Psychological Review
2015, Vol. 122, No. 4, 735–754

© 2015 American Psychological Association
0033-295X/15/$12.00 http://dx.doi.org/10.1037/a0039706

Titrating Decision Processes in the Mental Rotation Task

Alexander Provost
The University of Newcastle

Andrew Heathcote
The University of Newcastle and University of Tasmania

Shepard and Metzler's (1971) seminal mental-rotation task—which requires participants to decide if 1 object is a rotated version of another or its mirror image—has played a central role in the study of spatial cognition. We provide the first quantitative model of behavior in this task that is comprehensive in the sense of simultaneously providing an account of both error rates and the full distribution of response times. We used Brown and Heathcote's (2008) model of choice processing to separate out the contributions of mental rotation and decision stages. This model-based titration process was applied to data from a paradigm where converging evidence supported performance being based on rotation rather than other strategies. Stimuli were similar to Shepard and Metzler's block figures except a long major axis made rotation angle well defined for mirror stimuli, enabling comprehensive modeling of both mirror and normal responses. Results supported a mental rotation stage based on Larsen's (2014) model, where rotation takes a variable amount of time with a mean and variance that increase linearly with rotation angle. Differences in response threshold differences were largely responsible for mirror responses being slowed, and for errors increasing with rotation angle for some participants.

Keywords: mental rotation, LBA, cognitive model

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Spatial cognition affords the ability to make decisions based on mental manipulations of object representations. The mental rotation task introduced by Shepard and Metzler (1971) has had central place in the investigation of spatial cognition, and has been examined using approaches ranging from purely behavioral to evoked potentials (e.g., Hamm, Johnson, & Corballis, 2004; Provost, Johnson, Karayanidis, Brown, & Heathcote, 2013) and functional neuroimaging (see Zacks, 2008, for a review). Shepard and Metzler's task requires participants to make same versus different choices about pairs of two-dimensional projections of three-dimensional block figures, where one figure is a rotated version of the other that is either identical or a mirror image. Shepard and Metzler found that mean response time (reaction time; RT) increased linearly with rotation angle up to 180°, and suggested that participants performed a mental transformation analogous to perceiving a physical object rotating.

The linear RT effect applies to a wide variety of objects and to related paradigms, at least when the objects are sufficiently confusable and unfamiliar (Bethell-Fox & Shepard, 1988; Folk & Luce, 1987). In contrast, a curvilinear effect of rotation is found for mirror versus normal judgments about familiar alphanumeric characters (e.g., Cooper & Shepard, 1973; Corballis & McLaren, 1984). Regardless of linearity, mirror judgments in the pair-match and character tasks are often (e.g., Cooper & Shepard, 1973; Hamm et al., 2004; Provost et al., 2013), but not always (e.g., Jansen-Osmann & Heil, 2007; Larsen, 2014), slower than normal judgments by an amount that is the same at all rotation angles. Cooper and Shepard suggested this was due to the production of a normal response being more prepared than production of a mirror response. More recently Hamm et al. (2004) presented both behavioral and event-related potential evidence for picture-plane rotation ("spin") being followed by a faster rotation in depth ("flip") to normalize mirror-image character stimuli (see also Corballis & McMaster, 1996).

We develop and test a model of the cognitive processes involved in Shepard and Metzler's (1971) mental rotation task. In particular, we use an established cognitive model of choice RT, the linear ballistic accumulator (LBA; Brown & Heathcote, 2008), to separate out the processes involved in manipulating object representations from decision processes. Cooper and Shepard (1973) characterized the mental rotation task having four stages: (a) visual encoding,[1] (b) mental rotation, (c) response selection, and (d) response production. Rotation functions (i.e., plots of mean RT and sometimes error rates as a function of rotation angle) are commonly decomposed in terms of slope and intercept to identify effects associated with the rotation stage and the other stages, respectively (e.g., Wright, Thompson, Ganis, Newcombe, & Kosslyn, 2008).

By making the assumption that these stages are executed in series, and using the LBA to model the response-selection process, we are able to titrate out the effects of the decision stage, and so

This article was published Online First September 7, 2015.

Alexander Provost, School of Psychology, The University of Newcastle; Andrew Heathcote, School of Psychology, The University of Newcastle, and School of Medicine, University of Tasmania.

This work was supported by an Australian Research Council Professorial Fellowship. Thanks go to Blake Johnson and Scott Brown for advice.

Correspondence concerning this article should be addressed to Alexander Provost, School of Psychology, University Drive, Callaghan, New South Wales, Australia 2308. E-mail: alexander.provost@newcastle.edu.au

[1] When considering character stimuli a classification stage is sometimes added after the visual encoding stage (e.g., Corballis, 1988) based on evidence that rotation is required to differentiate normal and mirror alphanumeric characters but not to identify them.

provide a finer-grained decomposition of the mental rotation task than the regression approach. Our model is able to provide a more comprehensive account of behavior at the manifest (i.e., directly observed) level than any previous model of mental rotation, as it accounts for not only mean RT but also the full distribution of RT and the rates at which errors occur for both match and mismatch responses. Further, we were able to separately estimate the full distribution of the times required for the latent (i.e., not directly observed) mental rotation and decision processes.

In the first part of this article we provide a rationale for both our model's assumptions and the way we test it in the context of an examination of (a) brain and behavioral evidence about mental rotation, and of (b) previous models of mental rotation providing a quantitative account of mean RT (Kung & Hamm, 2010), errors (Kelley, Lee, & Wiley, 2000) or both (Larsen, 2014). In the second part of the article we describe our modeling framework in detail, propose a number of versions of both the rotation and decision stages within that framework, and evaluate the different versions against data collected in the final session of Provost et al.'s (2013) investigation of practice effects in the Shepard-Metzler pair-matching task. Provost et al.'s stimuli were designed so that, even though they were not familiar to participants prior to the experiment, rotation angle was well defined for both matching and mismatching stimuli. This characteristic greatly aids modeling at the distributional level, which otherwise requires extra assumptions about variability in rotation angles for mismatching stimuli, as we discuss in relation to Larsen's model.

Data from Provost et al.'s (2013) second experiment is also particularly useful as there is converging evidence—from event-related potentials and transfer testing on previously unseen stimuli—which strongly suggests that participants were using mental rotation rather than other strategies. This point is important because, as we discuss in relation to Kelley et al.'s (2000) and Kung and Hamm's (2010) models, it is likely that performance in what are nominally mental rotation tasks is often based on a mixture of strategies, including strategies that completely bypass any need for mental rotation.

The Process of Mental Rotation

Behavioral evidence from dual-task interference (Hyun & Luck, 2007) and evidence from electrophysiology (Prime & Jolicoeur, 2010) suggest that mental rotation involves transforming a representation held in visual short-term memory (STM). Neuroimaging studies indicate that visual motion areas in the posterior parietal cortex and superior posterior occipital cortex are activated during mental rotation tasks (Zacks, 2008), consistent with processing involving a continuous transformation performed on analog spatial representations of either whole objects or object parts (Just & Carpenter, 1985). Where the rotation involves imagining the transformation of effectors, such as hands or limbs, there is also evidence for activation in motor areas (Georgopoulos, Lurito, Petrides, Schwartz, & Massey, 1989; Thayer & Johnson, 2006; Zacks, 2008).

A close link with perceptual motion processing is indicated by Corballis and McLaren's (1982) finding that motion after effects interfere with mental rotation, with both the speed and preferred direction of mental rotation being modulated by motion adaptation (Heil, Bajric, Rösler, & Hennighausen, 1997). Shepard and Cooper (1982) summarize a range of evidence for a close link between perceived and imagined motion. For example, the linear effect of angle in mental rotation is consistent with Shepard and Judd's (1976) finding that the stimulus onset asynchrony associated with the breakdown in apparent rigid-rotation motion increases linearly with the angle between alternating images. Similarly, the existence of an upper limit on the rate of mental rotation is consistent with the existence of an upper bound on the angular velocity of apparent rotation (Farrell, Larsen, & Bundesen, 1982).

Commenting on mental transformations of both visual orientation and size, Larsen and Bundesen (2009) succinctly summarize the picture that emerges from behavioral and brain results: "Mental imagery seems to be a kind of simulated perception, not only psychologically but also neurophysiologically" (p. 533). They go on to state that both perceived and imagined rotation over large angles is made up of a sequence of smaller additive rotations. One of the rotation models we test, the *multiple-rotation model*, assumes that time for each smaller rotation has an independent exponential distribution, so overall rotation time has a gamma distribution. These assumptions, and the assumption that each of the additive rotations occurs at an equal expected rate, means that the gamma distribution's shape parameter increases linearly with rotation angle.

The second rotation model that we test, the *single-rotation model*, makes a very different assumption, that rotation occurs without any pauses and at a constant rate that varies according to a lognormal distribution. Ulrich and Miller (1993) discuss why the lognormal distribution (i.e., a random variable whose logarithm is normally distributed) provide a natural account of variability in process rates. These assumptions imply that rotation time also has a lognormal distribution with a scale parameter that increases linearly with rotation angle. Finally, we also consider a model that arises as a special case of the other two, the *deterministic-rotation model*, in which rotation-time variability is small enough relative to the variability in other stages that it can be neglected. In this case rotation time is a constant that increases linearly with rotation angle.

Strategies Used in Mental Rotation Tasks

Although mental rotation is one way that participants can obtain evidence on which to select a response, in some nominally mental-rotation tasks it is not the only way. This possibility is emphasized in the model of mirror versus normal decisions about letters proposed by Kung and Hamm (2010), and in the ACT-R based model of Kelley et al. (2000). Kelly et al. studied performance in a test developed by Shepard (1978), where participants have to choose which of three options is a rotated version of a fourth image. They focused on explaining accuracy, with the key findings being that the most common incorrect choice was a mirror-image rotation of the correct choice, especially for more complex images. This was modeled in ACT-R by assuming representations of mirror images were easily confused with correct representations, and so are more likely to be falsely retrieved, and that retrieval was more likely to fail altogether for more complex images. Verbal protocol analysis indicated that participants used a range of strategies such as shape and alignment comparisons, with mental rotation frequently reported as a strategy of last resort. This was captured in the ACT-R model by strategy selection being inversely

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

proportional to strategy difficulty as determined by the number of operations required to obtain a correct result using that strategy.

Kung and Hamm (2010) also postulated a mixture of rotation and nonrotation strategies, with an emphasis on modeling mean RT for correct responses, and in particular on explaining the curvilinear effect of angle found for normal versus mirror judgments about rotated alphabetic characters. Cooper and Shepard (1973) explained curvilinearity as occurring because rotation is faster for more familiar images, and that characters nearer to upright are more familiar. Kung and Hamm proposed instead—based on findings that characters at any angle can be identified without transformation (Corballis, 1988; Hamm & McMullen, 1998) – that participants could use letter identity to determine the polarity of a rotated character's horizontal axis and so determine without rotation if particular features were on the expected side. Curvilinearity reflects a mixture of this strategy, which takes a time that does not depend on rotation angle, and mental rotation, with an assumed linear effect on mean RT. The probability of the rotation strategy being used is assumed to increase linearly up to unity for an angle of 180°. Searle and Hamm's (2012) expansion of this proposal modeled individual differences in the degree of curvature through a power transformation of the linear change in the mixture probability.

Kung and Hamm (2010) also reported a correlation between the magnitude of slowing for mirror relative to normal mean RT and rotation rate. The correlation was interpreted as supporting Hamm et al.'s (2004) proposal that a "flip" is used to normalize mirror characters. They reported reasonable fits for many of their participants for an elaboration of their model that used a corresponding additive constant (i.e., an intercept) to model the difference between normal and mirror rotation functions. However, the additive assumption produced some misfit because their data contained a significant interaction, due to a shallower rotation function for the mirror condition. Searle and Hamm (2012) also found a trend for shallower mirror than normal slopes, but the interaction did not achieve significance ($p = .1$; Hamm, J. P. personal communication, January 13, 2015). Although Kung and Hamm cited one other study that found the same interaction to be significant (Duncombe, Bradshaw, Iansek, & Phillips, 1994) they commented that overall: "this interaction has rarely been replicated in the literature" (p. 212), and suggested that when it does occur might be because the spin and flip occur in cascade, rather than serially, and in a way that reduces the relative slowing for mirror images at larger angles.

Our focus in this article is modeling the mental rotation process and how it interacts with the response-selection process. In light of Kelley et al. (2000) and Kung and Hamm (2010) it is important that we test our model against data from a paradigm where participants use only a rotation strategy. The data collected in the final session of Provost et al.'s (2013) second experiment fulfills this criterion. Provost et al. investigated practice effects in Shepard and Metzler's (1971) pair-matching task, with stimuli that were two-dimensional projections of three-dimensional block figures with rotation in the picture plane. In this experiment participants practiced over five sessions with 320 pairs made up from 32 unique objects. The right pair member was either the same as, or a mirror image of, the left pair member, and rotated either 0°, 45°, 90°, 135° or 180°. The left pair member was always presented in the 0° orientation. In the final (sixth) session participants were tested on 320 trials using all 10 variants of 32 of the objects practiced in all previous sessions and 320 trials using transfer stimuli, made up of 10 variants of 32 objects that they had not previously seen.

Provost et al. (2013) found that both the intercept and slope of the mean-correct-RT rotation functions decreased markedly with practice, although the overall slope after practice remained substantial at 0.57 s per 180°, and it was almost identical for practiced (0.568 s) and transfer (0.576 s) stimuli. The latter finding, and almost identical intercepts (0.756 s and 0.746 s, respectively), suggests participants had learned a general purpose mental rotation skill rather than a strategy requiring specific knowledge about stimuli, such as that suggested by Kung and Hamm (2010). This conclusion was bolstered by event-related potential evidence. A component labeled the "rotation-related negativity" linearly modulated by rotation angle at the start of practice retained its linear modulation after practice, becoming if anything more strongly modulated and occurring earlier. In contrast, in Provost et al.'s first experiment, where participants practiced with a small set of stimuli, the modulation disappeared after practice, as did most of the behavioral effects of rotation angle, suggesting a different type of stimulus-specific skill acquisition. Hence, there is a strong case for asserting that data from the final session of the second experiment represents a fairly pure case of mental rotation.

To investigate the issue of match versus mismatch rotation functions we analyzed mean correct RT data from the final session of Provost et al.'s (2013) second experiment. In agreement with Hamm et al. (2004) and Kung and Hamm (2010), we found a positive correlation between rotation-function slope and the match versus mismatch difference ($r = .49$),[2] which approached significance despite there only being 11 participants, $t(9) = 1.7$, $p = .13$. Like Kung and Hamm, mismatch was significantly slower than match, $F(1, 10) = 44$, $p < .001$ (by 0.176 s, on average). We also replicated their finding of a significant interaction between angle and match versus mismatch factors, $F(4, 40) = 4.43$, $p = .02$, $\varepsilon = 0.57$, with a slightly larger slope for mismatch (180°/0.6 s) than match (180°/0.54 s).

Explaining Matching Versus Mismatching Differences

In light of these findings, in our model-based analysis of Provost et al.'s (2013) data we tested Hamm et al.'s (2004) "flip" hypothesis at the latent level. We did so by allowing for some extra rotation time for mismatch trials, where the extra time was assumed to be the same at all angles (i.e., an intercept effect). However, it is important to acknowledge two caveats to our investigation of the flip hypothesis. First, Provost et al.'s (2013) mirror stimuli—examples of which are shown in Figure 1—can only be normalized by flipping a part of the image (e.g., flipping the side of the long arm on which the shortest arm occurs in Figure 1b) rather than the whole image as is the case for characters. Second, there might also be less reason to flip the relatively unfamiliar objects used by Provost et al., as this is not required to make a correct response. A flip is not necessary for correct responding to alphanumeric characters either, but may occur auto-

[2] The correlation was based on averages over separate match and mismatch regressions. These analyses, and all of the other analyses reported in this article, collapsed over results for practiced and transfer stimuli; separate analyses produced almost identical results as this factor had little effect.

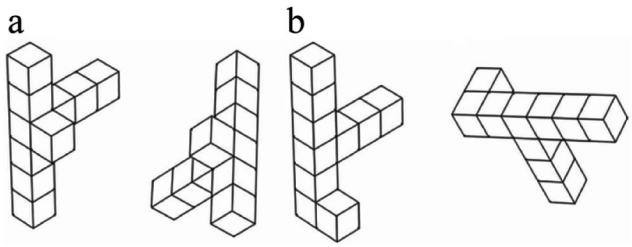

Figure 1. Examples of stimuli used in Provost el al.'s (2013) Experiment 2: (a) 180° matching or normal stimulus pair, and (b) right a 90° mismatching or mirror stimulus.

matically given the usual task with characters is to be sure about their identity rather than to determine whether they are in a mirror or normal form.

We contrasted the flip hypothesis with two other explanations of mirror versus normal differences, Cooper and Shepard's (1973) hypothesis about differences in response production, and a new account, response bias in the decision stage. Cooper and Shepard's original stage model (their Figure 11, p. 135) proposed that participants take approximately 0.1 s to prepare a match response before the stimulus appears. If, after the stimulus appears, they decide it is a match they produce the corresponding response, but if they decide it is a mismatch they must prepare the corresponding response. This preparation takes approximately 0.1 s, slowing mismatch-response production, and so their model predicts a constant slowing in mean RT at all angular displacements.

Our new proposal is that participants are more prepared to *select* a match response as well as, or instead of, being more prepared to actually *produce* a match response. In evidence accumulation models such as the LBA a choice is made when threshold amount of evidence favoring a particular response is obtained. Participants are thus more prepared to select a match response (i.e., they have a response bias) if the match accumulator has a lower threshold than the mismatch accumulator.

Although the response-selection and response-production accounts of mirror slowing do not predict Hamm et al.'s (2004) correlation, they are not necessarily incompatible with it. In particular, a correlation might arise from an alternative mechanism, proportional differences in individual speed, as is commonly assumed in aging research (e.g., Oberauer, 2005). Such proportional slowing would affect the size of slopes and intercepts in the same direction. Hamm et al. (2004) and Kung and Hamm (2010) tested for this possibility by correlating rotation rate with an index of general processing speed, mean RT in the upright normal-orientation condition. In neither case did they find a significant correlation, but Searle and Hamm (2012) did report a significant correlation ($r = .62$) when they used the average of mean RT for upright mirror and normal conditions as an index of general speed. However, this index could be contaminated by a flip for mirror upright, and when only upright normal is used as an index the correlation drops to 0.23 (J. P. Hamm, personal communication, January 13, 2015).

In Provost et al.'s (2013) data the correlation between rotation rate and mean RT in the upright normal-orientation condition was strong ($r = .82$, $p = .002$) and undiminished when the average of mean RT for upright mirror and normal conditions was used as an index of general speed ($r = .84$, $p = .001$). These results indicate that there may be some differences in the causes of mirror slowing in letter rotation and shape matching. We investigate this issue further in Provost et al.'s (2013) data, both by determining whether a model without the flip mechanism is able to produce a correlation between rotation time slopes and intercepts, and by testing all possible combinations of the flip, and response selection and production hypotheses.

Larsen's (2014) Model

The mental-rotation model proposed by Larsen (2014) is closest to the ones we develop here, as it is based on the idea of evidence accumulation and accounts simultaneously for accuracy and mean RT, although it does not address the distribution of RTs. It differs from our proposal, and that of Cooper and Shepard (1973), in that the mental rotation and response selection stages are interleaved rather than sequential. Participants are assumed to repeatedly rotate and decide for a match or mismatch, with each decision providing a unit of evidence that is accumulated until a threshold amount of evidence is obtained. Because the angular velocity of rotation is assumed to be constant, the rate of accumulation, and hence the expected time to achieve the threshold, increases linear with rotation angle for positive pairs.

Larsen (2014) fit his model to data from two experiments, one using a subset of five of Shepard and Metzler's (1971) stimuli and another using pairs of polygons where mismatching stimuli had a different shape rather than being mirror images. Participants were required to make decisions about both successively presented stimuli and simultaneously presented stimuli. In all cases a linear increase in mean RT with angle was obtained for matching stimuli, but mean RT for mismatching stimuli, and error rates for all stimuli, were largely unaffected by angle. Model fits assumed that, because there is no well-defined angular difference between mismatching stimuli, participants try a range of rotation angles based on the knowledge that matching pairs have rotation between 0° and 180°, resulting in an average 90° rotation for mismatching stimuli.

Provost et al.'s (2013) stimuli (see Figure 1) differed from those of Larsen (2014) in that they all have an obvious long main axis, with the main axis of the unrotated left pair member always presented upright. Hence, consistent with the linear increase in mismatching mean RT observed in their data, there is a well-defined rotation angle for all mismatching pairs. Because we fit the full RT distribution this is advantageous, as we cannot rely on the average-angle assumption used by Larsen to fit mean RT, but rather would have to specify a form for the distribution of angles.

Larsen's (2014) model also assumes that a probability mixture underpins reported noninteger estimates of evidence thresholds, which again would require of us extra specific assumptions about the mixture process. Nososky and colleagues faced an analogous dilemma with their model of STM scanning, which used a discrete random walk process with a few steps, similar to Larsen's (e.g., Nosofsky, Little, Donkin, & Fific, 2011). Like us they resolved this dilemma by moving to the LBA (Donkin & Nosofsky, 2012a, 2012b; Donkin, Nosofsky, Gold, & Shiffrin, 2013). In the LBA the starting point of evidence accumulation is assumed to vary over trials according to a uniform distribution, which is equivalent to having a uniform distribution of thresholds. The necessity of start-point variability receives independent verification, as it is

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

necessary for evidence accumulation models to account for the effect of speed versus accuracy instructions on the relative speed of correct and error responses. When instructions emphasize accuracy errors tend to be slower, whereas when they emphasize speed errors tend to be faster, which is accommodated by, respectively, a lesser or greater effect of start-point variability (Brown & Heathcote, 2008; Ratcliff & Rouder, 1998).

There are also potential challenges for modeling of Larsen's (2014) data in terms of a mixture of strategies. In his first experiment participants saw each of the five types of stimuli many times over the course of the experiment. This raises the possibility of stimulus-specific practice effects similar to those observed by Provost et al. (2013) in their first experiment. Further, changes induced by practice effects, which are at their most rapid early in practice (Heathcote, Brown, & Mewhort, 2000), add variability to performance, which potentially confounds fits of a model that assumes the data are stationary. We used only data from the last session of Provost et al. where it was clear that practice effects had reached asymptote and the data were indeed stationary. Second, the polygon stimuli in Larsen's second experiment share some similarities with Shepard's (1978) stimuli in that they may afford nonrotation strategies. Kelley et al. (2000) note such strategies may often be preferred to rotation, which would potentially confound a model assuming pure rotation. As previously stated, Provost et al.'s findings with respect to event-related potentials suggest that confounding from such a mixture is unlikely to be problematic for our analysis.

One point on which Larsen's (2014) model differs from many other evidence accumulation models is that it allows response criteria to change as a function of a stimulus property, rotation angle. Allowing thresholds to be affected by a stimulus property that is unknown to the participant until the trial begins is unusual in the evidence-accumulation modeling literature, where threshold change is viewed as a strategic process occurring on a slow time scale. Also, it would be circular to change the threshold contingent on a property that only becomes known as a result of evidence accumulation, or by more elaborate processing of that property.

However, there is precedent for allowing threshold to change based on a different and perhaps more salient aspect of the stimulus. For example, Bundesen's (1990) model allowed the threshold to vary with the number of items in a visual search display. As we assume that participants must at some level know the rotation angle before they begin accumulation, because they have to first rotate the image to upright, it is possible that they could also use this knowledge to change their threshold. In light of these considerations, we chose to compare models that both did and did not let the threshold be a function of angle. In light of the potential circularity, however, we did not let threshold vary between mirror and normal stimuli, although we did let it vary between mirror and normal accumulators in order to capture any potential response bias.

Larsen and Bundesen (2009) suggested that rotation-function slope increases for more difficult to discriminate complex stimuli (e.g., Folk & Luce, 1987) because participants increasingly tend to preform multiple rotations in an attempt to improve accuracy. Larsen (2014) suggested that any one rotation is error prone due to the limited capacity of visual STM. The idea of multiple rotations receives support as it helps to reconcile the fact that perceptually based estimates of the maximum angular rotation rate (e.g., Farrell et al., 1982) are much faster than rates observed in mental rotation experiments. Multiple error-prone rotations are compatible with our model if it is assumed they are all performed before evidence accumulation begins, with the net result being the encoding of a representation that subsequently provides a constant input to the evidence accumulation process.

Given this stipulation, we can accommodate the idea of multiple rotations, both in terms of effects on time and accuracy. For time, the additive basis of the gamma-distributed rotation time assumption is compatible with overall rotation time being the sum of times for a series of rotations. Accuracy effects can be represented in terms of the quality of the evidence provided by the final representation to the LBA evidence-accumulation process. The LBA makes decisions through a race between two units, one of which accumulates evidence for a "mirror" decision (i.e., that the right-hand stimulus is a mirror image of the left-hand stimulus after planar rotation) and one of which accumulates evidence for the alternative "normal" decision. Hence, evidence quality is naturally represented as modulating the size of the difference between accumulation rates for the correct unit (i.e., the normal accumulator for a matching stimulus or the mirror accumulator for a mismatching stimulus) and the incorrect unit (i.e., the mirror accumulator for a matching stimulus or the normal accumulator for a mismatching stimulus). In our model fits we allowed this difference to be a function of rotation angle in order to test whether the degree of rotation affected accuracy.

It is possible that the strength of the representation providing input to the LBA may also be affected by rotation angle. In the LBA strength of evidence is represented by the overall magnitude of rate parameters, and strength and quality can have largely selective effects on speed and accuracy, respectively. For example, increasing the rate of input to both the correct and incorrect accumulators equally would leave accuracy largely unaffected but would reduce the RT for all responses. Hence, we also allowed the overall magnitude of rate parameters to be a function of rotation angle, which is analogous to Larsen's (2014) assumption that the time to make a step in his discrete model increases with angle.

The combined assumptions of our model mean that there are three ways that an increased rotation angle can affect RT, through slowing the onset of evidence accumulation, through reducing the strength of the evidence driving accumulation or through changes in the evidence threshold. Model parameter estimates allow us to discern the relative influence of these three effects.

Stage Models

Perhaps the most striking difference between our proposal and that of Larsen (2014) is that we adopt Shepard and Cooper's (1973) assumption that mental rotation and response selection occur in sequential stages. Stage models have a long history in cognitive psychology (e.g., Sternberg, 1969), but overlapping or cascaded processing (e.g., McClelland, 1979) has also been proposed. In the present context, this suggests that mental rotation may begin before the visual encoding has finished. However, Ruthruff and Miller (1995) present evidence that any rotation occurring during the overlap is small due to interference between encoding and rotation processes. There is also evidence that response preparation can begin before rotation finishes (Band & Miller, 1997; Heil, Rauch, & Hennighausen, 1998), but again

interference may limit the effect of the overlap (see also Ruthruff, Miller, & Lachmann, 1995).

Shepard and Cooper (1973) confirmed that participants are able to complete mental rotation before making a response selection in a matching paradigm where they were cued with the identity and orientation of the upcoming letter stimulus. They found that the effect of rotation angle diminished and eventually disappeared as the cue-to-stimulus interval increased. Their interpretation was that participants were rotating a mental representation of the letter before the rotated letter stimulus appeared. When the stimulus appeared they directly matched it to the already appropriately rotated mental image in order to make a normal versus mirror-image judgment. Similarly Metzler and Shepard (1974) showed that participants could also engage the same constant-rate rotation process when they did not know the required angle ahead of time.

Of course, this does not mean that participants would use similarly sequential processing in an uncued mental rotation task. However, in our paradigm, like Shepard and Cooper's (1973), the required rotation angle is well defined without reference to any matching operation, so can be carried out in advance of the matching decision. Hence, the idea of completing rotation before response selection is at least plausible. In contrast, in the paradigms studied by Larsen (2014) multiple rotation and matching decisions may be necessary in order to determine when rotation should be terminated.

In light these considerations we also tested an alternative to the multiple-rotation model. This alternative *single-rotation* conceptualization of the rotation stage assumes that the time to complete rotation (t_{mr}), is given by the required rotation angle (A), divided by the rotation rate (r): $t_{mr} = A/r$. As in the multiple-rotation model, the rate of rotation it is assumed to be random variable that is independent of the rotation angle (i.e., a constant-rate assumption), and hence the distribution of rotation time is the inverse of the distribution of rotation rate. Ulrich and Miller (1993) pointed out that, under quite general conditions related to the central-limit theorem,[3] a lognormal rate distribution arises in a cascade process whose overall rate is the product of the rates of constituent processes. Hence, on the assumption that a cascaded process underpins the process of mental rotation, the distribution of the rotation rates will approximately follow a lognormal distribution.

Model Specification

Figure 2 summarizes our modeling framework. A single parameter, t_{er}, accounts for the sum of the times to encode the stimulus (t_e) and to produce a response (t_r). These times are assumed to have a much smaller variance than that associated with the other two stages, so t_{er} is well approximated by a constant value. We also assume t_{er} is unaffected by the rotation angle factor (A) or stimulus factor (S: normal vs. mirror). However, we did allow it to differ between accumulators in order to test Shepard and Cooper's (1973) response-production explanation of slower mirror than normal responses (i.e., $t_{r|mirror} > t_{r|normal}$, which implies $t_{er|mirror} > t_{er|normal}$, given $t_{e|mirror} = t_{e|normal}$). Overall RT is assumed to be the sum of t_{er} and the times to complete the mental-rotation stage (t_{mr}) and decision time (t_d), which is the time required by the LBA to select a response.

For the multiple-rotation model we chose the gamma distribution to describe rotation time as it is strictly positive, and so

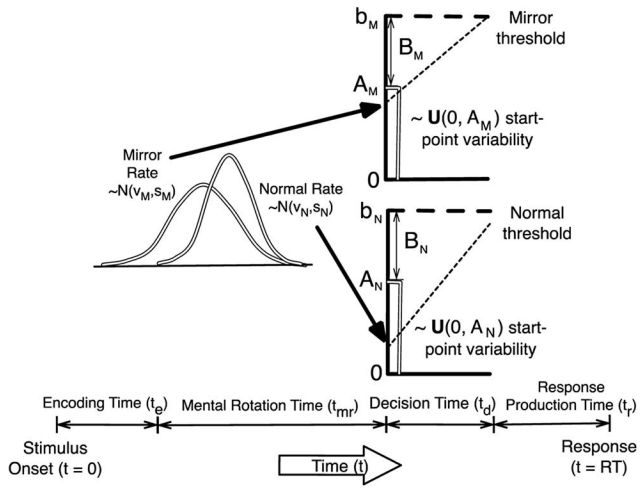

Figure 2. Model framework.

provides one suitable path—given time cannot be negative—to the limiting case of a normal distribution implied by the additive structure identified by Larsen and Bundesen (2009). That is, in the limit of adding many small rotation steps that take variable times, the total rotation time will be well approximated by a distribution with a symmetric normal shape. The shape of the gamma distribution is determined by its shape parameter ($k > 0$), where integer values correspond to the number of exponential distributions that are added together. Larger values of k (corresponding to a sum with many exponential terms) result in an increasingly good approximation to the normal distribution, whereas smaller values of k (corresponding to a sum with fewer terms) result in distributions with positive skew (i.e., a longer right than left tail).

The probability density of the gamma distribution at time t is given by $(e^{-\frac{t}{\theta}}t^{k-1})/(\Gamma(k)\theta^k)$. The exponential distribution is a special case of the gamma distribution when $k = 1$. $\Gamma(x)$ is the gamma function, a continuous generalization of a factorial (i.e., $\Gamma(x) = (x-1)!$). The gamma distribution, and hence rotation time, has mean $k\theta$, variance $k\theta^2$ and skewness $2/\sqrt{k}$. These expressions demonstrate that the mean and variance of rotation time increases linearly with angle, whereas skewness is a decreasing function of angle. Note that the linear increase in variance and the decrease in skew are general to multiple-rotation models that make a wide variety of assumptions about the forms of the distributions of times for each constituent rotation.

For the single-rotation model we assume that mental rotation occurs at a constant rate r on every trial, where r is a sample from a lognormal (hence always positive) distribution $R \sim LN(\mu,\sigma) = e^{N(\mu,\sigma)}$, where $N(\mu, \sigma)$ is a normal random variable with mean μ

[3] In a linear cascade the overall rate $R = \prod_i R_i$, where each constituent-stage has rate, R_i, which is a random variable with unspecified form except that it has a defined second moment. Given $\ln(R) = ln(\sum_i R_i)$, and that by the central-limit theorem the sum on the right hand side tends to a normal distribution, then R tends to a lognormal form. Simulations reported by Ulrich and Miller (1993) indicated that for a range of plausible positive rate distributions the lognormal approximation was quite accurate for as few as three or four stages.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

and standard deviation σ. For the lognormal the logarithm of the μ parameter determines the log-scale of the distribution (note there is no bound on μ) and σ>0 its shape. The random variable for the time for a rotation through angle A, is the angle divided by the rotation rate, so t_{mr} also has a lognormal distribution:

$$t_{mr} \sim A/e^{N(i,\acute{o})} = e^{lnA}e^{-N(\mu,\sigma)} = e^{lnA-N(\mu,\sigma)}$$

$$= e^{N(lnA-\mu,\sigma)} = LN(lnA-\mu,\sigma)$$

Note that *A* acts as a scaling parameter, so the mean, standard deviation and all quantiles (e.g., the median) of rotation time increase linearly with rotation angle. This also implies that, in contrast to the multiple-rotation model, skewness is unaffected by rotation angle. Note that these properties associated with scaling by angle apply to any single-rotation model no matter what form is assumed for the rate distribution. The single-rotation model is similar to the class of models generally used to explain the pervasive finding in RT distributions that standard deviation increases linearly with the mean (Wagenmakers & Brown, 2007). In the multiple-rotation model, in contrast, it is the variance that increases linearly with the mean.

To test Hamm et al.'s (2004) "flip" hypothesis about slower mirror responses we also fit versions of the rotation models in which the intercept for the linear effect of angle could be greater than zero for mirror stimuli, reflecting the time taken to flip a mirror stimulus to normal orientation. For all rotation models we assumed a zero rotation time for normal stimuli with an angle of zero, corresponding to the assumption that no rotation is required in this case. For all rotation models we also assume that when the rotation stage is finished it produces a representation that provides the input to the subsequent decision stage.

The input to the decision stage causes the summed evidence values in the accumulators, represented by dashed lines in Figure 2, to change with a rate (i.e., the slopes of the dashed lines) that varies randomly from trial to trial according to normal distribution, N(v, s). Figure 2 illustrates typical rate distributions associated with a normal stimulus, where the mean rate for the normal accumulator is greater than that for the mirror (i.e., $v_N > v_M$). It also illustrates a particular trial where the rates sampled from these distributions have the same ordering as the means (i.e., the dashed line for the normal accumulator is steeper than that for the mirror accumulator).

The modeling framework assumes that the mean rates for each accumulator could differ as a function of the angle (A) and stimulus (S), and their interaction. Drift rates are allowed this considerable flexibility on the assumption that the mental-rotation process can affect the information on which the mirror versus normal discrimination is based (Larsen, 2014). We tested this possibility by comparing models in which rate parameters could vary as a function of the angle and stimulus against more restrictive models that allowed for only angle effects, only stimulus effects, or neither effect.

Note that the rate distributions in Figure 2 overlap, so that for some trials the sampled rates may favor the incorrect accumulator (e.g., the mirror accumulator for a normal stimulus). The degree of overlap is modulated by the standard deviations (s) of the rate distributions. We assume that for the incorrect accumulator the standard deviation is fixed at a value of one, which makes the model identifiable (Donkin, Brown, & Heathcote, 2009). For a normal stimulus (as in Figure 2) this means $s_M = 1$, whereas for a mirror stimulus it means that $s_N = 1$. Figure 2 illustrates the pattern of standard deviations that occurs in many applications of the LBA (e.g., Heathcote, Loft, & Remington, 2015; Heathcote & Love, 2012; Rae, Heathcote, Donkin, Averell, & Brown, 2014); the correct accumulator has a smaller standard deviation than the incorrect accumulator. We freely estimated the value of the correct accumulator standard deviation, but assumed that value was the same for all conditions.

Even if on a particular trial an accumulator has a higher rate it may not necessarily win. An accumulator wins (and so the corresponding response is selected) when its evidence total reaches its threshold, b, before the other accumulator. This can occur when, as illustrated in Figure 2, the accumulator with a slower rate begins at a higher level. The initial level for each accumulator varies independently from trial to trial according to a uniform distribution over the range 0–A ($A \leq b$). As the distance between the top of the start-point distribution and the threshold ($B = b - A$) increases it becomes increasingly likely that the rate effect will dominate. Hence, changes in B cause a speed–accuracy trade-off, with higher values slowing decision time (t_d) but increasing accuracy.

Parameters associated with the accumulators are assumed to change slowly, contingent on factors such as the outcomes of previous decisions, so we estimated the same values of A for all conditions, which differ only in the nature of the stimuli. However, as previously discussed, we did allow B to change with rotation angle, and also allowed for the possibility that the normal and mirror accumulators have different thresholds. A lower threshold for the normal than mirror accumulator is predicted by our hypothesis that participants are more prepared to select normal responses than mirror responses.

In the first part of the next section we evaluate the different models without a flip in the rotation stage. These models still have considerable flexibility to explaining mirror slowing, either in terms of threshold, mean rate, or response production time (t_{er}) differences, or a combination of one or more of these. Our initial analysis focuses on comparing the different rotation models. We then investigate whether there are also rotation angle effects on the decision stage—mediated either by the mean rate of evidence accumulation and/or evidence thresholds—and whether mirror slowing is caused by threshold, mean rate, or response production time (t_{er}) differences, or a combination of two or more of these. Subsequently we examine the flip hypothesis, with our model-based analysis being augmented by a reanalysis of the event-related potential data collected by Provost et al. (2013).

Overall Model Evaluation

Our focus in this initial analysis is on discriminating between the multiple-rotation and single-rotation models of the rotation stage. We also consider a third *deterministic-rotation model*, which assumes there is negligible variability in the rotation stage. This possibility is compatible with Larsen and Bundesen's (2009) additive-step proposal about the time for a single-rotation process, and with Larsen's (2014) assumption that total rotation time is the sum of repeated rotation steps when it is assumed—in either case—that step times have little variability. It is also compatible with the single-rotation model when it is assumed that rotation rates do not vary substantially from trial to trial. Note that the

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

deterministic-rotation model is nested within (i.e., is a special case of) the two variable-rotation-time models in the limit of their scale and shape parameters (respectively) becoming zero.

In total the decision stage of all three models has 34 parameters. For each participant there is one A and one sv parameter, two t_{er} parameters, 10 B parameters and 20 v parameters. For B there are 5 parameters for the mirror and 5 for the normal accumulator. For v there are 10 parameters for each of the two accumulators, made up of a factorial combination of the 5-level angle factor and the two-level stimulus factor (normal vs. mirror). All three models also have a single slope parameter. This determines the rate at which the gamma shape parameter increases with angle for the multiple-rotation model and the rate at which the lognormal scale parameter increases for single-rotation model, and the rate at which rotation time increases for the deterministic-rotation model. Overall the deterministic model has 35 parameters. The other two rotation models both have one extra parameter, a scale parameter for the multiple-rotation model (θ) and a shape parameter (σ) for the single-rotation model. These parameters are the same across all rotation angles.

All models were fit to each subject's data separately, using the maximum-likelihood methods described in Donkin, Brown, and Heathcote (2011). Maximum-likelihood estimation minimizes the misfit between the model and data as measured by the deviance, which equals minus two times the maximized log-likelihood of the data. Table 1 reports deviance summed over participants (D). Smaller values of deviance indicate a better fit, but the absolute value of the deviance is not interpretable, as it has no natural zero point and its magnitude depends on the data's measurement units.

Where models are nested (i.e., a simpler model is a special case of a more complex model obtained by fixing some of its parameters) the difference between their deviance values is approximately distributed as $\chi^2(p_1 - p_0)$, where p_1 and p_0 are the number of parameters in the more complex and simpler models. Hence, it can be used to test whether the simpler model's fit is significantly worse than that of the more complex model. As the deterministic model is nested within the other two models it can be compared in this way. At the group level, it has a significantly worse fit than both the multiple-rotation model, $\chi^2(11) = 191$, $p < .001$, and the single rotation model, $\chi^2(11) = 144$, $p < .001$. This effect is highly consistent at the individual subject level, with significantly worse fit for 10 of the 11 participants. These results strongly support variability in the time to complete the rotation stage.

Although the two variable rotation-time models are not nested, they can be compared using the Akaike information criteria (AIC, Burnham & Anderson, 2004),[4] where the preferred model has the smaller AIC value. Unlike the deviance, AIC can be smaller for a less complex model (i.e., a model with fewer parameters), because it adds to the deviance a penalty term proportional to the number of model parameters (p). We use the small-sample corrected form, AICc, with the penalty term $2p \times (1 + (p + 1)/(N - p - 1))$, where N is the number of observations used to calculate the deviance. As N grows AICc approaches a maximum value of $2p$, so an extra parameter must decrease the deviance by 2 if its addition is to be supported.

AICc can be used to construct Akaike weights (w, Wagenmakers & Farrell, 2004), which can be interpreted as indicating the probability that one of a set of models is the best model. In the special case of comparing two, the probability that the model with a smaller AICc is best is $w = e^{\Delta/2}/(1 + e^{\Delta/2})$, where Δ is the amount by which it is smaller. The value of w provides a continuous and easily interpretable measure of the strength of evidence in favor of a model. For example, for $\Delta = 3$, $w \approx .8$, $\Delta = 6$, $w \approx .95$ or if $\Delta = 10$, $w \approx .99$. On a group level AICc provides very strong evidence favoring the multiple-rotation model over the single rotation model, $\Delta = 47$. At the individual level the results were more equivocal, with fairly strong evidence ($\Delta > 6$) favoring the multiple-rotation model for six participants and favoring the single rotation model for two participants, with the remaining three more equivocal.

In order to investigate whether it was the lognormal distributional assumption that caused the greater misfit we also fit single-rotation model which assumed the same gamma distribution for rotation time as the multiple-rotation model.[5] However, the results for this model were essentially the same as for the lognormal single-rotation model, but with a slightly worse fit (D = −3185).

In further analyses that look at the relative importance of different mechanisms in the decision stage we consider results for both the gamma multiple-rotation and lognormal single-rotation models. We do so in order to assess the sensitivity of our findings to assumptions about the rotation stage, but in light of its better fit, we emphasize results for the multiple-rotation model. Before proceeding with these comparative analyses, however, it is important to establish that the variable rotation models are able to provide a good descriptive account of the data in an absolute sense. We approach this question graphically, by looking at model fits to RT distribution (see Figure 3) and accuracy (see Figure 4) data for

Table 1
Model Evaluation for the Standard LBA Decision-Model Parameterization Combined With Four Rotation models

Rotation Model	p	Deviance	AICc
Gamma multiple-rotation	36	**−3237**	**−2396**
Lognormal single-rotation	36	−3190	−2349
Deterministic rotation	35	−3046	−2254

Note. Number of parameters per participant (p), sum of misfit measures over participants (deviance), the corrected group Akaike information criteria (AICc). The best (smallest) fit statistics are shown in bold. LBA = linear ballistic accumulator.

[4] We also examined the commonly used alternative Bayesian Information Criterion, BIC = D + p ln(N), where N is the number of data points. Because the BIC complexity penalty is typically much larger it prefers simpler models than AICc. In a Bayesian framework, assuming a unit information prior, BIC has asymptotically consistent model selection, but in nonasymptotic samples it tends to only allow for the large effects, ignoring smaller but still reliable differences. We found this to be the case in our data, with BIC selected models clearly providing a poor fit to the data. Hence, we report only results based on AICc, but note that BIC agreed on many major issues, such as rejection of the deterministic and no-rotation models and angle effects on mean rates.

[5] Like the lognormal single-rotation model, this model assumes the scale parameter, and hence the rotation-time standard deviation, increases linearly with rotation angle. This model assumes a rate of rotation that has an inverse-gamma distribution—a distribution used in Bayesian statistics that has a positively skewed shape that is quite similar to the gamma, and which is also governed shape and scale parameters. Note that we did not test a lognormal multiple-rotation model as we are not aware of corresponding psychological justification.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

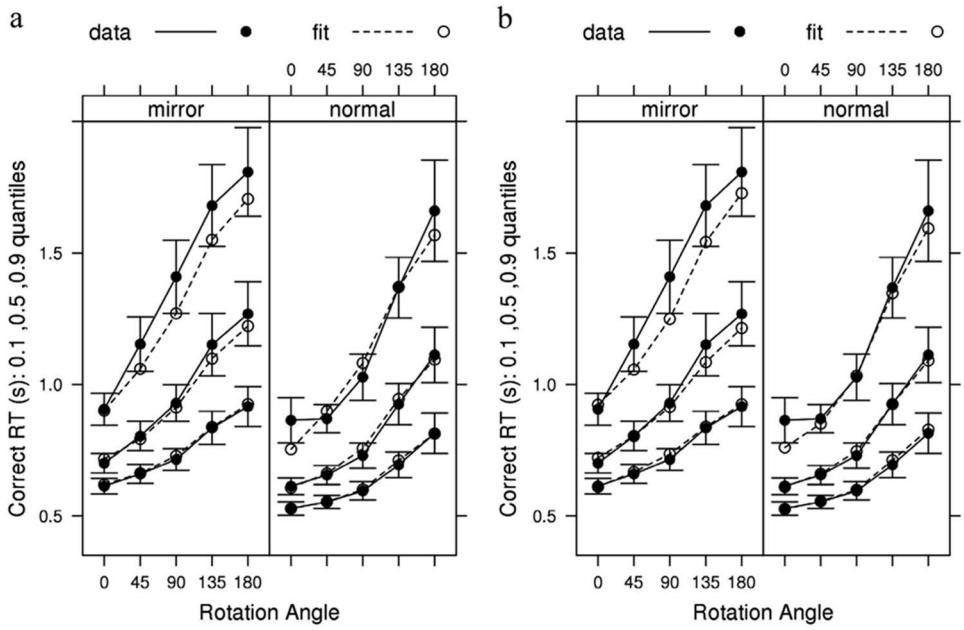

Figure 3. Observed reaction time (RT) quantiles for correct responses averaged over practiced and transfer items and participants with standard errors, and fits of (a) the gamma multiple-rotation model and (b) the lognormal single-rotation model.

both the gamma multiple-rotation model and the lognormal single-rotation model.

Figure 3 illustrates RT distribution for correct responses by plotting the average over participants in each condition of their median (i.e., 50th percentile, indicating the central tendency of the data), 10th percentile (indicating the speed of the fastest responses), and 90th percentile (indicating the speed of the slowest responses). The positive skew that is typical of RT distribution is evidenced by the greater gap between the median and 90th percentile than between the median and 10th percentile. Figure 3—and other figures showing model fits—display the data using solid lines to join solid symbols with standard error bars. The average model fits are shown as open symbols joined by dashed lines. It shows that both multiple and single rotation models provide a quite accurate and very similar account of RT distribution. In particular, both capture the shape of the angle effects, with a linear increase for mirror stimuli but a more curvilinear, upwardly accelerated increase for normal stimuli.

Figure 4 shows that both models also provide a fairly accurate and very similar account of RT distribution errors, although the

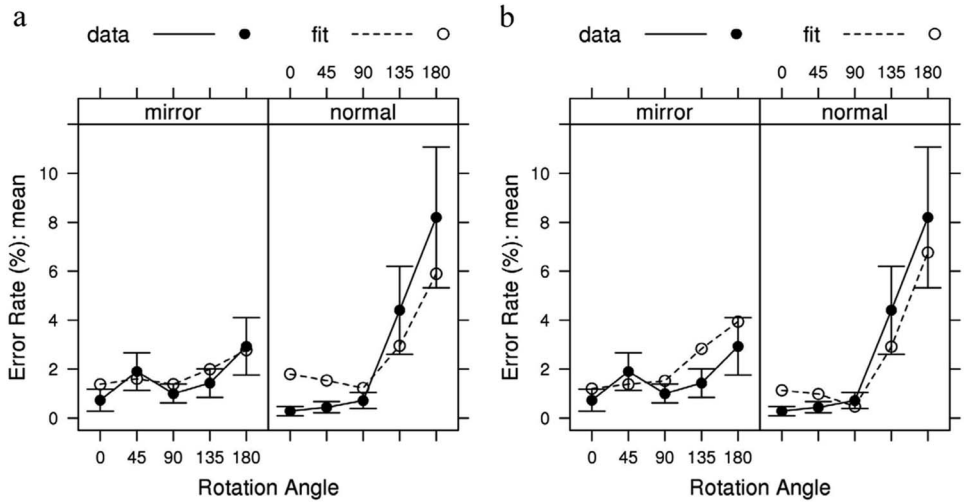

Figure 4. Observed error rates averaged over practiced and transfer items and participants with standard errors, and fits of (a) the top gamma multiple-rotation model and (b) the lognormal single-rotation model.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

multiple rotation model displays a slightly greater overestimate of normal error rates for small rotation angles and the single rotation model a slightly greater overestimate of mirror error rates for larger rotation angles. However, both models capture the general increase in errors with angle, as well as the steeper increase for normal than mirror stimuli. Although both the number of participants and error rates are fairly low, there is a significant[6] angle main effect, $F(4, 40) = 6.87$, $\varepsilon = .3$, $p = .02$, and the angle by stimulus interaction just achieves significance, $F(4, 40) = 4.22$, $\varepsilon = .33$, $p = .05$. Hence, it seems reasonable to require the model to capture these properties.

Parsimonious Models

The models examined in Table 1 are quite complex, and so run the risk of overfitting the data. That is, they may have excess flexibility, afforded by their large number of parameters, which allows them to accommodate random as well as systematic structure in the data, reducing their ability to accurately predict new data (Myung & Pitt, 1997). Overfitting might also allow wrong models to provide an apparently adequate description of the data. Further, tradeoffs between parameters can result in estimates that are not meaningful, so that parameter values do not provide a valid insight into the underlying psychological mechanisms. Hence, it is desirable to identify parsimonious models that strike a balance between fit and simplicity.

Donkin et al.'s (2011) fitting method requires fitting of all possible simplified variants of the most complex or "top" model (i.e., the models in Table 1). The variants are obtained by removing the effects of every possible combination of factors in the top model. This helps ensure that fits of complex top models are not suboptimal, by initializing fits of more complex models from the best fits of all models that are less complex by one factor, starting from the simplest model and working upward to the top model. Although this approach is computationally costly ($2^7 = 128$ variants of each type of model and participant), it has the additional advantage that AICc can then be used to select the best variant.

In the present context, selection among the model variants can answer three important questions. The first is, for their most parsimonious version (as determined by AICc), which type of rotation model is best? This question is important in case the top models of each type differ in their ability to overfit the data. The second question is, which factors have the strongest case for influencing each parameter (i.e., are retained in the selected model) and which the least (i.e., are dropped from the selected model)? The answer to this question provides insight into whether rotation angle affects rates, thresholds, or both, and whether there is a bias in the threshold and a difference between target and nontarget rates. The third question is, does the pattern of factor influence differ across models? Consistency indicates the answer to the second question is insensitive to detailed differences in assumptions among models.

We also added a new and very parsimonious model, which we call the no-rotation-stage model,[7] which corresponds to the model proposed by Larsen (2014). As the name implies, it assumes there is no rotation stage, but otherwise makes the same assumptions about the decision stage as previous models with one exception; that the mean rate of evidence accumulation is linearly proportional to the reciprocal of the rotation angle. The latter assumption arises from a multiple-rotation model if each rotation adds the same amount of evidence and takes a time linearly proportional to rotation angle. The no-rotation-stage model has only 22 parameters, 8 mean rate parameters, along with the same 10 threshold, two t_{er} and one s_v and one A parameter.

Model Selection

Table 2 shows the variants that were selected as best by AICc. Although the number of parameters estimated almost halves there is very little decrease in fit for either the multiple or single rotation models, with $\chi^2(176) = 82$, $p > .999$, in both cases. At the individual level no decreases in fit were significant at the 0.05 level. Only one factor was dropped from the no-rotation model, the effect of stimulus type on nondecision time, which produced a relatively small decrease in fit that only just achieved significance, $\chi^2(11) = 20$, $p = .045$, with only one of 11 decreases significant at the individual level.

The no-rotation-stage model was clearly out-performed relative to the multiple and single rotation models, with $\Delta = 93$ and $\Delta = 46$, respectively. Relative to the no-rotation-stage model, six participants had $\Delta > 6$ in favor of the multiple-rotation model and seven in favor of the single-rotation model. Only one participant had $\Delta > 6$ in favor of the no-rotation-stage model relative to the single-rotation model, with no cases of $\Delta > 6$ favoring the no-rotation-stage model relative to the multiple-rotation model.

Comparing variable rotation-time models, the multiple-rotation model was best, with the same $\Delta = 47$ as the comparison between the top models. At the individual level the results were also similar, again with five participants having $\Delta > 6$ in favor of the multiple-rotation model, two in favor of the single rotation model at the same level and the remaining three more equivocal. Note that (in results not shown) the deterministic-rotation model continues to do badly with $\Delta = 254$ relative to the multiple-rotation model. Hence, the answer to our first question based on the most parsimonious models appears to be the same as that based on the top models.

Table 2 also shows that the variable-rotation-time models both pick the same model variant, where rotation angle only affects the threshold and has no effect on mean rate. We examine these parameter effects in more detail below, but first review evidence for marked individual differences in the patterns of behavior they explain.

Individual Differences

Although the forgoing results suggest that the AICc selected models provide a parsimonious and coherent explanation of average performance, a question remains as to how they account for individual differences. Figures 5 and 6 plot the data and fits of the

[6] Where necessary we report Greenhouse–Geisser ε corrections and correspondingly corrected p values.

[7] We also fit a more general version of this model, where the mean rate free to vary arbitrarily with angle. This model but hardly fit any better ($D = -3107$) but had many more parameters (34) and so a much worse AICc (-2314). The most parsimonious variant of this model, which dropped the stimulus type effect on mean rate and so had 24 parameters, was also worse than the most parsimonious variant of the no-rotation-stage model (AICc $= -2524$).

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Table 2
Model Variants That Had the Lowest Group Akaike Information Criteria (AICc)

Best AICc Model Variant	*B*	*v*	t_{er}	*p*	Deviance	AICc
Gamma multiple-rotation	A, R	S, M	S	20	**−3155**	**−2700**
Lognormal single-rotation	A, R	S, M	S	20	−3108	−2653
No rotation stage	A, R	A, S, M		21	−3086	−2607

Note. "A" and "S" indicate the rotation angle and stimulus (mirror vs. normal) factors, respectively. "R" is a response-accumulator factor allowing the threshold (*B*) to differ between the mirror and normal accumulators. "M" is a "match" factor allowing the mean rate (*v*) to vary between the accumulator that matches the stimulus (e.g., the normal accumulator for a normal stimulus) and the accumulator that does not (e.g., the mirror accumulator for a normal stimulus). The best (i.e., smallest) fit statistics are shown in bold.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

AICc-selected multiple-rotation model for each participant (very similar results held for the single-rotation model). These graphs reveal marked individual differences, which are accounted for fairly well by the parsimonious AICc-selected model.

Figure 5, which plots correct RT quantiles, shows that rotation functions vary from fairly flat for Participant 238 to steeply increasing for Participant 262. The model captures important features such as the fact that even the fastest responses, represented by the 10th percentile, can increase markedly with angle, and that the spread of the quantiles also tends to systematically increase with angle. Figure 5 also shows some idiosyncratic features, such as the marked slowing in Participant 263 of the 90th percentile for unrotated normal stimuli, which appears to be entirely responsible for the corresponding misfit in Figure 3. Similarly, the misfit to the 90th percentile for mirror stimuli at middle angles evident in Figure 3 appears entirely due to Participants 001 and 262.

Figure 6 shows that there are even stronger individual differences in accuracy. These differences are, in the main, well accounted for by the AICc-selected model, although it shows an increased version of the pattern evident for normal stimuli in Figure 4 of overestimating errors for small angles and underestimation for larger angles. Participants can be divided into two approximately equal groups on the basis of accuracy, one of which displays virtually no effects on errors (237, 238, 246, 255, and 262), whereas the remainder display effects that can be quite marked, particularly for normal stimuli at 135° and 180°. This division is not reflected in speed, with participants from both the accurate (e.g., 246, 262) and error-prone (e.g., 001, 256) groups displaying strong angle effects on RT, supporting the case for different underlying causes for angle effects on errors (decision stage) and RT (rotation stage).

Parameter Effects

As illustrated in Figures 3 and 4, median RT for mirror stimuli is significantly slower than for normal stimulus (0.97 s vs. 0.807 s, respectively), $F(1, 10) = 33.7$, $p < .001$. There is a significant increase with angle for median RT, $F(4, 40) = 35.6$, $\varepsilon = .3$, $p < .001$, and errors, $F(4, 40) = 6.8$, $\varepsilon = .3$, $p = .02$, with a quicker increase with angle for normal than mirror stimuli, $F(4, 40) = 4.22$, $\varepsilon = .333$, $p = .05$. Effects on thresholds and mean rates in the AICc-selected mediate both differences. The difference in RT between normal and mirror stimuli is also mediated by nondecision time, but in the multiple-rotation model this effect, although significant, $F(1, 10) = 5.77$, $p = .037$, is minor (0.324 s vs. 0.294 s, respectively), explaining less than one fifth of the effect on the median. Nondecision time results were similar for the single-rotation model, but with a slightly smaller difference (0.016 s), which did not achieve significance, $F(1, 10) = 1.89$, $p = .2$.

For the AICc-selected multiple-rotation model, rates varied with stimulus type, but had little effect on mirror slowing, as for correct responses they were similar for mirror (1.8) and normal (1.9) stimuli. However, there was some evidence of a contribution to the error difference, with a marginally significantly larger difference between the rates for matching and mismatching accumulators for mirror (3.2) than normal (2.5) stimuli, $F(1, 10) = 3.33$, $p = .098$. The same pattern of significance held for the single-rotation model, but thresholds were a little higher (0.92 vs. 0.71) and matching rates (2.2 vs. 2.4) and rate differences (4.2 vs. 3.2) a little larger.

For the AICc-selected multiple-rotation model threshold estimates revealed a significant bias favoring normal responding, due to a larger threshold for the mirror (0.83) than normal (0.63) accumulator, $F(1, 10) = 5.15$, $p = .047$. Thresholds also increased significantly and in a positively accelerating manner with angle, $F(2, 40) = 7.36$, $\varepsilon = .29$, $p = .026$. There was a trend for a faster acceleration for mirror thresholds, but this did not achieve significance, $F(2, 40) = 1.33$, $\varepsilon = .33$, $p = .28$. However, this trend for faster acceleration was particularly marked in three participants whose responses to normal 180° stimuli was particularly error prone (001, 256 and 263). For these participants the normal bias was almost absent at 180° (5% relative to the average threshold) whereas it was very marked for lesser angles (48%). This change explains the ability of the model to account for their increased error rates for normal 180° stimuli (see Figure 6). The same pattern held for the AICc-selected single-rotation model's threshold estimates (e.g., 58% vs. 10% bias).

In order to illustrate the role played by the rotation-angle effects on the threshold, Figure 7 plots the fit of a multiple-rotation model that is the same as the AICc-selected model except with no angle effects on thresholds. This model is clearly unable to accommodate the upwardly accelerated effect of angle on normal RT, showing only the linear increase due to the rotation stage. It is also unable to accommodate the increase in errors with angle and its interaction with stimulus type. Again the same results are observed for the single-rotation model. Overall, these results suggest that both the AICc-selected models identify a response bias caused by threshold differences as the main cause of mirror responses being slower and more error prone. Changes in thresholds and response bias also explain the positively accelerating increase in RT and error rates with rotation angle for normal stimuli.

Rotation Models

Figure 8 illustrates the rotation-time distributions corresponding to the parameters from AICc-selected models averaged over participants. Figure 8a illustrates the decrease in rotation-time distribution skew as rotation-angle increases for the gamma multiple-rotation model, from more skewed than an exponential distribution for 90° or less to less skewed for greater rotation angles, with the shape parameter incrementing by 0.41 per 45°, and the scale parameter fixed at 0.256 s for all conditions. For the lognormal single-rotation model shown in Figure 8b, in contrast skew is

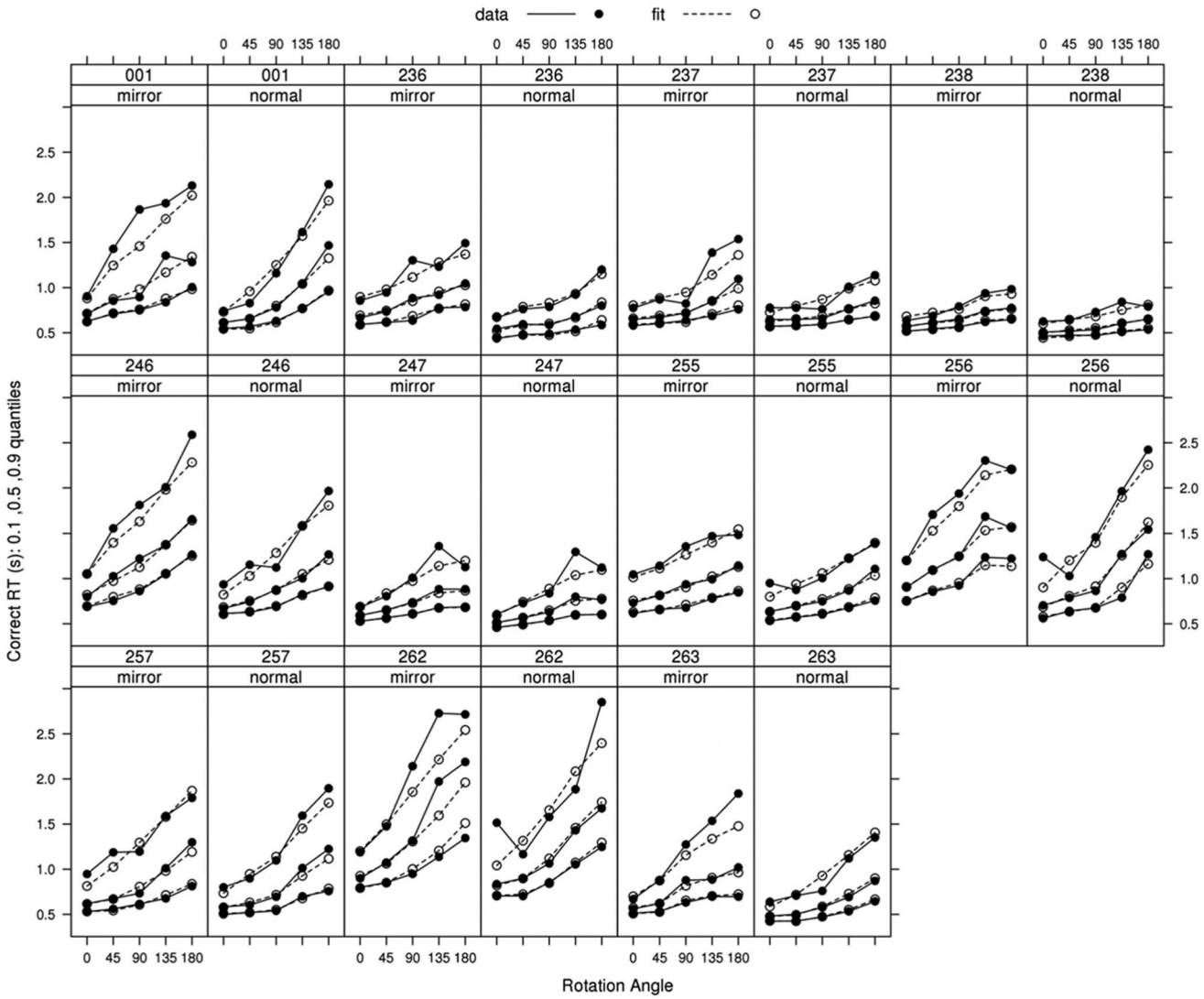

Figure 5. Observed reaction time (RT) quantiles collapsed over practiced and transfer items for each participant for correct responses and the AICc-selected gamma multiple-rotation model.

constant, with a shape parameter of 0.79, whereas scale increases by 0.062 per 45°.

Although the two sets of rotation models have very different shapes their central tendencies are very similar. For the multiple-rotation model the mean time for a 180° rotation is 0.42 s and for the single-rotation model it is 0.41 s. There were also strong individual differences in rotation-times according to both models. Mean rates were highly correlated between models ($r = .96$), and ranged from as fast as 0.6 ms/degree to as slow as 4.8 ms/degree.

As is evident in Figure 8, rotation time is more variable in the multiple-rotation model, with a standard deviation of 0.33 s at 180° compared with a standard deviation of 0.25 s for the single-rotation model. The higher threshold estimates for the single-rotation model—which result in greater decision time variability—explain why both models provide a very similar account of overall variability.

Mirror Slowing

The models examined so far explain stimulus (i.e., mirror vs. normal) mainly in terms of threshold (response bias) differences. Converging evidence for a bias toward normal responses is provided by an examination of the relative speed of RT for error and correct responses. As shown in Figure 9, for mirror stimuli error and correct RT medians were equivalent, whereas errors to normal stimuli were slow. This pattern is accommodated in the AICc-selected models (Figure 9 illustrates the fit of the multiple-rotation version) by larger threshold estimates for the mirror accumulator than normal accumulator. This is because errors to normal stimuli (i.e., mirror responses) will be slowed if the mirror threshold is larger. Similarly, errors to mirror stimuli (i.e., normal responses) will be speeded if the normal threshold is lower. As errors were relatively rare in most conditions for most participants it is difficult

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

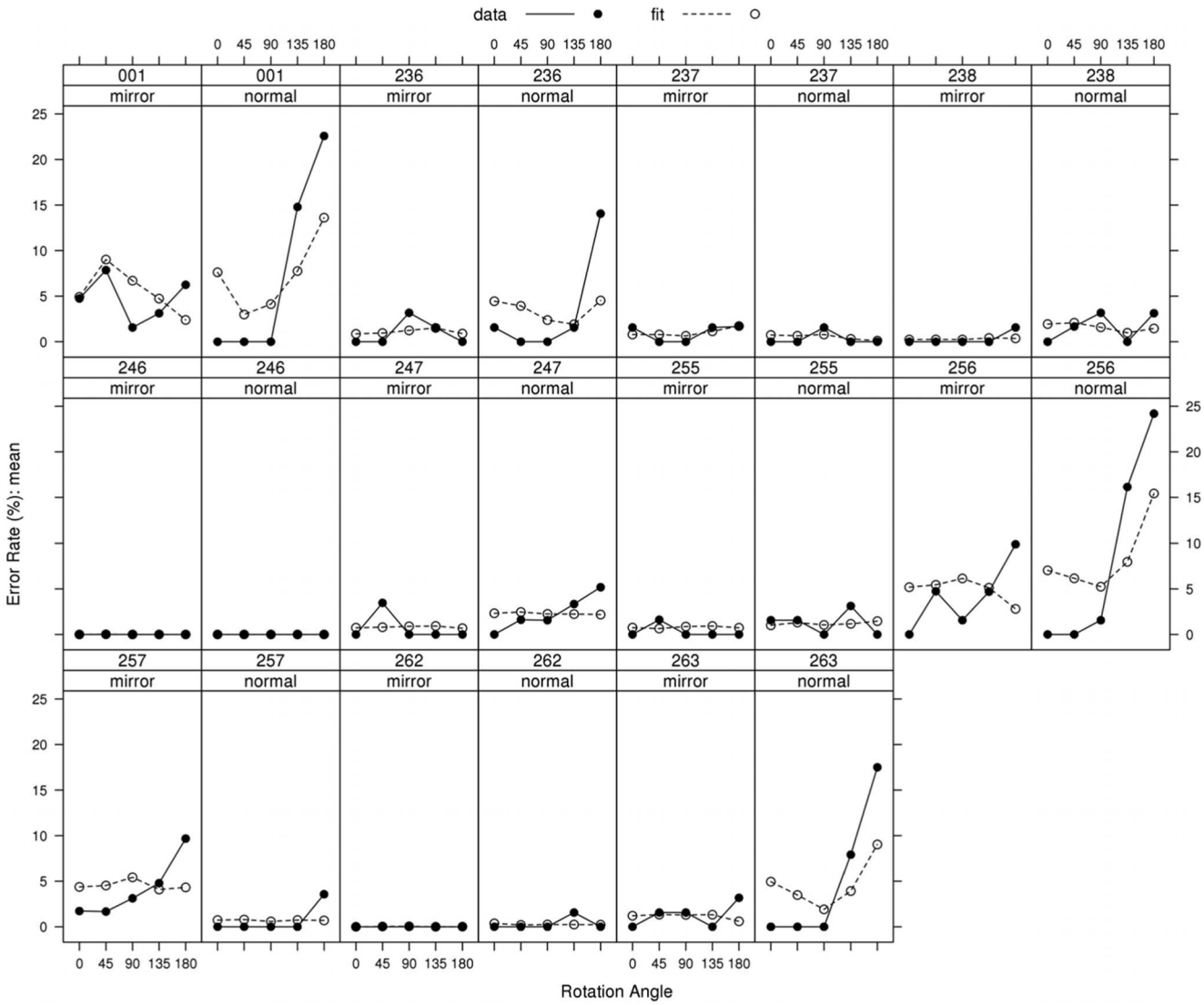

Figure 6. Observed error rates collapsed over practiced and transfer items for each participant, and the AICc-selected gamma multiple-rotation model.

to evaluate the model's account of error RT distribution, but Figure 9 shows a good fit to its central tendency.

We sought to test Hamm et al.'s (2004) flip hypothesis by adding an intercept parameter to both the single and multiple rotation models.[8] Table 3 reports results for both the top models and the AICc selected models. AICc selected exactly the same pattern of effects as for the models with no rotation intercept (see Table 2). However, the model selection results are fairly equivocal. At the group level there is a significant improvement in fit for the multiple-rotation model and particularly the single rotation model, both for the top, $\chi^2(11) = 25$, $p = .009$, and $\chi^2(11) = 77$, $p < .001$, and the AICc selected, $\chi^2(11) = 20$, $p = .045$ and $\chi^2(11) = 85$, $p < .001$, variants. However, the improvement was only significant for 5 of the 11 participants for the top variants and 4 participants for the AICc selected variants. In terms of AICc, at the group level there was mild evidence rejecting an intercept for the multiple rotation model, $\Delta = 3$, $w = 0.8$, but strong evidence favoring it for the single-rotation model, $\Delta = 68$. At the individual participant level over both models results were equivocal for the majority of participants. Three participants had $\Delta > 6$ favoring an intercept and three against for the multiple rotation model, whereas for the single rotation model three had $\Delta > 6$ favoring an intercept.

Given these unclear results, we sought converging evidence by examining the event-related potential data collected by Provost et al. (2013). Hamm et al. (2004) based their conclusion that there was extra rotation-related activity for mirror characters on analyses of both the Global Mean Field Power (GMFP) and topography of

[8] Note that the rotation model intercept affects the variability as well as the mean and so has a distinct effect to allowing nondecision time to vary between mirror and normal stimuli.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

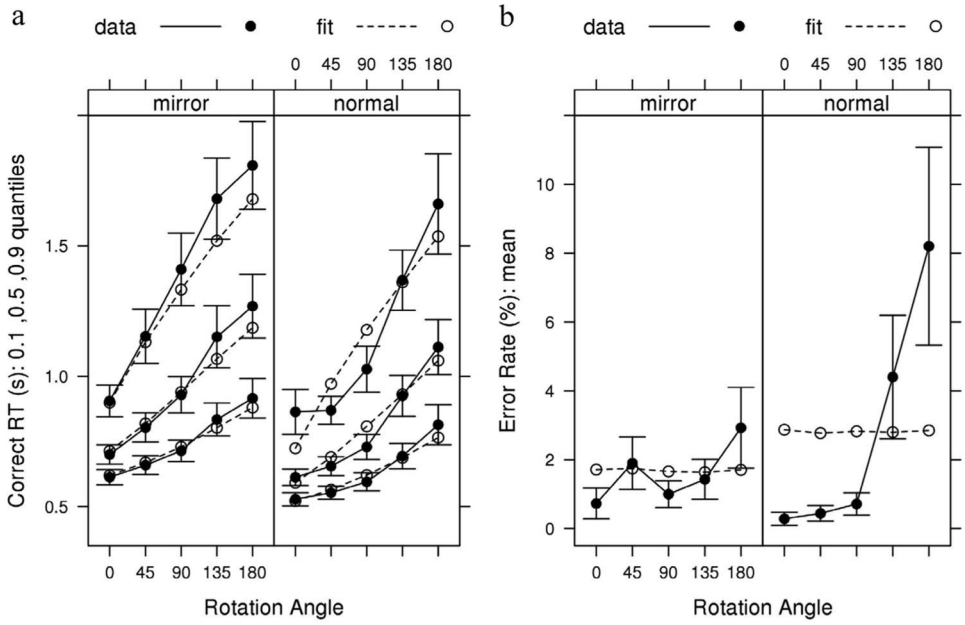

Figure 7. Observed (a) reaction time (RT) quantiles for correct responses and (b) error rates, both averaged over practiced and transfer items and participants with standard errors, and fits of the AICc-selected gamma multiple-rotation model.

their event-related potential data. They created an index of the flip by taking the difference between 0° normal and 0° mirror conditions, and found that the topography of this difference waveform (i.e., its distribution over the scalp) was consistent with a parietal peak consistent with an enhancement in the waveform associated with planar mental rotation, rotation-related negativity (their Figure 7). They also created an index of planar mental-rotation effect by taking the difference between 180° and 0° conditions, and

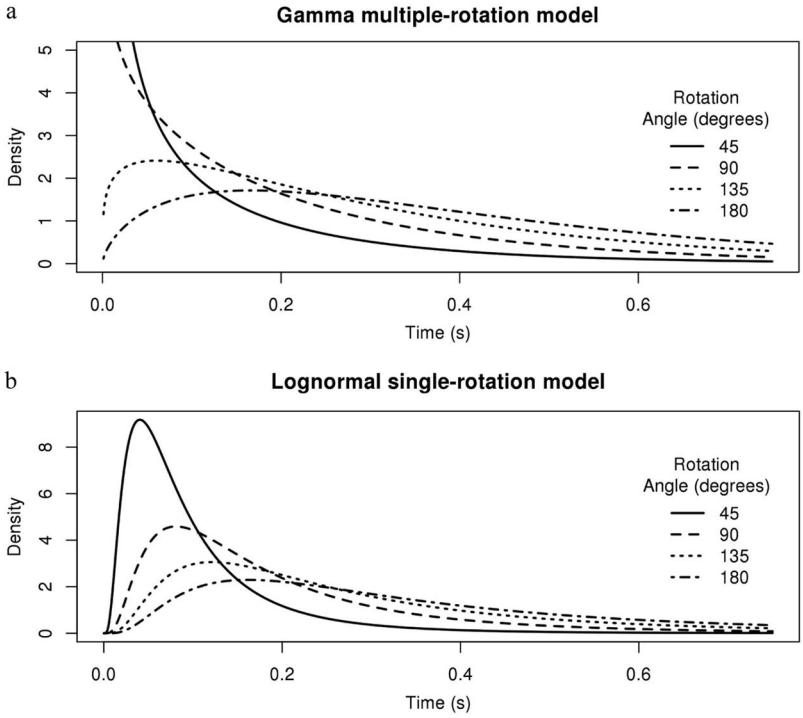

Figure 8. Rotation-time distributions corresponding to the average parameter estimates from the AICc-selected gamma multiple-rotation model and the lognormal single-rotation model.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

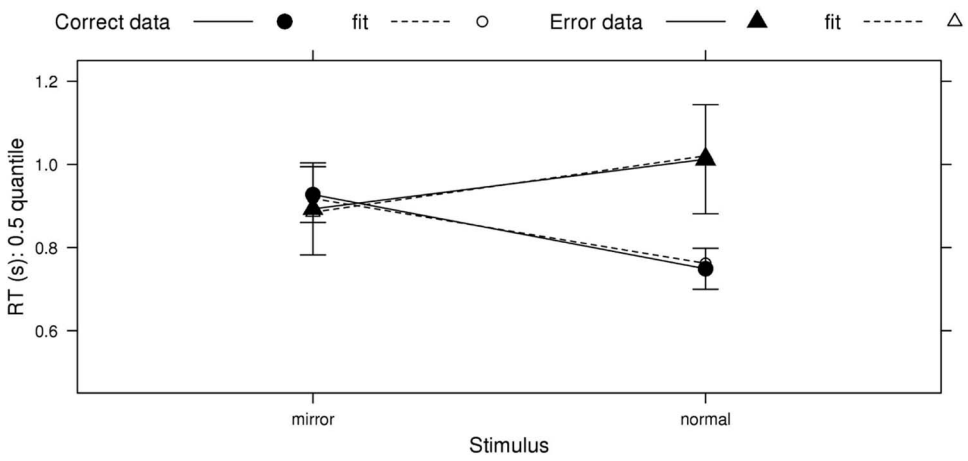

Figure 9. Observed reaction time (RT) medians collapsed over practiced and transfer items and rotation angle for correct and error responses averaged over participants, and standard errors and fits of the gamma multiple-rotation model.

found greater GMFP in this difference for mirror than normal stimuli between 0.5 s and 0.8 s poststimulus (their Figure 8d).

When we carried out a similar GMFP calculation on Provost et al.'s (2013) data for the last session of their second experiment we found a similar difference between 0.4 s and 0.8 s (see Figure 10). However, when we examined the topography of the 180° − 0° normal difference and 0° mirror − 0° normal difference waveforms in the 0.4-s to 0.8-s time window they were quite different (see Figure 11). The former topography is consistent with greater parietal rotation-related negativity for the 180° normal condition. However, in contrast to Hamm et al.'s (2004) findings for letters, the 0° mirror versus 0° contrast—indexing the mirror versus normal processing in Provost et al.'s (2013) data—produced quite a different topography, with more central and frontal negativity.

Hence our event-related potential results do not appear to be consistent with the flip topography found by Hamm et al. (2004) for letters. Instead, the 0° mirror versus normal differences is associated with a corresponding increase in fronto-central activation, rather than parietally maximal activation. This could be consistent with differences in the decision stage, such as a higher threshold for the mirror condition and/or the greater degree of evidence accumulation required to trigger a mirror response compared with a normal response.

Table 3
Fit Results For Models With Rotation intercepts, Both the Top Models and Model Variants That Had the Lowest Group Akaike Information Criteria (AICc)

Rotation Model	*p*	Deviance	AICc
Top Gamma multiple-rotation	37	−3262	−2396
AICc Gamma multiple-rotation	21	−3175	−2697
Top Lognormal single-rotation	37	**−3267**	−2401
AICc Lognormal single-rotation	21	−3193	**−2715**

Note. The best (i.e., smallest) fit statistics are shown in bold.

General Discussion

The enduring influence of Shepard and Metzler (1971) is evidenced by the fact that a conservative estimate of its citation count is well over 2000, and that it remains the most cited of Shepard's many influential works. Despite this intense and enduring interest, it is only very recently that researchers have attempted to provide a comprehensive quantitative account of behavior in the mental-rotation task. Kung and Hamm (2010) modeled mean RT (see also Searle & Hamm, 2012), Kelley et al. (2000) addressed errors, and Larsen (2014) modeled both. Our work extends this new direction by also taking into account the distribution of RT. This extension proved crucial to our aim of performing a "cognitive titration," that is, determining what portion of RT can be attributed to time to perform mental rotation as opposed to response selection and other processes.

We framed our analysis using Cooper and Shepard's (1973) decomposition of the mental rotation task into four stages: visual

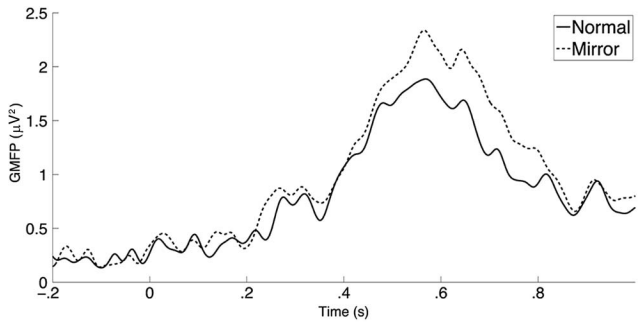

Figure 10. Global Mean Field Power (GMFP) calculated for subtraction waveforms, 180° normal − 0° normal, for the normal (unbroken line) and mirror 180° mirror − 0° normal (broken line) stimulus grand averages. GMFP is the standard deviation of all electrodes, and demonstrates in these subtraction waveforms the predominant activity for both normal and mirror rotation is between 0.4 and 0.8 s.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

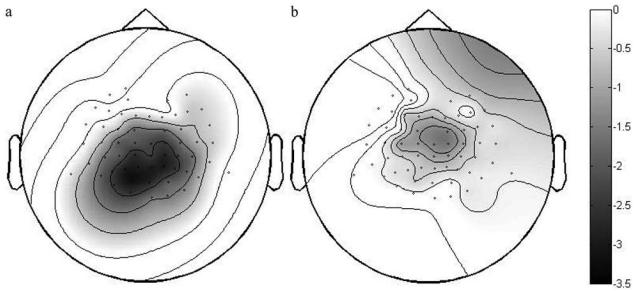

Figure 11. Topographical subtraction plots of (a) 180° normal − 0° normal conditions, representing the planar mental rotation effect on topography, and (b) 0° normal − 0° mirror conditions, representing the flip-effect on topography. The gray scale represents μV.

encoding, mental rotation, response selection, and response production. We assumed these stages occurred sequentially, with one stage finishing before the next commenced. An established evidence-accumulation theory—Brown and Heathcote's (2008) LBA—modeled response selection. We also borrowed a convention from the evidence-accumulation-modeling literature, where the times for visual encoding and response production were combined into a single parameter.

We compared a number of different models of the rotation stage, including a deterministic stage (i.e., no random variation in rotation time) and two models, the multiple-rotation and single-rotation models, where rotation time was variable, having a gamma or lognormal distribution, respectively. In the multiple-rotation model rotation angle linearly affected the shape parameter of the gamma distribution, whereas in the single-rotation model rotation angle linearly affected the scale of the lognormal rotation-time distribution. The deterministic-rotation model also assumed a linear angle effect, with linearity in all three cases being based on the assumption that rotation rate is a constant that does not vary with the amount of rotation done.

These models were tested against data reported by Provost et al. (2013) where there was strong converging evidence from event-related potential and transfer effects that participants used a pure mental rotation strategy. We first discuss how and why these comparisons favored the multiple-rotation model, and then address the implications of our results for understanding mirror slowing (i.e., slower responses to mirror than normal stimuli), and nonlinear effects on performance of rotation angle.

Models of Mental Rotation

Both models that assumed variability in rotation time were clearly better than the deterministic-rotation model. Hence, our results indicated that, if it is assumed there is a separate rotation stage that must finish before the decision stage begins, the time to complete the rotation stage varies appreciably from trial to trial. Empirically, rotation angle caused an increase not only in mean RT but also RT variability. These increases could have been explained by a decrease in evidence-accumulation rates as rotation angle increased. However, fits of both the multiple-rotation and single-rotation models attributed most of the increase in RT mean and variability to the rotation stage. The average estimated standard deviation of rotation time was about 0.3 s, with a somewhat smaller estimate for the single-rotation model (0.25 s) and a somewhat larger estimate for the multiple-rotation model (0.33 s).

There was much less difference between the two models in estimates of mean rotation time, with the average time to perform an 180° rotation estimated as approximately 0.4 s by both. Hence, both models predict a relatively fast mean rotation rate of 2.3 ms/degree, consistent with perceptually based estimates of the maximum angular rotation rate (e.g., Farrell et al., 1982). Assuming that half of the estimated average nondecision time of around 0.3 s is consumed by an initial stimulus encoding stage, these results suggest that mental rotation is occurring, on average, in a window from 0.25 s (for 45° stimuli) to 0.55 s (for 180° stimuli) poststimulus.

Converging evidence supporting the model-based estimates of rotation times comes from the neurosciences. Perhaps the most directly relevant is Harris and Miniussi's (2003) repetitive transcranial magnetic stimulation (rTMS) study. They found the effect of rotation angle on mean RT was increased in participants who received rTMS over the right parietal lobe, and that the slowing was maximal when stimulation was applied in a window 400 ms to 600 ms poststimulus. This window coincides with the maximal negative event-related potential modulation by rotation angle, the rotation-related negativity component (e.g., Heil, 2002). Provost et al. (2013) found rotation-related negativity in the final session of their second experiment that occurred earlier than usual, which they attributed to the effect of extensive prior practice. This brings the observed rotation-related negativities' timing into agreement with variable rotation time model's estimates of the timing of the rotation stage.

The multiple-rotation model implies that rotation time becomes less skewed for larger angles, with a variance that increases in proportion to the mean time. The single-rotation model implies a constant skew, and that it is the standard deviation that increases in proportion to the mean. Our results favored the multiple-rotation model over the single-rotation model, although this finding was not as strong as the rejection of the deterministic rotation model. We also showed that this conclusion did not depend strongly on the lognormal rotation-time assumed by the single-rotation model, as a version instead assuming a gamma distribution was also inferior to the multiple-rotation model, which also assumed a gamma distribution. This suggests it was the change in shape with rotation angle (i.e., decreasing skew) and the associated linear increase in variance—both of which consequences of the fundamental assumption of multiple additive rotation times—that were the source of the multiple-rotation model's advantage.

Although these results support the idea of multiple rotations occurring before decision processing starts, we did not find support for interleaved rotation and decision processing within an LBA framework. In particular, we tested an LBA model inspired by Larsen's (2014) model where each of a series of multiple rotations provides a unit of evidence, and the time for each addition being proportional to the required angle of rotation. These assumptions imply that the reciprocal of the LBA mean evidence accumulation rate is a linear function of the angle of rotation. Although this model is very parsimonious, it was inferior to both models that assume a rotation stage that takes a varying amount of time to complete. Thus, it appears that we can reject the idea of interleaved rotation and decision stages within an LBA framework. However,

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

clearly this does not mean we cannot reject Larsen's instantiation of interleaving, because we did not directly test it.

Once again, evidence from the neurosciences provides some further constraint on this issue. The assumption that rotation and decision processes are interleaved also implies that neural markers of rotation processing should continue to be evident right up until close to the point of response selection. Even if response production takes a relatively long time this means that the rotation-related negativity component should be evident until close to the response. This seems to be incompatible with Riečanský and Jagla (2008) and Riečanský et al. (2013), who both found the rotation-related negativity peak occurred approximately 0.4 s before response onset when event-related potentials are calculated relative to the response.

Overall, therefore, it appears that, at least in Provost et al.'s (2013) paradigm, where the required rotation angle is well defined for all stimuli, a rotation stage must be completed before response selection begins, consistent with Shepard and Cooper's (1973) results from a precued matching paradigm discussed earlier. Our rotation-stage models share with most theories the idea that mental rotation is an analogue of constant-rate physical rotation, and go beyond them in making commitments to the form of variability in rotation time. However, they remain abstract in the sense of not describing the computational or neural processes by which image information is transformed, and clearly it would be desirable to provide such extensions in the future.

By letting rotation angle affect the mean rate of evidence accumulation our analysis allowed for the possibility that effects arising from the rotation stage (e.g., distortions of information about the stimulus image) could affect accuracy in the decision stage. If our results had supported such effects our account of the rotation stage would also have been incomplete by not specifying the mechanisms causing such distortions. However, we did not find any effects of rotation angle on the inputs to the decision stage, suggesting that, at least for Provost et al.'s (2013) data where the images were relatively simple and accuracy quite high for most participants, that our purely temporal characterization of the rotation stage was sufficient.

Even so, it seems possible that with more complex stimuli the same may not be true. We believe that the development of computationally explicit models of the rotation stage will likely benefit from data with higher error rates obtained from paradigms using more confusable and complex normal and mirror stimuli. Certainly, mental rotation paradigms that produce higher error rates, and manipulations that affect error rates, will particularly benefit evidence-accumulation modeling.

Another direction in which our models could be extended is to the many paradigms in which participants do not know the required angle of rotation. Our multiple-rotation model is most compatible with such paradigms for the same reasons that motivated Larsen (2014) to originally propose the idea of multiple rotations. Our finding that the multiple-rotation model was preferred to a single-rotation model even in a paradigm where rotation angle is known suggests that multiple rotations are likely the default method of performing mental rotation tasks.

Clearly, however, further assumptions will be required to extend our instantiation of the multiple-rotation idea to such paradigms. This is particularly so with respect to mismatching stimuli, as a distribution of rotation angles will have to be specified, adding to the complexity of the model. It was for this reason that we avoided such paradigms in our initial exploration, but given the frequency with which they occur in the literature, and their likely importance in applied contexts, this is an important direction for future research.

Mirror Slowing

Cooper and Shepard (1973) attributed the common finding that responses to mirror stimuli are slower than responses to normal stimuli to greater preparation to execute a normal response. We performed a model-based test of this response-production hypothesis, by allowing the models nondecision time, which is partially constituted of response-production time, to differ between normal and mirror responses. We also tested an alternative explanation in terms of preparation in the decision stage, by allowing the evidence thresholds for normal and mirror accumulators to differ. Under this explanation mirror slowing corresponds to a higher threshold for selecting a mirror response. Hamm et al. (2004) proposed a very different explanation of mirror slowing for letters, that the rotation stage not only rotates mirror stimuli to upright but then also flips them to a normal view. We incorporated a parameter to measure the duration of a flip in our rotation models.

Our results indicated that a higher threshold for the mirror accumulator caused over 80% of the approximately 0.17-s mirror slowing observed on average in Provost et al.'s (2013) second experiment. The remainder was caused by a slightly faster nondecision time for normal stimuli. Our results also suggested there is not a strong case in this data for the occurrence of a flip—or to be more accurate a component flip—that normalizes right-hand mirror-image members of stimulus pairs to match the left-hand member. However, there are clearly many differences between Provost et al.'s stimuli and the letter stimuli for which evidence supporting the flip hypothesis has mainly been reported (e.g., the flip is of the full letter image, not a part, and letters are much more familiar and are intrinsically two dimensional rather than being projections of three dimensional objects). Hence, further research is required to determine whether the alternative explanations offered here are viable for letter stimuli.

Given the we did not find evidence for a flip, why was there a positive correlation observed between slopes and intercepts for mean correct RT observed in Provost et al.'s (2013) data ($r = .48$)? We previously speculated that this might be due to individual differences in speed having a proportional effect on slopes and intercepts. Consistent with this speculation, a correlation of $r = .51$ was found for mean correct RT predicted by the AICc-selected multiple-rotation model and $r = .6$ for the AICc-selected single-rotation model. Hence, it again appears that, at least for the data from the last session of Provost et al.'s second experiment, a threshold difference is sufficient to explain most aspects of mirror slowing.

Beyond Linear Rotation Effects

Our results rejected a role for drift rates in explaining angle effects, but indicated that—as suggested by Bundesen (1990) in the context of visual search and Larsen (2014) in the context of mental rotation—participants are able to adjust their thresholds contingent on a stimulus property, which in this case is rotation

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

angle. In particular, we found an upward curvilinear relationship between rotation angle and response thresholds, particularly for normal stimuli. These threshold effects explained corresponding upward curvilinear effect in error and correct mean RT rotation functions (i.e., plots of these measures against rotation angle). Again, these effects were most evident for normal stimuli, with the curvilinear mean RT effect superimposed on a larger linear increase attributed to mental-rotation time.

Kung and Hamm (2010, see also Searle & Hamm, 2012) proposed that a mixture of mental rotation and a second process insensitive to rotation angle explained the upward curvilinearity in RT rotation functions commonly found with character stimuli (e.g., Corballis & McMaster, 1996; Hamm et al., 2004). The second process is based on the ability to identify disoriented alphanumeric stimuli without mental rotation (Corballis, 1988; Hamm & McMullen, 1998), and so this explanation does not apply to Shepard and Metzler (1971) type stimuli. Also, Kung and Hamm did not use their mixture approach to model error rotation functions.

Error rotation functions are less often reported than RT rotation functions, but where they have been there is evidence for a similar pattern to the one we found, although this can be obscured by stimulus differences. Paschke, Jordan, Wüstenberg, Baudewig, & Müller (2012) found a strong upward curvilinear increase in errors for normal stimuli but a downward curvilinear function for mirror-stimulus errors. The same downward curvilinear pattern was found for mirror-stimulus RT, with the usual linear increase for normal-stimulus RT. The mirror RT result likely reflects the use of stimuli modeled on Shepard and Metzler's (1971), which do not have an upright orientation, and so rotation angle is ill defined for mirror stimuli (see also Jansen-Osman & Heil, 2007; Larsen, 2014).

Several studies using letter stimuli report some elements of the pattern we found. Kung and Hamm (2010) reported both the stronger upward curvature in errors for normal than mirror stimuli. There was also greater upward curvilinearity for normal than mirror stimuli in RT, although in contrast to our findings even mirror RT was quite curved. Corballis and McMaster (1996) found upward curvilinear error rotation functions, but with no clear normal versus mirror differences. They also reported upward curvature in RT functions that was larger for normal than mirror stimuli. Finally, Hamm et al. (2004) also reported evidence of upward curvature in error and RT functions, although again there was little difference in these effects between normal and mirror stimuli.

This summary indicates that there is potential for threshold effects to provide a unified account of nonlinearity in both error and RT rotation functions. Because this account is highly constrained by the relationship between errors and RT enforced by the evidence-accumulation model, it is eminently testable in future research, although the model would need to be extended to deal with cases where there is no well-defined upright orientation. Such research should take account of the possibility that there are strong individual differences in threshold effects, suggesting that different participants may adopt different strategies, resulting in individual variability in related effects on RT and accuracy.

More broadly, and in line with Kung and Hamm's (2010) approach, we believe a general account of mental rotation will require a framework in which participants may, potentially to differing degrees, use a mixture of strategies (see also Kelley et al., 2000). In that regard our models are advantageous in that they provide explicit values for the likelihood of joint choice and RT data. Likelihoods naturally support mixture modeling, because the likelihood of a mixture is a simple linear combination of the likelihoods of each component weighted by the probability that each occurs. Further, Turner and Sederberg (2014) recently expanded the generality of the likelihood approach, describing innovations enabling likelihoods to be approximated for a component that can only be simulated.

On several issues examined here, such as the flip hypothesis and the no-rotation-stage model, results from the neuroimaging literature provided a source of converging constraint. A further advantage of likelihood-based approaches is that they enable a more direct integration of behavioral and neuroimaging evidence, using joint Bayesian modeling (e.g., Turner et al., 2013). Quantitative approaches to integrating brain and behavioral results (see Forstmann, Wagenmakers, Eichele, Brown, & Serences, 2011) have been very productive in understanding of choice processes (e.g., Forstmann et al., 2008; van Maanen et al., 2011; Mansfield, Karayanidis, Jamadar, Heathcote, & Forstmann, 2011; van Vugt, Simen, Nystrom, Holmes, & Cohen, 2012). The same may well be true in the study of spatial skills, where there already exists a rich array of data from the neurosciences. The models of mental rotation developed here provide clear predictions about the nature and timing of cognitive process that will hopefully provide a platform for such integrative research in the future.

References

Band, G. P., & Miller, J. (1997). Mental rotation interferes with response preparation. *Journal of Experimental Psychology: Human Perception and Performance, 23,* 319–338. http://dx.doi.org/10.1037/0096-1523.23.2.319

Bethell-Fox, C. E., & Shepard, R. N. (1988). Mental rotation: Effects of stimulus complexity and familiarity. *Journal of Experimental Psychology: Human Perception and Performance, 14,* 12–23. http://dx.doi.org/10.1037/0096-1523.14.1.12

Brown, S. D., & Heathcote, A. (2008). The simplest complete model of choice response time: Linear ballistic accumulation. *Cognitive Psychology, 57,* 153–178. http://dx.doi.org/10.1016/j.cogpsych.2007.12.002

Bundesen, C. (1990). A theory of visual attention. *Psychological Review, 97,* 523–547. http://dx.doi.org/10.1037/0033-295X.97.4.523

Burnham, K. P., & Anderson, D. R. (2004). Multimodel Inference: Understanding AIC and BIC in Model Selection. *Sociological Methods & Research, 33,* 261–304. http://dx.doi.org/10.1177/0049124104268644

Corballis, M. C., & McLaren, R. (1982). Interaction between perceived and imagined rotation. *Journal of Experimental Psychology: Human Perception and Performance, 8,* 215–224. http://dx.doi.org/10.1037/0096-1523.8.2.215

Cooper, L. A., & Shepard, R. N. (1973). Chronometric studies of the rotation of mental images. In W. G. Chase (Ed.), *Visual information processing* (pp. 75–176). New York, NY: Academic Press. http://dx.doi.org/10.1016/B978-0-12-170150-5.50009-3

Corballis, M. C. (1988). Recognition of disoriented shapes. *Psychological Review, 95,* 115–123. http://dx.doi.org/10.1037/0033-295X.95.1.115

Corballis, M. C., & McLaren, R. (1984). Winding one's *p*s and *q*s: Mental rotation and mirror-image discrimination. *Journal of Experimental Psychology: Human Perception and Performance, 10,* 318–327. http://dx.doi.org/10.1037/0096-1523.10.2.318

Corballis, M. C., & McMaster, H. (1996). The roles of stimulus-response compatibility and mental rotation in mirror-image and left–right decisions. *Canadian Journal of Experimental Psychology/Revue Cana-*

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

dienne de Psychologie Experimentale, 50, 397–401. http://dx.doi.org/10.1037/1196-1961.50.4.397

Donkin, C., Brown, S. D., & Heathcote, A. (2009). The overconstraint of response time models: Rethinking the scaling problem. *Psychonomic Bulletin & Review, 16,* 1129–1135. http://dx.doi.org/10.3758/PBR.16.6.1129

Donkin, C., Brown, S. D., & Heathcote, A. (2011). Drawing conclusions from choice response time models: A tutorial using the linear ballistic accumulator. *Journal of Mathematical Psychology, 55,* 140–151. http://dx.doi.org/10.1016/j.jmp.2010.10.001

Donkin, C., & Nosofsky, R. M. (2012a). A power-law model of psychological memory strength in short- and long-term recognition. *Psychological Science, 23,* 625–634. http://dx.doi.org/10.1177/0956797611430961

Donkin, C., & Nosofsky, R. M. (2012b). The structure of short-term memory scanning: An investigation using response time distribution models. *Psychonomic Bulletin & Review, 19,* 363–394. http://dx.doi.org/10.3758/s13423-012-0236-8

Donkin, C., Nosofsky, R. M., Gold, J. M., & Shiffrin, R. M. (2013). Discrete-slots models of visual working-memory response times. *Psychological Review, 120,* 873–902. http://dx.doi.org/10.1037/a0034247

Duncombe, M. E., Bradshaw, J. L., Iansek, R., & Phillips, J. G. (1994). Parkinsonian patients without dementia or depression do not suffer from bradyphrenia as indexed by performance in mental rotation tasks with and without advance information. *Neuropsychologia, 32,* 1383–1396. http://dx.doi.org/10.1016/0028-3932(94)00071-9

Farrell, J. E., Larsen, A., & Bundesen, C. (1982). Velocity constraints on apparent rotational movement. *Perception, 11,* 541–546. http://dx.doi.org/10.1068/p110541

Folk, M. D., & Luce, R. D. (1987). Effects of stimulus complexity on mental rotation rate of polygons. *Journal of Experimental Psychology: Human Perception and Performance, 13,* 395–404. http://dx.doi.org/10.1037/0096-1523.13.3.395

Forstmann, B. U., Dutilh, G., Brown, S., Neumann, J., von Cramon, D. Y., Ridderinkhof, K. R., & Wagenmakers, E. J. (2008). Striatum and pre-SMA facilitate decision-making under time pressure. *Proceedings of the National Academy of Sciences of the United States of America, 105,* 17538–17542. http://dx.doi.org/10.1073/pnas.0805903105

Forstmann, B. U., Wagenmakers, E.-J., Eichele, T., Brown, S., & Serences, J. T. (2011). Reciprocal relations between cognitive neuroscience and formal cognitive models: Opposites attract? *Trends in Cognitive Sciences, 15,* 272–279. http://dx.doi.org/10.1016/j.tics.2011.04.002

Georgopoulos, A. P., Lurito, J. T., Petrides, M., Schwartz, A. B., & Massey, J. T. (1989, January 13). Mental rotation of the neuronal population vector. *Science, 243,* 234–236. http://dx.doi.org/10.1126/science.2911737

Hamm, J. P., Johnson, B. W., & Corballis, M. C. (2004). One good turn deserves another: An event-related brain potential study of rotated mirror-normal letter discriminations. *Neuropsychologia, 42,* 810–820. http://dx.doi.org/10.1016/j.neuropsychologia.2003.11.009

Hamm, J. P., & McMullen, P. A. (1998). Effects of orientation on the identification of rotated objects depend on the level of identity. *Journal of Experimental Psychology: Human Perception and Performance, 24,* 413–426. http://dx.doi.org/10.1037/0096-1523.24.2.413

Harris, I. M., & Miniussi, C. (2003). Parietal lobe contribution to mental rotation demonstrated with rTMS. *Journal of Cognitive Neuroscience, 15,* 315–323. http://dx.doi.org/10.1162/089892903321593054

Heathcote, A., Brown, S., & Mewhort, D. J. K. (2000). The power law repealed: The case for an exponential law of practice. *Psychonomic Bulletin & Review, 7,* 185–207. http://dx.doi.org/10.3758/BF03212979

Heathcote, A., Loft, S., & Remington, R. W. (2015). Slow down and remember to remember! A delay theory of prospective memory costs. *Psychological Review, 122,* 376–410. http://dx.doi.org/10.1037/a0038952

Heathcote, A., & Love, J. (2012). Linear deterministic accumulator models of simple choice. *Frontiers in Psychology, 3,* 292. http://dx.doi.org/10.3389/fpsyg.2012.00292

Heil, M. (2002). The functional significance of ERP effects during mental rotation. *Psychophysiology, 39,* 535–545. http://dx.doi.org/10.1111/1469-8986.3950535

Heil, M., Bajric, J., Rösler, F., & Hennighausen, E. (1997). A rotation aftereffect changes both the speed and the preferred direction of mental rotation. *Journal of Experimental Psychology: Human Perception and Performance, 23,* 681–692. http://dx.doi.org/10.1037/0096-1523.23.3.681

Heil, M., Rauch, M., & Hennighausen, E. (1998). Response preparation begins before mental rotation is finished: Evidence from event-related brain potentials. *Acta Psychologica, 99,* 217–232. http://dx.doi.org/10.1016/S0001-6918(98)00012-2

Hyun, J.-S., & Luck, S. J. (2007). Visual working memory as the substrate for mental rotation. *Psychonomic Bulletin & Review, 14,* 154–158. http://dx.doi.org/10.3758/BF03194043

Jansen-Osmann, P., & Heil, M. (2007). Suitable stimuli to obtain (no) gender differences in the speed of cognitive processes involved in mental rotation. *Brain and Cognition, 64,* 217–227. http://dx.doi.org/10.1016/j.bandc.2007.03.002

Just, M. A., & Carpenter, P. A. (1985). Cognitive coordinate systems: Accounts of mental rotation and individual differences in spatial ability. *Psychological Review, 92,* 137–172. http://dx.doi.org/10.1037/0033-295X.92.2.137

Kail, R., & Park, Y. S. (1990). Impact of practice on speed of mental rotation. *Journal of Experimental Child Psychology, 49,* 227–244. http://dx.doi.org/10.1016/0022-0965(90)90056-E

Kelley, T. D., Lee, F. J., & Wiley, P. (2000). *Developing an ACT-R model of spatial manipulation* (ARL Tech. Rep. No. 2179). Aberdeen, MD: Aberdeen Proving Ground.

Kung, E., & Hamm, J. P. (2010). A model of rotated mirror/normal letter discriminations. *Memory & Cognition, 38,* 206–220. http://dx.doi.org/10.3758/MC.38.2.206

Larsen, A. (2014). Deconstructing mental rotation. *Journal of Experimental Psychology: Human Perception and Performance, 40,* 1072–1091. http://dx.doi.org/10.1037/a0035648

Larsen, A., & Bundesen, C. (2009). Common mechanisms in apparent motion perception and visual pattern matching. *Scandinavian Journal of Psychology, 50,* 526–534. http://dx.doi.org/10.1111/j.1467-9450.2009.00782.x

Mansfield, E. L., Karayanidis, F., Jamadar, S., Heathcote, A., & Forstmann, B. U. (2011). Adjustments of response threshold during task switching: A model-based functional magnetic resonance imagining study. *The Journal of Neuroscience, 31,* 14688–14692. http://dx.doi.org/10.1523/JNEUROSCI.2390-11.2011

McClelland, J. L. (1979). On the time relations of mental processes: An examination of systems of processes in cascade. *Psychological Review, 86,* 287–330. http://dx.doi.org/10.1037/0033-295X.86.4.287

Metzler, J., & Shepard, R. N. (1974). Transformational studies of the internal representation of three-dimensional objects. In R. L. Solso (Ed.), *Theories in cognitive psychology: The Loyola symposium* (pp. 147–202). Oxford, England: Erlbaum.

Myung, I. J., & Pitt, M. A. (1997). Applying Occam's razor in modeling cognition: A Bayesian approach. *Psychonomic Bulletin & Review, 4,* 79–95. http://dx.doi.org/10.3758/BF03210778

Nosofsky, R. M., Little, D. R., Donkin, C., & Fific, M. (2011). Short-term memory scanning viewed as exemplar-based categorization. *Psychological Review, 118,* 280–315. http://dx.doi.org/10.1037/a0022494

Oberauer, K. (2005). Age differences and individual differences in cognitive function. In R. W. Engle, G. Sedek, U. von Hecker, & D. N. McIntosh (Eds.), *Cognitive limitations in aging and psychopathology* (pp. 44–72). New York, NY: Cambridge University Press.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Paschke, K., Jordan, K., Wüstenberg, T., Baudewig, J., & Müller, J. L. (2012). Mirrored or identical—Is the role of visual perception underestimated in the mental rotation process of 3D-objects? A combined fMRI-eye tracking-study. *Neuropsychologia, 50,* 1844–1851. http://dx.doi.org/10.1016/j.neuropsychologia.2012.04.010

Prime, D. J., & Jolicoeur, P. (2010). Mental rotation requires visual short-term memory: Evidence from human electric cortical activity. *Journal of Cognitive Neuroscience, 22,* 2437–2446. http://dx.doi.org/10.1162/jocn.2009.21337

Provost, A., Johnson, B., Karayanidis, F., Brown, S. D., & Heathcote, A. (2013). Two routes to expertise in mental rotation. *Cognitive Science, 37,* 1321–1342. http://dx.doi.org/10.1111/cogs.12042

Rae, B., Heathcote, A., Donkin, C., Averell, L., & Brown, S. (2014). The hare and the tortoise: Emphasizing speed can change the evidence used to make decisions. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40,* 1226–1243. http://dx.doi.org/10.1037/a0036801

Ratcliff, R., & Rouder, J. N. (1998). Modeling response times for two-choice decisions. *Psychological Science, 9,* 347–356. http://dx.doi.org/10.1111/1467-9280.00067

Riečanský, I., & Jagla, F. (2008). Linking performance with brain potentials: Mental rotation-related negativity revisited. *Neuropsychologia, 46,* 3069–3073. http://dx.doi.org/10.1016/j.neuropsychologia.2008.06.016

Riečanský, I., Tomova, L., Katina, S., Bauer, H., Fischmeister, F. Ph. S., & Lamm, C. (2013). Visual image retention does not contribute to modulation of event-related potentials by mental rotation. *Brain and Cognition, 83,* 163–170. http://dx.doi.org/10.1016/j.bandc.2013.07.011

Ruthruff, E., & Miller, J. (1995). Can mental rotation begin before perception finishes? *Memory & Cognition, 23,* 408–424. http://dx.doi.org/10.3758/BF03197243

Ruthruff, E., Miller, J., & Lachmann, T. (1995). Does mental rotation require central mechanisms? *Journal of Experimental Psychology: Human Perception and Performance, 21,* 552–570. http://dx.doi.org/10.1037/0096-1523.21.3.552

Searle, J. A., & Hamm, J. P. (2012). Individual differences in the mixture ratio of rotation and nonrotation trials during rotated mirror/normal letter discriminations. *Memory & Cognition, 40,* 594–613. http://dx.doi.org/10.3758/s13421-011-0172-2

Shepard, R. N. (1978). The mental image. *American Psychologist, 33,* 125–137. http://dx.doi.org/10.1037/0003-066X.33.2.125

Shepard, R. N., & Cooper, L. A. (1982). *Mental images and their transformations*. Cambridge: MIT Press.

Shepard, R. N., & Judd, S. (1976). Perceptual illusion of rotation of three-dimensional objects. *Science, 191*(4230), 952–954. http://dx.doi.org/10.1126/science.1251207

Shepard, R. N., & Metzler, J. (1971, February 19). Mental rotation of three-dimensional objects. *Science, 171,* 701–703. http://dx.doi.org/10.1126/science.171.3972.701

Sternberg, S. (1969). The discovery of processing stages: Extensions of Donders' method. *Acta Psychologica, 30,* 276–315. http://dx.doi.org/10.1016/0001-6918(69)90055-9

Thayer, Z. C., & Johnson, B. W. (2006). Cerebral processes during visuomotor imagery of hands. *Psychophysiology, 43,* 401–412. http://dx.doi.org/10.1111/j.1469-8986.2006.00404.x

Turner, B. M., Forstmann, B. U., Wagenmakers, E.-J., Brown, S. D., Sederberg, P. B., & Steyvers, M. (2013). A Bayesian framework for simultaneously modeling neural and behavioral data. *NeuroImage, 72,* 193–206. http://dx.doi.org/10.1016/j.neuroimage.2013.01.048

Turner, B. M., & Sederberg, P. B. (2014). A generalized, likelihood-free method for posterior estimation. *Psychonomic Bulletin & Review, 21,* 227–250. http://dx.doi.org/10.3758/s13423-013-0530-0

Ulrich, R., & Miller, J. (1993). Information processing models generating lognormally distributed reaction times. *Journal of Mathematical Psychology, 37,* 513–525. http://dx.doi.org/10.1006/jmps.1993.1032

van Maanen, L., Brown, S. D., Eichele, T., Wagenmakers, E.-J., Ho, T., Serences, J., & Forstmann, B. U. (2011). Neural correlates of trial-to-trial fluctuations in response caution. *The Journal of Neuroscience, 31,* 17488–17495. http://dx.doi.org/10.1523/JNEUROSCI.2924-11.2011

van Vugt, M. K., Simen, P., Nystrom, L. E., Holmes, P., & Cohen, J. D. (2012). EEG oscillations reveal neural correlates of evidence accumulation. *Frontiers in Neuroscience, 6,* 106. http://dx.doi.org/10.3389/fnins.2012.00106

Wagenmakers, E. J., & Brown, S. (2007). On the linear relation between the mean and the standard deviation of a response time distribution. *Psychological Review, 114,* 830–841. http://dx.doi.org/10.1037/0033-295X.114.3.830

Wagenmakers, E. J., & Farrell, S. (2004). AIC model selection using Akaike weights. *Psychonomic Bulletin & Review, 11,* 192–196. http://dx.doi.org/10.3758/BF03206482

Wright, R., Thompson, W. L., Ganis, G., Newcombe, N. S., & Kosslyn, S. M. (2008). Training generalized spatial skills. *Psychonomic Bulletin & Review, 15,* 763–771. http://dx.doi.org/10.3758/PBR.15.4.763

Zacks, J. M. (2008). Neuroimaging studies of mental rotation: A meta-analysis and review. *Journal of Cognitive Neuroscience, 20,* 1–19. http://dx.doi.org/10.1162/jocn.2008.20013

Received October 11, 2014
Revision received July 3, 2015
Accepted July 8, 2015 ■

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.