BRIEF REPORT

A model comparison approach reveals individual variation in the scope and control of attention

Daniel Schor[1] · **Alex S. Brodersen**[1] · **Bradley S. Gibson**[1]

Abstract

The maintenance capacity of working memory is known to be severely limited in scope. However, the reason this capacity varies across individuals remains unknown because it has proven difficult to estimate the maximum capacity of an individual's "scope of attention" (SoA) separate from their ability to achieve this maximum capacity due to temporary lapses in "attention control" (AC). The present study accomplished this separation by using a maximum likelihood framework to extract latent constructs representing SoA and AC from a whole-report version of the visual-array task. The results of two experiments ($N = 145$ and $N = 189$) showed that model fit was significantly greater when the model allowed both AC and SoA to vary across individuals relative to a model in which only AC was allowed to vary (and SoA was fixed). More importantly, the individual estimates of SoA and AC derived from this variable model suggested that (1) the observed range of SoA was found to be small across individuals, with 91% able to maintain a maximum of 3 or 4 items; (2) the consistency with which AC could be deployed was only weakly correlated with the magnitude of SoA; and (3) AC and SoA were both found to be significant predictors of fluid intelligence. Altogether, the present study clarified the nature of maintenance capacity and suggested that SoA and AC both need to be included in a mechanistic account of complex cognition.

Keywords Working memory capacity · Attention control · Scope of attention · Visual array task · Fluid intelligence · Latent variable model

A fundamental goal of cognitive science is to provide a mechanistic account of mental abilities. One common approach to this endeavor has been to specify how complex cognition arises from more primitive sources (Kyllonen & Christal, 1990). For the past several decades, working memory (WM) was considered to be one of the most important building blocks of higher-order cognition (Chuderski, Taraday, Nęcka, & Smoleń, 2012; Shipstead, Redick, Hicks, & Engle, 2012). In general, WM has been defined as a collection of processes responsible for temporarily maintaining the mental representations used in higher-level thought and action (Oberauer et al. 2018). As this definition suggests, however,

it has become increasingly clear that WM is itself a complex form of cognition whose functioning is determined by a collection of more basic mechanisms (Shipstead, Lindsey, Marshall, & Engle, 2014; Unsworth, Fukuda, Awh, & Vogel, 2014).

For instance, one of the benchmark properties of WM is that maintenance capacity appears to be severely limited in scope such that only a small set of approximately three to five items of information can be actively maintained at any given moment in time (Cowan, 2001; Oberauer et al., 2018). Two forms of attention have emerged as potential mechanisms to account for individual variation in this maintenance capacity. The "scope of attention" (SoA) reflects individual variation in the storage capacity of WM (Cowan et al., 2005), whereas, "attention control" (AC) reflects individual variation in the ability to engage on information that has been deemed relevant to current goals so as to prevent it from being replaced by internal or external sources of distraction (Engle, 2018; Shipstead, Harrison, & Engle, 2016). However, the extent to which individual variation in maintenance capacity reflects individual variation in SoA, AC, or both remains unresolved because it has proven difficult to estimate the maximum

---

✉ Bradley S. Gibson
bgibson@nd.edu

[1] Department of Psychology, 390 Corbett Hall, University of Notre Dame, Notre Dame, IN 46556, USA

capacity of an individual's SoA separate from their ability to achieve this maximum capacity due to temporary lapses in AC (see, e.g., Chuderski et al., 2012; Cowan et al., 2005; Luck & Vogel, 1997; Shipstead, Harrison, & Engle, 2015, for various perspectives).

Adam, Mance, Fukuda, and Vogel (2015) suggested an innovative approach to estimate the maximum capacity of an individual's SoA separate from temporary lapses in AC (see below for details). Of critical theoretical importance, they analyzed performance on a whole report version of the visual-array task (see Fig. 1) and found that a model in which SoA was fixed at three items was able to account for individual variation in maintenance capacity sufficiently well so long as the efficacy of AC was allowed to vary. The notion that individual variation in maintenance capacity can be accounted for by a variable AC and fixed SoA suggests that individual differences primarily reflect differences in the *consistency* with which individuals can focus attention to its maximum extent rather than differences in the *magnitude* of that extent. Such moment-to-moment fluctuations in the efficacy of AC are commonly known as "lapses of attention" (deBettencourt, Keene, Awh, & Vogel, 2019; Robison & Unsworth, 2019)

and may be driven by fluctuations in arousal/modulatory systems (Unsworth & Robison, 2017).

However, although Adam et al.'s (2015) model provided an innovative approach to estimate the maximum capacity of an individual's SoA separate from temporary lapses in AC, their conclusion that SoA was fixed across individuals cannot be stated with high confidence at the present point in time because it was based largely on the untested assumptions of their model. In particular, their conclusion suggesting that SoA was fixed across individuals was reached without evaluating the extent to which that model was superior to an alternative model in which the SoA and AC were both allowed to vary. This critical model comparison could not be performed because the parameter estimation procedure used in their model fitting process was not formalized in an appropriate statistical framework.

Accordingly, the present study measured the maintenance capacity of WM using two versions of a whole-report visual-array task and then extracted latent AC and SoA constructs by formalizing task performance within a maximum likelihood framework. Our unique formalization allowed us to compare Adam et al.'s (2015) "fixed model" with an alternative,
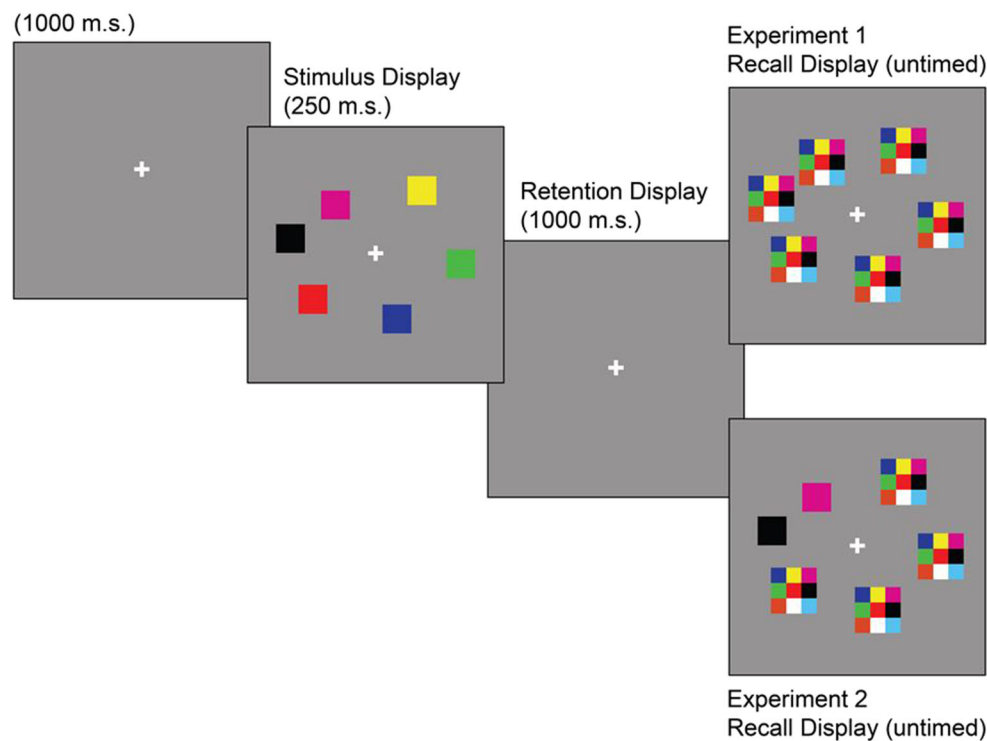

Fig. 1 A typical trial shown in the whole-report visual-array task used in Experiments 1 and 2. Each trial began with a fixation display (1,000 ms), followed by a stimulus display (250 ms) that consisted of six uniquely colored squares. The response display appeared following a retention interval (1,000 ms). The response display consisted of six response grids, each of which was centered at the location of a colored square shown in the preceding stimulus display. Each of the response grids consisted of a 3 × 3 array representing the nine possible colors that could have been shown on each trial. Participants were required to recall, in any order, the color of each of the six squares by clicking on the appropriate cell of the array. No feedback was provided. In Experiment 1, the response display did not visibly change after each recall response (upper response panel), whereas, in Experiment 2, the 3 × 3 array changed into the selected color following each recall response (lower response panel). The lower response panel shows the first two of six responses for the illustrated trial. (Color figure on line)

"variable model" in which AC and SoA were both allowed to vary across individuals. Providing an answer to this question is important because, if the variable model turns out to be the best-fitting model, then it will provide individual estimates of AC and SoA, which in turn can be used to clarify three additional important questions that have not been adequately addressed.

The first question concerned the characteristics of the SoA distribution. In particular, SoA is often described as ranging between three and five items (Oberauer et al., 2018), but are individuals who have a maximum capacity of five items just as likely as those who have a maximum capacity of three items? Or is the distribution of SoA dominated by one or two types of individuals? The second question concerned the relation between AC and SoA. In particular, do individuals who deploy AC more consistently also tend to have larger SoA, or are these two constructs largely unrelated? Finally, the third question concerned the extent to which AC and SoA each predict individual variation in other forms of complex cognition such as fluid intelligence (Gf), as measured here by the Raven's Advanced Progressive Matrices (RAPM) test. In particular, do both constructs make unique contributions to complex cognition, or is one construct more important than the other?

Method

Stimuli and procedure

Two versions of a whole-report visual-array task (Adam et al., 2015) were used to measure the maintenance capacity of WM across two experiments in this study. Differences between the two versions reflected the extent to which individual responses had to be maintained during recall (see Fig. 1 for details). The stimulus display consisted of six uniquely colored squares that were randomly selected without replacement on each trial from the following nine possible colors: red, green, blue, magenta, yellow, cyan, orange, white, and black. All stimuli appeared on a gray background. Each square subtended 0.92° of visual angle and appeared randomly in any location within the central 9.8° × 7.3° area of the computer screen, under the constraint that each square maintained a minimum distance of 2.83° of visual angle (center to center) from the other squares. Viewing distance (57 cm) was enforced with a chin rest. Participant responses were collected using the left button of a generic mouse. Participants completed 10 practice trials followed by 300 experimental trials, divided into 10 blocks of 30 trials. The dependent variable was the number of correct color–location pairs recalled on each trial (ranging from 0 to 6).

In addition, participants in Experiment 1 also completed the RAPM Set II items the day after they completed the whole-report visual-array task. After completing two practice problems from Set I, participants were instructed to complete the full exam, consisting of 36 problems. The participants were given a maximum of 1 hour to complete this self-paced exam. The primary outcome on this task was the number of correct responses (ranging from 0 to 36).

Participants

Justification for sample size is detailed in the Supporting Information. A total of 159 undergraduate students from the University of Notre Dame participated in Experiment 1, and a total of 197 undergraduate students from the same population participated in Experiment 2 (there was no overlap between the two experiments). All students participated in exchange for partial fulfillment of a course requirement or extra credit. All participants reported normal or corrected-to-normal vision, and all provided informed written consent prior to the beginning of the study.

Exclusion of participants

In the present study, an aberrant-response-detection procedure based on the cumulative sum of the mean-centered trial data was used to exclude participants who had greater than 20% of trials violate the tests of cumulative sum (Mason & Young, 2002). Although this procedure excluded participants based on the consistency of task performance, it was designed to distinguish between those who exhibited abrupt and lasting changes in performance over the course of the experiment versus those who had more temporary fluctuations in task performance (such as the two participants shown in Fig. 3, below). This procedure resulted in the exclusion of 10 participants in Experiment 1 and eight participants in Experiment 2. However, the conclusions reported below were identical regardless of whether this exclusion procedure was used or not.

Results

Performance on the whole-report visual-array task was formulated as follows:

$$Y_{ij} = Y_{ij}^* + \text{g}_{ij}, \tag{1}$$

where Y_{ij}^* is the number of color–location pairs that individual i remembered correctly on trial j, and g_{ij} is the residual term that represents the additional number of color–location pairs that this individual guessed correctly on this trial (see Supporting Information for full derivation of the distribution of guessing).

We assumed Y_{ij}^* followed a beta-binomial distribution; this distribution can be thought of as a hierarchical distribution

where Y_{ij}^* is a binomial random variable in which the number of possible items that could be maintained is specified by a parameter representing SoA, and the success probability associated with each of those outcomes follows a beta distribution that is specified by two parameters representing AC. More specifically, SoA was represented by the K_{max} parameter; this parameter was a discrete variable that ranged from 0 to 6 and represented the maximum number of color–location pairs that an individual is able to remember correctly at their peak capacity. In contrast, AC was represented by the α and β parameters. The α parameter was a continuous variable that may range from zero to infinity and represented the lapse rate of an individual. That is, the α parameter reflected the typical distribution of the number of correctly remembered items across all trials and determined the extent to which individuals performed at their peak capacity. Likewise, the β parameter was a continuous variable that also represented the lapse rate of an individual. However, as in Adam et al.'s (2015) original model, β was fixed to a value of 1 in the current model. Fixing β in this manner limited the distribution of full attention, partial lapses, and full lapses that could be captured by the model, but simplified the estimation procedure, which is described in more detail in the Supporting Information. Thus, individual variation in AC was captured exclusively by the α parameter in the current model.

Model comparison

A primary reason for the formalization of the statistical model was to evaluate the hypothesis that K_{max} was more or less equal across individuals by comparing the quantitative fit of that fixed model with the fit of an alternative, variable model in which α and K_{max} were both allowed to vary across individuals. To test these models, we employed the likelihood ratio test (LRT). Note that the fixed model was nested within the variable model. Let the log of the likelihood function for the above model be denoted $ll\left(\widehat{\alpha}, \widehat{k}_{max}\right)$; the likelihood ratio test statistic comparing the fixed model to the variable model is given by:

$$LRT = -2*\left[ll\left(\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{K}}_{max}\right) - ll\left(\widehat{\boldsymbol{\alpha}}, \widehat{K}_{max,global}\right)\right]. \qquad (2)$$

Asymptotically, the LRT statistic for nested models follows a χ^2 distribution with degrees of freedom equal to the difference in the number of parameters for the two models (Wasserman, 2013, p. 164). In this case, the variable model estimates K_{max} and α parameters for each person ($2n$); whereas, the fixed model estimates the α parameter for each person, and then adds a single global K_{max} parameter ($n + 1$). Thus, we have $LRT \sim \chi_{n-1}^2$. Rejection of the null hypothesis would provide evidence in favor of the variable model indicating that α and K_{max} both varied across individuals.

Figure 2 shows average correct performance on the whole-report visual-array task plotted as a function of α and K_{max} in Experiments 1 (top row) and 2 (bottom row). The left-hand column of Fig. 2 shows the results of the variable model, whereas, the right-hand column shows the results of the fixed model. Note that the best-fitting global value of K_{max} in the fixed model was equal to three items in both experiments, which corroborates the researcher-specified value used in the model reported by Adam et al. (2015).

Figure 2 also shows that average correct performance increased more across smaller α values than across larger α values, regardless of the value of K_{max}. Within the context of this statistical model, α values reflect the consistency of AC in that α values reflect the likelihood that K_{max} will be realized on any given trial. For instance, α values less than 1 reflect a positively skewed probability distributions in which there is higher probability that a participant will underachieve relative to their K_{max} potential, whereas α values greater than 1 reflect a negatively skewed probability distributions in which there is a higher probability that a participant will achieve their K_{max} potential.

LRT was performed to compare the variable and fixed models. As can be seen in Fig. 2, the variable model in which both parameters were allowed to vary accounted for individual differences in performance on the whole-report visual-array task significantly better than the fixed model in Experiment 1 (top row), $\chi^2(144) = 2{,}804.029, p < .001$, and in Experiment 2 (bottom row), $\chi^2(188) = 3{,}825.356, p < .001$. To evaluate model fit, we looked at the proportion of variance accounted for in the observed sample mean by the model implied mean. Standardized indices of model fit (e.g., Yuan, 2005) were not available for latent variable models of this type. Thus, we evaluated how well each model was able to replicate the sample mean of all trials via an R^2 statistic. The fit of the variable model in Experiment 1 was $R^2 = .98$, while the fit of the fixed model was $R^2 = .92$. Likewise, the fit of the variable model in Experiment 2 was $R^2 = .98$, while the fit of the fixed model was $R^2 = .91$.

Characteristics of the SoA distribution

Descriptive statistics for the 145 participants in Experiment 1 and 189 participants in Experiment 2 are shown in Table 1. The frequency of estimated K_{max} values (listed from one to six items) in the variable model of Experiment 1 was 4, 4, 109, 26, 2, and 0; similarly, the frequency of estimated K_{max} values in the variable model of Experiment 2 was 1, 17, 121, 48, 0, and 2. Clearly, a majority of participants in Experiment 1 (75.2%) and Experiment 2 (64.0%) had K_{max} estimates equal to three items within the context of the variable model, though a substantial number of participants in Experiment 1 (17.9%) and Experiment 2 (25.4%) had K_{max} estimates equal to four items.
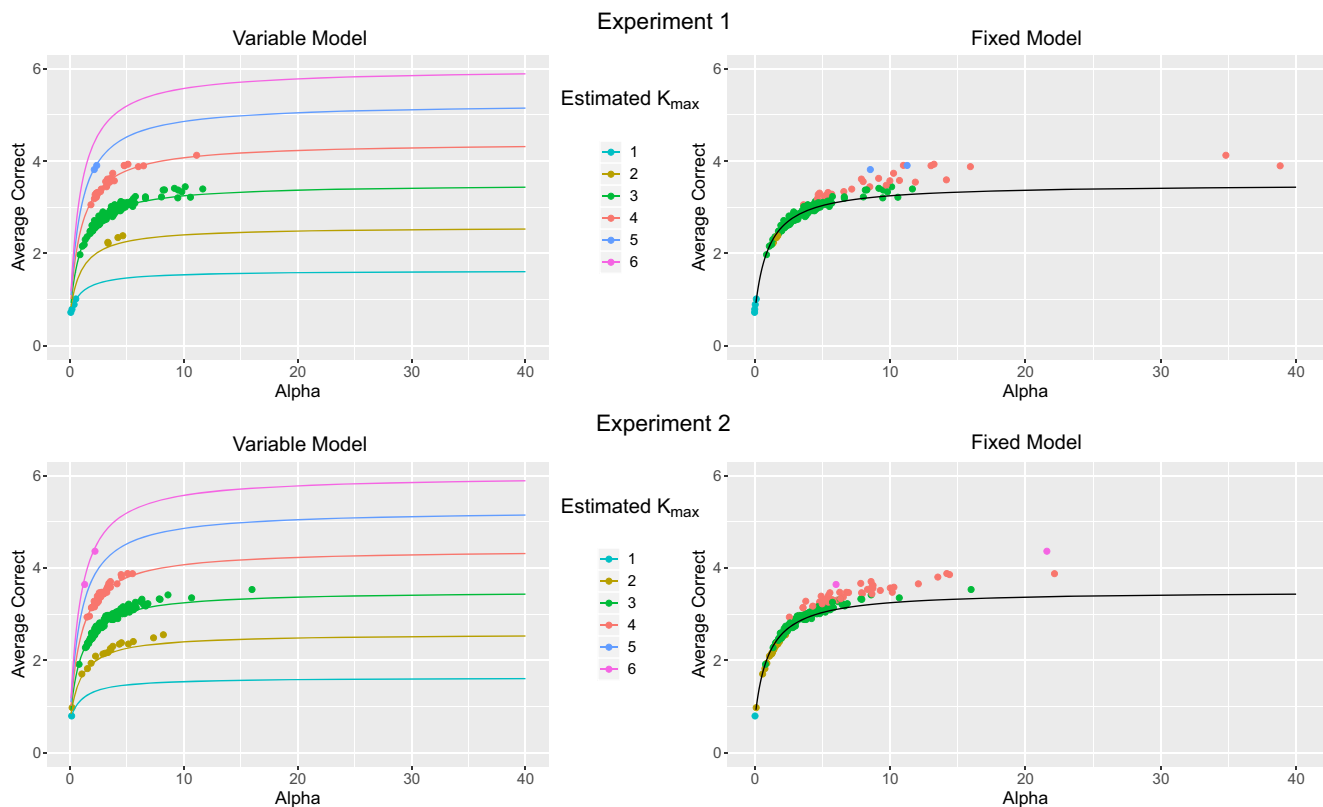
Fig. 2 Average correct performance on the whole-report visual-array task plotted as a function of α and K_{max} in Experiments 1 (top row) and 2 (bottom row). The left-hand column shows the results of the variable model, whereas the right-hand column shows the results of the fixed model. The best-fitting global value of K_{max} in the fixed model was equal to three items in both experiments. (Color figure online)

Thus, the distribution of SoA is dominated by individuals with K_{max} estimates equal to three items.

Relation between AC and SoA

There was a small, positive correlation observed between AC (α) and SoA (K_{max}) in Experiment 1 that did not attain significance, $r = .115$ (95% BCI [−0.318, 0.183]). Note that the 95% nonparametric bootstrap confidence intervals (BCI) were calculated from 1,000 resampled data sets to account for increased variability due to sampling error of the latent variable estimates. In contrast, there was a small, negative correlation observed between AC (α) and SoA (K_{max}) in Experiment 2 that did attain significance, $r = -.169$ (95% BCI [−0.311, −0.098]). Although the BCI did not include zero in Experiment 2, the fact that both correlations were opposite in direction and hovered near zero suggested that the true relation between AC and SoA is likely small in magnitude.

Table 1 Descriptive statistics for Experiments 1 and 2

	Experiment 1 (N = 145)			Experiment 2 (N = 189)		
	Mean (SD)	Skew (kurtosis)	Reliability[3]	Mean (SD)	Skew (kurtosis)	Reliability[3]
Items correct/trial	2.92 (0.54)	−1.26 (7.58)	.99	2.94 (0.48)	−0.84 (5.68)	.98
K_{max}[1]	3.14 (0.61)	−0.44 (6.92)	.87	3.19 (0.65)	0.59 (5.92)	.86
α[1]	3.73 (2.10)	1.63 (6.13)	.72	3.50 (1.85)	2.35 (14.04)	.73
RAPM[2]	25.39 (5.67)	−0.42 (−0.42)	.85	NA	NA	NA

[1] Estimate based on the variable model

[2] The RAPM numbers are based on 142 participants

[3] Reliability for RAPM and items correct/trial is Cronbach's alpha. Reliability for K_{max} and α is estimated as a function of the variance of the scores and the average of the squared bootstrap standard errors of the scores (see Equation 12 in Wang, Kolen, & Harris, 1997)

For both experiments, the integer value and restricted range of K_{max} could potentially cause attenuation in the estimated correlation.

Do AC and SoA predict Gf?

Linear regression analyses were performed on the data obtained from Experiment 1 to evaluate the relation between AC, SoA, and Gf as measured by the RAPM. This procedure treated the estimated latent variable values as known. The overall linear model was significant, $F(2, 139) = 10.53$, $p < .0001$. The overall R^2 of the model was .132. Individually, both SoA ($b = 2.51$, 95% BCI [1.23, 3.76]) and AC ($b = 0.58$, 95% BCI [0.073, 0.924]) had confidence intervals for the coefficients that did not include zero; the partial r^2 values for SoA and AC were .08 and .05, respectively.

General discussion

The present study shed light on the fundamental mechanisms underlying complex cognition by using a maximum likelihood framework to extract separate latent constructs representing SoA and AC from a whole-report version of the visual-array task. The results of two experiments consistently showed that model fit was significantly greater when the model allowed both AC and SoA to vary across individuals relative to a model in which only AC was allowed to vary (and SoA was fixed at an optimal level). This finding was important because it has the potential to clarify three interrelated issues that have not been adequately addressed by previous research.

The first issue concerned the characteristics of the SoA distribution. The distribution of discrete K_{max} values was found to be sharply peaked, such that the vast majority of individuals (91%) had K_{max} values equal to three or four items. Thus, although individual variation in K_{max} was found to be important, the observed range of this variation was also found to be small, perhaps explaining why individual variation in this parameter could be plausibly ignored in previous studies (Adam et al., 2015). We will return to this issue below.

The second issue concerned the relation between SoA and AC, which could now be examined within a single task. According to the variable model, performance was understood in terms of a binomial distribution in which K_{max} represented the number of possible items that could be maintained and α represented the success probability associated with those outcomes. Just as the number of times a coin can be flipped has no bearing on the success probability of those outcomes, the magnitude of K_{max} had little bearing on the value of α. Specifically, we observed only small correlations between K_{max} and α, and these correlations fell on either side of zero across the two experiments, even though there was nothing inherent in the statistical model that would have precluded such a relation to be found if it had existed.

The notion that K_{max} and α can be largely orthogonal parameters that jointly determine the *entire* distribution of the number of items correctly recalled on the whole-report visual-array task is important because it has become popular to use the extreme lower tail of this distribution—the number of trials in which an individual correctly recalled zero or one items—as a de facto measure of lapse rate (see, e.g., Adam & Vogel, 2017; Robison & Unsworth, 2019). Although interpretation of the number of 0s/1s would have been attributed primarily to variation in lapse rate within the context of the fixed model in which only AC was understood to vary (Adam et al., 2015), the interpretation of this number is more refined within the context of the variable model because the frequency of these (and all other) response categories are jointly determined by individual variation in both K_{max} and α. Hence, it is expected that the number of 0s/1s would correlate strongly with K_{max}, α, and the number of correctly recalled items (even if the number of 0s/1s were excluded from the average; cf. Robison & Unsworth, 2019). In this view, the relation between SoA and AC is reflected by the correlation between K_{max} and α, not the correlation between average correct and the number of 0s/1s (cf. Robison & Unsworth, 2019).

Another manifestation of the separate contribution that SoA and AC make to task performance is that two individuals could attain similar levels of performance on a single visual-array task for different reasons. For instance, the participant depicted in the top panel of Fig. 3 averaged 3.64 items correct, and the participant depicted in the bottom panel averaged 3.63 items correct. However, the former participant had a K_{max} estimate equal to 6 paired with an α estimate equal to 1.30; whereas, the latter participant had a K_{max} estimate equal to 4 paired with an α estimate equal to 3.56.

The present approach thus provides a more refined method for estimating individual variation in maintenance capacity than has been provided by average correct performance on visual array tasks (cf. Shipstead et al., 2014; Unsworth et al., 2014), which may have utility for understanding how SoA and AC undergird other forms of complex cognition such as Gf—the third issue to be addressed. Indeed, although SoA and AC were weakly related to each other in the present study, they were both found to be significant predictors of Gf. Altogether, AC and SoA were found to explain approximately 13% of the variance in Gf—an admittedly modest amount, though this magnitude was nonetheless in the range of task-level correlations that has been observed between average correct performance on the visual-array task and total scores on the RAPM task in previous studies (Shipstead et al., 2014; Unsworth et al., 2014).

However, it may be premature to judge the true relation between SoA and Gf, or between SoA and AC, on the basis of existing evidence because, in addition to providing a better
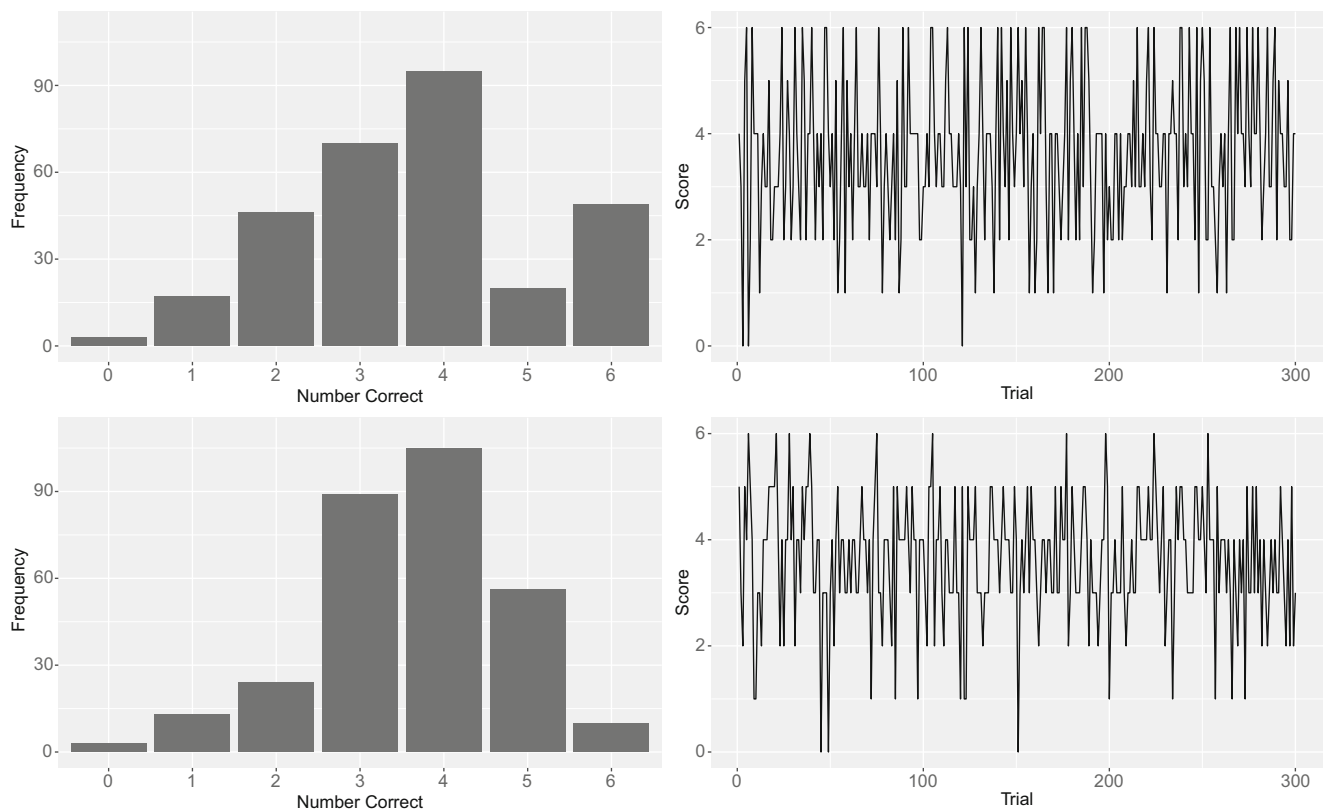
Fig. 3 Performance of two individual participants on the visual-array task. Left panels depict the frequency of trials in which each participant achieved zero to six items correct. Right panels depict fluctuations in number correct as a function of trial number (1 to 300)

estimate of SoA, the present findings suggested that typical samples are likely to contain very little individual variation in SoA. The rarity of extreme K_{max} values is an interesting question in its own right, but it also suggests that selective sampling techniques are needed to ensure that typical sample sizes contain adequate individual variation across the entire range of SoA. Accordingly, we have developed a 5–10 minute screening measure, based on the present formalization that enables researchers to identify individuals across the range of SoA in a timely and efficient manner (qubnk.com/whp/).

However, the strength of these three conclusions is limited by the relatively narrow scope of the fixed and variable models as well as by the validity of more general theoretical assumptions underlying the present approach. In particular, with respect to scope, the current study was designed to distinguish between fixed and variable models of visual-array-task performance, and it did not attempt to address a myriad of other important theoretical issues that are relevant to task performance, including the binding of colors and locations, the role of strategic factors, and the influence of long-term memory processes, to name a few. With respect to theoretical assumptions, the present study addressed the difference between the fixed and variable models within the framework of discrete capacity models, and it did not attempt to test this assumption against alternative models that assumed continuously varying memory states.

With these limitations in mind, the present study utilized the visual-array task to show that individual variation in SoA exists along with individual variation in AC. In so doing, the present study provided important new evidence that the magnitude of SoA can be separated from the consistency of AC, and that both constructs can make unique contributions to a mechanistic account of complex cognition. Finally, the present study has revealed limitations in the range of SoA used in previous individual difference studies, and it has provided a means of improving these studies by enabling identification of those in the tails of the SoA distribution.

References

Adam, K. C., Mance, I., Fukuda, K., & Vogel, E. K. (2015). The contribution of attentional lapses to individual differences in visual working memory capacity. *Journal of Cognitive Neuroscience*, *27*(8), 1601–1616. https://doi.org/10.1162/jocn_a_0081

Adam, K. C., & Vogel, E. K. (2017). Confident failures: Lapses of working memory reveal a metacognitive blind spot. *Attention, Perception, & Psychophysics*, *79*(5), 1506–1523. https://doi.org/10.3758/s13414-017-1331-8

Chuderski, A., Taraday, M., Nęcka, E., & Smoleń, T. (2012). Storage capacity explains fluid intelligence but executive control does not.

Intelligence, *40*(3), 278–295. https://doi.org/10.1016/j.intell.2012.02.010

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, *24*(1), 87–114. https://doi.org/10.1017/s0140525x01003922

Cowan, N., Elliott, E. M., Saults, J. S., Morey, C. C., Mattox, S., Hismjatullina, A., & Conway, A. R. (2005). On the capacity of attention: Its estimation and its role in working memory and cognitive aptitudes. *Cognitive Psychology*, *51*(1), 42–100. https://doi.org/10.1016/j.cogpsych.2004.12.001

deBettencourt, M. T., Keene, P. A., Awh, E., & Vogel, E. K. (2019). Real-time triggering reveals concurrent lapses of attention and working memory. *Nature Human Behavior*, *3*, 808–816. https://doi.org/10.1038/s41562-019-0606-6

Engle, R. W. (2018). Working memory and executive attention: A revisit. *Perspectives on Psychological Science*, *13*(2), 190–193. https://doi.org/10.1177/1745691617720478

Kyllonen, P. C., & Christal, R. E. (1990). Reasoning ability is (little more than) working-memory capacity?! *Intelligence*, *14*(4), 389–433. https://doi.org/10.1016/S0160-2896(05)80012-1

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, *390*(6657), 279–281. https://doi.org/10.1038/36846

Mason, R. L., & Young, J. C. (2002). *Multivariate statistical process control with industrial applications* (Vol. 9). Philadelphia, PA: Society for Applied and Industrial Mathematics (SIAM). https://doi.org/10.1137/1.9780898718461

Oberauer, K., Lewandowsky, S., Awh, E., Brown, G. D., Conway, A., Cowan, N., …. Ma, W. J. (2018). Benchmarks for models of short-term and working memory. *Psychological Bulletin*, *144*(9), 885–958. https://doi.org/10.1037/bul0000153

Robison, M. K., & Unsworth, N. (2019). Pupillometry tracks fluctuations in working memory performance. *Attention, Perception, & Psychophysics*, *81*, 407–419. https://doi.org/10.3758/s13414-018-1618-4

Shipstead, Z., Redick, T. S., Hicks, K. L., & Engle, R. W. (2012). The scope and control of attention as separate aspects of working memory. *Memory*, *20*(6), 608–628. https://doi.org/10.1080/09658211.2012.691519

Shipstead, Z., Lindsey D. R. B., Marshall, R. L., & Engle, R. W. (2014). The mechanisms of working memory capacity: Primary memory, secondary memory, and attention control. *Journal of Memory and Language, 72*, 116–141.

Shipstead, Z., Harrison, T. L., & Engle, R. W. (2015). Working memory capacity and the scope and control of attention. *Attention, Perception, & Psychophysics*, *77*(6), 1863–1880. https://doi.org/10.3758/s13414-015-0899-0

Shipstead, Z., Harrison, T. L., & Engle, R. W. (2016). Working memory capacity and fluid intelligence: Maintenance and disengagement. *Perspectives on Psychological Science*, *11*(6), 771–799. https://doi.org/10.1177/1745691616650647

Unsworth, N., Fukuda, K., Awh, E., & Vogel, E. K. (2014). Working memory and fluid intelligence: Capacity, attention control, and secondary memory retrieval. *Cognitive Psychology*, *71*, 1–26. https://doi.org/10.1016/j.cogpsych.2014.01.003

Unsworth, N., & Robison, M. K. (2017). A locus coeruleus-norepinephrine account of individual differences in working memory capacity and attention control. *Psychonomic Bulletin & Review*, *24*(4), 1282–1311. https://doi.org/10.3758/s13423-016-1220-5

Wang, T., Kolen, M. J., & Harris, D. J. (1997). Conditional standard errors, reliability, and decision consistency performance levels using polytomous IRT (Reliability issues with performance assessments: A collection of papers). *ACT Research Report Series 97(3)*, 13–40.

Wasserman, L. (2013). *All of statistics: A concise course in statistical inference*. New York, NY: Springer Science+Business Media.

Yuan, K. H. (2005). Fit indices versus test statistics. *Multivariate Behavioral Research*, *40*(1), 115–148. https://doi.org/10.1207/s15327906mbr4001_5

Open practices statement The data and model comparison are available from the Open Science Framework (https://osf.io/6w5hq/).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.