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 used in the classic psychological task of determining if two images depict the same object in different orientations. We compared three models on this mental rotation task, and found that a model based on an optimal experiment design for Bayesian quadrature is objectively more accurate and qualitatively more consistent with classic behavioral data than the other models we evaluated. 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 Figure 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 problems of allocating cognitive resources to solving computational problems can be analyzed using the methods of statistical decision theory [8, 9]. We apply this "rational metacognition" approach to the problem of mental rotation. We investigate several computational solutions to this task and qualitatively compare them to human mental rotation performance. In particular, we argue that performing 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). We show that recent work on methods for Bayesian quadrature [10, 11, 12] provides a way to solve this problem, outperforming simpler heuristics for determining the direction and extent of rotation.

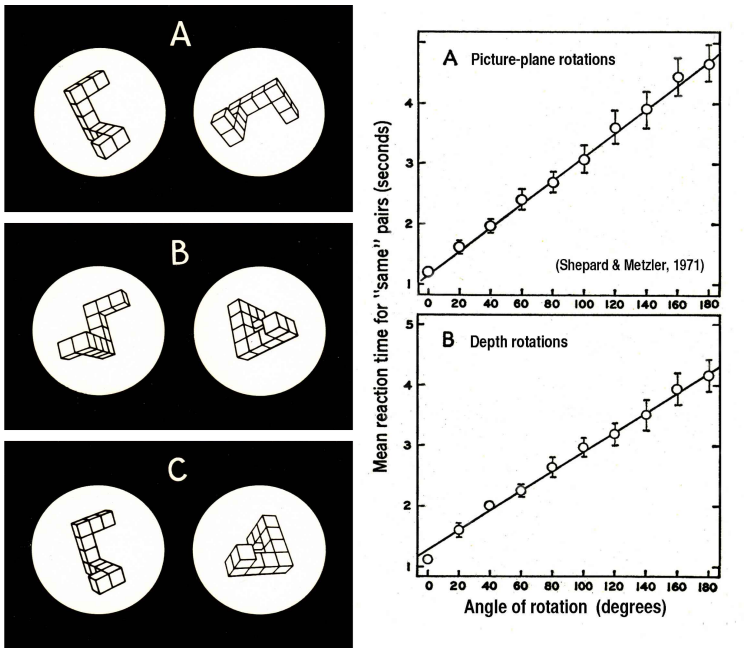


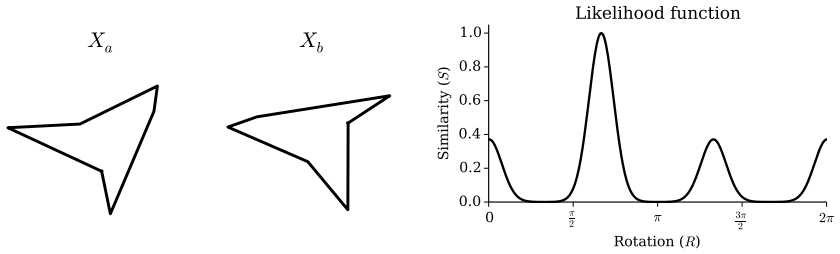
Figure 1: **Classic mental rotation task**. Participants in [6] saw stimuli such as those on the left, and judged whether each pair of shapes was the same shape in two different orientations ("same" pairs), or two different shapes ("different" pairs). **A** and **B** show "same" pairs, while **C** shows 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.

The plan of the paper is as follows. First, we give a brief overview of the work on mental rotation and active learning. Next, we present the problem domain and a computational-level analysis of how to solve it. We describe three different models which approximate this solution, evaluate their accuracy at approximating the computational solution, and qualitatively compare their behavior to that of the classic mental rotation results. We end with a brief summary and conclusion of our approach.

2 Background

There is a long and rich history of work on mental rotation beginning with Shepard and Metzler [6], who presented participants with pairs of images such as those shown in Figure 1 (left panel). These pairs of three-dimensional objects could either be: the same object rotated in the plane (Figure 1A), the same object rotated in depth (Figure 1B), or two different objects (Figure 1C; in this case, the "different" objects had reflective symmetry, but no rotational symmetry). Participants had to determine whether the two images were of the same object or not, and Shepard and Metzler famously found that response times for plane and depth rotations had a strong linear relationship with the minimum angle of rotation between the two objects (Figure 1, right panel). The conclusion was that participants were visually "rotating" the objects in their minds.

For many years, this idea was contested, with some researchers arguing the underlying cognitive processes were not visual in nature (e.g., [13]). In particular, Anderson [14] proved that the mental imagery debate could not be resolved on the basis of response time data alone. In recent years, however, significant efforts have investigated the neural underpinnings of the mental rotation phenomena, arguing that the results of brain imaging studies support the existence of visual imagery [7, 15]. We assume that mental imagery is indeed a visual process, and turn to the question of how, and when, it is used. Previous work has examined how people might use imagery [16, 17] to solve reasoning problems, and when they might use imagery as opposed to a symbolic rule [18]. Eye-tracking studies of people performing mental rotations suggest strategies that people might be using, such as aligning smaller features of the images [19]. People seem to use imagery methodically, which raises the question: what is the method?



(a) Example stimuli (b) Likelihood function

Figure 2: **Rotated shapes and their similarity. (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. **(b)** The approximate likelihood (similarity) of X_b given X_R, where X_R is a rotation of X_a by the angle R. The true angle of rotation between X_a and X_b is at $\frac{2\pi}{3}$, corresponding to the global maximum in the likelihood function.

Others studying people's pattern of behavior when engaged in self-directed learning have found that they tend to use a strategy of *active learning* [20, 21]. When given a choice as to the information they can obtain, people will not choose the information randomly (as is the case with a naïve Monte Carlo sampler), but according to some utility function such as information gain [21, 22]. We propose that such a strategy can be combined with the notion of mental imagery as a tool: an optimal model of mental rotation may, perhaps, rely on an active learning strategy. In the next sections, we investigate this idea with several different approaches.

3 Computational-level model

We begin by first analyzing mental rotation at Marr's *computational* level [3]: what is the problem to be solved, and what is the optimal way to do so? Formally, people are presented with two images, X_a and X_b, which are the coordinates of the vertices of two-dimensional shapes similar to those used by [23] (e.g., Figure 2a). Participants must determine whether X_a and X_b were generated from the same (albeit possibly transformed and permuted) original shape, i.e., whether $\exists R, M$ s.t. $X_b = MRX_a$, where M is a permutation matrix and 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: $\forall M, R\ X_b \neq MRX_a$ and h_1: $\exists M, R$ s.t. $X_b = MRX_a$. To compare the hypotheses, we need to compute the posterior for each:

$$p(h \mid X_a, X_b) \propto p(X_a, X_b \mid h)p(h). \tag{1}$$

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:

$$p(X_a,\ X_b \mid h_0) = p(X_a)p(X_b), \tag{2}$$

$$p(X_a,\ X_b \mid h_1) = \int_R \int_M p(X_a)p(X_b|X_a, R, M)p(R)p(M)\ \mathrm{d}M\ \mathrm{d}R. \tag{3}$$

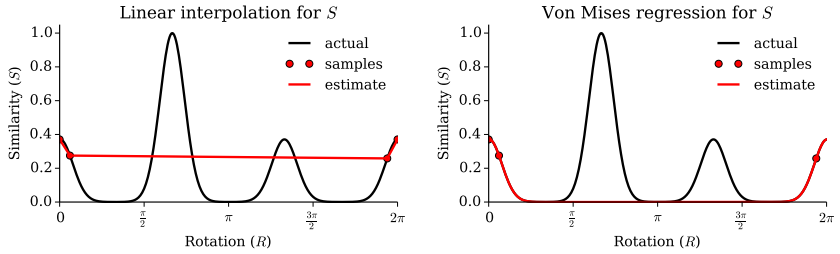
Under h_0, the likelihood is easy to compute because we assume that X_a and X_b are independent (Equation 2). When the configurations are the same, the likelihood becomes more complicated (Equation 3). For a small number of vertices, we can compute the integral over M by enumerating every possible mapping between X_a and X_b. After doing so, we obtain:

$$p(X_a,\ X_b \mid h_1) = \int_R p(X_a)p(X_b|X_a, R)p(R)\ \mathrm{d}R. \tag{4}$$

Once we have computed both likelihoods, we compute their ratio:

$$\ell = \frac{p(X_a, X_b \mid h_1)}{p(X_a, X_b \mid h_0)} = \frac{\int_R p(X_b \mid X_a, R)p(R)\ \mathrm{d}R}{p(X_b)}. \tag{5}$$

If $\ell < 1$, then h_0 is the more likely hypothesis. If $\ell > 1$, then h_1 is the more likely hypothesis.



(a) Naïve model (b) Parametric model

Figure 3: **Simple methods of estimating S.** In both cases, hill-climbing search is used until a maxima is found (in this case, at approximately $\frac{\pi}{6}$). The sampled points $\mathbf{R}$ (red circles) are then used to estimate S (black lines are the true S, red lines are the estimate). **(a)** Linear interpolation. The overall estimate of Z here will be too low, but certain angles will have a disproportionate contribution (e.g., between $\frac{\pi}{4}$ and $\frac{3\pi}{4}$. **(b)** Best fit of scaled Von Mises PDF parameters. As in (a), Z will be underestimated. However, the fit around the maximum near $\frac{\pi}{6}$ is much more accurate.

4 Metacognitive algorithms

Given the computational model defined above, we now explore the ways in which it can be expressed algorithmically. First, we define the prior probabilities over stimuli according to a generative procedure. 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). Thus, the prior over a shape X is $p(X) = n!\left(\frac{1}{2\pi}\right)^n$, which gives us the denominator in Equation 5. 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:

$$\begin{aligned} X_R = RX_a &= \tau(X_{R-r}, r) = \tau(\tau(X_{R-2r}, r), r) = \dots \\ &= \tau^{(\frac{R}{r})}(X_a, r). \end{aligned} \tag{6}$$

where r is a small angle, and $\tau^{(i)}$ indicates i recursive applications of τ. Using this sequential function, we get:

$$\begin{aligned} p(X_a, X_b \mid h_1) &= \int_R \int_X p(X_b|X)p(X|X_a, R)p(X_a)p(R)\,\mathrm{d}X\,\mathrm{d}R \\ &= \int_R \int_X p(X_b|X)\delta(\tau^{(\frac{R}{r})}(X_a, r) - X)p(X_a)p(R)\,\mathrm{d}X\,\mathrm{d}R \\ &= \int_R p(X_b|X_R)p(X_a)p(R)\,\mathrm{d}R \end{aligned} \tag{7}$$

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:

$$Z = \int_R S(X_b, X_R)p(R)\,\mathrm{d}R \approx \int_R p(X_b|X_R)p(R)\,\mathrm{d}R. \tag{8}$$

We define the similarity function S, which incorporates the different mappings as vertices, as follows. 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 $M = \{0, 1, \dots, n\}$, $M = \{n, 0, \dots, n-1\}$, and so on. Combining this with a Gaussian similarity metric, we obtain:

$$S(X_b, X_R) = \frac{1}{2n}\sum_{M\in\mathbb{M}}\prod_{i=1}^{n}\mathcal{N}(X_b[i] \mid (MX_R)[i], \Sigma) \tag{9}$$

where i denotes the i^{th} vertex. An example stimulus and corresponding S is shown in Figure 2.

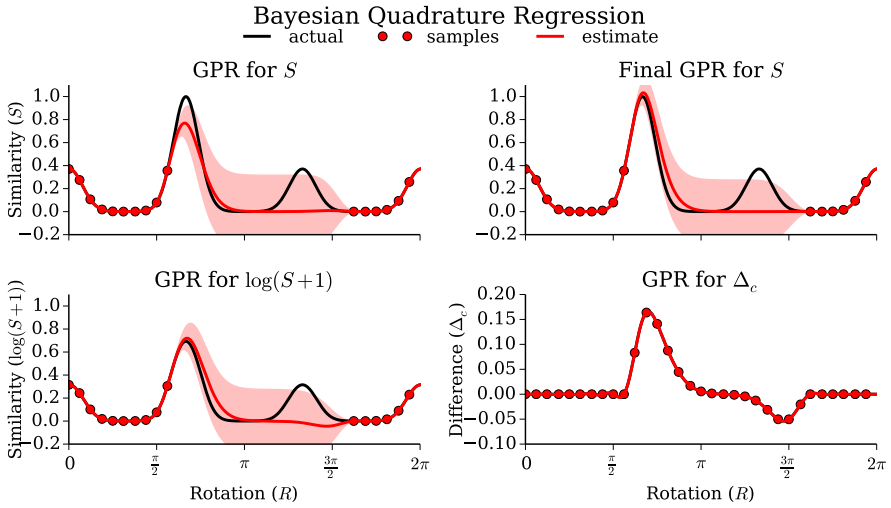


Figure 4: **Nonparametric model.** Each panel shows one step of the Bayesian Quadrature regression. Upper left: the original Gaussian Process (GP) regression for S. Lower left: the GP regression for $\log(S+1)$. Lower right: the GP regression for $\Delta_c = \mu_{\log S} - \log \mu_S$. Upper right: the adjusted regression for S, where the mean is equal to $\mu_S(1 + \mu_{\Delta_c})$. The model uses this final estimate to compute Z and will continue rotating until the variance of Z is low enough that a hypothesis may be accepted. This method allows the model to avoid local maxima such as the one near $\frac{\pi}{6}$, which causes trouble for the naïve and parametric models in Figure 3.

To summarize, the process of generating a mental image consists of computing a single X_R (as in Equation 6) 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. With this in mind, we now examine several approaches to solve this problem.

Gold standard To compare the accuracy of other models' estimates of Z, we computed a "gold standard" by evaluating $S(X_b, X_R)$ at 360 values of R spaced evenly between 0 and 2π and estimating the integral using the trapezoidal rule.

Naïve As a baseline, 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. Figure 3a shows an example of the naïve model's estimate of S.

Parametric (Von Mises) Another strategy is to assume a parametric shape for S and fit the appropriate parameters. If we assume that S is approximately unimodal when h_1 is true[1], then a reasonable form for the similarity function is that of a "circular Gaussian" or Von Mises distribution:

$$S(X_b, X_R) \approx h \cdot p(R \mid \hat{\theta}, \kappa) = \frac{h}{2\pi I_0(\kappa)} e^{\kappa \cos(R - \hat{\theta})} \tag{10}$$

where κ is the concentration parameter, $\hat{\theta}$ is the preferred direction, h is a scale parameter, and I_0 is the modified Bessel function of order zero. We fit these parameters by minimizing the mean squared error between this PDF and the computed values of S. To choose the sequence of rotations, we use the same hill-climbing strategy as in the naïve model. Figure 3b shows an example of the parametric model's estimate of S.

[1]This is not always a reasonable assumption, see e.g., Figure 3b.

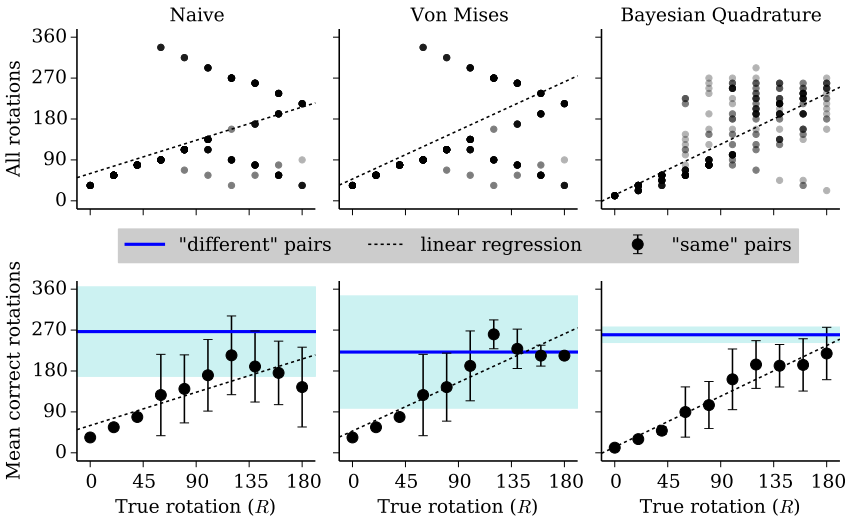


Figure 5: **Model rotations.** Top: each subplot shows the correspondence between the true angle of rotation (R) for "same" pairs and the amount of rotation performed by the model. Bottom: each subplot shows the models' mean rotations over stimuli pairs that were judged correctly. Black dots correspond to "same" pairs, and blue lines correspond to "different" pairs (for which the mean rotation is constant across true rotations, as the true rotation is undefined). Error bars/shaded regions indicate one standard deviation, and the dotted lines indicate the least-squares fit to the "same" pairs.

Nonparametric (Bayesian Quadrature) A more flexible strategy uses what is known as *Bayesian Quadrature* [10, 11] to estimate Z. Bayesian Quadrature allows us to compute 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. Because our data is circular, we use a periodic kernel [24]:

$$k(R, R') = h^2 \exp\left(-\frac{2\sin^2\left(\frac{1}{2}(R - R')\right)}{w^2}\right). \tag{11}$$

Bayesian Quadrature has its difficulties, however. While in our case S is a non-negative likelihood function, GP regression enforces no such constraint. In an effort to avoid this problem, [12] give a method to place a prior over the log likelihood[2], thus ensuring that $S = e^{\log S}$ will be positive[3]:

$$E[Z \mid \log S] = \int_{\log S} \left(\int_R \exp(\log S(X_b, X_R)) p(R)\, \mathrm{d}R\right) \mathcal{N}\left(\log S \mid \mu_{\log S}, \Sigma_{\log S}\right)\, \mathrm{d}\log S$$

where $\mu_{\log S}$ and $\Sigma_{\log S}$ are the mean and covariance, respectively, of the GP regression over $\log S$ given $\mathbf{R}$. We approximate this according to the method given in [12]:

$$\mu_Z = E[Z \mid S, \log S, \Delta_c] \approx \int_R \mu_S(1 + \mu_{\Delta_c}) p(R)\, \mathrm{d}R \tag{12}$$

where μ_S is the mean of a GP regression over S given $\mathbf{R}$; and μ_{Δ_c} is a regression 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].

[2]In practice, as in [12], we use the transform of $\log(S + 1)$. Additionally, we follow [12] and use a combination of MLII and marginalization to fit the w kernel parameters. The output scales are fixed at $h_S = \sqrt{0.1}$ and $h_{\log S} = \log(h_S + 1)$.

[3]We are not guaranteed positivity, however, as the approximation to the integral over $\log S$ (Equation 12) requires computing μ_S, which may have non-positive segments.

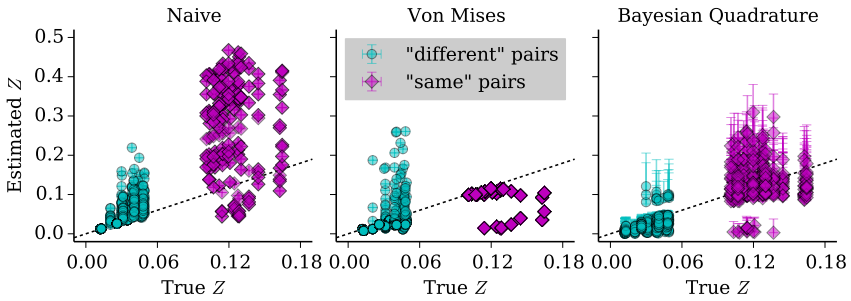


Figure 6: **Accuracy in estimating Z.** Each subplot shows the true ("gold standard") value of Z vs. the value estimated by the model. Black dotted lines indicate a perfect 1:1 correspondence between the true and estimated values. The naïve model (left panel) tends to overestimate. The parametric model (center panel) is very accurate for about half the "same" pairs, and severely underestimates the rest. The nonparametric model (right panel) maintains a decent correspondence with the true Z.

To start, we pick an initial direction of rotation which results in the higher value of S. At each step, we compute μ_Z and $\tilde{V}$ in order to estimate a distribution over the likelihood ratio ℓ:

$$p(\ell) \approx \frac{1}{p(X_b)} \mathcal{N}(Z \mid \mu_Z, \sigma_z).$$

We choose h_0 when $p(\ell < 1) \geq 0.95$, and h_1 when $p(\ell > 1) \geq 0.95$. Until one of these conditions are met (or the shape has been fully rotated), the model will continues to compute rotations and update its estimate of Z.

We additionally allow the model to change direction or "reset" based on an estimate of the posterior variance of Z given some new sample a. This is similar to the procedure given in [12], however we do not compute the full posterior variance. Instead, we compute the variance given only the current mean estimate of a. If this estimated variance is lowered more by "resetting", then the model will change directions. Thus, it is able to have an active say in which mental rotations are computed, unlike the hill-climbing procedure.

5 Results

We evaluated each model's performance on 20 shapes which had between 3 and 6 vertices, inclusive, (e.g., X_a in Figure 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", except that X_a was also reflected across the y-axis. Each shape was generated randomly, but under the constraint that the "same"/"different" pairs created from the shape could both be judged according to the "gold standard".

We considered three metrics of performance in particular:

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.

Rotations: For those "same" pairs which the model judged correct, 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 extent of the model's rotations ($|\mathbf{R}|$). The top row in Figure 5 shows individual points corresponding to true rotations vs. rotations by the models for all stimuli, including those judged incorrectly. The bottom row shows mean model rotations, across only those stimuli which were judged correctly. We structured the analysis in this way to better compare to the results of [6], which also excluded incorrectly judged stimuli.

Estimates of Z: How accurate was the model's estimate of Z? We defined this quantity as the mean squared error (MSE) between the model's estimate of Z and the "gold standard" value, where the error has scale such that $\text{MSE} = 0$ indicates no error and $\text{MSE} = 1$ indicates maximum error. Figure 6 shows plots of the true ("gold standard") value of Z vs. the model's estimate, for each model.

Naïve The naïve model's response error rate was $\mathrm{ME} = 0.18$, which is better than chance (equivalent to guessing randomly, i.e. $\mathrm{ME} = 0.5$). Closer inspection reveals that much of this comes from "different" pairs ($\mathrm{ME} = 0.29$), where the asymmetric linear interpolation may give an overestimate of Z (e.g., Figure 3a). This intuition is confirmed by the model's accuracy in estimating Z, which was $\mathrm{MSE} = 1.27$. As shown in the left panel of Figure 6, it overestimated Z for nearly all stimuli.

The correlation between the naïve model's average rotation and the true angle of rotation was $\rho = 0.82$ (Figure 5, bottom left). More complex patterns are revealed by examining the shape of the raw data in Figure 5, top left: the naïve model corresponds extremely well to the true angle of rotation for approximately $R < \frac{\pi}{2}$. This is unsurprising, because the closer the true angle is to zero, the less the model has to rotate, and the less likely it will get stuck on local maxima. Thus, it is more likely to locate the global maximum, which corresponds to the true angle of rotation. 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.

Parametric (Von Mises) The parametric model's error in determining whether the objects were identical was $\mathrm{ME} = 0.16$, making more errors on "same" pairs ($\mathrm{ME} = 0.19$) than "different" pairs ($\mathrm{ME} = 0.12$). Rather than overestimating as with the naïve model, however, the parametric model largely underestimated Z. This is unsurprising: because the Von Mises distribution only has a single peak, it necessarily underestimates Z for multimodal similarity functions. Overall, the parametric model had an error rate of $\mathrm{MSE} = 0.09$ in estimating Z, which was significantly more accurate than that of the naïve model.

The correlation between the parametric model's average rotation and the true angle of rotation was also higher ($\rho = 0.91$), though this is because more of the underestimated stimuli pairs were excluded from the analysis. The individual rotations performed by the parametric and naïve models were actually identical (see Figure 5, upper left and center) because both models use the same hill-climbing stopping strategy.

Nonparametric (Bayesian Quadrature) The nonparametric model was much more accurate in choosing the correct hypothesis than the other two models ($\mathrm{ME} = 0.04$). We note that this is very close to the 3.2% error rate reported by [6], though further experimentation is necessary to determine whether the types of errors people make align with those of the model.

The average rotations computed by the nonparametric model were strongly correlated with the true rotations ($\rho = 0.98$, see Figure 5, bottom right). Because the nonparametric model has the capacity to "reset", it could recover from rotating in the incorrect direction (e.g., Figure 4) and thus did not over-rotate as frequently. It also under-rotated less frequently: because the model continues to rotate until it is confident that its estimate of Z is accurate, it does not get stuck as easily on local optima. Indeed, it was largely successful, with $\mathrm{MSE} = 0.06$.

6 Conclusion

In this paper, we asked: how do people use mental simulation? We investigated the specific case of mental rotation, using rational analysis to characterize optimal strategies for performing mental rotation. We found that an adaptive, nonparametric model performs best at the classic rotation task [6], as it is able to actively monitor the confidence of its estimate and intelligently choose the direction of rotation. Two simpler models based on heuristics performed much worse: they did not maintain the linear relationship with the true angle of rotation due to lack of robustness against local maxima or incorrect rotation direction. More broadly, the nonparametric 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 [19]. These results further indicate that this may be another instance in which people do appropriately use available computational resources to solve the task at hand. Rational, adaptive approaches such as this may therefore be successful when applied to more complex cognitive processes above and beyond mental rotation.

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] Z. W. Pylyshyn, "The imagery debate: Analogue media versus tacit knowledge," *Psychological Review*, vol. 88, no. 1, pp. 16–45, 1981.

[14] J. R. Anderson, "Arguments concerning representations for mental imagery," *Psychological Review*, vol. 85, no. 4, p. 249, 1978.

[15] S. M. Kosslyn, "Aspects of a cognitive neuroscience of mental imagery," *Science*, vol. 240, no. 4859, pp. 1621–1626, 1988.

[16] M. Hegarty, "Mechanical reasoning by mental simulation," *Trends in Cognitive Sciences*, vol. 8, no. 6, pp. 280–285, 2004.

[17] D. L. Schwartz and T. Black, "Inferences Through Imagined Actions: Knowing by Simulated Doing," *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol. 25, no. 1, pp. 116–136, 1999.

[18] D. L. Schwartz and J. B. Black, "Shuttling between depictive models and abstract rules: Induction and fallback," *Cognitive Science*, vol. 20, no. 4, pp. 457–497, 1996.

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

[20] T. M. Gureckis and D. B. Markant, "Self-Directed Learning: A Cognitive and Computational Perspective," *Perspectives on Psychological Science*, vol. 7, pp. 464–481, Sept. 2012.

[21] D. B. Markant and T. M. Gureckis, "Does the utility of information influence sampling behavior?," in *Proceedings of the 34th Annual Conference of the Cognitive Science Society*, 2012.

[22] J. D. Nelson, "Towards a rational theory of human information acquisition," in *The Probabilistic Mind: Prospects for Bayesian Cognitive Science* (M. Oaksford and N. Chater, eds.), pp. 143–164, Oxford University Press, 2008.

[23] L. A. Cooper, "Mental Rotation of Random Two-Dimensional Shapes," *Cognitive Psychology*, vol. 7, pp. 20–43, 1975.

[24] C. E. Rasmussen and C. K. I. Williams, *Gaussian processes for machine learning*. MIT Press, 2006.