



HAL
open science

Intra-individual strategy shifts in Raven's matrices, and their dependence on working memory capacity and need for cognition.

Corentin Gonthier, Jean-Luc Roulin

► **To cite this version:**

Corentin Gonthier, Jean-Luc Roulin. Intra-individual strategy shifts in Raven's matrices, and their dependence on working memory capacity and need for cognition.. *Journal of Experimental Psychology: General*, 2020, 149 (3), pp.564-579. 10.1037/xge0000660 . hal-03288841

HAL Id: hal-03288841

<https://hal.science/hal-03288841>

Submitted on 26 Jan 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Intra-individual strategy shifts in Raven's matrices, and their dependence
on working memory capacity and need for cognition

Corentin Gonthier

Université Rennes 2

Jean-Luc Roulin

Université Savoie Mont Blanc

Corentin Gonthier, Université Rennes 2, LP3C EA 1285, 35000 Rennes, France

Jean-Luc Roulin, Univ. Grenoble Alpes, Univ. Savoie Mont Blanc, LPNC UMR CNRS 5105,
38000 Grenoble, France

Correspondence concerning this article should be addressed to Corentin Gonthier, Laboratoire
LP3C, Campus Villejean, Place du Recteur Henri Le Moal, CS 24307, 35043 Rennes Cedex,
France. E-mail: corentin.gonthier@univ-rennes2.fr

The data appearing in this manuscript were presented at the 23rd *Journées Internationales de
Psychologie Différentielle* conference (Luxembourg, June 2018).

©American Psychological Association, 2020. This paper is not the copy of record and may
not exactly replicate the authoritative document published in the APA journal. The final
article is available, upon publication, at: <https://doi.org/10.1037/xge0000660>

Abstract

Strategic behavior plays a key role in fluid intelligence tasks like Raven's matrices. Some participants solve items using the strategy of mentally constructing the answer (constructive matching), which is effective but costly for complex problems; other participants rely on the less accurate strategy of discarding potential answers (response elimination). While this process is relatively well-known, past research hints that intra-individual changes in strategy use may also take place as the task becomes increasingly difficult; however, intra-individual variability in Raven's matrices is poorly understood. The present study aimed to (1) test the hypothesis that participants dynamically shift between strategies during the course of Raven's matrices, as predicted by the literature, and (2) investigate the possibility that these shifts are moderated by individual differences in both ability and motivation. Two samples of 100 participants each completed Raven's advanced progressive matrices, and measures of working memory and need for cognition. The results confirmed that participants tended to turn to the less costly response elimination strategy as the difficulty of the task increased; this variability in strategy use predicted 78% of item-to-item variance in accuracy. Working memory capacity and need for cognition predicted strategy use, and working memory capacity additionally moderated the shift towards response elimination, so that only participants with both high working memory capacity and high need for cognition continued using constructive matching in the later part of the task.

Keywords

Fluid intelligence; Raven's Advanced Progressive Matrices; Strategy use; Intra-individual variability; Working memory capacity; Need for cognition

Highlights

- Strategy use was investigated after each item throughout Raven's APM.
- Participants tend to shift strategies as the difficulty of the task increases.
- Constructive matching is progressively replaced by response elimination.
- Sustained constructive matching depends on ability and motivation.

At first glance, intelligence tests may seem simple: a subject's performance is summed over all items to compute a total score, taken to reflect ability. Yet the simplicity of this process is often misleading. In a task such as Raven's progressive matrices, successive problems increase in perceptual complexity, number and difficulty of logical rules that have to be applied (Carpenter, Just, & Shell, 1990; Primi, 2001). This implies that items are not interchangeable, and intra-individual variability occurs: as the difficulty of the task increases, accuracy declines, and aspects of behavior such as response times, confidence and types of errors can vary. Individual items thus constitute a meaningful level of analysis. In this light, an intelligence task can be seen not as an indivisible test yielding a single score, but as a sequence of problems, whose succession induces positional effects such as modulation of exerted effort (Perret & Dauvier, 2018) and complex learning of one's successes and failures (Ren, Wang, Altmeyer, & Schweizer, 2014). One possibility, and the focus of the present study, is that successive problems also elicit intra-individual changes of strategy (Schunn & Reder, 2001).

Intra-individual variability in reasoning tasks is obscured by the traditional approach of computing a total score, but it is not a trivial issue. Multiple authors have recently stressed the importance of understanding intra-individual processes (see Borsboom, Cramer, Kievit, Scholten, & Franic, 2009; Nesselroade & Ram, 2004; Wang, Hamaker, & Bergeman, 2012). One reason is that variability can be viewed as an integral part of behavior, so that a complete theory has to be able to explain it (Underwood, 1975). In some cases, intra-individual variability can constitute a useful marker of individual differences, with the amount of variability indexing cognitive development (Siegler, 1994), cognitive aging (Fagot et al., 2018), or the effectiveness of abilities such as attention control (McVay & Kane, 2012). More generally, intra-individual processes are especially important in that they are reflections of the actual mechanisms leading to a response (Borsboom et al., 2009). This is particularly true of time-structured intra-individual variability (i.e., intra-individual variability that is not random,

but structured in relation to time; Ram & Gerstorf, 2009). In this sense, understanding how behavior changes across trials is necessary to understand the pathway a subject follows from stimulus to answer.

Intra-Individual Variability in Strategy Use

In the case of intelligence tasks, a key mechanism linking individual ability to response on an item is strategy use – the specific procedure participants use to perform the task (Lemaire & Reder, 1999). Visuo-spatial reasoning tasks like Raven's matrices, where participants have to find the missing piece in a visual display among several alternatives, are characterized by two major strategies: constructive matching and response elimination (Bethell-Fox, Lohman, & Snow, 1984; Snow, 1980; see also Hunt, 1974, for a possible visual strategy restricted to easy problems). Constructive matching involves mentally reconstructing the missing piece of the matrix, whereas response elimination involves systematically comparing the response alternatives with the matrix to find one that seems to fit. These two strategies constitute qualitatively different ways to solve the task, which affects patterns of performance: constructive matching is typically associated with better accuracy, longer time on the problems, higher confidence in the answers, and lesser sensitivity to salient distractors (e.g. Arendasy & Sommer, 2013; Mitchum & Kelley, 2010; Vigneau, Caissie, & Bors, 2006). Understanding variability in strategy use is therefore essential to understand the dynamics of success and failure in intelligence tasks (Kyllonen, Lohman, & Woltz, 1984; see also Schunn & Reder, 2001).

Individuals differ systematically in their tendency to use constructive matching and response elimination: high-ability subjects tend to demonstrate more constructive matching, which could partly explain their higher performance (Arendasy & Sommer, 2013; Bethell-Fox et al., 1984; Snow, 1978, 1980). However, there is good reason to suspect that strategy shifts also occur for the same individual throughout a reasoning task (Bethell-Fox et al., 1984; Snow,

1978, 1980; see also Lawson & Kirby, 1981). Intra-individual variability constitutes a major aspect of strategy use: subjects have been observed in multiple contexts to flexibly adapt their strategy to the problem at hand, which contributes to successful strategic behavior (Lemaire & Siegler, 1995; Schunn & Reder, 2001; Siegler & Shipley, 1995). In tasks like Raven's matrices, successive items increase in number of rules and elements that have to be combined to reconstruct the answer; for this reason, the conceptual cost of implementing constructive matching increases throughout the task (Mulholland, Pellegrino, & Glaser, 1980; Pellegrino & Glaser, 1980). Thus, individuals could be expected to adapt their strategy use by progressively shifting to response elimination (Bethell-Fox et al., 1984; Snow, 1978, 1980). In turn, this shift could contribute to the progressive decline in accuracy.

This prediction is supported by existing literature about intra-individual variability of behavior. Studies on decision making strategies have concluded that subjects tend to switch to simpler decision heuristics when faced with more complex problems (Klayman, 1985; Payne, 1976). This pattern can be described as a form of satisficing: subjects attempt to reach a "good enough" solution, instead of systematically trying to resolve all aspects of the situation. In essence, individuals tend to perform decision tasks based on a trade-off between effort and likelihood of success, using the most efficient strategy but not necessarily the most effective one (Klayman, 1985). Another line of literature suggests that the probability of using a strategy depends on its cost and its history of success (Lovett & Anderson, 1996; Verguts, Maris, & De Boeck, 2002). Even when a situation encourages subjects to always use the same strategy, presenting problems where this strategy fails tends to elicit strategy shifts (Luchins, 1942). Thus, as the complexity of items increases and the likelihood of producing a correct answer decreases, constructive matching should become less and less desirable in a reasoning task.

Given their hypothesized existence and their potential impact on performance, there has been surprisingly little interest in strategy shifts on reasoning tasks. In his initial account of

constructive matching and response elimination, Snow (1978, 1980) reported anecdotal observations that some subjects tended to use constructive matching on easy items, and to fall back on the less costly response elimination strategy on more difficult items. Unfortunately, this report detailed neither the results nor the exact tasks on which this pattern was observed. Snow's conclusions were confirmed in a later study on an analogy task (of the form "A is to B as C is to ?"), in a small sample of 35 high school students (Bethell-Fox et al., 1984). Using a combination of verbal reports, eye-tracking and analysis of response times data, the authors found evidence of less frequent use of constructive matching on more complex items. However, this conclusion was not supported by a more recent study using eye-tracking in a larger sample of 55 university students performing Raven's matrices: the results suggested stable strategy use, independently of item difficulty (Vigneau et al., 2006; see Table 1). Thus, it is still unclear whether participants actually demonstrate, on average, an intra-individual shift towards the simpler response elimination strategy in intelligence tasks.

Predictors of Intra-Individual Variability

Assuming that strategy shifts do occur in intelligence tasks, there is also good reason to suspect that individuals could systematically differ in their tendency to fall back on response elimination; in other words, that there exist inter-individual differences in intra-individual variability (see Reder & Schunn, 1999; Schunn & Reder, 1998, 2001). For example, Bethell-Fox and colleagues (1984) concluded that participants with higher fluid intelligence may be more likely to maintain the use of constructive matching throughout the task and less likely to shift to response elimination on complex items. More generally, the tendency to use complex decision strategies rather than simple heuristics has been conceptualized as depending on both ability and motivation (Petty & Cacioppo, 1986), and both ability and motivation have also been identified as possible predictors of changes of strategy (e.g. Reder & Schunn, 1999; Siegler & Shipley, 1995). In other words, individuals would be more likely to keep using a

complex strategy on a complex item if they have both the ability and the motivation to do so: this can be described as a cost-benefit account of mental effort (Sandra & Otto, 2018).

Which ability and motivation dimensions could influence a shift from constructive matching to response elimination? In the face of a recent study concluding that fluid intelligence is predicted by a combination of working memory capacity and need for cognition (Hill, Foster, Sofko, Elliott, & Shelton, 2016), these two constructs appear as particularly likely candidates.

Working memory capacity (WMC), or the ability to hold information in memory while performing concurrent processing, is not only one of the best predictors of performance in fluid intelligence tasks (Ackerman, Beier, & Boyle, 2005; Oberauer, Schulze, Wilhelm, & Süß, 2005); one study has also observed WMC to predict intra-individual variability of performance in Raven's matrices (Little, Lewandowsky, & Craig, 2014; but see Salthouse, 1993; Unsworth & Engle, 2005, for divergent results). Importantly, WMC is also a predictor of strategy use: participants with high working memory capacity use more complex and more effective strategies in a large range of tasks (e.g. Thomassin, Gonthier, Guerraz, & Roulin, 2015), and tend to use more constructive matching in Raven's matrices in particular (Gonthier & Thomassin, 2015; see also Jarosz & Wiley, 2012; Jastrzębski, Ciechanowska, & Chuderski, 2018). Some studies also found WMC to predict intra-individual variability in strategy use in cognitive tasks (Reder & Schunn, 1999; Schunn & Reder, 1998, 2001; but see Schunn, Lovett, & Reder, 2001). Because the increasing complexity of items in Raven's matrices can be understood as an increasing number of rules to simultaneously maintain in working memory (Carpenter et al., 1990), WMC can be expected to moderate the shift from constructive matching to response elimination: over more difficult items, it would be comparatively easier for participants with high WMC to continue using constructive matching to integrate all rule elements in working memory (see Bethell-Fox et al., 1984; Mulholland et al., 1980; Pellegrino

& Glaser, 1980). A single unpublished study tested this possibility, with inconclusive results (Jarosz, 2015).

Need for cognition (NFC), or the tendency to engage in and enjoy complex thinking, is also a predictor of performance in intelligence tests, and may contribute to complex performance above and beyond ability itself (Fletcher, Marks, & Hine, 2011; Hill et al., 2013, 2016; Toplak, West, & Stanovich, 2014; von Stumm, Hell, & Chamorro-Premuzic, 2011). NFC can be viewed as the motivation to expand effort to perform costly cognitive processing (e.g. Sandra & Otto, 2018). While it is straightforward to understand why this could lead to higher performance in intelligence tasks (Perret & Dauvier, 2018; Raven, Raven, & Court, 1998), this definition also suggests a relation with strategy use. Because constructive matching is viewed as more costly, participants with a high NFC may be more likely to engage in this strategy. This hypothesis does not seem to have been examined in the context of intelligence tasks, but prior studies have suggested that participants with low NFC tend to use more heuristic decision making strategies, especially under time pressure (Verplanken, 1993); low NFC has also been related to the tendency to use surface, rather than deep learning strategies (Evans, Kirby, & Fabrigar, 2003). Need for cognition could also conceivably moderate intra-individual shifts in strategy use: high motivation to engage in effortful cognitive processing could encourage participants with high NFC to keep using constructive matching, even in the face of increasing difficulty.

Rationale for the Present Study

The overarching goal of this study was to investigate patterns of intra-individual variability in one particular reasoning task, Raven's progressive matrices. Our first objective was to test the hypothesis that as the difficulty of the task increases, participants tend to shift to response elimination (presumably due to the increasing conceptual cost of the constructive matching strategy; Bethell-Fox et al., 1984; Snow, 1978, 1980). To this end, participants

completed Raven's Advanced Progressive Matrices (APM), and their responses and patterns of strategy use were assessed at the level of each successive item. Our second objective was to examine the possibility that intra-individual shifts in strategy use depend on individual differences in ability and motivation. Two candidate constructs, working memory capacity and need for cognition, were respectively assessed with a battery of complex spans and a self-report questionnaire.

Strategy use in the APM was primarily measured with verbal reports (Gonthier & Thomassin, 2015; Jastrzębski et al., 2018; Mitchum & Kelley, 2010). Because constructive matching use is moderately associated with longer time on the problem (Gonthier & Thomassin, 2015; Mitchum & Kelley, 2010; Vigneau et al., 2006; see also Foorman, Sadowski, & Basen, 1985; Partchev & De Boeck, 2012), response times were also investigated. Response times have the advantage of constituting an objective behavioral measure, and intra-individual changes in response times have been found to offer important insight into participants' engagement and invested effort (Goldhammer, Naumann, & Greiff, 2015; Perret & Dauvier, 2018). Critically, participants performed the APM with no time limit to allow for unconstrained strategy use (see Thomassin, Gonthier, Roulin, & Guerraz, 2015).

A particular feature of the current study was its strategy for data analysis. Modeling intra-individual variability is often difficult, both because the timecourse of behavior within a task is not necessarily linear, and because responses at successive timepoints are likely to be correlated within the same individual. Both issues can be solved with the use of Generalized Additive Mixed Models (GAMMs; Wood, 2017). GAMMs can be viewed as an extension of the general linear model, with two major advantages over traditional linear regression: the ability to model non-linear effects of predictors, without *a priori* hypotheses on the shape of these effects; and the inclusion of random effects at the individual level to model temporal autocorrelation of responses. Both features make GAMMs particularly suited to the analysis of

intra-individual variability. GAMMs have recently begun emerging in psychological studies of intra-individual variability (for an example in the field of intelligence, see Perret & Dauvier, 2018; for other examples, see Baayen, Vasishth, Kliegl, & Bates, 2017; Le Vigouroux, Pavani, Dauvier, Kop, & Congard, 2017; McKeown & Sneddon, 2014). In the present study, GAMMs were used to systematically model the timecourse of accuracy, response time, constructive matching and response elimination use throughout Raven's APM; as allowed by this type of analysis, the shape of these timecourses – linear or not – was directly estimated from the data.

Method

Statistical Power

The necessary sample size was estimated based on prior studies, due to the lack of either a consensual minimum sample size or a straightforward way to estimate statistical power for GAMM analyses. Intra-individual shifts in strategy use in a reasoning task were first described with $N = 35$ participants (Bethell-Fox et al., 1984), but they did not emerge in a larger sample of $N = 55$ (Vigneau et al., 2006). The correlation between strategy use and working memory capacity has been observed with $N = 95$ participants (Gonthier et al., 2015; see also Jarosz & Wiley, 2012, for data with $N = 35$), the interaction between working memory capacity and need for cognition was found with $N = 167$ participants (Hill et al., 2016), and an interaction between working memory capacity and item difficulty was found with $N = 130$ participants (Little et al., 2014). Based on this information, data collection was planned for 200 participants.

Participants

A sample of 200 undergraduate students at the University of Rennes 2 participated for course credit (161 females and 39 males; mean age = 19.94 years, $SD = 1.77$). All participants were native French speakers, and none had completed any of the experimental tasks before. All participants provided written informed consent prior to the experiment.

Materials

Raven's Advanced Progressive Matrices.

Participants completed the original version of Set II of Raven's APM (Raven et al., 1998). Each item is composed of a 3 x 3 matrix of black-and-white figures, where the bottom right piece is missing; participants are required to select the figure that logically completes the matrix, among eight possible answers. In order to maintain a reasonable testing time despite the added duration of strategy questions, each participant completed only half of the 36 items, as in prior studies (e.g. Gonthier et al., 2015; Jastrzębski et al., 2018; Unsworth, Redick, Lakey, & Young, 2010). Half the participants ($n = 100$) completed the odd-numbered items and the other half ($n = 100$) completed the even-numbered items, thus allowing for analyses over all items of the APM.

After each APM problem, participants were required to answer two questions about the strategies they used to figure out the missing piece of the matrix. The first item assessed the use of the constructive matching strategy ("After examining the drawing, you imagined the missing piece before looking for it among the possible answers"), and the other assessed the response elimination strategy ("You examined each possible answer in turn to decide whether it could be the missing piece"). Participants were asked to rate their agreement with each proposition on a 9-point Likert scale. The two items and their response scales were presented simultaneously on the same screen (with the constructive matching question above the other), so that participants had to answer both questions at once before moving on to the next item¹. Constructive matching and response elimination were first analyzed separately; a score

¹ Participants were also asked to rate their confidence in the fact that their answer was correct (as a follow-up to Mitchum & Kelley, 2010), on a visual analogue scale ranging from 0% to 100%. These data were not directly relevant to the present study and were not analyzed here, but they are available along with the main data file for interested readers.

reflecting effective strategy use was also computed as (constructive matching – response elimination) to combine both strategic behaviors with a single index.

Both strategy items were adapted from a questionnaire used in a prior study (Gonthier et al., 2015), which included two items per strategy. The decision to retain only one question per strategy was made because strategy use was assessed after each problem in the current study (Nezlek, 2017). Our concerns were both to limit testing time, and to limit the tediousness of answering the same set of questions on 20 consecutive problems (including two training problems): an 8-item (Gonthier et al., 2015) or a 10-item (Jastrzębski et al., 2018) strategy questionnaire would have had participants answer 160 or 200 identical questions, potentially decreasing participant engagement. The two items retained in the current study were chosen because they had higher correlation with total APM accuracy, and because anecdotal reports from participants questioned the suitability of the other question assessing response elimination.

An important question is whether this shortened strategy measure retained sufficient reliability for a study of individual differences, including intra-individual variation (e.g. Nezlek, 2017). At the between-subjects level, Cronbach's alphas computed across all trials were above .90 for both constructive matching and response elimination, indicating reliable inter-individual differences. To estimate of the reliability of intra-individual timecourses of strategy use, we used a form of the split-half method. A GAMM model was defined with only time as a predictor, including random timecourses for each subject. This model was estimated separately (using restricted maximum likelihood) based on half the items, and the predicted timecourses were extracted for each subject. Lastly, the correlation between the two estimated timecourses was computed for each subject. The median split-half correlation between estimated timecourses was $r = .98$ for accuracy, $r = .96$ for response times, $r = .83$ for constructive matching and $r = .85$ for response elimination. In other words, intra-individual

timecourses of strategy use appeared internally consistent despite being measured with a single item per time point.

Working memory capacity.

Working memory capacity was measured with the Composite Complex Span (CCS; Gonthier et al., 2016), a French-speaking battery of three classic complex span tasks: the reading span, symmetry span and operation span (see Unsworth, Heitz, Schrock, & Engle, 2005; Redick et al., 2012). In each trial of a complex span, participants have to alternate between solving simple problems (deciding whether sentences are correct, whether spatial displays are symmetrical, and whether mathematical operations are correct, respectively) and memorizing unrelated stimuli (digits, locations in a 4x4 grid, and consonants, respectively). At the end of a trial, all to-be-memorized stimuli have to be recalled in the correct order.

The CCS includes a total of 22 trials, with set sizes ranging from 3 to 8. Performance in each complex span is computed as the total number of stimuli recalled in the correct position (partial-credit load scoring: see Conway et al., 2005). A domain-general WMC estimate is then computed as the average performance over the three complex spans, after standardization. The CCS has demonstrated both internal consistency ($\omega_t = .86$) and convergent validity with the APM (Gonthier et al., 2016), and has been used with over 1500 participants.

Need for cognition scale.

Need for cognition was assessed with a French-speaking adaptation of the 18-item short form of the need for cognition scale (Cacioppo, Petty, & Kao, 1984). The French-speaking version includes 11 items and has demonstrated satisfying psychometric properties in samples of undergraduate students (Salama-Younes, 2011).

Procedure

Participants performed the testing session in groups of 2 to 12 individuals in a university computer room. The first task of the experimental session was the CCS, which lasted approximately 25 minutes. After a short break, participants completed two training items from Set I of the APM, followed by the rest of the APM task performed with unlimited time. The whole experimental session lasted approximately 60 min. Data collection was performed in two waves: a first wave of participants completed the odd-numbered items of the APM, and a second wave occurring several weeks later completed the even-numbered items.

Data Analysis

Data collected with the odd-numbered and even-numbered items of the APM were combined for all analyses, so as to estimate intra-individual variability over the whole task. Due to the risk of over-fitting posed by certain GAMM analyses, the two sets of items were also used for cross-validation to ensure stability of the results: for each statistical test, we confirmed that the effects and especially the functional form of the relationship between variables were comparable in the Odd and Even conditions. This was the case for all analyses reported here unless noted.

Statistical inference in GAMM analyses was performed based on approximate p -values (see Wood, 2017). The alternative strategy of comparing the fit of concurrent models gave similar results. For each predictor, we report the decision statistic (chi-square tests when modeling accuracy on the APM, F -tests for the other dependent variables), the corresponding p -value, and the effective degrees of freedom (*edf*: effective degrees of freedom equal to 1 reflect a linear relationship between predictor and dependent variable, values greater than 1 reflect a more complex trajectory). Indices of model fit such as the percentage of explained deviance are not reported, due to their being misleading in this context: position of the item within the APM (i.e., item difficulty) explained a very large share of variability in all cases.

GAMM analyses were performed using the *mgcv* package (Wood, 2017; version 1.8-16) for *R* (R Core Team; version 3.3.2). All dependent variables were modeled assuming a gaussian distribution, except for accuracy, which used a binomial distribution. Response times were log-transformed prior to statistical inference to account for positive skewness. Subject-level random effects were modeled as random intercepts, which was sufficient to keep lag 1 autocorrelation of the residuals below .10 for all variables, except for response times for which random slopes also had to be included. Models were fit using maximum likelihood. Basis dimension was adjusted so as to be sufficient for all analyses ($k = 18$ for the timecourse of accuracy and response times and $k = 9$ for the timecourse of strategy use). Smooths were modeled with the default classes – thin plate regression splines for simple smooths and cubic regression splines for interactions.

Results

Data files and analysis scripts detailing the syntax of tested models can be accessed via the Open Science Framework platform at <https://osf.io/pykxu/>.

Preliminary Analyses

The data were first screened for outliers. Three participants were excluded from the sample because they failed to perform Raven's matrices appropriately (spending less than 30 seconds per item on average and obtaining three correct answers or less), and three more participants were excluded because they failed to reach the criterion of minimal processing accuracy in the working memory task (Gonthier et al., 2016; see also Unsworth et al., 2005). The final sample size was $n = 97$ for the Odd condition and $n = 97$ for the Even condition.

Descriptive statistics are presented in Table 1. All measures demonstrated satisfying range and a distribution close to normal (except for response times on Raven's matrices, which were positively skewed). Participants in the Odd and Even conditions did not differ in terms of

working memory capacity, $t(192) = 0.17$, $p = .867$, or need for cognition, $t(192) = 0.32$, $p = .750$. Congruent with the literature, total scores on the APM correlated positively with both working memory capacity, $r(192) = .36$, $p < .001$, and need for cognition, $r(192) = .43$, $p < .001$. WMC was moderately correlated with NFC, $r(192) = .34$, $p < .001$. These correlations were in the same range as in prior studies (e.g. Ackerman et al., 2005; Hill et al., 2013, 2016).

The correlations between strategy items and other variables were of particular interest to ensure that our single-item strategy measures retained significant predictive power. Greater use of constructive matching (averaged over all items) was related to higher scores on the APM, $r(192) = .38$, $p < .001$, and to longer response times, $r(192) = .23$, $p = .001$; conversely, greater average use of response elimination predicted lower scores, $r(192) = -.32$, $p < .001$, and faster response times, $r(192) = -.28$, $p < .001$. These results are in line with prior studies (although with slightly lower correlations than for longer strategy measures; Gonthier & Thomassin, 2015; Mitchum & Kelley, 2010; Vigneau et al., 2006). Constructive matching and response elimination demonstrated a small negative correlation, $r = -.16$, $p = .026$, also consistent with prior studies (Gonthier & Thomassin, 2015; Jastrzębski et al., 2018, Experiment 1). None of these correlations differed between the Odd and Even conditions (all Bayes Factors > 50 in favor of the null). Overall, these results confirm that all measures functioned correctly in this sample.

Table 1

Descriptive statistics for all measures

Measure	<i>M</i>	<i>SD</i>	Range	Skewness	Kurtosis
APM – total score	10.08	3.29	2 to 18	0.06	-0.46
APM – average RT	47.06	23.74	13.21 to 139.97	1.63	3.16
APM – average CM	6.62	1.81	1.00 to 9.00	-1.02	0.68
APM – average RE	5.66	1.77	1.39 to 9.00	-0.39	-0.49
APM – average ESU	0.96	2.73	-8.00 to 7.11	-0.30	0.64
Working memory capacity	0.00	1.00	-2.13 to 2.21	-0.06	-0.61
Need for cognition	29.06	4.71	14 to 41	-0.19	0.41

Note. RT = Response Time (in seconds); CM = Constructive Matching; RE = Response Elimination; ESU = Effective Strategy Use (CM-RE). Possible values ranged from 0 to 18 for the total APM score, from 1 to 9 for the average use of constructive matching and response elimination, and from 11 to 44 for the need for cognition scale. Working memory capacity was standardized.

Intra-individual Variability in Raven's matrices

A first series of analyses investigated the timecourse of intra-individual change in Raven's matrices for the four indices of performance: accuracy, response times, constructive matching and response elimination. Intra-individual variability was indexed based on item ordinal position.

As expected, accuracy declined as a function of item position, $\chi^2 = 701.35$, $edf = 8.50$, $p < .001$. This decline is displayed in Figure 1A. The decline in performance throughout the task was not perfectly monotonic, with a slight increase in performance on items 30 to 33. This pattern also appeared in prior studies to varying extents (e.g. Little et al., 2014; Unsworth & Engle, 2005; Vigneau et al., 2006) and may reflect the fact that items of the APM are not perfectly ranked in order of increasing difficulty. Response times also varied throughout the task, $F = 48.58$, $edf = 15.27$, $p < .001$; this change had a more complex trajectory than accuracy, as depicted in Figure 1B. On average, response times first increased along with the

difficulty of the problems, then started to decrease around item 30, suggesting that participants tended to disengage from the task over the very difficult final items.

Critically, there was significant intra-individual variability in both constructive matching, $F = 35.39$, $edf = 6.56$, $p < .001$, and response elimination, $F = 42.34$, $edf = 3.74$, $p < .001$. The timecourse of strategy use is represented in Figure 1C. Ratings for both strategies were high throughout the task, suggesting that participants tended to approach problems strategically; the use of constructive matching declined as the task progressed, whereas the use of response elimination increased. On average, participants used constructive matching to a greater extent in the first half of the APM, but the balance between the two strategies gradually shifted – with a turning point around item 25, where response elimination became descriptively more prevalent (see Figure 1C). In sum, the results confirmed the conclusion of Bethell-Fox and colleagues (1984) that participants tend to turn away from the costly strategy of constructive matching as the task progresses.

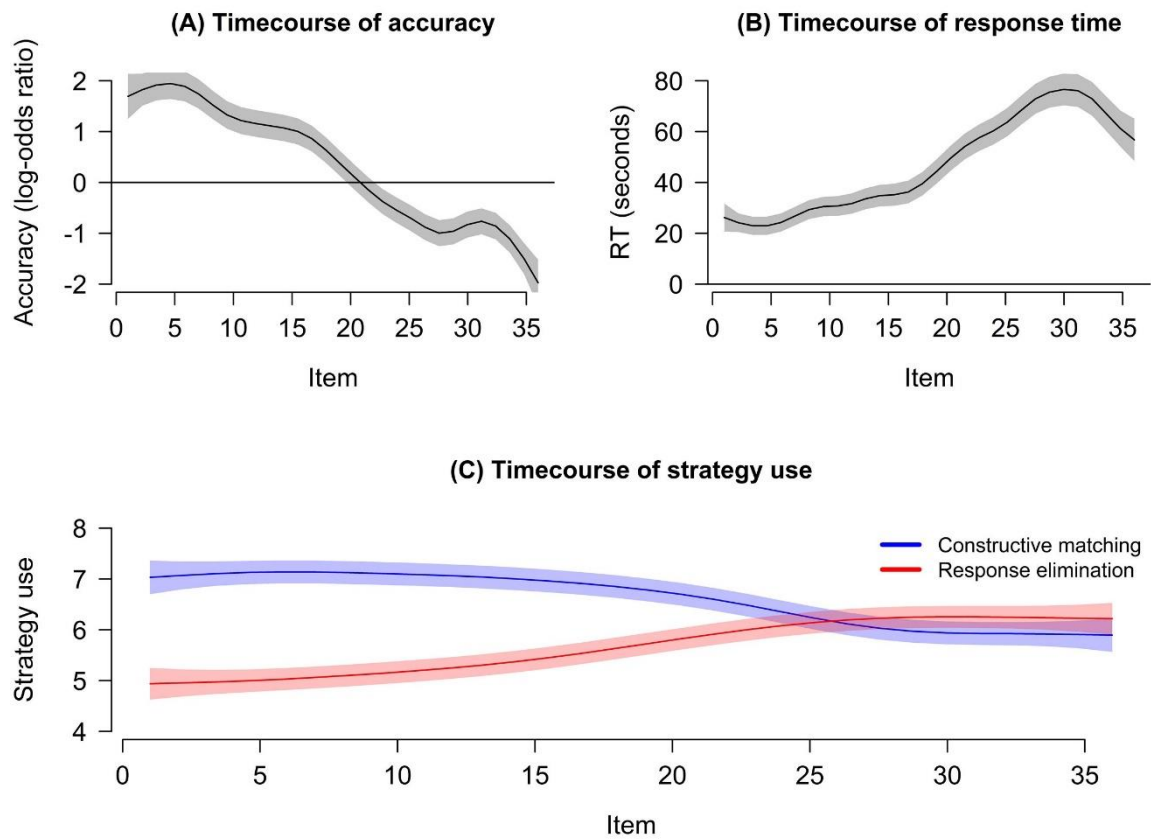


Figure 1. Timecourse of (A) accuracy, (B) response times, and (C) strategy use as a function of item position in the APM. Shaded regions around the prediction lines represent 95% confidence intervals. Accuracy in panel (A) is represented as the logarithm of the odds-ratio of giving a correct answer, with 0 indicating a 50% chance for the answer to be correct. Strategy use in panel (C) is expressed as the rating on a scale from 1 to 9.

In order to more closely investigate individual patterns of transition from constructive matching to response elimination, a complementary analysis examined the timecourse of the effective strategy score, computed as (constructive matching – response elimination), while allowing for random smooths at the subject level. Predicted values of the strategy score were then extracted for each participant at each time point. The results are displayed in Figure 2. Overall, predicted strategy scores were above zero (indicating greater constructive matching than response elimination) in 66% of all trials. For the first APM problem, predicted strategy scores were above zero for 153 participants (79%) and below zero for the remaining 41 participants (21%); by contrast, for the final APM problem, predicted strategy scores were below zero for more than half the sample (100 participants – 52%). All but 25 participants (13%) had a lower strategy score at the end of the task than at the beginning. Across all participants, the median point at which the timecourse of strategy scores crossed zero was item 24 (median absolute deviation = 3).

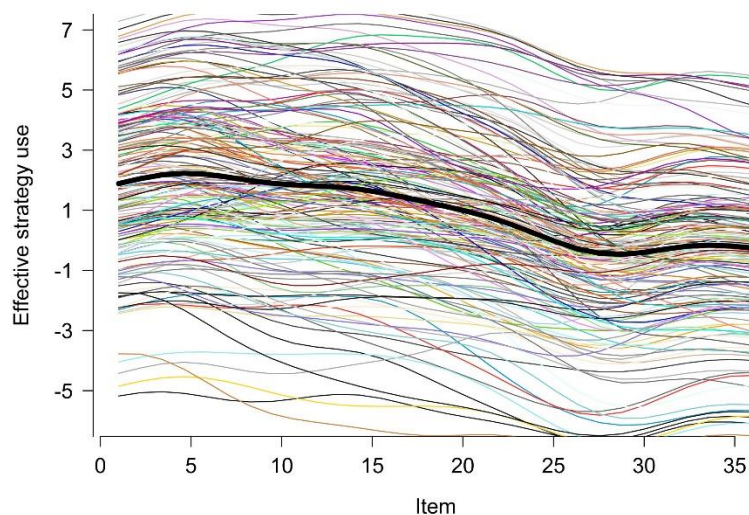


Figure 2. Timecourse of effective strategy use as a function of item position in the APM. Effective strategy use was computed as (constructive matching – response elimination). The thick black line represents the predicted timecourse collapsed across all subjects; thinner lines represent predicted timecourses for each subject.

To determine whether the timecourse of strategy use was consistent with the timecourse of the decline in accuracy, another complementary analysis tested the correlation between the two, averaging results at the item level. This analysis revealed that average accuracy on an item of the APM was very highly correlated with average use of constructive matching on this item, $r(34) = .82$, $p < .001$, and negatively correlated with average use of response elimination, $r(34) = -.85$, $p < .001$. Use of constructive matching on a given item had a high negative correlation with use of response elimination, $r(34) = -.77$, $p < .001$, confirming that the two had an inverse relationship. Summarizing strategic behavior with the effective strategy use index, computed as (constructive matching – response elimination), yielded the best prediction of average accuracy on an item with over 78% of explained variance, $r(34) = .89$, $p < .001$; this finding is summarized in Figure 3. These results confirm that the timecourse of the decline in accuracy closely matched the timecourse of decline in effective strategy use.

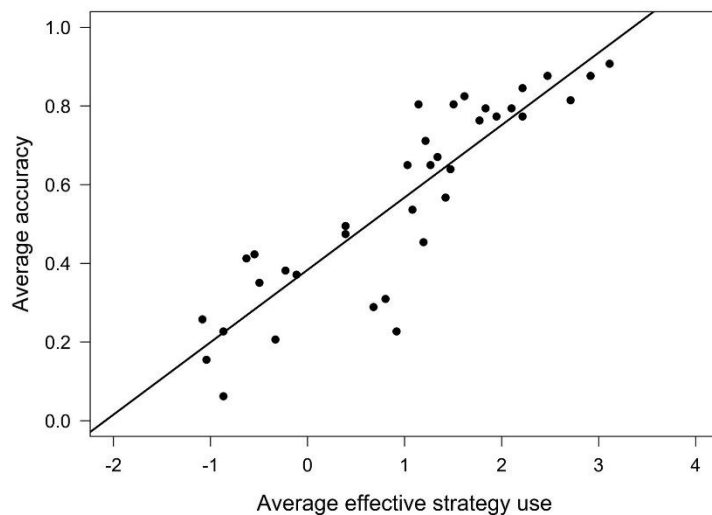


Figure 3. Correlation between effective strategy use and accuracy, averaged at the item level ($N = 36$ items). The solid line represents the slope of a linear regression line. Effective strategy use was computed as (constructive matching – response elimination).

Predictors of Intra-individual Variability

A second series of analyses aimed to understand how working memory capacity and need for cognition are related to intra-individual variability in strategy use on Raven's matrices. The following analyses tested the main effect of working memory on the various measures (accuracy, response time, strategy use), and the moderation of their timecourse by working memory. Interactions are represented by contour plots (Wood, 2017), with item number on the x-axis, WMC on the y-axis, and values of the dependent variable represented as different colors. These plots are easier to navigate horizontally (by reading the timecourse of the dependent variable for a particular value of WMC) or vertically (by reading the effect of WMC on the dependent variable for a particular item); an example is given in the caption of Figure 4.

The positive main effect of WMC on accuracy in the APM was significant, $\chi^2 = 29.54$, $edf = 1.00$, $p < .001$, with the estimated degrees of freedom equal to 1 indicating a linear relationship. The timecourse of accuracy throughout the task was significantly moderated by WMC, $\chi^2 = 5.28$, $edf = 1.02$, $p = .022$; this interaction is represented in Figure 4A. WMC was predictive of higher scores across the whole task, although this benefit was a little smaller over the final items².

The main effect of WMC on response time was significant, $F = 5.73$, $edf = 1.00$, $p = .017$; the timecourse of response time was also moderated by WMC, $F = 4.64$, $edf = 9.77$, $p < .001$. These results are represented in Figure 4B. WMC had little effect on response times over the first trials of the APM; however, a large effect of WMC emerged around item 25. Predicted values indicated that participants with high WMC (one standard deviation above the

² A complementary analysis based on point-biserial correlations was run to reproduce analyses reported in past literature (Little et al., 2014; Salthouse, 1993; Unsworth & Engle, 2005). Point-biserial correlations converged with the results of GAMM analyses in showing a progressive decrease of WMC-accuracy correlations throughout the task. This discrepancy with the results of Little and colleagues (2014) might be partly due to their use of a speeded APM task versus an unspeeded task here.

mean) spent on average 26.7 seconds on item 1 and 69.6 seconds on item 36; participants with low WMC (one standard deviation below the mean) spent on average 25.9 seconds on item 1 and 44.0 seconds on item 36. In other words, participants with high WMC devoted selectively more time to the more difficult items.

The effect of WMC on strategy use was first analyzed through the lens of the effective strategy use index. A high WMC predicted a higher strategy score, $F = 11.18$, $edf = 1.00$, $p < .001$, indicating that participants with higher working memory capacity tended to use constructive matching to a greater extent than response elimination (in line with the results of Gonthier et al., 2015). Critically, WMC also moderated the timecourse of the strategy score, $F = 2.94$, $edf = 6.36$, $p = .002$; this moderation is represented in Figure 4C. While all participants showed a progressive decrease in strategy score, corresponding to the shift from constructive matching to response elimination, participants with high WMC maintained a positive strategy score over more items – corresponding to extended maintenance of constructive matching.

Examining predicted values revealed that participants with low WMC (one standard deviation below the mean) showed a decrease in strategy scores after item 7, and maintained a constantly low strategy score over the second half of the task. By contrast, participants with high WMC (one standard deviation above the average) maintained a high strategy score over the first half of the task, and only demonstrated a decrease after item 15. Participants with the highest WMC (two standard deviations above the average) demonstrated an approximately constant strategy score from item 1 to item 21. In other words, participants with a high WMC tended to shift from constructive matching to response elimination later in the task.

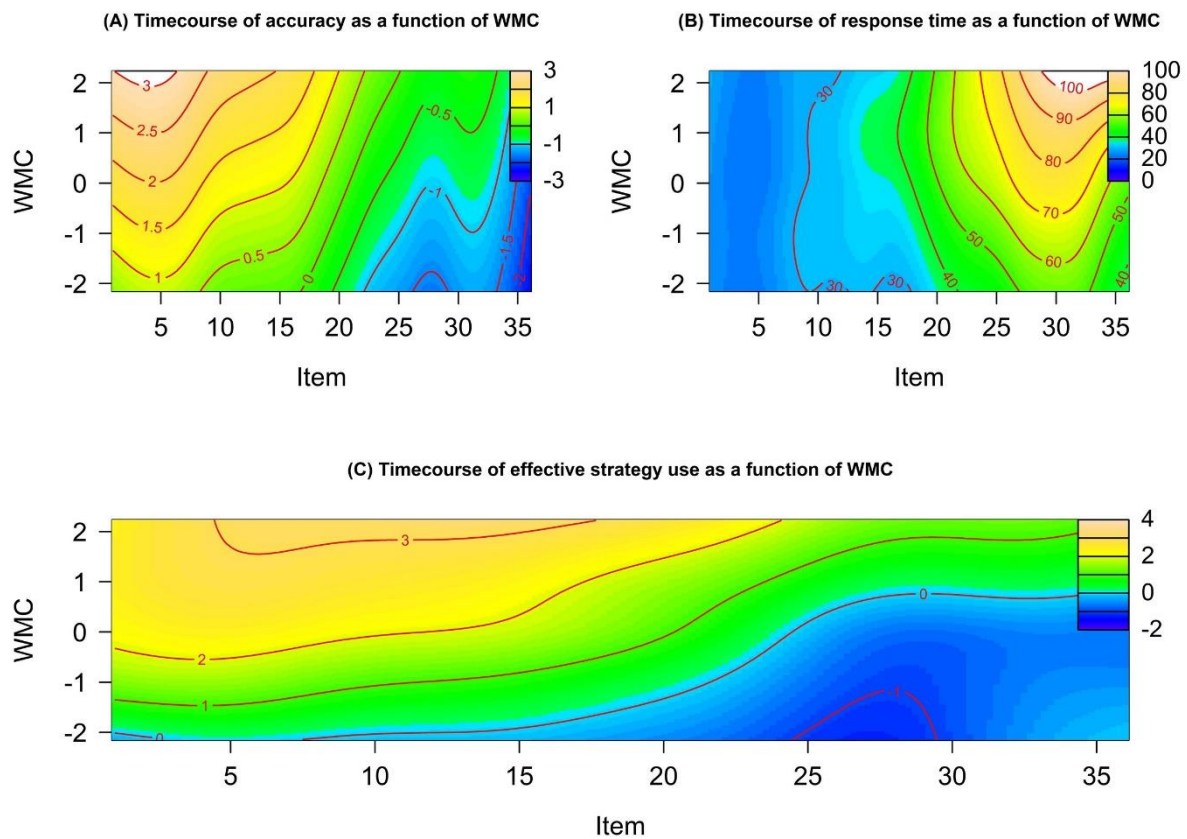


Figure 4. Timecourse of (A) accuracy, (B) response time, and (C) effective strategy use as a function of working memory capacity and item position in the APM. Accuracy is expressed as the logarithm of the odds-ratio of giving a correct answer, with 0 indicating a 50% chance for the answer to be correct; response times are expressed in seconds; effective strategy use is computed as (constructive matching – response elimination). *Example: the predicted level of effective strategy use with average WMC (equal to zero) is about 2 at item 10, and about 1 at item 20. The predicted level of effective strategy use at item 10 is about 2 with average WMC, about 1 for WMC one standard deviation below the mean, and about 0 with WMC two standard deviations below the mean.*

Follow-up analyses examining the constructive matching and response elimination scores separately confirmed these conclusions. For constructive matching, WMC demonstrated both a main effect, $F = 7.74$, $edf = 1.00$, $p = .005$, and a significant interaction with the timecourse, $F = 2.64$, $edf = 6.80$, $p = .004$, indicating that constructive matching was maintained over more items before declining in participants with high WMC (Figure 5A). Response elimination showed the opposite pattern: there was a main effect of WMC, $F = 5.11$, $edf = 1.00$, $p = .024$, and a significant interaction with the timecourse, $F = 3.30$, $edf = 6.08$, $p < .001$, reflecting the fact that participants with high WMC turned to response elimination later in the task – although all participants eventually tended to use response elimination to the same extent (Figure 5B).

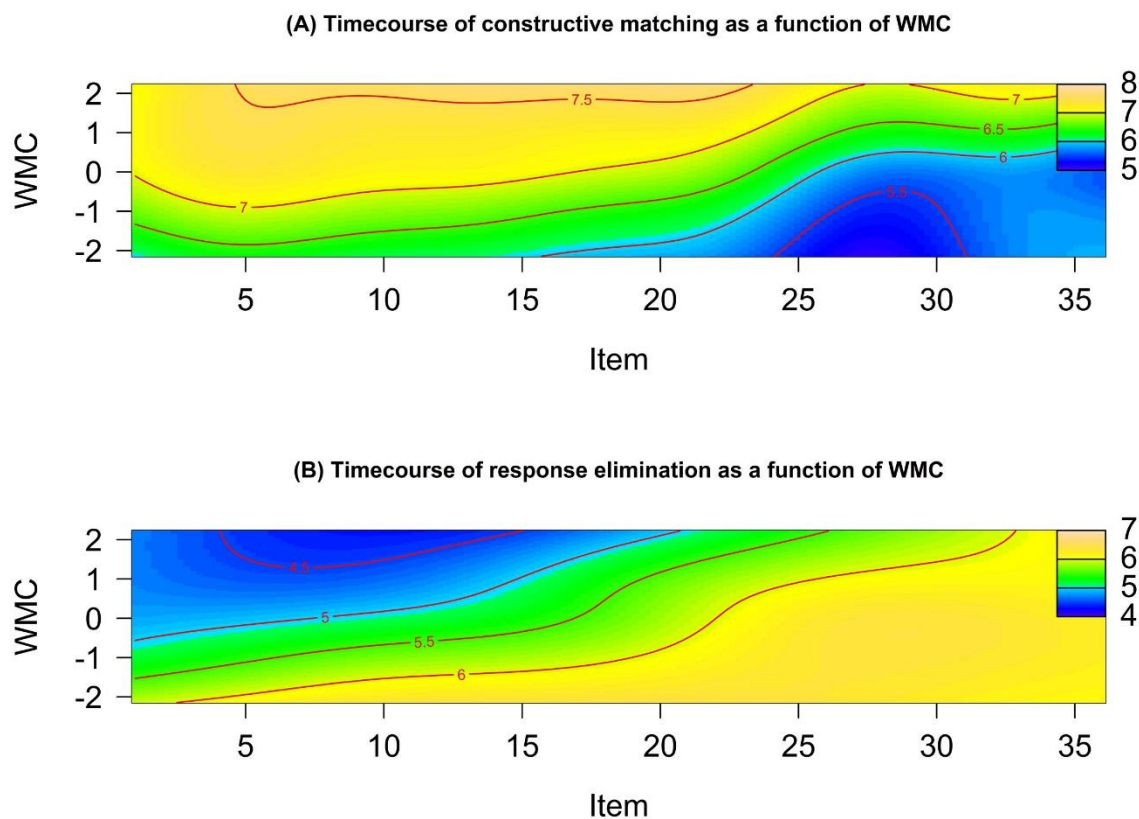


Figure 5. Timecourse of (A) constructive matching and (B) response elimination as a function of working memory capacity and item position in the APM.

The next series of analyses investigated the effect of need for cognition, with the same steps as for working memory. Need for cognition had a significant, linear main effect on accuracy, $\chi^2 = 42.29$, $edf = 1.00$, $p < .001$, such that participants with a high NFC performed higher throughout the task. However, the timecourse of accuracy did not vary significantly as a function of NFC, $\chi^2 = 0.28$, $edf = 1.00$, $p = .596$; this is represented in Figure 6A. As for response times, the effect of NFC was similar to that of working memory. The main effect of NFC did not reach significance, $F = 2.37$, $edf = 1.00$, $p = .124$, but NFC significantly moderated the timecourse of response time, $F = 6.68$, $edf = 5.15$, $p < .001$; as seen in Figure 6B, NFC did not influence response time in the first half of the APM, but participants with high NFC devoted significantly more time to problems in the more difficult second half of the task.

Like working memory, NFC had a significant positive relationship with the effective strategy use index, $F = 26.40$, $edf = 1.00$, $p < .001$; follow-up analyses confirmed that NFC had both a significant positive effect on constructive matching, $F = 28.12$, $edf = 1.00$, $p < .001$, and a significant negative effect on response elimination, $F = 5.81$, $edf = 1.00$, $p = .016$. However, the timecourse of the decline in effective strategy use did not significantly depend on NFC, $F = 1.33$, $edf = 1.87$, $p = .268$, as shown in Figure 6C. In other words, participants with high NFC tended to use comparatively more constructive matching than response elimination, to the same extent throughout the task.

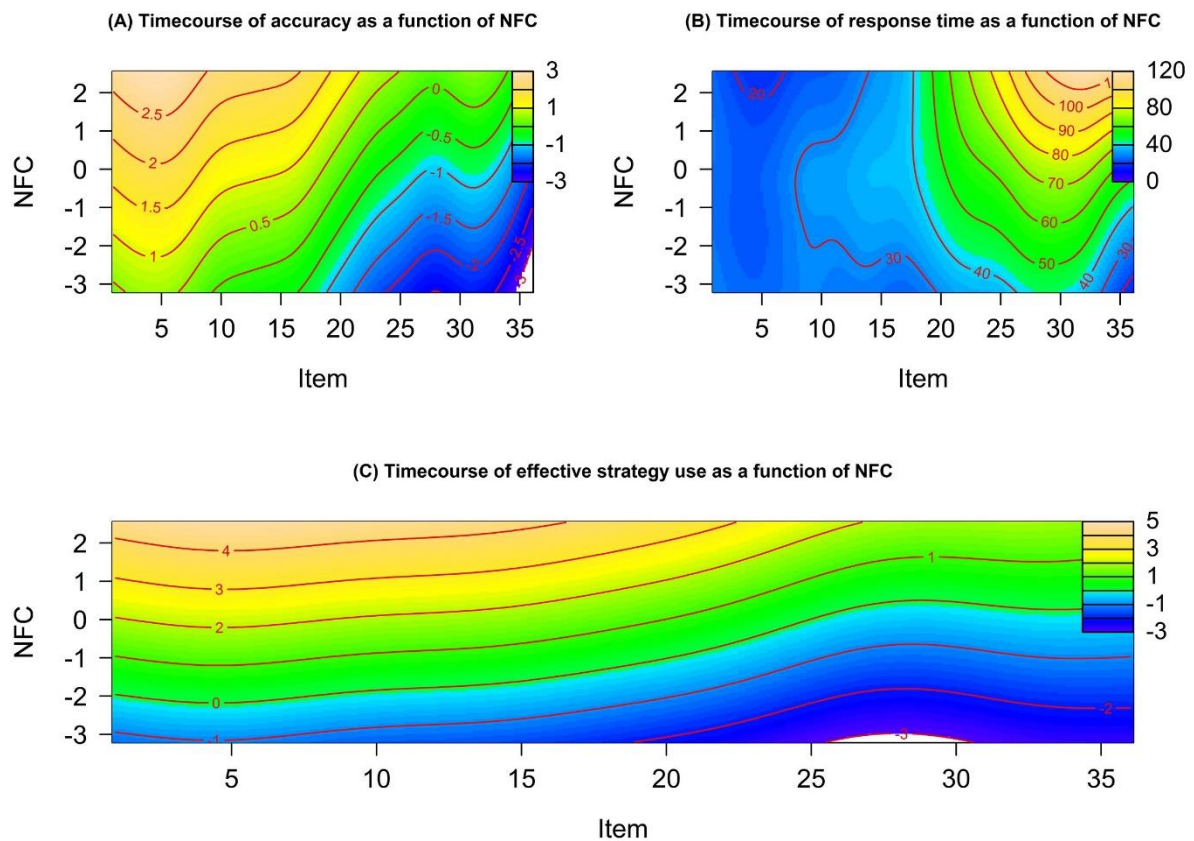


Figure 6. Timecourse of (A) accuracy, (B) response time, and (C) effective strategy use as a function of need for cognition and item position in the APM. Accuracy is expressed as the logarithm of the odds-ratio of giving a correct answer, with 0 indicating a 50% chance for the answer to be correct; response times are expressed in seconds; effective strategy use is computed as (constructive matching – response elimination).

A final series of analyses investigated the combined effects of WMC and NFC on Raven's matrices. For accuracy, both WMC and NFC retained a significant effect when tested within the same model, both $ps < .001$, but the two-way interaction between WMC and NFC did not reach significance, $\chi^2 = 0.66$, $edf = 1.00$, $p = .418$. The combined effects of WMC and NFC over time yielded a complex pattern depicted in Figure 7A. Over all items, the model consistently predicted a high accuracy only for those participants with a combination of high WMC and high NFC. This conclusion is conceptually compatible with the idea that successful performance can be described as the result of motivation and ability (Hill et al., 2016), although our results differed from those obtained by Hill and colleagues (2016), who found a significant interaction between WMC and NFC at the level of total scores in the APM. Replicating the same analysis in our own data also yielded a non-significant interaction, $\beta = .035$, $p = .584$, confirming our GAMM results. This discrepancy may be due to sampling differences or to random chance, given that the interaction found by Hill and colleagues was very weak even with $N = 167$ ($z = 2.01$, $p = .044$).

For response time, the results indicated a complex pattern of three-way interaction between WMC, NFC and item ordinal position, $F = 2.28$, $edf = 9.64$, $p = .009$; this pattern is represented in Figure 7B. Both WMC and NFC had little effect on response times on the earlier items of the APM; as the task progressed, the two predictors interacted in such a way that long response times were only observed at the combination of high WMC and high NFC.

Lastly, for effective strategy use, the analyses confirmed the results obtained when analyzing the two predictors separately, with a main effect of NFC and a moderation of the timecourse by WMC (both $ps < .01$). The three-way interaction between WMC, NFC and item ordinal position was marginally significant, $F = 2.75$, $edf = 2.89$, $p = .053$, although cross-validation showed that this interaction was non-significant in both the Odd and Even conditions (both $ps > .10$). The model consistently predicted a high strategy score only for those

participants with a combination of high WMC and high NFC, as depicted in Figure 7C – similar to the pattern observed for accuracy. Participants with high WMC again showed a slower decline in the rate of effective strategy use; this was especially true for participants who also had high NFC. By the final item of the APM, only participants with both WMC and NFC at least 1.8 standard deviation above the mean demonstrated a strategy score above 2.

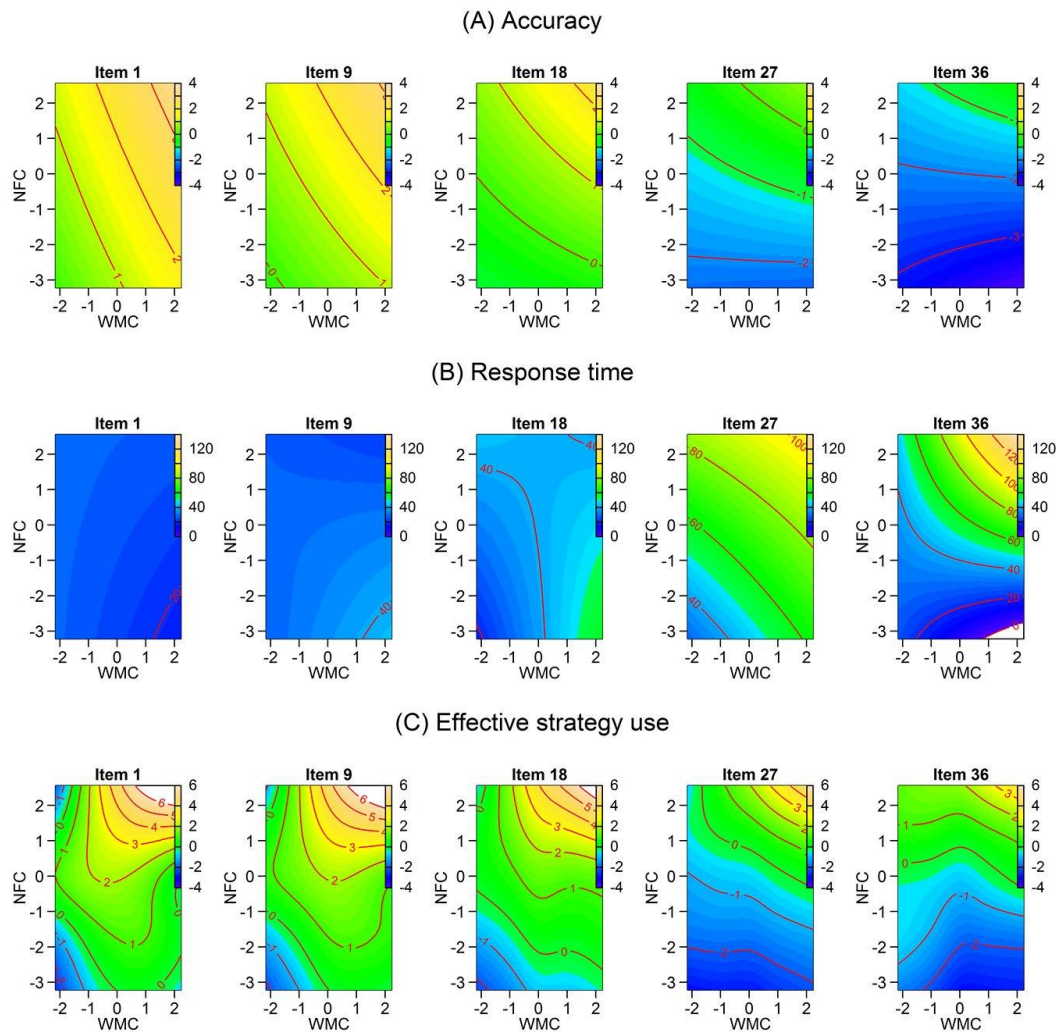


Figure 7. Timecourse of (A) accuracy, (B) response time, and (C) effective strategy use as a function of working memory capacity and need for cognition. Values of each dependent variable are represented horizontally for five evenly-spaced items of the APM. Accuracy is expressed as the logarithm of the odds-ratio of giving a correct answer, with 0 indicating a 50% chance for the answer to be correct; response times are expressed in seconds; effective strategy use is computed as (constructive matching – response elimination).

Discussion

The first objective of this study was to test the possibility of intra-individual strategy shifts in the course of Raven's matrices. On average, the constructive matching strategy dominated in the first two thirds of the task, but its use progressively decreased while the use of response elimination increased, so that response elimination was used to an equal or slightly greater extent over the later part of the task. The present results thus provide the first direct demonstration of a progressive shift away from constructive matching and towards response elimination in Raven's matrices, in line with predictions from the literature (Bethell-Fox et al., 1984; Snow, 1978, 1980).

A complementary analysis also revealed that the decline of effective strategy use was closely related to the decline of performance, with the item-to-item variation in strategy use predicting 78% of variance in accuracy. These findings highlight the importance of systematically investigating intra-individual variability, as a way to understand mechanisms of performance (Borsboom et al., 2009). They show that low accuracy is not exclusively tied to low ability: accuracy progressively declines throughout the task in tandem with the use of constructive matching, so that a lower use of constructive matching on a given item is accompanied by a corresponding drop in accuracy. Limitations of ability may contribute to this progressive decline in effective strategy use throughout the task (Bethell-Fox et al., 1984), or the two may be partly independent. In short, accuracy, ability and strategy use should be considered concurrently.

Our second objective was to investigate possible determinants of this intra-individual shift from constructive matching to response elimination. The results showed that only participants with both high working memory capacity and high need for cognition demonstrated sustained use of constructive matching throughout the task. Both high WMC and high NFC were predictive of greater use of constructive matching and lesser use of response

elimination overall. Working memory capacity additionally predicted the timecourse of strategy use, so that participants with high WMC tended to keep using constructive matching over more items and to shift to response elimination later in the task. In sum, our results confirm that behavior in a reasoning task, and more specifically the use of an effective strategy, depends on a complex interplay between ability and motivation (Hill et al., 2016; Sandra & Otto, 2018); these results also serve to confirm the existence of a meaningful relation between WMC and effective strategy use in Raven's matrices (Gonthier & Thomassin, 2015; Jarosz & Wiley, 2012; Jastrzębski et al., 2018).

The finding that WMC moderated the timecourse of decline in effective strategy use is more specifically in line with prior results suggesting a relationship between WMC and strategic adaptivity (Reder & Schunn, 1999; Schunn & Reder, 1998, 2001), and confirms that intra-individual variability can depend on inter-individual differences. In the context of Raven's matrices, it is especially compatible with the idea that constructive matching is contingent on working memory capacity, particularly for more difficult items (e.g. Bethell-Fox et al., 1984; Mulholland et al., 1980; Pellegrino & Glaser, 1980). Conversely, the non-significant moderation by NFC suggests that the perceived effort required to engage in constructive matching may be relatively constant across item difficulty.

These results must be put into perspective: the subtle non-linear moderation introduced by WMC was statistically significant and relatively stable over cross-validation, but it remained of limited magnitude, as seen in Figure 4C (and as reflected in relatively low decision statistics). Variability in strategy use remained largely dominated by the pattern of progressive decrease throughout the task and, secondarily, by the overall effect of WMC and NFC. Nonetheless, we believe this moderation to be conceptually important in that it contributes to understanding the mechanisms of strategy selection in reasoning tasks. Besides, the low magnitude of this moderation (as well as the non-significant moderation by NFC) may also

partly stem from a limitation of the present study: the use of only two items per problem to measure strategy use, which is accompanied by lower reliability (see also Jastrzębski et al., 2018), potentially decreasing relations with other measures. An alternative solution, which constitutes a possible extension to the present study, would be to investigate strategy use based on a behavioral measure such as eye-tracking; but this would be difficult to do on a scale large enough for a meaningful investigation of intra-individual variability, and prior results suggest that this method might fail to detect strategy shifts (Vigneau et al., 2006).

Modeling Strategy Selection in Raven's Matrices

Our study could seem to imply that the dominant behavior for strategy selection is maladaptive: in other words, that the average individual tends to select a strategy associated with failure, precisely when items become more difficult and require an effective strategy to be correctly solved. Such a conclusion would be at odds with the extensive literature showing that individuals flexibly adapt their strategies as best as possible to face the current problem (Lemaire & Siegler, 1995; Schunn & Reder, 2001; Siegler & Shipley, 1995). However, this apparent ambiguity is resolved when considering performance beyond just accuracy on the task: the present results suggest that maintaining constructive matching over the most difficult items is associated with disproportionately increased response times (see Perret & Dauvier, 2018, for a similar example), and, as illustrated by the interaction with WMC and as hypothesized by the literature, with a high working memory load (e.g. Bethell-Fox et al., 1984; Mulholland et al., 1980; Pellegrino & Glaser, 1980). In this sense, the shift towards response elimination is actually adaptive: rather than just accuracy, it optimizes the tradeoff between invested effort and likelihood of success.

When combined together, these data contribute to piecing together a possible model of strategy selection in reasoning tasks. In the seminal study on constructive matching and response elimination, Snow (1980) proposed an algorithmic model of strategy use including a

"strategy choice" step, but did not specify the details of how this strategy choice is performed by participants. The present results make it clear that selection depends at least on item difficulty, WMC and NFC. Based on prior works (e.g. Lovett & Anderson, 1996; Sandra & Otto, 2018; Verguts, Maris, & De Boeck, 2002) and our results, one could tentatively sketch the following model for Raven's matrices:

The most desirable strategy is dynamically selected for each item, with the desirability of a strategy being a function of (1) its perceived benefit, corresponding to the perceived likelihood of success when using this strategy (Lovett & Anderson, 1996); and (2) its perceived cost, which depends on (2a) the relative difficulty of using this strategy on the current problem, (2b) the participant's ability to implement this strategy, and (2c) the participant's motivation to expand the effort required by this strategy.

In the present study, (2a) was indexed by item ordinal position (i.e. item complexity), (2b) was indexed by working memory capacity, and (2c) by need for cognition. However, this list is not limitative. For example, certain features of a problem may facilitate response elimination (affecting parameter 2a; see Arendasy & Sommer, 2013, for examples), and a high fluid intelligence may facilitate rule induction, increasing the likelihood of engaging in constructive matching (affecting parameter 2b; see Bethell-Fox et al., 1984). Another possible factor would be the introduction of new logical rules on certain items (Wiley et al., 2011), which could not be investigated properly with the current design.

It is also the case that item ordinal position does not only index item complexity in Raven's matrices: later items may also be associated with participant fatigue and boredom, among others (which may affect parameter 2c – especially given that difficult items may be more discouraging when introduced after many other problems). Item order and difficulty are usually confounded in reasoning tasks, which allows participants to progressively learn task rules, but a recent work suggests that item order may be randomized without making the task

impractical (although at the cost of a drop in performance; Zeller, Reiß, & Schweizer, 2017). Determining whether the observed strategy shift is primarily caused by increasing difficulty or disengagement would be a possible extension of the present study.

Note that this proposed model of strategy choice does not require a deliberate, conscious decision by participants. On the contrary, subjects do not usually report explicitly choosing a strategy before each item in Raven's matrices (see e.g. Carpenter et al., 1990, for an example of verbal report). After considerable debate in the strategy literature about whether strategy choices require metacognition and/or awareness, it seems to be the case that metastrategic decisions are possible but not necessary (Cary & Reder, 2002). In the present case, the choice of a strategy through weighting of the determinants listed above could be essentially implicit. For example, the perceived likelihood of success with a given strategy could be based on a simple history of success and failures (Lovett & Anderson, 1996; Reder & Schunn, 1996); likewise, the perceived difficulty of using constructive matching could be just based on the perceptual complexity of the item (Primi, 2001).

Causal Relationships between Strategy Use and Performance

These results could be taken to suggest a direct causal role of strategy use on performance: participants fail more difficult items *because* they stop attempting to mentally reconstruct the answer; in other words, the progressive shift from constructive matching to response elimination contributes to failure in the later part of the task. On the other hand, a reverse interpretation is also possible: participants could stop using constructive matching on difficult problems because they are too complex for them to be able to reconstruct the answer. In other words, decrease in the use of constructive matching would be only a symptom of increasing complexity, not the cause of the corresponding decline in accuracy. In this case, changes in strategy use would be a consequence rather than a cause of limitations in ability.

In fact, the two possibilities could be true simultaneously. This interpretation is, perhaps, the most easily reconciled with the current results, and with divergent findings about the effects of strategy induction (Gonthier & Thomassin, 2015; Mitchum & Kelley, 2010). In this view, strategy selection would depend partly on a subjective choice (driven by factors such as motivation), which could be manipulated to improve performance; and partly on the objective complexity of the item, reflecting actual difficulty in implementing the strategy. Hypothetically, strategy selection and accuracy could even form a causal loop: being confronted with a complex item for which the answer cannot be easily reconstructed encourages participants to default to an easier strategy, which in turn decreases their understanding of the rules and increases the perceived complexity of the problems.

One way to explore the causal status of strategy use in greater depth is to induce participants to use a specific strategy. For example, participants can be induced to exclusively use constructive matching by having them draw the missing part of the matrix (Mitchum & Kelley, 2010; see also Duncan, Chylinski, Mitchell, & Bhandari, 2017). If the decrease in constructive matching is only a by-product of increasing difficulty and failure to deduce the logical rules, then preventing participants from using response elimination would not improve (and could even worsen) their performance throughout the task. Conversely, if decline in effective strategy use contributes to decline in performance, inducing participants to keep using constructive matching over difficult items should enhance their performance, which has important implications for practice. This possibility was partly supported by one study (Gonthier & Thomassin, 2015), but these results would definitely require replication: another study found no increase in performance when inducing constructive matching with a different method (Mitchum & Kelley, 2010).

A related point concerns the causal status of strategies in the relationship between working memory capacity, need for cognition, and reasoning performance. This question is

also left open by the present results. In a prior study, we have argued that strategies constitute a mechanism by which WMC affects performance in Raven's matrices: participants with high WMC perform higher partly *because* they use more constructive matching, as illustrated by the fact that strategy use can be observed to mediate the correlation between WMC and performance (Gonthier & Thomassin, 2015). The same could hold true for NFC. There could also be a form of conditional relationship: participants with high WMC perform higher if they use constructive matching. Of course, both WMC and NFC may exert an additional effect on performance through other mechanisms. For example, a recent study replicated the correlation between WMC and strategy use, but failed to observe a mediation by strategy use in a time-constrained version of the APM (Jastrzębski et al., 2018; see also Becker et al., 2016). This non-replication is only partly informative, as time pressure alters the nature of the task (Chuderski, 2013): it naturally represses strategic variability, discouraging participants from using complex strategies and decreasing the influence of WMC on strategy use (Thomassin et al., 2015; Beilock & DeCaro, 2007), and it may at the same time enhance the correlation between WMC and performance through a shared role of processing speed (Engle & Kane, 2004; Schweizer, Troche, & Rammsayer, 2018). This example illustrates that the relation between WMC and intelligence could be attributed to multiple mechanisms operating in parallel (see also Kovacs & Conway, 2016).

Multiple Strategies and the Unidimensionality of the APM

Intelligence tests like Raven's matrices are designed under the assumption of unidimensionality – with the idea that a subject's total score directly reflects their position on the continuum of a single underlying latent ability. Critically, this rationale breaks down when several different strategies exist. If there are multiple qualitatively different ways to solve a problem, and especially when certain strategies are more effective than others, differences between subjects cannot be summarized by their position on a single Euclidean dimension

(Hunt, 1974; for similar arguments, see DeShon, Chan, & Weissbein, 1995; Partchev & De Boeck, 2012). A subject's performance in the task must then be viewed not as the reflection of a unidimensional ability, but as the result of the combination of their ability and the way in which they process the problems. This issue is illustrated by the fact that two subjects using different strategies can obtain the same total score, despite having different levels of ability and basing themselves on different psychological techniques (Hunt, 1974).

This point has direct consequences for the clinical interpretation of intelligence scores. Failure on a given problem for a given subject should be interpreted as a function of whether it was reached through constructive matching or through response elimination. As an example, consider the case of a subject who fails a problem of medium difficulty, despite using constructive matching and spending a long time trying to reconstruct the correct answer. Clearly this failure does not have the same meaning as that of a subject with low need for cognition who defaulted to response elimination, and tried to solve the same problem based only on perceptual cues. In this situation, the score of the participant using response elimination may well be an underestimate of their actual reasoning ability. This possibility is reminiscent of Spearman's argument that an item measures g only when it is solved analytically (cited by Hunt, 1974). Note that this rationale is also valid when applied to items: the finding that difficult items in Raven's matrices tend to be solved using less constructive matching could explain why these items do not necessarily fit a unidimensional model of ability (Gallini, 1983; van der Ven & Ellis, 2000).

The existence of multiple processes has led some authors to propose treating the test as multidimensional (e.g. DeShon et al., 1995; Partchev & De Boeck, 2012). For the APM, this could mean computing two ability estimates: one for items solved through constructive matching, and another for items solved through response elimination. Although this is possible, it should be recognized that two strategies do not necessarily imply that the test measures two

dimensions. The basic steps of the two strategies are relatively similar (see Snow, 1980) and there is little reason to suspect that they tap into two different abilities (as could be the case, for example, with verbal versus visuo-spatial processing; DeShon, Chan, & Weissbein, 1995). It could just as well be the case that effectiveness using response elimination is not related to a specific ability that could yield reliable individual differences. Furthermore, evidence that subjects tend to use both strategies and to shift flexibly between the two – especially when one becomes ineffective – would rather tend to frame them as two complementary aspects of reasoning in the task.

Moreover, constructive matching and response elimination are not mutually exclusive for a given problem, although this is not always explicit in the literature. The non-perfect negative correlation between constructive matching and response elimination, as observed in this and prior studies (Gonthier et al., 2015; Jastrzębski et al., 2018), hints that participants may combine the two. Participants may well begin with trying to identify some rules in the matrix through constructive matching, then default to response elimination on the same item when this strategy fails; Snow's (1980) model of strategy use explicitly accounted for this by incorporating a loop allowing a participant to come back to the "strategy choice" step after evaluating the result of a given strategy. The very low response times on the most difficult problems indicate that participants do not always attempt to use constructive matching first, but such a combination could be expected for items of intermediate difficulty. Response elimination could also be used to check the result obtained via constructive matching. Lastly, it could be the case that participants try to identify one rule, then proceed to eliminate some response options based on this rule, then try to identify another rule, and so on; this pattern of dynamic shifting between the two strategies within a single problem was reported by Jarosz (2015).

What can we make of this? As described above, one possible way forward would be to remove the possibility of response elimination altogether by having subjects draw the missing answer (Mitchum & Kelley, 2010; Duncan et al., 2017). This removal of strategy-related variance could enhance the predictive ability of the task, but it could also have detrimental effects if there is a meaningful relation between individual differences in ability and strategy selection (see Gonthier & Thomassin, 2015). Another solution in applied settings could be to recognize the inherent complexity of reasoning tasks, to systematically assess strategy use along with accuracy, and to interpret a participant's performance in light of their use of constructive matching – keeping in mind that low constructive matching could be caused by low ability (Bethell-Fox et al., 1984).

Generalizability of the Results

It is likely that the observed progressive shift from constructive matching to response elimination would generalize to samples other than undergraduate students: Bethell-Fox and colleagues (1984) observed the same shift in a sample of high school students (14.5 to 17.5 years old), sampled to be representative in terms of fluid intelligence. However, care has to be taken in generalizing the precise trajectory of the shift reported in our study (with a tipping point around items 24 – 25). Because strategy use depends on levels of cognitive ability and motivation, it seems likely that a sample with lower NFC or lower WMC than university students would demonstrate an even more pronounced shift between the two strategies, with response elimination becoming the dominant strategy earlier in the task, and perhaps with a larger difference between the reported use of the two strategies by the end of the task.

A related question concerns the generalizability of these findings to other tasks. Anecdotal reports by Snow (1978, 1980) suggest that a similar shift could be observed in other visuo-spatial tasks, such as the paper folding test (PFT); Bethell-Fox and colleagues (1984) reported a similar shift towards response elimination in a visual analogy task. Both types of

tasks resemble Raven's matrices in that participants are required to find, among several possibilities, the correct piece to solve a visual problem. The same strategy shift should be found in other tests sharing these features, such as the Cattell culture fair intelligence test (Cattell, 1973). Moreover, the present results and the literature suggest that the basic mechanism – switching to an easier strategy when complexity of the task increases, due to a cost-benefit tradeoff – should exist in other reasoning tasks, even if they allow for different types of strategies; however, this topic remains to be explored. In fact, the types of strategies used in many other reasoning tasks (such as subtests of the Wechsler scales) are almost entirely unknown, not to mention their variability. This could represent a fruitful avenue for further research, leading to deeper understanding of dynamics of performance on intelligence tests.

Context of the Research

This study is part of a larger effort to better understand determinants of performance in intelligence tasks. Our project is especially focused on information processing mechanisms leading up to a participant's response, of which strategy use is a prominent example. In turn, strategic variability has a place in interpretation of intelligent performance in clinical settings, as demonstrated by prior studies in intellectual disability and giftedness. Future projects include the extension of analyses of strategic behavior to tasks other than Raven's matrices.

Acknowledgements

The authors thank Noémie Thomassin for her contributions to this study.

References

- Ackerman, P. L., Beier, M. E., & Boyle, M. O. (2005). Working memory and intelligence: The same or different constructs? *Psychological Bulletin*, *131*(1), 30-60. doi: 10.1037/0033-2909.131.1.30
- Arendasy, M. E., & Sommer, M. (2013). Reducing response elimination strategies enhances the construct validity of figural matrices. *Intelligence*, *41*(4), 234-243. doi:10.1016/j.intell.2013.03.006
- Baayen, H., Vasishth, S., Kliegl, R., & Bates, D. (2017). The cave of shadows: Addressing the human factor with generalized additive mixed models. *Journal Of Memory And Language*, *94*, 206-234. doi:10.1016/j.jml.2016.11.006
- Becker, N. Schmitz, F. Falk, A. M. Feldbrügge, J. Recktenwald, D. R. Wilhelm, O. Preckel, F. & Spinath, F. M. (2016). Preventing response elimination strategies improves the convergent validity of figural matrices. *Journal of Intelligence*, *4*(2). doi:10.3390/jintelligence4010002
- Beilock, S. L., & DeCaro, M. S. (2007). From poor performance to success under stress: Working memory, strategy selection, and mathematical problem solving under pressure. *Journal Of Experimental Psychology: Learning, Memory, And Cognition*, *33*(6), 983-998. doi:10.1037/0278-7393.33.6.983
- Bethell-Fox, C. E., Lohman, D. F., & Snow, R. E. (1984). Adaptive reasoning: Componential and eye movement analysis of geometric analogy performance. *Intelligence*, *8*(3), 205-238. doi: 10.1016/0160-2896(84)90009-6

- Borsboom, D., Cramer, A. J., Kievit, R. A., Scholten, A. Z., & Franić, S. (2009). The end of construct validity. In R. W. Lissitz, R. W. Lissitz (Eds.), *The concept of validity: Revisions, new directions, and applications* (pp. 135-170). Charlotte, NC, US: IAP Information Age Publishing.
- Cacioppo, J. T., Petty, R. E., & Kao, C. F. (1984). The efficient assessment of need for cognition. *Journal of Personality Assessment*, 48(3), 306–307. doi:10.1207/s15327752jpa4803_13
- 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(3), 404-431. doi: 10.1037/0033-295x.97.3.404
- Cary, M., & Reder, L. M. (2002). Metacognition in strategy selection. In P. Chambres, M. Izaute, & P.-J. Marescaux (Eds.), *Metacognition: Process, function and use* (pp. 63–77). Dordrecht: Kluwer Academic Publishers. doi:10.1007/978-1-4615-1099-4_5
- Cattell, R. B. (1973). *Measuring intelligence with the Culture Fair tests*. Champaign, IL: Institute for Personality and Ability Testing.
- Chuderski, A. (2013). When are fluid intelligence and working memory isomorphic and when are they not? *Intelligence*, 41(4), 244-262. doi:10.1016/j.intell.2013.04.003
- Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review*, 12(5), 769-786. doi: 10.3758/BF03196772

- DeShon, R. P., Chan, D., & Weissbein, D. A. (1995). Verbal overshadowing effects on Raven's Advanced Progressive Matrices: Evidence for multidimensional performance determinants. *Intelligence, 21*(2), 135–155. doi:10.1016/0160-2896(95)90023-3
- Duncan, J., Chylinski, D., Mitchell, D. J., & Bhandari, A. (2017). Complexity and compositionality in fluid intelligence. *Proceedings of the National Academy of Sciences of the United States of America, 114*(20), 5295–5299. doi:10.1073/pnas.1621147114
- Fagot, D., Mella, N., Borella, E., Ghisletta, P., Lecerf, T., & De Ribaupierre, A. (2018). Intra-individual variability from a lifespan perspective: A comparison of latency and accuracy measures. *Journal of Intelligence, 6*, 16. doi:10.3390/jintelligence6010016
- Fletcher, J. M., Marks, A. G., & Hine, D. W. (2011). Working memory capacity and cognitive styles in decision-making. *Personality And Individual Differences, 50*(7), 1136-1141. doi:10.1016/j.paid.2011.02.002
- Fletcher, J. M., Marks, A. G., & Hine, D. W. (2012). Latent profile analysis of working memory capacity and thinking styles in adults and adolescents. *Journal of Research in Personality, 46*(1), 40-48. doi:10.1016/j.jrp.2011.11.003
- Foorman, B. R., Sadowski, B. R., & Basen, J. A. (1985). Children's solutions for figural matrices: Developmental differences in strategies and effects of matrix characteristics. *Journal Of Experimental Child Psychology, 39*(1), 107-130. doi:10.1016/0022-0965(85)90032-3
- Gallini, J. K. (1983) A Rasch analysis of Raven item data. *The Journal of Experimental Education, 52*(1), 27-32. doi:10.1080/00220973.1983.11011869

- Goldhammer, F., Naumann, J., & Greiff, S. (2015). More is not always better: The relation between item response and item response time in Raven's matrices. *Journal of Intelligence*, 3, 21-40. doi:10.3390/jintelligence3010021
- Gonthier, C., & Thomassin, N. (2015). Strategy use fully mediates the relationship between working memory capacity and Raven's matrices. *Journal of Experimental Psychology: General*, 144(5), 916-924. doi: 10.1037/xge0000101.
- Gonthier, C., Thomassin, N., & Roulin, J.-L. (2016). The Composite Complex Span : French validation of a short working memory task. *Behavior Research Methods*, 48(1), 233-242. doi:10.3758/s13428-015-0566-3
- Hill, B. D., Foster, J. D., Elliott, E. M., Shelton, J. T., McCain, J., & Gouvier, W. D. (2013). Need for cognition is related to higher general intelligence, fluid intelligence, and crystallized intelligence, but not working memory. *Journal Of Research In Personality*, 47(1), 22-25. doi:10.1016/j.jrp.2012.11.001
- Hill, B. D., Foster, J. D., Sofko, C., Elliott, E. M., & Shelton, J. T. (2016). The interaction of ability and motivation: Average working memory is required for Need for Cognition to positively benefit intelligence and the effect increases with ability. *Personality And Individual Differences*, 98, 225-228. doi:10.1016/j.paid.2016.04.043
- Hunt, E. (1974). Quote the Raven? Nevermore. In L. W. Gregg (Ed.), *Knowledge and cognition*. Oxford, UK: Lawrence Erlbaum.
- Jarosz, A. (2015). *The relationship between working memory capacity and strategy use on tests of general fluid intelligence* (Unpublished doctoral dissertation). University of Illinois at Chicago, Chicago, IL, US.

- Jarosz, A. F., & Wiley, J. (2012). Why does working memory capacity predict RAPM performance? A possible role of distraction. *Intelligence*, *40*, 427–438. <http://dx.doi.org/10.1016/j.intell.2012.06.001>
- Jastrzębski, J., Ciechanowska, I., & Chuderski, A. (2018). The strong link between fluid intelligence and working memory cannot be explained away by strategy use. *Intelligence*, *66*, 44-53. doi:10.1016/j.intell.2017.11.002
- Klayman, J. (1985). Children's decision strategies and their adaptation to task characteristics. *Organizational Behavior And Human Decision Processes*, *35*(2), 179-201. doi:10.1016/0749-5978(85)90034-2
- Kovacs, K., & Conway, A. R. A. (2016). Process overlap theory: A unified account of the general factor of intelligence. *Psychological Inquiry*, *27*(3), 151–177. doi:10.1080/1047840X.2016.1153946
- Kyllonen, P. C., Lohman, D. F., & Woltz, D. J. (1984). Componential modeling of alternative strategies for performing spatial tasks. *Journal Of Educational Psychology*, *76*(6), 1325-1345. doi:10.1037/0022-0663.76.6.1325
- Lawson, M. J., & Kirby, J. R. (1981). Training in information processing algorithms. *British Journal Of Educational Psychology*, *51*(3), 321-335. doi:10.1111/j.2044-8279.1981.tb02490.x
- Le Vigouroux, S., Pavani, J., Dauvier, B., Kop, J., & Congard, A. (2017). Reactive or proactive? Age differences in the use of affective regulation strategies. *Psychology And Aging*, *32*(7), 621-627. doi:10.1037/pag0000197

- Lemaire, P., & Reder, L. (1999). What affects strategy selection in arithmetic? The example of parity and five effects on product verification. *Memory & Cognition*, 27(2), 364-382. doi:10.3758/BF03211420
- Lemaire, P., & Siegler, R. S. (1995). Four aspects of strategic change: Contributions to children's learning of multiplication. *Journal Of Experimental Psychology: General*, 124(1), 83-97. doi:10.1037/0096-3445.124.1.83
- 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). doi:10.3389/fpsyg.2014.00239
- Lovett, M. C., & Anderson, J. R. (1996). History of success and current context in problem solving: Combined influences on operator selection. *Cognitive Psychology*, 31(2), 168-217. doi:10.1006/cogp.1996.0016
- Luchins, A. S. (1942). Mechanization in problem solving: The effect of Einstellung. *Psychological Monographs*, 54(6), i-95. doi:10.1037/h0093502
- McKeown, G. J., & Sneddon, I. (2014). Modeling continuous self-report measures of perceived emotion using generalized additive mixed models. *Psychological Methods*, 19(1), 155-174. doi:10.1037/a0034282
- McVay, J. C., & Kane, M. J. (2012). Drifting from slow to “d'oh!”: Working memory capacity and mind wandering predict extreme reaction times and executive control errors. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 38(3), 525-549. doi:10.1037/a0025896

- Mitchum, A. L., & Kelley, C. M. (2010). Solve the problem first: Constructive solution strategies can influence the accuracy of retrospective confidence judgments. *Journal Of Experimental Psychology: Learning, Memory, And Cognition*, *36*(3), 699-710. doi:10.1037/a0019182
- Mulholland, T. M., Pellegrino, J. W., & Glaser, R. (1980). Components of geometric analogy solution. *Cognitive Psychology*, *12*(2), 252-284. doi:10.1016/0010-0285(80)90011-0
- Nesselroade, J. R., & Ram, N. (2004). Studying intraindividual variability: What we have learned that will help us understand lives in context. *Research In Human Development*, *1*(1-2), 9-29. doi:10.1207/s15427617rhd0101&2_3
- Nezlek, J. B. (2017). A practical guide to understanding reliability in studies of within-person variability. *Journal of Research in Personality*, *69*, 149–155. doi:10.1016/j.jrp.2016.06.020
- Oberauer, K., Schulze, 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*(1), 61-65. doi: 10.1037/0033-2909.131.1.61
- Partchev, I., & De Boeck, P. (2012). Can fast and slow intelligence be differentiated? *Intelligence*, *40*(1), 23–32. doi:10.1016/j.intell.2011.11.002
- Payne, J. W. (1976). Task complexity and contingent processing in decision making: An information search and protocol analysis. *Organizational Behavior & Human Performance*, *16*(2), 366-387. doi:10.1016/0030-5073(76)90022-2

- Pellegrino, J. W., & Glaser, R. (1980). Components of inductive reasoning. In R. E. Snow, P.-A. Federico, & W. E. Montague (Eds.), *Aptitude, learning, and instruction: Cognitive process analyses of aptitude* (Vol. 1, pp. 177-218). Hillsdale, NJ, US: Erlbaum.
- Perret, P., & Dauvier, B. (2018). Children's allocation of study time during the solution of Raven's progressive matrices. *Journal of Intelligence*, 6(1). doi:10.3390/jintelligence6010009
- Petty, R. E., & Cacioppo, J. T. (1986). The elaboration likelihood model of persuasion. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 19, pp. 123-205). New York, NY, US: Academic Press. doi:10.1016/S0065-2601(08)60214-2
- R Core Team (2016). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Ram, N., & Gerstorf, D. (2009). Time-structured and net intraindividual variability: Tools for examining the development of dynamic characteristics and processes. *Psychology and Aging*, 24(4), 778–791. <https://doi-org.distant.bu.univ-rennes2.fr/10.1037/a0017915>
- Raven, J., Raven, J. C., & Court, J. H. (1998). *Raven manual: Section 4, advanced progressive matrices*. Oxford, England: Oxford Psychologists Press.
- Reder, L. M., & Schunn, C. D. (1999). Bringing together the psychometric and strategy worlds: Predicting adaptivity in a dynamic task. In D. Gopher, A. Koriat, D. Gopher, A. Koriat (Eds.), *Attention and performance XVII: Cognitive regulation of performance: Interaction of theory and application* (pp. 315-342). Cambridge, MA, US: The MIT Press.

- Redick, T. S., Broadway, J. M., Meier, M. E., Kuriakose, P. S., Unsworth, N., Kane, M. J., & Engle, R. W. (2012). Measuring working memory capacity with automated complex span tasks. *European Journal of Psychological Assessment, 28*(3), 164-171. doi: 10.1027/1015-5759/a000123
- Ren, X., Wang, T., Altmeyer, M., & Schweizer, K. (2014). A learning-based account of fluid intelligence from the perspective of the position effect. *Learning And Individual Differences, 31*30-35. doi:10.1016/j.lindif.2014.01.002
- Salama-Younes, M. (2011). *Etudes socio-cognitives des besoins fondamentaux: Echelles de mesure et application sociocognitive pour une population d'étudiant de l'université* (Unpublished doctoral dissertation). Université Rennes 2, Rennes, France.
- Salthouse, T. A. (1993). Influence of working memory on adult age differences in matrix reasoning. *British Journal of Psychology, 84*(2), 171-199. doi: 10.1111/j.2044-8295.1993.tb02472.x
- Sandra, D. A., & Otto, A. R. (2018). Cognitive capacity limitations and Need for Cognition differentially predict reward-induced cognitive effort expenditure. *Cognition, 172*, 101-106. doi:10.1016/j.cognition.2017.12.004
- Schunn, C. D., Lovett, M. C., & Reder, L. M. (2001). Awareness and working memory in strategy adaptivity. *Memory & Cognition, 29*(2), 254-266. doi:10.3758/BF03194919
- Schunn, C. D., & Reder, L. M. (1998). Strategy adaptivity and individual differences. In D. L. Medin, D. L. Medin (Eds.), *The psychology of learning and motivation: Advances in research and theory*, Vol. 38 (pp. 115-154). San Diego, CA, US: Academic Press.

- Schunn, C. D., & Reder, L. M. (2001). Another source of individual differences: Strategy adaptivity to changing rates of success. *Journal Of Experimental Psychology: General*, *130*(1), 59-76. doi:10.1037/0096-3445.130.1.59
- Schweizer, K., Troche, S. J., & Rammsayer, T. H. (2018). Does processing speed exert an influence on the special relationship of fluid and general intelligence? *Personality And Individual Differences*, *131*, 57-60. doi:10.1016/j.paid.2018.04.022
- Siegler, R. S. (1994). Cognitive variability: A key to understanding cognitive development. *Current Directions in Psychological Science*, *3*(1), 1-5. doi:10.1111/1467-8721.ep10769817
- Siegler, R. S., & Shipley, C. (1995). Variation, selection, and cognitive change. In T. J. Simon, G. S. Halford, T. J. Simon, & G. S. Halford (Eds.), *Developing cognitive competence: New approaches to process modeling* (pp. 31-76). Hillsdale, NJ, US: Lawrence Erlbaum Associates, Inc.
- Snow, R. E. (1978). Eye fixation and strategy analyses of individual differences in cognitive aptitudes. In A. M. Lesgold, J. W. Pellegrino, S. D. Fokkema, & R. Glaser (Eds.), *Cognitive psychology and instruction* (pp. 299-308). New York: Plenum Press.
- Snow, R. E. (1980). Aptitude processes. In R. E. Snow, P.-A. Federico, & W. E. Montague (Eds.), *Aptitude, learning, and instruction: Cognitive process analyses of aptitude* (Vol. 1, pp. 27-63). Hillsdale, NJ, US: Erlbaum.
- Thomassin, N., Gonthier, C., Guerraz, M., & Roulin, J.-L. (2015). The hard fall effect: high working memory capacity leads to a higher, but less robust short-term memory performance. *Experimental Psychology*, *62*(2), 89-97. doi:10.1027/1618-3169/a000276

- Toplak, M. E., West, R. F., & Stanovich, K. E. (2014). Rational thinking and cognitive sophistication: Development, cognitive abilities, and thinking dispositions. *Developmental Psychology, 50*(4), 1037-1048. doi:10.1037/a0034910
- Unsworth, N., & Engle, R. W. (2005). Working memory capacity and fluid abilities: Examining the correlation between Operation Span and Raven. *Intelligence, 33*(1), 67-81. doi: 10.1016/j.intell.2004.08.003
- Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods, 37*(3), 498-505. doi:10.3758/BF03192720
- 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*(1), 111–122. <http://dx.doi.org/10.1016/j.intell.2009.08.002>
- van der Ven, A. H. G. S., & Ellis, J. L. (2000). A Rasch analysis of Raven's standard progressive matrices. *Personality and Individual Differences, 29*(1), 45–64. doi:10.1016/S0191-8869(99)00177-4
- Verguts, T., Maris, E., & De Boeck, P. (2002). A dynamic model for rule induction tasks. *Journal Of Mathematical Psychology, 46*(4), 455-485. doi:10.1006/jmps.2001.1400
- Verplanken, B. (1993). Need for cognition and external information search: Responses to time pressure during decision-making. *Journal Of Research In Personality, 27*(3), 238-252. doi:10.1006/jrpe.1993.1017

- Vigneau, F., Caissie, A. F., & Bors, D. A. (2006). Eye-movement analysis demonstrates strategic influences on intelligence. *Intelligence, 34*(3), 261-272. doi:10.1016/j.intell.2005.11.003
- von Stumm, S., Hell, B., & Chamorro-Premuzic, T. (2011). The hungry mind: Intellectual curiosity is the third pillar of academic performance. *Perspectives on Psychological Science, 6*(6), 574–588. doi:10.1177/1745691611421204
- Wang, L. P., Hamaker, E., & Bergeman, C. S. (2012). Investigating inter-individual differences in short-term intra-individual variability. *Psychological Methods, 17*(4), 567-581. doi:10.1037/a0029317
- Wiley, J., Jarosz, A. F., Cushen, P. J., & Colflesh, G. J. H. (2011). New rule use drives the relation between working memory capacity and Raven's advanced progressive matrices. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*(1), 256-263. doi: 10.1037/a0021613
- Wood, S. N. (2017) *Generalized Additive Models: An Introduction with R* (2nd edition). Boca Raton, FL, US: Chapman and Hall/CRC.
- Zeller, F., Reiß, S., & Schweizer, K. (2017). Is the item-position effect in achievement measures induced by increasing item difficulty? *Structural Equation Modeling, 24*(5), 745–754. doi:10.1080/10705511.2017.1306706