Psychological Research (2017) 81:901–909
DOI 10.1007/s00426-016-0796-8

ORIGINAL ARTICLE

A drop in performance on a fluid intelligence test due to instructed-rule mindset

Hadas ErEl[1,2] · **Nachshon Meiran**[3]

Received: 17 April 2016 / Accepted: 10 August 2016 / Published online: 17 August 2016

Abstract A 'mindset' is a configuration of processing resources that are made available for the task at hand as well as their suitable tuning for carrying it out. Of special interest, *remote-relation* abstract mindsets are introduced by activities sharing only general control processes with the task. To test the effect of a *remote-relation* mindset on performance on a Fluid Intelligence test (Raven's Advanced Progressive Matrices, RAPM), we induced a mindset associated with little usage of executive processing by requiring participants to execute a well-defined classification rule 12 times, a manipulation known from previous work to drastically impair rule-generation performance and associated cognitive processes. In Experiment 1, this manipulation led to a drop in RAPM performance equivalent to 10.1 IQ points. No drop was observed in a General Knowledge task. In Experiment 2, a similar drop in RAPM performance was observed (equivalent to 7.9 and 9.2 IQ points) regardless if participants were pre-informed about the upcoming RAPM test. These results indicate strong (most likely, transient) adverse effects of a *remote-relation* mindset on test performance. They imply that although the trait of Fluid Intelligence has probably not changed, mindsets can severely distort estimates of this trait.

✉ Hadas ErEl
hadaserel@gmail.com

1 Baruch Ivcher School of Psychology, Interdisciplinary Center (IDC), Herzliya 46150, Israel

2 School of Behavioral Science, Netanya Academic College, Netanya, Israel

3 Department of Psychology and Zlotowski Center for Neuroscience, Ben-Gurion University of the Negev, Beer-Sheva, Israel

Introduction

A 'mindset' may be described as "a configuration of processing resources that are made available for the task at hand as well as their suitable tuning for carrying it out" (ErEl & Meiran, 2011). Mindsets range from concrete (i.e., related to specific stimuli and responses or task sets, Grange & Houghton, 2014) to abstract (Galinsky & Kray, 2004; Gollwitzer, 1990; Kounios et al., 2006). The later type of mindsets is of special importance due to its pervasive nature (e.g., Gollwitzer, 1990).

Abstract mindsets may be subdivided according to their proximity to the task to be performed. This proximity is defined according to the degree of content overlap between the mindset inducing activity and the task to be performed. *Close-relation* mindsets influence performance on tasks that share specific features with the task used for the mindset activation, such as stimuli (Ginossar & Trope, 1987) and procedure (Luchins, 1942; Verguts & de Boeck, 2002). *Medium-relation* mindsets influence performance despite stimulus and procedure change between the inducing activity and the task as long as there is a direct relation between the mindset and some aspect of the tasks' content. An example for *medium-relation* mindsets is stereotype threat (Inzlicht & Schmader, 2011; Steele & Aronson, 1995), defined as the predicament arising from well-known negative stereotypes that can be experienced as self-evaluation threat (Steele, 1998; Steele & Aronson, 1995) leading to performance drop (Aronson, Fried, & Good, 2002; Nisbett et al., 2012b; but see Nguyen & Ryan, 2008). Notably, stereotype-related mindsets are not general, but are directly related to the task to be performed. For example, telling participants that men outperform women in math would lead to poorer math performance among women (Schmader, 2002). *Remote-relation* mindsets are characterized by a minimal overlap between the mindset induction and the task. Namely, they can be activated by conditions that appear only remotely related to the task to be performed. *Remote-relation* mindsets influence every task that involves the general control processes activated in the mindset, regardless of task content. Tasks involving the same control processes as those manipulated by the mindset may be drastically impacted by them (ErEl & Meiran, 2011; Freitas, Gollwitzer, & Trope, 2004; Markman, Lindberg, Kray, & Galinsky, 2007; Smith, Jostmann, Galinsky, & van Dijk, 2008). Examples for *remote-relation* mindsets are those invoked by counterfactual thinking (Galinsky & Kray, 2004), deliberative vs. implemental tasks (Fujita, Gollwitzer, & Oettingen, 2007; Gollwitzer, 1990, 2012), high and low construal levels (Förster, Friedman, & Liberman, 2004; Fujita, Trope, Liberman, & Levin-Sagi, 2006), divergent vs. convergent thinking (Fischer & Hommel, 2012), and instructed rules (ErEl & Meiran, 2011).

Mindsets have shown to affect problem solving and creativity (Fujita et al., 2006; Ginossar & Trope, 1987), rule finding, and executive functions (Aronson & McGlone, 2008; ErEl & Meiran, 2011). One common characteristic of these examples is their relation to Fluid Intelligence, representing reasoning and the ability to deal with novel, complex problems (Nisbett et al., 2012a). Specifically, rule finding and problem solving tests tend to show the highest correlations with general intelligence (Marshalek, Lohman, & Snow, 1983) and some executive functions (e.g., working-memory updating) show very high correlations with intelligence (e.g., Friedman et al., 2006). This relation implies the possibility of mindset influence on performance in tests used to measure Fluid Intelligence. This hypothesis gains some support from findings indicating that beliefs regarding intelligence (Aronson et al., 2002), stereotype threat (Aronson & McGlone, 2008; Brown & Day, 2006), and motivation (Duckworth et al, 2011) influence intelligence test performance. Yet, all these prior findings may be considered as reflecting *medium-relation* mindsets. An example is a mindset involving stereotypes regarding specific groups' intellectual ability (Aronson & McGlone, 2008) which involve a direct relation between the content of the stereotype used to produce the mindset and the intelligence test.

In this study, we note that the possibility that *Remote-relation* mindsets can potentially influence performance on a test of Fluid Intelligence if they share control processes with the Fluid Intelligence test (Heitz, Unsworth, & Engle, 2005). Specifically, successful performance on the Raven's Advanced Progressive Matrices (RAPM) (Raven & Court, 1998), among the best measures of Fluid Intelligence, has been associated with control processes, including rule generation, attention-shifting, and identifying the correct item (Carpenter, Just, & Shell, 1990; Conway, Cowan, Bunting, Therriault, & Minkoff, 2002; Rasmussen & Eliasmith, 2011). Critically, similar processes have been shown to be influenced by the mindset created by instructing participants to follow a simple classification rule for several trials, henceforth "instructed-rule mindset" (ErEl & Meiran, 2011, Experiments 4–6).

We thus hypothesized that this mindset would impair RAPM performance. This prediction is supported by ErEl and Meiran (2011) who explored an instructed-rule mindset impacts on rule generation. In that study, participants executed a concrete, classification task in which we provided the (simple) task rule in advance followed by a rule-generation task (Ashby, Alfonso-Reese, Turken, & Waldron, 1998), where participants needed to find the classification rule based on feedback. The results indicated that the number of trials required to find the rule approximately doubled when the rule-generation task was executed immediately after activating the instructed-rule mindset as compared to three control conditions. These included: Baseline (participants went directly to rule finding and did not perform an instructed-rule beforehand); just-instructed (participants received instructions for a rule, but did not execute the instructions); and prior rule finding (participants performed a rule-finding test with a different set of rules beforehand). This performance drop was seen even when the mindset induction and the rule generation involved completely different stimuli. In an effort to gain insight concerning the underlying processes, we tested the three functions that are involved in rule finding according to the COVIS model (Ashby et al., 1998). We showed that the mindset impaired all three functions, including hypothesis-generation (tested with a hypothesis fluency test), attention switching between rules (tested with a task-switching task), and feedback evaluation (tested using post-error slowing). These processes are needed, because rule-generation tasks require participants to generate hypotheses concerning what the rule might be, to process feedback and adjust performance when a hypothesis turns out to be wrong, and to switch from the current hypothesis to a new hypothesis. Importantly, since feedback evaluation is temporarily impaired participants take long to reactivate control processing when faced with a control-demanding task. This happens precisely, because they have missed the new signals from the environment that executive control processes are now required. Given that the aforementioned processes are also associated with performance on the RAPM (Carpenter et al., 1990; Rasmussen & Eliasmith, 2011; Verguts, De Boeck, & Maris, 1999), we predicted that RAPM performance would be impaired by the instructed-rule mindset.

A possible explanation for the drastic influences of this mindset is derived from neuroimaging studies implicating the prefrontal cortex in rule-finding tasks but not in rule application tasks (Crescentini et al., 2011). Furthermore, recent works suggest that people avoid as much as possible using prefrontal cortex resources (Kool & Botvinick, 2014; Shenhav, Botvinick, & Cohen, 2013). Accordingly, prefrontal resources are recruited only in the presence of relevant incentives (Jimura, Locke, & Braver, 2010), and are probably turned off, as when the environment signals that the demand for cognitive control resources is minimal (as in the case of the request to follow a simple instructed rule). Given that one of the impaired processes involves monitoring and feedback evaluation, participants are expected to take a non-trivial amount of time (and accumulated evidence that control processing is needed) to recover from the mindset. As a result, performance on subsequent tasks is impaired if they demand considerable cognitive control resources.

To summarize, *remote-relation* mindsets influence performance on tasks that do not bare resemblance (i.e., content overlap) to the tasks used to create the mindsets. Instructed-rule mindset is created by letting participants follow a simple classification rule, and this *remote-relation* mindset has shown to impair cognitive processes that are arguably involved in tests of Fluid Intelligence. We, therefore, predicted that instructed-rule mindset would (temporarily) impair performance on RAPM, a Fluid Intelligence test. We report two similar experiments demonstrating a drop in RAPM performance following a brief execution of a simple instructed rule.

Experiments

In Experiment 1, we measured RAPM performance and General Knowledge. We relied on Cattell's (1987) distinction between "Fluid Intelligence" and "Crystalized Intelligence" (representing acquired skills and knowledge). We reasoned that the instructed-rule mindset would impair performance on a Fluid Intelligence task (RAPM), but not on a crystalized intelligence task (General Knowledge). This prediction is based on the differences in demand for control processes associated with these different types of intelligence (e.g., Cole, Yarkoni, Repovš, Anticevic, & Braver, 2012). The predicted dissociation contributes to ruling-out general motivation (which would affect both tasks) as a candidate underlying process.

Experiment 2 was designed to replicate the results from Experiment 1 as well as to test whether participants can avoid the instructed-rule mindset effect if pre-informed that they would be required to execute the RAPM test immediately after the classification task.

Method

Participants

Experiment 1 included 56 undergraduate students from Netanya Academic College from the School of Behavioral Sciences who participated for partial course credit or from Computer Sciences department who participated for a payment of 40 NIS (roughly 10 EURO) per two 20–30 min sessions ($M = 23.21$ years and $SD = 2.82$) that were randomly assigned to the four groups (14 participants per group). Experiment 2 included 84 undergraduate students from the Baruch Ivcher School of Psychology, at the Interdisciplinary Center Herzliya, who participated for partial course credit ($M = 22.40$ years and $SD = 3.10$). Sample size was determined a priori only in Experiment 2, so that the power to detect a medium effect size ($\rho = 0.30$) in a planned two-tailed contrast was just over 0.80. The participants were randomly assigned to three groups (28 participants per group).

Design

Experiment 1 included two dependent measures (RAPM/General Knowledge) tested between participants and three independent variables: manipulation (with or without a classification rule execution—between participants), counterbalancing (of the assignment of the parallel forms to pre/post testing, between participants), and pre vs. post (within-participants). Experiment 2 focused on the RAPM measure. In this experiment, manipulation included an additional level: classification rule execution with notification regarding the following RAPM test.

Apparatus and stimuli

Both experiments were based on ErEl and Meiran's (2011) manipulation involving executing a newly instructed (simple) classification rule which applied to letter-digit stimuli (e.g., "A8"), composed of a letter (A, B, Y, and Z) and a digit (1, 2, 8, and 9) and presented in the middle of the screen. Participants responded by pressing the left key ("S") or the right key ("K"). The rule was: "If the letter in the stimulus is from the beginning of the alphabet, press the left key/if it's from the end of the alphabet, press the right key". The intelligence test items were individually presented on the monitor with participants keying the number corresponding to their answer, and then confirming their choice.

We generated two parallel forms for the pre- and post-RAPM (Set II) by dividing the test into even and odd items (18 items each). The assignment of forms to pre–post measurement was counterbalanced between participants.

Crystalized Intelligence was assessed with a General Knowledge test made of 50 four-choice trivia questions taken from internet sources (http://www.netrivia.co.il, http://www.mytrivia.co.il, and http://www.games.nana10.co.il). This choice was made due to the lack of a validated Hebrew test. Nonetheless, the questions were in line with the major semantic domains of the long-term memories for different kinds of information and knowledge suggested by Irwing, Cammock, and Lynn (2001) that are commonly used for content validity assessment in general knowledge tests (e.g., Batey, Furnham, & Safiullina, 2010). Two parallel forms of 25 items each were assigned to pre–post measurement (counterbalanced between participants). The pre–post correlation computed in the baseline group ($r = 0.64$, $p < 0.05$, Experiment 1) indicated acceptable reliability of this test.

Procedure

Participants provided written informed consent and were assigned to conditions according to the order in which they entered the experiment (see Fig. 1). They were tested twice (a week break between sessions). In Experiment 1, Session 1 involved a pre-test as baseline. In Session 2, the experimental group performed 12 consecutive rule execution trials in which the rule was provided, and after a break of 2 min performed the other half of the intelligence tests (RAPM/knowledge). The baseline group performed only the other half of the intelligence tests. In Experiment 2, general knowledge was omitted and the manipulation variable included the "rule execution with notification" condition, in which participants were notified at the beginning of Session 2 that the rule application phase would be immediately followed by the second part of the RAPM.

Results Experiment 1

The dependent variable was the number of correctly answered items. The pre–post correlation of the RAPM measure, as computed in the baseline group ($r = 0.24$, ns) indicated rather poor reliability of the half-test in this sample. Mean pre-test score on the RAPM was 8.75 items (SD = 2.04 items), which, when doubled (given that we used half of the test), is equivalent to $M = 17.50$ and SD = 4.08 items. These values are somewhere below the norm estimated in a sample from a prestigious university (e.g., Bors & Stokes, 1998, who reported $M = 22.17$, SD = 5.60 items) and considerably above that seen in a community sample of young adults ($M = 15.95$, Denney & Heidrich, 1990).

In a preliminary analysis, we included the counterbalancing condition, and since its interaction with the critical effect was tiny and non-significant, $\eta_p^2 = 0.01$, $p = 0.61$, we dropped it from the main analysis.

Importantly, there was a drop in RAPM performance in the experimental group ($M = 0.71$, SD = 2.23) as compared with the baseline group which improved from pre-test to post-test ($M = 1.07$, SD = 2.09). This group difference in change scores was significant, $F(1,26) = 4.77$, MSE = 4.68, $p = 0.038$, and $\eta_p^2 = 0.15$, reflecting an interaction between manipulation and pre–post of 1.88 items, roughly equivalent to 0.68 normative standard deviations (taken from Bors & Stokes, 1998, and given that this was a half-test) or an equivalent of 10.1 IQ points. In the general knowledge test, both groups showed a similar trend for improvement from pre-test to post-test ($M = 0.50$, 0.64 items, SD = 3.63, 1.93 items in the experimental and baseline groups, respectively) [$F(1,24) < 1.00$, ns., $\eta_p^2 < 0.001$]. Note that the lack of significant interaction effect for the general knowledge test could not be due to lack of statistical power, since it has proven to be much more reliable than the RAPM in this sample. Similarly, the low reliability of the RAPM in this sample does not undermine our finding, since low reliability means that the effect (which reached significance) is actually an under-estimate of the true effect.

Results Experiment 2

The pre–post correlation of the RAPM measure as computed in the Baseline group ($r = 0.64$, $p < 0.050$) indicates acceptable reliability of the half-RAPM in this sample. The mean pre-test RAPM score was $M = 9.67$ (SD = 2.54), which is equivalent to $M = 19.34$ (SD = 5.08) for the full RAPM. As in Experiment 1, after preliminary analysis, the counterbalancing condition was excluded from further analysis due to tiny and non-significant interaction with the critical effect ($\eta_p^2 = 0.01$ and $p = 0.760$).

More importantly, such as in Experiment 1, the baseline group improved from pre-test to post-test ($M = 0.68$, SD = 1.96), while the experimental groups (without notification and with notification) showed a performance drop ($M = 1.04$, 0.79, SD = 2.38, 2.28, respectively). This group difference reached significance [$F(2,81) = 4.88$, MSE = 4.91, $p = 0.010$, $\eta_p^2 = 0.11$] replicating the interaction between manipulation and pre–post on RAPM. Planned contrasts using the pooled error term indicated that while the difference between the experimental groups was non-significant [$F(1, 81) < 0.20$ and $\eta_p^2 < 0.003$], the difference between each of these groups and the baseline group was significant [$F(1, 81) = 8.38$, 6.11, $p = 0.020$,

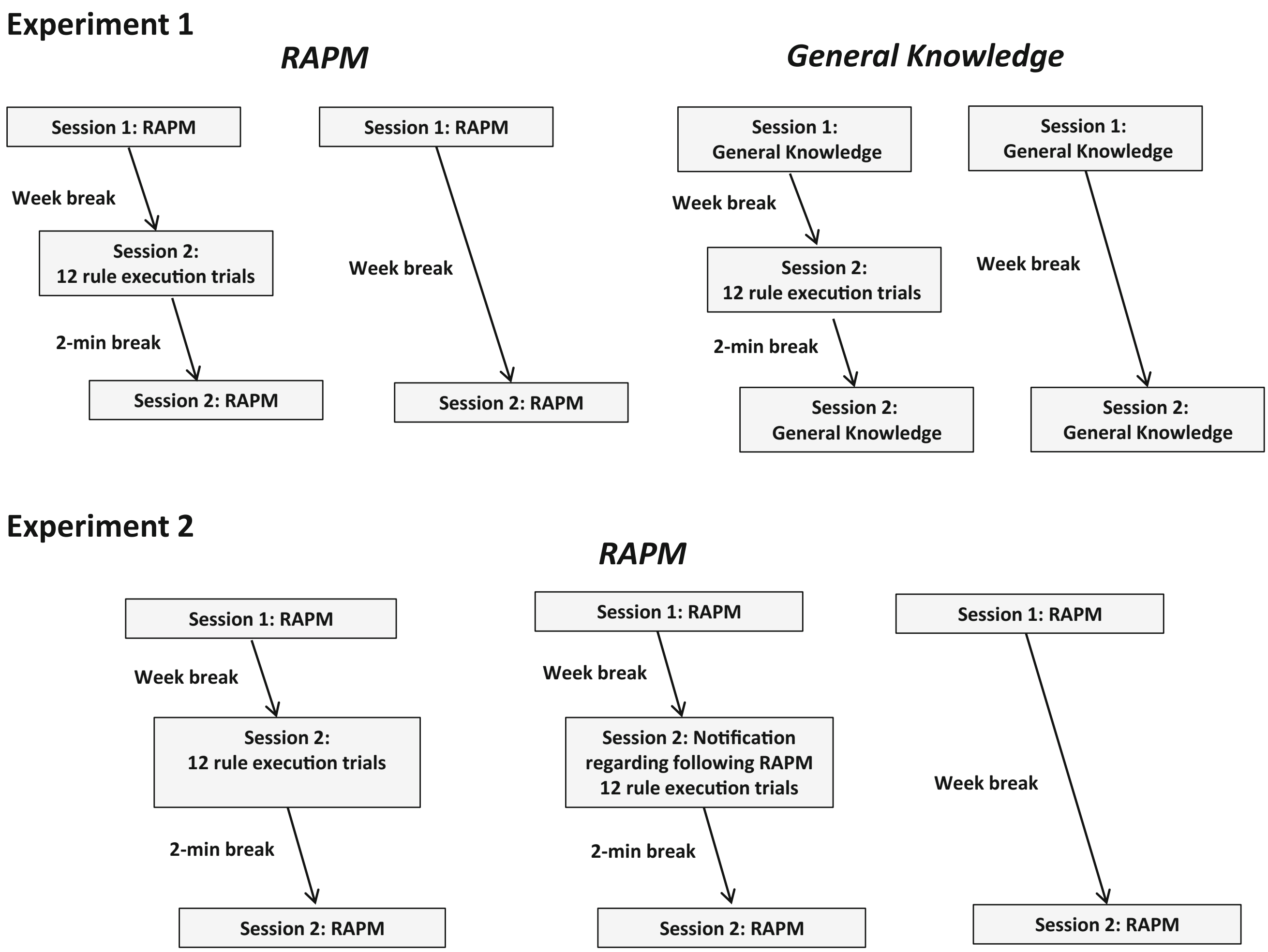

Fig. 1 Experimental procedure *RAPM* Raven's advanced progressive matrices

$\eta_p^2 = 0.09$ and 0.07]. This difference was 1.72 and 1.47, on average, equivalent to 0.62 and 0.52 normative SDs, or 9.2 and 7.9 IQ points.

Results Experiments 1 and 2

To quantify the strength of the evidence, we conducted a Bayesian analysis (using JASP, Love et al., 2015). The only model that reached Bayes factors larger than 3.00 (sometimes viewed as necessary for Bayesian evidence to be considered compelling, see Jeffreys, 1939/1961; Sprenger, 2013) when compared against the null hypothesis was the model, including both test (post-test–pre-test), group (baseline, rule application), and the interaction between them, $BF_{10} = 12.72$. This Bayes factor is considered as strong support for the model (Jeffreys, 1939/1961). Comparison between the model including the interaction to the model including both factors without the interaction yielded decisive support for the model including the interaction ($BF_{10} = 174.24$).

We additionally tested the significance of the performance drop ($M = 0.87$, $SD = 2.28$) in the experimental groups and the significance of the improvement in the baseline groups ($M = 0.80$, $SD = 1.99$) by comparing the effects to 0. Both effects were found to be significantly different from 0 ($t(69) = 3.19$, $p = 0.002$, $BF_{10} = 12.91$ for the drop in the experimental groups; $t(41) = 2.63$, $p = 0.012$, $BF_{10} = 3.75$, for the improvement in the baseline groups).

Time-course analysis

An interesting question raised by one of the reviewers concerns the time-course of the mindset effect. While the experiments in this study cannot address this question directly, some insight can be gained by examining the mindset effect as a function of the RAPM items' serial position. Clearly, simply testing the pre–post difference is not indicative, as the RAPM items are ordered by their difficulty. Hence, performance on a specific serial position

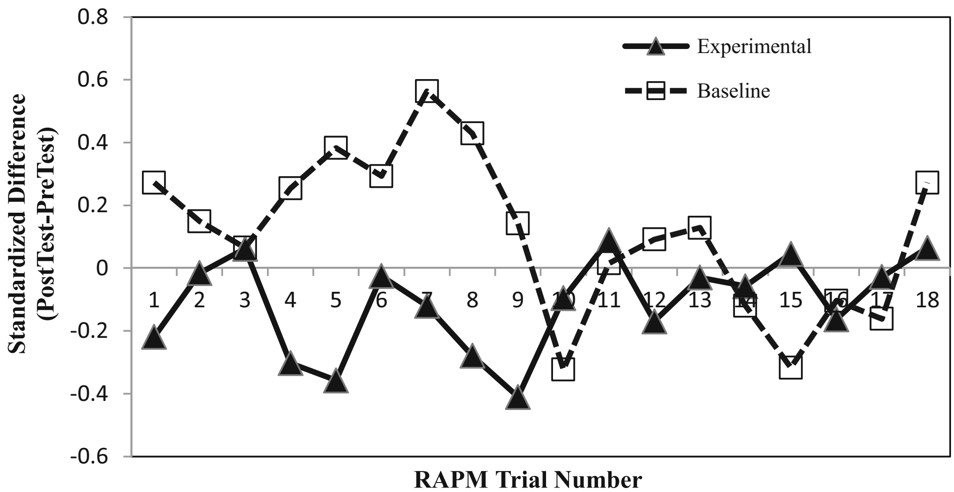

Fig. 2 Standardized RAPM performance difference (post-test minus pre-test proportion correct divided by their respective standard deviations) as a function of serial position and group. Results are pooled across experiments (and in Experiment 2, results in the "experimental" condition are pooled across the two respective groups) (see text for details)

may represent the mindset effect, but also the interaction between mindset and item difficulty. Moreover, mindset effects may be differentially compromised by ceiling and floor effects. In an attempt to partly address the last shortcoming, we standardized the post–pre difference by dividing it by the standard deviation of this serial position (floor and ceiling effects reflect in small standard deviations). Thus, what we present in Fig. 2 is the standardized post–pre differences, with positive differences indicating better post-test than pre-test performance. As evident from Fig. 2 (representing results pooled across experiments), the mindset effect was present in the early items (until Item #10, inclusive), but it then diminished.

There are several interpretations for this time-course effect. First, it is possible that the induced mindset faded as a function of time and/or interfering activity. Second, the mindset may have selectively influenced just the easy items. A third interpretation is that both time/activity and item difficulty played a role. Future studies should determine which one of these (or additional) possibilities is correct.

General discussion

In the present work, we predicted that the instructed-rule mindset would impair RAPM performance, since this test relies on processes that have been shown to be impaired by the mindset. Our results show that executing an instructed rule for only 12 trials, an intervention of about a minute, leads to a performance drop equivalent to 7.9–10.1 IQ points in RAPM performance as observed in three independent groups of participants. Informing the participants that the RAPM would follow the instructed-rule task was ineffective in preventing the drop. No such drop was observed in a general knowledge test, thus suggesting that executing the instructed-rule-impaired control processes and not semantic memory retrieval or motivation in general.

As the instructed-rule task used to induce the mindset did not resemble the RAPM content apart from merely involving the notion of rules, the associated mindset can be considered as involving a *remote-relation*. This mindset involves reduced usage of executive functions, possibly because of the clear signal from the environment that these processes are not required (as much) coupled with the inherent avoidance of using these resources (Kool & Botvinick, 2014; Shenhav et al., 2013). The mismatch between the control processes configuration of the mindset (specifically a decline in rule generation, attention switching, and feedback evaluation, ErEl & Meiran, 2011) and those required for RAPM execution is presumably the reason for the notable performance drop.

As noted by one reviewer, our claim that executive functions are turned off as a result of the instructed-rule manipulation is a bit extreme and unrealistic. A more realistic position can be made within theories viewing control as involving a balance between the need for task maintenance ("persistence") and flexibility (Dreisbach & Goschke, 2004; Goschke, 2003; Hommel, 2015). Accordingly, persistence involves avoiding interferences and shielding the goal from irrelevant information (Dreisbach, 2012). In contrast, generating novel hypotheses and switching between them requires flexibility, namely, being open to information that has previously been irrelevant. Importantly, these theories suggest that the balance between persistence and flexibility is a matter of degree and is adjusted according to task demands and the situational factors (Colzato, van der Wel, Sellaro, & Hommel, 2016; Dreisbach & Goschke, 2004; Hommel, 2015). These situational factors include meditation (Colzato et al., 2016) and positive mood (Ashby & Isen, 1999; Dreisbach &

Goschke, 2004; Steenbergen, Sellaro, Hommel, & Colzato, 2015). Applying this framework to the present study implies that executing an instructed-rule emphasizes persistence. This bias in favour of persistence is then carried over to the RAPM, for which the optimal bias should favour flexibility. More specifically, the balance between persistence and flexibility is one of the aspects of the configuration, which we call "mindset". Being so general in nature makes this balance a *remote-relation* mindset.

There are at least two reasons why the carryover effect has such a drastic influence. One has to do with the natural tendency to avoid the effort associated with the heavy use of the executive functions that are involved in flexibility (Jimura et al., 2010; Kool, McGuire, Rosen, & Botvinick, 2010; Kool & Botvinick, 2014; McGuire & Botvinick, 2010). The other is that persistence may involve shielding of the task goal from interference, and this shielding mode is highly adaptive in many everyday circumstances (Dreisbach & Haider, 2008; Dreisbach, 2012; Dreisbach & Haider, 2009; Reisenauer & Dreisbach, 2013). This may create a preference for persistence whenever the situation allows such a bias. As a result of this preference, shifting the balance from persistence to flexibility is especially costly (Reisenauer & Dreisbach, 2013).

Possible alternative accounts for the mindset influence include ego-depletion (Baumeister, Muraven, & Tice, 2000; Denney & Heidrich, 1990) or a convergent thinking mindset (Fischer & Hommel, 2012). We acknowledge the fact that the results of this study cannot refute these explanations. Yet, ErEl and Meiran's (2011) results using the exact same manipulation showed that a mindset activated by a different convergent task (i.e., rule-generation task) that seemed to involve greater ego-depletion than the current manipulation did not lead to a subsequent decline in performance (Experiment 1, ErEl & Meiran, 2011). Another alternative account is that the results are accidental. We note that apart from the strong support of the Bayesian analysis, the current manipulation has proven successful in the six experiments reported by ErEl and Meiran (2011), and in an additional unreported experiment using RAPM as a dependent measure (a replication of Experiment 1). Assuming independence between experiments, the chance of the present results to be a matter of luck is $0.05^3 = 0.000125$ or odds of 1–8000. Taking into account the additional seven experiments using the present manipulation, the chances are less than 1E$-$13, fairly small to say the least.

An interesting implication of the present results concerns the trait-vs.-state status of Fluid Intelligence tests. While we fully agree that Fluid Intelligence measures are very sensitive to trait, the present results indicate non-trivial state-related (probably transient) effects on a well-known and widely acceptable measure of Fluid Intelligence, following a very brief and seemingly minor intervention. In other words, the present results do not argue against Fluid Intelligence as a stable trait that was probably not affected by the mindset. Instead, the results should be taken to show that *estimates* of Fluid Intelligence might be severely biased by situational factors involving *remote-relation* mindsets.

From a broader perspective, we note that the *remote-relation* nature of the instructed-rule mindset can potentially affect any task involving high control demands. Instructed rules are an integral part of numerous everyday situations (e.g., education and training, labor market, organizations involving hierarchical structure, etc.). Hence, the potential impact of this mindset should be further studied in these socially important domains.

We wish to acknowledge some limitations of the present work. One limitation is that we drew conclusions regarding tests of Fluid Intelligence, but have tested our hypothesis with a specific test (RAPM). Whether the findings extend beyond this test is an open empirical issue. However, given that the mindset has already been shown to affect two different rule-finding tests, a task-switching test, a fluency test, a post-error slowing index, and now RAPM strongly suggests wide applicability. The second shortcoming is that we still do not know how long lasting is the influence of instructed-rule mindset. The third is that we had so far used only one type of manipulation to generate an instructed-rule mindset, and additional manipulations should be studied in the future. Finally, Fluid Intelligence tests are used in the entire range of this ability, and in many different cultures. However, our experiments had so far employed Israeli college/university samples. Thus, whether our conclusions apply to other populations is an additional open empirical question.

Compliance with ethical standards

Ethics approval The experiments presented in this study were approved by the Ethics Committees of the IDC and by the Ethics Committees of Netanya Academic College. Informed consent was obtained from all individual participants included in the study.
All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.
We wish to thank two anonymous reviewers for valuable comments and Dganit Avioz, Viki Abramov and Bat-El Elbaz for their help in data collection.

References

Aronson, J., & McGlone, M. S. (2008). Stereotype and social identity threat. In T. Nelson (Ed.), *The handbook of prejudice, stereotyping, and discrimination* (pp. 153–178). New York, NY: Psychology Press.

Aronson, J., Fried, C. B., & Good, C. (2002). Reducing the effects of stereotype threat on african american college students by shaping theories of intelligence. *Journal of Experimental Social Psychology, 38*, 113–125. doi:10.1006/jesp.2001.1491.

Ashby, F. G., Alfonso-Reese, L., Turken, A. U., & Waldron, E. M. (1998). A neuropsychological theory of multiple systems in category learning. *Psychological Review, 105*, 442–481. doi:10.1037/0033-295X.105.3.442.

Ashby, F. G., & Isen, A. M. (1999). A neuropsychological theory of positive affect and its influence on cognition. *Psychological Review, 106*, 529–550.

Batey, M., Furnham, A., & Safiullina, X. (2010). Intelligence, general knowledge and personality as predictors of creativity. *Learning and Individual Differences, 20*, 532–535. doi:10.1016/j.lindif.2010.04.008.

Baumeister, R. F., Muraven, M., & Tice, D. M. (2000). Ego depletion: a resource model of volition, self-regulation, and controlled processing. *Social Cognition, 18*, 130–150. doi:10.1521/soco.2000.18.2.130.

Bors, D. A., & Stokes, T. L. (1998). Raven's Advanced Progressive Matrices: norms for first-year university students and the development of a short form. *Educational and Psychological Measurement, 58*, 382–398. doi:10.1177/0013164498058003002.

Brown, R. P., & Day, E. A. (2006). The difference isn't black and white: stereotype threat and the race gap on Raven's Advanced Progressive Matrices. *Journal of Applied Psychology, 91*, 979–985. doi:10.1037/0021-9010.91.4.979.

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. doi:10.1037/0033-295X.97.3.404.

Cattell, R. B. (1987). *Intelligence: its structure, growth and action*. Oxford: North-Holland.

Cole, M. W., Yarkoni, T., Repovš, G., Anticevic, A., & Braver, T. S. (2012). Global connectivity of prefrontal cortex predicts cognitive control and intelligence. *The Journal of Neuroscience, 32*, 8988–8999. doi:10.1523/JNEUROSCI.0536-12.2012.

Colzato, L. S., van der Wel, P., Sellaro, R., & Hommel, B. (2016). A single bout of meditation biases cognitive control but not attentional focusing: evidence from the global–local task. *Consciousness and Cognition, 39*, 1–7.

Conway, A. R. A., Cowan, N., Bunting, M. F., Therriault, D. J., & Minkoff, S. R. B. (2002). A latent variable analysis of working memory capacity, short-term memory capacity, processing speed, and general fluid intelligence. *Intelligence, 30*, 163–184. doi:10.1016/S0160-2896(01)00096-4.

Crescentini, C., Seyed-Allaei, S., De Pisapia, N., Jovicich, J., Amati, D., & Shallice, T. (2011). Mechanisms of rule acquisition and rule following in inductive reasoning. *The Journal of Neuroscience, 31*, 7763–7774. doi:10.1523/JNEUROSCI.4579-10.2011.

Denney, N. W., & Heidrich, S. M. (1990). Training effects on Raven's Progressive Matrices in young, middle-aged, and elderly adults. *Psychology and Aging, 5*, 144–145. doi:10.1037/0882-7974.5.1.144.

Dreisbach, G. (2012). Mechanisms of cognitive control the functional role of task rules. *Current Directions in Psychological Science, 21*, 227–231.

Dreisbach, G., & Goschke, T. (2004). How positive affect modulates cognitive control: reduced perseveration at the cost of increased distractibility. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 30*, 343–353.

Dreisbach, G., & Haider, H. (2008). That's what task sets are for: shielding against irrelevant information. *Psychological Research, 72*, 355–361.

Dreisbach, G., & Haider, H. (2009). How task representations guide attention: further evidence for the shielding function of task sets. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 35*, 477–486. doi:10.1037/a0014647.

Duckworth, A. L., Quinn, P. D., Lynam, D. R., Loeber, R., & Stouthamer-Loeber, M. (2011). Role of test motivation in intelligence testing. *PNAS Proceedings of the National Academy of Sciences of the United States of America, 108*, 7716–7720. doi:10.1073/pnas.1018601108.

ErEl, H., & Meiran, N. (2011). Mindset changes lead to drastic impairments in rule finding. *Cognition, 119*, 149–165. doi:10.1016/j.cognition.2011.01.002.

Fischer, R., & Hommel, B. (2012). Deep thinking increases task-set shielding and reduces shifting flexibility in dual-task performance. *Cognition, 123*, 303–307. doi:10.1016/j.cognition.2011.11.015.

Förster, J., Friedman, R. S., & Liberman, N. (2004). Temporal construal effects on abstract and concrete thinking: consequences for insight and creative cognition. *Journal of Personality and Social Psychology, 87*, 177–189. doi:10.1037/0022-3514.87.2.177.

Freitas, A. L., Gollwitzer, P., & Trope, Y. (2004). The influence of abstract and concrete mindsets on anticipating and guiding others' self-regulatory efforts. *Journal of Experimental Social Psychology, 40*, 739–752. doi:10.1016/j.jesp.2004.04.003.

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.

Fujita, K., Gollwitzer, P. M., & Oettingen, G. (2007). Mindsets and pre-conscious open-mindedness to incidental information. *Journal of Experimental Social Psychology, 43*, 48–61. doi:10.1016/j.jesp.2005.12.004.

Fujita, K., Trope, Y., Liberman, N., & Levin-Sagi, M. (2006). Construal levels and self-control. *Journal of Personality and Social Psychology, 90*, 351–367. doi:10.1037/0022-3514.90.3.351.

Galinsky, A. D., & Kray, L. J. (2004). From thinking about what might have been to sharing what we know: the effects of counterfactual mind-sets on information sharing in groups. *Journal of Experimental Social Psychology, 40*, 606–618. doi:10.1016/j.jesp.2003.11.005.

Ginossar, Z., & Trope, Y. (1987). Problem solving in judgment under uncertainty. *Journal of Personality and Social Psychology, 52*, 464–474. doi:10.1037/0022-3514.52.3.464.

Gollwitzer, P. M. (1990). Action phases and mind-sets. In E. T. Higgins & R. M. Sorrentino (Eds.), *Handbook of motivation and cognition* (pp. 53–92). New York: Guilford Press.

Gollwitzer, P. M. (2012). Mindset theory of action phases. In P. Van Lange, A. W. Kruglanski, & E. T. Higgins (Eds.), *Handbook of theories of social psychology* (pp. 526–545). London: Sage.

Goschke, T. (2003). Voluntary action and cognitive control from a cognitive neuroscience perspective. In S. Maasen, W. Prinz, & G. Roth (Eds.), *Voluntary action. An issue at the interface of nature and culture* (pp. 49–85). Oxford, England: Oxford University Press.

Grange, J. A., & Houghton, G. (Eds.). (2014). *Task switching and cognitive control*. New York: Oxford University Press.

Heitz, R. P., Unsworth, N., & Engle, R. W. (2005). Working memory capacity, attention control, and fluid intelligence. In O. Wilhelm R. W. Engle (Ed.), *Understanding and measuring intelligence* (pp. 61–77). New York, NY: Sage Publications, Inc. doi:10.4135/9781452233529.n5.

Hommel, B. (2015). Between persistence and flexibility: The yin and yang of action control. In A. J. Elliot (Ed.), *Advances in motivation science* (pp. 33-67). New York: Elsevier. doi:10.1016/bs.adms.2015.04.003.

Inzlicht, M., & Schmader, T. (2011). *Stereotype threat: Theory, process, and application*. New York: Oxford University Press.

Irwing, P., Cammock, T., & Lynn, R. (2001). Some evidence for the existence of a general factor of semantic memory and its components. *Personality and Individual Differences, 30*, 857–871. doi:10.1016/S0191-8869(00)00078-7.

Jeffreys, H. (Ed.). (1939/1961). *The theory of probability* (1st/3rd ed.) Oxford, England: Clarendon Press.

Jimura, K., Locke, H. S., & Braver, T. S. (2010). Prefrontal cortex mediation of cognitive enhancement in rewarding motivational contexts. *PNAS Proceedings of the National Academy of Sciences of the United States of America, 107*, 8871–8876. doi:10.1073/pnas.1002007107.

Kool, W., & Botvinick, M. (2014). A labor/leisure tradeoff in cognitive control. *Journal of Experimental Psychology: General, 143*, 131–141.

Kool, W., McGuire, J. T., Rosen, Z. B., & Botvinick, M. M. (2010). Decision making and the avoidance of cognitive demand. *Journal of Experimental Psychology: General, 139*, 665–682. doi:10.1037/a0020198.

Kounios, J., Frymiare, J. L., Bowden, E. M., Fleck, J. I., Subramaniam, K., Parrish, T. B., & Jung-Beeman, M. (2006). The prepared mind: neural activity prior to problem presentation predicts subsequent solution by sudden insight. *Psychological Science, 17*, 882–890. doi:10.1111/j.1467-9280.2006.01798.x.

Love, J., Selker, R., Marsman, M., Jamil, T., Verhagen, A. J., Ly, A.,… Wagenmakers, E. J. (2015). JASP (version 0.6.6) [computer software]. Retrieved from https://jasp-stats.org.

Luchins, A. S. (1942). Mechanization in problem solving: The effect of *einstellung*. *Psychological Monographs, 54*, i-95. doi:10.1037/h0093502.

Marshalek, B., Lohman, D. F., & Snow, R. E. (1983). The complexity continuum in the radex and hierarchical models of intelligence. *Intelligence, 7,* 107–127

Markman, K. D., Lindberg, M. J., Kray, L. J., & Galinsky, A. D. (2007). Implications of counterfactual structure for creative generation and analytical problem solving. *Personality and Social Psychology Bulletin, 33*, 312–324. doi:10.1177/0146167206296106.

McGuire, J. T., & Botvinick, M. M. (2010). Prefrontal cortex, cognitive control, and the registration of decision costs. *Proceedings of the National Academy of Sciences of the United States of America, 107*, 7922–7926. doi:10.1073/pnas.0910662107.

Nguyen, H. D., & Ryan, A. M. (2008). Does stereotype threat affect test performance of minorities and women? A meta-analysis of experimental evidence. *Journal of Applied Psychology, 93*, 1314–1334. doi:10.1037/a0012702.

Nisbett, R. E., Aronson, J., Blair, C., Dickens, W., Flynn, J., Halpern, D. F., & Turkheimer, E. (2012a). Group differences in IQ are best understood as environmental in origin. *American Psychologist, 67*, 503–504. doi:10.1037/a0029772.

Nisbett, R. E., Aronson, J., Blair, C., Dickens, W., Flynn, J., Halpern, D. F., & Turkheimer, E. (2012b). Intelligence: new findings and theoretical developments. *American Psychologist, 67*, 130–159. doi:10.1037/a0026699.

Rasmussen, D., & Eliasmith, C. (2011). A neural model of rule generation in inductive reasoning. *Topics in Cognitive Science, 3*, 140–153. doi:10.1111/j.1756-8765.2010.01127.x.

Raven, J. C., & Court, J. H. (1998). *Raven's Progressive Matrices and Vocabulary scales Oxford*. England: Oxford Psychologists Press.

Reisenauer, R., & Dreisbach, G. (2013). The impact of task rules on distracter processing: automatic categorization of irrelevant stimuli. *Psychological Research, 77*, 128–138.

Schmader, T. (2002). Gender identification moderates stereotype threat effects on women's math performance. *Journal of Experimental Social Psychology, 38*, 194–201. doi:10.1006/jesp.2001.1500.

Shenhav, A., Botvinick, M., & Cohen, J. (2013). The expected value of control: an integrative theory of anterior cingulate cortex function. *Neuron, 79*, 217–240. doi:10.1016/j.neuron.2013.07.007.

Smith, P. K., Jostmann, N. B., Galinsky, A. D., & van Dijk, W. W. (2008). Lacking power impairs executive functions. *Psychological Science, 19*, 441–447. doi:10.1111/j.1467-9280.2008.02107.x.

Sprenger, J. (2013). Testing a precise null hypothesis: the case of Lindley's paradox. *Philosophy of Science, 80*, 733–744. doi:10.1086/673730.

Steele, C. M. (1998). A threat in the air: How stereotypes shape intellectual identity and performance. In J. L. Eberhardt S. & T. Fiske (Eds.), *Confronting racism: The problem and the response* (pp. 202-233). Thousand Oaks, CA: Sage Publications, Inc.

Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of african americans. *Journal of Personality and Social Psychology, 69*, 797–811. doi:10.1037/0022-3514.69.5.797.

Steenbergen, L., Sellaro, R., Hommel, B., & Colzato, L. S. (2015). Tyrosine promotes cognitive flexibility: evidence from proactive vs. reactive control during task switching performance. *Neuropsychologia, 69*, 50–55. doi:10.1016/j.neuropsychologia.2015.01.022.

Verguts, T., & de Boeck, P. (2002). The induction of solution rules in Raven's Progressive Matrices test. *European Journal of Cognitive Psychology, 14*, 521–547. doi:10.1080/09541440143000230.

Verguts, T., De Boeck, P., & Maris, E. (1999). Generation speed in Raven's Progressive Matrices test. *Intelligence, 27*, 329–345. doi:10.1016/S0160-2896(99)00023-9.