AMERICAN PSYCHOLOGICAL ASSOCIATION

Psychological Review

ISSN: 0033-295X

2023, Vol. 130, No. 3, 841–852
https://doi.org/10.1037/rev0000328

THEORETICAL NOTE

Measurement Models for Visual Working Memory—A Factorial Model Comparison

Klaus Oberauer
Department of Psychology–Cognitive Psychology, University of Zurich

Several measurement models have been proposed for data from the continuous-reproduction paradigm for studying visual working memory (WM): The original mixture model (Zhang & Luck, 2008) and its extension (Bays et al., 2009); the interference measurement model (IMM; Oberauer et al., 2017), and the target confusability competition (TCC) model (Schurgin et al., 2020). This article describes a space of possible measurement models in which all these models can be placed. The space is defined by three dimensions: (a) The choice of an activation function (von-Mises or Laplace), (b) the choice of a response-selection function (variants of Luce's choice rule or of signal-detection theory), (c) and whether or not memory precision is assumed to be a constant over manipulations affecting memory. A factorial combination of these three variables generates all possible models in the model space. Fitting all models to eight data sets revealed a new model as empirically most adequate, which combines a von-Mises activation function with a signal-detection response-selection rule. The precision parameter can be treated as a constant across many experimental manipulations, though it probably varies between individuals. All modeling code and the raw data modeled are available on the OSF: https://osf.io/zwprv/

Keywords: visual working memory, measurement model, model comparison, interference, signal-detection theory

Working memory (WM) for simple visual features is often studied with a continuous-reproduction task, in which participants are asked to remember an array of visual objects characterized by a simple visual feature varying on a continuous dimension, such as color or orientation. At test, one array item is selected at random as the target, and participants try to reproduce the target's feature on a continuous, usually circular, response scale, such as selecting its color on a color wheel (Blake et al., 1997; Wilken & Ma, 2004). Research using this task has benefited substantially from measurement models. *Measurement models* are models that are primarily used to measure theoretically meaningful latent variables—such as the precision of memory—as their free parameters. They are typically applied so that separate parameter estimates are obtained in each experimental condition, and compared between conditions. Measurement models can be distinguished from *explanatory models*, which aim to explain the effects of experimental manipulations, and therefore need to account for data across all conditions of an experiment with a single set of parameters. A number of explanatory models have been proposed for continuous-reproduction data, and the competition between them continues (Oberauer & Lin, 2017; Schneegans & Bays, 2017; Schneegans et al., 2020; van den Berg et al., 2014; van den Berg & Ma, 2018).

Measurement models for continuous-reproduction data have been spearheaded by the *mixture model* of Zhang and Luck (2008). The mixture model describes the distribution of responses as a mixture of two components: A circular normal distribution (i.e., a von-Mises distribution) centered on the target's true feature, and a uniform distribution on the circle. This model is motivated by the assumption that WM has a limited number of slots, each of which can represent one array item with a certain precision. When the target is represented in a slot, responses are assumed to be distributed around the true feature according to a von-Mises distribution. When the target is not maintained in a slot, the person has no information about it and must guess, leading to a uniform distribution of responses. The model has two free parameters, the precision of the von-Mises distribution, reflecting the precision of features stored in a slot, and the mixture weight of the von-Mises distribution, reflecting the probability that an array item is stored in a slot. The mixture model has soon been extended by Bays et al. (2009) by a third component of the mixture, reflecting responses centered on nontarget features in the array. The extended mixture model has a third free parameter for the mixture weight of nontargets.

An alternative approach was taken by Oberauer et al. (2017), who proposed a measurement model derived from the Interference Model of visual WM (Oberauer & Lin, 2017). The interference

This article was published Online First September 27, 2021.
Klaus Oberauer https://orcid.org/0000-0003-3902-7318
The research reported in this article was supported by a grant from the Swiss National Science Foundation (grant number 100014_126766) to the author. All raw data and model code can be found on the Open Science Framework (OSF): https://osf.io/zwprv/ (Oberauer, 2021).
Correspondence concerning this article should be addressed to Klaus Oberauer, Department of Psychology—Cognitive Psychology, University of Zurich, Binzmühlestrasse 14/22, 8050 Zürich, Switzerland. Email: k.oberauer@psychologie.uzh.ch

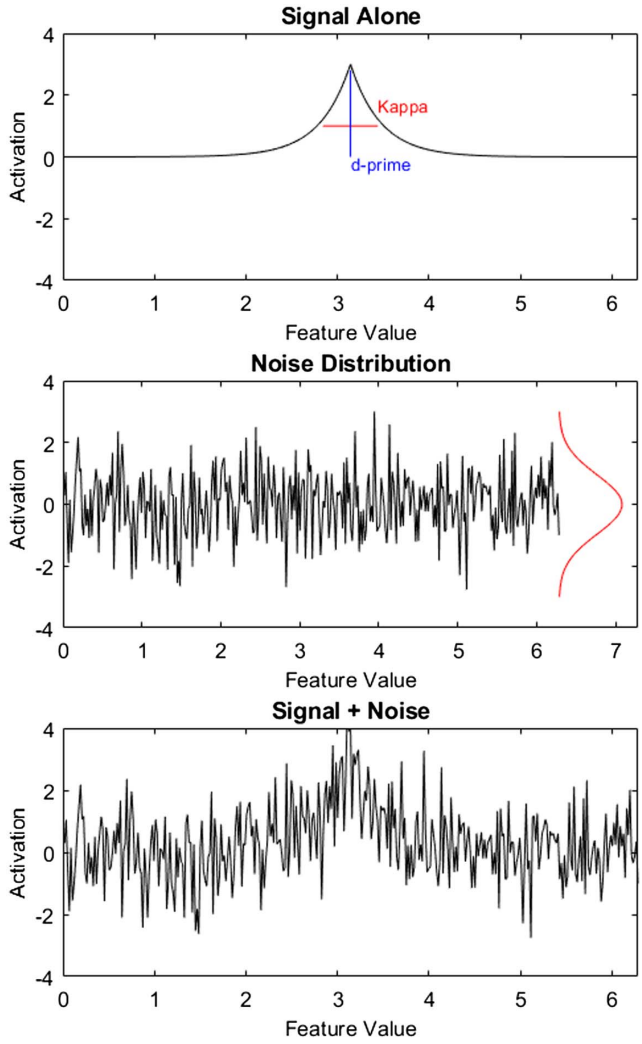
measurement model (IMM) builds on the assumption that access to an item in WM relies on cue-based retrieval. The probe identifying the test target—usually a marker of its location in the array—is used as a retrieval cue into WM, which results in a distribution of activation over all possible responses. This activation distribution consists of a mixture of von-Mises distributions centered on each of the features in the array, weighted by how strongly they are cued. The target feature is cued most strongly; nontarget features receive a larger weight when they are close to the target on the cueing dimension (usually spatial location). In addition, uniformly distributed background noise contributes to the activation distribution. The probability of choosing each response is obtained by a version of Luce's choice rule, which normalizes the activation distribution by dividing each response's activation by the sum of activations across all responses. Although the theoretical interpretation of the IMM differs from the interpretation of the mixture models, both the two-parameter and the three-parameter mixture model have been shown to be special cases of the IMM (Oberauer et al., 2017).

Recently, Schurgin et al. (2020) have proposed a new measurement model for continuous reproduction that they present as fundamentally different from existing models. Their target confusability competition (TCC) model is a signal-detection model that treats the continuous-reproduction task as a 360-alternative forced choice (360-AFC) task, with the 360° on the circular response scale as the possible response alternatives. According to the TCC model (illustrated in Figure 1), encoding an array increases the activation of each feature in the array by an amount d'. This increase in activation generalizes to neighboring features on the continuous scale according to a psychophysical scaling function, which maps the distance between features on the circular scale to their similarity. The authors demonstrate that this function can be measured through a perceptual-comparison experiment, in which people decide which of two features is more similar to a third (the standard). In line with decades of research on generalization and similarity, the psychophysical scaling function can be described well by an exponential function (Shepard, 1987). At test, the familiarity of each of the 360 response alternatives corresponds to the activation of that feature in memory, plus noise drawn from a standard normal distribution. The person then chooses the response option with the highest familiarity signal. The model has only one free parameter, d'. The TCC provides a very good fit to the typical shape of response distributions in continuous reproduction. Moreover, varying the model's only free parameter is sufficient for it to account well for variations of the response distribution across experimental conditions, such as variations of memory set size, presentation duration of the array, and the length of the retention interval.

Here I show that the TCC model is more similar to existing measurement models than it might appear from how Schurgin et al. (2020) present it. In particular, the TCC is conceptually very similar to the IMM: Both measurement models assume that responses in the continuous-reproduction task are driven by a distribution of activation over all possible responses (e.g., all possible colors in a color wheel), which is composed of a signal distribution that peaks over the correct response (i.e., the target feature) and a noise distribution that is uniformly distributed over all responses. The differences between TCC and the IMM lie in how these ideas are formalized. These differences pertain to three decisions: (a) The shape of the signal distribution over the response scale: Whereas the IMM assumes a von-Mises distribution, TCC assumes a distribution best approximated by a Laplace distribution (i.e., a distribution that falls off exponentially in both directions from its mean). (b) The response-selection rule: Whereas IMM applies a simple version of Luce's choice rule, normalizing the activation distribution to obtain the probability of each response, TCC uses a signal-detection framework. (c) Whereas in IMM, the dispersion of the signal distribution is interpreted as the precision of feature memory, and treated as a free parameter, in TCC, this parameter is interpreted as a scaling parameter translating feature differences into similarities, which is fixed by a separate psychophysical measurement.

Figure 1

In the TCC Model, the Activation of Each Possible Feature Value That Could be Chosen as Response (Bottom Panel) is the Sum of a Signal (Top Panel), and Normally Distributed Noise (Middle Panel)



Note. TCC = target confusability competition; The signal peaks at the target feature, scaled by d', and drops off approximately exponentially for other features as a function of their distance from the target feature; the steepness of that drop is governed by the rate parameter κ. See the online article for the color version of this figure.

An additional difference is that in the IMM the signal distribution is a weighted mixture of distributions centered on each feature in the array. The TCC as developed so far ignores the nontarget features and considers only the signal distribution centered on the target feature. However, Schurgin et al. (2020) write that the TCC model could be extended to incorporate nontarget signals in a straightforward manner. To simplify the comparison between IMM and TCC, I will initially focus on the simplest versions of both models, which ignore the nontarget items, and return to an extension to nontargets in the Discussion. The simplest version of the IMM considered here is mathematically equivalent to the two-parameter mixture model (Zhang & Luck, 2008).

Formally, these simple versions of both measurement models can be described by two general equations, one for the activation function that describes the memory signal $S(x)$ as a distribution of activation of feature values, and one for the response-selection function that translates the memory signal distribution into a probability distribution over possible responses:

$$S(x) = f_{act}(x - \theta);$$
$$P(x) = f_{response}(\mathbf{S}, x, noise). \quad (1)$$

Here, x is the response, θ is the target, and their difference reflects their distance on the response scale (for circular response scales, the angular distance).[1] For the IMM, the activation function is a von-Mises distribution function, vM, with mean θ and precision κ, whereas for the TCC, it is a Laplace distribution function, L, with mean θ and rate κ. In both cases, these functions are multiplied with a memory-strength parameter (c for the IMM, and d' for the TCC):

$$S_{IMM}(x) = c \cdot vM(x; \theta, \kappa) = c\frac{\exp(\kappa\cos(x-\theta))}{2\pi I_0(\kappa)};$$
$$S_{TCC}(x) = d' \cdot L(x; \theta, \kappa) = d'\frac{1}{2}\kappa\exp(-\kappa|x-\theta|). \quad (2)$$

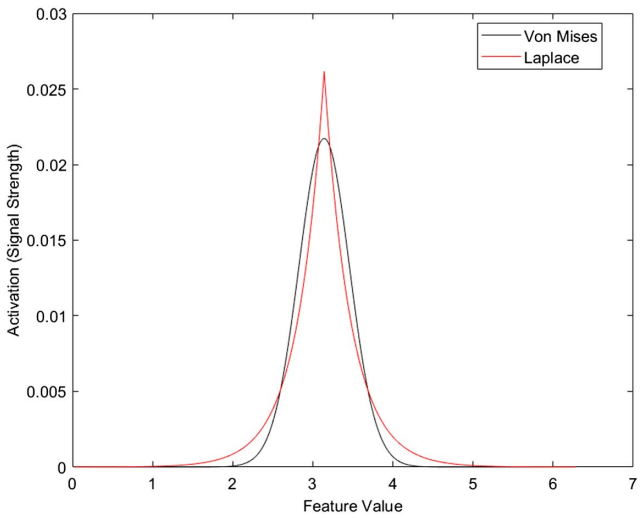
$I_0(\kappa)$ is the modified Bessel function of order 0. The shapes of the two functions are shown in Figure 2. The difference between the two activation functions can be understood as a difference in generalization gradients of activation over angular distance. Alternatively, the two activation functions can be understood as the same (exponential) generalization gradient on different scales for angular distance: The Laplace function represents the exponential generalization function on angular difference, whereas the von-Mises represents the exponential generalization function on distance measured as $1 - \cos(x - \theta)$, which is a measure of distance between two angles often used in circular statistics (Mardia & Jupp, 2000).[2] With this interpretation, the choice is on which distance measure is psychologically more adequate.

The response-selection function adds uniformly distributed noise to the signal distribution, and transforms the resulting activation distribution into a probability distribution that can be used as the likelihood distribution over possible responses. For the IMM, adding uniform noise consists of adding a constant to the signal, and the response-selection function is a simple Luce choice rule; for TCC it is a signal-detection rule, with noise drawn from a standard normal distribution:

$$P_{IMM}(x) = \frac{S(x)+1}{\sum(S(x)+1)};$$
$$P_{TCC}(x) = P(\arg\max(\mathbf{S}+\varepsilon) = x),$$
$$\text{with } \varepsilon \sim N(0, \mathbf{I}). \quad (3)$$

Figure 2

Illustration of the von-Mises Distribution ($\kappa = 10$) and the Laplace Distribution ($\kappa = 3$), Both Centered on π as Their Mean



Note. See the online article for the color version of this figure.

The two response-selection equations are more closely related than they appear, for the following reason. The signal-detection rule is a Thurstonian decision process of Type V, that is, a decision process that selects the item with the largest value after adding noise to each item's true value, where the noise has equal variance and covariance for all items. In signal-detection theory (SDT) the noise is assumed to be normally distributed, but there is no strong reason to choose this distribution. A Thurstonian process of Type V with noise from a Gumbel distribution (a.k.a. Double Exponential) is equivalent to a version of Luce's choice rule in which the signals are transformed by an exponential function (Luce, 1994; Yellott, 1977).[3] Hence, if we replace the Normal by the Gumbel in the TCC, we obtain:

$$P_{TCC}(x) = P(\arg\max(\mathbf{S}+\varepsilon) = x) = \frac{\exp(S(x))}{\sum\exp(S(x))};$$
$$P_{IMM}(x) = \frac{S(x)+1}{\sum(S(x)+1)} = P(\arg\max(\log(\mathbf{S})+\varepsilon) = x),$$
$$\text{with } \varepsilon \sim \mathcal{G}(0, \mathbf{I}). \quad (4)$$

The right-hand side of the P_{TCC} equation is equivalent to the IMM response-selection rule after submitting the activation distribution to an exponential transformation. The uniform noise term drops out because $\exp(S + 1) = \exp(S) \cdot \exp(1)$, and $\exp(1)$ is factored into the numerator and the denominator, canceling out. The uniform

[1] Following other models of continuous reproduction, including the TCC, I approximate the continuous response scale by a scale of 360 discrete response options one degree apart, so that x stands for one of the 360 possible responses. Assuming a memory signal distribution over a truly continuous feature scale, we can think of $S(x)$ as the integral of that signal distribution over the 1° interval of response option x.

[2] With C for the normalization constant $2\pi I_0(\kappa)$, we can write the von-Mises distribution function as $\frac{1}{C}\exp[\kappa\cos(x-\theta)] = \frac{1}{C}\exp[\kappa - \kappa(1-\cos(x-\theta))] = \frac{1}{C}\exp(\kappa)\exp[-\kappa(1-\cos(x-\theta))] = \frac{1}{C\exp(-\kappa)}\exp[-\kappa(1-\cos(x-\theta))]$.

[3] Luce (1994) attributes this insight to work by Holman and Marley in the early 1960s. I am grateful to Phil Smith for alerting me to the history of this equivalence.

background noise in the IMM is now represented by the standard Gumbel noise that is added to the signal value of each response option x. To conclude, the difference between the two response-selection functions can be eliminated by two steps: (a) exchange the normal noise by Gumbel noise in the TCC and (b) transform the activation values exponentially in the IMM.

One further difference between the IMM and the TCC is that in the IMM the steepness of the activation function is interpreted as reflecting the precision of feature memory, and therefore is treated as a free parameter (i.e., the precision parameter κ of the von-Mises distribution). By contrast, the authors of TCC regard it as a fixed property of the stimulus space, and therefore estimate it from psychophysical scaling data. This difference comes down to two questions: First, can the precision (or rate) parameter κ in the activation function be fixed to a constant across all experimental manipulations, unless the manipulations affect perception? If yes, the second question is whether this constant value can be estimated by psychophysical measurements on a perceptual task not involving memory. Here I investigate only the first question.

A Factorial Model Comparison

Taken together, we have three dimensions on which the two measurement models differ: (a) the choice of the activation function, (b) the treatment of κ as a free or fixed parameter, (c) the response-selection rule, with three options: (a) Luce's choice rule on the untransformed activation distribution, as in the IMM; (b) Luce's choice rule on the exponentially transformed activation distribution, which is equivalent to the signal-detection rule of TCC with Gumbel noise, and (c) the signal-detection rule with normal noise, as in TCC. The full combinations of these three decisions yield $2 \times 2 \times 3 = 12$ possible models.

I carried out a combinatorial model comparison of those 12 models, fitting them to eight data sets from continuous-response experiments. My aim is to identify which of the 12 measurement models fits the data best, and how the three dimensions on which the IMM and the TCC model differ contribute to differences in model fit. The data sets are the first three experiments from Oberauer and Lin (2017), as well as the five experiments testing visual WM reported by Schurgin et al. (2020). Table 1 gives an overview of the experimental manipulations in these experiments. I chose these experiments because together they cover most of the experimental variations investigated in continuous-reproduction experiments.

Table 1
Experiments Modeled

Source	Independent variables
Oberauer & Lin (2017), Experiment 1	Set size (1–8)
Oberauer & Lin (2017), Experiment 2	Set size (1–8) and pre-cue (neutral, valid, invalid)
Oberauer & Lin (2017), Experiment 3	Set size (1–8) and retro-cue (neutral, valid, invalid)
Schurgin et al. (2020), Experiment 1	Set size (1, 3, 6, or 8)
Schurgin et al. (2020), Experiment 2	Set size (1, 3, or 6) and retention interval (1, 3, or 5 s)
Schurgin et al. (2020), Experiment 3	Set size (1, 3, or 6) and encoding time (0.1, 0.5, or 1.5 s)
Schurgin et al. (2020), Experiment 4	Number of response alternatives (2, 8, 60, 360)
Schurgin et al. (2020), Experiment 5	Set size (2 or 5)

Note. The experiments were not numbered in Schurgin et al. (2020), I gave them numbers for ease of reference here.

First, I fitted the models separately to data from each participant in each experimental design cell, allowing both model parameters to vary freely between conditions. Second, I fitted the models to the data from each participant, but jointly to all experimental design cells of each experiment, allowing only the memory-strength parameter (c or d') to differ between conditions, while fixing κ to be equal for all conditions. The comparison of model fits with κ free to differ between conditions and those with κ fixed to be equal provides a test of the assumption of Schurgin and colleagues that κ is invariant across manipulations that affect memory. I fit the models by minimizing the deviance (i.e., -2 times the log-likelihood); the deviance of the model for a given data set is the sum of the deviances across all design cells from all participants. For the TCC model I used the version with uncorrelated noise, because it is computationally faster, and Schurgin et al. (2020) stated that it fit their data as well as the version with correlated noise. All modeling code (written in Matlab) and the raw data modeled are available on the Open Science Framework (OSF): https://osf.io/zwprv/

Table 2 shows the relative fit for all models with κ free to vary across conditions; Table 3 shows the relative fits for the same

Table 2
Relative Model Fits of the Six Models With κ Free to Vary Across Conditions: Difference of Each Model's Deviance From the Best-Fitting Model

Signal distribution	von-Mises	Laplace	von-Mises	Laplace	von-Mises	Laplace
Response rule	Luce	Luce	SDT (Gumbel)	SDT (Gumbel)	SDT (Normal)	SDT (Normal)
O&L Experiment 1	74	159	0	418	6	285
O&L Experiment 2	288	0	1	319	58	68
O&L Experiment 3	179	100	0	455	34	209
SWB Experiment 1	53	114	0	259	11	145
SWB Experiment 2	87	364	0	787	5	474
SWB Experiment 3	109	507	0	1,056	1	608
SWB Experiment 4	15	71	2	185	0	113
SWB Experiment 5	60	131	0	330	1	203

Note. SDT = signal-detection theory; O&L = Oberauer and Lin (2017); SWB = Schurgin et al. (2020).

Table 3

Relative Model Fits of the Six Models With κ Fixed Across Conditions: Difference of Each Model's Deviance From the Best-Fitting Model

Signal distribution	von-Mises	Laplace	von-Mises	Laplace	von-Mises	Laplace
Response rule	Luce	Luce	SDT (Gumbel)	SDT (Gumbel)	SDT (Normal)	SDT (Normal)
O&L Experiment 1	1930	1,227	0	428	13	292
O&L Experiment 2	1945	962	0	280	64	88
O&L Experiment 3	5,326	3,814	0	417	24	226
SWB Experiment 1	565	580	0	255	27	137
SWB Experiment 2	2,145	1,354	0	736	45	448
SWB Experiment 3	1,313	1,510	0	1,055	34	658
SWB Experiment 4	45	56	0	110	4	83
SWB Experiment 5	292	386	0	309	7	200

Note. SDT = signal-detection theory; O&L = Oberauer and Lin (2017); SWB = Schurgin et al. (2020).

models with κ fixed to be the same across conditions. The model in the first column in each table is the original IMM, and the model in the last column is the original TCC model. Because deviances are interpretable only as indicators of comparative model fit—their absolute values depend on the number of observations and the measurement scale, and are therefore not meaningful indicators of fit—both tables report the difference of each model's deviance to the deviance of the best-fitting model.

Across 10 of the 12 model comparisons, the model version with a von-Mises activation function, combined with an SDT decision rule with a Gumbel noise distribution had the best fit; in the remaining two cases it came second with a negligible difference to the winner on the deviance scale. That said, the fits of the corresponding model versions with Gumbel noise and with normal noise were always very close together, indicating that the choice of error distribution (Gumbel or Normal) does not matter much.

I next address the question whether the precision parameter κ can be fixed to be equal across experimental conditions. Table 4 presents the results of a comparison of each model version with κ free to vary across condition to the corresponding model version with κ fixed. I used AIC and BIC as model-fit measures, which compensate for the differences in model flexibility that arises from the different numbers of free parameters. Table 4 presents the differences in AIC and BIC between each model version with κ fixed versus free. For the models using the simple Luce choice rule (as in the IMM), AIC and BIC were mostly smaller—indicating a better fit—for the model version in which κ was free to vary across conditions. By contrast, for the model versions using an SDT response-selection function, the model comparisons were unanimously in favor of models with κ fixed across conditions. This result is consistent with the claim by Schurgin et al. (2020) that memory precision is a constant.

Table 4

Relative Model Fits of Models With κ Free Versus Fixed Across Conditions. First Entries in Each Cell are Differences in AIC, and Second Entries, Differences in BIC

Signal distribution	von-Mises	Laplace	von-Mises	Laplace	von-Mises	Laplace
Response rule	Luce	Luce	SDT (Gumbel)	SDT (Gumbel)	SDT (normal)	SDT (normal)
O&L Experiment 1	1,763	976	−93	−83	−86	−87
	1,140	353	−716	−707	−709	−710
O&L Experiment 2	1,516	822	−141	−179	−137	−120
	350	−344	−1307	−1345	−1301	−1286
O&L Experiment 3	5,074	3,640	−73	−111	−83	−56
	3,864	2,430	−1284	−1322	−1294	−1267
SWB Experiment 1	488	443	−24	−27	−8	−32
	257	213	−254	−258	−238	−262
SWB Experiment 2	1998	930	−60	−112	−20	−86
	1,268	200	−790	−842	−750	−816
SWB Experiment 3	1,121	919	−83	−83	−50	−34
	383	181	−821	−822	−788	−772
SWB Experiment 4	−167	−213	−200	−273	−194	−230
	−682	−727	−714	−788	−708	−743
SWB Experiment 5	207	230	−25	−46	−19	−29
	39	61	−193	−214	−188	−197

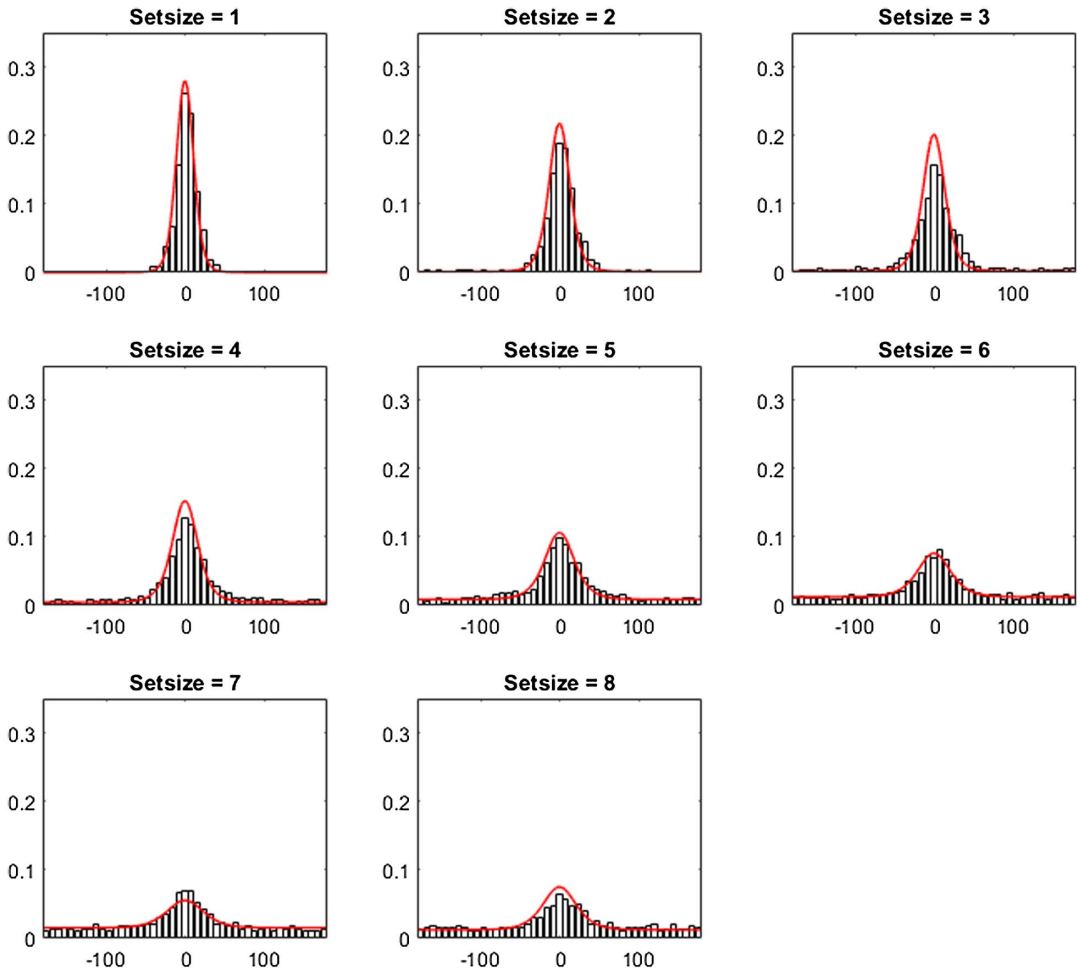
Note. AIC = akaike information criteroin; BIC = bayesian information criterion; O&L = Oberauer and Lin (2017); SWB = Schurgin et al. (2020); SDT = signal-detection theory; Differences are calculated by subtracting the fit index of the free κ version from that of the fixed κ version. Therefore, positive differences reflect a smaller AIC or BIC—and hence a better fit—for the model with κ free to vary across conditions.

Figures 3–5 show model fits of the best-fitting model (von-Mises activation function, SDT response-selection rule, Gumbel noise, κ fixed across conditions) to the error distributions for three experiments, selected to represent a broad range of experimental conditions. Figures 6–8 show parameter estimates from the six model versions in which κ was free to vary, for the same three experiments. I plotted only model versions with von-Mises activation function because their fit was usually superior to those with a Laplace function. In all model versions, the memory-strength parameters, c or d', declined with variables that make memory more difficult (i.e., larger set size, longer retention interval, shorter encoding time). In the model versions using the simple Luce choice rule, the precision parameter κ also declined with set size up to about four items. By contrast, in the models using an SDT response-selection rule (as in the TCC model), the behavior of κ was less systematic, and any influence of experimental conditions was small, consistent with the finding that κ is best fixed across conditions in these models.

The reason why κ can be considered a constant in the SDT-based models can be appreciated with the help of Figure 9. Remember that, with a Gumbel noise distribution, the SDT response-selection function is equivalent to Luce's choice rule on exp(S). Applying the exp function to the von-Mises distributed signal S makes it narrower, and it does so more strongly the higher the memory strength (c or d'; right panel of Figure 9). In this way, increasing memory strength (as is typically found with smaller set size) renders the predicted error distribution narrower to a stronger degree than in the untransformed signal (left panel of Figure 9). This is enough to accommodate the observed increase of precision in easier conditions (in particular, smaller set sizes), so that no increase of κ is necessary to accommodate that finding.

Figure 3

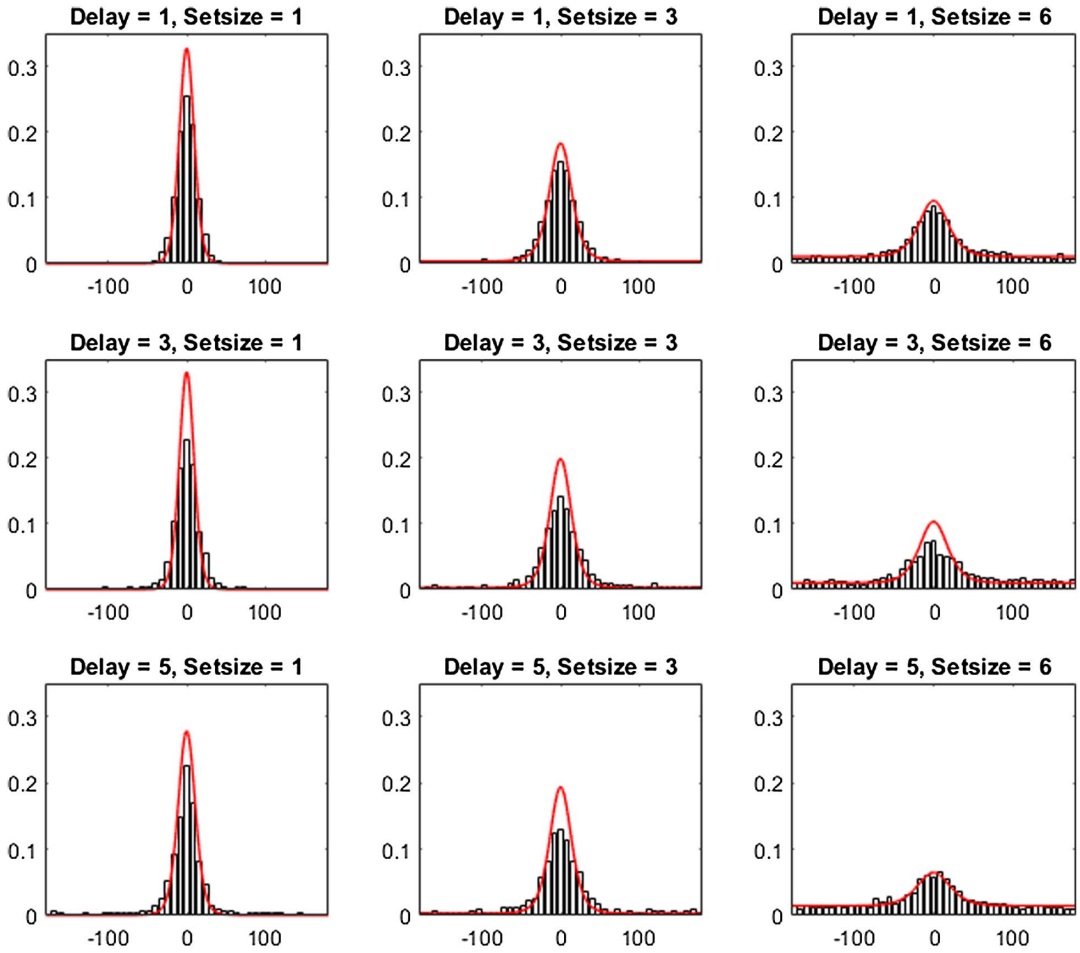
Error Distribution Histograms and Model Predictions of the Best-Fitting Model (von-Mises, SDT, Gumbel, κ Constrained to be Equal Across Conditions) Overlaid as Red Lines, for Experiment 1 (Oberauer & Lin, 2017, Experiment 1)



Note. SDT = signal-detection theory. See the online article for the color version of this figure.

Figure 4

Error Distribution Histograms and Model Predictions of the Best-Fitting Model (von-Mises, SDT, Gumbel, κ Constrained to be Equal Across Conditions) Overlaid as Red Lines, for Experiment 5 (Schurgin et al., 2020, Experiment 2), Varying the Retention Interval (Delay, in s) and Set Size



Note. SDT = signal-detection theory. See the online article for the color version of this figure.

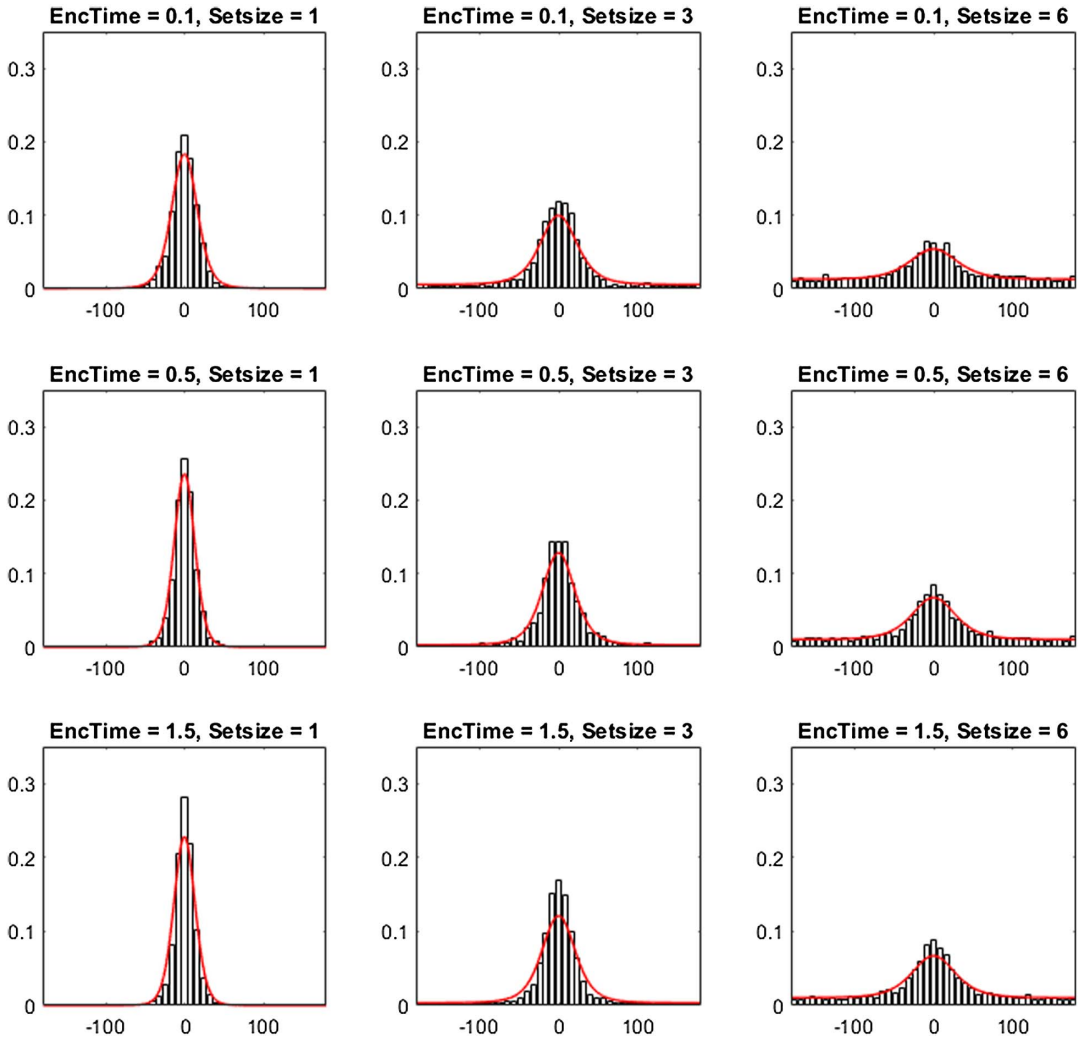
That said, before treating κ as a constant, we need to consider not only whether it varies across experimental conditions but also whether it varies between people. To assess individual differences, I focused on the best-fitting model version—the one with a von-Mises activation function combined with a SDT response-selection rule with Gumbel noise distribution—and compared the fit when κ is estimated freely for each person (but fixed across conditions) to one where κ was set to the mean, or to the median, of parameter estimates over participants in each experiment. The resulting AIC and BIC differences can be found in Table 5. Whereas by AIC, the model version allowing individual differences in κ was consistently superior, the results were more variable in light of BIC, which adds a larger penalty for having as many free κ parameters as participants.

Discussion

By introducing the TCC, Schurgin et al. (2020) broadened the space of measurement models considered for visual WM. Here I laid out this space explicitly, and explored the full combination of the three dimensions on which these models differ: The choice of the activation function (von-Mises or Laplace), the choice of response-selection rule (a simple Luce choice rule, an SDT with Gumbel noise distribution, and an SDT with normal noise distribution), and whether precision is considered a free or a fixed parameter. The IMM (Oberauer et al., 2017) and the mixture models (Bays et al., 2009; Zhang & Luck, 2008), which are mathematically equivalent to a constrained version of the IMM, inhabit one corner of this cube, whereas TCC inhabits the opposite corner.

Figure 5

Error Distribution Histograms and Model Predictions of the Best-Fitting Model (von-Mises, SDT, Gumbel, κ Constrained to be Equal Across Conditions) Overlaid as Red Lines, for Experiment 6 (Schurgin et al., 2020, Experiment 3), Varying the Encoding Time (EncTime, in s) and Set Size



Note. SDT = signal-detection theory. See the online article for the color version of this figure.

Across eight experiments, the model version fitting the data best was one not considered so far: A combination of a von-Mises activation function with an SDT response-selection rule. The choice of the SDT noise distribution appeared to matter very little. For practical reasons, I recommend using the Gumbel noise distribution because it makes the SDT model equivalent to one using Luce's choice rule on signal values after exponential transformation. When formalized in that way, fitting the model is faster by about two orders of magnitude than when using the SDT with normal noise. It will also be easier to implement that model in a Bayesian hierarchical framework, because we only need to replace $S(x) + 1$ by $\exp(S(x))$ in the Bayesian hierarchical IMM (Oberauer et al., 2017). For the same reason, expanding the model to include nontarget intrusions is straightforward because they are already included in the full IMM. Because of its usefulness, this model should have a name, and I propose to call it the signal-discrimination model (SDM).

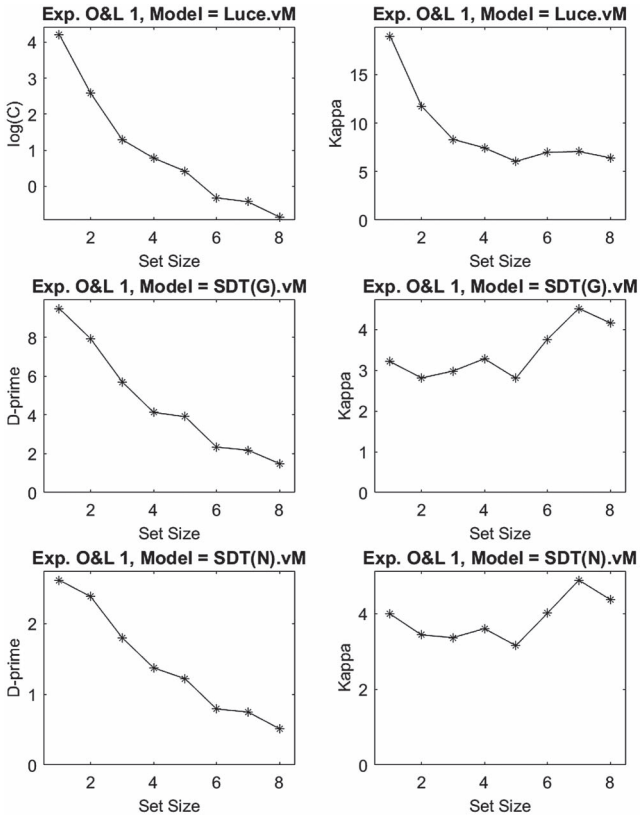
The full SDM, including the contribution of nontarget information to responses, is given by the following equations:

$$\begin{aligned} C(x) &= c\sum_{i}^{n}\exp(-s|y_i - y_\theta|)vM(x; x_i, \kappa); \\ A(x) &= a\sum_{i}^{n} vM(x; x_i, \kappa); \\ S(x) &= C(x) + A(x); \\ P(x) &= \frac{\exp(S(x))}{\sum\exp(S(x))}. \end{aligned} \tag{5}$$

In the full SDM, the signal $S(x)$ is the weighted sum of two components. The first, $C(x)$, is the activation of feature value x in response to a retrieval cue at the context of the target, y_θ. It is a

Figure 6

Median Parameter Estimates of the Model Versions With κ Free to Vary, Experiment 1 of Oberauer and Lin (2017), Abbreviated as O&L 1



Note. IMM = interference measurement model; The models are coded by their response-selection rule: Luce for the simple Luce rule as used in the IMM; SDT for signal-detection theory, with G for Gumbel noise distribution, and N for normal noise distribution. All models use a von-Mises (vM) activation function.

weighted sum of von-Mises distributions, one each centered on the true feature value x_i for each array item i in a set of n items. Each von-Mises distribution is weighted by an exponential function of their distance to the target on the context dimension (usually, the spatial location in the array). The steepness of that exponential function is governed by s, which represents the precision of memory on the context dimension. With this parameter the SDM can account for the fact that nontarget intrusions often come predominantly from nontargets close to the target on the context dimension (Bays, 2016a; Rerko et al., 2014). The second component, $A(x)$, is an unweighted sum of von-Mises distributions, one for each array item, and represents the degree of persistent activation, or reactivation, of all features in the array independent of the retrieval cue. The signal $S(x)$ is transformed exponentially before applying Luce's choice rule to obtain the probability of choosing response x. The full SDM has four free parameters: The weights of the two components, c and a, and the precision parameters on the context dimension and the feature dimension, s and κ, respectively.[4] The reduced version of the SDM considered above is obtained by setting $a = 0$ and s to a very high value, so that the component $A(x)$ drops out, and the von-Mises distributions centered on nontargets in component $C(x)$ are effectively given a weight of zero.

The parameters of the SDM have straightforward theoretical interpretations: The weight of the cue-based component, c, is the strength of binding between the target feature and the context that is used as a cue, and by implication, the degree to which the target is reactivated by the cue at retrieval. The weight of the cue-independent component, a, is the degree of activation of all array items at retrieval, which can be thought of as the strength of item memory, that is, memory for which items have been presented in the current array, regardless of their relation to each other or to a context. Both strength parameters are expressed relative to the standard deviation of the standard Gumbel distribution that characterizes the background noise added to the signal strength for each response option. The parameter s reflects the precision of memory representations on the context dimension, whereas κ reflects the precision of memory representations on the target-feature dimension.

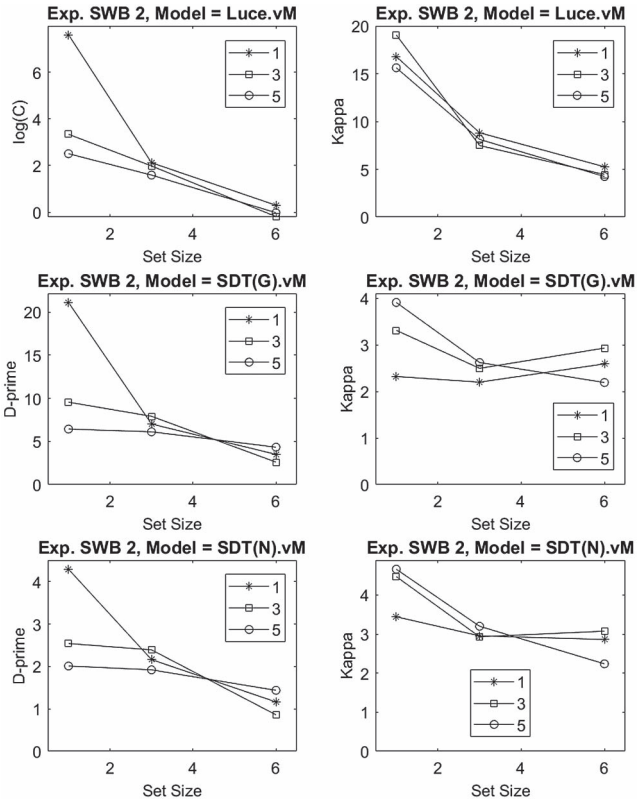
In addition to its empirical adequacy, the choice of a measurement model is often also motivated by the researchers' theoretical assumptions and their corresponding measurement aims. The mixture models, for instance, are inspired by the assumption of a discrete, slot-like capacity limit of visual WM, which entails two qualitatively different states of memory: Having a representation of a visual object with a certain precision, or having none at all. The mixture models offer an estimate of the proportions of these states, and thereby, of a person's capacity. By contrast, the IMM and the TCC are motivated by the assumption that every item of a memory set is encoded into visual WM with a degree of strength that varies on a continuous dimension. Whereas the IMM in its simple version (without nontarget intrusions) is equivalent to the Zhang and Luck (2008) mixture model (Oberauer et al., 2017), so that the choice between those two measurement models could not be informed by their relative empirical adequacy, this is not the case for the TCC, and also not for the SDM. Therefore, the present finding that the SDM (and the TCC) fit the data better than the IMM can be understood as evidence in favor of a memory model in which the memory strength of each item varies continuously, as opposed to a model assuming a dichotomy between in-memory states and zero-information states.

The finding that precision can be treated as a constant across experimental manipulations intended to affect memory is consistent with the possibility that it reflects a characteristic of the stimulus space (i.e., of how we perceive features on the relevant dimension), rather than of memory. If that is the case, then κ should be subject to selective influence by manipulations affecting perception. For instance, varying the contrast of Gabor patches (Bays, 2016b) or the thickness of ovals (Yoo et al., 2020) could be used to vary the perceptual precision of orientation in a continuous-reproduction task. Such a manipulation should selectively influence κ in the SDM without affecting other parameters.

[4] The parameter c in this formalization is the d' parameter in the SDT formalization of the model.

Figure 7

Median Parameter Estimates of the Model Versions With κ Free to Vary, Experiment 2 of Schurgin et al. (2020) Abbreviated as SWB 2



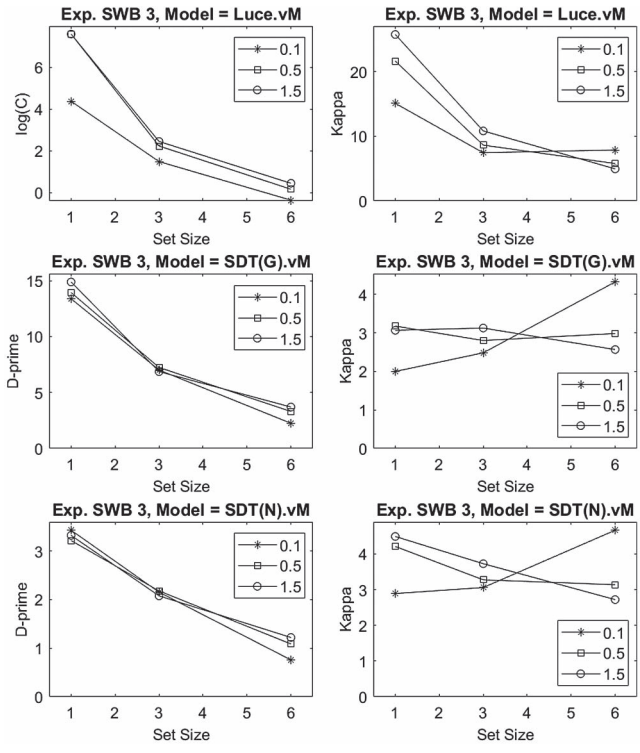
Note. The parameter distinguishing the plot lines is the duration of the retention interval (in s). For details see the legend of Figure 6.

For the purposes of a measurement model, treating κ as a constant is attractive because it enables researchers to estimate a single parameter, memory strength (c, or d'), as a measure of a person's visual-WM capacity, or of an experimental condition's difficulty. In light of the present individual-differences analysis, I advise against doing this when the model is used to measure the capacity of an individual, or capacity differences between groups, because it is more likely than not that there are individual differences in precision in addition to those in memory strength. A Bayesian hierarchical implementation of the SDM opens the possibility to incorporate both the knowledge of the relative constancy of κ across conditions, and the flexibility for estimating different κ parameters for different individuals (Lee & Vanpaemel, 2018): On the population level, we can set a very informative (i.e., narrow) prior for the population mean of κ, informed by the distribution of estimates from previous model fits. Based on the fits to the eight experiments considered here, a prior with a mean of 3.5 and a standard deviation of 1 would be a good starting point for that prior. On the level of individuals, each κ_j would be drawn from a distribution with the population mean, and a population standard deviation treated as a further free parameter.

The space of measurement models laid out here integrates several apparently incommensurable models. It is not meant to limit the space for further developments. The models considered here are deliberately held very simple, and therefore ignore potentially informative aspects of the data. For instance, continuous reproduction of visual features such as color have been shown to reflect biases toward the centers of categories (Bae et al., 2015), and Hardman et al. (2017) have proposed a measurement model that combines memory for categories and for continuous feature values. The perhaps most significant limitation of the present measurement models is that they account only for the distribution of responses, neglecting the distribution of response times. Recently, Smith (2016) and Ratcliff (2018) have extended the diffusion model of binary choices to response selection on continuous circular response scales; these models also potentially distinguish categorical and continuous information (Smith et al., 2020). These developments point to avenues for potential extensions of the SDM, through which it would be able to exploit a richer set of data to measure the strength and precision of memory representations.

Figure 8

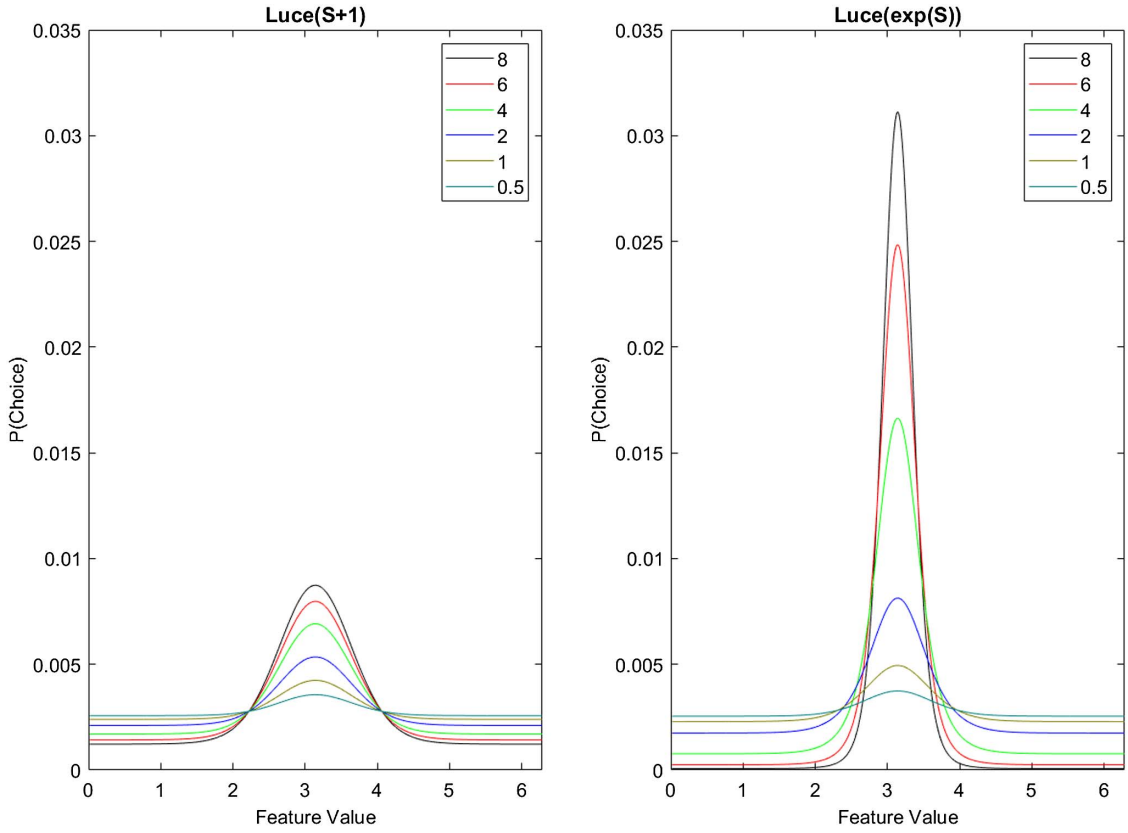
Median Parameter Estimates of the Model Versions With κ Free to Vary, Experiment 3 of Schurgin et al. (2020) Abbreviated as SWB 3



Note. The parameter distinguishing the plot lines is the encoding duration in seconds. For details see the legend of Figure 6.

Figure 9

Likelihood Function for Models With Luce's Choice Rule on von-Mises Distributed Signals Without (Left) and With Applying the exp Function (Right)



Note. The line parameter is the memory-strength parameter. See the online article for the color version of this figure.

Table 5

Relative Model Fits for the von-Mises SDT (Gumbel) Model With κ Constrained to be Equal Across Conditions, Compared to κ Set to the Mean, or the Median, Across Participants in Each Experiment

Experiment	Set κ to mean	Set κ to median
O&L Experiment 1	34	18
	−55	−71
O&L Experiment 2	98	115
	8	25
O&L Experiment 3	44	49
	−49	−44
SWB Experiment 1	52	45
	−25	−31
SWB Experiment 2	161	119
	70	27
SWB Experiment 3	111	124
	19	32
SWB Experiment 4	46	16
	−125	−156
SWB Experiment 5	322	95
	154	−73

Note. AIC = akaike information criteroin; BIC = bayesian information criterion; O&L = Oberauer and Lin (2017); SWB = Schurgin et al. (2020); SDT = signal-detection theory. First entries in each cell are differences in AIC, and Second entries, differences in BIC. Differences are calculated by subtracting the fit index of the version with κ free to vary between participants from that of the constant κ version. Therefore, positive differences reflect a smaller AIC or BIC—and hence a better fit—for the model with κ free to vary across participants.

References

Bae, G. Y., Olkkonen, M., Allred, S. R., & Flombaum, J. I. (2015). Why some colors appear more memorable than others: A model combining categories and particulars in color working memory. *Journal of Experimental Psychology: General*, *144*(4), 744–763. https://doi.org/10.1037/xge0000076

Bays, P. M. (2016a). Evaluating and excluding swap errors in analogue tests of working memory. *Scientific Reports*, *6*, Article 19203. https://doi.org/10.1038/srep19203

Bays, P. M. (2016b). A signature of neural coding at human perceptual limits. *Journal of Vision (Charlottesville, Va.)*, *16*(11), Article 4. https://doi.org/10.1167/16.11.4

Bays, P. M., Catalao, R. F. G., & Husain, M. (2009). The precision of visual working memory is set by allocation of a shared resource. *Journal of Vision (Charlottesville, Va.)*, *9*(10), Article 7 https://doi.org/10.1167/9.10.7

Blake, R., Cepeda, N. J., & Hiris, E. (1997). Memory for visual motion. *Journal of Experimental Psychology: Human Perception and Performance*, *23*(2), 353–369. https://doi.org/10.1037/0096-1523.23.2.353

Hardman, K. O., Vergauwe, E., & Ricker, T. J. (2017). Categorical working memory representations are used in delayed estimation of continuous colors. *Journal of Experimental Psychology: Human Perception and Performance*, *43*(1), 30–54. https://doi.org/10.1037/xhp0000290

Lee, M. D., & Vanpaemel, W. (2018). Determining informative priors for cognitive models. *Psychonomic Bulletin & Review*, *25*(1), 114–127. https://doi.org/10.3758/s13423-017-1238-3

Luce, R. D. (1994). Thurstone and sensory scaling: Then and now. *Psychological Review*, *101*(2), 271–277. https://doi.org/10.1037/0033-295X.101.2.271

Mardia, K. V., & Jupp, P. E. (2000). *Directional statistics*. Wiley.

Oberauer, K. (2021). Measurement models for visual working memory—A factorial model comparison. *PsyArXiv*. https://doi.org/10.17605/OSF.IO/ZWPRV

Oberauer, K., & Lin, H.-Y. (2017). An interference model of visual working memory. *Psychological Review*, *124*(1), 21–59. https://doi.org/10.1037/rev0000044

Oberauer, K., Stoneking, C., Wabersich, D., & Lin, H.-Y. (2017). Hierarchical Bayesian measurement models for continuous reproduction of visual features from working memory. *Journal of Vision (Charlottesville, Va.)*, *17*(5), Article 11. https://doi.org/10.1167/17.5.11

Ratcliff, R. (2018). Decision making on spatially continuous scales. *Psychological Review*, *125*(6), 888–935. https://doi.org/10.1037/rev0000117

Rerko, L., Oberauer, K., & Lin, H.-Y. (2014). Spatially imprecise representations in working memory. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, *67*(1), 3–15. https://doi.org/10.1080/17470218.2013.789543

Schneegans, S., & Bays, P. M. (2017). Neural architecture for feature binding in visual working memory. *The Journal of Neuroscience*, *37*(14), 3913–3925. https://doi.org/10.1523/JNEUROSCI.3493-16.2017

Schneegans, S., Taylor, R., & Bays, P. M. (2020). Stochastic sampling provides a unifying account of visual working memory limits. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(34), 20959–20968. https://doi.org/10.1073/pnas.2004306117

Schurgin, M. W., Wixted, J. T., & Brady, T. F. (2020). Psychophysical scaling reveals a unified theory of visual memory strength. *Nature Human Behaviour*, *4*(11), 1156–1172. https://doi.org/10.1038/s41562-020-00938-0

Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science*, *237*(4820), 1317–1323. https://doi.org/10.1126/science.3629243

Smith, P. L. (2016). Diffusion theory of decision making in continuous report. *Psychological Review*, *123*(4), 425–451. https://doi.org/10.1037/rev0000023

Smith, P. L., Saber, S., Corbett, E. A., & Lilburn, S. D. (2020). Modeling continuous outcome color decisions with the circular diffusion model: Metric and categorical properties. *Psychological Review*, *127*(4), 562–590. https://doi.org/10.1037/rev0000185

van den Berg, R., Awh, E., & Ma, W. J. (2014). Factorial comparison of working memory models. *Psychological Review*, *121*(1), 124–149. https://doi.org/10.1037/a0035234

van den Berg, R., & Ma, W. J. (2018). A resource-rational theory of set size effects in human visual working memory. *eLIFE*, *7*, Article e34963. https://doi.org/10.7554/eLife.34963

Wilken, P., & Ma, W. J. (2004). A detection theory account of change detection. *Journal of Vision (Charlottesville, Va.)*, *4*(12), Article 11 https://doi.org/10.1167/4.12.11

Yellott, J. I., Jr. (1977). The relationship between Luce's choice axiom, Thurstone's theory of comparative judgment, and the double exponential distribution. *Journal of Mathematical Psychology*, *15*(2), 109–144. https://doi.org/10.1016/0022-2496(77)90026-8

Yoo, A. H., Acerbi, L., & Ma, W. J. (2020). Uncertainty is maintained and used in working memory. *bioRxiv*. https://doi.org/10.1101/2020.10.06.328310

Zhang, W., & Luck, S. J. (2008). Discrete fixed-resolution representations in visual working memory. *Nature*, *453*(7192), 233–235. https://doi.org/10.1038/nature06860

Received January 20, 2021
Revision received July 24, 2021
Accepted August 12, 2021 ■