Intelligence 40 (2012) 278–295

Contents lists available at SciVerse ScienceDirect

Intelligence

journal homepage:

Storage capacity explains fluid intelligence but executive control does not

Adam Chuderski [a,*], Maciej Taraday [a], Edward Nęcka [a], Tomasz Smoleń [b]

[a] Jagiellonian University, Krakow, Poland
[b] Pedagogical University of Krakow, Poland

ARTICLE INFO

Article history:
Received 15 August 2011
Received in revised form 23 February 2012
Accepted 23 February 2012
Available online xxxx

Keywords:
Fluid intelligence
Working memory
Storage capacity
Executive control

ABSTRACT

We examined whether fluid intelligence (Gf) is better predicted by the storage capacity of active memory or by the effectiveness of executive control. In two psychometric studies, we measured storage capacity with three kinds of task which required the maintenance of a visual array, the monitoring of simple relations among perceptually available stimuli, or the quick update of information. Executive control was measured with tasks reflecting three executive functions, namely attention control, interference resolution, and response inhibition. Using structural equation modeling, we found that all storage tasks loaded on one latent variable, which predicted on average 70% of variance in Gf (Studies 1 and 2). On the contrary, neither interference resolution nor response inhibition was substantially related to Gf or to any other variable (Study 1). Although attention control predicted on average 25% of Gf variance (Studies 1 and 2), when storage capacity was statistically controlled for, attention control no longer significantly explained Gf.

1. Introduction

The last twenty years of research on individual differences in cognition have unquestionably enriched our understanding of fluid intelligence (Gf; also called fluid ability, fluid reasoning, or reasoning ability), one of the most important human abilities, which is closely related to general ability (*g* factor). Gf indicates how well (or how poorly) people reason and solve problems in novel, abstract tasks. The main observation shows that the capacity of working memory (WM), a cognitive mechanism responsible for active maintenance of information crucial for current processing, is the strongest predictor of Gf. Several studies (e.g., Ackerman, Beier, & Boyle, 2002, 2005; Colom, Abad, Rebollo, & Shih, 2005; Conway, Cowan, Bunting, Therriault, & Minkoff, 2003; Engle, Tuholski, Laughlin, & Conway, 1999; Kane et al., 2004; Süß, Oberauer, Wittmann, Wilhelm, & Schulze, 2002) estimated that working memory capacity (WMC), a latent variable being measured with either so-called complex span tasks or batteries of diverse WM tasks, shares a huge amount of common variance with Gf. According to different sources, this amount can be 50% (Kane, Hambrick, & Conway, 2005), 72% (Oberauer, Schultze, Wilhelm, & Süß, 2005), and even 92% (Colom, Rebollo, Palacios, Juan-Espinosa, & Kyllonen, 2004). However, *what* WM tasks really measure and *why* WM and Gf are so strongly related is very much disputed (e.g., Colom et al., 2005; Cowan, 2001; Engle & Kane, 2004; Kane, Conway, Hambrick, & Engle, 2007; Oberauer, Süß, Wilhelm, & Sander, 2007; Unsworth & Spillers, 2010). Several theories proposed different cognitive mechanisms presumed to underlie common variation in working memory and Gf.

In the present paper, we focus on two influential groups of theories of such mechanisms. One group of theories suggest that individual performance in both WM tasks and Gf tests depends on the quality of control over some cognitive processes like directing attention or triggering responses (ability henceforth referred to as *executive control*). On the contrary, the other group of theories propose that the capacity (henceforth called *storage capacity*) to simultaneously maintain the maximum possible amount of information in some kind of active memory is crucial for both WM and Gf. The general aim of our research was to confront the predictions of both approaches by estimating in one structural equations model (SEM) the coefficients of paths leading from latent variables

* Corresponding author at: Institute of Philosophy, Jagiellonian University, Grodzka 52, 31-044 Krakow, Poland.
E-mail address: adam.chuderski@gmail.com (A. Chuderski).

0160-2896/$ – see front matter © 2012 Elsevier Inc. All rights reserved.
doi:10.1016/j.intell.2012.02.010

representing either executive control or storage capacity to the fluid ability.

1.1. Executive control theories of working memory and fluid intelligence

1.1.1. Attention control (goal maintenance)

The attention control (also named executive attention) theory proposed by Engle, Kane, and their collaborators (e.g., Engle & Kane, 2004; Kane et al., 2007) suggests that individual performance in both WM tasks and Gf tests depends on the quality of domain-general control over attention. According to these authors, both highly capacious WM and correct fluid reasoning rely on the effective focusing of attention on task-relevant information, and on the blocking of potential distraction. Subjects with low attention control capabilities suffer from poor maintenance of task goals and from frequent capture by irrelevant stimuli and/or processes.

Some indirect evidence for the executive-attention theory came from strong relationships between Gf and latent variables believed to represent executive control, which were calculated from complex span task scores (Conway et al., 2003; Engle et al., 1999; Kane et al., 2004). More direct evidence consisted of significant correlations between WMC and the indices of attention control derived from tasks more evidently requiring attention. Example indices are error rates in the antisaccade task (Unsworth, Schrock, & Engle, 2004), in the flanker task (Heitz & Engle, 2007), and in incongruent trials of the high-proportion-congruent version of the Stroop task (Kane & Engle, 2003). Other instances include lapses of attention in a psychomotor vigilance task (Unsworth, Redick, Lakey, & Young, 2010) and in dichotic listening task (Colflesh & Conway, 2007). These errors/lapses were shown to result from the loss of the task goal (e.g., “don't read the word but name its color” in the Stroop) from attentional focus, while this goal should be endogenously maintained and protected from some dominant process (i.e., reading a word). When stimuli themselves remind a subject what the task is and goal neglect is unlikely, as in the low-proportion-congruent Stroop, then low-WMC participants commit a similar number of errors as high-WMC ones do (Kane & Engle, 2003). Moreover, several studies have demonstrated significant (though usually moderate) correlations between error rates in attention-demanding tasks and Gf scores (e.g., Buehner, Krumm, & Pick, 2005; Schweizer, Moosbrugger, & Goldhammer, 2005; Unsworth et al., 2010; Unsworth & Spillers, 2010).

1.1.2. Interference resolution

However, Kane and Engle (2003) found that though low- and high-WMC participants do equally well in the low-proportion-congruent condition of the Stroop task, the former demonstrate increased interference effects, measured as a difference in mean latency between incongruent and neutral trials. Moreover, a time to initiate a correct antisaccade, one that must be based on a properly maintained task goal, was also found to significantly predict WMC (Unsworth et al., 2004). So, according to Engle and Kane (2004), the individual efficiency of interference resolution, described as a gating mechanism that prevents distracting stimulation from entering the central stages of cognition and activating improper processes (Wilson & Kipp, 1998), is a distinct control-related factor, which also limits both WMC and Gf.

Some studies have provided evidence on the significant (though usually weak) relation between Gf and latency effects in interference-rich paradigms, such as the Stroop (Dempster & Corkill, 1999), the Navon task (Nęcka, 1999), and the flankers (Unsworth & Spillers, 2010; Unsworth et al., 2010).

1.1.3. Inhibition

Another influential executive control theory, which has been proposed by Hasher, Zacks, and collaborators (Hasher, Lustig, & Zacks, 2007; Hasher & Zacks, 1988; Lustig, May, & Hasher, 2001), points out that inhibitory abilities are the primary determinants of WMC and many other cognitive abilities. Simply, “…what most WM span tasks measure is inhibitory control” (Hasher et al., 2007, p. 231). This stance was based on observations showing that a decrease in imposed interference, which needs to be inhibited, significantly helps aging people to deal with WM tasks (Lustig et al., 2001). It was expected that the same factor may also help low-WMC young adults (Hasher et al., 2007).

In explanations of the results regarding tests of executive control, like the antisaccade, Stroop, vigilance, and dichotic listening tasks, the inhibition theory is consistent with, and indiscriminable from, the attention control and interference resolution theories. According to the inhibition theory, what people scoring high on IQ tests do within tests of executive control is just better inhibition of prosaccades, word reading, ruminations, or irrelevant channel, respectively. However, as the theory pertains to the concept of inhibition in a general way, it must also predict correlations between WMC/Gf and tasks that directly require inhibition. One such established task is a go/no-go paradigm, which consists of a rapid presentation of stimuli and a need to withheld responses for some previously defined no-go stimuli, while quickly responding for the go stimuli. The inhibition theory (see Hasher et al., 2007) suggests that such a motor inhibition may often be depleted, for example during individually varied non-optimal time of day. However, a recent study by Redick, Calvo, Gay, and Engle (2011) showed no significant correlation between the number of correct no-go trials and WMC. In another go/no-go study (McVay & Kane, 2009), mixed results on inhibition-WMC link were observed, as significant correlations were found only for some versions of the task (for a discussion see Redick et al., 2011). Also, Friedman et al. (2006) found no significant correlation between another hallmark test of inhibition, a stop-signal task (Verbruggen & Logan, 2008), and two measures of Gf. So, predictions of the inhibition theory, which regard relation of response inhibition to WM and Gf, have not yet been convincingly corroborated.

1.2. Storage-capacity theories of working memory and fluid intelligence

1.2.1. Capacity as the maximum number of chunks held in working memory

Some influential models of WM assume that a WM structure responsible for holding the most activated and most easily accessible information, called — depending on the particular model — “primary memory” (Unsworth & Engle, 2007), “the focus of attention” (Cowan, 2001), “activation buffer" (Davelaar, Gosher-Gottstein, Ashkenazi, Haarman, & Usher, 2005), or "the region of direct access" (Oberauer, 2002), is heavily limited by the maximum number of information chunks it is able to apprehend simultaneously. According to a thorough review by Cowan (2001), the capacity of such an active memory is four distinct chunks on average. Moreover, people seem to vary regarding storage capacity, which can range from two to six elements.

According to the capacity theory, how much information one can simultaneously maintain and process within WM constitutes the crucial basis of fluid reasoning. For example, Colom et al. (2005) demonstrated that the relation between intelligence and WMC, calculated from complex spans, becomes weak and unstable once a storage component has been partialled out. Moreover, in reanalysis of the data from Kane et al. (2004), Colom et al. showed that the latent variable assumed to reflect attention control (i.e., the one loading on complex spans) was statistically indistinguishable from the latent variable representing STM capacity. Additionally, the latter variable, when calculated from spatial versions of storage tasks, was as good a predictor of Gf as was the variable representing control (see also Colom, Abad, Quiroga, Shih, & Flores-Mendoza, 2008). Cowan, Fristoe, Elliott, Brunner, and Saults (2006) provided evidence that an index of individual storage capacity, estimated with a kind of simple span task which required deciding if one specified element of a visual matrix held in STM did or did not change in a matrix presented on a screen (a so-called two-array comparison task; Luck & Vogel, 1997), shared common variance with IQ scores above and beyond the amount of IQ variance co-shared with attention control.

1.2.2. Capacity as the construction and maintenance of temporary bindings

However, why precisely should capacity limits constrain fluid reasoning? According to Halford, Cowan, and Andrews (2007), the link between capacity and reasoning arises because the number of distinct items simultaneously held in active memory influences the number of relations that can be set between these items, resulting in differences in the complexity of reasoning process that an individual is able to carry out (Halford, Wilson, & Phillips, 1998).

This theory was further specified by Oberauer and his collaborators (e.g., Oberauer, Süß, Wilhelm, & Wittman, 2008; Oberauer et al., 2007; Süß et al., 2002), who proposed that the available capacity to access several items in WM at the same time constrains the process of setting flexible, temporary bindings between chunks held in WM, or between them and their corresponding positions within some mental structure. This structure may include relatively concrete coordinates, for example serial positions during recall, as well as abstract placeholders in some schema or solution's representation. Due to temporary bindings, a person is able to integrate completely new relational structures, which seem 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). According to Oberauer et al. (2008, p. 649), "the common denominator of working memory and other complex tasks such as reasoning could be the capacity to build and maintain […] bindings". In order to measure individual ability to process bindings, Oberauer et al. (2008) used tasks requiring constant monitoring of relations among a few stimuli, like testing if three rhyming words appear in a row, column, or diagonal line of the three-by-three matrix of words, or if any four out of several dots form a square. Oberauer et al. have demonstrated that scores in such tasks are excellent predictors of WMC and Gf, even when stimuli are accessible perceptually and need not be maintained in memory.

1.2.3. Updating

However, even highly capacious WM may not facilitate solving a fluid intelligence test if it contains information no longer relevant for the current step of reasoning. Effective checking and updating of WM contents has been identified on both functional and neuronal levels of analysis and has been assumed crucial for WM effectiveness (Smith & Jonides, 1999). Processes implicated in updating include adding relevant and deleting irrelevant contents of WM as well as altering them according to the requirements of the situation. Updating tasks may also impose control over distractors, when there is interference within updated information and participants have to carefully avoid "lure foils" — items familiar or similar to targets, but related to improper contexts. Several studies have confirmed moderate correlations between scores in updating tasks and Gf (e.g., Colom et al., 2008; Cowan, Morey, Chen, & Bunting, 2007; Friedman et al., 2006; Gray, Chabris, & Braver, 2003; Salthouse, 2005, Schmiedek, Hildebrandt, Lövdén, Wilhelm, & Lindenberger, 2009).

Updating has often been defined as an executive function (e.g., Friedman et al., 2006; Miyake, Friedman, Emerson, Witzki, & Howerter, 2000). However, updating also requires storage (i.e., contents being updated must be somehow maintained). In fact, updating tests have been found to correlate strongly with both complex (Ecker, Lewandowsky, Oberauer, & Chee, 2010; Schmiedek et al., 2009) and simple spans (Colom et al., 2008), so one can expect that updating is not any executive property of WM, but it mostly relies on available storage capacity.

1.3. The present studies

All the aforementioned theories have been corroborated to some extent, and there have been studies that analyzed joint influence of variables regarding control and capacity on Gf (e.g., Buehner, Krumm, Ziegler, & Pluecken, 2006; Colom et al., 2008; Cowan et al., 2006; Miyake, Friedman, Rettinger, Shah, & Hegarty, 2001; Oberauer et al., 2008; Schweizer et al., 2005; Süß et al., 2002; Unsworth & Spillers, 2010). However, to our knowledge, no study to date has accounted for the complete pattern of contributions to fluid intelligence of attention control, interference resolution, response inhibition, storage capacity, and updating. We believe that such a comprehensive test is very much needed. This paper reports two studies serving this aim.

Firstly, as there is some evidence for each of the presented views, only a full picture of the relationships between the proposed constructs enables us to examine if each construct predicts a distinct portion of Gf variance or whether it predicts Gf only because it covaries with some other genuine contributor (see Colom et al., 2008). Secondly, regression analyses (e.g., Cowan et al., 2006), psychometric analyses of reasoning tests (e.g., Embretson, 1995), computational models of Raven's test solver (e.g., Carpenter, Just, & Shell, 1990), and experimental studies of the interactive effects of memory load and cross-mapping interference on reasoning (e.g., Cho, Holyoak, & Cannon, 2007; Chuderska, 2010) have suggested that both capacity and control factors may jointly influence fluid ability. Thus, it is reasonable to investigate the simultaneous contribution of both these factors to Gf.

Another interesting question regards the relationships between the three manifestations of control (i.e., goal maintenance, interference resolution, and response inhibition). Are they separate factors or will they load onto one variable, reflecting a kind of general control ability? Previous studies provided mixed results about links between measures of control, for example showing strong correlation between interference and inhibition (Friedman & Miyake, 2004), but no correlations between scores from very similar interference tasks (Egner, 2008; Shilling, Chetwynd, & Rabbitt, 2002). One last question refers to the relation between storage capacity and updating. Is the updating variable identical to the capacity variable, as it appears to be in complex spans (Ecker et al., 2010; Schmiedek et al., 2009), or does updating reflect abilities above and beyond sheer capacity?

2. Study 1

2.1. Method

2.1.1. Participants

Volunteer participants were recruited via publicly accessible social networking websites, in order to obtain a wide range of intellectual abilities. Each participant provided an informed consent and was paid 30 Polish zloty (approx. 7 euro) for their close to four-hour participation in the study. A total of 98 women and 62 men participated. The mean age was 21.9 years ($SD = 2.7$, range 15–35). All participants had normal or corrected-to-normal vision and were not color blind.

2.1.2. Administered tests

Eleven computerized tasks and two paper-and-pencil intelligence tests were used. Software for the former was designed by the authors of the study. Standard PCs were used to run the tasks. Some features were common for all tasks. If not stated otherwise, all stimuli were presented at or around the center of the computer screen, in black on a light grey background. In no task were there any identical stimuli repetitions that could influence experimental effects (see Hommel, 1998). Except for this constraint, the presentation of trials was fully randomized. In all sessions, there was a start-up trial not included in the results. Depending on the task, one manual response out of one, two, three, or four possible response keys was required in each trial, with the exception that the stop-signal and go/no-go task required no response on stop/no-go trials. In order to minimize WM load imposed by instructions, and especially to avoid confounding capacity and control requirements, in each task involving more than one response key, short descriptions of S-R mappings were present at the bottom of the screen in a location not interfering with task stimuli, for the whole duration of a task. Each task was preceded by a detailed written instruction as well as by several training trials which provided participants with an accuracy feedback. Participants started experimental sessions when they decided they were ready. Each test lasted from 8 to 25 min, depending on the task. For each construct investigated, at least one measure included letter/numerical stimuli, and one contained figural stimuli.

2.1.2.1. Attention control tasks. In order to measure attention control, we used three Stroop-like tasks and the antisaccade task. Each Stroop-like task used bivalent stimuli. The stimuli in the color–word task (MacLeod, 1991) were four capital Polish words (approx. 5 cm × 2 cm in size) naming the colors red, green, blue, and black. Each word could be displayed in any of the four ink colors. In congruent stimuli the meaning of the word and the color of the ink were the same. In incongruent stimuli the former differed from the latter. The stimuli in the figure–word task (Hentschel, 1973) consisted of four figures (square, rhombus, circle, and oval, all approx. 4 cm in size). Each congruent stimulus contained a word meaning this very shape. Incongruent stimuli contained words meaning different shapes. The number Stroop task required counting digits (Fox, Shor, & Steinman, 1971). The task's stimuli were three, four, five, or six exemplars of a digit (0.6 × 0.8 cm in size) drawn from the set [3, 4, 5, and 6]. In congruent trials, the number of digits was in concord with the digits to be counted. In incongruent trials, the former and the latter differed.

Trials lasted until a response was given, or for a maximum of 3 s. The instruction in all three tasks was to avoid reading a word/digit and to press a response key that was assigned to a presented color/shape/number, respectively. Each task was administered in two sessions. In the first session, 90 congruent and 30 incongruent trials were presented. In the second session, 50 congruent and 70 incongruent trials were presented. DV for each task was the ratio of a proportion correct in all incongruent trials to a proportion correct in all congruent trials. The ratios informed us about performance decrease in the control-demanding condition, and not about general performance.

Each trial of the antisaccade task (Hallett, 1978) consisted of four events. First, a cue was presented for 1.5 s, reminding subjects that a target would be presented on the opposite side to the flashing square. Next, a fixation point was presented at the center of the screen for 1–2 s. Then, a rapidly flashing black square was shown on the left or right side of the screen, about 16 cm from the fixation point, for 100 ms. Finally, a small dark grey arrow (0.6 cm long), pointing left, down, or right, was presented in the location opposite the square for only 200 ms and was then replaced by a mask. The visual angle from both the square and the arrow to the fixation point was approximately 14°. The task was to look away from the flashing square, to detect the direction of the arrow, and to press the arrow key matching the stimulus arrow. Fifty trials were administered. DV was a proportion correct.

2.1.2.2. Interference resolution measures. The latencies from the second sessions of the three Stroop-like tasks were used for the calculation of interference measures. As in these sessions the congruent-proportion was lower (42%) than the incongruent-proportion (58%), we expected cases of goal neglect to be rare. Thus, we followed the logic of Kane and Engle (2003, 2004), who assumed that the relatively frequent incongruent trials support goal maintenance to such an extent that attention control need not be involved so much and the increased time of processing incongruent stimuli hardly reflects the time to recover from cases of goal neglect. On the contrary, it mostly represents the additional time needed to resolve the conflict (interference) between two opposite response tendencies. DV was the ratio of a mean RT in incongruent trials to a mean RT in congruent trials. Using ratios informed us about the relative increase of latencies in incongruent trials and allowed us to control for individual processing speed. We did not use neutral trials, as they might have complicated the tasks and thus could have blurred the interference expected to be present in incongruent trials (see MacLeod, 1991). Moreover, calculating the interference effect on the basis of congruent and incongruent trials should amplify the expected negative correlation between interference resolution and Gf, if any exists, as latency in congruent trials was shown to correlate positively with WMC (Kane & Engle, 2003), while latency in incongruent trials correlated negatively with WMC, and as WMC is a predictor of Gf..

2.1.2.3. Response inhibition tasks. Two tests of response inhibition were used: the stop-signal (Logan & Cowan, 1984) and the go/no-go (Eimer, 1993) tasks. In both tasks there was a training session aimed at inducing strong S-R associations. In the stop-signal task, 60 letters (each 1.5×2.0 cm in size) were presented. The participants were asked to press either one button if a letter was one of four specified vowels (i.e., A, E, O, and U) or the other button if a letter was one of four predefined consonants (i.e., B, K, N, and S). In the go/no-go task, 60 digits (1 to 8) were presented and participants were instructed to categorize digits as odd or even. Experimental sessions of each task differed from the training sessions only in that the participants were instructed to withhold their response on some prespecified trials. In the stop-signal task, the stop trials included a black border surrounding a digit stimulus, presented 200 ms after it. There were 120 go and 40 stop trials. In the go/no-go task, the no-go stimuli were the digits "1" or "2". There were 90 go and 30 no-go trials. Trials in each task lasted until a response was given, or for a maximum of 2 s. One-second interval followed each trial.

A note on the measurement of inhibition ability is necessary. Previously, a positive correlation between stop-signal delay (SSD) and the proportion of unsuppressed responses was observed (Logan & Cowan, 1984). So, performance in the stop-signal task was modeled as a race between the response execution process, triggered by a stimulus, and the "stop process", triggered by a stop-signal. In some psychometric studies (e.g., Friedman et al., 2006), individual stop-signal reaction time (SSRT), indicating the speed of stop process, was taken as a measure of inhibition. As a covert process, it had to be estimated indirectly from trials of various SSDs (Band, Molen, & van der Logan, 2003). In our opinion, using SSRT may confound relations among Gf, processing speed, and inhibition, because more intelligent people are generally cognitively faster (Ackerman et al., 2002). For example, for longer SSDs their go processes may be so advanced that they would make more errors than less intelligent people would make, which would influence the calculation of SSRTs. Thus, we think that accuracy of response suppression in the stop trials administered for short SSDs is a much better measure of inhibition than SSRT is. The former was used as DVs in both inhibition tasks.

2.1.2.4. Storage capacity measures. A modified two-array comparison task was used in order to estimate the so-called scope of attention (Cowan et al., 2006). Each of 90 self-paced trials consisted of a pattern of five, six, or seven stimuli. Each stimulus (approx. 2×2 cm in size) was one of ten Greek symbols (e.g., α, β, χ, and so on). The stimuli were presented simultaneously for the time equal to the number of them multiplied by 400 ms and then followed by a black square mask (1200 ms). Finally, another pattern was presented which was either identical to the first one (in a random 50% of trials) or differing by exactly one item (in the remaining trials). If the patterns differed, a square red border highlighted the new item. If the patterns were identical, a random item was highlighted. The task was to decide whether the highlighted item differed or not in two patterns.

We calculated DV for the two-array comparison task with a formula proposed by Cowan et al. (2006): $k = N \times (H - FA)$. It estimates the scope of attention (k) on the basis of the proportions of hits (H, correct responses to patterns with one item changed) and false alarms (FA, incorrect responses to unchanged patterns), on the assumption that a participant produced the correct hit or avoided the false alarm only if a marked item was transferred to her or his active memory (with the k/N chance, where N is a number of stimuli presented). If a non-transferred item was cued, then it was assumed that a participant guessed the answer. DV was a mean from ks for all N conditions. For supraspan Ns, which were used in the present study, the formula is believed to closely approximate the true storage capacity of individuals (Morey, 2011).

In order to measure relational integration, we used two no-memory versions of the monitoring task introduced by Oberauer et al. (2008), and modified them slightly. Each task consisted of the presentation of 80 patterns. Up to 20 patterns included stimuli that fulfilled some simple relation defined in the task's instruction. A participant's goal was to detect the specified relation. Responses given when stimuli did not form such a relation were interpreted as false alarm errors. The verbal version of the task consisted of a three-by-three array of two-letter syllables (3.0×2.5 cm in size), each syllable being composed of a capital consonant and one of four vowels (A, E, O, or U). Two subsequent arrays differed by exactly one syllable. Participants had to look for three syllables ending with the same vowel, located in one row, column, or diagonal line (i.e., this was a simpler version of a "rhyming" task). In the figural version, patterns of eight to twelve black dots (0.7 cm in size) were placed within a virtual ten by ten array. In a subsequent pattern, exactly two dots appeared, disappeared, or moved to other cells. The task was to detect if four dots formed a square (no matter what size it was). No rectangles whose length and width differed by only one dot could appear on the screen, to exclude the possibility of errors resulting from poor perceptual analysis.

In order to rule out the influence of either processing speed or the efficiency of the visual search, we allowed 250% in the letter task (5 s) and 300% in the figural task (6 s) more time for each trial than the time allowed in Oberauer et al. (2008) original study (2 s). Respective DVs were calculated as a number of correct responses for the specified relation minus one third of false alarm errors (as three times more non-target patterns occurred than target patterns).

2.1.2.5. Updating tasks. In order to place high demands on updating, we designed versions of two popular updating tasks that required frequent substitution of items and also included some kind of lure foils. The first test to be applied was a keep-track task (Yntema, 1963). In each trial of the task, participants watched several letters (1.0×1.5 cm in size), each presented serially for 800 ms, and placed in three distinct locations. The instruction was to track the most recent letter from each location and to update in memory a respective location if a new letter appeared there. After the presentation, a mask covering all locations was presented for 800 ms and then three letters were simultaneously shown in all locations. In eight trials, these were the most recent stimuli from respective locations. In another eight trials, all stimuli were different than the most recent stimuli from respective locations. In 16 lure trials, the stimuli were the most recent letters apart from one location, where a relatively recent but not the most recent stimulus was shown. The task was to decide whether the three stimuli were either all the most recent ones or at least one of them was not. DV was calculated as a mean proportion of correct responses in all trials.

The other updating test was the 2-back task. In each of three sessions, a sequence of 88 stimuli, drawn from a pool of 16 figures (each approx. 2.5×2.5 cm in size), was presented serially. Each item was shown for 2 s and followed by a 300 ms mask. There were eight targets in each session, namely figures that were identical to figures presented exactly two stimuli back, which required a response. Participants were informed that eight 3-back repetitions would also be presented and they had to withhold responses to these items. The neighboring stimuli repetitions could not overlap. Apart from targets and lures, no other stimuli could be repeated in less than eight previous trials. DV was calculated as a mean from the proportions of target hits and lure omissions.

2.1.2.6. Fluid intelligence tests. Each participant was given the standard paper-and-pencil test of fluid intelligence: Raven's Progressive Matrices Advanced Version (Raven, Court, & Raven, 1983). The test consists of 36 items that include a three-by-three matrix of figural patterns which is missing the bottom-left pattern, and eight response options which are the patterns which potentially match a missing one. The participant's task is to abstract the rules that govern the distribution of patterns (see Carpenter et al., 1990) and to apply these rules in order to choose the one and only right pattern. Another test of fluid reasoning administered was a test of analogical reasoning designed in our lab (Orzechowski & Chuderski, unpublished manuscript). It includes 36 figural analogies in the form 'A is to B as C is to X', where A, B, and C are types of relatively simple patterns 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, etc.), and X is an empty space. The task is to choose one figure, out of four alternatives, which relates to figure C, as B relates to A. The total numbers of correctly solved items in each test were taken as Raven and Analogies scores, respectively. A written instruction and an example item preceded each test.

2.1.3. General procedure

All computerized tasks were included in one session, which lasted approximately two hours. The order of tasks was fixed, as is the common procedure in such a kind of studies. Firstly, we applied the six tasks measuring control, as these measures may be influenced by automatization and fatigue of participants. Then the five capacity and updating tasks were applied, which seem to be more robust than executive control tests. Otherwise, the order maximized differences between subsequent tasks. The summary of procedure is presented in Table 1. The flanker task was applied, too, but because of a procedural error it was excluded from the study.

After a few minutes break, 40 min were allowed for the Raven test and another 30 min for the figural analogies. The participants could pause for a few minutes after any task. They were invited to consume tea/coffee and candies to maintain optimal levels of arousal and glucose.

The participants were tested in groups of two to four people. The computer room was dimly lit. Participants were encouraged to take comfortable sitting positions and to adjust keyboards and computer screens if necessary. During all computerized tasks each participant was equipped with headphones. Before each test, we verified that participants correctly understood the instructions.

2.1.4. Data screening and analysis

One participant omitted the go/no-go task and this datum was substituted with the DV's mean. No data substantially deviated from a particular DV's distribution, apart from 14 participants in the stop-signal task and another 13 in the go/no-go task, who suppressed no response. These few cases of null data resulted in substantial deviation of both DVs from the normal distribution. However, as we were especially careful that participants had understood the task instructions, these cases could not be attributed to procedural errors and most probably they represented genuine cases of goal neglect. Moreover, excluding these cases from SEM models mentioned below did not substantially change the parameters of these models, so we included these cases in further analyses. Due to ceiling effects, DVs from the stop-signal, go/no-go, and antisaccade tasks were submitted to arcsine transformations. In order to identify possible multivariate outliers, *D* statistic (Cook, 1977) was calculated for each variable. *D* values higher than 1 may indicate participants for whom one variable value influences relations among all the other variables. However, no data surpassed this criterion.

For CFA and SEM computations, we used Statistica software (version 9) with maximum-likelihood estimation. The goodness of fit of CFA and SEM 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

Table 1
The order of tasks and names of related latent variables, as applied in Studies 1 and 2.

Task	No. in Study 1	Variable name in Study 1	No. in Study 2	Variable name in Study 2
Color–word — acc.	1	Attention control		
Color–word — lat.	1	Interference resolution		
Figure–word — acc.	3	Attention control	11	Attention control
Figure–word — lat.	3	Interference resolution		
Number Stroop — acc.	5	Attention control	1	Attention control
Number Stroop — lat. Latency	5	Interference resolution		
Antisaccade	4	Attention control	5	Attention control
Stop-signal	2	Response inhibition		
Go/no-go	6	Response inhibition		
Letter two-array	9	Storage capacity	9	The scope of attention
Numerical two-array			6	The scope of attention
Figural two-array			4	The scope of attention
Monitoring letters	8	Storage capacity	3	Relational integration
Monitoring numbers			12	Relational integration
Monitoring shapes	11	Storage capacity	7	Relational integration
Tracking letters	10	Storage capacity	8	Updating
Counting numbers			10	Updating
Figural *n*-back	7	Storage capacity	2	Updating

Note. In Stroop-like tasks, abbreviation acc. refers to the proportion of accuracy rates, lat. refers to the proportion of latencies. The former was calculated from the first and second sessions of respective tasks, while the latter was based only on the second session. Only the tasks which were further analyzed are included.

(CFI), and the root mean square of approximation (RMSEA). We adopted the following criteria of the good fit of models: χ^2/df should not exceed 1.5, CFI should be higher than .95, and RMSEA should not surpass the value of .06.

2.1.5. Results

Table 2 presents a summary of descriptive statistics and reliabilities for all measures used in the study. Most measures had good or at least acceptable reliability, except for the tracking letters task and the latency in the color–word task (these reliabilities were .45 and .46, respectively). Especially, the latency in the number Stroop was not reliable (.12), and for that reason we discarded it from further analyses.

The matrix of correlations is presented in Table 3. Both Gf tests were strongly correlated ($r = .63$). They were weak to moderate ($rs = .23$–$.50$) but all significant correlations among all storage capacity measures, and between them and each Gf measure ($rs = .23$–$.51$). The attention control measures correlated insignificantly or at best weakly with Gf measures ($rs = .07$–$.22$). Some attention control and capacity measures correlated significantly (up to $r = .32$ in case of the antisaccade and shape monitoring tasks). From the two inhibition measures, only the go/no-go task was weakly related to Gf (mean $r = .22$). Interference resolution DVs did not correlate significantly with Gf. Correlations among DVs reflecting control were not significant, with a few exceptions.

Table 2
Descriptive statistics and reliabilities for measures used in Study 1.

Task	*M*	*SD*	Range	Skew	Kurtosis	Reliability
Color–word acc.	0.94	0.08	0.59–1.20	−1.89	4.67	.97
Color–word lat.	1.13	0.10	0.90–1.44	0.29	−0.09	.46
Figure–word acc.	0.99	0.03	0.89–1.05	−0.91	1.47	.89
Figure–word lat.	1.09	0.10	0.93–1.46	1.08	1.26	.58
Number Stroop acc.	0.97	0.05	0.66–1.08	−0.76	1.17	.95
Number Stroop lat.	1.10	0.09	0.77–1.28	−0.34	0.73	.12
Antisaccade	1.29	0.20	0.66–1.57	−0.62	0.48	.89
Stop signal	1.19	0.43	0.00–1.57	−1.89	2.88	.96
Go/no-go	1.18	0.39	0.00–1.57	2.04	4.01	.91
Letter two-array	3.02	1.35	0.00–5.87	−0.22	−0.58	.86
Monitoring letters	.81	.15	.28–1.00	1.08	1.16	.66
Monitoring shapes	.69	.20	.12–1.00	−0.53	−0.65	.88
Tracking letters	.57	.10	.17–.82	−0.52	1.20	.45
Figural 2-back	.78	.12	.48–1.00	−0.41	−0.43	.54
Raven	22	6.17	5–35	−0.41	−0.19	.85
Analogies	24	5.91	9–34	−0.36	−0.58	.83

Note. $N = 160$ for all tasks. In Stroop-like tasks, abbreviation acc. refers to the proportion of accuracy rates, lat. refers to the proportion of latencies. The stop signal, go/no-go, and antisaccade were arcsine transformed. Reliability = Cronbach's alpha, except for the color–word, figure–word, and number Stroop latency proportions, where reliabilities were split-half (random samples) correlations, adjusted with the Spearman-Brown prophecy formula. For Raven and analogies, reliability values regard a larger pool of participants ($N = 495$), who were tested in our lab with 40 min and 30 min versions of each test, respectively.

Before accounting for fluid reasoning, we tested a measurement model including five intercorrelated latent variables examined in our study, each loaded by respective measures. Additionally, we set a correlation path between error values of two measures regarding the figure–word task, in order to account for the significant correlation between these measures in the correlation matrix. Moreover, in order to eliminate negative variance observed in a preliminary model, we set two error values to one minus reliability of the go/no-go task (i.e., 1.0–.91 = .09) and the figure–word latency effect (i.e., 1.0–.58 = .42), respectively. The resulting model fitted the data very well ($N = 160$, $df = 56$, $\chi^2/df = 1.13$, CFI = .969, RMSEA = .018). However, the path coefficient between storage capacity and updating equaled one, indicating that both variables were statistically indistinguishable and this model was redundant. Thus, we deleted the updating variable and made three storage capacity and two updating tasks load onto one capacity variable. This model's fit was also very good ($N = 160$, $df = 60$, $\chi^2/df = 1.17$, CFI = .957, RMSEA = .022). All task loadings onto respective variables were significant (only in case of the color–word accuracy effect, the loading was marginally significant,

Table 3
Correlation matrix for measures used in Study 1.

Task	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.
1. Color–word acc.	–													
2. Color–word lat.	−.09	–												
3. Figure–word acc.	.11	−.10	–											
4. Figure–word lat.	**−.19**	**.25**	**−.30**	–										
5. Number Stroop acc.	.05	.01	**.17**	−.05	–									
6. Antisaccade	.06	−.05	**.21**	.02	**.21**	–								
7. Stop signal	−.01	−.04	.06	.01	.04	.04	–							
8. Go/no-go	.07	−.04	.12	−.10	.06	.01	**.35**	–						
9. Letter two-array	.08	.03	**.23**	.04	**.20**	**.29**	−.09	.09	–					
10. Monitoring letters	.07	−.02	**.27**	−.03	**.20**	**.21**	.09	**.19**	**.38**	–				
11. Monitoring shapes	.07	.05	.07	−.06	.03	**.32**	.06	**.23**	**.33**	**.38**	–			
12. Tracking letters	.05	−.04	.10	.02	.14	**.25**	.06	.07	**.35**	**.23**	**.32**	–		
13. Figural 2-back	.12	−.10	.10	−.02	.08	.13	.10	.14	**.50**	**.45**	**.34**	**.31**	–	
14. Raven	.15	−.10	**.22**	−.10	.11	.12	.13	**.30**	**.42**	**.44**	**.45**	**.32**	**.49**	–
15. Analogies	**.20**	−.07	.11	−.06	.14	**.22**	.03	.14	**.51**	**.35**	**.41**	**.23**	**.48**	**.63**

Note. $N = 160$ for all tasks. Abbreviation acc. refers to the proportion of accuracy rates, lat. refers to the proportion of latencies. Significant *r*s are marked in bold.

$p = .078$). Only two correlations between latent variables were significant: the one between capacity and attention control ($r = .63$), and the other between capacity and inhibition ($r = .24$). No other path approached the adopted level of significance. Further aggregation of the capacity and the attention control variables substantially decreased the fit of the model ($\Delta\chi^2 = 7.1$, $\Delta df = 3$), thus the four-variable measurement model was adopted as the resulting one.

In the search for the genuine predictors of Gf, we tested a few structural equation models. Firstly, we tested a model (Model 1) that included four intercorrelated exogenous variables, identified in the measurement model, and an endogenous (to-be-predicted) variable loaded by scores in both Gf tests. The model fitted well (N = 160, $df = 81$, $\chi^2/df = 1.23$, CFI = .956, RMSEA = .030). We present Model 1 in Fig. 1. No significant paths leading to Gf either from attention control, interference resolution, or response inhibition could be observed ($|rs| < .23$; $ps > .11$), while the path leading from storage capacity to reasoning ability ($r = .99$) was not statistically different from unity. However, Model 1 included some collinearity, specifically a strong correlation between storage capacity and attention control, which could influence their path coefficients leading to Gf.

A good method (see Colom et al., 2008; Oberauer et al., 2008) to test collinear Gf predictors against each other is to take one of the predictors as an endogenous variable, and to test how much variance common to both endogenous variables was left unexplained by the remaining predictor, by correlating disturbance terms of endogenous variables (i.e., their residual variances). In Model 2A, attentional control predicted the storage capacity and fluid intelligence variables. We tested if the residuals of capacity and Gf shared variance. It may be partially attributed to sheer maintenance of information, limited by capacity. Model 2A yielded an acceptable fit (N = 160, $df = 41$, $\chi^2/df = 1.45$, CFI = .949, RMSEA = .048). Attention control significantly predicted both capacity ($r = .63$) and fluid intelligence ($r = .43$). However, the correlation between capacity and Gf residuals was substantial ($r = .60$), indicating that 36% of their shared variance was left unexplained by attention control. In an analogous Model 2B (its fit was equal to Model 2A, as the structures of both models were identical), in which storage capacity was a predictor, correlation between attention control and Gf residuals, which could be associated with variance in executive control that was unexplained by the predictor, was not significant ($r = -.13$, $p = .202$). On the contrary, capacity was a very good predictor of attentional control ($r = .63$) and — especially — of fluid intelligence ($r = .87$), explaining more than 75% of variance in the latter variable. Models 2A and 2B are presented in Fig. 2.

In order to exclude a possible explanation of the observed data which postulates that storage capacity could be such an effective Gf predictor because it was loaded by two updating tasks, we estimated Model 2B′, which differed from Model 2B in that the two updating measures were excluded. The fit of Model 2B′ (N = 160, $df = 24$, $\chi^2/df = 1.53$, CFI = .951, RMSEA = .055) was virtually the same as the fit of Models 2A/2B, as was the path from capacity to Gf ($r = .91$). This time, capacity explained 82% of variance in fluid intelligence.

2.1.6. Discussion

Firstly, our results regarding the structure of control functions are in concord with previous studies, showing that different types of inhibition and conflict resolution are specific and distinct executive functions, tending to operate at a low level of human cognitive architecture (e.g., Egner, 2008; Keye, Wilhelm, Oberauer, & van Ravenzwaaij, 2009; Nigg, 2000; Shilling et al., 2002). Only Friedman and Miyake (2004) found a strong correlation between response inhibition and interference resolution, however in their study the Stroop task was chosen to load onto the response inhibition variable and it was loaded more strongly than the other two inhibition tasks (the antisaccade and stop-signal), while three tasks conceptually similar to the Stroop were chosen to load onto the interference variable. So, the correlation observed in Friedman and Miyake's study might have showed up only because the inhibition variable significantly reflected interference. In our study, no correlation between interference and inhibition was found. Attention control was not related to any of these variables either. Thus, no general executive control factor could be found.

Moreover, inhibition and interference did not account for any meaningful portion of Gf variance. Attention control did moderately predict fluid intelligence, but the analysis of

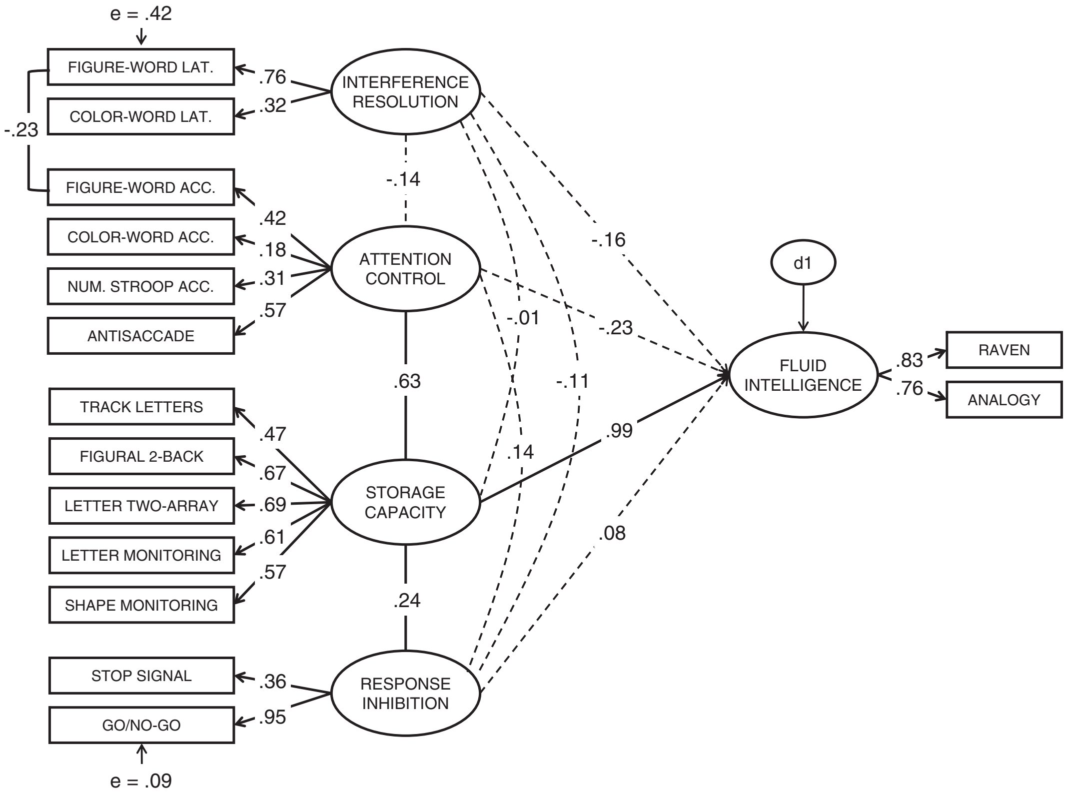

Fig. 1. The structural equation model (Model 1) relating the attention control, interference resolution, response inhibition, and storage capacity exogenous latent variables to the fluid intelligence endogenous latent variable. Boxes represent manifest variables, while large ovals represent latent variables. The small oval represents a disturbance term. Values between ovals and boxes represent relevant standardized factor loadings (all $ps<.01$, except for marginally significant loading of the color–word accuracy). Values between ovals represent path coefficients among latent variables. Error values of the go/no-go and the figure–word latency, which were fixed on the value of one minus reliability, are indicated by small arrows above/below the respective boxes. An arrow linking the figure–word accuracy and latency represents correlation of error terms for the figure–word task. The left part of the figure represents the measurement model for Study 1.

structural equations suggests that it was not a genuine predictor of Gf; rather, it correlated with intelligence only because it covaried with storage capacity.

Our results suggest that variance in updating can be perfectly predicted by variance in capacity. Updating did not account for any amount of Gf variance above and beyond the

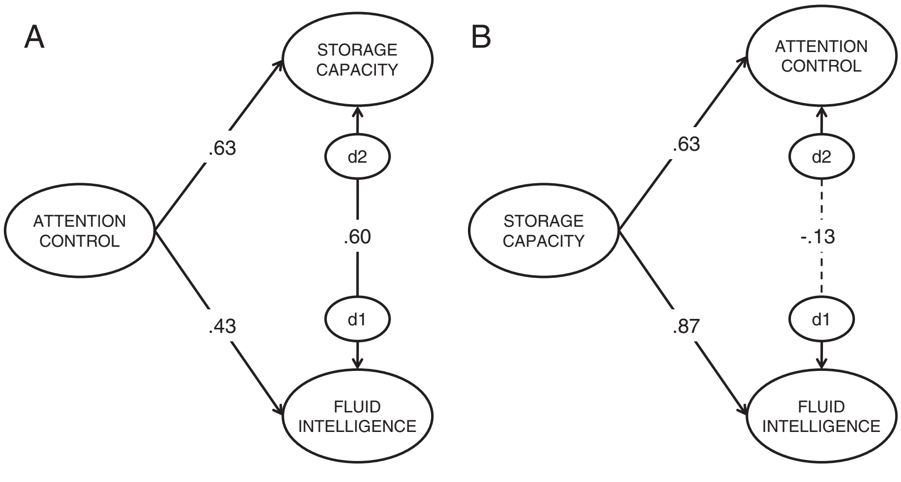

Fig. 2. Panel A: SEM Model 2A predicting the storage capacity and fluid intelligence endogenous variables with the attention control variable, and correlating disturbance terms for the two endogenous variables. Panel B: SEM Model 2B predicting the attention control and fluid intelligence endogenous variables with the storage capacity variable, and correlating disturbance terms for the two endogenous variables. In both panels, large ovals represent latent variables, small ovals represent disturbance terms, solid lines indicate paths significant at $p<.01$ level, while dashed lines represent insignificant paths ($p>.10$).

variance accounted for by scores reflecting maintenance of the pattern of a few items for a two-three seconds (i.e., the two-array comparison task) or construction and maintenance of temporary bindings among perceptually available items (i.e., the two monitoring tasks). So, the existence of a distinct executive function of updating is highly questionable. It seems that what our updating tasks measure, even though they include distracting stimuli (lures), is nothing more than storage capacity. A general WM storage capacity factor was easily identified, and it was properly loaded by varied WM tasks, which overtly seemed to put differing requirements on the cognitive system. Most importantly, storage capacity accounted for a substantial part of Gf variance. Thus, storage capacity seems to be the crucial factor determining fluid intelligence.

3. Study 2

In Study 1, the only genuine predictor of Gf, namely the storage capacity variable, was measured with only five tasks reflecting three different, though conceptually similar, cognitive processes (i.e., array comparison, relation monitoring, and information updating). In Study 2, we measured each of them with three tasks, so we were able to investigate the links between the related constructs and between them and Gf. We aimed to confirm that both the two-array comparison tasks and monitoring tasks would load on one variable, reflecting general storage capacity. We expected that the updating variable, again, would be fully dependent on storage capacity and no distinct updating function could be identified.

Moreover, we aimed to impose a higher load on attention control, in order to rule out the possibility that in Study 1 our attention control tasks insufficiently involved executive control. Attention control requirements of the Stroop/antisaccade tasks were increased due to the introduction of neutral/prosaccade trials (see Kane & Engle, 2003; Unsworth et al., 2004).

3.1. Method

3.1.1. Participants

The recruitment procedure was the same as in Study 1. This time, each participant was paid 50 Polish zloty (approx. 12 euro) for their seven-hour participation in the study. A total of 107 women and 71 men participated. The mean age was 22.6 years (*SD* = 3.3, range 17–39). All participants had normal or corrected-to-normal vision and were not color blind. All of them gave informed consent.

3.1.2. Administered tasks

In Study 2, we no longer investigated response inhibition or interference resolution, as they failed to predict any Gf variance, though for exploratory reasons we applied the Navon, stop-signal, and go/no-go tasks (not reported any further). We focused on the measurement of attention control, the scope of attention, relational integration, and updating. Twelve computerized tasks were used. The changes introduced to the battery of tasks in comparison to Study 1 are described below.

3.1.2.1. Attention control tasks. We used the figure–word, number Stroop, and antisaccade tasks. Neutral trials were introduced in the two former tasks, which consisted of figures filled with XXXXs or a few zero digits, respectively. In each task, 40 congruent, 40 neutral, and 60 incongruent trials were administered in the similar procedure as in Study 1. DVs were the ratios of the proportion of correct in incongruent trials to the proportion of correct in congruent and neutral trials. In the antisaccade, in a random half of all 80 trials we applied antisaccade trials, and the remaining ones were prosacccade trials, in which both the flashing square (this time presented for 200 ms in all trials) and the arrow (shown for 300 ms in all trials) appeared in the same location. DV was the ratio of the proportion of correct in antisaccade trials to the respective proportion in prosaccade trials.

3.1.2.2. Scope of attention tasks. We used the letter two-array comparison task and two other tasks of the same design. One task used two-digit numbers, and the other had figures as stimuli. Because encoding of figural stimuli may be more difficult, the figural task included set sizes of four, five, and six items, and the presentation time was increased to 500 ms per item, in comparison to the letter and number tasks. All three DVs were calculated as in Study 1.

3.1.2.3. Relational integration tasks. Along with the letter and shape monitoring tasks, which were applied as in Study 1, we also used the numerical monitoring task. It was identical to the letter task apart from the fact that the patterns consisted of three-digit numbers and the task required monitoring for three numbers ending with the same digit appearing in a row, column, or diagonal line. DVs were calculated as in Study 1.

3.1.2.4. Updating tasks. As in Study 1, we used the letter tracking task, this time with stimuli appearing either in three or in four locations, and the 3-back task (this time 2-back and 4-back stimuli repetitions were lures), with other task details and DV calculation analogous as in Study 1. We also used a version of a mental counter task (Larson, Merritt, & Williams, 1988). In this task, a stream of 180 stimuli (the numbers: “2”, “3”, or “4”, each 1.0 × 1.5 cm in size) was presented serially for 2.5 s apiece. The task was to count how many times each number was presented and to detect its fourth appearance. After this event, the participants had to reset their counter regarding a particular number and to start counting it from zero. If they missed the fourth occurrence, the cue to reset the respective counter was displayed. DV was the proportion of correct responses.

3.1.2.5. Fluid intelligence tests. Identical tests were used as in Study 1.

3.1.3. General procedure

Each participant was examined within one day, in three sessions. The first and third sessions consisted of computerized tasks, and were separated by a session including the Raven and figural analogies (and two questionnaires of anxiety, not related to the present study) as well as by a 20 minute break. As in Study 1, all tasks were administered in a fixed order, but as Study 2 lasted longer than Study 1, additional care was taken to minimize any effects of the procedure on the results. Firstly, to familiarize the participants with experimental settings and the computers, a start-up task (an easy task-switching test lasting a few minutes) was applied before starting the main experiment. Secondly, three tasks used to measure each latent variable were balanced in relation to the start-end of the experiment, so, on average, the amount of learning and the level of fatigue present during the measurement of each variable were comparable. Two consecutive tasks always measured different functions. The order of tasks in Study 2 and its comparison to Study 1 is presented in Table 1.

The participants were allowed 60 min in the Raven and 60 min in the figural analogies. These times were increased in comparison to Study 1 in order to rule out the possible interpretation that correlations between Gf and the computerized tasks could have arisen due to time pressure regarding intelligence tests (as the computerized tasks, even the self-paced ones, involve rapid presentation of stimuli). All other procedural details were the same as in Study 1.

3.1.4. Data screening and analysis

Only seven observations, which strongly deviated from DVs distributions, were noticed in data (in total of 178 participants multiplied by twelve DVs), and each such observation was substituted with mean either plus or minus three *SD*s, depending on whether it was above or below the mean, respectively. No data surpassed $D>1$ criterion for Cook's statistics. The cutting criteria for models were the same as in Study 1.

3.1.5. Results

Table 4 presents descriptive statistics and reliabilities for all measures used in Study 2. All measures had good reliability, except for the tracking letters task ($\alpha=.53$).

The matrix of correlations is presented in Table 5. All measures correlated at least marginally, with a few exceptions for the figure–word and numerical Stroop tasks.

Analogously to Study 1, an initial measurement model including intercorrelated latent variables, each loaded by respective measures, revealed a correlation between the scope of attention and updating not significantly different from unity ($r=.95$, $SE=.08$). One way to deal with this finding could be to collapse both variables into one latent variable, which would be loaded by the three two-array comparison and the three updating tasks. However, as we wanted to test a variable which could be unequivocally interpreted as reflecting the scope of attention, and also taking into account the results from Model 2B′ (i.e., no contribution of updating to Gf above the contribution of the scope), we decided to exclude updating measures from consecutive models.

The three-variable measurement model fitted very well (N = 178, $df=24$, $\chi^2/df=1.26$, CFI = .983, RMSEA = .034), with all measures loading on respective variables at least at $p<.02$ level. The model is depicted in Fig. 3. However, all variables were strongly correlated. Relational integration was related to both the scope ($r=.60$) and the control ($r=.67$) of attention, and the two latter were related even more strongly ($r=.87$). Fixing this latter correlation at unity did not decrease the fit ($\Delta\chi^2=0.6$; $\Delta df=1$), which suggests that one

Table 4
Descriptive statistics and reliabilities for measures used in Study 2.

Task	*M*	*SD*	Range	Skew	Kurtosis	Reliability
Figure–word	0.99	0.04	0.83–1.12	− 1.03	3.75	.93
Number Stroop	0.98	0.05	0.76–1.12	− 0.40	2.95	.91
Antisaccade	0.90	0.10	0.48–1.06	− 1.47	3.17	.79
Letter two-array	2.81	1.52	0.00–5.78	− 0.20	− 0.99	.88
Number two-array	3.01	1.10	0.00–5.60	− 0.29	− 0.29	.78
Figural two-array	2.85	1.01	0.00–4.78	− 0.51	− 0.07	.84
Monitoring letters	.74	.16	.15–1.00	− 0.82	0.16	.88
Monitoring numbers	.63	.17	.07–1.00	− 0.67	0.38	.76
Monitoring shapes	.63	.18	.18–0.99	− 0.16	− 0.65	.89
Tracking letters	.70	.12	.31–.97	− 0.32	0.02	.53
Counting numbers couters	.33	.16	.00–.73	0.29	− 0.46	.97
Figural 3-back	.54	.17	.09–.94	− 0.02	− 0.53	.84
Raven	23.83	5.63	9–36	− 0.34	− 0.10	.90
Analogies	26.35	4.76	11–35	− 0.61	− 0.05	.88

Note. $N=178$ for all tasks. Reliability = Cronbach's alpha. For Raven and analogies, reliability values regard a larger pool of participants ($N=324$), who were tested in our lab with 60 min versions of each test.

set of measures reflecting either the control or scope of attention was redundant. On the contrary, relational integration could not be equated with these variables, as the one-factor model was unacceptable (N = 178, $df=27$, $\chi^2/df=2.33$, CFI = .903, RMSEA = .086).

Again, we did not want to collapse the scope and control latent variables, as we aimed to interpret scope/control in a way consistent with studies and theories which introduced these constructs (Cowan et al.'s and Engle et al.'s works, respectively). We instead thought that a redundant set of measures should be eliminated. In order to test how the respective tasks predict intelligence, while in part getting rid of redundancy, we verified two alternative SEM models. Model 3A included the relational integration and attentional control exogenous variables, which both predicted a Gf endogenous variable. In Model 3B, Gf was predicted by relational integration and the scope of attention. Both models are presented in Fig. 4. Model 3A fitted very well (N = 178, $df=17$, $\chi^2/df=1.09$, CFI = .992, RMSEA = .021). However, relational integration was the only significant ($r=.65$) predictor of Gf, while the path from attentional control to Gf was weak ($r=.16$, ns.). Attention control significantly predicted Gf ($r=.57$) only when it was left as the sole Gf predictor. The fit of Model 3B was excellent (N = 178, $df=17$, $\chi^2/df=0.78$, CFI = 1.0, RMSEA = .00), and in this model Gf was significantly predicted by both the scope of attention ($r=.28$) and relational integration ($r=.58$).

Having eliminated the measures of attentional control from the Gf model, and aiming to estimate amounts of Gf variance accounted for by the scope of attention and relational integration, we calculated Model 4 which included three exogenous variables, one representing general storage capacity (loaded by all six tasks), and two representing variance

Table 5
Correlation matrix for measures used in Study 2.

Task	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.
1. Figure–word	–												
2. Number Stroop	**.15**	–											
3. Antisaccade	.09	**.19**	–										
4. Letter two-array	.09	**.19**	**.35**	–									
5. Number two-array	**.16**	**.20**	**.43**	**.61**	–								
6. Figural two-array	**.19**	.14	**.53**	**.57**	**.67**	–							
7. Monitoring letters	.00	**.22**	**.31**	**.33**	**.30**	**.35**	–						
8. Monitoring numbers	.14	.06	**.28**	**.24**	**.30**	**.36**	**.45**	–					
9. Monitoring shapes	.04	−.02	**.22**	**.23**	**.17**	**.21**	**.27**	**.32**	–				
10. Tracking letters	.10	**.17**	**.33**	**.51**	**.42**	**.37**	**.30**	**.31**	**.20**	–			
11. Counting numbers	.01	.09	**.23**	**.42**	**.39**	**.35**	**.31**	**.37**	**.20**	**.28**	–		
12. Figural 3-back	.00	.04	**.32**	**.26**	**.32**	**.36**	.14	.12	.11	.12	**.35**	–	
13. Raven	.05	.05	**.33**	**.36**	**.35**	**.44**	**.35**	**.36**	**.32**	**.19**	**.39**	**.17**	–
14. Analogies	.10	.12	**.28**	**.31**	**.37**	**.37**	**.39**	**.35**	**.27**	**.21**	**.40**	**.13**	**.56**

Note. N = 178 for all tasks. Significant *r*s are marked in bold.

specific to the two-array comparison tasks and the monitoring tasks, respectively. This model had an excellent fit (N = 178, $df = 11$, $\chi^2/df = 0.76$, CFI = 1.0, RMSEA = .00) and is presented in Fig. 5. The model indicates that the two-array comparison tasks and the relation monitoring tasks properly loaded onto one storage capacity variable, and they jointly accounted for 52% of variance in intelligence, while the scope of attention and relational integration (though this variable was significantly loaded only by the shape monitoring task) separately explained an additional 5 and 1% of Gf variance, respectively.

3.1.6. Discussion

Variance in attention control was not distinguishable from variance in the scope of attention. Replicating the results of Study 1, the attention control tasks were poor

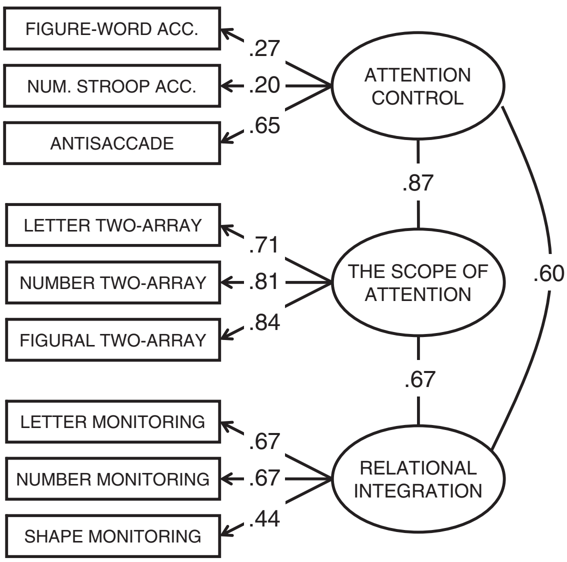

Fig. 3. The measurement model for Study 2, including the scope of attention, attention control, and relational integration latent variables. Boxes represent manifest variables, while ovals represent latent variables. Values between ovals and boxes represent relevant standardized factor loadings (all *p*s<.02). Values between ovals represent path coefficients among latent variables.

predictors of fluid ability. On the contrary, the storage capacity measures were excellent predictors of intelligence. The scope of attention and relational integration variables jointly contributed to more than half of Gf variance. Also, the prediction that the updating variable is statistically identical to the storage capacity variable has been fully confirmed.

4. General discussion

By the means of structural equation modeling, this work examined whether fluid intelligence is best explained by the three postulated executive control abilities (attention control, interference resolution, and response inhibition), by the capacity to store and update items and their bindings in active memory, or both. We also asked a question about relations among the control functions, and we tested the structure of variables reflecting different ways to operationalize the limitations of the active maintenance in working memory.

4.1. Relations among attention control, interference, inhibition, capacity, updating, and intelligence

In Study 1, we found that interference resolution and response inhibition were not related to fluid intelligence, neither in the case of direct measures (with an exception for the go/no-go task) nor — especially — in the case of latent variables. This null result should not be a surprise in face of the previous findings showing the lack of relations between executive functions and Gf (e.g., Colom et al., 2008; Friedman et al., 2006; Oberauer et al., 2008; Redick et al., 2011). Inhibition and interference were not significantly correlated nor related to attention control, so no general-domain control factor could be identified in Study 1. Inhibitory process cannot be considered as responsible for any Gf variance, contrary to previous suggestions made by Kane and Engle (2003) and Hasher et al. (2007). Although some other data suggest that inhibition of representations processed during reasoning takes place (e.g., Chuderska, 2010; Gentner & Toupin, 1986; Richland, Morrison, & Holyoak, 2006), any role of inhibition in higher cognitive processes seems to be more than just suppressing stimuli or responses and it surely needs more investigation in the future (MacLeod, 2007).

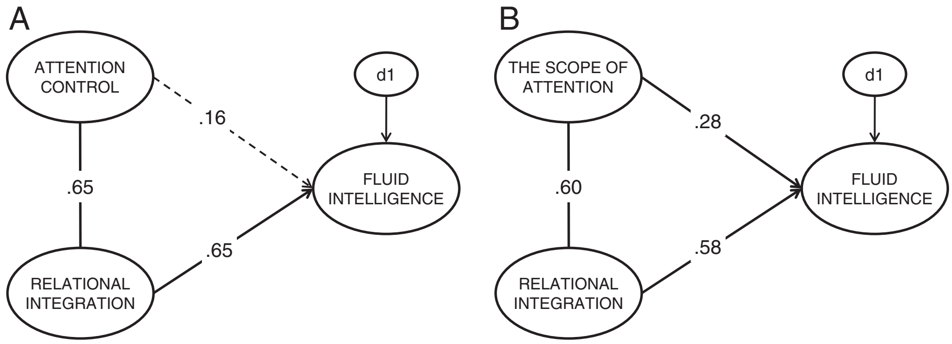

Fig. 4. Panel A: SEM Model 3A predicting the fluid intelligence endogenous variable with the attention control and relational integration variables. Panel B: SEM Model 3B predicting the fluid intelligence endogenous variable with the scope of attention and relational integration variables. In both panels, large ovals represent latent variables, small ovals represent disturbance terms, solid lines indicate paths significant at $p<.01$ level, while dashed lines represent insignificant paths ($p>.10$).

On the contrary, in both our studies, measures of attention control, storage capacity, and updating were correlated with fluid intelligence on both manifest and latent variable levels. However, the main contribution of this paper to the understanding of the cognitive basis of fluid intelligence consists of showing that the relation between fluid ability and attention control is no longer significant when the storage capacity is being statistically controlled for. When only attention control predicted intelligence, it indeed accounted for 18% (Study 1) and 33% (Study 2) of variance in Gf. These values are comparable to values obtained in studies aimed to corroborate the executive-attention theory of Gf (e.g., Unsworth & Spillers, 2010; Unsworth et al., 2010). That supports our claim that attention control was properly measured in the present study. However, attention control strongly correlated with storage capacity, and when capacity was entered as a Gf predictor, the link between attention control and fluid intelligence disappeared (see Models 1, 2B, and 3A). On the contrary, in all models presented in this paper, storage capacity was an excellent predictor of fluid intelligence, explaining from 58% (Study 2) to 82% (Study 1) of Gf variance. This last result replicates and extends previous findings showing that some form of active maintenance of information may be the fundamental mechanism underlying human intelligence (e.g., Buehner et al., 2005, 2006; Colom et al., 2008; Cowan et al., 2006; Oberauer et al., 2008).

Studies 1 and 2 raise a very interesting question: to what extent and why do the control and scope of attention correlate? A seminal work investigating the relation between these two attentional mechanisms, done by Cowan et al. (2006), indicates that they share about 12% of common variance and jointly contribute to a similar amount of variance in Gf, while the scope was able to explain another 15% of Gf variance, and the control — 10%. Cowan et al. concluded that although moderately related, both the scope and control of attention represent separate mechanisms. However, their

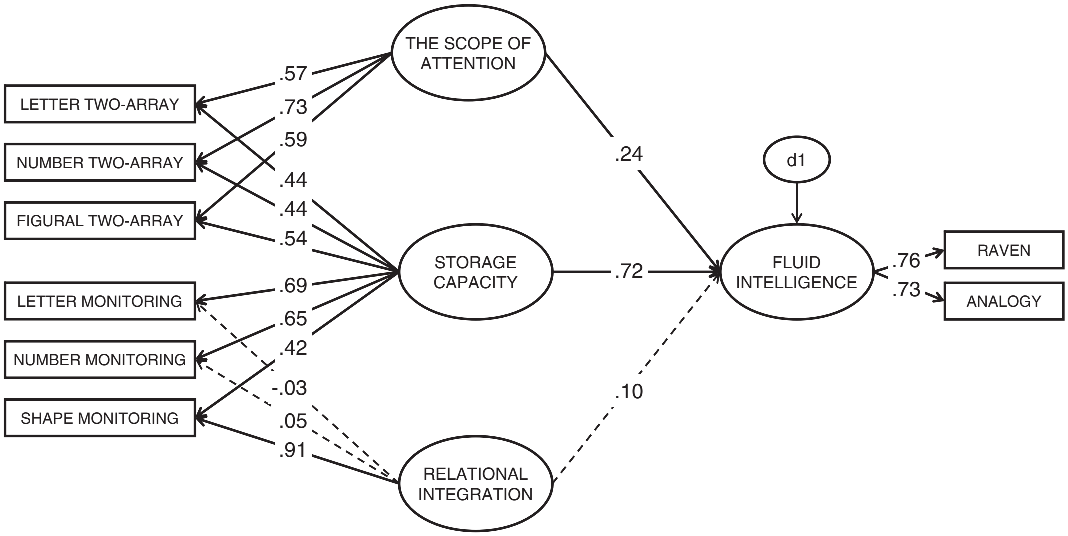

Fig. 5. SEM Model 4 predicting the fluid intelligence endogenous variable with the general storage capacity variable (loaded by all six tasks), and two variables representing variance specific to the two-array comparison tasks and the monitoring tasks, respectively. Boxes represent manifest variables, while large ovals represent latent variables. The small oval represents a disturbance term. Values between ovals and boxes represent relevant standardized factor loadings (all $ps<.01$, except for insignificant loadings of the two monitoring tasks onto the relational integration variable). Values between ovals represent path coefficients among latent variables. Solid lines indicate paths significant at $p<.01$ level, while dashed lines represent insignificant paths ($p>.10$).

conclusion was based on a single measure for each construct. So, task-specific variance could be responsible for the fact that the amount of variance shared by the scope and control of attention was only moderate. In our study, both constructs were measured by at least three tasks and were reflected on a latent level, which probably eliminated the effects of imperfect reliabilities and task-specific noise of these measures. So, it seems to us that the control and scope of attention are much more strongly related than Cowan et al. suggested.

One explanation of such a strong relation could be that some other factor, possibly a lower-level characteristic of attention, determines the effectiveness of both these manifestations of attention. However, in light of existing data and theories, it is difficult to identify such a factor, especially as neural mechanisms of the control versus scope of attention seem to be located in distinct structures of the brain (i.e., in the prefrontal versus parietal cortices, respectively; Osaka & Osaka, 2007).

A more plausible explanation suggests that accuracy in the incongruent/antisaccade conditions of the Stroop/antisaccade tasks does not reflect attention control in terms of the strong focusing on a task goal, but it reveals the capacity to actively maintain various task instructions and S-R mappings in WM. Such an interpretation may be drawn from Engle and Kane's (2004) own definition of executive attention, which is supposed to be "… the ability to maintain stimulus and response elements in active memory, particularly in the presence of events that would capture attention away from that enterprise" (p. 192). This definition says that (a) maintenance of task-relevant mappings is crucial for (measures of) executive attention and (b) this is an especially important process in face of distraction. However, maintenance may be much more important than handling distraction and mostly the former may determine scores in attention control tasks. Also, even if distraction matters, it is not certain how exactly it impairs maintenance. One possibility is that it may delete in an all-or-none manner all task-relevant chunks from WM because it has captured all attention (this seems to be Kane and Engle's default interpretation). Alternatively, distractors may gradually enter attention, "crowding out" task-relevant chunks one by one. If so, then individuals having larger scopes would be less impaired by distraction than people having smaller scopes because, given the same amount of distracting information, they would retain more slots occupied with information necessary for the task at hand. The large scope people may handle at the same time both relevant information needed to fulfill the task and distractors leaking into the scope.

In order to support this hypothesis, we reanalyzed the antisaccade task data from Studies 1 and 2 (procedural differences in antisaccade condition between both studies were negligible). We divided the participants into four groups, according to how many chunks of information they were able to maintain in the letter two-array comparison task, in order to see if the participants having the smallest scopes were especially distracted by the flashing square. Into the 1-chunk group, we included the participants who scored k estimates of 1.5 or less, which mean that they maintained on average around one chunk. The 2-chunk group's k estimates were between 1.5 and 2.5, the 3-chunk group scored ks between 2.5 and 3.5, and the 4 and 5-chunk group included people scoring ks higher than 3.5 (the 4- and 5-chunk groups were merged because only several participants scored ks higher than 4.5). Then, we analyzed the effect of the k estimate on the antisaccade accuracy (see Fig. 6), which was significant, $F(3, 334) = 16.11$, $\eta^2 = .13$, $p < .001$. However, the only significant contrast was between the 1-chunk and 2-chunk groups, $F(1, 334) = 20.89$, $p < .001$, while no significant differences in accuracy were found between the 2-, 3-, and 4 and 5-chunk groups (all $ps > .2$). So, maintaining the instruction of the antisaccade task (i.e., "look away from the cue") seems to be a tough activity only for people whose scope can handle only one chunk, which may be easily replaced by a distractor. Therefore, the limits of the scope of attention, and not the strength of attention control, seem to be the main source of individual differences in the antisaccade task. If so, then some conclusions drawn from the studies on that task (e.g., Unsworth, Spillers, Brewer, & McMillan, 2011; Unsworth et al., 2004) may require reconsideration. More generally, whether executive-attention tasks successfully predict scores in WM and Gf tests may depend on the number of elements these tasks require in WM (for a similar conclusion see Duncan et al., 2008).

Similar to attention control, the updating function does not seem to be a genuine Gf predictor either. Updating appeared to reflect exactly the same ability as the one measured by storage capacity. Because the two-array comparison tasks capturing storage limitations seem to be much simpler tools than updating tasks, as the former neither require manipulating the processed stimuli, nor do they include any interfering items (e.g., lures), as updating tasks do, we believe that a more complex process (updating) should be explained in terms of a simpler one (maintenance). No matter what distinct cognitive processes are implicated in updating (Ecker et al., 2010), the individual performance in updating tasks seems to be mostly limited by the number of items or item bindings which the participants are able to maintain in active memory. This result extends the data provided by Ecker et al. (2010) and by Schmiedek et al. (2009), by showing that updating is identical not only to complex span factor but also to simple spans. This result also allows us to rule out two possible reasons for strong correlation between complex spans and updating tasks considered by Schmiedek et al.,

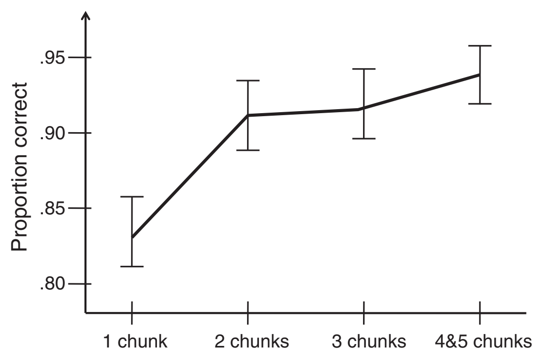

Fig. 6. The proportion of correct responses in the antisaccade task as a function of an average number of chunks that were supposedly held in the scope of attention (i.e., k estimate from the letter two-array comparison task) by participants in Studies 1 and 2. Error bars represent 95% confidence intervals.

namely that both tasks either load on executive control or require controlled search from long-term memory (LTM). The two-array comparison tasks put minimal demands on executive control, and transfer to LTM within these tasks was highly unlikely due to substantial time pressure, but the variable derived from these tasks was still identical to the updating variable. So, most probably neither control nor LTM operations determine scores in updating tasks.

4.2. Limitations and alternative explanations

Unfortunately, as a piece of psychometric research, the present study cannot tell precisely whether storage capacity depends mostly on the number of items which individuals were able to maintain (Cowan et al., 2007), or on the number of the item's bindings, supporting *ad hoc* mental structures, which can be flexibly constructed in WM and integrated into relations meaningful for the task at hand (Oberauer et al., 2007), or some interaction of both. We can only speculate about the high loadings of the two-array comparison tasks and the monitoring tasks on the common latent variable.

The discovery of relations required by the monitoring tasks, which is also the goal in the intelligence tests, seems to be extremely simple in the former tasks. One can hardly name the process leading to this discovery "a reasoning process". Rather, three pieces of information, for example "item X is present in upper-left location", "…is present in central-left location", and "…is present in bottom-left location", are being integrated into one piece providing information that "locations upper-left, central-left, and bottom-left form a column containing an item X". Once such a representation is formed in WM, it is simple to detect the match of the current pattern. So, the integration of (such simple) relations may not be the only factor crucial for the scores in monitoring tasks. Also, the maintenance of some prospective item-location bindings (e.g., those that partially satisfy the relation), while scanning the stimuli pattern for the remaining item-location pairs, may be important. Similarly, though the two-array comparison tasks look as if they mostly required active maintenance of as many stimuli as possible, the representation attended to may not only consist of a collection of matrix elements, but it could also include bindings of these items to particular locations within a matrix. When a border appears, it may initiate the finding of the particular binding of an item to the location defined by the border, instead of recollecting all items. So, fulfilling both kinds of task may involve a mixture of storage mechanisms, which combine zooming attention out in order to apprehend the maximum number of items (Cowan et al., 2007) and binding them to other information (Oberauer et al., 2007). For that reason, scores in those tasks may share a lot of common variance. Whether WM capacity refers to either the storage of items or the construction of their bindings (or both) may not be easy to disentangle with existing WM tasks.

Another limitation of our study concerns the fact that it relied on the two relatively similar and strongly intercorrelated tests of reasoning on figural material. This approach runs the risk that the construct of fluid intelligence operationalized in such a way may lack criterion validity. Numerous studies targeted the measurement of more broadly defined intelligence (e.g., Buehner et al., 2006; Colom et al., 2004, 2005; Oberauer et al., 2008; Süß et al., 2002). However, using a broad construct may lack theoretical meaning, which is especially evident in the case of the *g* factor, which has not yet gained univocal theoretical interpretation. We are aware of the relatively narrow definition of intelligence in our study and we do not exclude the possibility that a different pattern of relations may exist between a broader ability construct and the variables investigated in this study. However, at the same time we believe that successful reduction of intelligence to precise cognitive mechanisms — the scientific program which only started — should begin with a very well defined "intelligent process". If the resulting explanation is theoretically sound, the chance that it can be generalized is quite high.

Some recent studies (Shelton, Elliott, Matthews, Hill, & Gouvier, 2010; Unsworth, 2010; Unsworth, Brewer, & Spillers, 2009; Unsworth & Engle, 2007) have shown that LTM ability predicts fluid reasoning above and beyond the contribution of the scope of attention. Most radically, Mogle, Lovett, Stawski, and Sliwinski (2008) suggested that so-called secondary memory processes (controlled search and retrieval from LTM) fully mediate the relationship between WMC and Gf. Although it is highly unlikely that secondary memory is the sole mechanism underlying Gf (see Unsworth, 2010), the effectiveness of search and retrieval of information that surpassed available attentional resources might contribute to some Gf variance. Therefore, how can the findings on LTM be reconciled with the results of the present study?

In our opinion, there are two properties of widely used secondary memory tests, which make their univocal interpretation as measures primarily and directly reflecting LTM abilities problematic. Usually, such tests require sequential retrieval of a substantial amount of information, for example words from a 20-item list, or several word-number pairs (e.g., Mogle et al., 2008). These tasks, besides the searching of and retrieval from LTM, may also involve a great deal of relational integration of the retrieved elements. For example, when recalling words, one needs to bind each element to a tag declaring whether that element has just been retrieved (so it will not be recalled again) or whether it is still waiting for retrieval. Similarly, recall of associated pairs may require reconstruction of respective pair bindings in STM before a response is given. So, LTM abilities and some requirements on the very limited STM capacity may be confounded in many existing tests of secondary memory. Also, because of the long duration of each trial in a standard LTM test (e.g., approx. 1 min including responding in Mogle et al., 2008), at least in comparison to much shorter trials in tasks measuring storage capacity (usually lasting a few seconds, as in the present study), there is enough time within one LTM test trial to run diverse cognitive processes beyond the sole access to LTM. For example, any restructuring of memorized material surely requires construction of new bindings between encoded elements. Individual scores in LTM tests may indeed explain some amount of Gf variance, but in the face of our results we do not think that LTM is a fundamental mechanism underlying fluid intelligence. Moreover, the question of what LTM tests really measure surely needs a dedicated research program.

Correlational studies like the present one can only provide a large-scale map of relations between various cognitive abilities. In particular, no firm conclusions about causal relationships can be made on their basis. Nevertheless, the present study points out that investigation of storage capacity (but not of executive control) may be the most promising direction for research on the cognitive mechanisms responsible for individual differences in fluid ability. More studies, including advanced behavioral experimentation, brain imaging, and computational modeling (see Lewandowsky & Heit, 2006), have to be done in future in order to understand why some people excel in demanding intellectual and working memory tasks, while others fail in both.

4.3. Conclusions

The present study has supported the theories of fluid intelligence assuming that the crucial cognitive mechanism underlying fluid ability lies in storage capacity, which enables people to actively maintain distinct chunks of information and flexibly construct task-relevant bindings among them. In the two large-scale psychometric studies reported in this paper, storage capacity accounted for on average 70% of variance in fluid intelligence. On the contrary, no support has been found for the theories looking for the mechanisms responsible for intelligence in the domain of effectiveness of executive control. The measures of control were either not significantly related to fluid ability, as in the case of interference resolution and response inhibition, or fully depended on individual storage capacity, as in the case of attention control.

5. Role of the funding source

This research was supported by grant N106 2155/33 given to E. Nęcka by Polish Ministry of Science and Higher Education. The sponsoring institution had no role in the design of the study.

Acknowledgements

Authors thank Anna Anzulewicz, Krzysztof Cipora, Dominika Czajak, and Karolina Czernecka for conducting the experiments, and Mike Timberlake for correcting the manuscript.

References

Ackerman, P. L., Beier, M. E., & Boyle, M. O. (2002). Individual differences in working memory within a nomological network of cognitive and perceptual speed abilities. *Journal of Experimental Psychology. General*, *131*, 567–589.

Ackerman, P. L., Beier, M. E., & Boyle, M. O. (2005). Working memory and intelligence: The same or different constructs? *Psychological Bulletin*, *131*, 30–60.

Band, G. P. H., Molen, M. W., & van der Logan, G. D. (2003). Horse-race model simulations of the stop-signal procedure. *Acta Psychologica*, *112*, 105–142.

Buehner, M., Krumm, S., & Pick, M. (2005). Reasoning = working memory ≠ attention. *Intelligence*, *33*, 251–272.

Buehner, M., Krumm, S., Ziegler, M., & Pluecken, T. (2006). Cognitive abilities and their interplay: Reasoning, crystallized intelligence, working memory components, and sustained attention. *Journal of Individual Differences*, *27*, 57–72.

Carpenter, P. A., Just, M. A., & Shell, P. (1990). What one intelligence test measures: A theoretical account of the processing in the Raven Progressive Matrices test. *Psychological Review*, *97*, 404–431.

Cho, S., Holyoak, K. J., & Cannon, T. D. (2007). Analogical reasoning in working memory: Resources shared among relational integration, interference resolution, and maintenance. *Memory and Cognition*, *35*, 1445–1455.

Chuderska, A. (2010). Executive control in analogical mapping: Two facets. In S. Ohlsson, & R. Catrambone (Eds.), *the Proceedings of the 32nd Annual Conference of the Cognitive Science Society* (pp. 2749–2754). Austin, TX: Cognitive Science Society.

Colflesh, G. J. H., & Conway, A. R. A. (2007). Individual differences in working memory capacity and divided attention in dichotic listening. *Psychonomic Bulletin & Review*, *14*, 699–703.

Colom, R., Abad, F. J., Quiroga, M. A., Shih, P. C., & Flores-Mendoza, C. (2008). Working memory and intelligence are highly related constructs but why? *Intelligence*, *36*, 584–606.

Colom, R., Abad, F. J., Rebollo, I., & Shih, P. C. (2005). Memory span and general intelligence: A latent variable approach. *Intelligence*, *33*, 623–642.

Colom, R., Rebollo, I., Palacios, A., Juan-Espinosa, M., & Kyllonen, P. C. (2004). Working memory is (almost) perfectly predicted by *g*. *Intelligence*, *32*, 277–296.

Conway, A. R. A., Cowan, N., Bunting, M. F., Therriault, D. J., & Minkoff, S. R. (2003). A latent variable analysis of working memory capacity, short-term memory capacity, processing speed, and general fluid intelligence. *Intelligence*, *30*, 163–183.

Cook, R. D. (1977). Detection of influential observations in linear regression. *Technometrics*, *19*, 15–18.

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

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.

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: Oxford University Press.

Davelaar, E. J., Gosher-Gottstein, Y., Ashkenazi, A., Haarman, H. J., & Usher, M. (2005). The demise of short-term memory revisited: Empirical and computational investigations of recency effects. *Psychological Review*, *112*, 3–42.

Dempster, F. N., & Corkill, A. J. (1999). Individual differences in susceptibility to interference and general cognitive ability. *Acta Psychologica*, *101*, 395–416.

Duncan, J., Parr, A., Woolgar, A., Thompson, R., Bright, P., Cox, S., Bishop, 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.

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.

Egner, T. (2008). Multiple conflict-driven mechanisms in the human brain. *Trends in Cognitive Sciences*, *12*, 374–380.

Eimer, M. (1993). ERPs elicited by Go and Nogo stimuli: Effects of attention and stimulus probability. *Biological Psychology*, *35*, 123–138.

Embretson, S. E. (1995). The role of working memory capacity and general control processes in intelligence. *Intelligence*, *20*, 169–189.

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*, *44*. (pp. 145–199)New York, NJ: 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.

Fox, L. A., Shor, R. E., & Steinman, R. J. (1971). Semantic gradients and interference in naming color, spatial direction, and numerosity. *Journal of Experimental Psychology*, *91*, 59–65.

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.

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.

Gray, J. R., Chabris, C. F., & Braver, T. S. (2003). Neural mechanisms of general fluid intelligence. *Nature Neuroscience*, *6*, 316–322.

Halford, G. S., Cowan, N., & Andrews, G. (2007). Separating cognitive capacity from knowledge: A new hypothesis. *Trends in Cognitive Sciences*, *11*, 236–241.

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

Hallett, P. E. (1978). Primary and secondary saccades to goals defined by instructions. *Vision Research*, *18*, 1279–1296.

Hasher, L., Lustig, C., & Zacks, R. T. (2007). Inhibitory mechanisms and the control of attention. In A. R. A. Conway, C. Jarrold, M. J. Kane, A. Miyake, & J. N. Towse (Eds.), *Variation in working memory* (pp. 227–249). New York, NY: Oxford University Press.

Hasher, L., & Zacks, R. T. (1988). Working memory, comprehension, and aging: A review and a new view. In G. H. Bower (Ed.), *The psychology of learning and motivation*, *Vol. 22*. (pp. 193–225)New York, NY: Academic 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.

Hentschel, U. (1973). Two new interference tests compared to the Stroop Color–Word Test. *Psychological Research Bulletin, Lund University*, *13*, 1–24.

Hommel, B. (1998). Event files: Evidence for automatic integration of stimulus-response episodes. *Visual Cognition*, *5*, 183–216.

Hummel, J. E., & Holyoak, K. J. (2003). A symbolic–connectionist theory of relational inference and generalization. *Psychological Review*, *110*, 220–264.

Gentner, D., & Toupin, C. (1986). Systematicity and surface similarity in the development of analogy. *Cognitive Science*, *10*, 277–300.

Johnson-Laird, P. N. (1999). Deductive reasoning. *Annual Review of Psychology*, *50*, 109–135.

Kane, M. J., Conway, A. R. A., Hambrick, D. Z., & Engle, R. W. (2007). Variation in working memory capacity as variation in executive attention and control. In A. R. A. Conway, C. Jarrold, M. J. Kane, A. Miyake, & J. N. Towse (Eds.), *Variation in working memory* (pp. 21–48). Oxford: Oxford University Press.

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.

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, & Boyle (2005). *Psychological Bulletin*, *131*, 66–71.

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.

Keye, D., Wilhelm, O., Oberauer, K., & van Ravenzwaaij, D. (2009). Individual differences in conflict-monitoring: Testing means and covariance hypothesis about the Simon and the Eriksen Flanker task. *Psychological Research*, *73*, 762–776.

Lewandowsky, S., & Heit, E. (2006). Some targets for memory models. *Journal of Memory and Language*, *55*, 441–446.

Logan, G. D., & Cowan, W. B. (1984). On the ability to inhibit thought and action: A theory of an act of control. *Psychological Review*, *91*, 295–327.

Larson, G. E., Merritt, C. R., & Williams, S. E. (1988). Information processing and intelligence: Some implications of task complexity. *Intelligence*, *1*, 131–147.

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, *390*, 279–281.

Lustig, C., May, C. P., & Hasher, L. (2001). Working memory span and the role of proactive interference. *Journal of Experimental Psychology. General*, *130*, 199–207.

MacLeod, C. M. (1991). Half a century of a research on the Stroop Effects: An integrative review. *Psychological Bulletin*, *109*, 163–203.

MacLeod, C. M. (2007). The concept of inhibition in cognition. In D. S. Gorfein, & C. M. MacLeod (Eds.), *Inhibition in cognition* (pp. 3–23). Washington, DC: American Psychological Association.

McVay, J. C., & Kane, M. J. (2009). Conducting the train of thought: Working memory capacity, goal neglect, and mind wandering in an executive-control task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*, 196–204.

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.

Miyake, A., Friedman, N. P., Rettinger, D. A., Shah, P., & Hegarty, M. (2001). How are visuospatial working memory, executive functioning, and spatial abilities related? A latent-variable analysis. *Journal of Experimental Psychology. General*, *130*, 621–640.

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.

Morey, R. D. (2011). A Bayesian hierarchical model for the measurement of working memory capacity. *Journal of Mathematical Psychology*, *55*, 8–24.

Nęcka, 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. 161–181). Washington, DC: APA.

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.

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

Oberauer, K., 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.

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: Oxford University Press.

Oberauer, K., Süß, H. -M., Wilhelm, O., & Wittman, W. W. (2008). Which working memory functions predict intelligence? *Intelligence*, *36*, 641–652.

Orzechowski, J., & Chuderski, A. *Test analogii obrazkowych [A figural analogies test]*. Unpublished manuscript. Jagiellonian University, Krakow.

Osaka, M., & Osaka, N. (2007). Neural bases of focussing attention in working memory: An fMRI study based on individual differences. In N. Osaka, R. H. Logie, & M. D'Esposito (Eds.), *The cognitive neuroscience of working memory* (pp. 99–118). Oxford: Oxford University Press.

Raven, J. C., Court, J. H., & Raven, J. (1983). *Manual for Raven's Progressive Matrices and vocabulary scales (Section 4: Advanced Progressive Matrices)*. London: H. K. Lewis.

Redick, T. S., Calvo, A., Gay, C. E., & Engle, R. W. (2011). Working memory capacity and go/no-go task performance: Selective effects of updating, maintenance, and inhibition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*, 308–324.

Richland, L. E., Morrison, R. G., & Holyoak, K. J. (2006). Children's development of analogical reasoning: Insights from scene analogy problems. *Journal of Experimental Child Psychology*, *94*, 249–273.

Salthouse, T. (2005). Relations between cognitive abilities and measures of executive functioning. *Neuropsychology*, *19*, 532–545.

Schweizer, K., Moosbrugger, H., & Goldhammer, F. (2005). The structure of the relationship between attention and intelligence. *Intelligence*, *33*, 589–611.

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.

Shelton, J. T., Elliott, E. M., Matthews, R. A., Hill, B. D., & Gouvier, Wm. D. (2010). The relationships of working memory, secondary memory, and general fluid intelligence: Working memory is special. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*, 813–820.

Shilling, V. M., Chetwynd, A., & Rabbitt, P. M. (2002). Individual inconsistency across measures of inhibition: An investigation of the construct validity of inhibition in older adults. *Neuropsychologia*, *40*, 605–619.

Smith, E. E., & Jonides, J. (1999). Storage and executive processes in the frontal lobe. *Science*, *283*, 1657–1661.

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.

Unsworth, N. (2010). On the division of working memory and long-term memory and their relation to intelligence: A latent variable approach. *Acta Psychologica*, *134*, 16–28.

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.

Unsworth, N., Redick, T. S., Lakey, C. E., & Young, D. L. (2010). Lapses in sustained attention and their relation to executive control and fluid abilities: An individual differences investigation. *Intelligence*, *38*, 111–122.

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.

Unsworth, N., & Spillers, G. J. (2010). Working memory capacity: Attention, memory, or both? A direct test of the dual-component model. *Journal of Memory and Language*, *62*, 392–406.

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.

Unsworth, N., Spillers, G. J., Brewer, G. A., & McMillan, B. (2011). Attention control and the antisaccade task: A response time distribution analysis. *Acta Psychologica*, *137*, 90–100.

Verbruggen, F., & Logan, G. D. (2008). Response inhibition in the stop-signal paradigm. *Trends in Cognitive Sciences*, *12*, 418–424.

Waltz, J. A., Knowlton, B. J., Holyoak, K. J., Boone, K. B., Mishkin, F. S., de Menezes Santos, M., Thomas, C. R., & Miller, B. R. (1999). A system for relational reasoning in human prefrontal cortex. *Psychological Science*, *10*, 119–125.

Wilson, S. P., & Kipp, K. (1998). The development of efficient inhibition: Evidence from directed-forgetting tasks. *Developmental Review*, *18*, 86–123.

Yntema, D. B. (1963). Keeping track of several things at once. *Human Factors*, *5*, 7–17.