq. 10).

For a typical instance of reinforcement learning and active perception, the active control strategy converged to nearly optimal perception within $\sim 10^3$ decision trials (Fig. 6; $\alpha/\beta = 0.2$). The decision threshold $\theta_{\rm id}$ and fixation point $x_{\rm fix}$ converged close to their optimal values (Figs 6D,E; red lines), validated with brute force optimization of the reward function (over $\sim 10^7$ trials). The fixation point converged to the center of the range, consistent with the brute-force results in Fig. 5, while the decision threshold converged to a suitable value to balance mean decision times and errors. Accordingly, the mean decision error $\bar{e}_{\rm id}$ and decision time $\bar{T}$ approached their optimal values, with noise due to the stochastic decision making (Figs 6A,B), while reward also increased stochastically to around its optimal value (Fig. 6C).

For many instances of reinforcement learning and active Bayesian perception, the active control strategy converged to nearly optimal perception over a range of risk parameters (Fig. 7; $0 < \alpha/\beta < 0.4$). This risk parameter represents the relative risk of delaying the decision ($\alpha$) versus making an error ($\beta$). All parameters, including the decision threshold, fixation point, rewards, decision error and decision time reached values near to optimal after 5000 trials (Fig. 7; red plots, validation with $\sim 10^7$ trials) over a range of risk parameters that give a broad span of speed-accuracy tradeoffs. In accordance, the final mean reward was close to its optimal value (Fig. 7C).

Therefore, reinforcement learning and active perception combine naturally to give a robust method for achieving optimal perception. The converged parameters values controlling active perception depend on the relative risk of speed versus accuracy. Shifting the balance of risk towards accuracy (smaller $\alpha/\beta$), results in larger decision thresholds and longer decision times, while the converse occurs with placing the risk in speed (larger $\alpha/\beta$). Concurrently, the fixation point is tuned to optimize both quantities, and converges to the central position apart from very brief decisions when the active perception strategy becomes irrelevant (for large $\alpha/\beta$).

## IV. Discussion

In this paper, we combined active Bayesian perception with reinforcement learning and applied this method to an example task in robot touch: perceiving object identity from its curvature using tapping motions of a biomimetic fingertip from an unknown initial contact location. Active perception with fixation point control strategy can give robust and effective perception; however, the decision time and acuity depend strongly on the choice of fixation point and belief threshold, necessitating some way of tuning these parameters. Introducing a reward function based on the Bayes risk of the decision outcome and considering each combination of threshold and fixation point as an action, allowed use of standard reinforcement learning methods for multi-armed bandits. The system could then learn the appropriate belief threshold to balance the risk of making mistakes versus the risk of reacting too slowly, while tuning the fixation point to optimize both quantities.

These results demonstrate that optimal robot behavior for a perceptual task can be tuned by appropriate choice of reward function. Following work on optimality in sequential analysis [3], we used a linear Bayes risk parameterized just by the relative risk of speed versus accuracy. The system then learned to make quick but inaccurate decisions when decision time was risky compared with errors, and accurate but slow decisions when errors were more risky than decision times, analogously to perceptual decision making in animals [1]. We emphasize that our general approach does not depend on the specifics of the reward function, with the actual choice representing the task aims and goals. Imagine, for example, a production line of objects passing a picker that must remove one class of object: if the robot takes too long, then objects pass it by, and if it makes mistakes, then it picks the wrong objects; both of these outcomes can be evaluated and used to reward or penalize the robot to optimize its behavior.

A key step in our combination of active perception and reinforcement learning was to interpret each active perception strategy (parameterized by the threshold and fixation point) as an action. We could thus employ standard techniques for multi-armed bandits [21], which generally worked well, and for reasons of simplicity and pedagogy we used a greedy method with optimistic initial values. Although it is beyond the scope of this paper, we expect that efficient use of the reward structure could significantly reduce exploration and hence regret (reward lost while not exploiting). For example, the reward is generally convex in the decision threshold, which could be used to constrain the value estimates.

In future work, we will study scaling our method to the many degrees of freedom necessary for practical purposes in robotics. Looking forward, we propose that optimal active Bayesian perception via reinforcement can give a general approach to robust and effective robot perception.

*Acknowledgements:* We thank Kevin Gurney, Ashvin Shah and Alberto Testolin for discussions, and the organizers of the 2012 FIAS school on Intrinsic Motivations for hosting NL and GP while some of this work was carried out.

## References

[1] J.I. Gold and M.N. Shadlen. The neural basis of decision making. *Annual Reviews Neuroscience*, 30:535–574, 2007.

[2] R. Bogacz and K. Gurney. The basal ganglia and cortex implement optimal decision making between alternative actions. *Neural computation*, 19(2):442–477, 2007.

[3] A. Wald. *Sequential analysis*. John Wiley and Sons (NY), 1947.

[4] N.F. Lepora, C.W. Fox, M.H. Evans, M.E. Diamond, K. Gurney, and T.J. Prescott. Optimal decision-making in mammals: insights from a robot study of rodent texture discrimination. *Journal of The Royal Society Interface*, 9(72):1517–1528, 2012.

[5] N.F. Lepora, M. Evans, C.W. Fox, M.E. Diamond, K. Gurney, and T.J. Prescott. Naive bayes texture classification applied to whisker data from a moving robot. *Neural Networks (IJCNN), The 2010 International Joint Conference on*, pages 1–8, 2010.

[6] N.F. Lepora, J.C. Sullivan, B. Mitchinson, M. Pearson, K. Gurney, and T.J. Prescott. Brain-inspired bayesian perception for biomimetic robot touch. In *Robotics and Automation (ICRA), 2012 IEEE International Conference on*, pages 5111 –5116, 2012.

[7] N.F. Lepora, U. Martinez-Hernandez, H. Barron-Gonzalez, M. Evans, G. Metta, and T.J. Prescott. Embodied hyperacuity from bayesian perception: Shape and position discrimination with an icub fingertip sensor. In *Intelligent Robots and Systems (IROS), 2012 IEEE/RSJ International Conference on*, pages 4638–4643, 2012.

[8] N.F. Lepora, U. Martinez-Hernandez, and T.J. Prescott. Active touch for robust perception under position uncertainty. In *Robotics and Automation (ICRA), 2013 IEEE International Conference on*, pages 3005–3010, 2013.

[9] N.F. Lepora, U. Martinez-Hernandez, and T.J. Prescott. Active bayesian perception for simultaneous object localization and identification. In *Robotics: Science and Systems*, 2013.

[10] N.F. Lepora, U. Martinez-Hernandez, and T.J. Prescott. A SOLID case for active bayesian perception in robot touch. In *Biomimetic and Biohybrid Systems*, pages 154–166. Springer, 2013.

[11] S.D. Whitehead and D.H. Ballard. Active perception and reinforcement learning. *Neural Computation*, 2(4):409–419, 1990.

[12] S.D. Whitehead and D.H. Ballard. Learning to perceive and act by trial and error. *Machine Learning*, 7(1):45–83, 1991.

[13] L. Chrisman. Reinforcement learning with perceptual aliasing: The perceptual distinctions approach. In *Proceedings of the National Conference on Artificial Intelligence*, pages 183–188, 1992.

[14] J. Peng and B. Bhanu. Closed-loop object recognition using reinforcement learning. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 20(2):139–154, 1998.

[15] L. Paletta and A. Pinz. Active object recognition by view integration and reinforcement learning. *Robotics and Autonomous Systems*, 31(1):71–86, 2000.

[16] F. Deinzer, C. Derichs, H. Niemann, and J. Denzler. A framework for actively selecting viewpoints in object recognition. *Int Journal of Pattern Recognition and Artificial Intelligence*, 23(04):765–799, 2009.

[17] H. Saal, J. Ting, and S. Vijayakumar. Active estimation of object dynamics parameters with tactile sensors. In *Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on*, pages 916–921, 2010.

[18] S. Minut and S. Mahadevan. A reinforcement learning model of selective visual attention. In *Proceedings of the fifth international conference on Autonomous agents*, pages 457–464, 2001.

[19] D. Ognibene, G. Pezzulo, and G. Baldassarre. How can bottom-up information shape learning of top-down attention-control skills? In *Development and Learning (ICDL), 2010 IEEE 9th International Conference on*, pages 231–237, 2010.

[20] A. Borji, M. Ahmadabadi, B. Araabi, and M. Hamidi. Online learning of task-driven object-based visual attention control. *Image and Vision Computing*, 28(7):1130–1145, 2010.

[21] R.S. Sutton and A.G. Barto. *Reinforcement learning: An introduction*. Cambridge University Press, 1998.

[22] A. Schmitz, P. Maiolino, M. Maggiali, L. Natale, G. Cannata, and G. Metta. Methods and technologies for the implementation of large-scale robot tactile sensors. *Robotics, IEEE Transactions on*, 27(3):389 –400, 2011.

[23] M. Evans, C. Fox, N. Lepora, M. Pearson, J. Sullivan, and T. Prescott. The effect of whisker movement on radial distance estimation: a case study in comparative robotics. *Frontiers in neurorobotics*, 6, 2012.

[24] R.D. Luce. *Response times: Their role in inferring elementary mental organization*. Oxford University Press, USA, 1991.