Mental Rotation as Bayesian Quadrature

Jessica B. Hamrick and Thomas L. Griffiths
Department of Psychology, University of California, Berkeley
Berkeley, CA 94720
{jhamrick,tom_griffiths}@berkeley.edu

Abstract

Given a computational resource–for example, the ability to visualize an object rotating–how do you best make use of it? We explored how mental simulation should be used in the classic psychological task of determining if two images depict the same object in different orientations. We compared two models on this mental rotation task, and found that a model based on an optimal experiment design for Bayesian quadrature is qualitatively more consistent with classic behavioral data than a simpler model. We suggest that rational models which adaptively exploit available resources are promising in their ability to characterize metacognitive processes like mental simulation.

1 Introduction

One of the challenges of solving any computational problem is determining how best to use the available computing resources. For example, a computer can render complex graphics faster by recognizing that this kind of computation should be carried out by a specialized graphics processor. The same challenge arises in designing an intelligent agent: how should the agent make best use of its computing resources? Recent research on rational models of human cognition has provided insight into the nature of the computational problems that human beings need to solve (e.g., [1, 2]), but leaves open the question of how people allocate their resources in solving those problems. In this paper, we take a step towards addressing this question, applying rational analysis (in the spirit of [3, 4, 5]) to one aspect of human metacognition: the use of mental simulation.

Consider the images on the left in Fig. 1. In each panel, are the two depicted objects identical (except for a rotation), or distinct? When presented with this mental rotation task, people default to a strategy in which they visualize one object rotating until it is congruent with the other [6]. There is strong evidence for such "mental simulation": we can imagine three-dimensional objects in our minds and manipulate them, to a certain extent, as if they were real [7]. However, the use of mental simulation is predicated on determining appropriate parameters to give the simulation, analogous to determining exactly what computation should be passed to a graphics processor. In the case of the classic mental rotation task, we might ask: How do people know which way to rotate the object? When should one stop rotating and accept the hypothesis that the objects are different?

Recent work in cognitive science has shown how the allocation of cognitive resources to solving computational problems can be analyzed using the methods of statistical decision theory [8, 9]. We suggest that this "rational metacognition" approach may also be applied to the problem of mental rotation. Specifically, we hypothesize that mental rotation can be framed as integration over a probability distribution, with the direction of rotation becoming an optimal experiment design problem (or in machine learning parlance, an active learning problem). In an initial investigation into this hypothesis, we find that recent methods for Bayesian quadrature [10, 11, 12], in contrast to a simpler heuristic model, provide a possible solution to determining the direction and extent of rotation.

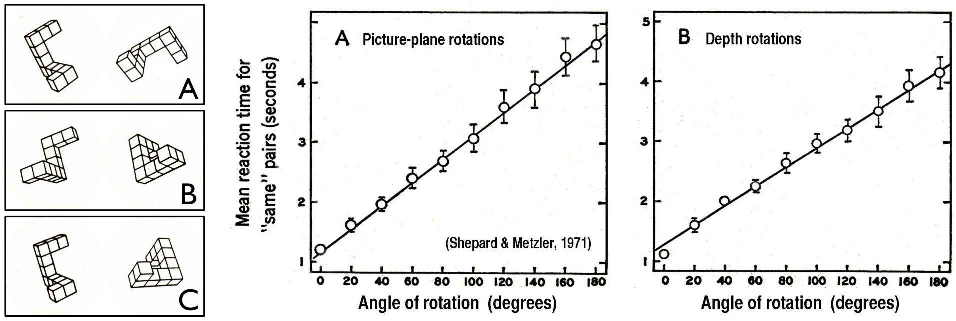


Figure 1: **Classic mental rotation task**. Participants in [6] saw stimuli such as those on the left, and judged whether each pair of shapes was the same shape in two different orientations ("same" pairs), or two different shapes ("different" pairs). **A** and **B** show "same" pairs, while **C** shows a "different" pair. The plots on the right indicate mean response times on the mental rotation task, exhibiting a strong linear relationship with increasing variance as a function of true rotation.

2 Computational-level model

We begin by analyzing mental rotation at Marr's *computational* level [3]: what is the problem to be solved? Formally, people are presented with two images, X_a and X_b, which are the coordinates of the vertices of 2D shapes (e.g., Fig. 2a). Participants must determine whether X_a and X_b depict the same shape, i.e., whether $\exists R$ s.t. $X_b = RX_a$, where R is a rotation matrix. We can formulate the judgment of whether X_a and X_b have the same origins by deciding about two hypotheses, h_0: $\forall R\ X_b \neq RX_a$ and h_1: $\exists R$ s.t. $X_b = RX_a$. To compare the hypotheses, we need to compute the posterior for each: $p(h \mid X_a, X_b) \propto p(X_a, X_b \mid h)p(h)$. Assuming the hypotheses are equally likely *a priori*, the prior term $p(h)$ will cancel out when comparing h_0 and h_1, thus allowing us to focus on the likelihoods, which are $p(X_a,\ X_b \mid h_0) = p(X_a)p(X_b)$ for h_0 and $p(X_a,\ X_b \mid h_1) = \int_R p(X_a)p(X_b|X_a, R)p(R)\ \mathrm{d}R$ for h_1. From these likelihoods, we compute the ratio ℓ which is given by $\ell = \left(\int_R p(X_b \mid X_a, R)p(R)\ \mathrm{d}R\right) / p(X_b)$. If $\ell < 1$, then h_0 is the more likely hypothesis. If $\ell > 1$, then h_1 is the more likely hypothesis.

3 Algorithmic approximation

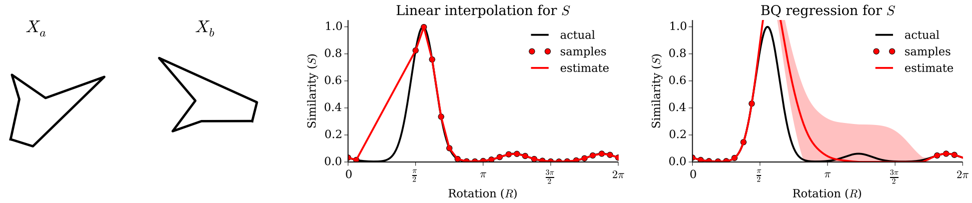
We define the prior probability of shape X to be $p(X) = n! \left(\frac{1}{2\pi}\right)^n$ according to a generative procedure.[1] This gives us the denominator of ℓ. Computing the numerator is more difficult, as we we cannot compute $p(X_b|X_a, R)$ directly. Instead, we introduce a new variable X_R denoting a mental image, which approximates RX_a. The X_R are generated by repeated application of a function τ, i.e. $X_R = RX_a = \tau(X_{R-r}, r) = \tau(\tau(X_{R-2r}, r), r) = \ldots = \tau^{(\frac{R}{r})}(X_a, r)$, where r is a small angle, and $\tau^{(i)}$ indicates i recursive applications of τ. Using this sequential function, we get:

$$
\begin{aligned}
p(X_a, X_b \mid h_1) &= \int_R \int_X p(X_b|X)p(X|X_a, R)p(X_a)p(R)\ \mathrm{d}X\ \mathrm{d}R \\
&= \int_R \int_X p(X_b|X)\delta(\tau^{(\frac{R}{r})}(X_a, r) - X)p(X_a)p(R)\ \mathrm{d}X\ \mathrm{d}R = \int_R p(X_b|X_R)p(X_a)p(R)\ \mathrm{d}R \quad (1)
\end{aligned}
$$

However, the exact form of $p(X_b|X_R)$ is still unknown. We approximate it with a similarity function $S(X_b, X_R)$, and denote the resulting integral as $Z = \int_R S(X_b, X_R)p(R)\ \mathrm{d}R \approx \int_R p(X_b|X_R)p(R)\ \mathrm{d}R$. We define the similarity based on Gaussian similarity and possible mappings M of the vertices,[2] i.e. $S(X_b, X_R) = \frac{1}{2n}\sum_{M \in \mathbb{M}} \prod_{i=1}^{n} \mathcal{N}(X_b[i] \mid (MX_R)[i], \Sigma)$ where i denotes the i^{th} vertex. An example stimulus and corresponding S is shown in Fig. 2.

[1] A set of n vertices could be chosen in any of $n!$ different ways, and each vertex is located at a random angle (between 0 and 2π) and radius (between 0 and 1).

[2] Because the vertices are connected in a way which forms a closed loop, we need only consider $2n$ mappings of the n vertices (we assume uncertainty for which is the "first" vertex, and then which of its two neighbors is the "second"). So, the possible orderings are of the form $M = \{0, 1, \ldots, n\}$, $M = \{n, 0, \ldots, n-1\}$, etc.



(a) Example stimuli (b) Naïve model (c) Bayesian Quadrature Model

Figure 2: **Example model behavior.** **(a)** An example stimulus in which the shapes differ only by a rotation. All stimuli consist of three to six vertices centered around the origin, and edges which create a closed loop from the vertices. The true angle of rotation between X_a and X_b is at $\frac{2\pi}{3}$. **(b-c)** Likelihood function and naïve (b) and BQ (c) model estimates. The sampled points $\mathbf{R}$ (red circles) are then used to estimate S (black lines are the true S, red lines are the estimate).

To summarize, the process of generating a mental image consists of computing a single X_R and then computing $S(X_b, X_R)$. We denote the sequence of rotations computed by this procedure as $\mathbf{R} = \{R_1, R_2, \dots\}$. However, this sequence cannot be arbitrary, as mental rotation is computationally demanding. Our goal is to minimize the number of rotations $|\mathbf{R}|$ while still obtaining an estimate of Z that is accurate enough to choose the correct hypothesis.

Naïve As a lower bound on performance, we defined a naïve model which performs a hill-climbing search over the similarity function until it reaches a (possibly local) maximum. Once a maximum as been found, the model computes an estimate of Z by linearly interpolating between sampled rotations (e.g., Fig. 2b).

Bayesian Quadrature A more flexible strategy uses what is known as *Bayesian Quadrature* (BQ) [10, 11] to estimate Z. BQ computes a posterior distribution over Z by placing a Gaussian Process (GP) prior on the function S and evaluating S at a particular set of points. However, while S is a non-negative likelihood function, GP regression enforces no such constraint. [12] give a method to place a prior over the *log* likelihood, thus ensuring that $S = e^{\log S}$ will be positive, i.e. $E[Z \mid \log S] = \int_{\log S} \left(\int_R \exp(\log S(X_b, X_R)) p(R)\, \mathrm{d}R \right) \mathcal{N}\left(\log S \mid \mu_{\log S}, \Sigma_{\log S}\right) \, \mathrm{d}\log S$, where $\mu_{\log S}$ and $\Sigma_{\log S}$ are the mean and covariance, respectively, of the GP over $\log S$ given $\mathbf{R}$. We approximate this according to the method given in [12], i.e. $\mu_Z = E[Z \mid S, \log S, \Delta_c] \approx \int_R \mu_S (1 + \mu_{\Delta_c}) p(R)\, \mathrm{d}R$, where μ_S is the mean of a GP over S given $\mathbf{R}$; and μ_{Δ_c} of a GP over $\Delta_c = \mu_{\log S} - \log \mu_S$ given $\mathbf{R}_c$, which consists of $\mathbf{R}$ and a set of intermediate *candidate points* c as described in [12]. The variance is $\tilde{V}(Z|S, \log S, \Delta_c)$ as defined in Equation 12 of [12].

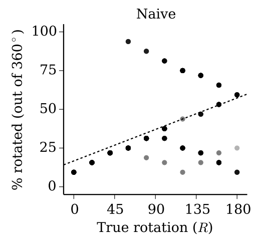
We pick the initial direction of rotation which results in the higher value of S. From then on, at each step we compute μ_Z and $\tilde{V}$ to estimate a distribution over the likelihood ratio ℓ, i.e. $p(\ell) \approx \frac{1}{p(X_b)} \mathcal{N}(Z \mid \mu_Z, \sigma_z)$. We choose h_0 when $p(\ell < 1) \geq 0.95$, and h_1 when $p(\ell > 1) \geq 0.95$.[3] Until one of these conditions are met (or the shape has been fully rotated), the model will continue to compute rotations and update its estimate of Z. Additionally, the model will change direction if doing so would lower the *expected* posterior variance of Z given some new sample a.[4] Thus, it is able to actively change its strategy, unlike the hill-climbing procedure.

4 Results

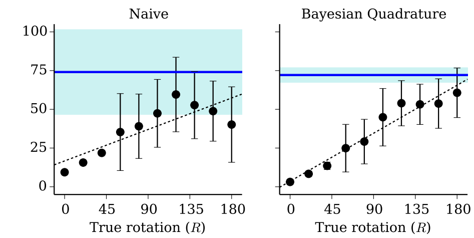
We evaluated each model's performance on 20 randomly generated shapes which had between three and six vertices, inclusive, (e.g., X_a in Fig. 2a). For each shape, we computed 18 "same" and 18 "different" stimuli pairs, with R spaced at $20'$ increments between 0 and 360, as in [6]. "Same" pairs were created by rotating X_a by R; the same was true for "different" pairs, except that X_a

[3] We chose a threshold of 0.95 because the standard confidence interval is 95%. In the future, however, this threshold could be fit to human data.

[4] This is similar to the full posterior variance calculation given in [12], however we compute the variance given only the current mean estimate of a.



(a) **All stimuli**. Each subplot shows the correspondence between the true angle of rotation (R) for "same" pairs and the amount of rotation performed by the model.



(b) **Correct stimuli**. Each subplot shows the models' mean rotations over stimuli pairs that were judged correctly. Black dots correspond to "same" pairs, and blue lines to "different" pairs.

Figure 3: **Model rotations.** Error bars/shaded regions indicate one standard deviation, and the dotted lines indicate the least-squares fit to the "same" pairs.

was also reflected across the y-axis. To gauge performance, we looked at *response error rates*: how accurate was the model at choosing the correct hypothesis? This was defined as the mean error (ME), or fraction of times the model chose incorrectly. We additionally looked at *rotations*: for those "same" pairs which the model judged correctly, how correlated were the model's rotations with the true angles of rotation? We quantified this using the Pearson's correlation coefficient ρ for the true rotation, R, versus the number of steps/rotations take by the model, $|\mathbf{R}|$. Fig. 3 shows true rotations vs (3a) the number of steps/rotations taken by the model for individual stimuli, and (3b) the average number of steps taken for each true rotation. This latter analysis can be qualitatively compared to the results of [6], as shown in Fig. 1.

Naïve The naïve model's response error rate was $\mathrm{ME} = 0.18$, which is better than chance (where chance is equivalent to guessing randomly, i.e. $\mathrm{ME} = 0.5$). The correlation between the naïve model's average rotation and the true angle of rotation was $\rho = 0.82$ (Fig. 3b, left). As shown in Fig. 3b (right), the naïve model corresponds extremely well to the true angle of rotation when $R < \frac{\pi}{2}$; this is because it needs to rotate less and is therefore less likely to get stuck on local maxima. For $R > \frac{\pi}{2}$, we see an increasing tendency to under-rotate due to getting stuck on local maxima, as well as a tendency to over-rotate if the wrong direction was initially chosen.

Bayesian Quadrature The BQ model was much more accurate in choosing the correct hypothesis than the naïve model ($\mathrm{ME} = 0.04$). The number of rotations computed by the BQ model were strongly correlated with the true rotations ($\rho = 0.98$), a result which is qualitatively similar to that exhibited by humans (Fig. 1 vs. Fig. 3b, right). Because the BQ model has the capacity to "reset", it could recover from rotating in the incorrect direction (e.g., Fig. 2c) and thus did not over-rotate as frequently as the naïve model. The BQ model also under-rotated less frequently because it rotates until it is confident in its estimate of Z and thus does not get stuck on local optima.

5 Conclusion

How do people allocate their mental simulation resources? We performed an initial investigation into the specific case of mental rotation, using rational analysis to characterize optimal strategies. We demonstrated that the classic mental rotation task [6] presents a non-trivial computational problem and cannot be solved with a simple, heuristic-based model. In contrast, an adaptive, Bayesian Quadrature model provides answers to puzzling questions surrounding the incremental nature of mental rotation: which way should the object be rotated, and for how long? This model formalizes these answers in a way that is qualitatively consistent with human behavior, both in response time linearity [6] and variability [13]. Although this research is still in its first stages, these initial results support the idea that mental rotation may be another instance in which people do appropriately use available computational resources to solve the task at hand.

Acknowledgments This research was supported by ONR MURI grant number N00014-13-1-0341, and a Berkeley Fellowship awarded to JBH.

References

[1] N. Chater and M. Oaksford, "Ten years of the rational analysis of cognition," *Trends in Cognitive Sciences*, vol. 3, no. 2, pp. 57–65, 1999.

[2] J. B. Tenenbaum, C. Kemp, T. L. Griffiths, and N. D. Goodman, "How to grow a mind: Statistics, structure, and abstraction," *Science*, vol. 331, pp. 1279–1285, 2011.

[3] D. Marr, *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. Henry Holt and Company, 1983.

[4] J. R. Anderson, *The adaptive character of thought*. Hillsdale, NJ: Erlbaum, 1990.

[5] R. N. Shepard, "Toward a universal law of generalization for psychological science," *Science*, vol. 237, pp. 1317–1323, Sept. 1987.

[6] R. N. Shepard and J. Metzler, "Mental Rotation of Three-Dimensional Objects," *Science*, vol. 171, no. 3972, pp. 701–703, 1971.

[7] S. M. Kosslyn, W. L. Thompson, and G. Ganis, *The Case for Mental Imagery*. Oxford University Press, 2009.

[8] F. Lieder, T. L. Griffiths, and N. D. Goodman, "Burn-in, bias, and the rationality of anchoring," in *Advances in Neural Information Processing Systems*, 2012.

[9] E. Vul, N. D. Goodman, T. L. Griffiths, and J. B. Tenenbaum, "One and done? Optimal decisions from very few samples," in *Proceedings of the 31st Annual Conference of the Cognitive Science Society*, pp. 148–153, 2009.

[10] P. Diaconis, "Bayesian numerical analysis," *Statistical Decision Theory and Related Topics IV*, vol. 1, pp. 163–175, 1988.

[11] A. O'Hagan, "Bayes–Hermite quadrature," *Journal of Statistical Planning and Inference*, vol. 29, no. 3, pp. 245–260, 1991.

[12] M. A. Osborne, D. Duvenaud, R. Garnett, C. E. Rasmussen, S. J. Roberts, and Z. Ghahramani, "Active Learning of Model Evidence Using Bayesian Quadrature," in *Advances in Neural Information Processing Systems*, 2012.

[13] M. A. Just and P. A. Carpenter, "Eye fixations and cognitive processes," *Cognitive Psychology*, vol. 8, pp. 441–480, 1976.