FROM HEURISTIC TO OPTIMAL MODELS IN NATURALISTIC VISUAL SEARCH

Angela Radulescu,
James Hillis
Facebook Reality Labs
Redmond, WA, USA
{angelar}@princeton.edu
{jmchillis}@fb.com

Bas van Opheusden, Frederick Callaway,
Thomas L. Griffiths
Departments of Psychology and Computer Science
Princeton University
Princeton, NJ, 08544, USA
{svo,fredcallaway,tomg}@princeton.edu

ABSTRACT

Effective use of limited computational resources is a hallmark of intelligent systems. Here we explore how people make use of limited perceptual resources in a naturalistic visual search task. We hypothesize that people optimally trade off performance against the cost of sampling information. To formalize this hypothesis, we frame the problem of attention allocation in visual search as a meta-level Markov decision process (meta-MDP) and show that a classic Bayesian model of visual search can be interpreted as a heuristic policy for this meta-MDP. We test the heuristic policy against gaze data from 21 human participants in a virtual reality (VR) visual search study, finding that human gaze trajectories share qualitative structure with trajectories simulated from the model.

Keywords: resource rationality, visual search, virtual reality

1 INTRODUCTION

One hallmark of human intelligence is our ability to solve complex problems with limited computational resources. Resource-rational analysis attempts to understand how this is possible by framing cognitive processes as solutions to optimization problems, where the objective function explicitly trades off external utility with internal computational cost (Griffiths et al., 2015; Lieder & Griffiths, 2019). This approach has generated models of a range of cognitive processes, explaining observations from research on judgment (Lieder et al., 2012; 2018), decision-making (Gul et al., 20 18; Callaway & Griffiths, 2019), and planning (Callaway et al., 2018). However, one area in which resource-rational analysis has not been widely applied is perception.

Perception research generally focuses on highly constrained settings in which people seem to optimally solve the statistical inference problem posed by making sense of sensory input (Hillis et al., 2004; Alais & Burr, 2019; Landy et al., 2012; Ernst & Banks, 2002). But the optimization problem in real world perceptual-motor tasks is more complex, demanding selective processing of sensory signals. Imagine for instance trying to find your keys in a messy bedroom. It is not possible to take in the full scene and immediately identify the keys. Instead, you must sequentially sample the scene with the high-resolution area of the eye (the fovea) and allocate higher-level visual processing resources to specific features, such as the color of the keychain.

Here, we propose that human strategies for solving this problem—*naturalistic visual search*—can be understood using the framework of resource-rational analysis: people decide where to look, and what features to look for, in a way that minimizes the computational cost of visual search.

2 A RESOURCE-RATIONAL MODEL OF VISUAL SEARCH

The key intuition behind our model is that visual search can be posed as a sequential decision problem. That is, visual search consists of a sequence of decisions about where to fixate next, where each such decision takes into account the information gained from previous fixations. The field of

rational metareasoning formalizes this intuition by framing the problem of computational resource allocation as a *metalevel Markov decision process* (meta-MDP; Russell & Wefald, 1991; Hay et al., 2012). Like a standard MDP, a meta-MDP consists of a set of states, a set of actions, a transition function, and a reward function. But the states correspond to beliefs, and the actions correspond to computations. The transition function defines how computations update beliefs. Finally, the reward function both penalizes computation, and rewards good decisions made based on accurate beliefs. Below, we specify a meta-MDP for visual search.

- **Latent state**: We assume that a visual scene is defined as a set of objects at specific locations, each object being represented by a feature vector in low-dimensional space. We further assume (for simplicity) that the agent knows the location of each object, but not its features. We can thus represent the latent (unknown) state as a matrix $\boldsymbol{A}$ such that the value of feature f for object o is A_{of}.
- **Beliefs**: A belief is a distribution over $\boldsymbol{A}$. For tractability, we assume independent Gaussian beliefs. A belief can thus be represented with mean and precision matrices, $\boldsymbol{F}$ and $\boldsymbol{J}$, such that $p(A_{of}) = \mathcal{N}(F_{of}, 1/J_{of})$.
- **Computations**: A computation corresponds to looking at the center of object o.
- **Actions**: An action is either $\perp$, a special termination action, or performing a computation.
- **Transition**: Formally, a computation takes a measurement $\boldsymbol{X}$ of the features of objects near the center of gaze, and incorporates that measurement into the belief using Bayesian cue combination. To specify this transition:
  1. Compute an object attentional mask $\boldsymbol{g}_o$, in which the attention paid to all objects in the scene exponentially decreases with their distance to o.
  2. Compute the measurement precision $\boldsymbol{J}_{\text{meas}} = \boldsymbol{g}_o \mathbf{1}^T$ such that the features of objects near the center of gaze have the highest precision.
  3. Sample a measurement $\boldsymbol{X} \sim \mathcal{N}\left(\boldsymbol{A}, 1/\boldsymbol{J}_{\text{meas}}\right)$.
  4. Perform the Bayesian update
  $$\boldsymbol{F} \rightarrow \frac{\boldsymbol{F} \odot \boldsymbol{J} + \boldsymbol{X} \odot \boldsymbol{J}_{\text{meas}}}{\boldsymbol{J} + \boldsymbol{J}_{\text{meas}}} \tag{1}$$
  and $\boldsymbol{J} \rightarrow \boldsymbol{J} + \boldsymbol{J}_{\text{meas}}$, where $\odot$ represents the element-wise product.
- **Reward**: The agent incurs a cost $-c$ for each computation. When the agent terminates computation, it calculates a posterior over which object o is the target given its beliefs,
  $$\boldsymbol{p}_o \propto \exp\left(\frac{1}{2}\sum_f \left[\log\left(1 + \boldsymbol{J}_{of}\right)\boldsymbol{F}_{of}^2 - \boldsymbol{J}_{of}\left(\boldsymbol{F}_{of} - f_f^{\text{target}}\right)^2\right]\right) \tag{2}$$
  This calculation assumes that the true values of object features are all mutually independent and Gaussian with mean 0 and variance 1. The agent then selects $\text{argmax}[\boldsymbol{p}_o]$ as its target report and receives a final reward of $R = 1$ if this is correct and $R = 0$ otherwise.

The classic *ideal observer* model of visual search proposed in Najemnik & Geisler (2005) can be interpreted as a policy for this meta-MDP with a fixed myopic decision rule in which the agent always selects $\text{argmax}\boldsymbol{p}_o$ as its fixation target, and chooses $\perp$ whenever $\max[\boldsymbol{p}_o]$ exceeds a threshold. We can prove that this policy is optimal for simplified versions of the meta-MDP, however it is unclear if that is also the case for the general version. To address this question, we trained artificial agents with deep reinforcement learning to solve the meta-MDP in environments similar to those experienced by the human observers (see Supplement for additional details). We found that even with extensive training, these agents cannot outperform the ideal observer. Moreover, after training, these agents increasingly take the same action as the ideal observer. These results suggest that we can use the ideal observer policy as a computationally efficient stand-in for the true optimal policy.

3 Experimental paradigm

Many studies of visual search are designed to test the human ability to use specific features to find objects (Wolfe, 2010). While these studies have identified some image features that drive eye movements, such stimuli are are inconsistent with the multidimensional structure of real-world environments. To circumvent this limitation, we conducted a study of visual search in virtual reality (VR).

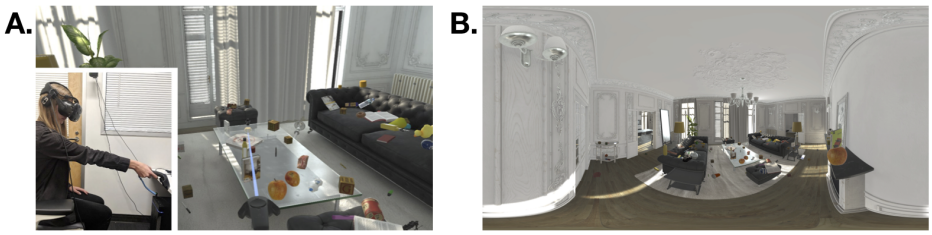

Figure 1: **Visual Search in Virtual Reality**. A: Egocentric & 3^{rd} person view of participant shown in the inset. B: 360-degree equirectangular projection of scene in A.

26 participants viewed VR scenes generated with the Unity game engine through a head-mounted display (HTC Vive VR), equipped with a Tobii Pro VR eye tracker (sampling frequency: 120Hz), while holding a handheld controller (HTC Vive VR). Each participant performed 300 trials of visual search for a target in a cluttered room with 55-112 distractors and under an 8 second deadline (Figure 1A, see Supplement for additional details). In some of the trials, visual recommendations, in the form of transparent blue blobs, were presented. Data from those trials are not reported here. To analyze gaze trajectories, we transformed the raw gaze sample coordinates to the pixel space of 360 degree equirectangular projections from the pre-specified viewpoints in each scene (Sitzmann et al., 2018) (Figure 1B). This transformation is necessary in order to simultaneously take into account participants' head and eye movements.

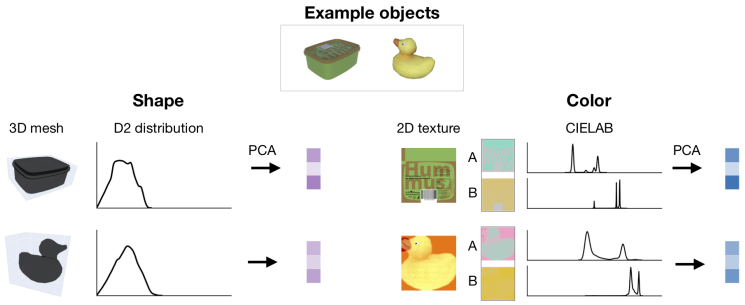

Figure 2: **Feature extraction**. Top: a plastic box and a rubber duck, two example objects rendered with Unity that were used in our study. Bottom left: mesh representation and corresponding D2 distributions. Bottom right: 2D texture and corresponding CIE L*a*b* color distributions.

4 Feature Extraction

A challenge for any decision-theoretic model of goal-directed behavior is reducing the perceptual input to a low-dimensional state representation that can provide task-relevant reward signals (Niv, 2019; Leong et al., 2017; Radulescu et al., 2019). In the context of naturalistic visual search, we drew from work in visual cognition showing that eye-movements are guided by objects and object shape, color and texture (Wolfe & Horowitz, 2017). We quantified shape as the D2 distribution over the 3D mesh of each object (Osada et al., 2002) (Figure 2, left). We quantified color by converting the 2D texture of each object to the CIE L*a*b* color space and extracting the A (green-red) and B (blue-green) channels (Figure 2, right). We then applied Principal Components Analysis (PCA) to the shape and color of all objects, computed low-dimensional projections in the space defined by the principal components, and measured similarity between objects in this space (see Supplement for additional details and a validation of the PCA procedure and similarity metric).

5 Results

Participants successfully found the target on 76% of trials, with a median response time on correct trials of 2.89s (IQR: 1.99-4.44s). Preliminary simulations of human gaze data using the ideal-

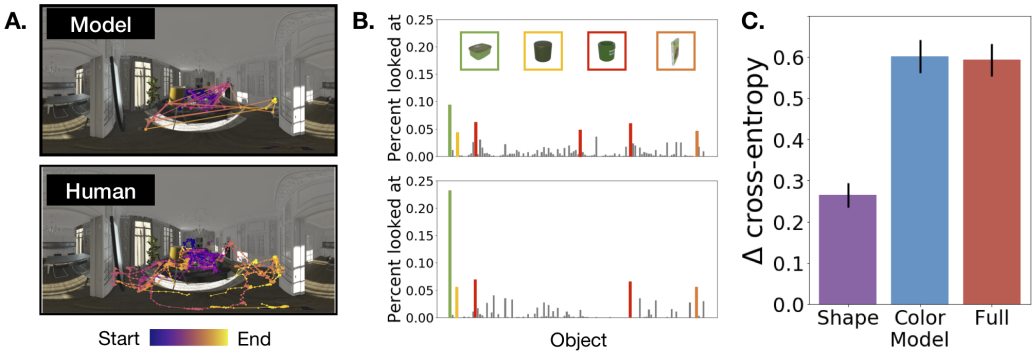

Figure 3: **Modeling results**. A, top: gaze trajectories obtained by simulating the ideal-observer heuristic policy in one of the living room scenes. A, bottom: raw gaze trajectories of 8 participants searching in the same scene. B, top: relative looking times to all objects averaged over 100 simulations of the ideal-observer for the scene on the left. B, bottom: relative looking times to all objects averaged over participants whose gaze trajectories are shown on the left. The green bar represents the target and yellow, red and orange are three prominent distractors. C: Model comparison between variants of the ideal-observer model that use either shape, color or both features when computing the reward.

observer heuristic policy showed that the dynamics of search are consistent with a sequential sampling policy that maximizes the probability of the next fixation landing on the target given current beliefs (Figure 3A). Notably, both the model and humans display "inhibition of return", in that they tend to sample objects in the same parts of space, and then focus attention elsewhere (Najemnik & Geisler, 2005). The model and the human also tend to get distracted by the same objects, and distractors that consistently capture attention across both model and human search episodes are similar to the target (Figure 3B, Supplementary figure 2). This suggests that, as predicted by the model, people incorporate similarity to target as an intrinsic reward signal into their policies for sequentially sampling the environment during visual search. Finally, a model comparison revealed that an ideal-observer model which only uses color to compute reward is most predictive of the distribution of objects that people look at (Figure 3C, see Supplement for details of the model comparison procedure). In other words, for this task, color is more important for human search policies than shape. We note however that the termination rule for the heuristic policy of the ideal observer is set by an arbitrary criterion. An optimal criterion may be exposed in ongoing work aimed at identifying the optimal policy.

6 Discussion and ongoing work

We framed visual search as a sequential sampling problem, formalized as a meta-MDP with a low-dimensional perceptual representation. We tested an "ideal observer" model, which always looks at the object it believes to be the most likely target. We used deep reinforcement learning to show that this model is close to the optimal policy of the meta-MDP, and provided evidence that its policy matches some aspects of human eye movements. Another promising avenue for future work is to investigate alternative schemes for automated extraction of low-dimensional feature representations of objects, such as representations that emerge in deep convolutional neural network models of the the ventral visual stream, or neural network models trained on object segmentation and discrimination (Yamins et al., 2014; Fan et al., 2019; Wu et al., 2018). Using PCA instantiates a simple hypothesis for the kinds of representations that might underlie object discrimination based on parallel feature maps (Treisman & Gelade, 1980), but does not capture more complex invariances (e.g. to lighting, rotation) that our visual system may have learned over the course of interacting with the environment. Our results highlight the potential of framing goal-directed behaviors in naturalistic environments as meta-reasoning problems. This approach may shed light on the longstanding question of how humans dynamically adapt the structure of their environment into perceptual representations that are useful across a wide range of tasks.

REFERENCES

Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Yangqing Jia, Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dandelion Mané, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. URL `https://www.tensorflow.org/`. Software available from tensorflow.org.

David Alais and David Burr. Cue Combination Within a Bayesian Framework. In *Multisensory Processes*, pp. 9–31. Springer International Publishing, Cham, March 2019.

Frederick Callaway and Tom Griffiths. Attention in value-based choice as optimal sequential sampling. 2019.

Frederick Callaway, Falk Lieder, Priyam Das, Sayan Gul, Paul M Krueger, and Tom Griffiths. A resource-rational analysis of human planning. In *CogSci*, 2018.

Doug ER Clark, Jonathan R Corney, Frank Mill, Heather J Rea, Andrew Sherlock, and Nick K Taylor. Benchmarking shape signatures against human perceptions of geometric similarity. *Computer-Aided Design*, 38(9):1038–1051, 2006.

Wolfgang Einhäuser, Merrielle Spain, and Pietro Perona. Objects predict fixations better than early saliency. *Journal of Vision*, 8(14):18–18, 2008.

Marc O Ernst and Martin S Banks. Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, 415(6870):429–433, January 2002.

Judith E Fan, Jeffrey D Wammes, Jordan B Gunn, Daniel LK Yamins, Kenneth A Norman, and Nicholas B Turk-Browne. Relating visual production and recognition of objects in human visual cortex. *Journal of Neuroscience*, 2019.

Thomas L Griffiths, Falk Lieder, and Noah D Goodman. Rational use of cognitive resources: Levels of analysis between the computational and the algorithmic. *Topics in cognitive science*, 7(2): 217–229, 2015.

Sergio Guadarrama, Anoop Korattikara, Oscar Ramirez, Pablo Castro, Ethan Holly, Sam Fishman, Ke Wang, Ekaterina Gonina, Neal Wu, Efi Kokiopoulou, Luciano Sbaiz, Jamie Smith, Gábor Bartók, Jesse Berent, Chris Harris, Vincent Vanhoucke, and Eugene Brevdo. TF-Agents: A library for reinforcement learning in tensorflow, 2018. URL `https://github.com/tensorflow/agents`.

Sayan Gul, Paul M Krueger, Frederick Callaway, Thomas L Griffiths, and Falk Lieder. Discovering rational heuristics for risky choice. In *The 14th biannual conference of the German Society for Cognitive Science*, 20 18.

Nicholas Hay, Stuart Russell, David Tolpin, and Solomon Eyal Shimony. Selecting computations: Theory and applications. In *Proceedings of the 28th Conference on Uncertainty in Artificial Intelligence*, 2012.

James M Hillis, Simon J Watt, Michael S Landy, and Martin S Banks. Slant from texture and disparity cues: Optimal cue combination. *Journal of Vision*, 4(12):1–26, December 2004.

Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

Nikolaus Kriegeskorte, Marieke Mur, and Peter A Bandettini. Representational similarity analysis-connecting the branches of systems neuroscience. *Frontiers in systems neuroscience*, 2:4, 2008.

Michael S Landy, Martin S Banks, and David C Knill. Ideal-Observer Models of Cue Integration. In *Sensory Cue Integration*, pp. 5–29. Oxford University Press, September 2012.

Yuan Chang Leong, Angela Radulescu, Reka Daniel, Vivian DeWoskin, and Yael Niv. Dynamic interaction between reinforcement learning and attention in multidimensional environments. *Neuron*, 93(2):451–463, 2017.

Falk Lieder and Thomas L Griffiths. Resource-rational analysis: understanding human cognition as the optimal use of limited computational resources. *Behavioral and Brain Sciences*, pp. 1–85, 2019.

Falk Lieder, Tom Griffiths, and Noah Goodman. Burn-in, bias, and the rationality of anchoring. In *Advances in neural information processing systems*, pp. 2690–2798, 2012.

Falk Lieder, Thomas L Griffiths, and Ming Hsu. Overrepresentation of extreme events in decision making reflects rational use of cognitive resources. *Psychological review*, 125(1):1, 2018.

Jiri Najemnik and Wilson S Geisler. Optimal eye movement strategies in visual search. *Nature*, 434 (7031):387–391, 2005.

Yael Niv. Learning task-state representations. *Nature neuroscience*, 22(10):1544–1553, 2019.

Robert Osada, Thomas Funkhouser, Bernard Chazelle, and David Dobkin. Shape distributions. *ACM Transactions on Graphics (TOG)*, 21(4):807–832, 2002.

Angela Radulescu, Yael Niv, and Ian Ballard. Holistic reinforcement learning: the role of structure and attention. *Trends in cognitive sciences*, 2019.

Stuart Russell and Eric Wefald. Principles of metareasoning. *Artificial Intelligence*, 49(1-3):361–395, 1991.

John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*, 2017.

Vincent Sitzmann, Ana Serrano, Amy Pavel, Maneesh Agrawala, Diego Gutierrez, Belen Masia, and Gordon Wetzstein. Saliency in vr: How do people explore virtual environments? *IEEE transactions on visualization and computer graphics*, 24(4):1633–1642, 2018.

Anne M Treisman and Garry Gelade. A feature-integration theory of attention. *Cognitive psychology*, 12(1):97–136, 1980.

Manasij Venkatesh, Joseph Jaja, and Luiz Pessoa. Comparing functional connectivity matrices: A geometry-aware approach applied to participant identification. *NeuroImage*, 207:116398, 2020.

Jeremy M Wolfe. Visual search. *Current biology*, 20(8):R346–R349, 2010.

Jeremy M Wolfe and Todd S Horowitz. Five factors that guide attention in visual search. *Nature Human Behaviour*, 1(3):1–8, 2017.

Zhirong Wu, Yuanjun Xiong, Stella Yu, and Dahua Lin. Unsupervised feature learning via non-parametric instance-level discrimination. *arXiv preprint arXiv:1805.01978*, 2018.

Daniel LK Yamins, Ha Hong, Charles F Cadieu, Ethan A Solomon, Darren Seibert, and James J DiCarlo. Performance-optimized hierarchical models predict neural responses in higher visual cortex. *Proceedings of the National Academy of Sciences*, 111(23):8619–8624, 2014.

A Supplementary material

A.1 Additional experimental details

We recruited 26 participants via web advertisement. Participants provided informed consent consistent with the Declaration of Helsinki. Experimental sessions lasted approximately 60 minutes, for which we compensated participants $50. We excluded data from 5 participants for data quality issues, yielding a dataset of 21 participants.

A trial was defined by a combination of one of 6 possible rooms (kitchen, living room, bathroom, study or one of two bedrooms), one of 5 pre-determined viewpoints, and one of 10 pre-generated target/distractor sets. Each participant experienced the set of 300 unique trials this procedure generated but in different random order. For each participant, assistance was provided on 100 trials.

Each trial began by showing the target object for 4 seconds. The participant was then put into the room from a perspective selected *a priori*. At any time during the 8-second search period, the participant could report having found the target by pointing and clicking with a handheld controller (Figure 1A). If the participant did not find the target with the 8-second period, the trial ended.

We determined the identity, size, location and orientation of the target object and distractor objects with a number of constraints: (1) each object rested on a stable surface (e.g. floor, table, etc); (2) none of the distractors had the same object identity as the target, or are too similar; (3) at least 50% of the target was visible from the participant's viewpoint; (4) the visible area of the target is at least 3 degrees of visual angle squared.

A.2 Additional annotation and feature extraction details

We did not apply other processing or smoothing to gaze data, and we did not segment trajectories into fixations and saccades, since saccade detection algorithms are developed for head-fixed participants viewing 2D screens and may not readily extend to free viewing in 3D virtual reality environments.

In post-processing, we annotated each raw gaze sample with the label of the object at the center of gaze. We extracted this object label by recording the object mesh collider hit by a ray in the Unity environment in the direction of gaze. We found that about 50% of gaze samples landed either on task-relevant objects (i.e., the target or a distractor), suggesting that objects are a strong cue for gaze, consistent with previous literature (Einhäuser et al., 2008). In this paper, we restrict analysis to those gaze samples which hit task-relevant objects.

The D2 shape distribution and CIE L*a*b* color representation both yield high-dimensional representations of the shape and color of each object, constrained by the number of times pairs of points are sampled from the mesh (in the case of D2) and the pixel size of the texture (in the case of CIE L*a*b*). For computing the shape distribution, we randomly sampled 500,000 pairs of points on the mesh surface. For converting color texture to CIE L*a*b*, we resampled each texture to 2048*2048px, and concatenated the A and B channels.

Due to its ease of computation and robustness in the presence of rotations, translations and other shape perturbations, the D2 distribution has been widely used in the computer graphics and computer vision literature as a metric for discriminating between shapes of different classes (Osada et al., 2002). Importantly, data from human similarity judgments suggests the D2 distribution captures some aspects of human shape perception (Clark et al., 2006).

The CIE L*a*b* color space describes all the colors visible to the human eye, and can thus serve as a metric space for color similarity that roughly matches that of human color perception. Because color was extracted from 2D texture files, this representation is likely to also include some textural elements such as smoothness or graininess.

A.3 Feature extraction validation

To validate the choice of PCA as a dimensionality reduction step in our model, we generated separate shape and color representational dissimilarity matrices (RDMs) for all the objects in used in the experiment by computing the pairwise Euclidean distance between shape and color vectors (Kriegeskorte et al., 2008; Venkatesh et al., 2020). We used used either the full representation (RDM-full), or a representation based on a restricted number of principal components (RDM1, RDM2, etc). We then performed a permutation test to obtain a null distribution for the Pearson distance between the intact and randomly permuted full RDM. This distribution provides an estimate of the upper bound on the Pearson distance we might expect between different RDMs. Taking the distance of the full RDM to itself as a lower bound, we computed and plotted the distance between RDM-full, RDM1, RDM2, etc.

We found that object shape and color similarity is largely preserved for low-dimensional projections in the space of principal components (Figure A1). For instance, two objects that are similar when

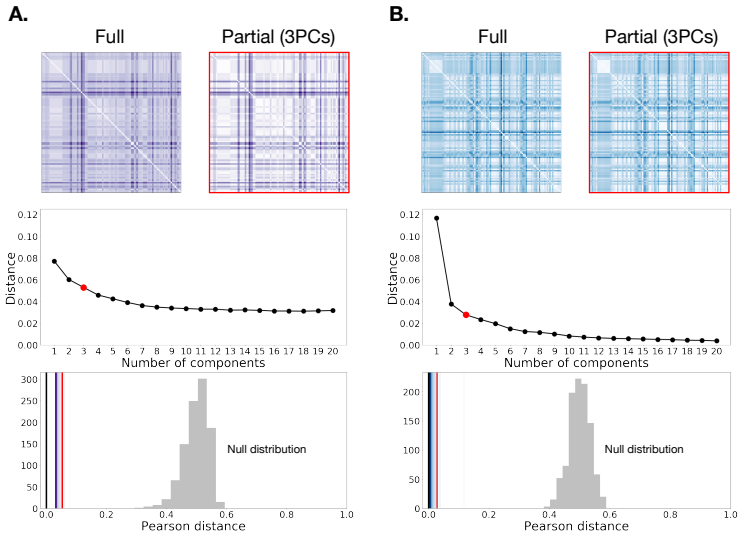

Figure A.1: Feature extraction validation.

computing similarity in the space defined by the full shape/color representation are also similar when computing similarity in the space defined by the first three principal components (Figure A1 top row). We found that the distance between full and partial RDMs sharply decreases with the number of components used (Figure A1 middle row), And for both shape and color, RDMs1-20 were well outside the null distribution obtained via permutation testing (Figure A1 bottom row). After visual inspection of how objects are distributed in the space of principal components, we chose a 3D projection for each feature. In other words, every object's shape and color was represented as a 3D vector of real numbers obtained by projecting the shape and color data in the space defined by the first 3 principal components.

A.4 SIMILARITY METRIC VALIDATION

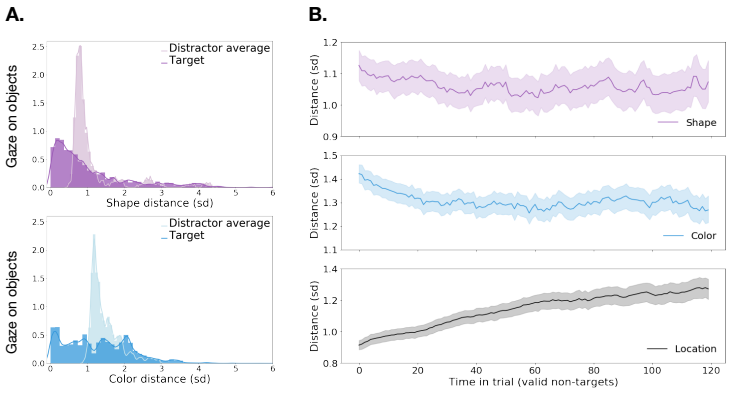

Figure A.2: Similarity metric validation.

According to our model, the probability that an object is the target decreases with distance along each feature, weighted by the precision of that feature. In other words, similarity to target serves as an internal reward signal. In alignment with this assumption, we also found that similarity between objects computed from the low-dimensional shape and color representations is predictive of human gaze behavior. Participants were more likely to look at objects that are similar to the target rather

than to a "prototypical" distractor (Figure A2-A). Moreover, the dissimilarity between the objects people fixated at and the target decreased over time for both shape and color, but not for location. This suggests that people preferentially look at objects close to the target in representational space, although these objects are not physically close (Figure A2-B). These results validated similarity to target in the low-dimensional feature space defined by color and shape as a useful signal that guides participants' search policies.

A.5 Learning metalevel policies with deep reinforcement learning

In order to identify near-optimal policies for the meta-MDP, we trained artificial agents to maximize the expected cumulative rewards in an ensemble of visual search environments. For each of the 300 scenes presented in the human task, we generated one artificial environment in which the target was the same as in the human experiment. We also created additional environments by considering the same scene, but with a different object assigned to be the target. We only generated such scenes if that object was unique among the object set, resulting in an ensemble of 7022 environments.

In order to ensure that each environment has the same number of objects, which is necessary for our network architecture, we add 'phantom objects' to each environment with less than the maximum number of objects (113). For these phantom objects, we set the true feature values to zero.

To train our artificial agents, we employed neural networks implemented in tensorflow (Abadi et al., 2015) with the tf-agents library (Guadarrama et al., 2018). We used proximal policy optimization (PPO) (Schulman et al., 2017) to define the loss function, and trained using Adam (Kingma & Ba, 2014) with exponential learning rate decay.

Since PPO is an actor-critic method, we specified two neural networks which only differed in the output layer (Figure A.3). Both networks consist of a preprocessing layer which represents inputs (see below), then a series of 3 densely connected layers of 113 units each, followed by an output layer (a single value for the critic, an action distribution layer for the actor).

The preprocessing layer first represents a belief states in the meta-MDP as a list of 4 tensors

- F, a 113×6 matrix of the mean feature values (see Section 2).
- J, a 113×6 matrix with the corresponding precisions.
- f_{target}, a 1×6 vector with the true values of the target object's features.
- $\boldsymbol{x}_o$, a 113×2 matrix with the horizontal and vertical coordinates of each object, normalized so that the screen maps to the interval $[-1, 1]$. For phantom objects, we set $\boldsymbol{x}_o$ to $(0, 0)$, the screen center.
- $\boldsymbol{p}_o$, a 113×1 vector with the output of Equation 2. We set $\boldsymbol{p}_o$ to 0 for phantom objects, then re-normalize to ensure the sum of $\boldsymbol{p}_o$ over non-phantom objects equals 1. Although $\boldsymbol{p}_o$ is a deterministic function of F, J and f_{target} and therefore provides no additional information to the agent, we include it in the state representation to accelerate learning.

The preprocessor then passes each of these tensors through a single dense layer with 113 units and concatenates the outputs into a single tensor which represents the belief state.

In preliminary results with this setup, we found that the learned agents were not able to achieve returns exceeding those of the ideal observer heuristic. Thus, we performed two additional modifications to the state and action space to allow the network to more easily learn policies that match or outperform the ideal observer. First, when passing a belief state to the preprocessing layer, we re-order objects by rank order according to the posterior, and we apply the same re-mapping for the action space. In other words, the first rows of F and J, and the first entry of $\boldsymbol{x}_o$, always correspond to the object that is most likely to be the target according to the belief state; action 1 corresponds to fixating on that object. Therefore, the heuristic policy can easily be represented and learned by the network. Additionally, we initialize network with a bias towards picking action 1, thereby encouraging it to explore policies similar to the ideal observer. With these modifications, we found that the artificial agents are able to learn policies similar to the heuristic agent, with similar returns, but not outperform the ideal observer heuristic. Although these deep RL results reflect a work in progress, the difficulty of exceeding the heuristic agent's return suggests that the ideal observer heuristic is a good policy for the meta-MDP and potentially close to optimal.

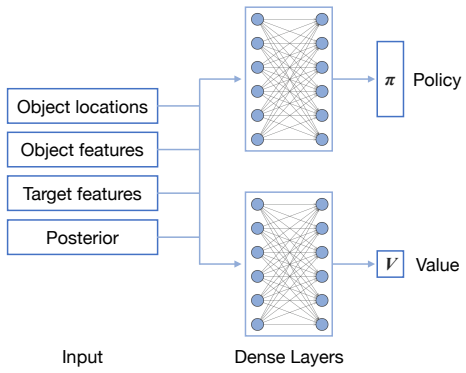

Figure A.3: **Network architecture**. A policy and value network each received a set of inputs consisting of object locations, object features, target features and the posterior over which object is the target.

A.6 Model comparison

To compare different variants of the ideal-observer model, we simulated 100 fixation trajectories per scene. We then computed the proportion of times that the model looked at each object for a given scene. To compare each resulting scene-wise probability distribution with the empirical distribution, we used cross-entropy, fitting a lapse rate for each individual participant. We report this cross-entropy metric averaged across scenes, and subtracted from a baseline model which assumes uniform relative looking times across all objects. Note that in its current iteration, the model does not account for potential variation across participants or scenes in the object kernel. In ongoing work, we are employing Approximate Bayesian Computation methods to obtain posterior probability distributions for the kernel width and lapse rate parameters of the model.