

The Contribution of Working Memory to Fluid Reasoning: Capacity, Control, or Both?

Adam Chuderski

Jagiellonian University in Krakow

Edward Nęcka

Jagiellonian University in Krakow and Warsaw School of
Social Sciences and Humanities

Fluid reasoning shares a large part of its variance with working memory capacity (WMC). The literature on working memory (WM) suggests that the capacity of the focus of attention responsible for simultaneous maintenance and integration of information within WM, as well as the effectiveness of executive control exerted over WM, determines individual variation in both WMC and reasoning. In 6 experiments, we used a modified *n*-back task to test the amount of variance in reasoning that is accounted for by each of these 2 theoretical constructs. The capacity of the focus accounted for up to 62% of variance in fluid reasoning, while the recognition of stimuli encoded outside of the focus was not related to reasoning ability. Executive control, measured as the ability to reject distractors identical to targets but presented in improper contexts, accounted for up to 13% of reasoning variance. Multiple analyses indicated that capacity and control predicted non-overlapping amounts of variance in reasoning.

Keywords: working memory, attentional capacity, executive control, fluid reasoning, *n*-back task

Fluid reasoning tests are believed to give a good measure of general fluid intelligence (Gf),¹ which reflects the inter-individually varied but intra-individually constant unacculturated ability for abstract reasoning on novel material. In terms of the material's novelty, Gf differs from crystallized intelligence (Gc), which involves the effective application of previously acquired knowledge (Cattell, 1971). Gf is closely related to the concept of general intellectual ability (Spearman's *g*; Gustafsson, 1984), and both Gf and *g* factors are significant predictors of professional and personal success in human life (Gottfredson, 1997).

Working memory (WM) is a basic cognitive mechanism responsible for the active maintenance of information for its ongoing processing. The construct of WM, introduced to psychology by Baddeley and Hitch (1974), refers to a kind of cognitive "engine" involved in various mental processes. It is usually believed to be responsible for holding and manipulating temporary solutions, structures, subgoals, or subproducts of thinking, before the final

result is reached. Its most important feature is its severely limited capacity (Cowan, 2001; Daneman & Carpenter, 1980).

Over the last 20 years numerous studies (e.g., Colom, Abad, Rebollo, & Shih, 2005; Conway, Cowan, Bunting, Theriault, & Minkoff, 2002; Engle, Tuholski, Laughlin, & Conway, 1999; Kane et al., 2004; Kyllonen & Cristal, 1990; Martínez et al., 2011; Süß, Oberauer, Wittmann, Wilhelm, & Schulze, 2002) have proved that fluid reasoning is strongly correlated with working memory capacity (WMC). Analyses of the results of multiple psychometric studies estimate that WMC and Gf share from 50% (Kane, Hambrick, & Conway, 2005) to more than 70% (Oberauer, Schultze, Wilhelm, & Süß, 2005) of variance, thus making WMC the strongest single predictor of fluid reasoning. Consequently, most WM researchers (e.g., Conway, Kane, & Engle, 2003; Oberauer, Süß, Wilhelm, & Sander, 2007) purport that the cognitive processes underlying and constraining WMC also determine the efficiency of fluid reasoning captured by Gf tests (but see Ackerman, Beier, & Boyle, 2005, for a contrasting opinion). So, an important goal for cognitive science is to understand the cognitive and biological mechanisms of WM and explain why people who excel in working memory tasks also easily solve demanding intellectual tests, while others fail in both.

Two main strands of theoretical proposals considering the WMC-Gf link are most influential. Proponents of the first ("a capacity approach") suggest that both WMC and reasoning depend on the amount of information that can be maintained and integrated within WM by some kind of attentional capacity. Researchers taking the other stance ("a control approach") believe that both WMC and Gf factors are determined by the efficiency of control

Adam Chuderski, Institute of Philosophy, Jagiellonian University in Krakow, Krakow, Poland; Edward Nęcka, Institute of Psychology, Jagiellonian University in Krakow, and Warsaw School of Social Sciences and Humanities, Warsaw, Poland.

This research was supported by Grants N106 2857/32 (Experiments 1 and 2) and N106 417140 (Experiments 3–6), the latter funded by the National Science Centre of Poland. Portions of this research (Experiments 1 and 2) were presented during the 14th meeting of the European Society for Cognitive Psychology, Leiden, the Netherlands, 2005, and the 29th International Congress of Psychology, Berlin, Germany, 2008. We thank Maciej Taraday, Anna Anzulewicz, Krzysztof Cipora, Dominika Czajak, and Katarzyna Raczy for their help in conducting the experiments. We are also grateful to Mike Timberlake for correcting the text.

Correspondence concerning this article should be addressed to Adam Chuderski, Institute of Philosophy, Jagiellonian University, Grodzka 52, 31-044 Krakow, Poland. E-mail: adam.chuderski@gmail.com

¹ Keeping in mind the fact that *fluid intelligence/Gf* constitute rather theoretical/statistical concepts while *fluid reasoning* and *reasoning ability* refer more to a cognitive process/task/faculty involving Gf, we use these four terms interchangeably throughout the text.

over the information kept in WM. Both the capacity and the control approaches enjoy considerable empirical and theoretical support.

However, most of the above cited studies on the Gf–WM link have relied on batteries of relatively complex tasks (e.g., so-called complex span tasks that combine memory storage with some extra processing; Daneman & Carpenter, 1980). Therefore, it is difficult to interpret unequivocally the observed links as reflecting capacity, control, or both. The present research attempts to prove that comparably high correlations between Gf and WM measures may also be observed with a much simpler task, and the various scores on that task may be better interpretable in terms of either capacity or control. Moreover, the crucial methodological aspect of this study is an experimental manipulation regarding the task, which makes WM–Gf correlations appear and disappear. Because many cognitive tasks correlate with Gf due to the phenomenon of positive manifold (Van Der Maas et al., 2006) or coincidence of task requirements (e.g., task complexity, instruction, need for vigilance, novelty of situation), the identification of cognitive mechanisms generating such correlations is difficult. If by application of a specific experimental manipulation we are able to control the strength of correlation between WM task performance and Gf, we can infer that this very manipulated variable reflects the specific aspects of a WM mechanism that are related to Gf (also see Hambrick, Kane, & Engle, 2005).

The general aim of the present article is to use a simple WM task and the manipulations influencing the strength of the WMC–Gf correlation in order to simultaneously examine the involvement of attentional capacity and control processes in operation of WM and to test if both the capacity and control approaches to the explanation of fluid reasoning can be supported.

Cognitive Basis of Fluid Reasoning

The idea that some attentional capacity determines intelligence is obviously not a new one. Its roots may be found in Spearman's (1927) "mental energy" metaphor of the *g* factor, which described it as a kind of resource that is supposed to saturate all mental activities to some extent. The most thorough examination of attentional limitations within WM was carried out by Cowan (1995, 2001). On the basis of evidence from behavioral, psychometrical, developmental, neurobiological, and formal modeling studies, he argued that working memory capacity is determined by the limit of WM's *focus of attention*, also referred to as *primary memory capacity* or *attentional scope*. The focus is structurally distinct from the other mechanism contributing to WM operation, namely, the *activated part of long-term memory* (LTM), also called *secondary memory*. Capacity of the focus is defined as the number of attentionally activated chunks of information easily available for direct use by ongoing cognitive processes. The proper estimation of the number of chunks that attention can span during a given experimental situation depends on the reduction of chunk associations and the elimination of mnemonic strategies, which exploit activated LTM and can artificially inflate the amount of available information. According to Cowan (2001), the average capacity of the focus in healthy adults is four chunks, although it varies among people from two to six chunks, and an individual's span is closely related to Gf level. Two studies by Cowan (Cowan et al., 2005; Cowan, Morey, Chen, & Bunting, 2007) confirmed the close

relationship between capacity estimates and intelligence and maturation.

Duncan et al. (2008) also proposed that the Gf level results from some limited capacity, but they explain the nature of this capacity in a somehow broader sense. According to them, WMC does not reveal the capacity limit of a particular storage buffer, but rather it reflects the total capacity of the cognitive system to code multiple demands and components of a novel task. When such a task is complex, for instance, while working on difficult fluid reasoning test items, its different components compete for attention and the most vulnerable ones may be lost, leading eventually to some aspects or goals regarding the task being neglected (Duncan, Emslie, Williams, Johnson, & Freer, 1996).

Halford and his collaborators (Halford, Cowan, & Andrews, 2007; Halford, Wilson, & Phillips, 1998) agreed that adult humans can process about four independent entities in parallel but disagree that this limitation is the sole storage constraint. The authors proposed that the nature of WM capacity limitation consists of the maximum number of processing dimensions (objects, variables, etc.) that can be simultaneously interrelated. Such a number defines the relational complexity imposed by a task. The higher this relational complexity, the more complex representations have to be formed in WM. In case of difficult relational problems used in fluid reasoning tests, the size of representation can easily exceed the available capacity of the majority of people. In accordance with Halford's and Hummel and Holyoak's ideas, Oberauer et al. (2007) proposed that reasoning ability is determined by the number of elements that can be simultaneously bound to a task-relevant structure, such as positions in a mental array during some spatial task or time tags during a serial recall task. These elements can be directly accessible within an active component of WM. Variance in the capacity for such bindings is reflected by WM measures. Additionally, such a capacity determines success in reasoning tasks, because they require the binding of numerous elements into an appropriate mental structure, as well as quick and accurate access to this structure when necessary (Oberauer et al., 2007; Oberauer, Süß, Wilhelm, & Wittman, 2008).

Executive control (or *cognitive control*) processes are believed to be responsible for the organization and coordination of other mental states and processes in accordance with the internal goals of an individual (Monsell & Driver, 2000). Numerous studies have shown significant correlations between fluid reasoning and the indices of cognitive control obtained from various tasks, for example involving updating (Friedman et al., 2006; Gray, Chabris, & Braver, 2003; Salthouse, 2005); the inhibition of salient distractors, unwanted thoughts, or prepotent responses (Brewin & Beaton, 2002; Dempster & Corkill, 1999; Gray et al., 2003; Nęcka, 1999); and dual-task coordination (Ben-Shakhar & Sheffer, 2001; Chuderski & Nęcka, 2010). Although there is no consensus yet concerning how many distinct types of executive processes really exist (see Collette & Van der Linden, 2002; Engle & Kane, 2004; Friedman & Miyake, 2004; Miyake, Friedman, Emerson, Witzki, & Howerter, 2000; Nigg, 2000), nor what the neurobiological mechanisms of control look like (Braver, Gray, & Burgess, 2007; Duncan et al., 2000), the proponents of the control approach agree that the quality of human executive control plays the crucial role in higher cognitive processes.

The most influential theory about the link between executive control, WMC, and fluid intelligence is that of Engle, Kane,

Conway, and their collaborators (Engle & Kane, 2004; Kane & Engle, 2002). According to these authors, an executive process determining an individual's effectiveness in both WM and reasoning tasks relies on the proper use of domain-general attentional control, consisting of focusing attention on crucial task-relevant information, rather than spanning attention on all available information. Such a process allows for goal-directed behavior, especially under interference, distraction, or conflict. The proposed link between attention control and WMC has been verified by numerous executive tests, for instance proving that high-WMC individuals, compared to low-WMC individuals, were faster and more accurate on antisaccades (Kane, Bleckley, Conway, & Engle, 2001; Unsworth, Schrock, & Engle, 2004), produced smaller error rates in incongruent trials using a high-congruent version of the Stroop test (Kane & Engle, 2003) and the flanker task (Heitz & Engle, 2007), and more effectively suppressed distractors in a dichotic listening task (Conway, Cowan, & Bunting, 2001).

Most studies that related attentional control to fluid intelligence exploited latent variable modeling. Their results indicated that path coefficients between latent variables reflecting Gf and complex span tasks were equal to $r = .49$ (Engle et al., 1999), $r = .52$ (Kane et al., 2004), and could even reach $r = .60$ (Conway et al., 2002). However, these data should be interpreted with caution, as calculation of the control latent variable required additional assumptions. For example, Kane et al. (2004) assumed that this variable represented the variance common for both simple and complex span tasks, while in the two other studies cited, this variable reflected the variance unique to complex span tasks (i.e., after storage variance had been partialled out).

Additional support for the control approach comes from two studies showing that the Gf–WMC correlation does not depend on the memory load imposed by WM tasks (Salthouse & Pink, 2008; Unsworth & Engle, 2005). In the latter study, correlation coefficients between WM task performance and Gf were similar for two-item and seven-item memory set size conditions, suggesting that it was not the storage capacity factor that determined the value of these correlation coefficients.

Also some neuroimaging studies (Burgess, Braver, & Gray, 2006; Burgess, Gray, Conway, & Braver, 2011; Gray et al., 2003) suggested a strong correlation between reasoning scores and the neuronal activity of the postulated control mechanisms, like the prefrontal or anterior cingulate cortices, during performance in WM tasks. Also electroencephalography (EEG) data have shown that high-WMC brains are better than low-WMC ones in filtering irrelevant items out of WM (Vogel, McCollough, & Machizawa, 2005) and the latter more slowly recover from attentional capture (Fukuda & Vogel, 2011).

A promising research examines the joint contribution of both capacity and control limitations to fluid reasoning. For example, Cowan et al. (2007) proposed that attentional mechanism is flexible, and depending on task requirements it can zoom out to span as many items as possible (e.g., in a recall task), or it can zoom in and focus on a single goal in order to protect it against distraction (e.g., in a task requiring resolving interference). Cowan, Fristoe, Elliott, Brunner, and Saults (2006) showed that two measures, one loading the scope of attention and the other capturing attentional control, shared 12% of variance in intelligence, although scope and control contributed separately to an additional 15% and 10%, respectively. Furthermore, two studies demonstrated the interac-

tive effects of memory load and cross-mapping interference on reasoning (e.g., Cho, Holyoak, & Cannon, 2007; Chuderska, 2010). All these results suggest that the partially overlapping mechanisms responsible for both control and capacity may be the cornerstone of fluid reasoning (see also Schweizer, Moosbrugger, & Goldhammer, 2005).

Measurement of Working Memory

The important issue considers how to properly measure the effectiveness of WM. Traditional short-term memory (STM) tasks have long been believed to mostly involve automatic storage mechanisms (e.g., the phonological loop) and yield low correlations with reasoning (see Unsworth & Engle, 2007). So, WMC was often measured with various versions of complex span tasks (Daneman & Carpenter, 1980). Although these tasks have useful psychometric properties and predict Gf (Conway et al., 2005), as noted in the introduction, they are too complex to be useful tools for understanding how WM mechanisms underlie reasoning. Fortunately, recent studies have shown that also some simple WM tasks can be proper measures of WM, substantially correlating with reasoning. For example, Oberauer, Süß, Schulze, Wilhelm, and Wittmann (2000) demonstrated that simple span tasks, which involve memorizing spatial relations, revealed both high loadings on WM factor and WM–Gf correlations comparable to those obtained with complex span tasks. Kane et al. (2004) showed that tasks requiring only the storage of spatial information correlated stronger with Gf measures than did verbal complex spans. Even simple verbal spans correlated substantially with reasoning, provided that memory load was high enough (Unsworth & Engle, 2006) or rehearsal and chunking were successfully blocked (Cowan et al., 2005).

Two relatively simple WM tasks, which successfully predicted intellectual abilities (Cowan et al., 2005, 2007; Friedman et al., 2006; Gray et al., 2003; Hockey & Geffen, 2004; Kane, Conway, Miura, & Colflesh, 2007; Roberts & Gibson, 2002; Salthouse, 2005; Schmiedek, Hildebrandt, Lövdén, Wilhelm, & Lindenberger, 2009), are an *n*-back task (Kirchner, 1958; Mackworth, 1959) and a running memory task (Pollack, Johnson, & Knaff, 1959). The former requires matching the current item in a continuous stream of stimuli with an item that occurred *n* items ago, while the latter requires recall of the most recent items from a serially presented list of unpredictable length. In the present study, we decided to design a task that combines advantages of both these paradigms.

Although the validity of the *n*-back task as a WM measure has been questioned, as it was unrelated to complex span tasks and both tasks accounted for unique amounts of Gf variance (Kane et al., 2007; Roberts & Gibson, 2002), that fact might simply result from superficial differences in both tasks' requirements (i.e., recognition vs. recall) or the use of single tasks. When either the recall version of the *n*-back task was used (Shelton, Metzger, & Elliott, 2007) or complex spans and *n*-back scores from more than one task were aggregated (Shamosh et al., 2008), they shared a quarter of variance. Moreover, in two recent studies, latent variables representing memory updating, which included *n*-back task scores, were related to variables reflecting either complex-span (Schmiedek et al., 2009) or simple-span (Chuderski, Taraday, Nęcka, & Smoleń, 2012) tasks, and the observed correlation was close to

unity. In our opinion, the most “non-WM” feature of the standard n -back task is the requirement of continuous responding after each stimulus presentation. This makes the n -back procedure more a decisional than a WM task. We believe that even a standard recognition version of the n -back task can be a suitable measure of WM, provided that it is properly modified in such a way that participants do not have to engage in continuous decision making, so they can focus on processes specific to WM, like memory encoding, updating, and searching.

So, in our study we used a modified version of the n -back task, which involved relatively rare responding as in the running memory task, but—unlike the latter task—which relied on a recognition paradigm (i.e., there was no interference from recall procedure). Such a task may be administered in two ways (McElree, 2001). Within a so-called inclusion condition, a participant is asked to retain in memory n the most recent stimuli out of a stream of stimuli and to indicate if a current stimulus (a target) in the stream matches any of these n stimuli. By increasing n values of item repetitions, the load on attentional capacity can be intensified until its limit is exceeded. Cowan (2001, p. 89) noticed that such a task could appropriately be taken as a measure of capacity limits, and in the present article we aimed to test that expectation. Within a so-called exclusion condition, a participant is asked to retain n of the most recent stimuli but to indicate only when a target matches the n -back stimulus. Increasing n also increases the load on attentional capacity, but additionally such a task enables the imposition of a significant load on control mechanisms. This goal is reached by presenting distractor items (so-called lures), which are identical to items placed at positions different than n . Lures activate the tendency for accepting them (i.e., committing false alarms), which must be suppressed. So, the false alarm rate is commonly believed to be an index of the (in)efficiency of control (e.g., Burgess et al., 2011; Gray et al., 2003). Thus, using the exclusion version, the contribution of both target and lure conditions to the prediction of reasoning scores can be compared. This was also examined in the present article.

Goals and Rationale of the Study

Assuming that WM is a simpler cognitive mechanism underlying a more complex process of reasoning, the specific goals of the present article may be defined as the search for both capacity limits and executive control constraints of WM that contribute to the effectiveness of fluid reasoning. In order to achieve this goal, two questions should be precisely answered.

The first question concerns the kind of WM capacity related to reasoning: whether it is the total capacity of WM or the capacity of a particular part of WM (the focus of attention). We assumed that participants would often maintain a small number of recent items in their focus of attention, while less recent items would be usually moved outside the focus (i.e., to the activated part of LTM). Taking into consideration the results of Cowan (2001) and Usher, Cohen, Haarmann, and Horn (2001), who estimated the attentional capacity to around three or four items, we expected that participants would often keep within their foci of attention items up to the 2- or 3-back position (i.e., the currently presented stimulus, the 1- and 2-back items, and sometimes the 3-back item), but they would rarely be able to hold the 4- or 5-back items (n s larger than five were not examined as they usually lead to floor effects) within

their attention, because it would require the capacity of five or six items, respectively. Thus, if it is the higher capacity of the (already very limited) focus of attention that makes people more intelligent, reasoning scores and target hit rates will be more strongly (positively) correlated at small n positions (especially at the 2- and 3-back) than at large n positions (i.e., 4- and 5-back). Alternatively, if fluid ability depends on the total working memory capacity expressed as the maximum number of items that can be retrieved from more than one memory area (see Duncan et al., 2008; Unsworth & Engle, 2006, 2007), then one may expect the strongest correlations in cases of large values of n . Another possible result is that no effect of n factor on the Gf–WM correlation would be observed, suggesting that more intelligent participants outperform less intelligent ones generally, for reasons that the present study would be unable to identify decisively but that could be in concord with the control approach (Unsworth & Engle, 2006).

The empirical evidence regarding WM–Gf correlations as the function of memory load seems to be ambiguous. Some data show that significant correlations between WM task performance and fluid reasoning appear only when memory load is high. Unsworth and Engle (2006) analyzed correlations between Raven’s score and simple spans at different memory loads and found that correlations increased with rising memory sets, from $r = .12$, in case of a two-item memory load, to $r = .45$, for loads of seven items. In cases of tasks more similar to ours, Hockey and Geffen (2004) used the visuospatial n -back task with 0- (i.e., matching predefined target), 1-, 2-, and 3-back exclusion conditions. The authors observed increasing WM–reasoning correlations (r s from non-significant to .27) as a function of n . However, their result could be influenced by the ceiling effect (94% correct) for 0-back and 1-back items. Within the letter n -back task, Kane et al. (2007) observed significant correlations with Raven’s test only for the 3-back exclusion condition (as did Gray et al., 2003) but not for the 2-back one. Two studies reported opposite findings. Salthouse (2005) found significant correlations with Gf for 0-, 1-, and 2-back conditions of a digit n -back task (he did not test higher values of n). When the running memory task was applied, the recall from the 2-back position correlated significantly with Gf, while the result from the 3-back position did not. However, Salthouse’s data may have been caused by the fact that his research involved elderly participants (up to 95 years old). Also Friedman et al. (2006) reported a significant correlation ($r = .28$) with Gf for the spatial 2-back task.

In our pilot study (Chuderski & Chuderska, 2009), with the n -back task and an analogical reasoning test, we found the largest difference in WM scores between high-Gf and low-Gf participants in case of the values of n equaling one, two, and three, but not in case of the value of n equal to four. Therefore, we expect to obtain a similar result in the present study. This outcome would support the hypothesis that the effectiveness of reasoning is related to some very capacity-limited part of WM (most probably the focus of attention), but not to the total capacity of WM.

The second question regards whether executive control over WM, which is known to be significantly related to reasoning (e.g., Burgess et al., 2011; Gray et al., 2003), contributes to Gf jointly with WM capacity, with them both predicting overlapping amounts of variance in reasoning, or whether they contribute independently, predicting distinct amounts of Gf variance. If con-

trol and capacity contribute to shared variance, and accepting the aforementioned assumption that differences in Gf-related capacity are captured only by small n conditions, then one should expect stronger (negative) correlation between Gf and the false alarm rate in cases of these n conditions than in cases of larger values of n . This would mean that larger capacity somehow facilitates lure rejection, simply because information that a lure is at a position different than n would be more precisely represented in the focus of attention than outside the focus. On the contrary, if control and capacity independently contribute to reasoning then one should expect no effect of n on the correlation between false alarm rate and Gf. By systematically manipulating the n and target/lure conditions of our n -back task, we aimed to test these two opposing hypotheses on (joint vs. independent) contribution to fluid reasoning of WM capacity and control over WM.

Furthermore, the complex pattern of correlations between Gf and WM measures—if found—should reveal not only the cognitive basis of fluid reasoning, but it may tell us something about the structure of WM itself. If some aspects of WM correlate with reasoning but some others do not, then there are good arguments for functional dissociation of these aspects.

We present six experiments. The first three experiments use the inclusion version of the n -back task and investigate the effect of n on the strength of correlation between the target hit rates and reasoning scores. These experiments aimed to provide an answer to the first question. The next three experiments exploit the exclusion version in order to, first, generalize that answer onto this version and, second, to test the pattern of correlations among reasoning scores and target hits and false alarms for varied n positions, and so to provide the answer to the second question.

Experiment 1

The first experiment served to validate our modified n -back task by testing if there would be any effect of increased n on the task performance. Regarding fluid reasoning, we attempted to observe how the value of n would influence the correlation between reasoning and that performance.

We adopted the modified inclusion n -back task from the study by McElree (2001). In his study, McElree used sequences of six- to 15-letter stimuli; a fast presentation rate (900 ms per stimulus); 1-, 2-, and 3-back conditions; and a two-choice response set. The target was always the last stimulus in a sequence of an unpredictable length. We made two major changes to the task to make it more suitable for our purposes. First, in order to capture the process of information maintenance unaffected by mnemonic strategies, we used two-digit numbers as stimuli, because their rehearsal takes longer and their chunking is more difficult. Second, we applied the go/no-go methodology: Participants were instructed to respond only when they detected item repetition and to refrain from responding when the current stimulus did not match any recent item. As a result, participants did not have to respond in every trial and so they could focus on processes specific to WM.

Method

Participants. A total of 101 students recruited from several colleges in Lodz, Poland, participated in the study (44 females; M age = 21.44 years, SD = 2.11; age ranged from 18 to 34 years).

For a 2-hr session each person received the equivalent of about €5 in Polish zloty plus a CD-ROM encyclopedia. In this and all subsequent experiments, all participants reported normal or corrected-to-normal vision, they were instructed to take the most comfortable sitting position during the experiment, they completed WM and Gf tests in groups of two to five people, and during the computerized task they were equipped with headphones.

Materials and procedure. Stimuli were 72 out of 90 possible two-digit numbers, excluding nine numbers containing zero as well as nine palindromes. Each stimulus was 1.5×1.0 cm in size and was presented in the center of a PC computer screen, in black on a green background. In each trial, a sequence of 4 to 12 stimuli (randomly) was serially displayed for 1,000 ms each. The last stimulus in a sequence (a target) was always identical to an item presented randomly two, three, or four (i.e., the values of n) items ago. There was only one such repetition in a given trial. After the last item in each sequence had been displayed, a black square 1.5×1.0 cm in size was presented for 1,000 ms, in order to inform participants that the recent stimulus had been a target and that the next trial was about to start. Each session consisted of nine trials, three trials at random per each n value, starting with a fixation point presented for 1,000 ms. All nine targets differed from one another, and none of the remaining 60 items could match any other stimulus in the whole session. In each session, a presentation of one trial per each possible sequence length (i.e., in the 4–12 range) was intended. However, due to an error in the software controlling the experiment, there was one additional six-stimulus trial, but there was no nine-stimulus trial. The sequence of trial lengths was also random. There were 10 sessions in the test, all followed by short breaks. Two training sessions preceded the experiment.

The participants were instructed to track the four most recent numbers and to press the space bar only when they detected that the current item was identical to an item presented two, three, or four items ago. They were required to refrain from responding in any other case. Only responses that occurred more than 200 ms after a target appeared on the screen and within 200 ms of a target disappearing from the screen were accepted as correct hits. Participants' reactions out of this time frame, together with reactions to non-repeated items, were treated as errors. Errors were signaled with a beep sound.

The inclusion condition does not contain any lures, and the errors cannot be assigned to any particular value of n , so only a total error rate is reported. As these errors constituted a kind of false alarm error, henceforth we will call these errors *false alarms to non-repeated stimuli*. This will help us to distinguish them from false alarm errors committed in the cases of lures (introduced in Experiments 4–6), which will shortly be referred to as *false alarms*. False alarms to non-repeated stimuli reflected a bias toward responding in cases when the participants were uncertain whether a non-repeated stimulus had or had not been repeated. The rates of these false alarms were subtracted from hit rates in order to get an unbiased estimate of individual accuracy in responding to targets. Such a correction accounted for the fact that a participant's probability of guessing to respond when she or he was uncertain whether the current trial consisted of a repetition or not did, respectively, increase her or his hit rate in comparison to the "real" hit rate she or he would score if she or he responded only when being certain.

Before the computerized WM task, each participant was given the standard paper-and-pencil test of fluid reasoning: Raven's Progressive Matrices Advanced Version (Raven, Court, & Raven, 1983). The total number of correctly solved items (out of all 36 items) within 40 min was taken as an estimate of an individual's reasoning ability (the reasoning score).

Results

Reliabilities of the consecutive hit rates of the n -back task, estimated with Cronbach's alpha, were $\alpha_2 = .80$, $\alpha_3 = .69$, and $\alpha_4 = .70$ (the subscripts indicate respective values of n). Raven's reliability (calculated from a larger sample of 1,129 participants) was $\alpha = .87$. As this article focuses on correlational analyses, relevant means and standard deviations for consecutive values of n for this and all subsequent experiments are presented in the Appendix. The rate of false alarms to non-repeated stimuli was relatively low ($M = 0.08$, $SD = 0.05$), and for further analyses it was used in order to correct the individual hit rates. Repeated-measures analysis of variance (ANOVA) revealed the main effect of n factor, $F(2, 200) = 128.53$, $\eta^2 = .56$, $p < .001$, indicating that accuracy significantly decreased with increasing n . All three hit rates were strongly correlated ($r_s > .61$, $p_s < .001$).

Gf level correlated significantly with the 2-back ($r = .33$, $p = .001$) and 3-back ($r = .25$, $p = .013$) hit rates, while the correlation of the reasoning score with the 4-back hit rate ($r = .10$, $p = .317$) was not significant. A respective r value for the 2-back condition was significantly higher than the one for the 4-back condition, $t(99) = 2.37$, $p = .001$.

Discussion

Experiment 1 yielded two important results. First, increasing the value of n largely decreased performance, replicating the results commonly observed within the n -back paradigm. This effect suggests that 3- and 4-back trials imposed a high load on WM. Second, the significant difference in correlation values between Gf and the 2- and 4-back hit rates is consistent with the hypothesis predicting that the Gf-WM correlation is highest at small values of n . Following our assumptions from the introduction, we interpret this result as indicating that in the majority of trials most participants were generally unable to hold the 4-back targets within their foci of attention, and the more capacious attention of some of them probably did not help them to recognize these targets. On the contrary, the 2-back targets, and probably the 3-back ones, were more often kept within the focus, and the supposedly more capacious attention of more intelligent participants contributed to their better recognition of these targets in comparison to the less intelligent ones. The reliabilities of respective n -back conditions were satisfactory and—most important—comparable, so they could not account for the observed differences in correlation strengths.

However, the main problem with this first study was that the stimulus presentation rate was probably too fast for the applied compound stimuli, resulting in a relatively low hit rate in comparison to the accuracy observed in studies on the n -back task administered with longer presentation periods (e.g., Hockey & Geffen, 2004; Kane et al., 2007). Elsewhere, one of us has shown (Chuderski, Stettner, & Orzechowski, 2008) that the differences in the use of attention that appear within the Sternberg task depend on

time pressure. So, in order to reject the possibility that substantial time pressure narrowed participants' actual capacity, more time for responding was allowed in the next study. We also wanted to test if the 1-back hit rates would be related to Gf.

Experiment 2

Method

Participants. A total of 102 young people (some were not students) participated in the study (52 females; M age = 22.80 years, $SD = 3.62$; age range 18–36 years). They were recruited by newspaper ads and flyers in Krakow, Poland. The same reward applied as in Experiment 1.

Materials and procedure. Stimuli were the same as in Experiment 1, with the exception that they were presented on a light gray background. In order to make this task more similar to the standard n -back task, we no longer used squares that separated the sequences of stimuli. Instead, one stream of 80 stimuli was presented serially to participants in each session. Each item was presented for 1,000 ms and then a mask was shown for another 1,000 ms. There were 16 targets in one session, 4 per each n -back condition (i.e., 1-, 2-, 3-, 4-back), placed randomly in the stream of stimuli with regard to n . The neighboring stimulus repetitions could not overlap. All 16 targets differed from one another, and none of the remaining 64 items could match any other stimulus in the whole session. Each session started with a fixation point presented for 1,000 ms. There was one training session and eight experimental sessions, each followed by a short break. The participants were instructed to remember the four most recent numbers and to press the space bar only when they detected stimulus repetition one, two, three, or four items ago, and to withhold responses in any other case. Correct responses had to occur within 250 ms of a target appearing on the screen and within 250 ms of a respective mask disappearing from the screen. All remaining responses were treated as errors.

In Experiments 2–6, we used two tests of fluid reasoning. A novel analogy test (Orzechowski & Chuderski, 2007) included figural analogies in the form "A is to B as C is to X," where A, B, and C are types of figures; A is related to B according to two, three, four, or five latent rules (e.g., symmetry, rotation, change in size, color, thickness, number of objects); and X is an empty space. The task is to choose one figure from four that relates to Figure C as B relates to A. Before the WM task, each participant was given Raven's Advanced Progressive Matrices and the figural analogy test. The reasoning score (Gf) was calculated as the mean of z scores of the total number of correctly solved items within 60 min in Raven's test and the total number of correctly solved items within 60 min in the analogy test.

Results

Raven's test scores ranged from 2 to 36 ($M = 22.14$, $SD = 8.35$). Analogy test scores ranged from 3 to 36 ($M = 23.97$, $SD = 8.08$). The analogy test used herein was highly reliable ($\alpha = .86$; based on a sample of 1,129 participants), which was comparable to the reliability of Raven's test (see Experiment 1). Both Gf test scores correlated strongly ($r = .85$, $p < .001$).

Reliabilities of the n -back task were $\alpha_1 = .42$, $\alpha_2 = .54$, $\alpha_3 = .60$, and $\alpha_4 = .60$. These values were unsatisfactory. The hit rates were again corrected by subtracting the individual rates of false alarms to non-repeated stimuli, which this time were relatively high ($M = 0.15$, $SD = 0.15$), from each hit rate. Again, there was a significant main effect of n factor, $F(3, 303) = 125.46$, $\eta^2 = .55$, $p < .001$.

All correlations among hit rates were significant ($r_s > .50$, $p_s < .001$). The correlations between Gf and the 1-back ($r = .50$), 2-back ($r = .47$), as well as the 3-back ($r = .32$) were significant ($p_s < .001$), while the correlation between Gf and the 4-back was not ($r = .10$, $p = .318$). The correlation coefficient for the 1-back was significantly higher than the r value regarding the 4-back, $t(100) = 4.40$, $p < .001$.

Discussion

The increase in time allowed for responding resulted in higher levels of accuracy, comparable to other n -back task studies. Experiment 2 replicated and extended the results of Experiment 1. The correlation coefficients between hit rates and reasoning were the highest for the 1- and 2-back targets and then significantly decreased as the value of n increased. This result is consistent with our hypothesis suggesting that only attentional capacity, including a very limited number of items, contributed to fluid reasoning. Performance under the largest WM load, which was presumably based on retrievals from WM area outside the focus, was not related to reasoning.

It is interesting that even the 1-back hit rate correlated with reasoning as strongly as did the 2-back rate, though the 1-back condition imposed a relatively low load on WM. According to our assumptions, maintaining the 1-back items required attentional capacity of two items (i.e., the current and 1-back items), so assuming a mean capacity equal to three or four items, most participants should have been able to fulfill such a condition within their foci of attention, and it should not have yielded particularly strong correlations with reasoning. However, maybe in the n -back task, due to its dynamic nature and the compound stimuli we used, some participants were not able to fully use their capacity, and they attentionally processed a smaller number of items than they would have processed in a less dynamic task. For this reason, the 1-back condition might also appropriately differentiate between participants who did and did not use the focus of attention to maintain the 1-back targets.

An unexpected outcome of Experiment 2 was the low reliability of the hit rates. However, the rates that yielded the weakest correlations with reasoning (i.e., the 3- and 4-back), had higher reliability than more strongly correlating rates (i.e., 1- and 2-back), and the former approached the reliability level accepted in psychometric studies (both $\alpha_s = .60$). So, it is unlikely that the observed differences in the correlation strength between hit rates and reasoning were caused by differences in reliabilities of the former. On the contrary, the differences in reliability might have attenuated the observed effect, which would be even stronger if the reliabilities of all hit rates were similar.

Experiment 3

Experiment 3 aimed to extend the results regarding the 4-back condition onto the 5-back condition. Moreover, having more conditions and a larger sample, we intended to test our aforementioned

interpretations with structural equation modeling (SEM). We aimed to introduce one latent variable representing the hit rates at small n positions, assumed to fall within the focus of attention, and the other variable representing hit rates at large n positions, presumably reflecting encoding outside the focus, and to relate both these variables to the variable reflecting reasoning. We also wanted to compare such a model with a one-factor model, including one WM variable loaded by all hit rates.

Method

Participants. Participants were recruited via publicly accessible social networking websites. A total of 136 people participated (89 females; M age = 23.2 years, $SD = 4.4$; range 17–44). Each participant was paid the equivalent of about €6 in Polish zloty.

Materials and procedure. We used the same stimuli as in Experiments 1 and 2. This time a stream of 75 stimuli was presented serially to participants in each session. Each item was presented for 1,800 ms and was followed by a mask that was shown for 1,000 ms. There were 15 targets in each session, three per each n -back condition (i.e., 1-, 2-, 3-, 4-, and 5-back), placed randomly in the stream of stimuli with regard to n . There was one training session and five experimental sessions, each followed by a short break. Also, in this and subsequent experiments, a “warm-up” computerized task (dissimilar from the n -back task) was applied for several minutes, allowing the participants to get familiar with a computerized procedure. All other procedural details and the calculation of the dependent variables were identical to Experiment 2. In Experiment 3, 20 min were allowed for solving Raven’s test, and 15 min were given for the figural analogy test, in order to shorten the duration of the experiment.

For SEM computations, we used Statistica software (Version 9) with the maximum-likelihood estimation. The goodness of fit of the present and subsequent models was evaluated with three measures: chi-square value divided by the number of degrees of freedom (χ^2/df), Bentler’s comparative fit index (CFI), and the root-mean-square error of approximation (RMSEA). We adopted the following commonly postulated criteria of the acceptable fit of models: χ^2/df should not exceed 2.0, CFI should be higher than the value of .92, and RMSEA should not surpass the value of .08.

Results

Raven’s test scores ranged from 2 to 29 ($M = 18.75$, $SD = 4.85$). Analogy test scores ranged from 3 to 31 ($M = 17.63$, $SD = 5.23$). Reliabilities of the n -back task were $\alpha_1 = .52$, $\alpha_2 = .50$, $\alpha_3 = .46$, $\alpha_4 = .63$, and $\alpha_5 = .63$, and they were still unsatisfactory. The hit rates were corrected by subtracting the (relatively low) individual rates of false alarms to non-repeated stimuli ($M = 0.07$, $SD = 0.02$). Again, the main effect of n factor was found, $F(4, 540) = 221.44$, $\eta^2 = .62$, $p < .001$.

The significant (all $p_s < .02$) correlations between the reasoning score and hit rates regarded the 1-back ($r = .21$), 2-back ($r = .26$), and 3-back ($r = .28$) hit rates, while the 4-back ($r = .16$) and 5-back ($r = .13$) conditions did not yield significant correlations ($p_s > .07$). The (highest) correlation coefficient regarding the 3-back was significantly higher than the (lowest) r value of the 5-back condition, $t(134) = 1.74$, $p = .042$.

Correlations among hit rates and Gf test scores, used in SEM computations, are presented in Table 1. They indicate that the

Table 1
Correlation Matrix for Hit Rates and Gf Test Scores in
Experiment 3

Variable	1	2	3	4	5	6	7
1. 1-back hit rate	—						
2. 2-back hit rate	.40	—					
3. 3-back hit rate	.35	.51	—				
4. 4-back hit rate	.19	.39	.58	—			
5. 5-back hit rate	.15	.29	.54	.62	—		
6. Raven	.15	.24	.21	.07	.04	—	
7. Analogies	.23	.23	.29	.21	.19	.60	—

Note. $N = 136$. Significant r s are marked in bold. Gf = general fluid intelligence.

longer the distance between conditions (e.g., the 1- and 5-back vs. the 4- and 5-back), the weaker the correlations between them. Exploratory factor analysis (varimax rotation) extracted two factors, one (eigenvalue = 1.46) loading the 1-, 2-, and (moderately) 3-back hit rates, and the other (eigenvalue = 2.16) loading the 3-, 4-, and 5-back hit rates.

The above data suggest that two latent variables may underlie performance in the n -back task in Experiment 3. One variable may reflect the effectiveness of the focus of attention, and it should be loaded by the 1-, 2-, and 3-back hit rates, while the other variable may represent the accuracy of retrievals from the activated part of LTM, and it should be loaded by the 3-, 4-, and 5-back hit rates. The former variable was expected to strongly correlate with the reasoning variable, which represented variance shared by Raven's and the figural analogy tests, while the latter variable may not be a significant predictor of reasoning. The overlap regarding the 3-back hit rate reflected individual differences in attentional capacity (Cowan, 2001). We assumed that some more capacious participants would be able to fulfill 3-back targets within their foci of attention, while others would not be able to do that, and by linking the 3-back hit rate to both variables, we were able to

account for this fact. Such a model was estimated and is presented in Figure 1.

The model's fit was excellent ($N = 136$, $df = 10$, $\chi^2/df = 0.96$, CFI = 1.0, RMSEA = .00). Both latent variables moderately correlated ($r = .55$), which is understandable as all their manifest variables reflected one and the same task. The variable representing the focus of attention explained 15.2% of variance in reasoning, while the other variable did not account for any Gf variance. Elimination of the path between activated LTM and reasoning did not influence the model's fit ($\Delta df = 1$, $\Delta\chi^2 = 0.08$). An alternative model, which included one latent variable loaded by all five hit rates, was unacceptable ($N = 136$, $df = 13$, $\chi^2/df = 2.56$, CFI = .923, RMSEA = .122) and definitely worse than the two-factor model ($\Delta df = 3$, $\Delta\chi^2 = 23.68$).

Discussion

The presented data again indicated that the n -back task is a good predictor of fluid reasoning only when n does not exceed three, and these data are consistent with the interpretation that only the capacity of severely limited focal attention contributes to reasoning. This time, the 1-back hit rate correlated more weakly with Gf than did the 2- and 3-back rates (though such a difference was not significant). Prolonging the trials allows us to reject the hypothesis that assumed that participants did not hold 4- or 5-back items in their WM due to time pressure. Contrary to the studies showing an important role of LTM in fluid ability (e.g., Mogle, Lovett, Stawski, & Sliwinski, 2008; Unsworth, Spillers, & Brewer, 2010), our study did not confirm any link between the activated part of LTM (related to performance on 4- and 5-back targets) and reasoning.

Again, although the reliabilities of hit rates were relatively low, the rates with the highest relative reliability yielded the lowest correlations with reasoning. This fact, as in Experiment 2, makes unlikely any interpretation suggesting that the observed effect is related to differences in reliability. One possible cause of the low reliabilities observed in Experiments 2 and 3 could be the fact that

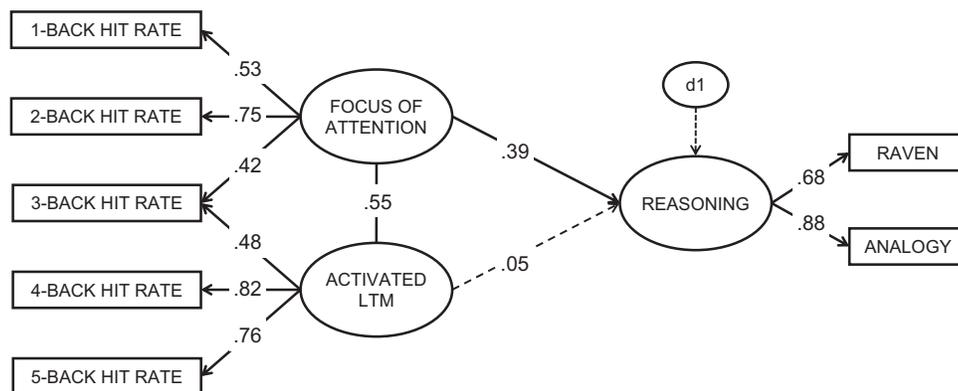


Figure 1. The structural equation model for Experiment 3, relating the focus of attention and activated long-term memory (LTM) exogenous latent variables to the fluid reasoning endogenous latent variable. Boxes represent manifest variables (see explanation in the text). Large ovals represent latent variables, while the small oval represents a disturbance term. Values between ovals and boxes represent relevant standardized factor loadings (all p s < .01). Values between ovals represent path coefficients among latent variables: The solid lines represent p s < .001, while the dashed line represents $p > .05$.

the processing in the inclusion version of n -back task is relatively weakly constrained, meaning that participants had to track more than one n -back position at once and their relevant coping strategies may have differed (in Experiment 1, due to only three n -back conditions and the fast presentation rate, this might not have been such a problem). If so, then the task's exclusion version, which requires tracking only one n -back position at a time, should give more reliable measures.

Experiment 4

In Experiment 4, we switched the procedure to the exclusion version of the n -back task. In the inclusion version, participants are required to simultaneously focus their attention on a few items. One question is whether we can also observe the effect of the n factor on the strength of WM-Gf correlation when participants' attention is focused on one prespecified value of n . Surely, items at more recent positions than the n -back have to be somehow maintained before they become n -back items. So, in the exclusion version, it can be expected that even though participants can allocate their attention to only one (n -back) item, increasing the value of n would still reduce the chance that this item would be maintained in the focus of attention. Thus, a decreasing strength of correlation between reasoning and hit rates in the function of n should be observed. In order to test that expectation we examined the 2-, 3-, and 4-back conditions separately, and we informed participants about which condition was currently being tested.

Another aim of this experiment was to examine responses to lures, which were introduced at the 1- and 5-back positions in all conditions. We asked if false alarm rates related to these lures would be significantly correlated with n -back performance in target trials and with reasoning.

It should be noted that in the exclusion version of the n -back task, the hit and false alarm rates of an individual are influenced not only by a bias related to guessing whether an item was or was not repeated (as indicated by rates of false alarms to non-repeated stimuli) but also by a bias related to whether individuals are more likely to interpret detected stimulus repetitions as targets (matching n) or lures (not matching n). However, in a particular condition of the n -back task paradigm, it is not possible to directly estimate such an individual bias (e.g., with use of the β parameter of signal detection theory) nor to derive a bias-free index of discriminability between lures and targets (e.g., d' parameter), because, by definition, targets and lures have to be placed at disjoint n positions in a stimuli sequence, and as the n effect on response accuracy is substantial, one cannot directly compare commission and omission errors at unequal n positions. For example, equal rates of false alarms for sub- n -back lures and hits for n -back targets would not indicate zero discriminability between the former and the latter, because a hypothetical hit rate at sub- n -back position would be higher than that at n -back, and thus the discriminability would in fact be positive (the reverse would be true in case of supra- n -back lures). Nonetheless, some approximate estimation of β and d' indices (as recommended by reviewers of a previous version of this article) seemed to be necessary in regard to our exclusion studies. So, regardless of the aforementioned objections, we undertook the following procedure in accounting for biases toward targets/lures.

First, in all subsequent experiments, we verified for a given n condition if a mean discriminability of lures and targets, both located at the same position, is satisfactory, because poorly discriminable lures most probably would not sufficiently engage control processes, and the corresponding false alarm rate would not be a valid indicator of executive control. So, we computed d' indices using mean hit and false alarm rates from different groups/experiments. Although the latter procedure did not allow us to compute individual values of d' , on the assumption that the n -back procedure was relatively similar among consecutive experiments and it yielded comparable hit rates for the analogous conditions (see Appendix), d' indices computed in such a way should approximate the true mean discriminability at particular n positions. We were interested in whether a mean d' level was substantially above zero, indicating that participants were really able to identify lures as lures (and then the differences in suppressing responses to these lures can be validly interpreted), or if it was close to zero, suggesting that they messed up targets and lures.

Second, we computed individual d' values for each n -back condition regarding targets, using false alarm rates from the same condition. Although these estimates deviated from the true values of discriminability for the aforementioned reasons, they more or less approximated the individual bias-free level of performance at various ns . As we were not interested in absolute values of these estimates but used them only in correlational analyses (i.e., in relation to Gf), such an approximation seemed acceptable.

Finally, using the same hit and false alarm rates as in individual d' calculation, we also approximated decisional biases in each n -back condition. We did so by dividing the mean false alarm rate observed in that condition by the mean sum of rates of false alarms and omissions (i.e., the latter equal to one minus hit rate) in that condition, no matter what lure positions were applied. We aimed to test whether the expected advantage of intelligent participants over less intelligent ones in some conditions of the exclusion n -back task would or would not emerge due to their more liberal or more conservative bias. As we were again interested only in relative differences in the values of β regarding intelligence level, such an approximation also seemed acceptable. In addition to analyses regarding the d' and β parameters, we separately analyzed data on hit and false alarm rates.

Method

Participants. A total of 135 participants, who were tested in Experiment 3, were examined with the exclusion version immediately after completing the inclusion one (one remaining male person resigned).

Materials and procedure. Stimuli were the same as in Experiment 3. A stream of 84 stimuli was presented serially in each session. Three sessions were used for each n condition (2-, 3-, and 4-back). There were 10 targets in each session. Also, three 1-back repetitions and three 5-back repetitions were included in each session, acting as lures. No other stimuli could be repeated within the same session. Participants were informed of how many items back the targets would appear in a given session, and they were instructed not to respond to the 1- and 5-back lures. Also, a proper instruction was presented at the bottom of the screen for the entire duration of a session (e.g., in the 3-back, "respond only to item repetitions separated by exactly two other items"). We expected

that rejecting lures would involve control to some extent, because in the preceding inclusion task participants had learned to respond to the 1- and 5-back targets. Half the participants fulfilled the task in the following order: the 2-, 3-, and 4-back. The other half of participants encountered the reversed order of sessions. All other procedural details and dependent variables calculations were identical to Experiment 3.

Results

The reasoning scores from Experiment 3 were adopted. Reliabilities of the hit rates equaled $\alpha_2 = .88$, $\alpha_3 = .86$, and $\alpha_4 = .84$. They were substantially higher than in the two preceding experiments and were very satisfactory. Reliabilities of the false alarm rates were $\alpha_1 = .84$ and $\alpha_5 = .65$. The hit rates were corrected by subtracting the individual rates of false alarms to non-repeated stimuli in the respective n -back conditions. False alarm rates were corrected accordingly using the individual rates of false alarms to non-repeated stimuli averaged over all conditions ($M = 0.07$, $SD = 0.02$). The main effect of the n factor on hit rate was significant, $F(4, 540) = 221.44$, $\eta^2 = .62$, $p < .001$, as was the n effect on false alarm rate regarding the 1-back versus 5-back lures, $F(4, 540) = 6.23$, $\eta^2 = .04$, $p = .014$. The latter effect indicated that the 5-back false alarm rate was higher than the 1-back rate.

In order to calculate the mean d' values for the 1- and 5-back positions, we used the 1- and 5-back hit rates from Experiment 3. The d' value for the former position equaled $d' = 2.39$, while such a value for the latter position was only $d' = 0.66$. So, the 1-back lures were easily discriminable, most probably because they constituted direct repetitions of stimuli and their inadequate position could be clearly identified, while the 5-back ones could not be discriminated so well, and most probably the more 4-back items one could recognize, the more 5-back alarms the participant had committed due to the greater number of detected 5-back repetitions.

The latter conclusion was supported by inspecting the correlations among hit/false alarm rates and scores on Gf tests, presented in Table 2. It shows that the 3- and 4-back hit rates positively correlated with the 5-back false alarm rate, while the 1-back false alarm rate correlated negatively with the 2-back hit rate. Also, all three hit rates strongly correlated, and there was also a significant correlation between the 1- and 5-back false alarm rates. Analysis of correlations between Gf and the individual d' values for 2-, 3-, and 4-back positions yielded significant r s (all $ps < .01$); however

the 2-back ($r = .44$) correlation was significantly stronger than the 3-back ($r = .22$) and the 4-back ($r = .27$), with $t(133) = 2.08$, $p = .02$, for the latter comparison. All individual d' values moderately correlated ($.34 < r_s < .44$, $ps < .001$). The analysis of β parameters for the 2-, 3-, and 4-back positions did not indicate any significant correlation with Gf (all $r_s < |.095|$, $ps > .27$).

All correlations between the reasoning score and hit rates in all n -back conditions were significant and equaled $r = .36$, $p < .001$; $r = .25$, $p = .004$; and $r = .18$, $p = .038$, in the 2-, 3, and 4-back conditions, respectively. The correlation regarding the 2-back was significantly stronger than the 4-back correlation, $t(133) = 2.23$, $p = .014$. The correlation between Gf and the 1-back false alarm rate was significant ($r = -.22$, $p = .011$), while the 5-back correlation was not ($r = -.05$, $p = .549$).

Exploratory factor analysis (varimax rotation) applied to the hit and false alarm rates extracted two factors, one (eigenvalue = 2.07) highly loaded three hit rates, while the other (eigenvalue = 1.44) highly loaded the false alarm rates. So, accordingly, we tested an SEM model in which all hit rates loaded onto one latent variable, both false alarm rates loaded onto another latent variable, and these two variables correlated and predicted the reasoning variable. However, such a model yielded an unacceptable fit ($N = 135$, $df = 11$, $\chi^2/df = 4.21$, CFI = .838, RMSEA = .159). Second, taking into account the results of that d' analysis, which showed that it was difficult for participants to distinguish the 5-back lures and 4-back items, we altered the model by allowing the 5-back false alarm rate to load onto both exogenous variables. This helped only a little, and the modified model also fit poorly ($N = 135$, $df = 10$, $\chi^2/df = 3.58$, CFI = .882, RMSEA = .140).

Finally, we tested an SEM model including three latent variables, one aimed to reflect the focus of attention, the second probably representing the activated part of LTM (i.e., analogously to the model estimated in Experiment 3), and the last variable intended to account for executive control. We allowed the 2- and 3-back hit rates to load onto the focus of attention variable, while the 3-back and 4-back hit rates and the 5-back false alarm rate loaded onto the activated part of LTM. Both false alarm rates loaded onto the executive control variable. All these latent variables were allowed to correlate, and each was linked by a directed path to the reasoning variable. The model's fit was excellent ($N = 135$, $df = 7$, $\chi^2/df = 0.58$, CFI = 1.0, RMSEA < .001). However, activated LTM again predicted a negligible amount of reasoning variance ($r = .07$, $p = .613$), and elimination of this path did not influence the fit ($\Delta df = 1$, $\Delta\chi^2 = 0.25$). The resulting model is presented in Figure 2. The one-factor model definitely fit worse than this model ($\Delta df = 3$, $\Delta\chi^2 = 32.12$). The focus of attention variable moderately correlated with the activated LTM ($r = .38$, $p < .001$) and executive control ($r = .32$, $p = .016$) variables. The path between the focus of attention and reasoning variables was highly significant ($r = .40$, $p = .001$), while the path between control and reasoning was only marginally significant ($r = .22$, $p = .087$).

Discussion

Experiment 4 yielded three new results. First, also in the exclusion condition, the correlation between reasoning and hit rates weakened with increasing n , which again favors an interpretation

Table 2
Correlation Matrix for Hit/False Alarm Rates and Gf Test Scores in Experiment 4

Variable	1	2	3	4	5	6	7
1. 2-back hits	—						
2. 3-back hits	.50	—					
3. 4-back hits	.33	.56	—				
4. 1-back false alarms	-.24	-.08	.02	—			
5. 5-back false alarms	.08	.21	.48	.31	—		
6. Raven	.30	.23	.17	-.13	-.01	—	
7. Analogies	.35	.21	.15	-.25	-.09	.60	—

Note. $N = 135$. Significant r s are marked in bold. Gf = general fluid intelligence.

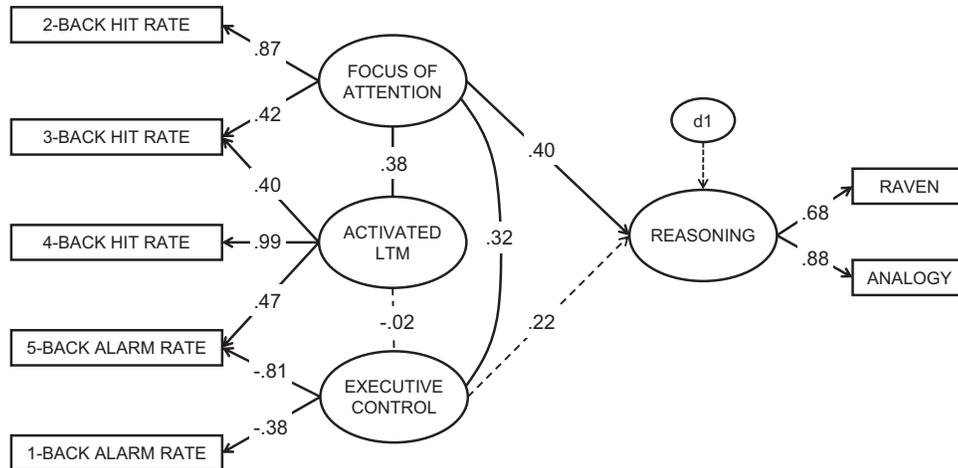


Figure 2. The structural equation model for Experiment 4, relating the focus of attention and executive control exogenous latent variables to the fluid reasoning endogenous latent variable. Boxes represent manifest variables (see explanation in the text). Large ovals represent latent variables, while the small oval represents a disturbance term. Values between ovals and boxes represent relevant standardized factor loadings (all $ps < .01$). Values between ovals represent path coefficients among latent variables: The solid lines represent $ps < .001$, while the dashed lines represent $ps > .05$. LTM = long-term memory.

that reasoning ability is related to the limited capacity of the focus of attention.

Second, the d' parameter analysis indicated that the 5-back lures were difficult to identify as lures and they were often treated as targets. As such, they did not constitute effective distractors and the lack of significant correlation between their rejection rate and reasoning cannot be meaningfully interpreted.

Finally, the 1-back lures, contrary to the 5-back ones, were easily identifiable and seemed to work as effective distractors. Most probably, as direct repetitions, their exact n -back position could be easily assessed. This view is supported by the fact that in Experiments 2 and 3 the 1-back items were almost perfectly detected ($M = 0.91$ on average). However, even though it seems that participants should have been able to correctly recognize almost all 1-back lures as lures, they committed a surprisingly high rate of false alarms to these lures ($M = 0.18$) and varied substantially in how they rejected them ($SD = 0.20$; range 0.0–1.0). Such a result suggests that identification of an item as a 1-back lure was not always sufficient to reject it. The role of executive control might be in helping to suppress reactions to lures, which otherwise would be made regardless of a participant's knowledge that they are indeed lures. The involvement of executive control would be analogous to the case of the Stroop task, when, regardless of participants' knowledge that the task consists of naming colors, the control has to prevent them from reading words. How well participants rejected such lures did correlate with reasoning, which suggests that such a kind of control may be another contributor to fluid intelligence. The SEM analysis suggested that in contributing to reasoning the variable reflecting control is independent from the variance reflecting capacity. However, the former factor accounted for three times less Gf variance as the latter.

Experiment 5

In this experiment, we applied the between-subjects design with regard to the n factor (i.e., the 2-back vs. 3-back groups). We used

$n - 1$ lures (i.e., 1-back and 2-back, respectively) and $n + 1$ lures (i.e., 3-back and 4-back, respectively), in order to test if the 2-, 3-, and 4-back lures could also yield significant correlations between false alarms and reasoning as did the 1-back lures, or if they could not yield it, similar to 5-back lures. Finally, we used figural stimuli instead of numerical ones. The former are more difficult to maintain and should substantially reduce chunking and rehearsal, thus amplifying correlations between the n -back task performance and reasoning.

The experimental procedure was further extended by introducing a new task modeled on Luck and Vogel's (1997) two-array comparison task, which is considered (Cowan et al., 2006; Rouder, Morey, Morey, & Cowan, 2011) to allow for reliable estimation of individual focus of attention capacity. The original task required memorizing an array of items. Then, after a retention interval, the array was repeated, but there was 50% chance that one of the items would change. The task was to indicate whether the item had changed or not. In Cowan et al.'s (2006) version of this task, one of the items in the second array was marked by a cue indicating that, if any of the items had changed, it was the marked one.

The formula, which estimates the sheer capacity of the focus of attention as the proportion of hits (H , correct responses for arrays with one item changed) and the proportion of false alarms (FA , incorrect responses for unchanged arrays), in the cued version of a two-array task, was proposed by Cowan (2001). Accordingly, the capacity of the focus is estimated to be k items (out of N items of a memory load), on the assumption that a participant produces a correct hit or avoids a false alarm only if a cued item is transferred to his or her focus (with the k/N chance). If a non-transferred item is cued, then a participant guesses the answer. Consequently, the following formula evaluates the capacity of the focus of attention: $k = N \times (H - FA)$.

With the use of the two-array comparison task and the estimation of k , we aimed to test our assumption that accuracy in low n conditions of the n -back task is related to individual differences in

capacity of the focus of attention. We assumed that the 2-back condition involves maintenance of a minimum of three items in the focus: a currently presented stimulus, a 1-back one (which is needed for a next trial), and a 2-back one (which has to be compared with a current stimulus). Similarly, in the 3-back condition a minimum of four items have to be maintained in the focus in order to provide high accuracy. Of course, participants with insufficient capacity may still respond correctly on some occasions using their activated LTM. We hypothesized that in the 2-back condition participants showing k values of three and more would display higher accuracy than ones presenting k s below three, while in the 3-back condition only participants possessing k values of four or more would surpass less capacious ones.

Method

Participants. A total of 275 people, who were recruited via publicly accessible social networking websites, participated (166 females; M age = 22.5 years, SD = 3.6; range 17–40). Each participant was paid the equivalent of €7 in Polish zloty. The participants were assigned to two groups (the 2-back group had 135 people and the 3-back group had 138 people) according to the order in which they entered the laboratory. Data from two participants (one in each group) were excluded because of enormously high error rates (M = 0.70 and M = 0.49), greatly exceeding the error rates of all other participants, which suggested that these two participants did not follow the task instructions.

Materials and procedure. In the n -back task, stimuli were 16 simple black figures (e.g., a square, a circle, a rhombus, an arrow, a cross), each approximately 2.5×2.5 cm in size, presented for 1,500 ms plus a 300-ms mask. A total of 88 stimuli were presented serially to participants in each session. Four sessions were used in each group, preceded by some training. Each session that was completed by the 2-back group included eight 2-back targets, four 1-back lures, and four 3-back lures. In the 3-back group, eight 3-back targets, four 2-back lures, and four 4-back lures were included in each session. No other stimuli could be repeated within the few most recent items. Participants were instructed to respond to either 2-back or 3-back repetitions (depending on the group) and to suppress responses to all other repetitions. As only one n condition was applied to each participant, there were no hints presented during trials. Calculation of the dependent variables was identical to that in previous experiments.

Each of the 90 trials of the two-array comparison task consisted of a virtual, four by four array filled with a few stimuli (i.e., only some cells in the array were filled). The stimuli were 10 Greek symbols (e.g., α , β , χ), each 2×2 cm in size. The number of stimuli within the array varied from five to seven items. The array was presented for the time equal to the number of its items multiplied by 400 ms and then followed by a black square mask of the same size as the array, presented for 1,200 ms. In a random 50% of trials, the second array was identical to the first one, while in the remaining trials both differed by exactly one item at one location. If they differed, the new item was highlighted by a square red border. If they were identical, a random item was highlighted. The task was to press one of two response keys depending on whether the highlighted item differed or not in the two arrays. The second array was shown until a response was given or 4 s elapsed.

The trials were self-paced. The dependent variable was a mean of k values calculated for each set size.

In Experiment 5, we allowed 40 min for Raven's Progressive Matrices, and 30 min were given for the figural analogy test.

Results

The Raven test scores ranged from 1 to 36 (M = 22.77, SD = 6.40). Analogy test scores ranged from 10 to 36 (M = 25.32, SD = 5.45). Reliabilities of the hit rates were α_2 = .85 and α_3 = .78. Reliabilities of the false alarm rates were α_1 = .92, α_2 = .68, α_3 = .76, and α_4 = .64. All these values ranged from good to acceptable. The hit and false alarm rates were corrected by subtracting the individual rates of false alarms to non-repeated stimuli, which were low (M = 0.06, SD = 0.06).

There was the main effect of the n factor in both targets and lures, $F(1, 271) = 53.80$, $\eta^2 = .17$, $p < .001$, and $F(1, 271) = 35.32$, $\eta^2 = .12$, $p < .001$, respectively. The former effect reflected a higher hit rate in the 2-back than in the 3-back condition, while the latter effect indicated that the 2- and 4-back false alarm rates observed in the 3-back group were significantly higher than the 1- and 3-back false alarm rates seen in the 2-back group (M = 0.41 vs. M = 0.20, respectively). We also calculated the d' indices using the 2- and 3-back hit rates from the present experiment and the 1- and 4-back hit rates from Experiment 3. The resulting values were $d' = 2.39$, $d' = 0.78$, $d' = 0.87$, and $d' = 0.51$, for the 1-, 2-, 3-, and 4-back conditions, respectively.

The correlational matrix for lure and hit rates is presented in Table 3. Matching the analysis of d' values, it shows significant positive correlations between the hit rates and the 2-, 3-, and 4-back false alarm rates but not between the former and the 1-back false alarm rate. Both d' values and those correlations indicate that participants imperfectly distinguished 2-, 3-, and 4-back lures from targets. The individual d' value significantly correlated with reasoning both in the 2-back group ($r = .51$, $p < .001$) and in the 3-back group ($r = .26$, $p = .002$), which differed significantly, $t(271) = 3.20$, $p < .001$. The correlations between the β parameter and reasoning were significant neither in the 2-back group ($r = -.04$, $p = .652$) nor in the 3-back group ($r = .03$, $p = .705$).

The 2-back hit rate correlated with reasoning score significantly more strongly ($r = .50$, $p < .001$) than did the 3-back hit rate ($r = .25$, $p = .004$), $t(271) = 3.16$, $p < .001$. In the 2-back group, the 1- and 3-back false alarm rates significantly correlated with reasoning ($r = -.22$ and $r = -.20$, respectively, both $ps < .03$). On the contrary, in the 3-back group, correlations of the 2- and 4-back false alarms with Gf were substantially weaker and not significant ($r = -.05$ and $r = .04$, respectively, both $ps > .5$).

The mean k value equaled $M = 3.25$ ($SD = 1.36$, range 0.0–5.78). In order to present the differences in n -back performance related to the k estimate, we divided participants into five groups. The $k = 1$ group ($n = 38$) included participants with k values lower than 1.5, who were assumed to be able to track on average only the current item. The $k = 2$ group ($n = 31$) included participants in the range 1.5–2.5, who were assumed to be able to hold on average two items in their foci of attention. The $k = 3$ ($n = 68$) and $k = 4$ ($n = 88$) groups were formed accordingly. The $k = 5$ ($n = 48$) group included participants possessing k values of 4.5 or higher. The means for all k groups and both n -back conditions are presented in Figure 3. Subsequently, hit rates in the n -back task

Table 3
Correlation Matrix for Hit/False Alarm Rates and Gf Test Scores in Experiment 5

Variable	1	2	3	4	5	6	7	8	9
1. 2-back hits	—								
2. 1-back false alarms	-.19	—							
3. 3-back false alarms	-.15	.36	—						
4. 3-back hits				—					
5. 2-back false alarms				.40	—				
6. 4-back false alarms				.53	.49	—			
7. <i>k</i> value	.48	-.12	.03	.29	.03	.14	—		
8. Raven	.49	-.22	-.19	.24	-.01	.06	.30	—	
9. Analogies	.40	-.17	-.17	.19	-.08	.02	.16	.60	—

Note. $N = 135$ and $N = 138$ for the 2- and 3-back groups, respectively. Significant *r*s are marked in bold. Gf = general fluid intelligence.

were submitted to ANOVA, with *k* (1–5) and *n*-back (2 and 3) groups as factors. The effect of the *k* group was highly significant, $F(4, 263) = 13.29$, $\eta^2 = .17$, $p < .001$, indicating that accuracy in both tasks were strongly related. Most important, the interaction between *k* and *n* group factors was also significant, $F(4, 263) = 4.43$, $\eta^2 = .04$, $p = .048$. It showed that in the 2-back group, there was no contrast between the $k = 1$ and $k = 2$ groups ($F = 0.05$), while there was a highly significant contrast between the $k = 2$ and $k = 3$ groups, $F(1, 263) = 7.06$, $p = .008$ (further contrast between the $k = 3$ and $k = 4$ groups was significant, $F[1, 263] = 4.32$, $p = .038$, and the one between the $k = 4$ and $k = 5$ groups was close to significance, $F[1, 263] = 3.48$, $p = .063$). On the contrary, in the 3-back group, the $k = 2$ and $k = 3$ groups did not differ significantly ($F = 0.03$), while the $k = 3$ and $k = 4$ groups did, $F(1, 263) = 6.63$, $p = .011$ (the $k = 4$ and $k = 5$ groups contrast

was not significant, $F = 0.25$). Contrary to the hit rates, no effect of the *k* group on the false alarm rates has been found ($F = 0.83$).

Data regarding the 2-back group, indicating good predictability of reasoning by hit rates as well as sufficient identifiability of lures, were used in order to calculate another SEM model, relating the attentional capacity and executive control variables to the reasoning variable. We aimed to replicate the results of SEM analysis done in Experiment 4. In the present model, the capacity variable was loaded by the 2-back hit rates and also by another measure of capacity, namely, the value of *k*. The control variable was loaded by the 1- and 3-back lures. These assumptions were supported by exploratory factor analysis (varimax rotation), which showed one factor (eigenvalue = 1.57) yielding high loadings ($>.82$) on the 2-back hit rates and *k* values but negligible loadings ($<.125$) on both false alarm rates, and another factor (eigen-

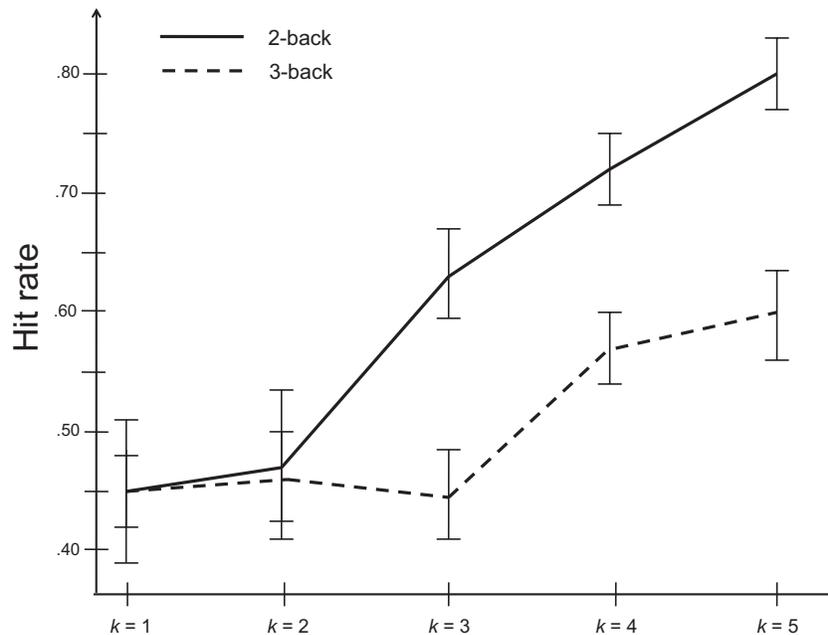


Figure 3. The error-corrected hit rates in five groups differing in the estimated capacity of the focus of attention (*k* values), in the 2-back (solid line) and the 3-back (dashed line) conditions, in Experiment 5. Error bars represent 1 *SE* of the mean.

value = 1.25) yielding high loadings ($>.80$) on false alarms but negligible loadings ($<.05$) on the hit rate and the value of k . According to the good-fitting model ($N = 135$, $df = 7$, $\chi^2/df = 1.76$; CFI = .969, RMSEA = .071), presented in Figure 4, attentional capacity and executive control did not significantly correlate ($r = .13$, $p = .291$), and the former variable explained 62.4% of variance in reasoning ($r = .79$, $p < .001$), while the latter accounted for 13.7% of that variance ($r = .37$, $p < .001$). The correlation between capacity and control could be eliminated without significant loss of fit ($\Delta df = 1$, $\Delta\chi^2 = 1.07$). We compared the above model with the one-factor model, which included a latent variable on which all four manifest variables loaded. Such a model appeared to be completely unacceptable (CFI = .452, RMSEA = .296).

Discussion

In the figural version of the n -back task, the correlation between accuracy in the task and the reasoning score also appeared to be significantly stronger for a lower value of n than for its higher value. This result clearly replicated the data regarding the numerical version of the task, observed in Experiments 1–4. Additionally, the analysis of individual k values, estimated with the two-array comparison task, presumably reflecting how many items could be held within the focus of attention, indicates that such a capacity substantially varied in our sample, from one (if we assume that several zero values of k mostly reflected little involvement in fulfilling the task) up to almost six items. The analysis of relation between performance on both WM tasks provided a high level of correspondence between the assumed capacity requirements of the 2- versus 3-back conditions and the special advantages of the $k = 3$ and $k = 4$ groups, respectively. This result supports such an interpretation of processing requirements of the n -back task, suggesting that the capacity of three items is needed

for maintaining the 2-back targets, while the capacity of four items is needed to attentionally process the 3-back targets.

False alarms in the 1- and 3-back lure conditions significantly contributed to reasoning, while 2- and 4-back ones did not. Together with results regarding the 5-back lures in Experiment 4, these data suggest that in order to effectively reject lures, participants have first to correctly identify them as lures and not as targets. In the 2-back group, as indicated by much lower false alarm rates in comparison to the 3-back group, differentiating lures from targets was relatively effective. As noticed when discussing Experiment 4, the 1-back lures were probably easily identified because they were direct repetitions of stimuli. The 3-back lures could have been relatively well identified because their position directly followed a position being tested for targets. In such trials, after a participant decided with high accuracy that a 2-back item was not a target, she or he encountered as the next stimulus a repetition of the target item, which as such could not be a 2-back one. Because of the good distinctiveness of lures in the 2-back group, false alarm rates in this group most probably properly reflected the efficiency of executive control and thus correlated significantly with reasoning. The contribution of executive control to Gf was independent from the contribution of capacity. The former was much smaller than the latter, but it was still significant.

On the contrary, in the 3-back group, lures seemed to be often indistinguishable from targets, probably due to inferior access to the more remote n positions. Data regarding lures most probably confounded executive control (when lures were correctly identified as lures) with storage capacity (when they were taken as targets). As lures and targets yielded opposite correlations with reasoning (i.e., negative vs. positive ones), it resulted in close-to-zero correlations of respective false alarm rates with Gf. As was previously noticed, such data cannot be meaningfully interpreted.

Of course, all the above regularities were probabilistic, so sometimes even $k = 1$ or $k = 2$ participants could have been able to

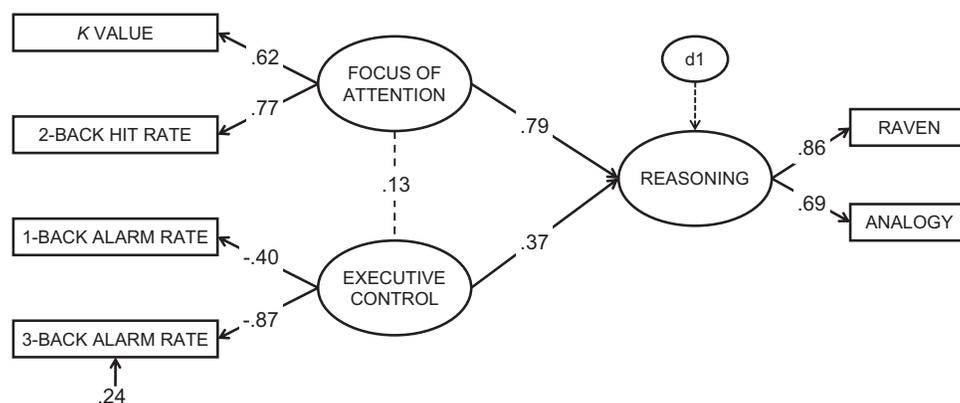


Figure 4. The structural equation model for the 2-back group of Experiment 5, relating the focus of attention and executive control exogenous latent variables to the fluid reasoning endogenous latent variable. Boxes represent manifest variables (see explanation in the text). Large ovals represent latent variables, while the small oval represents a disturbance term. Values between ovals and boxes represent relevant standardized factor loadings (all $ps < .01$). Values between ovals represent path coefficients among latent variables: The solid lines represent $ps < .001$, while the dashed line represents $p > .05$. Note that because of the negative variance found in the initial model, the error parameter for the 3-back false alarm rate was fixed at a value of 1 minus its reliability.

maintain the 2-back or 3-back items in the focus (e.g., because they did not update the 1-back items, their attention was temporarily spanned beyond the mean capacity, or some items were chunked). A methodological conclusion follows that the nature of contribution to reasoning of any assumed measure of executive control derived from a WM task depends on the load on WM that is implicated in the processing of corresponding items. Only a rejection rate of easily identifiable and thus prepotent distractors directly reflects the effectiveness of executive control. Maybe, in the context of the n -back task, including larger distances between targets and lures will make the former and the latter more distinguishable. Only then will false alarms for lures at larger n -back positions than the 1-back strongly correlate with reasoning. This prediction is the subject of the next study.

Experiment 6

This small-sample follow-up study had one aim: to test if 4-back lures, when placed at a larger distance from targets (i.e., the 2-back ones), would be more correctly rejected than the 4-back lures in Experiment 5, and, if so, whether their rejection rate would be a significant predictor of reasoning.

Method

Participants. A total of 75 people, who were recruited via publicly accessible social networking websites, participated (47 females; M age = 23.5 years, SD = 4.5; range 18–46). Each participant was paid the equivalent of €6 in Polish zloty.

Materials and procedure. Stimuli were identical to those in Experiment 5. A total of 80 stimuli were presented serially to participants in each of two sessions (preceded by training). Each session included four 2-back targets and twelve 4-back lures. The participants were warned that lures are frequent and that they should be carefully rejected, while 2-back items should be responded to. All other procedural details, including Gf test administration time and the calculation of dependent variables, were identical to those of Experiment 5.

Results

Raven's test scores ranged from 6 to 35 (M = 21.73, SD = 6.02). Analogy test scores ranged from 5 to 36 (M = 23.83, SD = 6.30). Both Gf test scores correlated significantly (r = .74, p < .001).

Cronbach's alpha equaled α_2 = .64 in hit rates (note that only eight targets were used) and α_4 = .90 in false alarms. Participants committed a false alarm rate almost two times lower than the analogous 4-back false alarm rate in Experiment 5 (see the Appendix). The hit and false alarm rates were corrected by subtracting the individual rates of false alarms to non-repeated stimuli, which in any case were very low (M = 0.04, SD = 0.02). The mean d' value for the 4-back position, which was calculated with use of the 4-back hit rate in Experiment 3, was d' = 0.99, which suggests acceptable discriminability. The correlation between reasoning and individual d' equaled r = .47, p < .001, while reasoning and β did not correlate significantly (r = -.04). The hit and false alarm rates did not correlate significantly (r = -.09, p = .402) either, and, most important, each rate significantly and

substantially correlated with the reasoning score (r = .36, p = .002, r = -.37, p = .001, respectively).

Discussion

The results of Experiment 6 suggest that false alarm rates regarding some lures placed at positions more than 1-back, as observed in Experiments 4 and 5, were not proper measures of executive control, because these lures were not effective enough as distractors and consequently they did not yield significant correlations with reasoning. When lures and targets were placed on distinctive n positions, namely, ones differing by n = 2, the former predicted reasoning as well as did the latter. The present experiment seems to support our previous conclusions, with regard to the fact that some form of executive control, operationalized as the ability to reject distinctive distractors in WM, may be a genuine explanatory factor of fluid reasoning, and it contributes to intelligence independently from WM capacity, expressed as the number of items held in the focus of attention.

General Discussion

Summary of Results

The presented research tested the theories that integrate capacity and control approaches in order to explain the cognitive basis of fluid reasoning. Summing up the results, in Experiments 1–6, the indices of performance, which were assumed to reflect storage capacity within WM, explained a substantial part of variance in fluid reasoning. In Experiments 4–6, some of the indices of performance, which were intended to measure the efficiency of control over distraction within WM, also predicted a significant but relatively smaller amount of variance in reasoning ability. As much as three quarters of variance in reasoning could be explained by these two factors (i.e., 76.1% in Experiment 5). The modified n -back task has been confirmed as a valid measure of WMC.

Storage Capacity and Fluid Reasoning

Regarding the first question, namely, whether either the total capacity of WM or the capacity of a particular substructure of WM is related to fluid reasoning, the results consistently suggest the latter possibility. In all six experiments, hit rates in our WM task substantially correlated with reasoning when the load imposed on WM was relatively low (i.e., n did not exceed three), while the respective correlation was substantially decreased or even eliminated in cases of larger values of n (Experiments 1–5). Such a pattern of results was also supported by two SEM models (Experiments 3 and 4), which included a null path between reasoning and the latent variable reflecting response accuracy for items at high n -back positions, while the variable related to items at low n -back positions predicted a substantial amount of variance in reasoning.

Why is accuracy at small n values strongly correlated with reasoning, while at large ones it is not? In our view, the 2-back condition is the most sensitive to the substantial individual differences in attentional capacity that we observed in the results from the two-array comparison task. Comparison of 2-back performance among participants with different k values indicates that half of the sample, namely, people with capacity lower than the mean k

equating around three items, were most probably not able to span their attention onto 2-back targets on most occasions, as such spanning required at least the capacity of three items (i.e., holding a current stimulus and two preceding items). So, in coping with the 2-back repetitions, those people possibly had to rely on the activated part of LTM, and due to that fact their n -back accuracy was decreased in comparison to the other half of the sample, which had a capacity surpassing the mean. Moreover, if we assume that in our n -back task participants did not always use their maximum capacity (i.e., the one estimated by their value of k), but due to the dynamic nature of this task and its compound stimuli were often able to hold in the focus of attention a lower number of stimuli, then a strong correlation between reasoning and the 1-back performance (as observed in Experiment 2) is also coherent with the above explanation.

On the contrary, the hit rates regarding large n values seem to poorly capture the individual differences in attentional capacity. In the 3-back condition, only the capacity of at least four items allowed the 3-back targets to be kept within attention (as indicated by increased accuracy in the $k = 4$ and $k = 5$ groups), so the requirements of that condition seem to exceed the capacity of most of the sample, and the 3-back targets were mostly retrieved from activated LTM. As only a minority of participants contributed to the observed individual differences in the 3-back hit rate, its correlation with Gf was usually much weaker than the analogous correlation of the 2-back (see Experiments 1, 2, 4, and 5). Extrapolating that line of reasoning, in the 4-back and, especially, 5-back conditions, hardly anyone could maintain n -back items in the focus, so the respective hit rates did not reflect individual differences in the capacity of the focus of attention, and it should be no surprise that there were no significant correlations regarding these conditions and Gf (except for a weak correlation regarding the 4-back in Experiment 4). This occurred even though reliabilities regarding the 4- and 5-back targets were comparable to those at lower n positions.

If the above interpretations are valid, then the observation of relatively strong correlations of the 2-back hit rates with reasoning, and the lack of significant correlations in most cases of the 4- and 5-back hit rates, leads to the aforementioned explanation that differences in reasoning scores depend on differences in the very limited capacity of the focus of attention, while they are not related to differences in the accuracy of retrievals from outside the focus. Hence, the presented data seem to speak for a much more important contribution to fluid processing of information by the focus of attention (primary memory) than by the activated part of LTM (secondary memory).

This thesis contrasts with interpretations by Mogle et al. (2008), Unsworth and Engle (2006), and Unsworth et al. (2010), who concluded that fluid intelligence is also related to the efficacy of search within secondary memory (e.g., to the narrowing of a search set). However, the paradigms used in our study (the n -back and two-array comparison tasks) and the ones used in the cited studies (the complex span and free recall tasks) differed significantly, for example by using the procedure of either recognition or recall, respectively, or by allowing relatively short (the n -back) or long (the free recall task) duration of time for encoding stimuli. In fact, in complex and long-lasting recall tasks, the use of STM and LTM may be substantially interleaved, so an issue of what the complex

span and free recall tasks really measure and what is a relation of LTM to reasoning surely needs further research.

Why is the focus of attention so important for fluid cognition, while WM outside the focus is not? Although the present study cannot give an answer to that question, it would be interesting to consider a few options. The first one concerns the possibility that access to information in the focus may be much better than access to activated LTM. The former may yield rich representations, for example, a complete set of bound features of an item (e.g., an item's identity and its n -back position in the n -back task), while the latter may only indicate the global familiarity of an item (e.g., showing if it was/was not encoded, but not at what n -back position). Such an explanation would be coherent with our data showing that more recent targets yielded higher accuracy than did less recent ones, and also with data on the low discriminability of the 4- and 5-back lures from comparable targets in Experiments 5 and 4, respectively. However, in Experiment 5, the 2-back lures were also poorly discriminated, though they should have been processed in the focus of attention. Moreover, the 4-back lures of Experiment 6, presumably processed out of the focus, were rejected at a high rate. It seems that information on a serial position associated with a particular item can also be retrieved from activated LTM, at least in some cases. So, these results suggest that there seems to be more to the story concerning the relation between reasoning and the subparts of WM than just the ease of access to these subparts.

Another possibility concerns the fact that some inherent characteristic of the focus of attention may make accessing the focus only relatively more efficient than accessing the activated LTM, but it may be crucial for reasoning. A good candidate for such a characteristic, as described in the introduction, would be the ability of the focus of attention to represent arbitrary, relational structures (Halford et al., 1998, 2007), for example, due to constructing, maintaining, and transforming the flexible, temporary bindings among elements of such structures (Hummel & Holyoak, 2003; Oberauer et al., 2007). Operating easily on such structures seems to be crucial for both deductive reasoning (e.g., integration of the premises into mental models; Johnson-Laird, 1999) and inductive thinking (e.g., mapping elements of source and target during analogy making; Hummel & Holyoak, 2003; Waltz et al., 1999).

Executive Control and Fluid Reasoning

As to the second question posed in the introduction, our results suggest that some form of control over distraction seems to be an important prerequisite of effective reasoning, as false alarm rates significantly correlated with scores on intelligence tests. However, this observation pertains only to lures that were sufficiently distinctive from targets. On the contrary, results regarding indiscriminable lures are not meaningful, as they most probably reflected mixed cases when participants identified lures as lures but failed to reject them and cases when participants did not identify them as lures and responded to them in the belief that they were proper targets. Due to the latter fact, weak positive correlations between respective false alarm and hit rates were observed in cases of lures yielding low values of d' .

When the analysis of false alarm data was narrowed only to properly discriminated lures, there was barely any relation between control and capacity, as those false alarm rates did not significantly correlate with adequate hit rates (if any correlations were signifi-

cant, they were weak and negative). Also, the strength of Gf correlations regarding those properly identified lures hardly depended on the value of n : Correlation for the 1-back lures in Experiments 4 and 5 equaled $r = -.22$, for the 3-back (Experiment 5) it equaled $r = -.20$, and for the 4-back (Experiment 6) it was $r = -.37$. In light of these results, the contribution of control to reasoning seems to be relatively independent from the contribution of attentional capacity. This conclusion is further supported by two SEM models (Experiments 4 and 5), which both estimated relatively weak paths ($r = .36$ and $r = .13$, respectively) between control and capacity. Although these results cannot exclude some overlap between capacity and control aspects of WM in predicting fluid reasoning, the amount of this overlap seems to be substantially smaller than the previous research suggested (e.g., Cowan et al., 2006).

What is the role of executive control in coping with lures, if any? One possibility (suggested by a reviewer) could be that rejecting lures depends solely on retrieving the proper n -back position of a repeated stimulus: If the retrieved serial position equals n indicated in the task's instruction, a repetition is accepted, while if they mismatch, a (lure) repetition is rejected. Although hit and false alarm rates would rely on the same process (i.e., the better one retrieves an n -back position, the better one both accepts targets and rejects lures), because of individual differences in bias for guessing targets/lures in cases when the serial position cannot be retrieved, the correlation between both these measures could be artifactually decreased (as noted by another reviewer), leading to a (false) identification of two distinct factors (i.e., capacity and control) in SEM models.

However, two results suggest that false alarms produced in response to highly available lures were caused not only by improper assessment of their positions. First, as noted previously, the average rate of false alarms to the 1-back lures (Experiments 4 and 5) was surprisingly high ($M = 0.18$), given that the 1-back repetitions were highly available to participants (e.g., a mean hit rate in Experiments 2 and 3 was $M = 0.91$) as well as the fact that they were also highly discriminable from targets (in Experiments 4 and 5, $d' = 2.39$). Second, if the high rate of false alarms to those highly discriminable lures was caused by imperfect assessment of serial positions of repetitions, then such an invalid process should also influence the 2-back hit rates (i.e., participants would omit some targets due to retrieving their wrong positions). On the contrary, in our data, the 2-back hit rates in the exclusion version, which were highly dependent on proper assessment of serial positions, were comparable to the 2-back hit rates in the inclusion version, in which serial positions needn't be assessed at all ($M = 0.75$ vs. $M = 0.76$, in Experiments 1–3 and Experiments 4–6, respectively). In other words, introducing the 1-back lures did not decrease the rate of acceptance of the 2-back targets. If participants knew that the 1-back repetitions were lures and the 2-back repetitions were targets (as suggested by these two results), why did they produce so many false alarms to the 1-back lures?

In our interpretation, within the go/no-go method that was applied, participants displayed a reflexive tendency to respond to whichever repetition they detected, and on some occasions this tendency could not be overridden even if they assessed that it was related to an improper n -back position. Suppressing such a tendency would be a kind of executive control process, analogous (though most probably not identical) to processes commonly be-

lieved to resolve conflicts in other interference tasks. Such an interpretation seems to be supported by neuroimaging studies showing that human brains react differently to targets than to lures, with the latter activating regions implicated in resolution of conflicts (e.g., Burgess et al., 2011; Chatham et al., 2011; Gray et al., 2003). In our view, the individual efficiency of some kind of executive control did influence the individual false alarm rates to highly distinguishable lures (while there was a lesser involvement of control in the rejection of poorly distinguished lures, because, as they were often taken as targets, they did not cause a high level of conflict). However, the presented experiments do not allow for identification of what the control process responsible for coping with conflicts related to lures is exactly, and examination of the characteristics of such a process surely needs a dedicated study (see Chatham et al., 2011; Ecker, Lewandowsky, Oberauer, & Chee, 2010; Szmalec, Verbruggen, Vandierendonck, & Kemps, 2011).

Here, only speculation on the role of control in reasoning can be drawn. If the attentional capacity allowing for information processing during reasoning is so limited, then, while dealing with difficult problems, people surely have to break them down into several smaller sub-problems and then build the final solution from the respective partial solutions (this is exactly how the influential LISA model of reasoning works; see Hummel & Holyoak, 2003). So, efficient control may be needed for the composition of final solutions. For example, while choosing appropriate sub-products (e.g., hypotheses, relations) one usually needs to avoid superficially matching mental representations that are dominant but not relevant to the problem (e.g., ones consistent with common sense but not following the premises). Instead, a problem solver may have to retrieve representations from memory that violate some well-learned schemata and semantic relations while still being goal relevant. The process of rejecting lures during performance on the n -back task could be somehow related to the more general control processes required for fluid reasoning, which allow for goal-directed selection of information. Due to better control processes of this kind, some people may be both more efficient problem solvers in the fluid reasoning tests and more accurate participants in WM tasks requiring the rejection of distractors. Such speculation seems to be supported by recent data indicating that proper selection and rejection of features of structures processed during reasoning is crucial for finding correct solutions (e.g., Cho et al., 2007; Chuderska, 2010).

Some Alternative Explanations of the Effect of n on the Correlation of Hit Rate and Reasoning

At least three alternative explanations of the data indicating the difference in correlation between Gf and hit rates on smaller versus larger n positions are possible. One potentially plausible explanation, already noted, refers to the more frequent use of the phonological loop by intelligent people, as the phonological loop is an important mechanism in some WM models (e.g., Baddeley & Hitch, 1974). However, in light of knowledge on the effects of rehearsal on WM performance, this hypothesis seems less likely than the one relating to the limits of the focus of attention and the effectiveness of control. Rehearsal, as a relatively passive mnemonic mechanism, blurs rather than increases the differences in working memory performance. As far as we know, no published

results have shown that intelligent people use the phonological loop more often or more efficiently. On the contrary, it is often suggested (e.g., Cowan et al., 2007) that rehearsal in WM tasks decreases the strength of relation between WMC and intelligence. Moreover, in Experiments 3 and 4, sufficient time (2.8 s) for rehearsing the two recent stimuli was provided, but it did not influence the correlation between 2-back accuracy and reasoning compared to the remaining experiments.

Another explanation concerns the more efficient recollection processes of intelligent people.² Recollection is believed to be a relatively slow mode of access to memory, which consists of the retrieval of all the associated information about an object (e.g., its specific features like color or size) as well as the contextual data associated with it (e.g., its position in the memory set). Recollection contrasts with the relatively fast, familiarity-based process of recognition, which yields only a general, unitary value indicating the confidence that the given object has or has not been coded in memory (Yonelinas, 2002). One possible explanation of the presented data states that due to the more efficient retrieval of contextual information about the precise *n*-back position of a stimulus (i.e., more efficient recollection), intelligent people accepted more targets and rejected more distractors than did less intelligent ones. No such difference would pertain to a familiarity-based process. However, this hypothesis does not need to contradict the hypothesis of limited attentional capacity, because recollection is often believed to only make use of the focus of attention, while familiarity processes probably operate on the whole WM, including its activated long-term components outside the focus (Oberauer, 2005). Assuming that reasoning is related to recollection, but recollection operates only on the focus of attention, which varies in capacity among individuals, would yield predictions analogous to those following the assumption of the limited and varied capacity of the focus underlying *Gf*. As our work was not aimed to precisely examine the processes occurring in the *n*-back task performance, but only to generally relate the structure of WM that is most commonly advocated in the literature, we only acknowledge the former, more specific explanation of our data and stay with a more general interpretation. An examination of exact mechanisms due to which *n*-back items are better processed in the focus of attention than outside of it requires a dedicated study.

The last alternative explanation that we considered regards possible differences in coping strategies in the *n*-back task. The use of either low-control or high-control strategies by participants in this task was suggested on the basis of both post-experimental questionnaires (Lovett, Daily, & Reder, 2000) and behavioral indices (Juvina & Taatgen, 2007). As less intelligent participants might have been aware of their worse cognitive functioning and could have perceived the *n*-back task as very difficult, they might have decided not to “invest” their attentional resources but, instead, to exploit some less demanding (but also less effective) processing strategy. Although we cannot fully reject this alternative, as no control over strategies was exerted in our studies, it should be noted that in Lovett et al.’s (2000) study, participants reporting a low-control strategy in the *n*-back task displayed worse performance (than high controllers) in cases of the largest memory load (i.e., the 3- and 4-back conditions) and not in cases of low memory load (i.e., the 1- and 2-back conditions). In Experiments 1–5, we found the opposite results in this regard. Strategy use and adaptation within WM tasks should surely be intensively studied

(for an example, see Braver et al., 2007), but any contribution of processing strategy to the pattern of correlations found in our work seems unlikely.

Conclusion

The *n*-back is a relatively simple task, but processing in this test involves numerous cognitive processes and is very difficult to analyze (Ecker et al., 2010; Oberauer, 2002; Szmalec et al., 2011). However, on the basis of the complex pattern of correlations between reasoning and the *n*-back task performance observed in our experiments, we could identify the two factors, namely, the capacity of attentional maintenance of information, as well as the efficacy of control over distracting representations, which crucially contribute to the level of performance on this task. Most important, the individual differences in both these factors also contributed to fluid reasoning, and they did so in a relatively independent way. So, most probably, one may find people having both capacious WM and some problems with control over WM, as well as low-capacity efficient controllers. An interesting question for future research is whether deficits in one factor can be overcome with the other factor. For example, maybe some low-capacity people are able to carefully select information that enters their foci of attention and then to avoid wasting their scarce capacity on information unrelated to the current task, and vice versa, maybe the cognition of high-capacity individuals can suffer less from breakdowns of control.

Although the aim of this article did not consist of testing any theoretical model of WM, the complex pattern of observed correlations between *Gf* and our WM measures is consistent with the “standard model” of WM (O’Reilly, 2006), which assumes at least the bipartite structure of WM and the control processes operating upon it (e.g., Cowan, 1995; Engle & Kane, 2004; Unsworth & Engle, 2007; Oberauer et al., 2007). Each of these presumed elements of WM yielded a different relation to reasoning: from the strong one in the case of the focus of attention, through the moderate/weak one in the case of control processes, to the zero correlation in the case of activated LTM.

² We thank Klaus Oberauer for suggesting this interpretation.

References

- Ackerman, P. L., Beier, M. E., & Boyle, M. O. (2005). Working memory and intelligence: The same or different constructs? *Psychological Bulletin*, *131*, 30–60. doi:10.1037/0033-2909.131.1.30
- Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G. H. Bower (Ed.), *Recent advances in learning and motivation* (Vol. 8, pp. 47–89). New York, NY: Academic Press.
- Ben-Shakhar, G., & Sheffer, L. (2001). The relationship between the ability to divide attention and standard measures of general cognitive abilities. *Intelligence*, *29*, 293–306. doi:10.1016/S0160-2896(00)00056-8
- Braver, T. S., Gray, J. R., & Burgess, G. C. (2007). Explaining the many varieties of working memory variation: Dual mechanisms of cognitive control. In A. R. A. Conway, C. Jarrold, M. J. Kane, A. Miyake, & J. N. Towse (Eds.), *Variation in working memory* (pp. 76–106). Oxford, England: Oxford University Press.
- Brewin, C. R., & Beaton, A. (2002). Thought suppression, intelligence, and

- working memory capacity. *Behaviour Research and Therapy*, 40, 923–930. doi:10.1016/S0005-7967(01)00127-9
- Burgess, G. C., Braver, T. S., & Gray, J. R. (2006). Exactly how are fluid intelligence, working memory, and executive control related? Cognitive neuroscience approaches to investigating the mechanisms of fluid cognition. *Behavioral and Brain Sciences*, 29, 128–129. doi:10.1017/S0140525X0625903X
- Burgess, G. C., Gray, J. R., Conway, A. R. A., & Braver, T. S. (2011). Neural mechanisms of interference control underlie the relationship between fluid intelligence and memory span. *Journal of Experimental Psychology: General*, 140, 674–692. doi:10.1037/a0024695
- Cattell, R. B. (1971). *Abilities: Their structure, growth, and action*. Boston, MA: Houghton Mifflin.
- Chatham, C. H., Herd, S. A., Brant, A. M., Hazy, T. E., Miyake, A., O'Reilly, R., & Friedman, N. P. (2011). From an executive network to executive control: A computational model of the *n*-back task. *Journal of Cognitive Neuroscience*, 23, 3598–3619. doi:10.1162/jocn_a_00047
- Cho, S., Holyoak, K. J., & Cannon, T. D. (2007). Analogical reasoning in working memory: Resources shared among relational integration, interference resolution, and maintenance. *Memory & Cognition*, 35, 1445–1455. doi:10.3758/BF03193614
- Chuderska, A. (2010). Executive control in analogical mapping: Two facets. In S. Ohlsson & R. Catrambone (Eds.), *Proceedings of the 32nd annual conference of the Cognitive Science Society* (pp. 2749–2754). Austin, TX: Cognitive Science Society.
- Chuderski, A., & Chuderska, A. (2009). Constraints on analogical reasoning: Working memory capacity or executive control? In D. Gentner, K. Holyoak, & B. Kokinov (Eds.), *New frontiers in analogy research* (pp. 62–71). Sofia, Bulgaria: NBU.
- Chuderski, A., & Nečka, E. (2010). Intelligence and cognitive control. In A. Gruszka, G. Matthiws, & B. Szymura (Eds.), *Handbook of individual differences in cognition: Attention, memory, and executive control* (pp. 263–282). New York, NY: Springer. doi:10.1007/978-1-4419-1210-7_16
- Chuderski, A., Stettner, Z., & Orzechowski, J. (2008). Bipartite models of working memory. In V. Sloutsky, B. Love, & K. McRae (Eds.), *Proceedings of the 30th annual meeting of the Cognitive Science Society* (pp. 1531–1536). Austin, TX: Cognitive Science Society.
- Chuderski, A., Taraday, M., Nečka, E., & Smoleń, T. (2012). Storage capacity explains fluid intelligence while executive control does not. *Intelligence*, 40, 278–295. doi:10.1016/j.intell.2012.02.010
- Collette, F., & Van der Linden, M. (2002). Brain imaging of the central executive component of working memory. *Neuroscience and Biobehavioral Reviews*, 26, 105–125. doi:10.1016/S0149-7634(01)00063-X
- Colom, R., Abad, F. J., Rebollo, I., & Shih, P. C. (2005). Memory span and general intelligence: A latent variable approach. *Intelligence*, 33, 623–642. doi:10.1016/j.intell.2005.05.006
- Conway, A. R. A., Cowan, N., & Bunting, M. F. (2001). The cocktail party phenomenon revisited: The importance of working memory capacity. *Psychonomic Bulletin & Review*, 8, 331–335. doi:10.3758/BF03196169
- Conway, A. R. A., Cowan, N., Bunting, M. F., Theriault, D. J., & Minkoff, S. R. (2002). A latent variable analysis of working memory capacity, short-term memory capacity, processing speed, and general fluid intelligence. *Intelligence*, 30, 163–182. doi:10.1016/S0160-2896(01)00096-4
- Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review*, 12, 769–786. doi:10.3758/BF03196772
- 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. doi:10.1016/j.tics.2003.10.005
- Cowan, N. (1995). *Attention and memory: An integrated framework*. New York, NY: Oxford University Press.
- Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, 24, 87–114. doi:10.1017/S0140525X01003922
- Cowan, N., Elliott, E. M., Saults, J. S., Morey, C. C., Mattox, S., Hismatullina, A., & Conway, A. R. A. (2005). On the capacity of attention: Its estimation and its role in working memory and cognitive aptitudes. *Cognitive Psychology*, 51, 42–100. doi:10.1016/j.cogpsych.2004.12.001
- Cowan, N., Fristoe, N. M., Elliott, E. M., Brunner, R. P., & Saults, J. S. (2006). Scope of attention, control of attention, and intelligence in children and adults. *Memory & Cognition*, 34, 1754–1768. doi:10.3758/BF03195936
- Cowan, N., Morey, C. C., Chen, Z., & Bunting, M. (2007). What do estimates of working memory tell us? In N. Osaka, R. H. Logie, & M. D'Esposito (Eds.), *The cognitive neuroscience of working memory* (pp. 43–58). Oxford, England: Oxford University Press.
- Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, 19, 450–466. doi:10.1016/S0022-5371(80)90312-6
- Dempster, F. N., & Corkill, A. J. (1999). Individual differences in susceptibility to interference and general cognitive ability. *Acta Psychologica*, 101, 395–416. doi:10.1016/S0001-6918(99)00013-X
- Duncan, J., Emslie, H., Williams, P., Johnson, R., & Freer, C. (1996). Intelligence and the frontal lobe: The organization of goal-directed behavior. *Cognitive Psychology*, 30, 257–303. doi:10.1006/cogp.1996.0008
- Duncan, J., Parr, A., Woolgar, A., Thompson, R., Bright, P., Cox, S., . . . Nimmo-Smith, I. (2008). Goal neglect and Spearman's *g*: Competing parts of a complex task. *Journal of Experimental Psychology: General*, 137, 131–148. doi:10.1037/0096-3445.137.1.131
- Duncan, J., Seitz, R. J., Kolodny, J., Bor, D., Herzog, H., Ahmed, A., . . . Emslie, H. (2000, July 21). A neural basis for general intelligence. *Science*, 289, 457–460. doi:10.1126/science.289.5478.457
- Ecker, U. H. K., Lewandowsky, S., Oberauer, K., & Chee, A. E. H. (2010). The components of working memory updating: An experimental decomposition and individual differences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36, 170–189.
- Engle, R. W., & Kane, M. J. (2004). Executive attention, working memory capacity, and a two-factor theory of cognitive control. In B. Ross (Ed.), *The psychology of learning and motivation* (Vol. 44, pp. 145–199). New York, NY: Elsevier.
- Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. A. (1999). Working memory, short term memory, and general fluid intelligence: A latent variable approach. *Journal of Experimental Psychology: General*, 128, 309–331. doi:10.1037/0096-3445.128.3.309
- Friedman, N. P., & Miyake, A. (2004). The relations among inhibition and interference control functions: A latent-variable analysis. *Journal of Experimental Psychology: General*, 133, 101–135. doi:10.1037/0096-3445.133.1.101
- Friedman, N. P., Miyake, A., Corley, R. P., Young, S. E., DeFries, J. C., & Hewitt, J. K. (2006). Not all executive functions are related to intelligence. *Psychological Science*, 17, 172–179. doi:10.1111/j.1467-9280.2006.01681.x
- Fukuda, K., & Vogel, E. K. (2011). Individual differences in recovery time from attentional capture. *Psychological Science*, 22, 361–368. doi:10.1177/0956797611398493
- Gottfredson, L. S. (1997). Why *g* matters: The complexity of everyday life. *Intelligence*, 24, 79–132. doi:10.1016/S0160-2896(97)90014-3
- Gray, J. R., Chabris, C. F., & Braver, T. S. (2003). Neural mechanisms of general fluid intelligence. *Nature Neuroscience*, 6, 316–322. doi:10.1038/nn1014
- Gustafsson, J.-E. (1984). A unifying model for the structure of intellectual abilities. *Intelligence*, 8, 179–203. doi:10.1016/0160-2896(84)90008-4
- Halford, G. S., Cowan, N., & Andrews, G. (2007). Separating cognitive capacity from knowledge: A new hypothesis. *Trends in Cognitive Sciences*, 11, 236–242. doi:10.1016/j.tics.2007.04.001

- 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*, 803–831. doi:10.1017/S0140525X98001769
- Hambrick, D. Z., Kane, M. J., & Engle, R. W. (2005). The role of working memory in higher-level cognition: Domain-specific vs. domain-general perspectives. In R. J. Sternberg & J. Pretz (Eds.), *Intelligence and cognition* (pp. 104–121). New York, NY: Cambridge University Press.
- Heitz, R. P., & Engle, R. W. (2007). Focusing the spotlight: Individual differences in visual attention control. *Journal of Experimental Psychology: General, 136*, 217–240. doi:10.1037/0096-3445.136.2.217
- Hockey, A., & Geffen, G. (2004). The concurrent validity and test-retest reliability of a visuospatial working memory task. *Intelligence, 32*, 591–605. doi:10.1016/j.intell.2004.07.009
- Hummel, J. E., & Holyoak, K. J. (2003). A symbolic-connectionist theory of relational inference and generalization. *Psychological Review, 110*, 220–264. doi:10.1037/0033-295X.110.2.220
- Johnson-Laird, P. N. (1999). Deductive reasoning. *Annual Review of Psychology, 50*, 109–135. doi:10.1146/annurev.psych.50.1.109
- Juvina, I., & Taatgen, N. A. (2007). Modeling control strategies in the *n*-back task. In R. L. Lewis, T. A. Polk, & J. E. Laird (Eds.), *Proceedings of the 8th International Conference on Cognitive Modeling* (pp. 73–78). Oxford, England: Psychology Press.
- Kane, M. J., Bleckley, M. K., Conway, A. R. A., & Engle, R. W. (2001). A controlled-attention view of working-memory capacity. *Journal of Experimental Psychology: General, 130*, 169–183. doi:10.1037/0096-3445.130.2.169
- Kane, M. J., Conway, A. R. A., Miura, T. K., & Colflesh, G. J. H. (2007). Working memory, attention control, and the *n*-back task: A question of construct validity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*, 615–622. doi:10.1037/0278-7393.33.3.615
- Kane, M. J., & Engle, M. J. (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. doi:10.3758/BF03196323
- Kane, M. J., & Engle, R. W. (2003). Working-memory capacity and the control of attention: The contributions of goal neglect, response competition, and task set to Stroop interference. *Journal of Experimental Psychology: General, 132*, 47–70. doi:10.1037/0096-3445.132.1.47
- Kane, M. J., Hambrick, D. Z., & Conway, A. R. A. (2005). Working memory capacity and fluid intelligence are strongly related constructs: Comment on Ackerman, Beier, and Boyle (2005). *Psychological Bulletin, 131*, 66–71. doi:10.1037/0033-2909.131.1.66
- Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., & Engle, R. W. (2004). The generality of working memory capacity: A latent variable approach to verbal and visuo-spatial memory span and reasoning. *Journal of Experimental Psychology: General, 133*, 189–217. doi:10.1037/0096-3445.133.2.189
- Kirchner, W. K. (1958). Age differences in short-term retention of rapidly changing information. *Journal of Experimental Psychology, 55*, 352–358. doi:10.1037/h0043688
- Kyllonen, P. C., & Cristal, R. E. (1990). Reasoning ability is (little more than) working memory capacity?! *Intelligence, 14*, 389–433. doi:10.1016/S0160-2896(05)80012-1
- Lovett, M. C., Daily, L. Z., & Reeder, L. M. (2000). A source activation theory of working memory: Cross-task prediction of performance in ACT-R. *Cognitive Systems Research, 1*, 99–118. doi:10.1016/S1389-0417(99)00012-1
- Luck, S. J., & Vogel, E. K. (1997, November 20). The capacity of visual working memory for features and conjunctions. *Nature, 390*, 279–281. doi:10.1038/36846
- Mackworth, J. F. (1959). Paced memorizing in continuous task. *Journal of Experimental Psychology, 58*, 206–211. doi:10.1037/h0049090
- Martínez, K., Burgaleta, M., Román, F. J., Escorial, S., Shih, P. C., Quiroga, M. A., & Colom, R. (2011). Can fluid intelligence be reduced to ‘simple’ short-term storage? *Intelligence, 39*, 473–480. doi:10.1016/j.intell.2011.09.001
- McElree, B. (2001). Working memory and focal attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*, 817–835. doi:10.1037/0278-7393.27.3.817
- Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., & Howerter, A. (2000). The unity and diversity of executive functions and their contributions to complex “frontal lobe” tasks: A latent variable analysis. *Cognitive Psychology, 41*, 49–100. doi:10.1006/cogp.1999.0734
- Mogle, J. A., Lovett, B. J., Stawski, R. S., & Sliwinski, M. J. (2008). What’s so special about working memory? An examination of the relationships among working memory, secondary memory, and fluid intelligence. *Psychological Science, 19*, 1071–1077. doi:10.1111/j.1467-9280.2008.02202.x
- Monsell, S., & Driver, J. (2000). Banishing the control homunculus. In J. Driver & S. Monsell (Eds.), *Attention and Performance XVIII: Control of cognitive processes* (pp. 3–31). Cambridge, MA: MIT Press.
- Nejcka, E. (1999). Learning, automaticity, and attention: An individual differences approach. In P. L. Ackerman, P. C. Kyllonen, & R. D. Roberts (Eds.), *Learning and individual differences: Process, trait, and content determinants* (pp. 161–184). Washington, DC: American Psychological Association. doi:10.1037/10315-007
- Nigg, J. T. (2000). On inhibition/disinhibition in developmental psychology: Views from cognitive and personality psychology and a working inhibition taxonomy. *Psychological Bulletin, 126*, 220–246. doi:10.1037/0033-2909.126.2.220
- 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. doi:10.1037/0278-7393.28.3.411
- Oberauer, K. (2005). Binding and inhibition in working memory: Individual and age differences in short-term recognition. *Journal of Experimental Psychology: General, 134*, 368–387. doi:10.1037/0096-3445.134.3.368
- Oberauer, K., Schultze, R., Wilhelm, O., & Süß, H.-M. (2005). Working memory and intelligence—their correlation and their relation: Comment on Ackerman, Beier, and Boyle (2005). *Psychological Bulletin, 131*, 61–65. doi:10.1037/0033-2909.131.1.61
- Oberauer, K., Süß, H.-M., Schulze, R., Wilhelm, O., & Wittmann, W. W. (2000). Working memory capacity—Facets of a cognitive ability construct. *Personality and Individual Differences, 29*, 1017–1045. doi:10.1016/S0191-8869(99)00251-2
- Oberauer, K., Süß, 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). Oxford, England: Oxford University Press.
- Oberauer, K., Süß, H.-M., Wilhelm, O., & Wittman, W. W. (2008). Which working memory functions predict intelligence? *Intelligence, 36*, 641–652. doi:10.1016/j.intell.2008.01.007
- O’Reilly, R. C. (2006, October 6). Biologically based computational models of high-level cognition. *Science, 314*, 91–94. doi:10.1126/science.1127242
- Orzechowski, J., & Chuderski, A. (2007). *Test analogii obrazkowych* [A figural analogies test]. Unpublished manuscript, Jagiellonian University, Krakow, Poland.
- Pollack, I., Johnson, L. B., & Knaff, P. R. (1959). Running memory span. *Journal of Experimental Psychology, 57*, 137–146. doi:10.1037/h0046137
- Raven, J. C., Court, J. H., & Raven, J. (1983). *Manual for Raven’s Progressive Matrices and Vocabulary Scales (Section 4: Advanced Progressive Matrices)*. London, England: Lewis.
- Roberts, R., & Gibson, E. (2002). Individual differences in sentence

- memory. *Journal of Psycholinguistic Research*, 31, 573–598. doi:10.1023/A:1021213004302
- Rouder, J. N., Morey, R. D., Morey, C. C., & Cowan, N. (2011). How to measure working memory capacity in the change detection paradigm. *Psychonomic Bulletin & Review*, 18, 324–330. doi:10.3758/s13423-011-0055-3
- Salthouse, T. A. (2005). Relations between cognitive abilities and measures of executive functioning. *Neuropsychology*, 19, 532–545. doi:10.1037/0894-4105.19.4.532
- Salthouse, T. A., & Pink, J. E. (2008). Why is working memory related to fluid intelligence? *Psychonomic Bulletin & Review*, 15, 364–371. doi:10.3758/PBR.15.2.364
- Schmiedek, F., Hildebrandt, A., Lövdén, M., Wilhelm, O., & Lindenberger, U. (2009). Complex span versus updating tasks of working memory: The gap is not that big. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35, 1089–1096. doi:10.1037/a0015730
- Schweizer, K., Moosbrugger, H., & Goldhammer, F. (2005). The structure of the relationship between attention and intelligence. *Intelligence*, 33, 589–611. doi:10.1016/j.intell.2005.07.001
- Shamosh, N. A., DeYoung, C. G., Green, A. E., Reis, D. L., Johnson, M. R., Conway, A. R. A., . . . Gray, J. (2008). Individual differences in delay discounting: Relation to intelligence, working memory, and anterior prefrontal cortex. *Psychological Science*, 19, 904–911. doi:10.1111/j.1467-9280.2008.02175.x
- Shelton, J. T., Metzger, R. L., & Elliott, E. M. (2007). A group-administered lag task as a measure of working memory. *Behavior Research Methods*, 39, 482–493. doi:10.3758/BF03193017
- Spearman, C. (1927). *The abilities of man*. London, England: Macmillan.
- Süß, H. M., Oberauer, K., Wittmann, W. W., Wilhelm, O., & Schulze, R. (2002). Working memory capacity explains reasoning ability—and a little bit more. *Intelligence*, 30, 261–288. doi:10.1016/S0160-2896(01)00100-3
- Szmalc, A., Verbruggen, F., Vandierendonck, A., & Kemps, E. (2011). Control of interference during working memory updating. *Journal of Experimental Psychology: Human Perception and Performance*, 37, 137–151. doi:10.1037/a0020365
- Unsworth, N., & Engle, R. W. (2005). Working memory capacity and fluid abilities: Examining the correlation between operation span and Raven. *Intelligence*, 33, 67–81. doi:10.1016/j.intell.2004.08.003
- Unsworth, N., & Engle, R. W. (2006). Simple and complex memory spans and their relation to fluid abilities: Evidence from list-length effects. *Journal of Memory and Language*, 54, 68–80. doi:10.1016/j.jml.2005.06.003
- Unsworth, N., & Engle, R. W. (2007). The nature of individual differences in working memory capacity: Active maintenance in primary memory and controlled search from secondary memory. *Psychological Review*, 114, 104–132. doi:10.1037/0033-295X.114.1.104
- Unsworth, N., Schrock, J. C., & Engle, R. W. (2004). Working memory capacity and the antisaccade task: Individual differences in voluntary saccade control. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30, 1302–1321. doi:10.1037/0278-7393.30.6.1302
- Unsworth, N., Spillers, G. J., & Brewer, G. A. (2010). The contributions of primary and secondary memory to working memory capacity: An individual differences analysis of immediate free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36, 240–247. doi:10.1037/a0017739
- Usher, M., Cohen, J. D., Haarmann, H., & Horn, D. (2001). Neural mechanism for the magical number 4: Competitive interactions and nonlinear oscillation. *Behavioral and Brain Sciences*, 24, 151–152. doi:10.1017/S0140525X01583922
- Van Der Maas, H. L. J., Dolan, C. V., Grasman, R. P. P., Wicherts, J. M., Huizenga, H. M., & Raaijmakers, M. E. J. (2006). A dynamical model of general intelligence: The positive manifold of intelligence by mutualism. *Psychological Review*, 113, 842–861. doi:10.1037/0033-295X.113.4.842
- Vogel, E. K., McCollough, A. W., & Machizawa, M. G. (2005, November 24). Neural measures reveal individual differences in controlling access to working memory. *Nature*, 438, 500–503. doi:10.1038/nature04171
- Waltz, J. A., Knowlton, B. J., Holyoak, K. J., Boone, K. B., Mishkin, F. S., de Menezes Santos, M., . . . Miller, B. R. (1999). A system for relational reasoning in human prefrontal cortex. *Psychological Science*, 10, 119–125. doi:10.1111/1467-9280.00118
- Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language*, 46, 441–517. doi:10.1006/jmla.2002.2864

(Appendix follows)

Appendix

Mean Hit Rates and Mean False Alarms for Lures (Where Available), All With Standard Deviations, of Respective n Conditions of Experiments 1–6

Experiment—Variable	$n = 1$		$n = 2$		$n = 3$		$n = 4$		$n = 5$	
	<i>M</i>	<i>SD</i>								
E1—hit rates			.67	.17	.56	.15	.46	.15		
E2—hit rates	.89	.13	.80	.13	.71	.14	.65	.16		
E3—hit rates	.93	.10	.78	.16	.64	.19	.56	.22	.47	.22
E4—hit rates			.81	.17	.54	.21	.44	.19		
E4—false alarm rates	.18	.20							.23	.16
E5—hit rates			.76	.17	.54	.18				
E5—false alarm rates	.18	.26	.47	.21	.22	.19	.36	.19		
E6—hit rate			.72	.22						
E6—false alarm rate							.20	.22		

Note. E = Experiment.

Received December 30, 2010
Revision received March 10, 2012
Accepted March 21, 2012 ■