Journal of Experimental Psychology: General
2009, Vol. 138, No. 1, 64–87

0096-3445/09/$12.00 DOI: 10.1037/a0014738

Accessing Information in Working Memory: Can the Focus of Attention Grasp Two Elements at the Same Time?

Klaus Oberauer and Svetlana Bialkova
University of Bristol

Processing information in working memory requires selective access to a subset of working-memory contents by a focus of attention. Complex cognition often requires joint access to 2 items in working memory. How does the focus select 2 items? Two experiments with an arithmetic task and 1 with a spatial task investigate time demands for successive operations that involve 2 digits or 2 spatial positions, respectively. When both items used in an operation have been used in the preceding operation, latencies are shortened. No such repetition benefit (arithmetic) or a much smaller benefit (spatial) was found when only 1 item was repeated. The results rule out serial access to the 2 items, parallel access by expanding the focus, and parallel access by splitting the focus. They support the notion that 2 items are accessed by chunking them, so that they fit a focus limited to 1 chunk.

Keywords: attention, working memory, relations

Complex cognitive activities such as reasoning, problem solving, planning, and text comprehension require the manipulation of complex structural representations with multiple interrelated elements, such as mental models (Gentner & Stevens, 1983; Johnson-Laird, 1983), goal structures (Hayes-Roth & Hayes-Roth, 1979), and propositional structures (Kintsch & van Dijk, 1978). Most theories of complex cognition assume that these representations are assembled and manipulated in working memory (WM), a system with a limited capacity for temporary maintenance and manipulation of information (Baddeley, 1986; Cowan, 2005). This view is supported by consistent evidence for high correlations between measures of WM capacity and performance in complex cognitive tasks (for a review see Conway, Kane, & Engle, 2003).

Despite the broad consensus that WM is crucial for holding available and manipulating information in complex cognition, surprisingly little is known about how the system accomplishes these functions. Most experimental research on WM, rooted in the literature on human memory, investigates the system's maintenance function (for recent reviews see Lewandowsky & Farrell, 2008; Jonides et al., 2008), whereas little research has been devoted to how information in WM is manipulated. The purpose of this article is to contribute to our understanding of how the WM system manipulates information in complex cognition.

Klaus Oberauer and Svetlana Bialkova, Department of Experimental Psychology, University of Bristol, Bristol, United Kingdom.

This work was supported by Economic and Social Research Council Grant RES-000-23-1527 to Klaus Oberauer. We thank Tim Jones for assistance with collecting the data of Experiment 3 and Anna Lelievre for assistance with recording the instructions for Experiment 3.

Correspondence concerning this article should be addressed to Klaus Oberauer, Department of Experimental Psychology, University of Bristol, 12a Priory Road, Bristol BS8 1TU, United Kingdom. E-mail: k.oberauer@bristol.ac.uk

Our starting point is the observation that the manipulation of WM contents typically involves, at any given processing step, only a subset of all elements currently held in WM, whereas the other elements must be maintained without influencing or being affected by the operation. This requires a mechanism for selective access to subsets of WM contents. For instance, when we mentally add two three-digit numbers, we may start by adding the units first, then progressing to the tens and the hundreds. WM must remember all six digits of the given numbers, and all intermediate results, but at the same time make sure that only selected contents enter each operation. Likewise, when manipulating a mental model of a spatial array of objects—for example, when thinking through a series of hypothetical moves on a chess board—WM must make sure that the imagined movement of one object is applied to that object exclusively while maintaining the position of other objects unchanged.

The Focus of Attention in Working Memory

Simultaneously holding a number of representations and selectively accessing some of them requires a *focus of attention* to be deployed to a subset of the current contents of WM. The proposal of an attentional focus in WM has been made by several authors (Cowan, 1988, 1995; Garavan, 1998; McElree & Dosher, 1989; Oberauer, 2002). The concept of a focus of attention in WM is based on an understanding of attention as a mechanism, or collection of mechanisms, for selecting representations for (cognitive) action (Allport, 1987). Selection of representations from WM bears many similarities to selection from perceptual input (Lepsien & Nobre, 2006), and therefore it is likely that similar, if not the same, mechanisms are involved in both selection functions.

Many authors proposing a focus of attention in WM assume that the focus is limited to a single mental object or chunk

(Garavan, 1998; McElree & Dosher, 1989; Oberauer, 2002).[1] This assumption is supported by the available evidence, which comes from two basic findings. First, McElree and his colleagues have found that, in short-term recognition tasks such as the Sternberg (1969) task, the last item presented is accessed at a faster rate than all earlier items (for a review see McElree, 2006). McElree argued that the last item presented is still held in the focus of attention and thus can be directly compared to the probe, whereas earlier items have to be retrieved into the focus. The increased rate of access extends to the last group of three items defined by a common semantic category, possibly because the common category facilitates chunking the three items into one mental object (McElree, 1998).

The second finding suggesting a focus of attention limited to one element comes from the object-switch paradigm introduced by Garavan (1998). He asked participants to count geometrical shapes of two categories. Shapes were displayed one by one, such that each display required updating one counter in WM. The time people took for each counting step was about 300 ms longer when the current shape was from a different category than the preceding shape, compared to when a shape from the same category was shown again. Garavan interpreted these switch costs as the time to switch the focus of attention from one mental object (i.e., counter) to the other.

Later research has extended this finding, using an arithmetic task in which participants hold up to four digits in WM and perform a sequence of arithmetic operations on them (Oberauer, 2003). At each step, the operation sign and one argument of the equation (i.e., the second addend or subtrahend) were displayed on the screen, such as "+ 3" or "− 7." The other argument (i.e., the first addend or subtrahend) had to be taken from WM, and the spatial location or the color of the stimulus specified which of the digits currently held in WM had to be accessed for each step. Switching from one digit in WM to another incurred a time cost that increased with the memory set size, that is, the number of digits to be maintained available for access.

These findings are compatible with a focus of attention that is limited to a single mental object or chunk at any time—but then, the experimental tasks never required access to more than one element from WM. Many everyday cognitive activities, however, require access to more than one element in WM as input for a cognitive operation. For example, the multidigit mental addition discussed above requires, in the first step, access to the units of both numbers, and when both numbers are held in WM, this means access to two elements in WM. In general, access to two or more elements is required for any cognitive operation that involves integration of information from several independent representations (as in mental arithmetic), the building and processing of relations (e.g., computing verb–noun attachments in sentence comprehension or building spatial mental models of the relative locations of objects in spatial deductive-reasoning tasks), or the comparison of entities (e.g., comparing stimuli along several feature dimensions in categorization and inductive reasoning). The examples listed here show that access to more than one element in WM is required for many processes that play crucial roles in complex cognitive tasks such as multidigit arithmetic, deductive and inductive reasoning, and language comprehension. This observation raises the possibility that the focus of attention appears limited to a single chunk only when the experimental task does not require more, whereas it can expand, possibly to an upper limit, when the task demands it (Cowan et al., 2005). The degree of flexibility of the focus of attention in WM has important implications for how the system manipulates information—as we elaborate below, a severely limited focus implies strictly sequential processing of all elements in WM, whereas a more flexible focus enables some degree of parallel processing. Therefore, understanding the limitations and capabilities of the focus of attention in WM is important to inform process models of complex cognition.

Six Hypotheses

To summarize, the question we address in this article is: How does the focus of attention provide selective access to two elements in WM when the task requires it? We consider six hypotheses, derived by the hypothesis tree in Figure 1. The tree emerges from a succession of theoretical decisions. The first decision, represented by the top-most branching, is whether the focus is structurally limited to a single chunk. Assuming that it is, the next decision is whether a chunk can consist of only one element in WM, or whether two (or more) elements can be chunked ad hoc, thus forming a single chunk that can be held in the focus.

If ad-hoc chunking is not possible, the focus is limited to accessing one element at a time. Thus, when an operation requires access to two elements from WM, they have to be accessed one after the other. For instance, adding together two digits, both of which are held in WM, would involve accessing one, sending it to the arithmetic module, then accessing the other and sending it to the arithmetic module. This modus of processing presupposes that the arithmetic module has a buffer in which the first digit can be held until the second one arrives. The arithmetic module would receive as a further ingredient the instruction to add or to subtract; this instruction comes from a task set that we assume to be represented outside WM. The two digits can be accessed in two orders, so there are two hypotheses to consider under this branch; *serial-access 1–2* assumes that the focus starts with the first digit and then moves on to the second, whereas *serial-access 2–1* assumes the reverse order.

If ad-hoc chunking is possible, the two digits in WM required for an arithmetic operation can be bundled into a chunk, which can then be selected and forwarded to the arithmetic module as a single object. We define a *chunk* as a unit that contains information of separable elements, but whose elements cannot be accessed or manipulated separately unless the chunk is unpacked (Halford, Wilson, & Phillips, 1998; Miller, 1956). Therefore, a chunk can be taken into the focus of attention only as a whole and can be used only as a whole. We distinguish two variants of the chunking hypothesis that differ in their predictions, as we show below. The *ordered-chunking* hypothesis assumes that elements (e.g., digits) are bundled in a fixed order within the chunk, so that each is assigned a fixed role (1st vs. 2nd digit) in the arithmetic equation. The *free-chunking* hypothesis assumes that elements are not linked to a fixed order within the chunk, so that the chunk can be reused while the elements in it swap their roles.

[1] We refer to the contents of WM as elements, (mental) objects, or chunks. By elements we mean the nominal items used in a task, such as digits, letters, or words. By chunks and (mental) objects we mean a unitized representation that can bundle together several elements (Miller, 1956). We assume that by default, each element is encoded into WM as a chunk, but elements can be chunked together to form a unitized object, as we explain below.

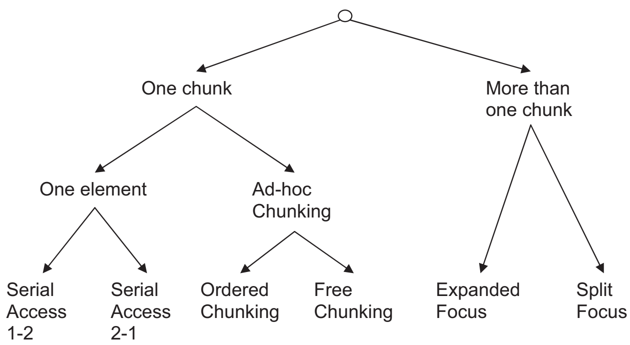

Figure 1. Hypothesis tree with theoretical decision points leading to the six alternative hypotheses about how the focus of attention accesses two elements in working memory.

Assuming that the focus is not limited to a single chunk, we have two further theoretical options. One, the *expanded-focus* hypothesis, is that the focus of attention can expand to accommodate as many elements as are needed for an operation, up to a limit (Cowan et al., 2005). Thus, in the mental addition task both digits would be taken into the focus and sent to the arithmetic module simultaneously. This hypothesis differs from the chunking hypothesis in that the two digits held in the focus remain independent objects, which can be removed from the focus (and replaced by other objects) independent of each other.

The final hypothesis—the second alternative under the multiple-chunk branch—is that the focus of attention is flexibly adapted to task demands in a way that goes beyond expanding its capacity. Whereas in an expanded focus each element is interchangeable, the *split-focus* hypothesis assumes that the focus of attention is split into separate foci, one for each element that needs to be accessed for the current task. Each focus is dedicated to a particular role in the task. For example in mental addition or subtraction, one focus would select the first argument of the equation (i.e., the digit before the operation sign) and the other would select the second argument (i.e., the digit after the operation sign). The two arguments would be sent to the arithmetic module simultaneously along separate channels, thus avoiding cross-talk between the arguments.

In the following section, we introduce the experimental paradigm we used to investigate dual access to WM and then derive predictions from the six hypotheses for that paradigm.

The Dual-Access Paradigm

The dual-access paradigm is an extension of the paradigm used by Oberauer (2003) to study the focus of attention in WM. We asked participants to hold in WM four digits, each displayed in a different color. The colors, which always occurred in the same order during presentation of the digits, serve as retrieval cues by which individual digits can be accessed. Thus, participants had to hold in mind the temporary bindings of four digits to their colors, or to their corresponding serial positions.

A series of 11 equations was then presented, each of which required access to two digits in WM. The equations were displayed as a plus or minus sign, flanked by two color patches in place of the two addends or subtrahends (see Figure 2). The colors indicated which digits must be retrieved from WM. For example, the first equation in Figure 2 is “BLUE – RED”; the blue and red patches serve as retrieval

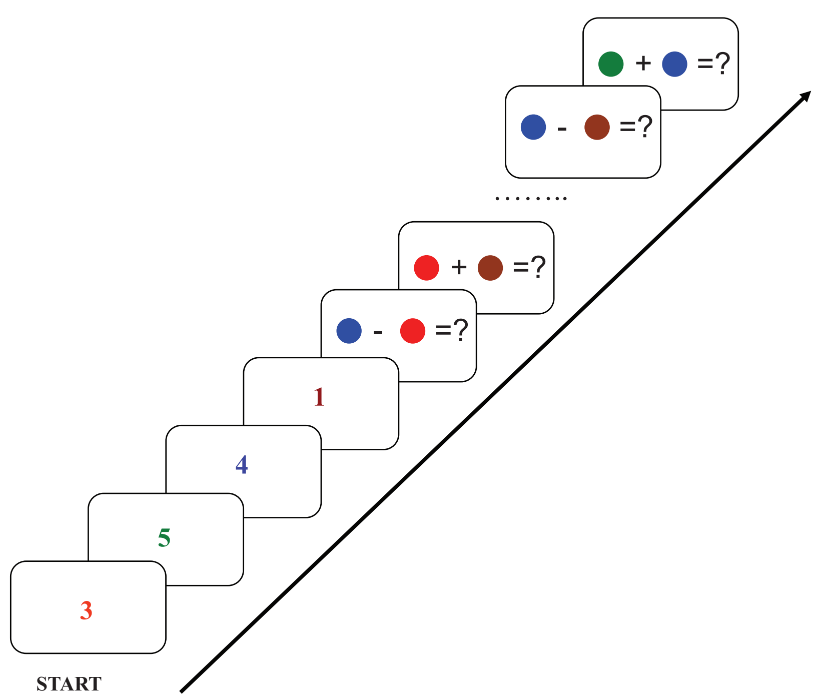

Figure 2. An example of a trial of the dual-access condition in Experiments 1 and 2. Equations consisted of two colored circles; each trial had 11 equations; the first equation served to initialize the sequence and was not analyzed.

cues for the numbers 4 and 3, respectively, so the required response is 1. The dependent variable is the time until participants enter the result of each equation.

From one equation to the next, the first argument can be the same as in the preceding equation (repeat condition [rep]), it can switch to the digit that has been the second argument in the preceding equation (switch-to-other condition [swo]), or it can be a digit that has not been used in the preceding equation (switch-to-new condition [swn]). The same three levels of the switching variable apply to the second argument of the equation.

By crossing the two switching variables we obtain nine conditions, illustrated in Figure 3. For example, let equation $n - 1$ be "RED + BLUE." When the following equation n is "RED − GREEN," it would be sorted into the design cell "repetition of first argument, switch-to-new of second argument," because the first argument, RED, is repeated, whereas the second argument switches from BLUE to GREEN, and GREEN was not used in equation $n - 1$. (We refer to this condition as rep–swn; the first acronym refers to the switching condition of the first argument, and the second acronym refers to the switching condition of the second argument). When instead equation n is "BLUE – RED," it would be sorted into the "switch-to-other, switch-to-other" (swo–swo) cell, because the new first argument switches to the color that has been used for the other (i.e., the second) argument on the preceding equation, namely BLUE; likewise, the new second argument switches to the color that has been used for the other (i.e., the first) argument in the preceding equation, namely RED. Two of the nine conditions in Figure 3 result in using the same digit for both arguments, implying that only one digit has to be accessed from WM. Therefore, we excluded these two conditions from the design, leaving seven design cells.

Predictions

We next derive predictions for the response latencies in the seven design cells from the six hypotheses introduced above. The predictions are models of the latencies, formulated as patterns of long, intermediate, and short latencies (coded as 1, .5, and 0, respectively) across the seven conditions. We use these models as predictors in a multilevel regression analysis, as explained below. The basic assumption guiding the predictions is that the information last held in the focus of attention when solving equation $n - 1$ is still in the focus when work on equation n commences. When that information can be reused, a repetition benefit is predicted. Conversely, when the information needed to work on equation n must be retrieved into the focus by using a new color–digit binding held in WM, a switch cost is incurred.

The Serial-Access Models

The first two hypotheses assume that the focus of attention selects the digits one by one, beginning either with the first or the second argument. Serial-access 1–2 assumes that the focus starts with the first argument and then moves on to the second. After entering the result of equation $n - 1$, the focus will hold that equation's second argument. If the following equation n uses that digit as its first argument, no focus switch is necessary, whereas in all other conditions, the focus must switch to another digit as the first argument of equation n. Thus, serial-access 1–2 predicts a repetition advantage for the condition in which the first argument of equation n involves a switch to the second argument of equation $n - 1$ (conditions swo–swo and swo–swn). The predicted pattern of this and all the following models is given in Table 1, and illustrated in Figure 4. (Table 1 also contains two further serial-access models, *serial-access 1–1* and *serial-access 2–2*, which are relevant only for the single-access conditions discussed in the context of Experiment 1 below.)

Serial-access 2–1 starts with accessing the second argument of equation $n - 1$, and thus, the focus ends up holding that equation's first argument. A repetition benefit is predicted for the case where the first argument of equation $n - 1$ becomes the second argument of equation n, in other words, when the second argument of equation n involves a switch to the other argument (i.e., conditions swo–swo and swn–swo). In all other cases, the focus must access a new digit, and hence, no differences are predicted between the other conditions.

The Chunking Models

The chunking hypothesis assumes that the two digits used in equation $n - 1$ are packed into a chunk. We distinguish two versions of the chunking model, ordered chunking and free chunking. In *ordered chunking*, the two digits are chunked in a fixed order, and thus are bound to their argument roles—the first element in the chunk is assigned to the first argument in the equation, and the second element is assigned to the second argument in the equation. Because

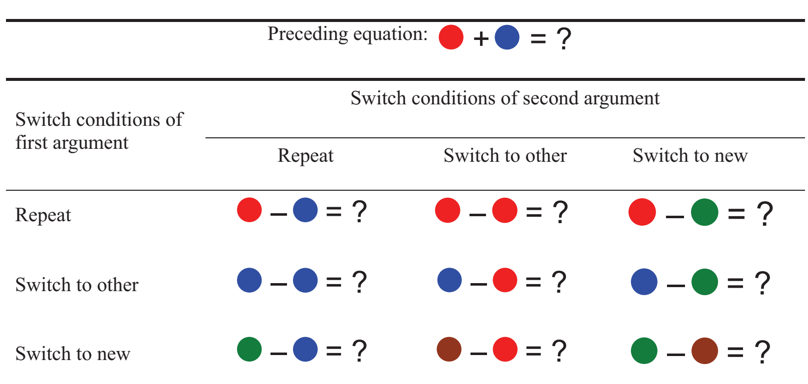

Figure 3. Switch conditions of the first and the second argument in Experiments 1 and 2.

Table 1
Predictors for Experiments 1 and 2

Argument	Condition						
	rep–rep	rep–swn	swo–swo	swo–swn	swn–rep	swn–swo	swn–swn
1st	Repeat	Repeat	Other (2nd)	Other (2nd)	New	New	New
2nd	Repeat	New	Other (1st)	New	Repeat	Other (1st)	New
Predictor							
Serial 1–2	1	1	0	0	1	1	1
Serial 2–1	1	1	0	1	1	0	1
Expand	0	0.5	0	0.5	0.5	0.5	1
Split	0	0.5	1	1	0.5	1	1
Ordered chunk	0	1	1	1	1	1	1
Free chunk	0	1	0	1	1	1	1
Serial 1–1[a]	0	0	1	1	1	1	1
Serial 2–2[a]	0	1	1	1	0	1	1

Note. The first two indented rows indicate the switching condition of the first and the second argument, respectively, for each of the seven joint conditions at the heads of the columns. The remaining rows represent the predictors, which contain a 1 for the conditions with no repetition benefit, 0.5 for conditions with a small repetition benefit, and 0 for conditions with a large repetition benefit.
[a] For single-access conditions only.

elements in a chunk cannot be manipulated individually, their order cannot be flipped around without unpacking the chunk. This model predicts a repetition benefit only for the condition in which the whole ordered chunk is repeated, that is, when both arguments are repeated in the same roles (i.e., condition rep–rep). In all other conditions, a new chunk must be formed, and this costs extra time, regardless of whether one of the digits is repeated from the preceding equation.

The second chunking model, *free chunking*, assumes that the two digits are not ordered within a chunk. In this model, the chunk formed for equation $n - 1$ can be reused when both digits from equation $n - 1$ are used again, either in the same roles (i.e., condition rep–rep) or with a reversed role assignment (i.e., condition swo–swo). In all other conditions, a new chunk has to be formed and selected by the focus of attention. In particular, the focus cannot reuse the digit selected for one argument of equation $n - 1$ but select a new digit for the other argument, because the digits in a chunk cannot be kept in the focus individually. Therefore, all conditions that involve access to at least one digit that has not been used in equation $n - 1$ require the formation and selection of a new chunk.

The Expanded-Focus Model

If the focus of attention is expanded to take in both digits as separate objects, they will both be in the focus after completion of equation $n - 1$. When both of them are used again in equation n, either in the same argument roles (condition rep–rep) or in swapped roles (condition swo–swo), a large repetition advantage is predicted, because the focus does not have to access a new digit at all. When one of them is used again, but the other argument requires a digit not used in equation $n - 1$ (conditions rep–swn, swo–swn, swn–rep, and swn–swo), then a smaller repetition benefit is to be expected because the focus needs to access only one new digit. Different from the chunking models, the expanded-focus model assumes that the focus can drop one digit while maintaining the other, thus reaping a partial benefit from partial repetitions. No repetition gain is predicted only for the case in which both arguments of equation n are digits not used in equation $n - 1$ (condition swn–swn).

The Split-Focus Model

If the focus is split into two separate foci, each of which is dedicated to an argument role, repetition benefits will be reaped only when a digit of equation $n - 1$ is repeated in equation n in the same argument role. If both arguments are repeated (condition rep–rep), the repetition benefit should be maximal. If one argument is repeated (conditions rep–swn and swn–rep), there should be a smaller benefit, because only one of the two foci needs to switch. If the two foci are switched one after the other, switching one of them takes half as long as switching both of them. If the two foci are switched in parallel, the time for switching two of them is the larger of the durations of each switch. The switch durations for each focus are random variables, and the expected value of the maximum of two random variables exceeds that of both variables individually; therefore, switching only one focus would be faster than switching both. No repetition benefit is to be expected in the conditions in which both foci need to access a new digit (conditions swo–swo, swo–swn, swn–swo, and swn–swn).[2]

[2] The split-focus model can be emulated by a strategy in which the focus of attention selects one digit at a time but does so in random order, rather than in a constant order as assumed in the serial-access models outlined above. If the focus accesses the digits of equation $n - 1$ in random order, it will end up holding the first digit half of the time and holding the second digit half of the time. If the digit held in the focus at the end of equation $n - 1$ is repeated in the same argument role in equation n, it can immediately be reused, yielding a repetition benefit. This happens with a 50% chance in the conditions with one repeated digit (i.e., rep–swn and swn–rep) and with certainty in the condition in which both digits are repeated in their roles (condition rep–rep).

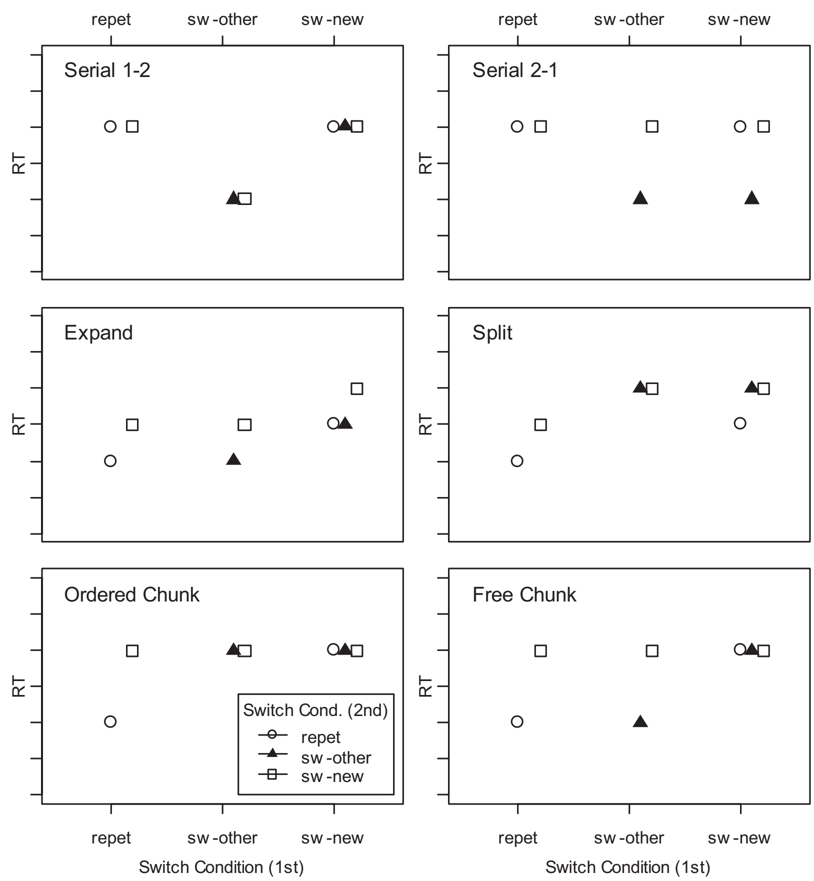

Figure 4. Predicted patterns of reaction times (RT) according to the six predictors for the dual-access paradigm (see top rows in Table 1). The *x*-axis represents the switching condition with regard to the first argument, and the parameter represents the switching condition with regard to the second argument. The serial predictors assume that the focus holds a single element at any time (Serial 1–2: focus moves from first to second argument; Serial 2–1: focus moves from second to first argument). Expand reflects the assumption that the focus expands to hold both elements without distinguishing their roles. Split reflects the assumption that the focus holds both elements but separates them by their roles in the equation. Ordered chunking means that the focus chunks both elements while preserving their roles in the equation. The chunk can be reused only as a whole and only with the same role assignments of both elements. Free chunking means that the focus chunks both elements without preserving their roles; the chunk can be reused only as a whole, but the roles can be swapped. repet = repetition; sw-other = switch to other; sw-new = switch to new.

Statistical Analysis Strategy

Our research strategy is one of model selection rather than null-hypothesis testing. As argued by statisticians for a long time (e.g., Cohen, 1995; Gigerenzer, 2004; Howson & Urbach, 1993; Wagenmakers, 2007), null-hypothesis testing is a weak method of scientific inference that should be used only as a last resort in a situation where neither theory nor prior empirical work permits formulation of more precise alternative hypotheses than that there is an effect versus that there is not. Fortunately, here we are in a better position, because we can select between six alternative models. Therefore, we fit the models to the data and selected between them on the basis of their relative success, measured by their likelihood (i.e., the probability of the data under the assumption of the model).

Our model selection procedure considers not only pure cases but also the possibility that different people behave according to different models, or even that the same person uses processes specified by different models on different trials. Therefore, we also fit mixture models to the data. Specifically, we compare linear regression models using different subsets of predictors from Table 1 to predict the mean latencies across the seven conditions. We estimate the regression weights of each predictor as free parameters; these weights reflect the relative contribution of each predictor in the mixture models.

The regression models were applied in the framework of linear mixed-effects (LME) models (Pinheiro & Bates, 2000), using the *lme* function in the *nlme* package (Pinheiro, Bates, DebRoy, & Sarkar, 2005) in *R* (R Development Core Team, 2005). Mixed-effects models are applied to the data on two (or more) levels. On a first level, a model predicts the group mean latencies of the seven conditions, that is, the fixed effects of the experimental manipulations. On a second level, the model predicts the deviation of individual participants from the group mean in each condition, that is, the random effects. The random effects are modeled by a standard deviation for each parameter. In this way, the model is applied to $N \times 7$ data points (i.e., each individual's mean reaction time [RT] in each of the seven conditions) using only two free parameters (one mean and one standard deviation) for each predictor, plus one mean and one standard deviation for the intercept. (LME can also estimate correlations between parameters, but for simplicity we fixed all correlations to zero).

For example, imagine that some of the participants used serial access, starting with the first argument, on most trials. Their individual patterns of RTs across the seven conditions would look like the predicted pattern of the serial-access 1–2 model in Figure 4. Other participants would use free chunking most of the time, and their individual RT patterns would therefore look like the free chunking pattern in Figure 4. When averaged across all participants, the means of the seven conditions would fall somewhere between the means predicted by the two patterns. When we enter the serial-access 1–2 predictor and the free chunking predictor as fixed effects into the regression model, both of them would account for a substantial part of the variance between conditions (within each subject). Their regression weights would reflect the relative size of the two groups of participants. As long as we use only fixed effects, the mean RTs in each condition would be predicted to be the same for each individual:

$$RT_i = A + B_1S_i + B_2C_i + E. \tag{1}$$

In Equation 1, the subscript i represents the experimental condition, A is the intercept, E is the residual, and S_i and C_i are the predictions of the two model predictors (serial-access 1–2 and free chunking, respectively) for condition i, with their respective regression weights, B_1 and B_2.

Individual differences are accounted for by random effects. First, participants differ in their overall speed, and this effect is captured by the random effect of the intercept. Second, participants differ in the degree to which they behave according to one or the other model, and this is captured by the random effect of the regression weights. The full model therefore is

$$RT_{ij} = A + a_j + (B_1 + b_{1j})S_i + (B_2 + b_{2j})C_i + E. \tag{2}$$

Here, a_j is the deviation of the individual intercept of participant j from the mean intercept A; b_{1j} and b_{2j} reflect the deviations of the weights for each predictor for participant j from the mean weights. Thus, if participant j refers to a participant who uses only free chunking, b_{1j} will have a substantial negative value, so that $B_1 + b_{1j}$ is close to zero, and b_{2j} will have a large positive value, so that $B_2 + b_{2j}$ is close to 1. Importantly, the a_j, b_{1j}, and b_{2j} values are not estimated independently for each participant. Rather, they are implied by a normal distribution of parameters across subjects, with means A, B_1, and B_2 (i.e., the fixed effects) and their respective standard deviations (i.e., the random effects).

We select between alternative models on the basis of their likelihood ratio,[3] which can be tested for significance when models are nested (using the *anova* function in *R*) and on the basis of the Akaike information criterion (AIC) and the Bayes information criterion (BIC), which are both based on the log-likelihood and in addition penalize for the number of free parameters (for accessible introductions to maximum-likelihood-based model selection see Glover & Dixon, 2004; Myung, 2003). A predictor was included in a model if and only if it improved the BIC; because the BIC is the most conservative of the three criteria, this implied in all cases that the included predictor also improved the AIC and was significant. We refer to a predictor that surpassed the inclusion criterion as *substantial*.[4]

The most important advantage of LME models over conventional test statistics such as analysis of variance (ANOVA) is that they account for all systematic sources of variance in an experiment simultaneously. This has made them an invaluable tool, for instance, in psycholinguistic studies for simultaneously accounting for random effects across subjects and across items (Baayen, Davidson, & Bates, 2008; Quené & van den Bergh, 2008).[5] Closer to the present study, LME models enable the application of regression models to within-subject variance without having to estimate all regression weights separately for each individual (Hoffman & Rovine, 2007). For instance, Kliegl (2007) predicted fixation durations in reading from experimental manipulations and naturally occurring sources of variance across items. Compared to applying a single regression model to the group means, this method has the advantage in that it takes individual differences in the direction and size of effects into account, rather than modeling an artificial "average participant," and that it builds on many more data points, thereby increasing statistical power. Compared to separate regression models for each individual, this method saves free parameters and thereby reduces the chance of over-fitting (i.e., fitting to noise). In addition, it safeguards against distortion of parameter estimates that arises from outliers and from unreliability in individual data sets (Hoffman & Rovine, 2007) by imposing a normal distribution on the random effects across individuals (as well as items).

The present analysis strategy takes this approach one step further by predicting within-subject variance (i.e., the fixed effects of the experimental manipulation) by theoretically motivated models, rather than the experimentally manipulated variables. Each model predictor specifies the expected pattern of RTs or error rates across all seven experimental conditions (in terms of *high* = 1, *intermediate* = .5, *low* = 0; cf. Tables 1 and 5), analogous to a contrast code. In this way we can analyze all effects in our structurally incomplete design jointly, rather than having to break them down into many pairwise comparisons. Most important, however, LME enables us to test theoretically motivated models against each other, rather than testing individual effects predicted by one or another model against the theoretically uninteresting null hypoth-

[3] Likelihood ratios are meaningful only when the *lme* function is called with setting *method* = *ML*, so this is what we did (Pinheiro & Bates, 2000).

[4] Within the framework of Bayes statistics, differences between models in BIC can be translated into a Bayes factor (Wagenmakers, 2007). According to the guidelines introduced by H. Jeffreys, the strength of evidence reflected by a Bayes factor between 1 and 3 is "barely worth mentioning," from 3 to 10 it is "substantial," from 10 to 30 it is "strong," between 30 and 100 it is "very strong," and above 100 it is "decisive" (as cited in "Bayes factor," n.d.). The smallest improvement in BIC for any predictor included in our models had a Bayes factor of 3.16, which brings it into range of being supported by "substantial" evidence.

[5] In the present context, random effects of items are not included because the population of items (e.g., arithmetic equations in Experiments 1 and 2) is limited and is almost exhaustively sampled in our experiments.

esis (see Gigerenzer, Krauss, & Vitouch, 2004, for a particularly poignant example of how testing a model against null, rather than against a competing model, can lead to the wrong conclusion).

Because we are interested in model comparison, not in the testing of individual effects, we do not present conventional null-hypothesis tests for differences between conditions. Readers worried about whether our effects might have resulted from chance alone can glean the relevant information from the confidence intervals in the figures. All confidence intervals are computed for within-subjects comparisons by the method of Bakeman and McArthur (1996). Confidence intervals that do not overlap reflect differences significant at $p = .01$, and confidence intervals overlapping by less than 50% reflect differences significant at $p = .05$ (Cumming & Finch, 2005).

Experiment 1

The main goal of Experiment 1 was to provide data from the dual-access paradigm to test the six models outlined above and their mixtures. A second goal was to compare the dual-access paradigm to two analogous single-access conditions. One single-access condition involves access to the first argument in each equation from WM while the second argument is displayed on the screen. We refer to this as the single-access (first) condition. The other single-access condition is the reverse, with the first argument presented on the screen and the second taken from WM; this is called the single-access (second) condition. A direct comparison of dual-access and single-access conditions is of interest because it affords a comparison of attentional processes directed to WM contents and attentional processes directed to perceptual information in an otherwise identical task. If the same attentional processes apply to digits presented on the screen as to digits in WM, the same model should apply to the single-access and the dual-access conditions.

We included both single-access conditions to test the possibility of a cognitive asymmetry between the two arguments of an equation. All experiments conducted so far with the object-switch paradigm were single-access conditions in which the first argument had to be taken from WM, whereas the second argument was either displayed on the screen, as in the arithmetic tasks of Oberauer (2003), or was a constant addition of 1, as in the counter task of Garavan (1998). The results from these experiments are compatible with a focus of attention that holds the first argument only: Switch costs were found when the first argument changed from equation $n - 1$ to equation n but not when the second argument changed (Oberauer, 2003). This could mean that the focus of attention holds only the argument retrieved from WM, not the argument presented visually. Alternatively, it could mean that the focus holds only the first argument of an equation. This hypothesis could be justified by assuming that the second argument, together with the operation sign, forms a parameter of the task set that specifies which operation is applied to the first argument. For example, the equation "2 + 3" would be parsed into the object "2" held in the focus of attention and the task set "add 3."

To test the two alternative interpretations of the object-switch effect in single-access conditions, we included the new condition in which the first argument was given perceptually (i.e., a digit displayed on the screen) and the second argument was to be taken from WM. If the focus always holds the first argument, then we should find an object-switch effect for the first argument in this condition but not for the second.

For the analysis of the single-access paradigms we included two further serial-access models. *Serial 1–1* assumes that the focus of attention holds only the first argument. Thus, a repetition benefit is expected when the first argument is repeated, regardless of what happens with the second argument (i.e., in conditions rep–rep and rep–swn). All other conditions should be equally slower than these two conditions. Analogously, *serial 2–2* assumes that the focus only ever holds the second argument. Therefore, a repetition benefit is expected when the second argument is repeated (i.e., conditions rep–rep and swn–rep), and all other conditions are assumed to be equal and slower than these two conditions. These two predictors are presented in the bottom two rows of Table 1 and illustrated in Figure 5.

Method

Participants. Twenty-four participants (16 women and 8 men) between the ages of 17 and 35 years took part in Experiment 1. All had full color vision and normal or corrected-to-normal vision.

Materials. Thirty-six types of equations resulted from fully crossing the four within-subjects variables: (a) object switching

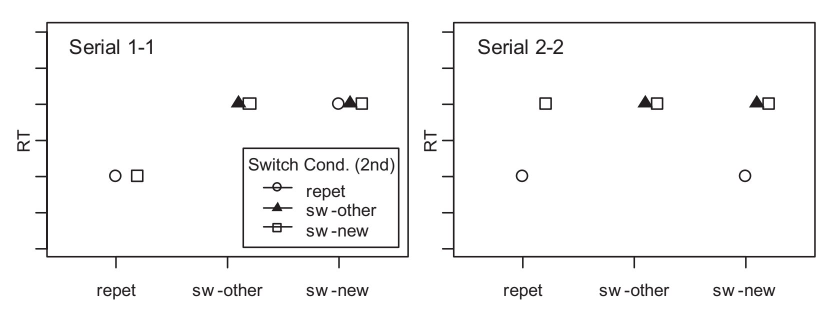

Figure 5. Predicted patterns of reaction times (RT) according to the additional two predictors for the single-access conditions (see bottom two rows of Table 1). The *x*-axis represents the switching condition with regard to the first argument, and the parameter represents the switching condition with regard to the second argument. Serial 1–1 means that the focus holds the first argument, and Serial 2–2 means that the focus holds the second argument of the equation. repet = repetition; sw-other = switch to other; sw-new = switch to new.

pertaining to the first argument with three levels: repetition, switch to the second argument of the previous equation, switch to a new digit; (b) object switching pertaining to the second argument with three levels: repetition, switch to the first argument of the previous equation, switch to a new digit; (c) operation sign with two levels: plus or minus; and (d) operation sign switch with two levels: repeat or switch. Equations with two identical arguments (i.e., $x + x = 2x$ and $x - x = 0$) were excluded as experimental conditions. Complete repetitions, in which both arguments and the operation sign were identical to the preceding equation (i.e., $x + y$ on step n and $x + y$ on step $n + 1$), were also excluded. This left 25 conditions.

A computer program generated all trials. Each trial consisted of four memory digits, followed by 11 equations to be performed with these digits. All equations had results between 1 and 9. The first equation could not be classified with regard to the object-switch conditions and therefore was not included in any analyses. The remaining 10 equations were classified into 1 of the 25 conditions. The program generated trials such that, across the 19 trials of a block, all 25 conditions occurred about equally often (i.e., 7 ± 1 repetitions per condition per block). Apart from this constraint on overall frequency, the 25 kinds of equations followed each other in random order. Trials were generated for 10 blocks, three each for the single-access (first) and single-access (second) condition and four for the dual-access condition. The same algorithm constructed equations for all three access conditions. Therefore, even in the two single-access conditions the digits used in the equations, including those presented on the screen, were limited to the set of four digits held in WM. This constraint ensures full comparability of the single-access and the dual-access conditions.

Procedure. The three access conditions were carried out in separate sessions, with one session each for the single-access (first) and the single-access (second) condition and two sessions for the dual-access condition. The order of the four sessions was counterbalanced by a Latin square design, and the sessions were scheduled on different days. The single-access sessions consisted of three blocks of 19 trials each. The dual-access sessions consisted of only two blocks with 19 trials each, because we anticipated that trials in this condition would take longer than those in the other two access conditions. At the beginning of each session there was a practice block consisting of 10 trials. These trials were excluded from analysis.

Each trial was initiated by the word "START" appearing on the screen for 1 s. Then 100 ms after it disappeared, the four digits to be held in memory were presented, one by one, each in a different color. The four colors always appeared in the same order: red, green, blue, and brown. Participants were instructed to remember which digit was presented in which color and to move on after each digit by pressing any of the number keys on the computer keyboard. After the fourth digit disappeared, 11 equations were displayed one by one in the center of the screen. In the dual-access condition, each equation consisted of two colored circles, separated by a plus or a minus sign and followed by an equals sign and a question mark (see Figure 2 for an example trial). In the single-access (first) condition, only the first argument was represented by a colored circle, whereas the second argument was given as a digit. In the single-access (second) condition, the second argument was represented by a colored circle, and the first argument was given as a digit.

Participants were instructed to complete the equation by replacing each colored circle by the digit that had been presented in that color. They entered the solution to each equation through the number keys on the computer keyboard, and 100 ms later the next equation was displayed. Once participants had provided an answer for the last equation in a trial, the word "START" was displayed after an intertrial interval of 100 ms, signaling the beginning of the next trial. At the end of each block, participants were encouraged to take a short break.

Results

In the condition with both digits repeated (rep–rep), the operation sign had to switch from plus to minus or vice versa to avoid repeating the exact equation, including the result. To avoid confounds of operation sign switch with any of the object-switch variables, we therefore limited our analyses to equations with an operation sign switch in all conditions.

Accuracy was very good in the single-access conditions, with 3.0% and 2.6% errors in the single-access (first) and single-access (second) conditions, respectively. In the dual-access condition participants committed 8.8% errors. Our main focus of interest was on the response latencies. We analyzed only latencies of correct responses. After eliminating extreme outliers (>50 s), we defined outliers as latencies exceeding an individual's mean by 3 intra-individual *SD*s. These outliers (1.8% to 1.9% of correct latencies across conditions) were discarded. The remaining latencies were averaged across operation sign for each participant, resulting in means for the seven object-switch conditions for which the predictors were defined in Table 1. These are the data submitted to the mixed-effects models.

Our modeling strategy was a variant of the forward regression strategy, entering predictors one by one until no further improvement in fit could be achieved. At each step we selected the predictor that improved model fit the most. Thus, we first selected the predictor with the best individual fit, as assessed by the model's log-likelihood, and in the second step added each of the other predictors to it in turn, evaluating all the resulting two-predictor models. If exactly one of these two-predictor models led to a substantial increase of fit over the best single-predictor model, we selected that two-predictor model as the final model of the fixed effects. If more than one two-predictor model superseded the best single-predictor model, we tested combinations of the predictors that were successfully added in the second step until no further substantial improvement could be reached by adding predictors. Entering predictors by their order of strength ensures that we account for the data with the smallest possible number of predictors, and trying out the addition of all remaining predictors ensures that we do not miss any possible improvement of fit from one of them. Throughout these steps, each predictor was entered with two free parameters, one for its mean regression weight (i.e., the fixed effect) and one for its standard deviation (i.e., the random effect). In a final step, we attempted to remove the standard deviation term for each predictor in the best-fitting model. When that did not lead to a substantial loss of fit, the standard deviation was permanently removed, leaving only the fixed effect for that predictor.

The predictors in Table 1 were applied to the three access conditions (i.e., the two single-access conditions and the one dual-access condition) separately. The best-fitting models are summarized in Table 2. The table contains information about the regression weight of each predictor. Because the predictors are scaled from 0 (for the

Table 2
Model Statistics for Experiment 1

Predictor	Mean weight (*SD*)	95% CI	L-ratio	ΔBIC	ΔAIC
Single-access, first argument from WM (AIC = 123.2, BIC = 142.0, R^2_{adj} = .92)					
Intercept	2.22 (0.58)	1.97, 2.47			
Serial 1–1	0.28 (0.42)	0.08, 0.47	64.2	54.0	60.2
Ordered chunking	0.20 (0)	0.08, 0.33	9.5	4.5	7.5
Single-access, second argument from WM (AIC = 7.8, BIC = 32.8, R^2_{adj} = .97)					
Intercept	1.87 (0.53)	1.64, 2.09			
Split focus	0.21 (0.29)	0.03, 0.38	23.3	13.1	19.3
Serial 2–2	0.13 (0.20)	0.01, 0.24	16.6	6.3	12.6
Ordered chunking	0.38 (0)	0.28, 0.49	43.3	38.2	41.3
Dual-access, both arguments from WM (AIC = 250.2, BIC = 265.8, R^2_{adj} = .93)					
Intercept	3.02 (0.76)	2.70, 3.34			
Free chunking	1.38 (0.56)	1.12, 1.63	213.6	202.8	209.1

Note. Mean weights are unstandardized regression weights for fixed effects; they reflect the size of the effect of each predictor on the reaction time scale. Their standard deviations are estimates of the random effects (when zero, the random effect was dropped from the model). CI = upper and lower limits of confidence intervals for the estimation of fixed effects, computed by the *intervals* function of *nmle*; L-ratio = likelihood ratio of complete model to model with effect removed (all *p*s < .001); ΔBIC and ΔAIC = changes in Bayes information criterion (BIC) and Akaike information criterion (AIC), respectively, when removing effect (positive values mean decrease of fit); WM = working memory.

conditions with shortest predicted latencies) to 1 (for conditions with longest predicted latencies), these weights reflect how much (in seconds) the latencies in the fastest and the slowest conditions are predicted to differ by the effect of that predictor in the model. The table also contains information about the loss of fit when each predictor was removed from the model. In all cases, the loss of fit was substantial (and also highly significant; all p values $< .01$). The likelihood ratio for each predictor indicates how much more likely the data are under the full model, compared to the reduced model with that predictor removed. The change in AIC and BIC reflect the same likelihood ratio after subtracting the penalties for the different numbers of free parameters in the two models. For all models we also report the adjusted R^2, which corrects for the number of free parameters:

$$R^2_{adj} = 1 - \frac{\sum_{i=1}^{n}(d_i - \hat{d}_i)^2/(n-k)}{\sum_{i=1}^{n}(d_i - \bar{d})^2/(n-1)},$$

where d_i represents the observed values, $\hat{d}_i$ is the predicted values, $\bar{d}$ is the mean, n is the number of data points, and k indicates the number of free parameters.

The mean latencies are presented by the two object-switch variables in Figures 6 (single-access conditions) and 7 (dual-access condition); the layout of these figures corresponds to that of the individual predictors in Figures 4 and 5. The fixed-effect estimates of the best-fitting models are represented by an asterisk accompanying each data point.

As predicted from previous research, serial 1–1 was a substantial predictor for the single-access (first) condition. The serial 1–1 model assumes that the focus of attention holds only the first argument, which is the one to be selected from WM, and is not engaged with the visually presented second argument. In addition, the ordered-chunking model also was a substantial predictor. This model assumes that the focus of attention holds both arguments, packing them into a chunk.

The results for the single-access (second) condition are to a large degree complementary. The predictor serial 2–2 reflects the assumption that the focus of attention holds only the second argument, which in this condition is the one selected from WM. The ordered-chunking model also was a substantial predictor in this condition. Different from the single-access (first) condition, here the split-focus model accounted for an additional portion of variance.

The dual-access condition showed a different pattern (see Figure 7). The free-chunking model was the only predictor that accounted for a substantial proportion of variance across conditions.

Error rates were low and therefore of secondary interest; they were analyzed primarily to check for speed–accuracy trade-offs. Table 3 summarizes the mean error rates across conditions. Error rates were submitted to the same model selection analysis as latencies, with the only difference being that we entered predictors into not a linear but a logistic regression, using the *nlme* function in the *nlme* package. The analysis of error rates for the single-access conditions indeed revealed a hint of speed–accuracy trade-off. In the single-access (first) condition, the best-fitting model included serial 1–1 as the only substantial predictor, entering with a negative regression weight, thus predicting larger error rates in the two conditions with repetition of the first argument (rep–rep and rep–swn). Analogously, in the single-access (second) condition, the best-fitting model included serial 2–2 with a negative regression weight.[6] Both these effects go against those ef-

[6] This model also included a substantial effect of serial 1–1, but this predictor contributed only a substantial random effect, not a substantial fixed effect, and therefore can be ignored with regard to speed–accuracy trade-offs.

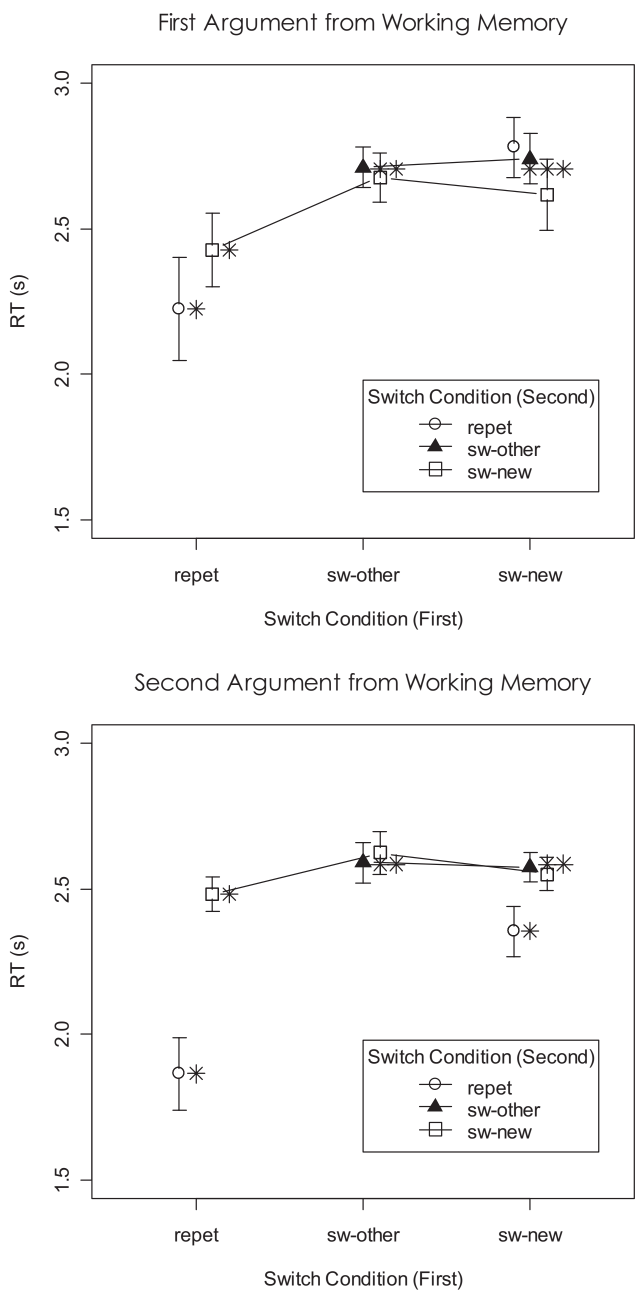

Figure 6. Reaction times (RTs) in Experiment 1, single-access conditions (top: first argument retrieved from working memory; bottom: second argument retrieved from working memory). Error bars are 95% confidence intervals for within-subjects comparisons computed by the method of Bakeman and McArthur (1996). Predictions of the best-fitting model are presented as asterisks. repet = repetition; sw-other = switch to other; sw-new = switch to new.

fects observed in the latencies, where the corresponding predictors had positive weights, reflecting faster latencies when the arguments accessed from WM were repeated. Whereas the latencies showed substantial repetition benefit, the error rates showed a repetition cost. Numerically, the repetition cost was tiny (see Table 3), so that speed–accuracy trade-off is unlikely to account completely for the large repetition benefit in latencies. Nevertheless, the consistency of this finding across conditions is noteworthy.

The analysis of error rates from the dual-access condition resulted in the free-chunking predictor as the only substantial predictor, which had a positive regression weight. Thus, the error data in this condition confirmed the pattern observed in the latency data.

Discussion

The three access conditions showed distinctly different patterns of latencies and error rates. The data patterns in the two single-access conditions yielded evidence for two modes of operation for the focus of attention. One is that the focus deals only with the digit to be selected from WM. The other mode is to chunk both digits together into an ordered chunk. Repeating the whole chunk in the same order resulted in a benefit that went above what would be expected from the effects of repeating individual digits. This added benefit of repeating both digits cannot be due to a repetition of the whole equation because the operation sign always changed from one equation to the next, so that in one equation, the two digits had to be added, and in the other, they had to be subtracted from each other.

In the single-access (second) condition, we found that the split-focus predictor contributed substantially to explaining the experimental effects. This means that not only repeating the digit from WM but also repeating the visually presented digit resulted in a latency advantage on its own. These two advantages were independent from each other but limited to repetitions in the same argument role—repeating the visually presented digit as the digit

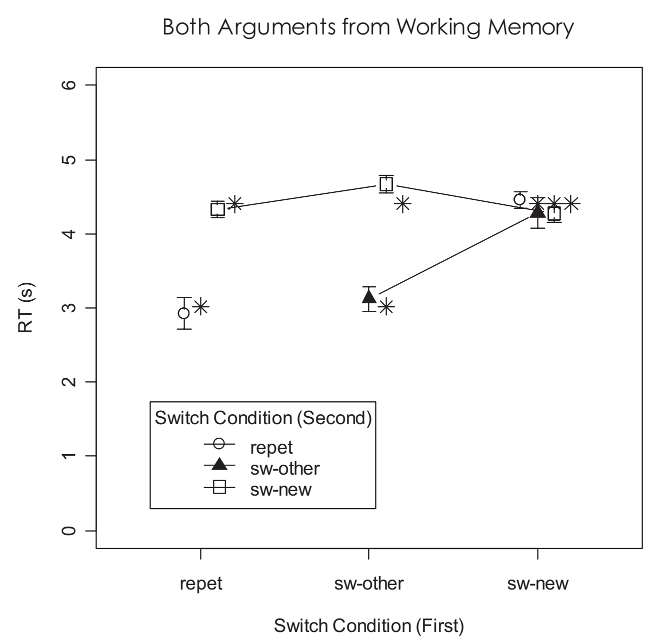

Figure 7. Reaction times (RTs) in Experiment 1, dual-access condition (both arguments retrieved from working memory). Error bars are 95% confidence intervals for within-subjects comparisons. Predictions of the best-fitting model are presented as asterisks. repet = repetition; sw-other = switch to other; sw-new = switch to new.

Table 3
Error Proportions in Experiment 1

Access condition	Switch condition						
	rep–rep	rep–swn	swo–swo	swo–swn	swn–rep	swn–swo	swn–swn
First argument	.029	.040	.025	.022	.025	.034	.034
Second argument	.028	.032	.022	.020	.043	.022	.018
Both arguments	.082	.091	.074	.089	.093	.103	.082

Note. rep = repeat condition; swn = switch-to-new condition; swo = switch-to-other condition.

from WM in the next equation, or the other way around, did not help. A comparable repetition benefit for the visually presented digit was not found in the single-access (first) condition and was also not found in Experiment 1 of Oberauer (2003), which realized a single-access (first) condition. Thus, the repetition benefit for the visually given digit seems to occur only when that digit takes the role of the first argument in the equation. One possible explanation for the split-focus effect is that the focus of attention—on the trials in which digits are not chunked—accesses either one or both digits in a serial fashion, with a preference for focusing on the WM digit and a preference for focusing on the first digit of the equation. In the single-access (first) condition, the two preferences converge on a strong bias of accessing the first digit most of the time. In the single-access (second) condition, the two preferences guide the focus to different digits. In most cases, the focus would end up holding the second digit—the one selected from WM—but in some cases, it would hold the first digit, and these cases would create the benefit for conditions in which the first digit is repeated. This benefit would be additive with the other effects and therefore would emulate a split-focus effect (see Footnote 2).

Consistently across both single-access conditions, we found a small amount of speed–accuracy trade-off for the serial-access models. When the digit taken from WM was repeated, latencies were substantially shorter, whereas error rates slightly increased. We can only speculate about the cause of this unexpected effect. Perhaps participants, upon noticing a repetition of the color cue, sometimes reused not only the digit associated to that color that they still held in the focus of attention but inadvertently also reused the task set (i.e., the operation sign) or the other argument of the preceding equation, or both. Such a tendency could come about from associations between task sets and the objects they have been applied to (Koch, Prinz, & Allport, 2005; Waszak, Hommel, & Allport, 2003). Reusing the operation sign as well as the digits results in a fast response, because no new result needs to be computed, but the result would be wrong in all trials included in our analysis because we included only trials in which the operation sign changed from the preceding equation.

The dual-access condition resulted in a pattern clearly distinct from that of the two single-access conditions. The data unambiguously point towards chunking as the only processing mode when both digits had to be selected from WM. Different from the single-access conditions, the free-chunking predictor rather than the ordered-chunking predictor explained the data best, implying that the digits packed into a chunk are not bound to their argument roles, and repeating both digits in swapped roles (condition swo–swo) yielded a benefit nearly as large as that obtained with repeating them in the same roles (condition rep–rep).

Why, when chunks can have a free order in the dual-access condition, do they appear to be ordered in the two single-access conditions? There is a simple explanation: The chunk itself does not impose an order to its elements. A repetition benefit, however, can be obtained only when the participant recognizes that the chunk built for solving equation $n - 1$ can be reused for equation n. In the dual-access condition, chunk repetition is signaled by repetition of both colors in the equation, regardless of their order. Seeing "BLUE + GREEN" following "GREEN + BLUE" indicates that the old chunk can be reused, and no retrieval of digits from WM is needed to realize that. In the single-access conditions, in contrast, repetition of both digits in swapped roles would not be recognizable. When "BLUE + 3" follows "GREEN – 2," participants cannot see that the blue circle refers to the "2" they hold in WM without retrieving that digit. Retrieving the blue digit, however, means to pull it into the focus of attention, and there it replaces the chunk used in the preceding equation. Therefore, in the single-access conditions, repeating a chunk with swapped argument roles does not lead to a repetition benefit. Only when both digits are repeated in their same argument roles (condition rep–rep) are the equations visibly similar, and participants can reap the benefit of chunk repetition. Therefore, the ordered-chunk predictor captures the chunk repetition benefit in the single-access conditions.

Experiment 2

In Experiment 2 we focused on the dual-access condition. We presented equations—consisting of an operation sign flanked by two colored circles—in two parts, varying the stimulus onset asynchrony (SOA) between them. In the condition with SOA = 1,000, the color representing the first argument was displayed 1,000 ms before the color representing the second argument. With SOA = –1,000, the color representing the second argument was displayed 1,000 ms before the color representing the first argument. The third condition, with SOA = 0, was identical to the dual-access condition of Experiment 1; that is, both colors were presented simultaneously. Latencies were always measured from the moment the full equation was presented; thus, the asynchronous SOA conditions provide a potential preview benefit compared to the SOA = 0 condition.

The purpose of the SOA manipulation was to test a prediction of the chunking models. When the WM system uses ad-hoc chunking, it needs information from both arguments of the equation to decide whether to reuse the chunk left in the focus from the preceding equation or to build a new chunk. When the two parts are presented asynchronously, the system can use one of two strategies: Either it waits until the second part is presented and then processes the whole equation as in the SOA = 0 condition or it forfeits the potential benefit of being able to reuse the chunk from the preceding equation and starts building a new chunk with the digit that is already cued by the color presented first. The waiting strategy implies that the pattern of latencies should look the same in all three SOA conditions. In particular, the preview of half the equation in the asynchronous conditions should yield no benefit compared to the SOA = 0 condition. The alternative strategy, starting a new chunk, comes at a risk. In those conditions in which the old chunk cannot be reused, starting a new one as soon as the first color cue appears creates a head start compared to the SOA = 0 condition, resulting in faster measured latencies. In those conditions, however, in which the old chunk could be reused (i.e., conditions rep–rep and swo–swo), this strategy leads to a loss of the potential repetition benefit that is expected for SOA = 0. As a result, the profile of latencies across the seven object-switch conditions should be relatively flat in the asynchronous conditions compared to SOA = 0.

A second reason for manipulating SOA is that asynchronous presentation of the retrieval cues creates a condition that disadvantages chunking and strongly encourages serial access. Whereas with chunking the focus of attention has little chance of reaping a preview benefit, with serial access the focus can fully process the digit cued first and then proceed to the digit cued second. For instance, with SOA = 1,000, focusing first on the first argument and then moving to the second argument fits perfectly with the sequence of presentation. This condition, therefore, is ideally suited for the serial 1–2 model. Likewise, the negative SOA condition provides ideal conditions for the serial 2–1 model, in which the focus moves from the second to the first argument. If the WM system could flexibly choose between chunking and serial access, the present asynchronous conditions should therefore lead to a larger contribution of the serial-access predictors and a reduced contribution of the chunking predictor compared to the SOA = 0 condition. Conversely, if chunking is a robust modus of operation for accessing two elements from WM that is not easily adapted to the opportunities of the task environment, the chunking predictor should be equally dominant in all three SOA conditions.

Method

Participants. Twenty participants from the University of Bristol community (12 women and 8 men) between the ages of 17 and 35 years took part in Experiment 2. All had full color vision and normal or corrected-to-normal vision.

Materials and procedure. Materials and procedure were the same as in Experiment 1 with the following exceptions: All equations consisted of two colored circles. The equations were presented in two parts. In trials with SOA = 1,000, the colored circle on the left side, together with the operation sign, appeared 1,000 ms before the circle on the right side. In trials with SOA = −1,000, the circle on the right side, together with the operation sign, appeared 1,000 ms before the circle on the left side. In trials with SOA = 0, the whole equation was presented at once, exactly as in the dual-access condition of Experiment 1. SOA was varied between trials in a random order but held constant for all equations within one trial. All latencies were measured from the moment the whole equation was presented, so that the asynchronous SOA conditions provide a partial preview of the equation, the processing of which did not count toward the measured latency.

Experiment 2 consisted of three sessions (approximately 1 hr each) on different days. Each session consisted of two blocks of 19 trials each. At the beginning of each session there were 10 practice trials, which were excluded from analysis.

Results

Error rates were comparable to the dual-access condition in Experiment 1, with 9.9% for the zero SOA, 9.4% for the positive SOA, and 9.6% for the negative SOA. Latencies of error responses were excluded from analysis. Outliers, defined as in Experiment 1, were removed (2.4%, 1.7%, and 1.4% in the conditions with zero, positive, and negative SOA, respectively).

The mean RTs for the SOA = 0 condition are displayed in Figure 8; those for the positive and the negative SOA conditions are shown in Figure 9. We tested the predictions outlined in the introduction to Experiment 2 by fitting regression models to the three SOA conditions simultaneously, thus accounting for 3 × 7 means. The modeling strategy was to start with a very basic model informed by the results of Experiment 1 and enriching it step by step with predictors that test specific hypotheses. The basic model consisted of the free-chunking predictor applied in all SOA conditions and two predictors capturing general preview benefits in the two asynchronous SOA conditions. The free-chunking predictor, applied equally to all SOA conditions, represents the assumption that the focus of attention always waits until the full equation is presented. The preview predictor for SOA = 1,000 was set to −1 for all seven object-switch conditions with positive SOA and

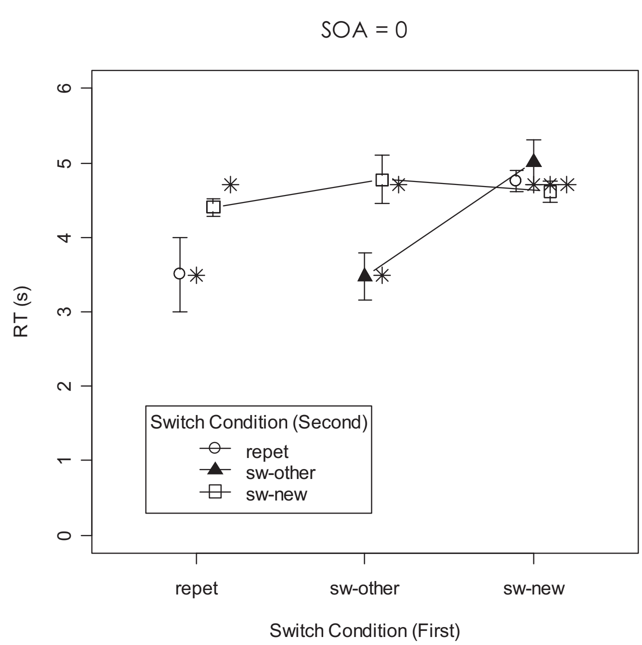

Figure 8. Reaction times (RTs) in Experiment 2, SOA = 0. Error bars are 95% confidence intervals for within-subjects comparisons. Predictions of the best-fitting model are presented as asterisks. SOA = stimulus onset asynchrony; repet = repetition; sw-other = switch to other; sw-new = switch to new.

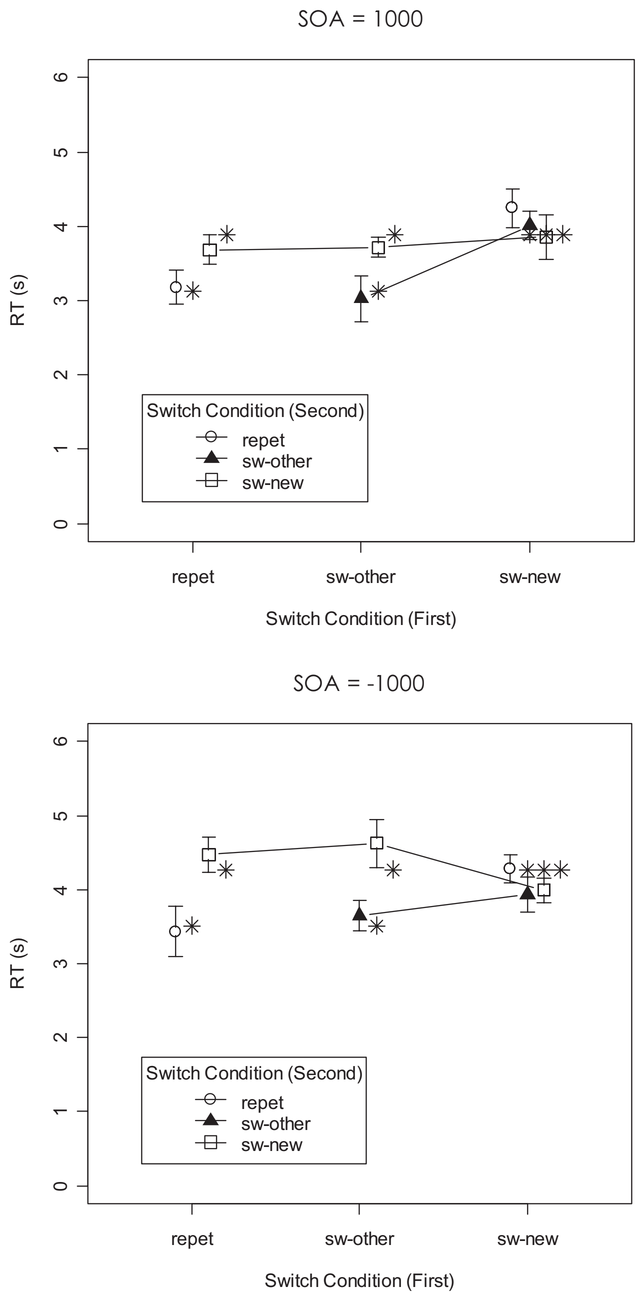

Figure 9. Reaction times (RTs) in Experiment 2, nonzero stimulus onset asynchrony (SOA) conditions (top: SOA = 1,000 ms; bottom: SOA = −1,000 ms). Error bars are 95% confidence intervals for within-subjects comparisons. Predictions of the best-fitting model are presented as asterisks. repet = repetition; sw-other = switch to other; sw-new = switch to new.

to 0 for all conditions with the other two SOA values. This preview predictor therefore represents the advantage from previewing the first argument. Likewise, the preview predictor for SOA = −1,000 was set to −1 for the seven object-switch conditions with SOA = −1,000, and to 0 for all conditions with the other two SOA values, thus representing the advantage of previewing the second argument. The general preview predictors were meant to reflect preview benefits that were constant for all object-switch conditions. For instance, participants could use the preview time for perceptual processing of the first-presented color cue and anticipatory eye movements to the location of the other color cue. They could also use the preview for task-set switching between addition and subtraction, because the operation sign was always included in the preview stimulus.

Starting from this basic model with the free-chunking predictor and the two general previous predictors, we first added two serial-access predictors tailored to the affordances of the asynchronous SOA conditions. A *preview-serial 1–2* predictor was applied to the positive SOA condition only, because this condition was ideally suited for the focus to move from the first to the second argument. Doing so would result in a benefit in the positive SOA condition, relative to SOA = 0, specifically in the conditions in which the focus can repeatedly use the second argument of equation $n-1$ as the first argument of equation n. To capture this benefit, the preview-serial 1–2 predictor was set to −1 in the two conditions that afford this (i.e., swo–swo and swo–swn) and to 0 in all other conditions. Thus, the preview-serial 1–2 predictor equals the serial 1–2 predictor in Table 1 after subtracting 1, applied to the positive SOA condition only. Likewise, we applied the predictor *preview-serial 2–1* to the negative SOA condition to account for the benefit expected in that condition when the focus moved from the second to the first argument. A benefit is expected when the first argument of equation $n-1$ is repeated as the second argument of equation n (i.e., conditions swo–swo and swn–swo). In the negative SOA condition only, these two conditions were set to −1 in that predictor; all other conditions were set to 0. We added these two predictors individually to the basic model. In both cases, this led to no substantial improvement of fit, with a likelihood ratio of 2.07 ($p = .35$) for adding preview-serial 1–2 and of 0.86 ($p = .64$) for adding preview-serial 2–1. In both cases, AIC and BIC increased, confirming that the added free parameters were not worth the negligible increase in likelihood. We conclude that the asynchronous presentation did not induce serial access.

Next, we added the *preview-free-chunking* predictor to the basic model. This predictor captures the expected effect of starting a new chunk upon seeing the preview stimulus in the conditions with positive and negative SOA. When starting a new chunk in the preview period, latencies should be shorter than in the SOA = 0 condition, except in the two conditions in which the chunk from equation $n-1$ could have been reused (i.e., rep–rep and swo–swo), in which case latencies should be longer than those in the SOA = 0 condition. Thus, preview-free-chunking is the mirror image of free-chunking and is computed as .5 minus free-chunking for the positive and the negative SOA conditions; in the zero SOA condition this predictor is set to zeros throughout because this condition affords no preview. Thus, the preview-free-chunking predictor represents the interaction between SOA (zero vs. asynchronous) and the free-chunking predictor. Adding the preview-free-chunking predictor resulted in a substantial improvement of fit (see Table 4). Its regression weight of 0.46 implies that, in the asynchronous SOA conditions, the net advantage of the conditions that afforded reusing the previous chunk over those conditions that did not was diminished by 460 ms. The final model is summarized in Table 4; its fixed-effect estimates are shown as asterisks in Figures 8 and 9.

Table 4
Model Statistics for Experiment 2

Predictor	Mean weight (*SD*)	95% CI	L-ratio	ΔBIC	ΔAIC
Joint model (AIC = 792.1, BIC = 830.9, R^2_{adj} = .91)					
Intercept	3.49 (1.23)	2.91, 4.07			
Preview (SOA = 1,000 ms)	0.59 (0)	0.44, 0.73	61.4	55.4	59.4
Preview (SOA = −1,000 ms)	0.21 (0)	0.07, 0.35	8.3	2.3	6.3
Free chunking	1.22 (0.73)	0.83, 1.61	22.8	16.9	20.9
Preview free chunking	0.46 (0)	0.21, 0.71	12.9	7.0	10.9

Note. Mean weights are unstandardized regression weights for fixed effects; they reflect the size of the effect of each predictor on the reaction time scale. Their standard deviations are estimates of the random effects (when zero, the random effect was dropped from the model). CI = upper and lower limits of confidence intervals for the estimation of fixed effects, computed by the *intervals* function of *nmle*; L-ratio = likelihood ratio of complete model to model with effect removed (all *p*s < .001); ΔBIC and ΔAIC = changes in Bayes information criterion (BIC) and Akaike information criterion (AIC), respectively, when removing effect (positive values mean decrease of fit); SOA = stimulus onset asynchrony.

To investigate potential speed–accuracy trade-offs, we applied all predictors in Table 1 to the error proportions. No predictor led to an improvement of model fit over the null model (i.e., a model with only the intercept). Thus, error rates did not differ substantially across conditions in a way correlated with any of our predictors.

Discussion

The results of the second experiment provide a clear confirmation of chunking as the dominant mode of operation in the dual-access paradigm with a mental arithmetic task. Despite creating conditions that were particularly suitable for serial access and not favorable to chunking, the single-access predictors did not account for any systematic variance in the data. Rather, the free-chunking predictor accounted for the largest part of the variance across all three SOA conditions. In addition, the data confirmed a prediction following from the chunking hypothesis: With asynchronous presentation, the advantage of those conditions that afford reusing the chunk from the preceding equation should diminish relative to the SOA = 0 condition, because on some trials participants forfeit the possible advantage of reusing the existing chunk and start over building a new chunk without full knowledge of the equation. This prediction was reflected by the preview-free-chunking predictor, which turned out to capture a substantial part of the variance in the data.

We conclude that, when the task requires access to two digits from WM, the focus consistently chunks the two digits into a free, not-ordered chunk, even in conditions where chunking implies sacrificing potential benefits—either those of using the preview time for processing or those of reusing the chunk carried over from the preceding equation. The robustness of the chunking model suggests that, at least in the dual-access paradigm with mental arithmetic, chunking is not a strategy chosen among alternatives but rather the only option the focus actually has. The alternative models—obtaining the two digits from WM serially, expanding the focus, or splitting it—might just not be possible.

Expanding the focus and splitting the focus might not be feasible strategies for the dual-access paradigm simply because the focus of attention is structurally limited to holding one chunk at a time. Serial access to two digits from WM might be unfeasible because when the focus moves to the second digit, it loses the first digit. To keep both digits the focus would have to place the digit selected first into a buffer before moving to the second digit. If no such buffer is available, the serial-access model is not suited for solving the problem of selecting two digits from WM. Serial access might still be an option in conditions where only one digit has to be retrieved from WM and the other is perceptually given, as in the single-access conditions of Experiment 1. For instance, a perceptually given digit could be processed by a separate attention mechanism tuned to perceptual input rather than the contents of WM, and thus, each digit would be selected by its own attentional device. Alternatively, one could argue that no attention is needed for the visually presented digit because the visually displayed part of the equation is the only digit presented on the screen at any time, and thus, no selection is required. The situation is different for the digit to be taken from WM, because four digits are held in WM, and the relevant one has to be selected from them. A fruitful avenue toward investigating the interplay of perceptual attention and attention to WM would be to increase the selection demand for the perceptually given information in the single-access conditions of Experiment 1.

Experiment 3

In the final experiment we investigated access to two locations in WM in a spatial task. Participants remembered the locations of four elements (the digits 1 to 4) in a 4 × 4 grid. They updated the digit's locations according to instructions such as "put 2 to the left of 3." In more general terms, each updating step involved moving one digit, called the *moved digit*, into the cell immediately left of, right of, above, or below the other digit, called the *relatum*.

We selected this task for three reasons. First, it provides a generalization of our investigation to spatial WM. Second, the new task places different demands on the focus of attention than the arithmetic task, as we explain below. Third, the relational instructions used in this experiment bridge the present work on access to WM with research on reasoning with spatial relations; we elaborate on this link in the *Discussion* section for this experiment.

The spatial updating task differs from the arithmetic task in that it does not require retrieval of information about two elements. In the spatial task it is not necessary to access the current locations of both digits. Rather, the focus must access the current location of the relatum and, in addition, must select the as-yet empty new location of the moved digit to bind that digit to this new location. Knowledge of the old location of the moved digit is not necessary, because wherever it was before, the moved digit's new location is fully determined by the relatum's location and the spatial relation in the instruction. Therefore, a repetition benefit should be expected if and only if the relatum of instruction n is held in the focus of attention after completion of instruction $n - 1$.

There are four scenarios to be considered as variants of this hypothesis; these scenarios map onto four different sets of predictors for the regression analysis. The predictors for Experiment 3 are summarized in Table 5. In the first scenario, the focus always holds only the relatum of each instruction. In that case, a repetition benefit would be found only when the relatum of instruction $n - 1$ is repeated as the relatum of instruction n. This prediction is captured by predictor *Serial R–R* in Table 5 (because the relatum is the second digit mentioned in each instruction, this predictor corresponds to *Serial 2–2* in Table 1). This scenario is implausible, however, because it does not explain how WM can place the moved digit in its new position without ever focusing on it.

The second scenario is that the focus of attention starts on the relatum and moves from there to the new position of the moved digit, following the direction in the instruction. At the completion of instruction $n - 1$, the focus would be on the location of the moved digit. A repetition benefit would be expected if the moved digit of instruction $n - 1$ figures as the relatum of instruction n. For instance, when given "put 2 above 3" as instruction $n - 1$, the focus would move from 3 to 2. If instruction n is "put 1 to the left of 2," the focus is already on the relatum 2 and can directly move left to find the new location of 1. This process would be faster than the retrieval of a new relatum, such as when instruction n is "put 1 to the left of 4." The predicted pattern of repetition benefits following from the second scenario is represented by the predictor *Serial R–M* in Table 5 (corresponding to *Serial 2–1* in Table 1).

The third scenario is that the focus of attention starts on the relatum of instruction $n - 1$ and then expands to take in the location of the moved digit as well. At the end of instruction $n - 1$, it would hold both the relatum and the moved digit. Thus, at the end of an instruction "put 2 above 3," the focus would hold two elements, the locations of 2 and 3. If the relatum of instruction n matches either the relatum or the moved digit of the preceding instruction, a repetition benefit ensues (e.g., for "put 4 to the right of 3," as well as for "put 4 to the right of 2"). This prediction is represented by the new predictor *relatum*, which is set to 0 for all conditions in which any of the digits used in instruction $n - 1$, regardless of their role, are repeated as the relatum of instruction n.

The fourth scenario is that the focus starts at the relatum of instruction $n - 1$ and then stretches out in the direction specified by the spatial relation to build a chunk that integrates both the relatum and the moved digits in its new location. Thus, at the end of an instruction such as "put 2 above 3," the focus would hold a chunk that represents the relation "two-above-three" as a single unit. A repetition benefit would be expected if the following instruction n uses both digits again in the same roles, as for instance, "put 2 below 3." This instruction could be followed by rotating the chunk in space without having to unpack it. Thus, the fourth scenario can be captured by the ordered-chunking predictor.

If the fourth scenario is correct, an additional repetition benefit could also be expected if instruction n uses one of the elements included in the chunk as its relatum. For instance, following "put 2 above 3," the next instruction could be "put 4 below 3." The focus of attention could not use the "two-above-three" chunk again; it has to abandon that chunk and start again by searching for the location of the relatum, 3. That relatum, however, is easy to find, compared to a new digit, because it must be close in space to the "two-above-three" chunk. The old chunk therefore serves as a very effective search cue for the new relatum's location. Thus, repeating an element of the chunk built for instruction $n - 1$ as the relatum of instruction n can be expected to be faster than retrieving a new digit's location as the relatum. The repetition benefit from having the next relatum in close vicinity is captured by the predictor relatum introduced above; this predictor codes a benefit for all conditions in which the relatum of instruction n is one of the elements involved in instruction $n - 1$. In the context of the fourth scenario, the predictor relatum is assumed to represent a relatively small effect in addition to an effect of ordered chunking.

Table 5
Predictors for Experiment 3

	Condition						
Argument	rep–rep	rep–swn	swo–swo	swo–swn	swn–rep	swn–swo	swn–swn
Moved (M)	Repeat	Repeat	Other (R)	Other (R)	New	New	New
Relatum (R)	Repeat	New	Other (M)	New	Repeat	Other (M)	New
Predictor							
Serial R–R	0	1	1	1	0	1	1
Serial R–M	1	1	0	1	1	0	1
Ordered chunking	0	1	1	1	1	1	1
Relatum	0	1	0	1	0	0	1
Expand	0	0.5	0	0.5	0.5	0.5	1
Split	0	0.5	1	1	0.5	1	1

Note. The first two indented rows indicate the switching condition of the relatum (R) and the moved (M) digit, respectively, for each of the seven joint conditions at the heads of the columns. The remaining rows represent the predictors, which contain a 1 for the conditions with no repetition benefit, 0.5 for conditions with a small repetition benefit, and 0 for conditions with a large repetition benefit.

In addition to the four predictors representing these four scenarios, we also tested the hypothesis that an expanded focus or a split focus take in both digit locations at the same time. These hypotheses are represented by the *expand* and the *split* predictors, respectively, as in the first two experiments.

Method

Participants. Twenty-one participants (16 women and 5 men) between the ages of 19 and 34 years took part in Experiment 3. All had full-color and normal or corrected-to-normal vision.

Design and materials. The design involved the same seven switch conditions as in the first two experiments, generated by crossing two object-switch variables: (a) switching of the moved digit, with three levels: repetition, switch to the relatum of the previous step, switch to a new digit; and (b) switching of the relatum, with three levels: repetition, switch to the moved digit of the previous step, switch to a new digit. Two of the nine cells of that design resulted in using the same digit as relatum and moved digit and were thus excluded, leaving seven design cells. Each design cell could be realized with four spatial relations, "to the left of," "to the right of," "above," and "below." Thus, the complete design had 28 cells.

A computer program constructed 12 blocks of 14 trials each. A trial consisted of the initial locations of the four digits in the 4 × 4 grid, selected at random with the constraint that no two digits occupy the same cell, and nine updating instructions, selected at random with the constraint that the new location of the moved digit never fall outside the grid and that two digits never occupy the same location. All updating instructions except the first were classified into one of the 28 design cells, and the program constructed trials such that each design cell was repeated four times in each block of 14 trials.

Instructions for all possible combinations of moved digit (1 to 4), relatum (1 to 4), and spatial relation were digitally recorded by a female English speaker. All sound files with the instructions were edited to 2 s of spoken duration by stretching or squeezing the original sounds using the *Transform* function of Cool Edit 2000 software (Audacity Developer Team, 2007).

Procedure. The sequence of events in a trial is illustrated in Figure 10. The word "START" (displayed on the screen for 1 s) initiated the beginning of a trial. After 100 ms, the digits 1 to 4 were presented simultaneously in four random locations in a 4 × 4 grid. Participants had to try to remember which digit was presented in which cell of the grid. Once participants had learned the positions of all four digits, they pressed the space bar to continue, whereupon the digits were erased. Then nine instructions were played (one by one) through headphones, telling participants to move one of the digits to a new position. For instance, if the first instruction was "put 2 above 1," participants would mentally move the 2 to the cell directly above the 1. Participants had to try to remember the new position of the digit, as well as the old positions of the other three digits. Participants were instructed to press the space bar once they were ready to continue with the next instruction. Latency was measured from the end of the spoken instruction to the participant's response. The interval between the response to instruction n and the onset of instruction $n + 1$ was 200 ms. The empty black grid remained on the screen until the response to the last instruction.

After the response to the last instruction in the series, an empty red grid appeared on the screen, and a series of four questions was

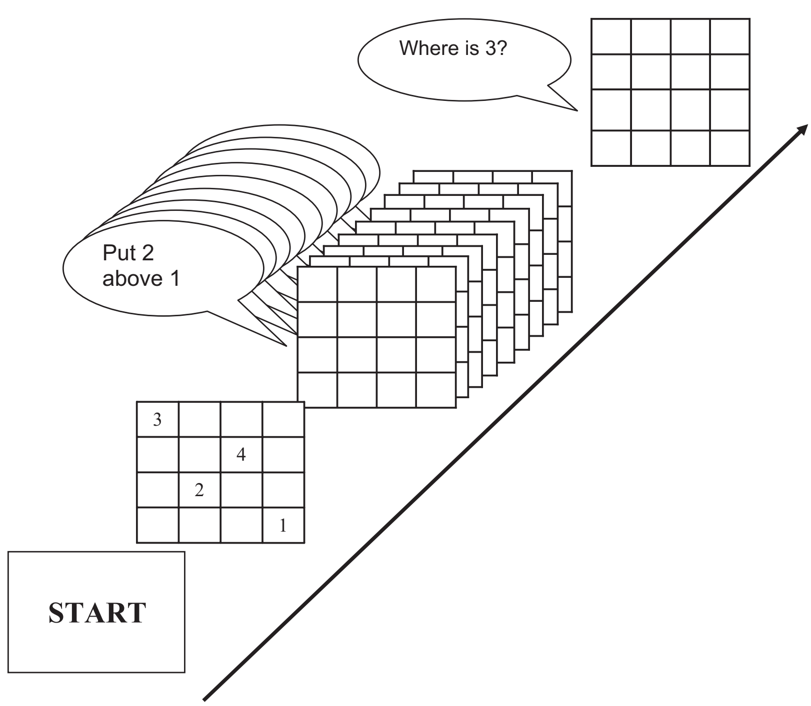

Figure 10. An example of a trial in Experiment 3. Each trial had nine instructions; the first served to initialize the sequence and was not analyzed.

played through the headphones, for example, "Where is 3?" Participants had to click with the mouse on the final position of the digit asked for in the question. The next question was played 200 ms after that response. Following the mouse click to the last question in a trial, a summary of the accuracy for the recall of the final position of the digits (e.g., "3 out of 4 correct") was displayed for 1,500 ms. After an intertrial interval of 200 ms, a new trial began with the word "START" displayed on the screen for 1 s.

There were four experimental sessions of 1 hr each, scheduled on different days. Each session consisted of three blocks of 14 trials each. Because pilot testing showed that the spatial task was more difficult than the arithmetic task, we regarded the full first session as a practice session and excluded it from analysis. In addition, each session started with three practice trials, also excluded from data analysis.

Results

Accuracy could not be assessed for individual updating steps in each trial, as in Experiments 1 and 2, but only for the recall of each digit's final position. On average, participants reported 2.67 (*SD* = 0.91) of the 4 final positions correctly. One participant with exceptionally low accuracy (0.63 out of 4 correct) was excluded (the results did not change qualitatively with this participant included). Limiting analysis to only perfectly correct trials would have eliminated nearly all trials. We therefore placed a cutoff such that at least 3 out of 4 positions must be recalled correctly for a trial to be included. Outliers, defined as in the previous experiments, were removed (2.0%).

The mean RTs for the seven conditions are displayed in Figure 11. The best-fitting model included the ordered-chunking predictor and the relatum predictor; its fixed-effect estimates are shown as asterisks in Figure 11. This model, summarized in Table 6, corresponds to the fourth scenario in the introduction to this experiment: The focus of attention starts implementing each instructed movement by focusing on the relatum, then stretches out to build an ordered chunk that includes the new location of the moved digit. When the next instruction involves both digits involved in that chunk in the same order (i.e., condition rep–rep), a substantial repetition benefit of more than 600 ms ensues because the existing chunk can be reused. When only the next instruction's relatum is one of the digits included in the chunk, the focus cannot use the old chunk, but it can find the relatum that forms the seed for the new chunk easily. The benefit of having the relatum of the new chunk nearby the location of the old chunk is reflected by the relatum predictor. In this experiment it amounted to about 400 ms.

We also tested the other three scenarios discussed in the introduction to this experiment. They all fit much worse than the fourth scenario; the best alternative (the third scenario, with the relatum predictor only) had a fit of 30.0 BIC units worse than the fourth scenario.

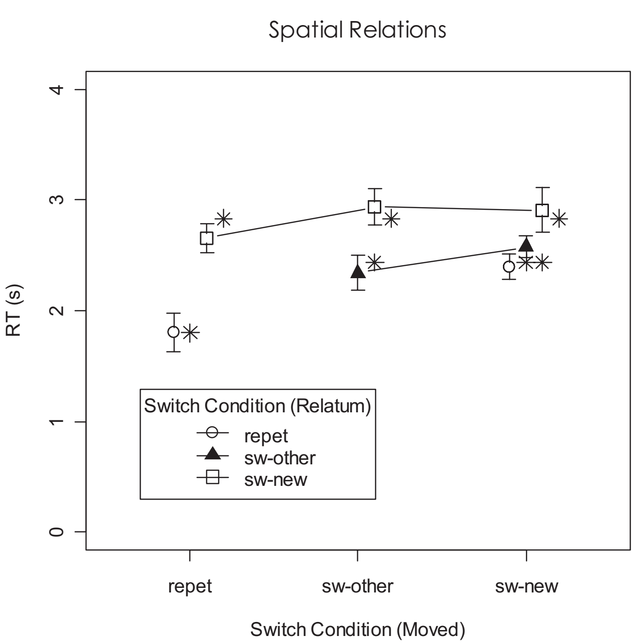

Figure 11. Reaction times (RTs) in Experiment 3 for spatial relations. Error bars are 95% confidence intervals for within-subjects comparisons. Predictions of the best-fitting model are presented as asterisks. repet = repetition; sw-other = switch to other; sw-new = switch to new.

Discussion

The results of Experiment 3 support, once again, the hypothesis that the focus of attention builds chunks to accomplish relational processing. Given the instruction to place a moved digit A in a specific spatial relation with a relatum B, the focus starts selecting the relatum B as a point of reference and uses the information in the relational term (e.g., "to the left of") to find the new location of the moved digit A. Rather than moving to the new location, leaving B behind, the focus stretches out to include both A and B. This does not mean, however, that both digits and their locations are held in the focus as separate entities, as would be the case if the focus expanded. Rather, the two digits and their spatial relation are integrated into an ordered chunk. The evidence supporting ordered chunking over expansion comes from the finding that the repetition benefit is substantially larger when both digits are repeated in their roles (condition rep–rep) than when both are repeated in swapped roles (condition swo–swo) or when only one of them is repeated. In addition, repeating the relatum yielded a substantial repetition benefit even when the moved digit changed, whereas repeating only the moved digit resulted in hardly any repetition benefit. This asymmetry of repetition benefits for the two elements is captured by the relatum predictor. It arises because the present task requires retrieving the current location of the relatum but not the location of the moved digit.

The task used in Experiment 3 is related to tasks of spatial comprehension and spatial reasoning, in which people build spatial mental models from descriptive statements such as "the school is to the left of the hospital" or "on top of the red block is the blue block" (Goodwin & Johnson-Laird, 2005; Huttenlocher, 1968; Mani & Johnson-Laird, 1982). These mental models are constructed in WM (Oberauer, Weidenfeld, & Hörnig, 2006), and their construction is likely to involve processes similar to those used in the present spatial updating task. In particular, each descriptive statement can be interpreted as an instruction to place one entity relative to another entity that serves as its relatum. Hörnig, Oberauer, and Weidenfeld (2005) have shown that this process is easier when the relatum of a statement is already part of the mental

Table 6
Model Statistics for Experiment 3

Predictor	Mean weight (*SD*)	95% CI	L-ratio	ΔBIC	ΔAIC
	Best-fitting model (AIC = 226.7, BIC = 241.4, R^2_{adj} = .93)				
Intercept	1.80 (1.24)	1.22, 2.38			
Ordered chunking	0.64 (0)	0.44, 0.84	35.5	30.5	33.4
Relatum	0.40 (0)	0.26, 0.53	28.0	23.0	26.0

Note. Mean weights are unstandardized regression weights for fixed effects; they reflect the size of the effect of each predictor on the reaction time scale. Their standard deviations are estimates of the random effects (when zero, the random effect was dropped from the model). CI = confidence intervals for the estimation of fixed effects, computed by the *intervals* function of *nmle*; L-ratio = likelihood ratio of complete model to model with effect removed (all $ps < .001$); ΔBIC and ΔAIC = changes in Bayes information criterion (BIC) and Akaike information criterion (AIC), respectively, when removing effect (positive values mean decrease of fit).

model constructed from a preceding statement, compared to when the relatum is a new entity. They called this observation the *relatum = given* principle. For instance, consider the two statements "A is to the left of B" and "C is on top of B." The first statement gives rise to a mental model of the relation between A and B. The second statement is relatively easy to process because the existing model can be incremented by placing C on top of its relatum B. For comparison, consider "A is on the left of B" and "B is below C." Here, the relatum of the second statement is new, and processing is more difficult. Importantly, a processing benefit for the second statement is also observed when its relatum was not the relatum of the first statement, such as in "A is on the left of B" and "C is below A." Thus, a processing benefit is found whenever the relatum of the second statement is already given as part of the first statement, regardless of its role in the first statement. The same principles seem to govern reasoning with temporal relations such as "the train stopped before the conductor fell," and comparative relations such as "Jim is faster than John" (Oberauer, Hörnig, Weidenfeld, & Wilhelm, 2005).

The relatum = given principle governing integration of descriptive spatial statements corresponds exactly to the effect of the relatum predictor in the present experiment: An instruction is processed faster if its relatum is one of the entities used in the preceding instruction, regardless of what role that entity had in the preceding instruction. We feel justified, therefore, to generalize our conclusion about WM processes from the present experiment to the integration of descriptive sentences as studied by Hörnig et al. (2005) and other researchers investigating relational reasoning (e.g., Goodwin & Johnson-Laird, 2005).

Reasoning with relations can thus be described as follows. Given a statement of the form "A is related to B" (where A and B can refer to objects and events and "is related to" can stand for any specific spatial, temporal, or comparative relation), the focus of attention first establishes a representation of the relatum B in WM and then adds A in the prescribed relation to B, thus forming a chunk of the A–B relation. When the next statement uses one of the two already given elements, A or B, as relatum, then the focus can find this element easily and use it as the seed for a new chunk. For instance, a second statement "C is below A" is processed by abandoning the A–B chunk, finding A close by, and then stretching out to the location below A, where C is placed. As a result, the focus now holds a new chunk of the A–C relation. In contrast, when the second statement introduces a new entity as its relatum, such as "A is above C," the focus has to establish C as a new object in a new location in WM. Then the focus stretches out in the direction indicated by the relation, searching for A above C—if A is not found, the focus has to select a new place for C and try again. Eventually, the focus finishes holding a chunk of the A–C relation, but the process leading there is more complicated and time demanding. An alternative strategy is to translate "A is above C" into "C is below A"; this again costs extra time.

General Discussion

Working memory (WM) is often characterized as a system for short-term maintenance and processing of information. Surprisingly little research has been done on how the information temporarily held in WM is processed. The present work asks how the information to be used in cognitive operations is selected from the set of representations currently held in WM. The mechanism selecting representations for cognitive operations is called the focus of attention in WM. The specific question addressed in this work is how the focus of attention selects representations for processes that involve more than one element, such as adding two digits, integrating two words in a sentence, placing one object to the left of another in space, determining the sequence of two action steps, or comparing two entities. Such processes form the basic steps of many complex cognitive tasks such as solving equations, language comprehension, reasoning about relations, action planning, and inductive reasoning. They pose a challenge for any theory that assumes that the focus of attention is strictly limited to holding a single representational unit (i.e., a single chunk) at a time.

We tested a number of alternative models, and mixtures of models, of how the focus of attention selects two elements in WM. The best-fitting mixture of models varied across experiments, but they all shared a model that assumed chunking of two elements into a single unit. Thus, our results converge on the conclusion that the focus of attention uses ad-hoc chunking to select two elements in WM for processing.

Ad-Hoc Chunking and Relational Processing

The empirical signature of chunking in the dual-access paradigm is a repetition benefit that is limited to transitions in which both elements are repeated. This benefit for repetitions of both elements has been observed in all conditions of our experiments. In the experiments with the arithmetic task, the repetition benefit was found regardless of whether the two repeated elements changed their roles in the equations, thus pointing to a free chunk that does not preserve the binding between each digit and its role. In the experiment with the spatial task, the repetition benefit from chunking was limited to repetitions of both digits in the same roles as relatum and as moved digit, respectively. The latter finding points to an ordered chunk in which elements are bound to their roles, and when the roles are swapped, the chunk cannot be reused.

This difference between the arithmetic and the spatial task can be explained by their different demands. In the arithmetic task, the roles of digits in the equation are not strictly necessary to compute the result. This is obvious for addition, but it holds also for the subtraction equations in our experiment, when the following assumption is made: Rather than taking "$x - y$" as input and computing the result, the arithmetic module might take the unordered pair "x, y" as input and return their absolute difference. The sign of the difference is immaterial because participants knew that the results of all equations were limited to the range of 1 to 9. In the spatial task, in contrast, the roles of relatum and moved element are essential for correct placement of the elements: "Put A above B" results in a different arrangement than "put B above A." Therefore, the chunks formed in the arithmetic task did not need to include role bindings, whereas those formed in the spatial task did.

A Narrow Focus of Attention in Working Memory

The chunking hypothesis is compatible with the assumption that the focus of attention is limited to hold one chunk at a time (Garavan, 1998; McElree, 2001; McElree & Dosher, 1989; Oberauer, 2002). Therefore, the present results provide evidence for the assumption of a focus of attention in WM that is limited to a single representational unit, while at the same time demonstrating that a unit does not necessarily correspond to a nominal unit of the task (e.g., digits in the arithmetic task). The WM system apparently has remarkable flexibility in forming ad-hoc chunks. Using a recognition paradigm, McElree (1998) has already provided first evidence that the focus of attention can hold several items when their semantic relatedness supports chunking them (for further supporting evidence, see McElree, 2006).

The present results call into question an alternative view of the focus of attention. In the theory of Cowan, the focus of attention can flexibly expand to hold up to four chunks (Cowan et al., 2005). If that were the case, we should expect that the expand predictor would carry a substantial weight in explaining repetition benefits in the present experiments; this was not what we found.

The lack of empirical support for the expand predictor also speaks against an explanation of repetition benefits (or object-switch costs) in terms of a recency gradient of activation. All previous research on the object-switch effect has used a single-access paradigm. With that paradigm it is not possible to rule out an explanation of the object-switch effect by a recency gradient: The object last accessed might be more available than other objects simply because it was most recently used and therefore is still activated highest. Such an explanation does not require the concept of a focus of attention at all.

With the dual-access paradigm, the recency-based account of object-switch effects can be distinguished from an explanation referring to a focus of attention. If the speed of accessing an object is a function of only its current level of activation, which depends on the recency of its use, then repetition benefits for recently used objects should be independent of each other. That is, if one of the two objects figuring as arguments in the preceding operation is reused in the present operation, a repetition benefit should occur, regardless of whether the other object from the preceding operation is also reused. In other words, there should be no interaction between the repetition status of one argument and the repetition status of the other argument. This purely additive model of repetition benefits is just what the expand predictor represents. The data, in particular those of Experiments 1 and 2, clearly show that repetition benefits are not additive.[7] A repetition benefit was found only when both objects were reused jointly, not when only one of them was repeated. This pattern clearly rules out an explanation of repetition benefits on the basis of a recency gradient alone, and by implication, strongly supports the notion of a focus of attention.

This is not to say that a recency gradient of activation or availability does not exist. There is compelling evidence for a recency gradient of availability for items outside the focus of attention, which is evident both in latencies and accuracies (Oberauer, 2003, 2006; Verhaeghen & Basak, 2005). A full explanation of the speed of access to information in WM will have to include the notion of a focus of attention that speeds up access to the chunk it holds as a whole and a recency gradient of accessibility for information outside the focus. Studies measuring full speed–accuracy trade-off functions (McElree, 2006; McElree & Dosher, 1989) have shown that the repetition benefit for a chunk in the focus of attention reflects an increase in processing rate, whereas the benefit from accessing recent items outside the focus reflects only a higher asymptote of the speed–accuracy trade-off function for more recent items.

Finally, the lack of support for the expand predictor also helps to rule out an explanation of object-switch effects in terms of perceptual priming. In our experiments, as in most previous experiments on the object-switch effect, repeated access to the same object involves cueing with the same cue as before. In Experiments 1 and 2, colors served as cues to digits, and in Experiment 3, spoken digits served as cues to spatial locations. It could be that repetition benefits arise simply because processing of the cues is facilitated by their repetition (Li et al., 2006; for an analogous argument in the task-switching literature see Mayr & Kliegl, 2003). In the dual-access paradigm, repetition benefits from priming of cues, like those from activation of individual objects, should be additive, so that repeating one cue should yield a partial benefit in between repetition of both cues and switching of both cues. This translates into the pattern predicted by the expand model, which was not supported by our data.

[7] At the request of one reviewer, we computed conventional ANOVAs to test the critical interaction. As can be seen in Figure 4, for ordered chunking, an interaction is predicted only for repetition versus switch to a new element; for free chunking, this interaction is also predicted for switch-to-other versus switch-to-new. In all cases, both interactions were significant (smallest $F = 6.7$, largest $p = .018$), with the exception of the interaction for swo versus swn in Experiment 3 ($F = 3.7$, $p = .07$), for which the model analysis identified ordered chunking instead of free chunking. Thus, the ANOVAs are in agreement with the model-based analyses.

A Broader View on the Engine of Thought

Zooming out from the issues discussed so far, we can place the present work in the context of a broader question: What is WM good for, and how does it accomplish its function? WM is the system responsible for holding information available for complex cognitive activities such as language comprehension, planning, and reasoning. All these activities involve the representation and processing of relational information, that is, information on how two or more elements (e.g., objects, events, words) are related to each other. For instance, in language comprehension we constantly relate verbs to subjects and objects and build mental models of what is spoken about by relating agents to actions and their objects, relating objects to each other in space, relating events to each other in time, and making causal connections (Gernsbacher, 1991; Zwaan, Magliano, & Graesser, 1995). Similar processes are involved in deductive reasoning, where the meaning of a set of premises is represented by mental models of their possible referents (Johnson-Laird, 1983). In inductive reasoning tasks, such as the Raven matrices task, finding out the rule governing a set of elements requires comparing two or more elements and building a representation of their relation on the dimension of comparison. Likewise, in analogical reasoning tasks, a structural representation of the base domain must be constructed and mapped on a representation of the target domain (Gentner, 1989).

The reliance on relational representations is the common denominator of those tasks that have been identified as depending highly on WM capacity (Oberauer, Süβ, Wilhelm, & Sander, 2007). Therefore, two minimal requirements for WM as a system to support activities such as language comprehension and reasoning are that it can hold relational representations and that it has a mechanism for processing relations. The embedded three-components framework (Oberauer, 2002, 2007) provides a sketch of a system that meets these requirements; it has motivated the present work, and we use it here to explain how the WM system enables complex cognition.

The framework assumes three components of WM: the *activated part of long-term memory*, the *region of direct access*, and the *focus of attention*. Holding relational representations is assumed to be the function of the direct-access region. This component is responsible for building and maintaining temporary bindings between representations of content elements (e.g., objects, events, persons, words, numbers) and their locations in a cognitive coordinate system (e.g., their location in space and time), as well as bindings between content elements and their roles in schematic structures (e.g., the role of agent in a proposition or the role of effect in a causal schema). In the present Experiments 1 and 2, for instance, temporary bindings between digits and their colors must be established and maintained.[8] In Experiment 3, the digits must be bound to their positions in space. These bindings are maintained in the region of direct access.

The focus of attention has the function of providing selective access to the representations that are to be processed. Thus, the focus is the interface between the contents of WM (i.e., the representations held in the direct-access region) and the processing system. Cognitive operations acting on relational representations must be sensitive to relations and must be able to manipulate relations. Therefore, not only individual elements but also their relations must be selected by the focus of attention as input, and received as output, of operations. This means that the focus must select at least two elements, together with their roles in their relation. For example, in a spatial reasoning task one cognitive operation could be to build a representation of the premise "A is to the left of B." The procedure for this operation needs as input information about the location of B, together with its role as the relatum, and it produces as output information about the locations of both A and B. Thus, the focus starts by accessing the location of B, then stretches out to include the new location of A, and ends with holding both A and B in their respective locations. This outcome is fixed in the direct-access region of WM by binding A to its new location in mental space; once that is done, the focus of attention can select new representations. Likewise, in an inductive reasoning task, one operation would be to compare two objects, A and B. This operation needs as input both A and B, together with the information that, say, B is to serve as the relatum (i.e., the standard of comparison) and A is the compared object. The operation produces as output information about the relation between the two objects, for instance, "A is larger than B." This result would be established in the direct-access region of WM by binding A to a location above B on the size dimension in a mental feature space.

We have seen that the focus of attention uses ad-hoc chunking to accommodate relational information within its limit to hold only a single chunk. Besides offering a way around the limitation of the focus, ad-hoc chunking also enables building and evaluating relations between relations (Halford et al., 1998). For instance, the spatial-location task can be extended to include instructions such as "put A between B and C" (for an extension of the relatum = given principle to reasoning with "between," see Hörnig, Oberauer, & Weidenfeld, 2006). To process this instruction, the focus would form a chunk of the B–C relation and use it as the relatum for A. As a result, a new chunk is formed that represents the whole B–A–C complex. Likewise, the result of one comparison between two objects can in turn be compared to the result of another comparison, resulting in judgments such as, "the size difference between A and B is larger than that between C and D." This kind of reasoning, dear to us all, underlies the evaluation of interactions between experimental variables. Similarly, analogical reasoning involves noticing a similarity relation between two relations, as in, "horse is to cart as motor is to car," or, "the focus of attention in WM relates to cognitive action as visual attention relates to physical action." Thus, chunking affords a way of building representations of higher order relations with a mechanism that does not process anything more complex than a binary relation.

The ad-hoc chunks held in the focus include the objects on which cognitive actions operate, and their relations, but not the information specifying the actions themselves. For instance, in the arithmetic task of Experiments 1 and 2, the two digits are combined into a chunk but the operation sign is not. In the spatial task of Experiment 3, the two object locations are chunked, but the instructed direction in which the focus of attention stretches out from the relatum to find the moved object's location is not part of the chunk. Otherwise, the chunk could not be used again with a repetition benefit when the operation sign, or the spatial direction,

[8] Alternatively, digits could be bound to serial positions, because the colors were always presented in the same order.

is changed. In Experiments 1 and 2, equations with repeated signs were excluded, and in Experiment 3, instructions with repeated directions were rare, and yet we found repetition benefits when both elements in WM were repeated across successive operations; this would not be possible if every change of operation sign, or of movement direction, implied abandoning the currently selected chunk. We conclude that the information guiding the cognitive action is not included in the chunk but rather used by the processing system to specify the appropriate task (i.e., addition or subtraction in the arithmetic task and stretching up, down, left, or right in the spatial task).

Narrow Focus—Narrow Mind?

Why is the focus of attention limited to a single chunk? Wouldn't the WM system be much more efficient if the focus could hold several independent elements at the same time, as described by the expanding-focus model or the split-focus model? The reason for the one-chunk limit could be the function of the focus of attention, which is to select representations for processing. Selection of one representation means exclusion of other candidates. When several independent representations are selected at once, as they would be by an expanded focus, they could be confused with each other or, in the case of distributed neural representations, overlay each other to create a new pattern that does not resemble any of them individually (the so-called "superposition catastrophe"; cf. Bowers, 2002). Even with a split-focus architecture in which two objects are selected but assigned different roles within a focus with two compartments would not avoid confusion and cross-talk completely unless the separation of the two objects is perfect. The brain is a highly interactive system in which complete compartmentalization of representations, in particular when they come from the same category (such as two digits, or two spatial locations within the same grid), is difficult and perhaps impossible. Chunking, in contrast, does not create confusion or superposition problems because a chunk is a single, new representation in which the components do not exist as separate entities (for discussions of how chunking can be accomplished in neural networks, see Halford et al., 1998; Plate, 2003). Thus, the limitation of the focus of attention to a single chunk should not be bemoaned as a capacity limit that restricts our cognitive efficiency but rather as a means for maximizing the efficiency of the selection of contents for processing. Selecting a single representation at the exclusion of others might simply be the cleanest solution to that problem.

The narrow limit on the focus of attention matches a corresponding narrow limit on the side of the processing system: Under most circumstances, the processing system can conduct only one cognitive operation at a time (Pashler, 1994; Pashler, Johnston, & Ruthruff, 2000). Again, this limitation seems to arise not from a structural limitation (i.e., a hard-wired "bottleneck") but reflects an organization of the cognitive system that minimizes cross-talk (Logan & Gordon, 2001; Oberauer & Kliegl, 2004). This implies that in situations where cross-talk is unlikely—for instance, when two cognitive operations are dissimilar and operate on dissimilar inputs—the constraint to do one operation at a time can be relaxed. Indeed, with practice on a dual-task combination of an arithmetic task and a spatial task, a majority of young adults accomplished parallel processing without dual-task costs (Göthe, Oberauer, & Kliegl, 2007; Oberauer & Kliegl, 2004). By analogy, we can speculate that the limit of one chunk at a time in the focus of attention also can be relaxed when the contents to be selected are sufficiently dissimilar to avoid any danger of confusing them (e.g., selecting one digit and one spatial position at the same time) and when participants receive a substantial amount of practice on the task.

What does this theoretical analysis imply for the limits of WM capacity? A focus limited to a single chunk might seem like a prime candidate for explaining why WM capacity is limited. This would be misleading, however. Measures of WM capacity such as the reading-span task (Daneman & Carpenter, 1980) or the operation-span task (Turner & Engle, 1989) reflect the number of elements people can recall after brief retention intervals filled with processing of unrelated material or of the memory items themselves. In the embedded-components framework, the main determinant of the ability to maintain several elements available for processing or recall is the capacity of the direct-access region, that is, its ability to hold the bindings of all these elements to their retrieval cues (in most cases, their list positions). Individual differences in WM capacity reflect primarily differences in the ability to maintain several such bindings simultaneously with little interference between them (Oberauer & Kliegl, 2006; Oberauer et al., 2007). It is not obvious how the limited capacity of the focus of attention would have an impact on people's ability to recall items after an intervening processing episode. As we argued above, it is not clear whether a less limited focus would be advantageous, or rather increase the risk of cross-talk in processing. The narrow focus might well be a feature of optimal design. If this is so, we should not expect to find individual differences in the capacity of the focus, at least not within typical populations, because deviating from it would put efficient processing at risk.

The present work does, however, point to another potential source of individual differences, namely the ability to chunk elements ad hoc for the processing of relations or for the integration of information from several sources. The ability to rapidly chunk the elements that need to be accessed for a given task could be regarded as one aspect of the ability to control attention, which Engle, Kane, and colleagues proposed to be the common factor underlying WM capacity and its correlates (Engle, Kane, & Tuholski, 1999; Kane & Engle, 2002). The dual-access paradigm, and the models associated with it, provide a tool for diagnosing the extent to which people chunk elements in a given dual-access task; this tool can be applied in future work to investigate individual differences in the processes involved in selective access to WM.

To conclude, we propose that WM is a system for making relational representations available for processing. It interacts with a processing system that determines what cognitive operations to apply to the contents of WM. The two systems interact through the focus of attention. By its limit to a single chunk, the focus creates a narrow channel for this interaction. The channel is narrow not because of structural or resource limits but as a means to ensure efficient selection without cross-talk. Despite its being narrow, the channel is open for relational information as long as that information is neatly packed into chunks.

References

Allport, A. (1987). Selection for action: Some behavioral and neurophysiological considerations of attention and action. In H. Heuer & A. F. Sanders (Eds.), *Perspectives on perception and action* (pp. 395–419). Hillsdale, NJ: Erlbaum.

Audacity Developer Team. (2007). *Audacity* (Version 1.3 beta). Available from http://audacity.sourceforge.net

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59,* 390–412.

Baddeley, A. D. (1986). *Working memory.* Oxford, United Kingdom: Clarendon Press.

Bakeman, R., & McArthur, D. (1996). Picturing repeated measures: Comments on Loftus, Morrison, and others. *Behavioral Research Methods, Instruments, & Computers, 28,* 584–589.

Bayes factor. (n.d.). Retrieved March 2, 2008, from Wikipedia: http://en.wikipedia.org/wiki/Bayes_factor

Bowers, J. S. (2002). Challenging the widespread assumption that connectionism and distributed representations go hand-in-hand. *Cognitive Psychology, 45,* 413–445.

Cohen, J. (1995). The earth is round ($p < .05$). *American Psychologist, 49,* 997–1003.

Conway, A. R. A., Kane, M. J., & Engle, R. W. (2003). Working memory capacity and its relation to general intelligence. *Trends in Cognitive Sciences, 7,* 547–552.

Cowan, N. (1988). Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information-processing system. *Psychological Bulletin, 104,* 163–191.

Cowan, N. (1995). *Attention and memory: An integrated framework.* New York: Oxford University Press.

Cowan, N. (2005). *Working memory capacity.* New York: Psychology Press.

Cowan, N., Elliott, E. M., Saults, J. S., Morey, C. C., Mattox, S., Hismjatullina, A., et al. (2005). On the capacity of attention: Its estimation and its role in working memory and cognitive aptitudes. *Cognitive Psychology, 51,* 42–100.

Cumming, G., & Finch, S. (2005). Inference by eye: Confidence intervals and how to read pictures of data. *American Psychologist, 60,* 170–180.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior, 19,* 450–466.

Engle, R. W., Kane, M. J., & Tuholski, S. W. (1999). Individual differences in working memory capacity and what they tell us about controlled attention, general fluid intelligence, and functions of the prefrontal cortex. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 102–134). Cambridge, United Kingdom: Cambridge University Press.

Garavan, H. (1998). Serial attention within working memory. *Memory & Cognition, 26,* 263–276.

Gentner, D. (1989). The mechanisms of analogical learning. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 199–241). Cambridge, United Kingdom: Cambridge University Press.

Gentner, D., & Stevens, A. L. (1983). *Mental models.* Hillsdale, NJ: Erlbaum.

Gernsbacher, M. A. (1991). Cognitive processes and mechanisms in language comprehension: The structure building framework. In G. Bower (Ed.), *The psychology of learning and motivation* (Vol. 27, pp. 217–263). New York: Academic Press.

Gigerenzer, G. (2004). Mindless statistics. *Journal of Socio-Economics, 33,* 587–606.

Gigerenzer, G., Krauss, S., & Vitouch, O. (2004). The null ritual: What you always wanted to know about significance testing but were afraid to ask. In D. Kaplan (Ed.), *The Sage handbook of quantitative methodology for the social sciences* (pp. 391–408). Thousand Oaks, CA: Sage.

Glover, S., & Dixon, P. (2004). Likelihood ratios: A simple and flexible statistic for empirical psychologists. *Psychonomic Bulletin & Review, 11,* 791–806.

Goodwin, G. P., & Johnson-Laird, P. N. (2005). Reasoning about relations. *Psychological Review, 112,* 468–493.

Göthe, K., Oberauer, K., & Kliegl, R. (2007). Age differences in dual-task performance after practice. *Psychology and Aging, 22,* 596–606.

Halford, G. S., Wilson, W. H., & Phillips, S. (1998). Processing capacity defined by relational complexity: Implications for comparative, developmental, and cognitive psychology. *Behavioral and Brain Sciences, 21,* 803–864.

Hayes-Roth, B., & Hayes-Roth, F. (1979). A cognitive model of planning. *Cognitive Science, 3,* 275–318.

Hoffman, L., & Rovine, M. J. (2007). Multilevel models for the experimental psychologist: Foundations and illustrative examples. *Behavior Research Methods, 39,* 101–117.

Hörnig, R., Oberauer, K., & Weidenfeld, A. (2005). Two principles of premise integration in spatial reasoning. *Memory & Cognition, 33,* 131–139.

Hörnig, R., Oberauer, K., & Weidenfeld, A. (2006). Between reasoning. *Quarterly Journal of Experimental Psychology, 59,* 1805–1825.

Howson, C., & Urbach, P. (1993). *Scientific reasoning: The Bayesian method* (2nd ed.). Chicago: Open Court.

Huttenlocher, J. (1968). Constructing spatial images: A strategy in reasoning. *Psychological Review, 75,* 550–560.

Johnson-Laird, P. N. (1983). *Mental models.* Cambridge, United Kingdom: Cambridge University Press.

Jonides, J., Lewis, R. L., Nee, D. E., Lustig, C. A., Berman, M. G., & Moore, K. S. (2008). The mind and brain of short-term memory. *Annual Review of Psychology, 59,* 193–224.

Kane, M. J., & Engle, R. W. (2002). The role of prefrontal cortex in working-memory capacity, executive attention, and general fluid intelligence: An individual-differences perspective. *Psychonomic Bulletin & Review, 9,* 637–671.

Kintsch, W., & van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review, 85,* 363–394.

Kliegl, R. (2007). Toward a perceptual-span theory of distributed processing in reading: A reply to Rayner, Pollatsek, Drieghe, Slattery, and Reichle (2007). *Journal of Experimental Psychology: General, 136,* 530–537.

Koch, I., Prinz, W., & Allport, A. (2005). Involuntary retrieval in alphabet–arithmetic tasks: Task-mixing and task-switching costs. *Psychological Research, 69,* 252–261.

Lepsien, J., & Nobre, A. (2006). Cognitive control of attention in the human brain: Insights from orienting attention to mental representations. *Brain Research, 1105,* 20–31.

Lewandowsky, S., & Farrell, S. (2008). Short-term memory: New data and a model. In B. H. Ross (Ed.), *The psychology of learning and motivation* (Vol. 49, pp.1–48). London: Elsevier.

Li, Z., Bao, M., Chen, X., Zhang, D. R., Han, S., He, S., et al. (2006). Attention shift in human verbal working memory: Priming contribution and dynamic brain activation. *Brain Research, 1078,* 132–142.

Logan, G. D., & Gordon, R. D. (2001). Executive control of visual attention in dual-task situations. *Psychological Review, 108,* 393–434.

Mani, K., & Johnson-Laird, P. N. (1982). The mental representation of spatial descriptions. *Memory & Cognition, 10,* 181–187.

Mayr, U., & Kliegl, R. (2003). Differential effects of cue changes and task changes on task-set selection costs. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29,* 362–372.

McElree, B. (1998). Attended and non-attended states in working memory: Accessing categorized structures. *Journal of Memory and Language, 38,* 225–252.

McElree, B. (2001). Working memory and the focus of attention. *Journal* *of Experimental Psychology: Learning, Memory, and Cognition, 27,* 817–835.

McElree, B. (2006). Accessing recent events. In B. H. Ross (Ed.), *The psychology of learning and motivation* (Vol. 46, pp. 155–200). San Diego, CA: Academic Press.

McElree, B., & Dosher, B. A. (1989). Serial position and set size in short-term memory: The time course of recognition. *Journal of Experimental Psychology: General, 118,* 346–373.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63,* 81–97.

Myung, I. J. (2003). Tutorial on maximum likelihood estimation. *Journal of Mathematical Psychology, 47,* 90–100.

Oberauer, K. (2002). Access to information in working memory: Exploring the focus of attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28,* 411–421.

Oberauer, K. (2003). Selective attention to elements in working memory. *Experimental Psychology, 50,* 257–269.

Oberauer, K. (2006). Is the focus of attention in working memory expanded through practice? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32,* 197–214.

Oberauer, K. (2007). Activation, binding, and selective access: An embedded three-component framework for working memory. In N. Osaka, R. Logie, & M. D'Esposito (Eds.), *The cognitive neuroscience of working memory* (pp. 351–368). Oxford, United Kingdom: Oxford University Press.

Oberauer, K., Hörnig, R., Weidenfeld, A., & Wilhelm, O. (2005). Effects of directionality in deductive reasoning: II. Premise integration and conclusion evaluation. *Quarterly Journal of Experimental Psychology, 58*(A), 1225–1247.

Oberauer, K., & Kliegl, R. (2004). Simultaneous execution of two cognitive operations: Evidence from a continuous updating paradigm. *Journal of Experimental Psychology: Human Perception and Performance, 30,* 689–707.

Oberauer, K., & Kliegl, R. (2006). A formal model of capacity limits in working memory. *Journal of Memory and Language, 55,* 601–626.

Oberauer, K., Süβ, H.-M., Wilhelm, O., & Sander, N. (2007). Individual differences in working memory capacity and reasoning ability. In A. R. A. Conway, C. Jarrold, M. J. Kane, A. Miyake, & J. N. Towse (Eds.), *Variation in working memory* (pp. 49–75). New York: Oxford University Press.

Oberauer, K., Weidenfeld, A., & Hörnig, R. (2006). Working memory capacity and the construction of spatial mental models in comprehension and deductive reasoning. *Quarterly Journal of Experimental Psychology, 59,* 426–447.

Pashler, H. (1994). Dual-task interference in simple tasks: Data and theory. *Psychological Bulletin, 116,* 220–244.

Pashler, H., Johnston, J. C., & Ruthruff, E. (2000). Attention and performance. *Annual Review of Psychology, 52,* 629–651.

Pinheiro, J. C., & Bates, D. M. (2000). *Mixed-effect models in S and S-Plus.* Berlin: Springer.

Pinheiro, J. C., Bates, D. M., DebRoy, S., & Sarkar, D. (2005). *nlme: Linear and nonlinear mixed effects models. R package* (Version 3.1–73). Available from http://cran.r-project.org/web/packages/nlme/index.html

Plate, T. A. (2003). Convolution-based memory models. In L. Nadel (Ed.), *Encyclopedia of cognitive science* (pp. 824–828). London: Nature.

Quené, H., & van den Bergh, H. (2008). Examples of mixed-effects modeling with crossed random effects and with binomial data. *Journal of Memory and Language, 59,* 413–425.

R Development Core Team. (2005). *R: A language and environment for statistical computing.* Vienna: R Foundation for Statistical Computing. Available from http://www.R-project.org

Sternberg, S. (1969). Memory scanning: Mental processes revealed by reaction-time experiments. *American Scientist, 57,* 421–457.

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language, 28,* 127–154.

Verhaeghen, P., & Basak, C. (2005). Ageing and switching of the focus of attention in working memory: Results from a modified *N*-back task. *Quarterly Journal of Experimental Psychology, 58*(A), 134–154.

Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems of *p* values. *Psychonomic Bulletin & Review, 14,* 779–804.

Waszak, F., Hommel, B., & Allport, A. (2003). Task-switching and long-term priming: Role of episodic stimulus-task bindings in task-shift costs. *Cognitive Psychology, 46,* 361–413.

Zwaan, R. A., Magliano, J. P., & Graesser, A. C. (1995). Dimensions of situation model construction in narrative comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 386–397.

Received April 4, 2008
Revision received November 10, 2008
Accepted November 13, 2008 ■