Abstract

Researchers have proposed that solving complex reasoning problems, a key indicator of fluid intelligence, involves the same cognitive processes as solving working memory tasks. This proposal is supported by an overlap of the functional brain activations associated with the two types of tasks and by high correlations between interindividual differences in performance. We replicated these findings in 53 older participants but also showed that solving reasoning and working memory problems benefits from different configurations of the functional connectome and that this dissimilarity increases with a higher difficulty load. Specifically, superior performance in a typical working memory paradigm (n-back) was associated with up-regulation of modularity (increased between-network segregation), whereas performance in the reasoning task was associated with effective downregulation of modularity. We also showed that working memory training promotes task-invariant increases in modularity. Because superior reasoning performance is associated with downregulation of modular dynamics, training may thus have fostered an inefficient way of solving the reasoning tasks. This could help explain why working memory training does little to promote complex reasoning performance. The study concludes that complex reasoning abilities cannot be reduced to working memory and suggests the need to reconsider the feasibility of using working memory training interventions to attempt to achieve effects that transfer to broader cognition.

INTRODUCTION

Fluid intelligence is central to solving complex reasoning tasks, that is, to identifying common patterns and applying logic to extrapolate them to novel problems. The cognitive processes that determine performance in such tasks have been a matter of debate for decades (Little, Lewandowsky, & Craig, 2014; Chuderski, 2013; Colom, Abad, Quiroga, Shih, & Flores-Mendoza, 2008; Oberauer, Süß, Wilhelm, & Wittmann, 2012; Conway, Kane, & Engle, 2003). Task analyses have led some researchers to propose that reasoning ability depends on and is almost isomorphic with the capacity to maintain and manipulate information in working memory (Little et al., 2014; Carpenter, Just, & Shell, 1990). In keeping with this line of reasoning, psychologists have emphasized the demands that both working memory and reasoning tasks place on executive control (Wiley, Jarosz, Cushen, & Colflesh, 2011; Gray, Chabris, & Braver, 2003). The high correlation between interindividual differences in working memory and complex reasoning scores (Little et al., 2014; Chuderski, 2013) and observations that these tasks elicit similar functional brain activity in the frontoparietal control network (CN; Clark, Lawlor-Savage, & Goghiari, 2017; Hearne, Mattingley, & Cocchi, 2016; Krawczyk, 2012; Colom, Karama, Jung, & Haier, 2010; Lee et al., 2006; Gray et al., 2003; Duncan et al., 2000) are also consistent with this idea. Yet, there are reasons to believe that cognitive processes other than working memory ability are involved in complex reasoning. Among them are exploratory cognitive functions such as perspective shifting and aspects of creativity that may facilitate rule discovery and insight. The latter refers to the spontaneous realization that a difficult problem has a simple solution if perceived from a conceptually different or new perspective. Although these processes are much less studied, their importance in solving complex reasoning tasks has been demonstrated in psychometric studies (Necka, Zak, & Gruszka, 2016; Benedek, Jauk, Sommer, Arendasy, & Neubauer, 2014).

Functional brain imaging studies of large-scale brain dynamics allow us to disentangle some of the key principles underlying higher cognition. Shaped by evolution to enhance adaptability and efficient information processing, the brain appears to function as a critical system achieving optimal performance near phase transitions from order to disorder (Chialvo, 2010). This, in turn, requires a hierarchical modular structure balancing segregation and integration of its functional subdivisions (Hilgetag & Hutt, 2014; Moretti & Munoz, 2013; Wang & Zhou, 2012). Modularity has been used as a global
beneficial to rule dis-
covery and abstraction (Cao, Li, Li, & Li, 2016). In contrast,
(Friston, Breakspear, & Deco, 2012)
networks. This enhanced communication may in turn
communication of the CN with the DMN and other brain
mance may be linked to a low-modular configuration of
for complex reasoning tasks that may require a higher degree
of between-network integration. We employed large-scale
network analysis to investigate these hypotheses. The
study used fMRI during one reasoning and working mem-
ory task. The spatial reasoning task was developed to be
similar to the classic matrix reasoning tests, and a classical
figural updating task (n-back) was used to probe working
memory. Both tasks were performed at three different dif-
ficulties (Figure 1) in a sample of healthy older adults (n = 53; 
age = 65–75 years) before (i.e., at pretest) and after
(i.e., at posttest) working memory training (n = 27) and
active control training (perceptual matching; n = 26).

METHODS

Participants

For the present analysis, the sample consisted of 53 par-
participants who had brain scans of acceptable quality from
two fMRI scanning sessions, one before and one after
4 weeks of cognitive training. Imaging data quality criteria
were (i) no motion artifacts in the raw data noticeable to
the naked eye, (ii) max frame-wise displacement below
3 mm, and (iii) less than 10% of missing volumes after
the de-spiking procedure. The study was approved by
the regional ethical review board in Stockholm (DNR:
2014/2188-31/1) and was conducted in accordance with
the revised Declaration of Helsinki (2013). Informed con-
sent was obtained from all participants before enroll-
ment. Standardized physical and neuropsychological
examinations were also carried out at screening. The par-
ticipants were randomly assigned to one of the two alter-
ative training programs—(1) working memory (active; 
n = 28, age = 68.89 ± 3.13 years, education = 15.17 ± 
3.25 years, number of women = 14, attendance = 19.07 ±
0.94 session) and (2) perceptual matching (control; n = 
25, age = 69.48 ± 2.96 years, education = 15.02 ± 4 years,
number of women = 14, attendance = 19 ± 1.58 session)—
and were blind to the hypotheses about the two protocols.
The randomization was accomplished in an automated
manner using R programming language (R Core Team,
2015) for each study wave separately, with age, sex, and
fluid intelligence used as stratifies.

The main study targeted 123 participants, and imaging
data were acquired for a random subsample of 78 par-
ticipants, 25 of whom were excluded from the analyses
in keeping with the image quality criteria (see above;
see Nilsson, Lebedev, Rydström, & Lövdén, 2017, for details). As a part of the study, half of the participants received transcranial direct current stimulation, whereas the other half received sham stimulation. Because the main study found no effects of stimulation on any of the collected cognitive measures, we did not estimate corresponding effects of stimulation on the imaging data in the present set of analyses but accounted for stimulation as an additional nuisance regressor during modeling.

Cognitive Training

Between pretest and posttest assessments with cognitive tasks and imaging, a JAVA-based computerized cognitive training battery was administered over 20 sessions (4 weeks, weekends off), each lasting 40 min. In the active program, four tasks trained two crucial facets of working memory: updating (running span and n-back tasks) and switching (rule- and task-switching tasks), each with four randomly varying stimuli sets to minimize the use of stimuli-specific strategies. The control intervention was focused on perceptual matching speed, employing four versions of the same task. In the task, participants were asked to assess the similarity of the presented figures and shapes, responding as quickly and accurately as possible. The training batteries were designed to be adaptive, such that the difficulty of the tasks increased systematically with improvements in performance so that the participants always trained at the highest level reached. The batteries were designed and carefully pilot-tested to balance the length and scope of the training, and the working memory training (active) and perceptual matching training (control) groups did not differ in the average level reached by the participants or in overall attendance. The cognitive training took place the week after an extensive 5-day assessment of cognitive functions. See Nilsson et al. (2017) for details and a comparative description of the two batteries.

MRI

Structural and fMRI scans were collected with the GE Discovery MR-750 3.0-T scanner with the 32-channel research head coil located at the Karolinska University Hospital, MRI Center in Solna, Sweden. Structural images were acquired employing a standardized contrast-enhanced T1 spoiled gradient BRAVO sequence with 0.94-mm³ isotropic voxels, a field of view of 256 mm (256 × 256 matrix), repetition time/inversion time/echo time = 5,688/450/2.492 msec, and flip angle = 12°. fMRI data were collected using echo-planar T2* imaging to measure the BOLD signal (voxel size = 3 mm³, echo time/repetition time = 30/2000 msec, flip angle = 80°).

fMRI Tasks

Two tasks were administered inside the magnetic resonance scanner to evaluate connectivity patterns during
working memory and reasoning. A figural n-back task assessed working memory performance. Three 1-min blocks were administered: 1-, 2-, and 3-back (each repeated three times) with a 4-sec fixation screen indicating the rule for the forthcoming level (9.54 min, in total). See Figure 1 for an illustration. For this task, accuracies were calculated as the average value of true positives and true negatives with a chance level of 50%. The second task was developed to evaluate the functional state involved in complex visuospatial reasoning. Participants were instructed to complete the presented pattern by selecting one of the four solutions shown at the bottom of the screen. One of the four answers was always correct. Like the n-back task, this paradigm had three levels of difficulty: A, perceptual matching; B, simple analytical (single rule); and C, difficult analytical (with at least two nontrivial rules), similar to the implementation proposed by Yamada et al. (2012). Each block was repeated three times in a pseudorandom order during the scanning session (9 task blocks + 9 fixation screens, in total). The reasoning task was self-paced to better match typical fluid intelligence paradigms (such as Raven’s progressive matrices) with fixed (1-min) blocks (9.54 min of total acquisition time including fixation blocks). The participants were explicitly instructed not to guess. For the reasoning paradigm, accuracy was defined as the fraction of correctly solved problems relative to the total number of attempts per block. In a number of simulations with randomly varying answers and total number of responses, chance level accuracy was estimated at 25%.

In both tasks, the participants demonstrated consistent accuracy drop as the difficulty level increased. In all the analyses, performance was measured as total task accuracy, which was defined as the sum of block-specific accuracies.

Both tasks were administered twice—before and after the cognitive training—and were programmed and presented using PsychoPy software, Version 1.84 (Peirce, 2007).

Cognitive Assessment at Pretest and Posttest

We evaluated the effects of working memory training via an extensive testing battery completed over four sessions, each lasting about 180 min. Preintervention testing took place 2 weeks before the training started, and posttesting was completed the week after the intervention. The battery was identical on both occasions and included a number of diverse cognitive tests measuring switching (task- and rule-switching), updating (n-back, running span), attention (temporal expectancy, perceptual matching), episodic memory (verbal and spatial recall), spatial (WASI, Raven’s, BETA matrices), and verbal (Syllogisms, BIS Analogies, ETS Kit inference) reasoning. See Nilsson et al. (2017) for a detailed description of the cognitive assessment.

MRI Processing

A standardized SPM12-based (Wellcome Trust Center for Neuroimaging, UCL, London) pipeline was implemented in the MATLAB R2013a environment (Anonymous, 2013).

Each participant’s T2* imaging data underwent subsequent steps for slice-timing correction, spatial realignment, and registration to standardized Montreal Neurological Institute space using a population-specific template produced from the set of T1 images with the Diffeomorphic Anatomical Registration Through Exponentiated Lie Algebra algorithm (Ashburner, 2007).

An extended 24-parameter model (Friston, Williams, Howard, Frackowiak, & Turner, 1996) was employed for head motion correction. The images were also smoothed (FWHM = 6 mm), de-trended, and, for the connectivity analyses, also band-pass filtered (0.008–0.12 Hz). The resulting output was used to calculate connectivity matrices (one matrix per each task and load level) based on temporal (Pearson r) correlation between the regions parcellated in accordance with the Craddock’s 200-ROI scheme (Craddock, James, Holtzheimer, Hu, & Mayberg, 2012). We did not employ global signal regression because of high risks of removing task-relevant signal variance (see Supplement S3 from the Supplement available at http://dx.doi.org/10.1101/202630).

No between-task differences were found in mean frame-wise displacement (meanFD_{n-back} = 0.09 ± 0.04, meanFD_{reasoning} = 0.11 ± 0.09), and neither of the cognitive measures significantly correlated with frame-wise displacement. Nevertheless, we repeated our analyses with scrubbing, which had a very minor impact on the results and did not change any inferences.

After removing negative weights, the adjacency matrices were then used to estimate whole-brain network modularity as implemented in the Brain Connectivity Toolbox (Rubinov & Sporns, 2010). The Louvain community detection method generalized for weighted graphs was used in the study as a measure of between-network segregation. The method attempts to solve an optimization problem by maximizing modularity (Q) given the observed graph weights:

\[
Q = \frac{1}{2m} \sum_{ij} \left[ A_{ij} - \frac{k_i k_j}{2m} \right] \delta(c_i, c_j),
\]

where:

- \( A_{ij} \) = weight (connectivity strength) between Nodes i and j
- \( k_i, k_j \) = sum of the weights attached to Nodes i and j
- \( m \) = sum of all graph weights
- \( c_i, c_j \) = communities of the nodes
- \( \delta \) = Kronecker delta

When there are few edges between the communities and within-cluster density is high, Louvain modularity is close to 1, but when the number of within-modular connections is comparable with the one from a random
graph, modularity usually approaches 0. Practically, it measures level of modular segregation in a graph.

**Statistical Analyses**

After preprocessing and cleaning, behavioral and imaging measures underwent outlier detection and checks for distribution normality. Between-group comparisons of demographic data used two-sample *t*-tests for continuous variables and binomial tests for sex ratios.

Next, we conducted a number of analyses to evaluate the reliability and validity of the collected measures. Pre- and post-Pearson correlation coefficients were calculated for the behavioral data to evaluate reliability. To address the problem of task validity, we analyzed updating and spatial reasoning domains in latent space, employing structural equation modeling (SEM) of the two constructs with offline measures and data collected inside the MRI scanner (see Figure 2). Each of the offline latent variables (LVs) was modeled with three tasks (updating: *n*-back and two running span tasks; spatial reasoning: Raven’s progressive matrices, matrix reasoning tasks from the Wechsler Abbreviated Scale of Intelligence, and BETA). Online (fMRI) LVs were formed with two indicators each (updating: 2- and 3-back; spatial reasoning: “simple analytical” and “difficult analytical”; 1-back and reasoning “A” accuracies were excluded from the SEM because of ceiling effects). The variance of the LVs was constrained to 1 to make path coefficients equivalent to Pearson correlations. SEM was also repeated for block-specific measures of in-scanner performance (2-back and “simple analytical”; 3-back and “difficult analytical”).

The next block was focused on evaluating task-related BOLD activity and modularity at baseline. Initially, we estimated first-level contrasts in both tasks: (a) 2-back – 1-back, (b) 3-back – 1-back, (c) reasoning “B” – reasoning “A,” and (d) reasoning “C” – reasoning “A.” Then, we performed a second-level analysis of the activations, which yielded block-specific statistical maps. Once the smoothness of the data had been estimated, adjustment for multiple tests was carried out using Monte Carlo simulations (*n* = 5000) as implemented in the AlphaSim algorithm (Ward, 2000), with an initial cluster-forming threshold of *p* < 0.005. In addition, individual block-specific activation patterns were matched with a meta-analytical map derived using Neurosynth software [www.neurosynth.org/; Yarkoni, Poldrack, Nichols, Van Essen, & Wager, 2011]. This procedure employs voxelwise Pearson correlation as a measure of similarity (Smith et al., 2009) between the statistical map and the meta-analytical image generated for a specific feature of interest. In the present analysis, the search was completed for the terms “working memory” (*n* = 901 studies) and “reasoning” (*n* = 147). The search was also completed for “executive control” (*n* = 157), which was used as a third independent term to evaluate validity of the activation patterns.

Next, we embarked on the analyses of large-scale connectivity, evaluating the differences in functional connectivity between the tasks and between-task correlation of modularity across different loads. Subsequently, we conducted a number of mass univariate tests of raw functional connectivity to identify which of the networks and between-network interactions were involved in the reported global effects. This was accomplished by fitting a linear model that evaluated correlations between block-specific functional connectivity and total task performance.

The last analytical block addressed effects of working memory training on performance and large-scale functional connectivity. We started by evaluating the main effects of time on task performance and then estimated Load × Group × Time effects in each of the two tasks.

**Figure 2.** Results of the SEM analysis. Circles represent modeled LVs; rectangles, measured variables (indicators); omnidirectional arrows, dependence; and bidirectional arrows, correlation. SR = spatial reasoning; UPD = updating; SR-B and SR-C = spatial reasoning (fMRI) blocks “B” and “C.” Numbers next to the bidirectional arrows represent standardized covariances (i.e., correlations) between the latent factors at pretest (black) and posttest (gray). CFI = comparative fit index; SRMR = standardized root mean residual; WASI = matrix reasoning tasks from the Wechsler Abbreviated Scale of Intelligence.
using linear mixed-effects models (random effect: “participants”). Standardized effect sizes were also calculated and reported as the mean difference in performance divided by the pretest standard deviation.

Modularity was analyzed as overall pre/post change as well as Time × Group interactions. At this stage, a full-interaction model was fitted to evaluate the effects of working memory training in two tasks (modularity ~ Load × Task × Time × Group × WM performance, random effect: “participants”). Finally, a subsequent analysis of raw functional connectivity was conducted to localize the reported global effects by estimating Group × Time effects on connectivity for each block.

Statistical Software and Toolboxes
All statistical analyses of behavioral data and modularity were performed with the R programming language, Version 3.2.2 (R Development Core Team, 2015), and “nlme” toolbox (Pinheiro, Bates, DebRoy, Sarkar, & R Core Team, 2016). In all models, age, sex, and the first three principal components from the motion parameters were introduced as nuisance covariates.

Mass univariate analyses of BOLD activation patterns were carried out in SPM12 (Wellcome Trust Center for Neuroimaging, UCL).

The Network-Based Statistics Toolbox (Zalesky, Fornito, & Bullmore, 2010) was used to analyze raw functional connectivity.

SEM was conducted using Ωnyx software (von Oertzen, Brandmaier, & Tsang, 2015).

Data Availability
Raw MRI and behavioral data can be requested from the authors and transferred for specific analysis projects that are in line with the original ethical approval. This requires a data use agreement, which effectively transfers the confidentiality obligations of the institution (Karolinska Institutet) at which the original research was conducted to the institution of the recipient of the data. An R script specifying main analytical and plotting steps is available at github.com/alex-lebedev.

RESULTS
Cognitive Measures
Initial analyses aimed to probe the reliability and validity of the cognitive tasks. Between-person differences in performance on the in-scanner tasks (Figure 1) were relatively stable from before to after cognitive training, with pre–post correlations of \( r = .57 (p < .001) \) for \( n \)-back and \( r = .6 (p < .001) \) for total reasoning performances averaged across the difficulty levels of the tasks. Corresponding values were equivalent between the groups: \( r_{\text{pre–post}} = .54/.47 (z = -0.44, p = .66) \) in the control/active groups for \( n \)-back and \( r_{\text{pre–post}} = .57/.65 (z = -0.44, p = .66) \) in the control/active for the reasoning task. These results indicate satisfactory lower bounds of reliability.

The validity of the in-scanner tasks was evaluated with SEM (see Methods). The \( n \)-back task performed in the scanner (UPD-fMRI) correlated strongly (pre: \( r = .65, p < .001 \); post: \( r = .79, p < .001 \)) with a latent factor of working memory updating (formed by tasks assessed outside the scanner; see Figure 2 and Methods for a more detailed description of the testing battery). Performance on the reasoning task administered in the scanner also correlated strongly (pre: \( r = .59, p < .001 \); post: \( r = .65, p < .001 \)) with a reasoning factor (formed by three tasks assessed outside the scanner). Cross-correlations were equivalent. Performance on the in-scanner \( n \)-back correlated in a statistically significant way with the reasoning factor (pre: \( r = .71, p < .001 \); post: \( r = .67, p < .001 \)); and performance on the in-scanner reasoning task, with the working memory updating factor (pre: \( r = .78, p < .001 \); post: \( r = .67, p < .001 \)). The reasoning and working memory updating factors were highly correlated (pre: \( r = .89, p < .001 \); post: \( r = .72, p < .001 \)) and so were the in-scanner tasks (pre: \( r = .52, p < .001 \); post: \( r = .45, p < .001 \)). Overall, the reliability as well as discriminant and convergent validity of the tasks used in the scanner were thus satisfactory.

Task-related BOLD Activity and Modularity at Baseline
Both tasks produced similar and reliable load-related activations in the control network (CN) (see Figure 3) that are typical for working memory and reasoning tasks. Moreover, similarity (voxel-wise Pearson \( r \); see Methods for details) of individual activation patterns to the meta-analytical map derived using the Neurosynth package (see Figure S1 from the Supplement available at

![Figure 3. Overlap of the task-related activation patterns. Overlap of the activation patterns in the \( n \)-back and spatial reasoning tasks (\( p_{\text{FWE}} < .05 \)). L1 = load 1 (1-back and reasoning A); L2 = load 2 (2-back and reasoning B); L3 = load 3 (3-back and reasoning C).](http://www.example.com/figure3.png)
executive control was statistically significant across all loads (significantly larger than 0 at \( p < .001 \)) and positively correlated with participants’ performance during \( n \)-back (2-back: \( r = .34, p = .01 \); 3-back: \( r = .43, p = .001 \)) and simple analytical reasoning tasks (reasoning B: \( r = .30, p = .03 \); reasoning C: \( r = .19, p = .18 \)). These results indicate that both paradigms reliably engaged the executive control circuit and that the magnitude of the engagement significantly correlated with successful task solving in all active conditions except for the difficult analytical block (“C”) of the reasoning task, in which engagement of the circuit had a limited nonsignificant effect on performance.

Despite similarities in the activation patterns, we observed a statistically significant effect of load, \( t(208) = -2.7, p < .01 \), and Load × Task effect, \( t(208) = 2.74, p = .007 \), on modularity (\( Q \)). These effects were further qualified by a Load × Task × Performance effect, \( t(208) = 3.07, p = .002 \). Although performance was implemented as a continuous variable in these analyses, to visualize these effects in an accessible manner, we performed a median split on performance (Figure 4). The resulting graph shows that individuals who performed well on the \( n \)-back task had a tendency to increase modularity (i.e., between-network segregation) with load, whereas individuals with better performance in the more complex reasoning task downregulated modularity with task complexity. In other words, superior performance in the difficult working memory task (\( n \)-back) was associated with higher modularity (between-network segregation), whereas superior performance in the complex reasoning task was associated with lower modularity (see also Figure 5A). To further assess these effects, we selected a subset of participants (\( n = 17 \)) who performed high on both tests.

Figure 4. Shifts in modularity associated with successful solving of working memory and reasoning tasks. Different trajectories of modularity (\( Q, y \) axis) change as a function of difficulty load (Loads 1–3, \( x \) axis) in good (thick lines) versus poor (thin dashed lines) performers (median splits, for illustration purposes); dashed arrows represent increasing difference between good and bad performers across the difficulty loads. No task = fixation blocks; Load 1 = 1-back and reasoning A; Load 2 = 2-back and reasoning B; Load 3 = 3-back and reasoning C.

Figure 5. Different associations between large-scale functional connectome and performance in \( n \)-back and reasoning tasks. The plot shows associations of participants’ performance with modularity (A) and raw (node pairwise) functional connectivity (B). In B, blue lines show significant negative associations, and red lines indicate positive associations between connectivity and performance at \( p_{\text{FDR}} < .05 \). The identified pattern suggests that better \( n \)-back performance is associated with increased segregation between the control network (CN) and default mode network (DMN) communities, whereas superior performance in the complex reasoning task is linked with enhanced connectivity of the CN with the sensory community and the DMN. Load 1 = 1-back and reasoning A; Load 2 = 2-back and reasoning B; Load 3 = 3-back and reasoning C.
defined as above-median performance both in \( n \)-back and reasoning. Load × Task effect was significant in this group, \( t(82) = 3.28, p = .001 \), whereas participants performing poorly on both tasks (\( n = 18 \)) did not exhibit such pattern, \( t(87) = 3.28, p = .001 \). Thus, individual differences in performance were, as predicted, related to different (segregated vs. integrated) configurations of the functional connectome in the working memory and complex reasoning tasks, especially at higher loads.

To identify which of the networks and between-network interactions drive these relations with global modularity, we conducted a number of subsequent tests of raw functional connectivity (Zalesky et al., 2010). The results presented a similar pattern of findings, which suggest that superior \( n \)-back performance is linked to increased segregation between the CN and the default mode network (DMN), whereas superior performance in the complex reasoning task is linked with more active cross-talk of the CN with the sensory community and the DMN (Figure 5B). Note that modularity during the fixation condition did not differ in a statistically significant way between the tasks, \( t(52) = 0.34, p = .736 \). Moreover, both tasks induced prominently less modular states than the fixation condition: \( n \)-back, \( t(52) = 6.2, p < .001 \); reasoning, \( t(52) = 7.0, p < .001 \) (Figure 4). Yet, all difficulty loads were consistently associated with modular states: Mean modularity was significantly (\( p < .001 \)) and consistently larger than the one derived from a set (\( n = 250 \)) of simulated random networks with equivalent properties (see Supplement S2 available at http://dx.doi.org/10.1101/202630), both as a total effect and when taken separately within specific tasks and loads.

**Effects of Working Memory Training on Performance and Modularity**

The main effect of time (pretest vs. posttest) on performance was statistically significant, which indicated general improvements in \( n \)-back and reasoning performances (0.57 SD for \( n \)-back, \( t(52) = 4.97, p < .001 \); 0.64 SD for the reasoning task, \( t(52) = 4.02, p < .001 \); Figure 6). Working memory training led to larger increases in \( n \)-back performance at higher loads than perceptual matching training did (Load × Group × Visit: \( t(155) = 2.77, p < .01 \)). Tests comparing gains from working memory training with gains from perceptual matching training at different loads indicated a statistically significant effect at 3-back only. There was a 0.88-SD improvement in the working memory training group and a 0.1-SD improvement in the control group, \( t(46.331) = -2.01, p = .026 \). It is worth noting, however, that this comparison would not survive strict \( p \) threshold correction for multiple testing (0.05/3 tests = 0.016).

We found that the active and control interventions affected modularity differently, with a statistically significant Group × Time interaction, \( t(316) = 3.5, p < .001 \).
(Figure 7). Taken separately, working memory training resulted in a significant increase in modularity, 0.31 SD, \( t(161) = 3.3, p = .001 \), whereas perceptual matching training did not result in statistically significant changes in modularity. Instead, it exhibited a trend toward reducing it, \(-0.2 \text{ SD}, t(155) = -1.8, p = .07\). The Group × Load × Time effect, however, was nonsignificant, which is consistent with a load-invariant effect of working memory training on modularity. The Group × Task × Time interaction was statistically significant, \( t(310) = 2.25, p = .025 \), indicating that effects of training on modularity were somewhat larger for the reasoning task.

Longitudinal analyses of raw functional connectivity revealed that working memory training reduced connectivity of the frontoparietal network mostly with the sensorimotor community and some components of the default mode/temporal cluster, with more marked effects seen in the reasoning paradigm (Group × Time t contrast: \( p_{\text{FDR}} < .05 \); Figure 8).

Given that the two tasks benefited from opposite modes of large-scale brain dynamics while training intervention resulted in load-invariant increases in modularity, we performed a set of additional analyses to explore whether individual differences in training-induced changes in working memory performance and large-scale functional organization were related to changes in reasoning performance. To do this, we first fit a simple model using behavioral data collected offline and observed that overall gains were negatively related between working memory (trained) and reasoning (transfer) composites \( (r = -0.28, \Delta \text{SR}_{\text{Composite}} \sim \Delta \text{WM}_{\text{Composite}}; t(51) = -2.01, p = .04) \), and this relationship was trend-significant after the main effect of Training Group had been removed, \( t(51) = -1.95, p = .06 \). For the in-scanner performances, however, a corresponding effect was nonsignificant, \( r = .17, t(51) = 1.3, p = .2 \).

Subsequently, exploring whether training-induced increases in modularity during reasoning task had negative relationships with changes in reasoning performance did not yield any significant results for any of the difficulty loads. However, we found that change–change correlations between modularity during n-back task and reasoning performance gradually went down across the three loads \( (r_{\text{Load1}} = .19, r_{\text{Load2}} = .07, r_{\text{Load3}} = -.15) \), but this was a trend that did not reach conventional levels of statistical significance, as shown by the Load × Modularity effect on performance, \( t(158) = 1.75, p = .08 \).

**DISCUSSION**

Employing methods of large-scale network analyses, we found that, despite high correlations in the behavioral data and substantial overlap of brain activation patterns, performance in complex reasoning and working memory tasks was associated with different modes of large-scale connectivity. Better performance in a typical working memory task (n-back) was associated with higher modularity (i.e., increased between-network segregation), whereas better performance in a reasoning task was associated with lower modularity (i.e., more interactions between the major brain networks) at a higher task.
complexity. Furthermore, working memory training modified functional connectome in both tasks in a similar way, that is, by increasing modularity. Perceptual matching training, on the other hand, did not have any significant impact on it. The observed effects appeared to be driven by integration within the frontoparietal CN and its segregation from the other communities.

It has been proposed that balanced modular organization of the brain is necessary for the emergence of the critical dynamics important for adaptive information processing and learning (Barbey, 2018; Cocchi, Gollo, Zalesky, & Breakspear, 2017; Hilgetag & Hutt, 2014; Moretti & Munoz, 2013; Wang & Zhou, 2012; de Arcangelis & Herrmann, 2010). Previous research has indicated that most cognitive tasks generally tend to perturb these modular dynamics by shifting them to a more integrated state of increased between-network communication (Cohen & D’Esposito, 2016; Liang et al., 2016; Vatansever et al., 2015; Dwyer et al., 2014; Stanley et al., 2014; Küźbichler et al., 2011). We observed a similar pattern of general modularity reductions from rest to task in this study. However, we also showed that performance on working memory and complex reasoning tasks is associated with modularity in different ways, such that between-network interplay benefits complex reasoning, whereas sustaining a more modular organization benefits working memory performance.

The observed divergent pattern of shifts in connectivity is generally in line with the literature, which suggests that automated tasks requiring outward attentional focus tend to favor segregation of the DMN and the frontoparietal CN (Piccoli et al., 2015; Jolles, van Buchem, Crone, & Rombouts, 2013; Hampson et al., 2010), whereas tasks requiring novel and creative problem solving benefit from enhanced between-network interplay ( Hearne et al., 2016; Beaty et al., 2014). It may be that lower modularity is associated with better complex reasoning performance because it relates to an exploratory mode of cognition (Senden et al., 2012) beneficial for rule discovery and abstraction (Cao et al., 2016). In contrast, such a mode of cognition may hinder working memory performance, which relies on a more modular network configuration with stabilized dynamics of the frontoparietal network coupled with effective suppression of DMN activity. These distinct associations were more marked at a higher task complexity. The findings fully support the network neuroscience theory of human intelligence suggesting that the tasks that engage broader cognitive functions benefit from dynamic shifts in connectivity patterns toward a more random reconfiguration, unlike the ones relying on specific cognitive abilities that rather benefit from a more regular mode of large-scale organization of the connectome (Barbey, 2018). This possibility is also in line with the recent findings that performance on simple tasks is positively related to resting state modularity, whereas performance on complex tasks is negatively related to resting state modularity (Yue et al., 2017).

It is worth noting that the circuit described in the influential parieto-frontal integration theory (P-FIT) of intelligence represents a broader community than the frontoparietal network discussed in this study and also includes components of the default mode and visual clusters (Jung & Haier, 2007). The reported results thus do not contradict P-FIT. Instead, our observation that superior reasoning performance favors enhanced cross-talk of the frontoparietal regions with occipital and medial pFCs is nicely in line with the P-FIT model.

Working memory training produced marked increases in global modularity, thus promoting the pattern of large-scale dynamics that was associated with better working memory performance before training. Interestingly, the same pattern of modularity changes was observed for the reasoning task, although no improvements in total reasoning performance were observed. This suggests that working memory training fosters a mode of solving the reasoning tasks that relies more on working memory and less on an exploratory mode of cognition. Such an interpretation is consistent with the trends for training-related improvements on the “simple analytical” (B) reasoning tasks, which may benefit from superior modularity (because the task demands attention, but the rules for solving it are simple to detect and abstract), but not on the “complex analytical” (C) tasks, which may benefit from less modular dynamics (because they require complex pattern discovery). It should be acknowledged that the ability to solve the complex analytical blocks is generally more representative of fluid intelligence. Therefore, despite trends in training-related differences in performance improvements on the simple analytical part of our reasoning task, our study does not support the thesis that it is possible to improve the spatial aspect of fluid intelligence with working memory training, at least in healthy older people. This conclusion is generally in line with the literature, which shows that working memory training has either no effects at all (Melby-Lervag, Redick, & Hulme, 2016) or, at best, small effects (Au et al., 2015) on reasoning. Our results expand these observations: They suggest that the reason working memory training does not improve fluid intelligence is because it shifts large-scale brain dynamics in ways that are not beneficial for solving novel and complex reasoning problems.

In summary, although interindividual differences in reasoning and working memory performance, and most other cognitive tasks too, are positively correlated, the brain may solve these tasks in partly different ways. Rather than indicating strong links between the processes involved in reasoning and working memory, the correlated interindividual difference could be produced by initially unrelated cognitive domains that support each other in learning and development over the life span (van der Maas et al., 2006). Indeed, it is clear that higher working memory capacity does not necessarily imply superior intelligence. A particularly illustrative example is the
observation that chimpanzees have, in fact, better visual working memory than humans (Inoue & Matsuzawa, 2007), which suggests that the cognitive advancement of the human species was unlikely driven by increased storage capacity. Despite the fact that our study did not find changes in modularity induced by working memory training to actually be detrimental to reasoning abilities, it generally supports the idea that storage capacity may, in fact, compete with some aspects of complex reasoning, such as insight and deep information processing capabilities (DeCaro, Van Stockum, & Wieth, 2016). Our results are also broadly in line with the recently proposed idea of a trade-off in biological systems, in which high-modular organization provides evolutionary benefits over shorter time scales and low-modular systems exhibit greater fitness over longer time scales (Yue et al., 2017; Chen & Deem, 2015). Notably, imposing time constraints on reasoning problems appears to increase the correlation between participants’ performance on such tasks with their performance on working memory tests (Chuderski, 2013). The trade-off hypothesis (Yue et al., 2017; Chen & Deem, 2015) and our results are in accordance with the observation that working memory training studies tend to show transfer effects when reasoning tasks are administered under more stringent time constraints (Jaeggi, Buschkuehl, Jonides, & Perrig, 2008; possibly benefitting from high-modular states), whereas those that administer the tasks in a traditional, time-relaxed manner (which benefit from low-modular dynamics) do not report such effects (Colom et al., 2013).

This study had a number of limitations. First, one could argue that the tasks we used inside the scanner do not properly represent corresponding abilities of working memory and fluid intelligence or, at least, that the tasks only tap into selected aspects of these abilities. For example, some tasks that are supposed to characterize the same domain of working memory (e.g., complex span and n-back) actually have relatively low correlation and may therefore represent different aspects of working memory (Redick & Lindsey, 2013). Meanwhile, some researchers have questioned the plausibility of this argument, as several studies demonstrate that the performances on the two tasks converge when working memory is modeled as a latent construct (Jaeggi, Buschkuehl, Jonides, & Perrig, 2008; Schmiedek, Hildebrandt, Jonides, & Perrig, 2010). This is still of crucial importance for this study, considering that both resting state and task-related modularity have been shown to be differently associated with performance on working memory tasks, depending on their complexity, with positive correlation found for visual working memory performance and a trend-level negative correlation observed for a complex span task (Alavash, Doebler, Holling, Thiel, & Giessing, 2015).

Another potential limitation of the study is the absence of counterbalancing of the fMRI tasks, which may have affected the reliability of the results because of carry-over effects of the n-back paradigm on reasoning. However, consistent engagement of the control circuit in both tasks and strong associations between modularity and performance, together with the occurrence of longitudinal changes in the expected direction, strongly support plausibility of the proposed interpretations. It is also worth noting that, although we tried to minimize the difference between the tasks in aspects that are not central to the research question, the stimuli still differed in a number of ways, besides their reasoning and working memory demands, which in turn might confound some of our results. Finally, this study was conducted on a sample of older adults, and the results therefore may not be generalizable to younger populations, especially because community structure of the brain undergoes substantial reconfiguration over different periods of the life span (Geerligs et al., 2015; Onoda & Yamaguchi, 2013). Decreased signal-to-noise ratio of brain activity found in older participants (D’Esposito, Deouell, & Gazzaley, 2003) may foster a different mechanism underlying rest-to-task perturbation of functional connectome in this population, perhaps imposing higher demands on the aging brain to upregulate segregation of the frontoparietal network to allow superior filtering of the neural noise, which in turn appears to have detrimental effects on working memory capacity (Bays, 2015), while facilitating exploratory states of brain dynamics (Ghosh, Rho, McIntosh, Kotter, & Jirsa, 2008). Further studies may want to focus on how differences in the amount of neural noise influence reconfiguration of the functional connectome depending on task complexity. This, in turn, may help to explain a discrepancy of findings from imaging studies of working memory, some of which show positive associations between modularity and performance (Stevens et al., 2012), whereas others report negative relationships (Cohen & D’Esposito, 2016).

Despite these issues, we conclude that reasoning and working memory performances in older adults have different associations with large-scale functional brain connectivity at a higher task complexity. We also found that working memory training fosters task-invariant increases in functional network modularity and does not improve performance on complex reasoning tasks. Instead, complex reasoning tasks benefit from a higher degree of between-network integration. Together, these results indicate that complex human reasoning abilities cannot be reduced to working memory and add weight to the conclusion that working memory training interventions will not result in improvements that transfer to broader cognition.

Acknowledgments
We thank Anders Rydström, Linda Lidborg, Jakob Norgren, Helena Franzén, and Simon Peyda for help with data collection and Marie Helsing for help with recruitment and organization. Thanks also to Alireza Safati, Goran Papenberg, and Lars Nyberg for comments on an earlier draft of this article. This
research has received funding from the European Research Council under the European Union’s Seventh Framework Programme (FP7/2007-2013)/ERC grant agreement #617280-REBOOT. Martin Lövdén was also supported by a “distinguished younger researcher” grant from the Swedish Research Council (446-2013-7189).

Reprint requests should be sent to Alexander V. Lebedev, Department of Neurobiology, Care Sciences and Society, Aging Research Center (Karolinska Institutet & Stockholm University), Gävlegatan 16, 113 30 Stockholm, Sweden, or via e-mail: alexander.lebedev@ki.se.

REFERENCES


Chuderski, A. (2013). When are fluid intelligence and working memory isomorphic and when are they not? *Intelligence*, 41, 244–262.


Little, D. R., Lewandowsky, S., & Craig, S. (2014). Working memory capacity and fluid abilities: The more difficult the item, the more more is better. Frontiers in Psychology, 5, 239.


