Optimal strategies in mental rotation tasks

Jessica B. Hamrick
Department of Psychology
University of California, Berkeley
Berkeley, CA 94720
jhamrick@berkeley.edu

1 Introduction

Consider the objects 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 [1]. There is strong evidence for such *mental imagery* or *mental simulation*: we can imagine three-dimensional objects in our minds and manipulate them, to a certain extent, as if they were real [2]. However, the use of mental simulation is predicated on determining appropriate parameters to give the simulation, and people's cognitive constraints may furthermore place limitations on the duration or precision of simulation. One hypothesis for how these issues are handled argues that people use a rational solution, meaning that it is optimal under existing constraints [3, 4, 5], such as processing speed, memory, or perceptual acuity.

In the case of the classic mental rotation task, we might ask: in what direction should the object be rotated? What are the requirements for "congruence"? When should one stop rotating and accept the hypothesis that the objects are different? We investigate several computational solutions to this task and qualitatively compare them to general mental rotation results. Of particular interest is a model which actively reasons about the amount of evidence it has, and performs mental rotations until it reaches a certain level of confidence. We contrast this model with simpler approaches which are less computationally demanding, but which employ less flexible strategies.

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 discussion of the strengths and weaknesses of each model, as well as directions for future work.

2 Background

There is a long and rich history of work on mental rotation beginning with [1], who presented participants with pairs of images such as those shown in Figure 1. These pairs three-dimensional objects could either be: the same object rotated in 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 [1] 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. 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. [6]). In particular, [7] proved that the mental imagery debate could not be resolved on the basis of response time data alone. In recent years, however, a significant

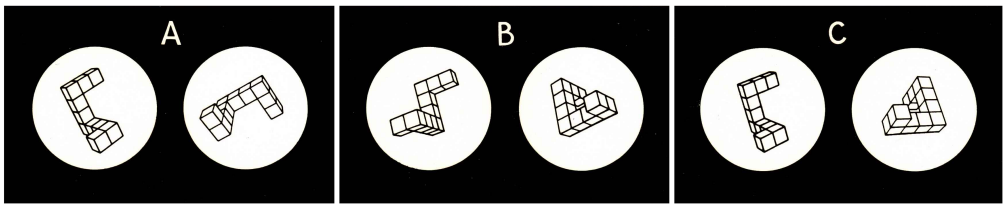


Figure 1: **Classic mental rotation task**. Participants in [1] saw stimuli such as these, and judged whether each pair of shapes was the same shape in two different orientations, or two different shapes. **A** shows a "plane rotation", **B** shows a "depth rotation", and **C** shows two distinct objects.

amount of work has investigated the neural underpinnings of the mental rotation phenomena, arguing that the results of brain imaging studies support the existence of visual mental rotation [8, 2].

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 [9, 10] to solve reasoning problems, and when they might use imagery as opposed to a symbolic rule [11]. People seem to use imagery methodically, which begs the question: what is the method?

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* [12, 13]. 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 [14, 13]. 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 strategy. In the next sections, we investigate this idea with several different approaches.

3 Computational-level model

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 [15] (e.g., see 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: $\forall M, R\ X_b \neq MRX_a$
- h_1: $\exists M, R$ s.t. $X_b = MRX_a$

To compare the hypotheses, we need to compute the posterior:

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

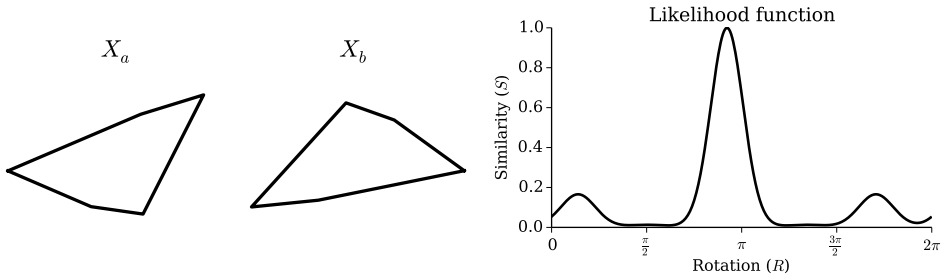
Assuming the hypotheses are equally likely *a priori*, the prior term $p(h_i)$ will cancel out when comparing h_0 and h_1, thus allowing us to focus on the likelihoods:

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

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

Under h_0, the likelihood is easy to compute because we assume that X_a and X_b are independent (Equation 2). However, 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:

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



(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 five vertices centered around the origin, and four 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 slightly less than π, corresponding to the global maximum in the likelihood function.

However, 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 sequentially by repeated application of a function τ:

$$\begin{aligned} X_R &= RX_a \\ &= \tau(X_{R-r}, r) \\ &= \tau(\tau(X_{R-2r}, r), r) \\ &\dots \\ &= \tau^{(\frac{R}{r})}(X_a, r) \end{aligned} \tag{5}$$

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

$$\begin{aligned} p(X_a, X_b \mid h_1) &= \int_R \int_X p(X_b|X)p(X|X_a, R)p(X_a)p(R)\,\mathrm{d}X\,\mathrm{d}R \\ x &= \int_R \int_X p(X_b|X)\delta(\tau^{(\frac{R}{r})}(X_a, r) - X)p(X_a)p(R)\,\mathrm{d}X\,\mathrm{d}R \\ &= \int_R p(X_b|X_R)p(X_a)p(R)\,\mathrm{d}R \end{aligned} \tag{6}$$

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

$$\ell = \frac{p(X_a, X_b \mid h_1)}{p(X_a, X_b \mid h_0)} = \frac{\int_R p(X_b \mid X_R)p(R)\,\mathrm{d}R}{p(X_b)} \tag{7}$$

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

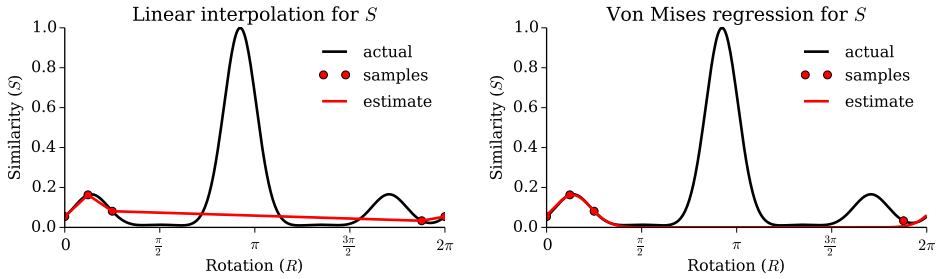
4 Implementation

We define the prior probabilities over stimuli based on how they are generated. For shapes with n vertices, each vertex is at a random angle with a radius random chosen between 0 and 1, in polar coordinates, and could be chosen in any of $n!$ different ways. Thus, the prior over a shape X is:

$$p(X) = n! \left(\frac{1}{2\pi}\right)^n \tag{8}$$

which gives us the denominator in Equation 7.

Computing the numerator of Equation 7 is more difficult, as we do know $p(X_b|X_R)$. We approximate it with a similarity function $S(X_b, X_R)$, which also takes into account the different possible



(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.

mappings of vertices:

$$Z = \int_R S(X_b, X_R) p(R)\, \mathrm{d}R \approx \int_R p(X_b \mid X_R) p(R)\, \mathrm{d}R \tag{9}$$

Because the vertices are connected in a way which forms a closed loop, we need only consider n mappings, M, of the n vertices (we assume the uncertainty is only in which is the "first" vertex). Thus, the possible orderings are of the form $M = \{0, 1, \ldots, n\}$, $M = \{n, 0, \ldots, n-1\}$, and so on. This gives us an explicit form for the similarity function:

$$S(X_b, X_R) = \frac{1}{n} \sum_{M \in \mathbb{M}} \prod_{i=1}^{n} \mathcal{N}(X_b[i] \mid (MX_R)[i], \Sigma) \tag{10}$$

where i denotes the i^{th} vertex. An example of S for the stimuli shown in Figure 2a is illustrated in Figure 2b.

This process of mental rotation (i.e., generating a single X_R (as in Equation 5) and then computing $S(X_b, X_R)$) is a computationally demanding cognitive process. Thus, our goal is to minimize the number of rotations 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 estimating Z.

In each of these models, we denote the computed rotations to be a set $\mathbf{R} = \{R_1, R_2, \ldots\}$.

4.1 Gold standard

To compare the accuracy of other models' estimates of Z, we computed a "gold standard"[1] by evaluating $S(X_b, X_R)$ at 360 values of R spaced evenly between 0 and 2π.

4.2 Naïve model

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 prior to estimating Z.

4.3 Parametric (Von Mises) model

Another strategy is to assume a parametric shape for S and fit the appropriate parameters. When h_1 is correct, it is likely that the function will be approximately unimodal (shapes with rotational

[1]This only gives an accurate estimate of Z, which is itself an approximation, and is thus not necessarily the true value of the numerator in Equation 7.

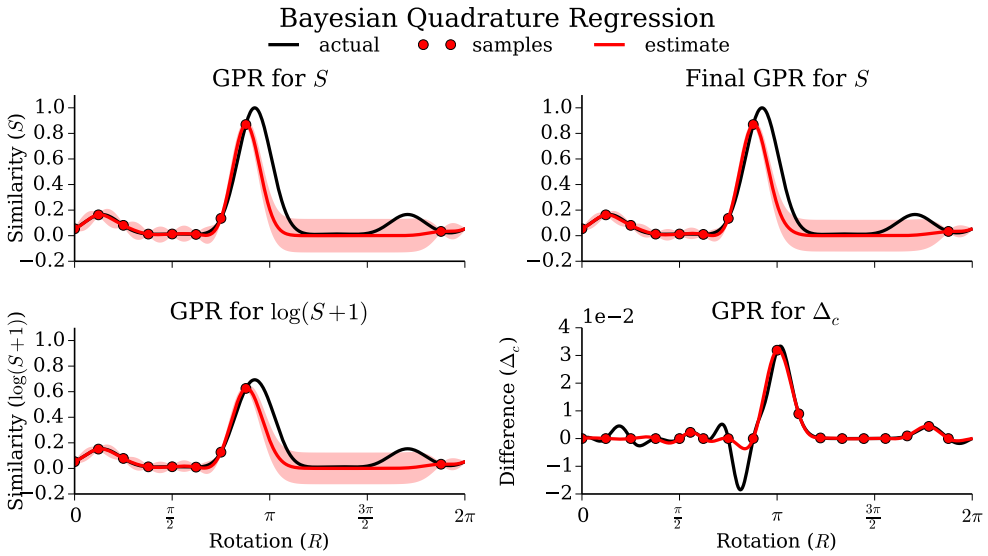


Figure 4: **Nonparametric model.** Each panel shows one step of the Bayesian Quadrature regression. Upper left: the original Gaussian Process (GP) regression for S. Lower left: the GP regression for $\log(S+1)$. Lower right: the GP regression for $\Delta_c = \mu_{\log S} - \log \mu_S$. Upper right: the 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.

symmetry would be multimodal). A reasonable assumption, then, is that the likelihood follows a "wrapped Gaussian", or Von Mises, distribution:

$$S(X_b, X_R) \approx h \cdot p(R \mid \hat{\theta}, \kappa) = \frac{h}{2\pi I_0(\kappa)} e^{\kappa \cos(R - \hat{\theta})} \tag{11}$$

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 this parametric model, prior to estimating Z.

4.4 Nonparametric (Bayesian Quadrature) model

A more flexible strategy uses what is known as *Bayesian Quadrature* [16, 17] 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 can use a periodic kernel [18]:

$$k(R, R') = h^2 \exp\left(-\frac{2\sin^2\left(\frac{1}{2}(R - R')\right)}{w^2}\right) \tag{12}$$

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, [19] give a method which involves instead placing a prior over the log likelihood[2], thus ensuring that $S = e^{\log S}$

[2]In practice, as in [19], we use the transform of $\log(S + 1)$. Additionally, while [19] use a combination of MLII and marginalization to fit the kernel parameters, we set the output scale $h_S = 0.125$ and $h_{\log S} = \log(h_S + 1)$, and use MLII to fit all other parameters.

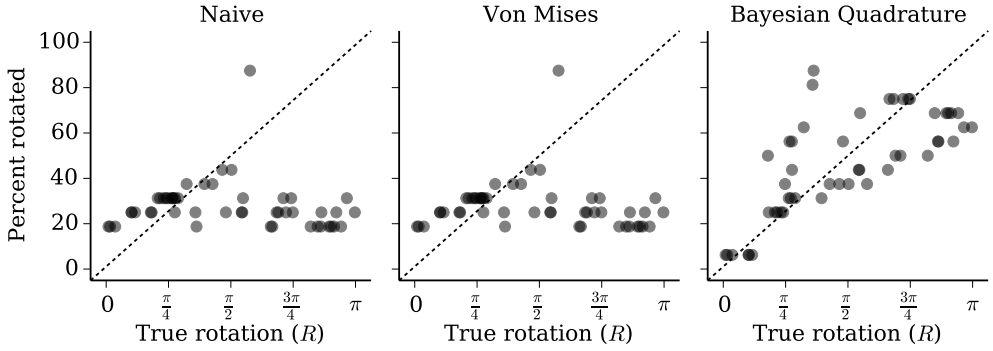


Figure 5: **Model rotations.** Each subplot shows the correspondence between the true angle of rotation (R) for h_1 stimuli and the amount of rotation performed by the model. Dotted lines indicate a perfect 1:1 relationship. The naïve and parametric models (left and center panels, respectively) are identical; they show modest agreement until roughly $\frac{\pi}{2}$, after which agreement decreases as they tend to encounter local maxima. The nonparametric model (right panel) exhibits a significant correlation between the true angle and the amount the model rotates.

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 [19]:

$$\mu_Z = E[Z \mid S, \log S, \Delta_c] \approx \int_R \mu_S (1 + \mu_{\Delta_c}) p(R) \, \mathrm{d}R \tag{13}$$

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*[4] c as described in [19]. The variance of the estimate of Z is given by:

$$\sigma_Z = \mathrm{Var}(Z \mid S, \log S, \Delta_c) = \int_R \int_{R'} \mu_S(R) \mu_S(R') \Sigma_{\log S}(R, R') p(R) p(R') \, \mathrm{d}R \, \mathrm{d}R' \tag{14}$$

To choose $\mathbf{R}$, we pick an initial direction of rotation which results in the higher value of S, and then continue rotating in that direction until the variance in Equation 14 is low enough that we are confident about the likelihood ratio ℓ:

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

Specifically, we choose h_0 when $p(\ell < 1) \geq 0.95$, and we choose 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 continue to compute rotations and update its estimate of Z.

5 Results

We evaluated each model's performance on 50 randomly generated shapes (such as X_a in Figure 2a). For each shape, we picked a random angle R and generated two stimuli: a h_0 stimulus, by reflecting X_a across the y-axis and rotating it by R; and a h_1 stimulus, by simply rotating X_a by R.

We considered three metrics of performance in particular:

[3]We are not guaranteed positivity, however, the approximation to the integral over $\log S$ (Equation 13) requires computing μ_S, which may have non-positive segments.

[4]These candidate points do not require evaluating the true S, only the GP estimates of S and $\log S$.

- *Rotations*: for stimuli where h_1 is correct, how correlated are the true angles of rotation with the rotations performed by the model? This is defined to be the Pearson's correlation coefficient ρ between R and $|\mathbf{R}|$. Figure 5 shows individual points corresponding to true rotation vs. rotation by the model, for each model.
- *Responses*: how accurate is the model at choosing the correct hypothesis? This is defined to be the mean error (ME), or fraction of times the model picks the incorrect hypothesis.
- *Estimates of Z*: how accurate is the model's estimate of Z? This is defined to be the mean squared error (MSE) between the model's estimate of Z and the "gold standard" value, where the error has scale such that $\mathrm{MSE} = 0$ indicates no error, and $\mathrm{MSE} = 1$ indicates maximum error. Figure 6 shows plots of the true (gold standard) value of Z vs. the model's estimate, for each model.

Gold standard The gold standard always evaluates S at all angles, so the correlation between it and the true angle of rotation is undefined. The mean error of responses was $\mathrm{ME} = 0.01$, indicating that Z is a sufficient approximation of $\int_R p(X_b \mid X_R)p(R)\,\mathrm{d}R$.

Naïve model The correlation between the number of rotations computed by the naïve model and the true angle of rotation was $\rho = -0.11$. The shape of the data is more interesting (see Figure 5, left panel): the naïve model actually corresponds quite well to the true angle of rotation for $R < \frac{\pi}{2}$ ($\rho = 0.64$). 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 generally corresponds to the true angle of rotation. For $R > \frac{\pi}{2}$, we see an increasing tendency to under-rotate ($\rho = -0.42$); this is likely because it finds a local maxima and ends prematurely.

The naïve model's response error rate was better than chance, but still quite high, with $\mathrm{ME} = 0.23$. Closer inspection reveals that the bulk of this comes from h_1 stimuli ($\mathrm{ME} = 0.36$ vs. $\mathrm{ME} = 0.1$ for h_0 stimuli). This is, as above, probably related to finding only local maxima: if the model finds a local maxima which is low enough, the area under the estimated curve will not be large enough to accept h_1.

This is reflected in the accuracy in estimating Z as well ($\mathrm{MSE} = 0.14$). Figure 6 (left panel) shows individual estimates for each stimulus, color-coded by true hypothesis. The model underestimates Z for significant portion of the h_1 stimuli. On the other h_1 stimuli, the naïve model actually tends to overestimate because linear interpolation does not account for symmetry of global maxima (e.g., Figure 3a).

Parametric (Von Mises) model Like the naïve model, the parametric model uses hill-climbing as a sampling strategy, so they have identical correlations with the true angle of rotation. However, given the same $\mathbf{R}$, the parametric model estimates Z differently, thus giving rise to different response and estimated Z error rates. The response error for the parametric model was higher than the naïve model, with $\mathrm{ME} = 0.26$. Breaking this down, we see that the error for h_1 stimuli is $\mathrm{ME} = 0.5$: in other words, the parametric model is at chance when determining whether two identical shapes are the same. This is because it never overestimates (see Figure 6, center panel). So, it accurately estimates Z for stimuli which the naïve model would overestimate ($\mathrm{MSE} = 0.07$), but still has poor performance when Z is underestimated. Indeed, if we exclude stimuli for which the naïve model overestimates ($Z \geq 0.3$), the error for the naïve model lowers to match that of the parametric model.

Nonparametric (Bayesian Quadrature) model The rotations computed by the nonparametric model were strongly correlated with the true rotations ($\rho = 0.73$, see Figure 5, right panel). This is because the nonparametric model does not get stuck as easily on local optima: it will continue rotating until it is confident that its estimate of Z is accurate.

Correspondingly, the nonparametric model is much more accurate in choosing the correct hypothesis ($\mathrm{ME} = 0.06$). As with the naïve and parametric models, it mostly errs on h_1 stimuli ($\mathrm{ME} = 0.1$). This appears to mostly be the result of choosing the incorrect initial direction of rotation.

Similarly, because the nonparametric model actively attempts to obtain an accurate estimate of Z, its estimate of Z is indeed fairly accurate ($\mathrm{MSE} = 0.10$). This is slightly higher than the nonparametric

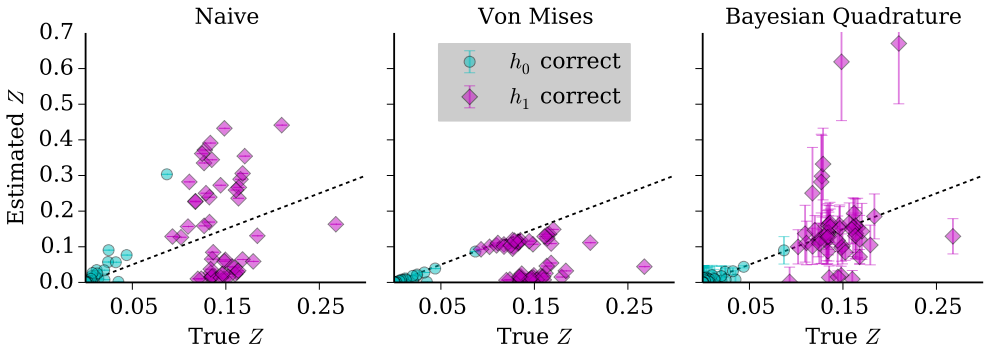


Figure 6: **Accuracy in estimating Z.** Each subplot shows the true ("gold standard") value of Z vs. the value estimated by the model. Black dotted lines indicate a perfect 1:1 correspondence between the true and estimated values. All models perform well for h_0 stimuli, but exhibit different behavior for h_1 stimuli. The naïve model (left panel) tends to either greatly overestimate, or underestimate. The parametric model (center panel) is very accurate for about half the h_1 stimuli, and severely underestimates the rest. The nonparametric model (right panel) maintains a decent correspondence with the true Z, with the exception of a handful of outliers.

model, but is due to a few outliers. If we again exclude stimuli for which Z is greatly overestimated ($Z \geq 0.3$), the nonparametric model's error decreases significantly ($\text{MSE} = 0.03$).

6 Discussion

We investigated strategies for performing the mental rotation task [1] in two dimensions, and found that a nonparametric model that enforces a positivity constraint on the likelihood function performs best, as it is able to actively monitor the confidence of its estimate. The simpler models performed much worse: they did not maintain the linear relationship with the true angle of rotation due to lack of robustness against local maxima, and they were inaccurate at identifying whether the two shapes were the same or not.

One option for improving the performance of the simple models would be to fit a threshold value, below which local maxima would be ignored. However, this strategy would be rather brittle, for if the distribution of shapes changed, the threshold would have to be re-learned. It is possible that people exhibit this behavior, but we cannot make any assumptions one way or the other without empirical data.

A further strength of the nonparametric model over the other models, however, is that it is likely to generalize well to three dimensions. One aspect of [19] which we have not yet explored is their main contribution of *active sampling*, in which samples are iteratively selected to maximally decrease the expected variance of Z. Due to the sequential constraint of mental rotation, this strategy is not particularly useful in two dimensions, as there are only every two directions in which to rotate. In three dimensions, however, there are an infinite number of directions that could be chosen after every step.

We conclude that, from this initial survey, models of mental rotation which take an "active", directed approach seem well-suited to explaining human behavior in these tasks. Future work will collect empirical data from participants to perform a more quantitative analysis of these modeling tools.

References

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

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

[3] 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.

[4] 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.

[5] T. L. Griffiths, E. Vul, and A. N. Sanborn, "Bridging Levels of Analysis for Probabilistic Models of Cognition," *Current Directions in Psychological Science*, vol. 21, no. 4, pp. 263–268, 2012.

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

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

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

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

[10] 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.

[11] 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.

[12] 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.

[13] 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.

[14] 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.

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

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

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

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

[19] 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.