Validating the relation-monitoring task as a measure of relational integration and predictor of fluid intelligence

Joel E. Bateman¹ · Kate A. Thompson¹ · Damian P. Birney¹

Published online: 19 June 2019
© The Psychonomic Society, Inc. 2019

Abstract

The relation-monitoring task (RMT) has demonstrated a remarkable ability to predict higher-order cognitive abilities such as fluid intelligence, despite its apparent simplicity: It requires no storage over time and no advanced mental manipulation. Instead, the task is theorized to measure relational integration: the process of constructing mental relations between independent elements. Although several studies have established a link between the RMT and fluid intelligence, few studies have investigated the task parameters that contribute to the task's ability to predict higher-order performance. In the present experiment, we manipulated relational complexity and attentional-control demands by varying visual interference and the amount of new information presented on each trial. Even the most basic version of the task (loading primarily on relational integration) explained substantial variance in fluid intelligence, above and beyond the variance already predicted by traditional working memory tasks. We extended prior results by suggesting an incremental effect of attentional-control demands that contributes (but is not imperative) to the RMT's relationship with fluid intelligence. These findings support the relational integration hypothesis, the theory that what fundamentally limits fluid intelligence is the capacity for relational integration.

Keywords Working memory · Reasoning · Fluid intelligence · Attention · Relational integration

The *relation-monitoring task* (RMT) involves monitoring a grid (typically 3×3) of periodically changing stimuli (e.g., words or digits) and detecting relational matches that may appear across rows or columns, according to a predetermined match rule (e.g., three numbers end in the same last digit), before the array is updated with new stimuli. This simple task is hypothesized to load on a capacity for *relational integration* (Chuderski, 2014; Oberauer, Süß, Wilhelm, & Wittman, 2008): the ability to connect multiple elements within working memory (WM). Relational integration is thought to be the cornerstone of higher-order intelligence (Bateman & Birney, 2019; Halford, Wilson, & Philips, 1998; Oberauer et al., 2008), required in well-established measures of fluid intelligence (Gf; Raven, 1989), and forming the premise of analogical reasoning tasks (Sternberg, 1977). Indeed, the RMT has demonstrated a remarkable ability to predict performance on intelligence tasks (Chuderski, 2014; Krumm et al., 2009; Oberauer et al., 2008), despite it involving no explicit (i.e.,

controlled) storage of information over time. This implies that the often-cited link between WM and Gf (Ackerman, Beier, & Boyle, 2005) may inadequately capture the important role of relational integration in Gf. However, because the RMT also involves rapid scanning, it is difficult to rule out theories of attentional control altogether (Engle & Kane, 2004). Although Chuderski's experimental manipulations of visual interference have indicated that attentional control has minimal impact on RMT performance, these results are preliminary, and the theoretical aspects of the task are still largely equivocal.

The purposes of the present report are (a) to replicate previous findings demonstrating that the RMT predicts Gf over and above classic WM tasks and (b) to more comprehensively understand the factors that influence RMT performance and their relationship to Gf. To this end, three theoretically aligned RMT manipulations were developed and implemented. First, we varied the complexity of the relations to be integrated, because the capacity to deal with complexity has been recognized as a core determinant of intellectual function (Birney & Bowman, 2009; Stankov, 2000). Second, the amount of new information present in each trial was manipulated, in an attempt to tease apart the roles of visual scanning and attentional control in the RMT and its relationship with Gf. Finally, to explore the role of inhibition, we manipulated the amount of

✉ Joel E. Bateman
jbat2136@uni.sydney.edu.au

¹ School of Psychology, University of Sydney, Sydney, New South Wales, Australia

visual interference presented in each trial (Chuderski, 2014). In the following sections, we describe the background of the RMT and detail the rationale for these manipulations. Our experiment replicates prior research demonstrating the RMT's remarkable ability to predict Gf and reveals that attentional control does contribute—though is not imperative—to this ability. Instead, it appears that the core demand of the task—the ability to bind multiple elements into an integrated relation—is what is paramount to the relationship with Gf.

The RMT was originally featured in Oberauer, Süß, Wilhelm, and Wittman's (2003) analysis of WM. Participants are presented with a 3×3 array of three-digit strings (see Fig. 1). In the standard version of the task, participants are asked to validate whether there is a row or column in which a particular rule holds (e.g., all digit strings end in the *same* last digit). In Oberauer et al. (2008), a reanalysis of the data revealed strong correlations with latent constructs of intelligence, particularly Gf, typified by Raven's-style (Raven, 1989) abstract-reasoning tasks. Buehner, Krumm, Ziegler, and Pluecken (2006) and Krumm et al. (2009) found similar correlations between the RMT and Gf. The strong overlap between the RMT and Gf is supportive of a theory that we are terming the *relational integration hypothesis*: the theoretical inference that performance in Gf is most fundamentally and ultimately limited by the capacity for relational integration, a sentiment that is being shared by a growing number of researchers (Bateman & Birney, 2019; Bateman, Birney, & Loh, 2017; Chuderski, 2014; Halford et al., 1998; Oberauer et al., 2008). The RMT appears to be an ideal exemplar of the relational integration hypothesis, given its remarkably simple concept and administration, and equally remarkable correlations with Gf.

Given the impressive RMT-related findings, it is perhaps surprising that it was not until Chuderski (2014) that a more formal analysis of the task was conducted, to better understand the basis of these correlations. Chuderski manipulated the complexity of the relations to be considered, by including a five-match condition (i.e., each relation involved binding five elements in the array for comparison, rather than the typical three) and by introducing a *different* match rule. The standard

match rule, up to that point, had required searching for identical stimuli (e.g., all digit strings end in the *same* last digit), whereas the *different* rule involved searching for distinct digits (e.g., all digit strings end in *different* last digits). The five-match condition produced an interaction effect with the *different* condition, such that performance dropped substantially more moving from three- to five-match with the *different* rule than with the *same* rule. Chuderski hypothesized that this was because nonidentical digits could not be chunked in the same way that identical digits could be, leading to a much higher concurrent relational-processing load moving from three to five digits to be integrated. The results of this manipulation strongly suggested that the task was primarily demanding relational integration. This was further supported by Chuderski finding no impact of visual interference (where high interference involved arrays with many identical digits) on task difficulty. Together with the facts that (a) there is typically sufficient time to fully scan the array (over 5 s) and (b) not all stimuli are replaced when the array is updated (we henceforth refer to this feature as *string-preservation*, meaning that some strings are preserved from trial to trial), this indicates that the task does not load heavily on attentional control. Chuderski again found an overall good correlation between the RMT and Gf ($r = \sim .41$), though none of the experimental manipulations appeared to impact the magnitude of this correlation. Thus, it seemed that even the most basic form of the task could produce a valid measure of relational integration and, by extension, fluid intelligence.

In the present study we aimed to replicate and extend our understanding of the features of the RMT that contribute to its success in predicting Gf. The end goal was a clearer appreciation of the importance of relational integration in higher-order cognition. Three theoretically aligned RMT manipulations were investigated—cognitive complexity, attentional control, and inhibition. Below, we consider each in turn.

Cognitive complexity To corroborate Chuderski's (2014) finding on match complexity, we also incorporated the *same* versus *different* manipulation, but added an additional novel *ascending* condition. Before explaining the *ascending*

044	742	426	516	301	005
045	481	330	774	248	926
752	082	252	140	268	444
Match			No match		

Fig. 1 Examples of two arrays with the *same* match rule from the relation-monitoring task. In the “match” example, all three number strings in the bottom row end with the same digit: 2. In the “no-match” example, there are no rows or columns in which all three number strings end in the same digit

manipulation, it is worth reiterating how *same* and *different* manifest in the task, what they convey theoretically, and why *ascending* could help round out the complexity manipulations. In the *same* condition, the match rule is “all strings within a row or column end in the same digit,” whereas in the *different* condition, the match rule is “all strings within a row or column end in different digits.” The *same* condition has a lower theoretical relational complexity (Halford et al., 1998) than the *different* condition, because the first two end-digits in a *same* match [*same*(4,4,4)] can be systematically chunked together [*same*(4,4)], and distinguishing between these first two end digits is not paramount to verifying whether the third end-digit (4) is also part of the relation—we need only know that the first two digits are the same and that both of them are 4. Contrarily, a different match [*different*(5,8,7)] cannot be chunked, because, although together the first two end-digits can form the relation [*different*(5,8)], their unique identities must be kept available in order to verify that the third digit (7) is different from both the 5, as in [*different*(5,7)], and the 8, as in [*different*(8,7)]. Thus, according to the chunking principle in relational-complexity theory (Halford et al., 1998), *different* matches require more complex, ternary relational integration than the binary relational integration involved in a *same* match. If we were to replicate earlier findings (Chuderski, 2014) demonstrating that the *different* match is more difficult than the *same* match, this would provide supporting evidence that the demands of the task might well lie in relational complexity. However, it is also possible that the first two identical digits in the *same* condition are easier to chunk simply because they are identical, meaning that lower-level visual identification strategies could be used for chunking, whereas the *different* condition necessitates higher-order relational integration. To clarify this question, we included the *ascending* match condition: “at least one row or column has strings all ending in consecutively ascending digits.” Theoretically, an *ascending* match should have the same complexity as a *same* match (binary), because the first two digits could be chunked together. For instance, consider the *ascending* relation [*ascending*(2,3,4)]. According to the relational systematicity principle (Halford et al., 1998, p. 808), the relational information between the 2 and 3 can be systematically chunked as [*ascending*(2,3)], because we do not need to know the difference between 2 and 3 in order to integrate the following relation [*ascending*(3,4)]. Rather, we need only know that both separate binary relations are ascending. To reiterate the earlier complexity analysis, this is different from the relation [*different*(5,8,7)] because each element of 5, 8, 7 must be kept available in order to verify that each digit is different from both of the other two digits. Thus, whereas both *ascending* and *different* have three unique elements involved in the relation, the effective complexity is only higher than *same* for *different*, and not for *ascending*. If the task primarily demands relational integration, we should see no

difference in performance between *same* and *ascending*, but *different* should follow the same substantial drop in performance as in prior studies.

Attentional control (scanning) Our second core experimental manipulation concerned scanning demands. In the past, the RMT has always involved string preservation, in which some of the nine strings present in the current array carry over to the next array (Chuderski, 2014; Krumm et al., 2009; Oberauer et al., 2008), reducing the amount of new information presented in each new array. Theoretically, this helps minimize the total attentional-control demands and maximize the relational integration demands. Operationally, the number of new stimuli that must be attended to is reduced, and the primary demand remaining is to rapidly bind the strings into the target (match) relation. Although this task feature is theoretically meaningful, there has yet to be a clear experimental manipulation to determine how much this feature (attentional control) actually contributes to performance and the relation with Gf. Kane, Bleckley, Conway, and Engle (2001) proposed that the ability to actively maintain goal-relevant information in the face of irrelevant information is what connects WM tasks to Gf. Further findings by Kane and Engle (2003) in the Stroop task suggested that poor goal maintenance was a major factor in low-WM participants struggling on WM tasks. Being frequently bombarded with arrays of completely new information in the RMT would require the abilities to rapidly determine which strings are goal-relevant and to efficiently dismiss irrelevant strings, on top of the relational integration demands already present in the task. To test the effects of attentional control through goal maintenance, our experiment included both a *string-preserve* condition (in which some strings persisted between arrays) and a *string-replace* condition (in which all strings were replaced between arrays). The *string-preserve* and the *string-replace* conditions should only differ in their associations with Gf, insofar as the attentional-control demands of the task are related to Gf. In other words, if the *string-replace* condition were to significantly increase the relationship to Gf as compared to *string-preserve*, this would indicate that attentional control is a significant component of Gf. To go one step further, if the task relied on *string-replace* in order to correlate with Gf, this would indicate that the RMT’s ability to predict Gf is driven entirely by attention control, rather than by the core relational integration demands of the task.

Inhibition (visual interference) Our final manipulation was to follow Chuderski’s (2014) work on visual interference, which manipulated the number of identical digits in the array. In each of Chuderski’s arrays, one of the target string-ending digits was duplicated in non-string-ending positions. The theoretical idea was that these identical digits would act as distractors by increasing the similarity of the targets (end-position digits)

and distractors (non-end-position digits), demanding not just relational integration but the ability to inhibit distracting interference in the visual search process. In particular, they should adversely affect the *same* condition more than the *ascending* or *different* conditions, because the identical digits would be crucial to the *same* match but not to either of the other two matches (Chuderski, 2014). Although Chuderski found no impact of interference on mean scores, one potential issue with his implementation of interference was that the high number of distractors (12 out of a possible 24, when excluding the progenitor's string) could actually cue the participant to the target end-position digits, and thus cancel out any detrimental impact of the visual similarity. We explored this potential limitation by including three levels of interference with a similar target–duplication system: No interference involved no duplicated digits, low interference involved six duplicates (a novel condition), and high interference involved 12 duplicates. We predicted that low interference, but *not* high interference, would produce a deficit in performance, because it would cause some visual interference without overtly cueing the participant to the target digits. Inhibition is also related to attentional control (Engle, 1996). Although *inhibition* usually refers to associative activation in long-term memory (Hasher & Zacks, 1988), the explicit visual inhibition caused by visually identical stimuli can also have a strategic component related to task performance, through purposeful avoidance of the allocation of attention to distracting elements (Lu et al., 2017). Thus, a difference between the interference conditions in predicting variance in Gf could still represent the contribution of visual inhibition in the RMT's relationship to Gf.

Aims and hypotheses

The present study was conducted in order to systematically manipulate the task features of the RMT, to determine what makes the task so good at predicting Gf over and above classic “store-and-process” WM measures. We extended on Chuderski's (2014) manipulations by further investigating the roles of attentional-control demands through an elaborated visual interference manipulation and by comparing string preservation with string replacement. We also further manipulated complexity through the addition of the *ascending* match type, to remove the confound of identical digits contributing to lower-order visually oriented chunking. To provide a stronger conclusion for determining what the RMT shares with Gf, we included three Gf measures, so we could form a latent Gf measure. We also included two classic criterion measures of WM: a complex-span and an *n*-back measure. It was predicted that the extent of the relationship with Gf would be largely determined by the capacity of the RMT to measure relational integration (the *relational integration hypothesis*). Specifically, we hypothesized (H1) that the

different condition would increase both the difficulty of the task and the relationship to Gf, as compared to *same* and *ascending*, in that *different* matches would require a higher relational complexity to integrate, whereas *same* and *ascending* would have the same theoretical complexity, and so should produce similar performance. We also hypothesized (H2) that *string-replace* conditions would add an additional unique component predicting Gf over *string-preserve*, in line with the additional attentional-control demands required in dealing with a full array of new strings. However, in line with the relational integration hypothesis, both versions of the task would predict Gf (i.e., *string-replace* would not be necessary for the RMT to predict Gf). Finally, we hypothesized (H3) that *low interference* but not *high interference* would decrease performance on the task, because low interference would visually interfere with participants without overtly cueing them to the target. Again, in line with the relational integration hypothesis, all versions of the task would predict Gf. In summary, in all cases, we predicted that the relational integration demands of the RMT would predict Gf over and above the two criterion WM measures, which primarily measure classic “store-and-process” demands, and that further increases would be concomitant with the respective theoretical demands of the manipulations.

Method

Participants and procedure

A total of 105 participants took part in exchange for course credit. Five of the participants were excluded due to unacceptably low scores on at least one measure, indicating that they did not understand the task instructions or were purposely not engaging in the task.¹ Of the remaining 100 participants, 67 were female and 33 were male, with an average age of 19.47 years ($SD = 2.12$). Participants undertook six tasks: the RMT, three measures of Gf (Advanced Progressive Matrices [APM], letter series, Latin square task), a complex span task (operation span), and an *n*-back task (spatial *n*-back). Participants completed the tasks in a random order in 90-min sessions, in groups of up to eight, in computer labs at the University of Sydney.

¹ Of the five participants excluded, two scored 0 for the Latin square task–dynamic completion, two scored 0 for the letter series, and one scored 1 for the Advanced Progressive Matrices (all of these scores were more than three standard deviations below their respective means, with the distribution plots demonstrating clear outliers). All three tasks included items ranging in difficulty, including particularly easy items that were expected to be trivial for university-level adults.

Measures

Relation-monitoring task

The RMT involved presenting a continuous 3×3 array of three-digit number strings. The task was to respond (with the spacebar) whenever an array matching the current match rule was presented (see Fig. 1). If the array did not match the current rule, the participant was to wait for the next array, which would replace some or all strings (depending on the condition) with new ones. Each array was presented for 5.5 s, with a 100-ms interval.

There were three experimental manipulations, balanced across one another: complexity (same/ascending/different), string preservation (string-preserve/string-replace), and interference (none/low/high). Each manipulation is detailed in the paragraphs below. Participants completed a total of six blocks, with a unique complexity and string-preservation combo: *same-replace*, *same-preserve*, *ascending-replace*, *ascending-preserve*, *different-replace*, and *different-preserve*. Each block had 36 test trials, half of which were matches. The score was derived through the proportion of correct hits on match trials minus the proportion of false alarms on no-match trials (e.g., 15/18 correct matches and 4/18 incorrect false alarms would lead to a score of $[(.83 - .22) = .61]$ for that block). The three levels of the interference condition were balanced within each block.

RMT: Complexity The three match rules (representing complexity) are demonstrated in Fig. 2. The *same* condition involved matches in which three strings in a row or column ended in the same digit. The *different* condition involved matches in which three strings in a row or column ended in different digits. For the new *ascending* condition, a match occurred whenever three strings in a row or column ended in consecutively ascending digits. Participants were given instructions and practice on each match type, including specific instructions on the *ascending* condition that made it clear that the ascending digits must be in consecutive order (from top to bottom for columns, from left to right for rows). A reminder of the current match rule was always present to the left of the array.

RMT: String preservation The string-preservation parameter was manipulated by comparing the score of *preserve* blocks against *replace* blocks, averaged across complexity. In *preserve* trials, one to four strings (at random) persisted from one array to the next—this replicated Chuderski’s (2014) methodology. In *replace* trials, all strings were always replaced with new ones on each new array.

RMT: Interference The final manipulation was interference, with three levels: *Int-0*, *Int-1*, and *Int-2* (corresponding to no, low, and high interference, respectively). These levels are demonstrated in Fig. 3. *Int-0* were regular trials with no duplicated digits. *Int-1* involved one random string-ending digit duplicating six times across the array in non-string-ending positions. *Int-2* was similar, except that the progenitor digit duplicated 12 times. In match trials, the progenitor digit was always one involved in the target match, whereas in nonmatch trials, it was a random string-ending digit. *Int-0* and *Int-2* replicated Chuderski (2014), whereas *Int-1* was a novel addition. Each level of interference was presented an equal number of times per block, such that for each of the 18 matches and each of the 18 nonmatches in a block, there were six *Int-0*, six *Int-1*, and six *Int-2* arrays, distributed randomly in terms of complexity and string preservation.

Raven’s Advanced Progressive Matrices

Participants completed an abbreviated, 20-item version of Raven’s APM (odd items + items 34 and 36 in the original test) as an indication of Gf (Raven, 1989). Participants had 20 min to solve as many items as possible. This 20-item version has shown excellent reliability as a shortened variant of the APM (Bateman & Birney, 2019), because it is sufficient for participants to learn and apply the rules that govern APM items (Bui & Birney, 2014).

Letter series

Participants had 4 min to complete as many of 15 letter series items as they could. Each item involved a patterned sequence of letters followed by an underscore, to indicate that the task

189	749	039	612	363	914	793	654	424
294	265	783	420	843	627	004	598	271
010	148	763	089	156	721	364	868	304
SAME			ASCENDING			DIFFERENT		

Fig. 2 Examples of match arrays for each complexity condition. *Same*: At least one row or column has strings all ending in the same digit (match: top row). *Ascending*: At least one row or column has strings all ending in

consecutively ascending digits (match: top row). *Different*: At least one row or column has strings all ending in different digits (match: middle row)

189	749	039	885	886	428	226	222	680
294	265	783	448	886	791	223	222	226
010	148	763	839	302	847	227	672	328
Interference: 0			Interference: 1			Interference: 2		

Fig. 3 Examples of arrays with the interference manipulation. 0: No digits are replaced. 1: Six digits are replaced by a random string-ending digit. 2: Twelve digits are replaced by a random string-ending digit. Int-1

was to complete the pattern by inserting a single letter at the end of the sequence (Horn & Cattell, 1967). As in the APM, the items become progressively more difficult.

Latin square task

The Latin square task (LST) was developed to assess relational complexity (Birney, Halford, & Andrews, 2006), but it has since seen use as a Gf measure, following strong correlations to classic Gf measures such as APM (Birney, Bowman, Beckmann, & Seah, 2012). In the LST, participants are presented with an incomplete 4×4 matrix, partially filled with four types of shapes (a circle, square, triangle, and cross) and including one target “?” cell. Participants are informed of the one defining rule of the LST: that each row and column may only contain *one* of each of the four shapes from the set. The task is to determine which of the four shapes should be in the marked target cell. Items vary primarily in difficulty through complexity (i.e., how many rows and columns must be considered in order to derive the target cell).

For this implementation of the task, we administered 24 items split evenly by complexity. In addition, half of these items were standard LST items, whereas half were *dynamic completion* (DC) items, in which participants could dynamically fill nontarget cells of the matrix as they solved for the target cell (Bateman et al., 2017). For the present purposes, it is only necessary to know that both standard and DC items load on relational complexity and show strong correlations to classic Gf measures.

Operation span

Participants completed the operation span (OSPAN) task with set sizes of three, four, five, and six (two sets of each). In each set, participants alternated between memorizing a letter and verifying the truth of a mathematical operation. Once all letters for that set had been presented, participants attempted to recall the letters in the order they had been presented. Scores were calculated as the total number of correct letters recalled (*OSPAN letters*) rather than the number of correct letters in fully recalled sets

shows a low level of visual overlap, caused by the duplicated 8s, whereas Int-2 shows a high level of visual overlap, caused by the duplicated 2s

(*OSPAN capacity*). The partial scoring of OSPAN letters was preferred because it accounts for the same variance picked up by absolute scoring with OSPAN capacity, but it also accounts for additional variance that would otherwise be discarded (Redick et al., 2012).

Spatial n-back

Participants completed a spatial version of the *n*-back task, with two blocks of two-back and two blocks of three-back trials. In each block, participants were presented with a 3×3 cell matrix. Every 2 s, a blue square would flash for 1 s inside a random cell. The participant's task was to respond whenever a blue square appeared that was on the same cell as a blue square from *n* steps back (e.g., two squares back in the two-back condition). The score was derived as the number of hits minus the number of false alarms, then averaged across the four blocks.

Results

RMT manipulations: Performance effects

Descriptives are presented in Tables 1 and 2. The six RMT blocks demonstrated acceptable internal consistency, $\alpha = .79$, despite differences in the match complexity of the conditions, which a repeated measures analysis of variance (ANOVA) determined to be significant, $F(2, 198) = 222.11, p < .001$. Two planned contrasts revealed that the matches with lower relational complexity (*same* and *ascending*) had higher performance than the matches with higher complexity (*different*), $F(1, 99) = 236.90, p < .001, \eta_p^2 = .71$, and the *same* condition also had higher performance than the *ascending* condition, $F(1, 99) = 201.93, p < .001, \eta_p^2 = .67$.

For interference (digits duplicated), we observed no main effect on performance in a repeated measures ANOVA, $F(2, 182) = 1.24, p > .05$, indicating that neither low nor high interference decreased performance relative to no interference.

Table 1. Descriptives ($n = 100$)

Measure	Mean	SD
Raven's APM	.61	.19
Letter Series	.69	.12
OSPAN (Letters)	.83	.18
<i>N</i> -back DV*	2.42	1.64
Latin Square Task	.82	.16
LST-Basic	.80	.18
LST-DC	.83	.18
RMT Grand Total	.67	.13
RMT Same Total	.87	.11
RMT Same	.86	.14
RMT Same-P	.87	.11
RMT Asc Total	.64	.18
RMT Asc	.62	.20
RMT Asc-P	.65	.21
RMT Diff Total	.50	.20
RMT Diff	.50	.22
RMT Diff-P	.51	.24

**N*-back DV reflects the raw score (block average hits minus false alarms), since task was not itemized. RMT conditions marked with “-P” indicate the *preserve* variant, whereas those without the suffix indicate the *replace* variant

RMT prediction of Gf: Controlling for WM

The purpose of this set of analyses was to verify that the RMT correlated with Gf over and above the two criterion WM measures, complex span and *n*-back. A Gf factor was derived through principal axis factoring with varimax rotation on the three Gf measures: APM ($\alpha = .80$), letter

series ($\alpha = .75$), and LST ($\alpha = .76$).² This factor accounted for 65% of the variance in the three measures, with an eigenvalue of 1.95 and factor loadings of .774 for LST, .691 for APM, and .602 for letter series.

As is demonstrated in Fig. 4, the RMT had a considerable $r = .61$ with Gf—this is in comparison with $r = .41$, as reported by Chuderski (2014). As can be seen in Table 3, the *n*-back also correlated with Gf ($r = .53$), but the OSPAN did not. As in past research (Redick & Lindsey, 2013), the *n*-back and OSPAN also did not correlate with each other ($r = -.02$). Table 4 provides the full correlation matrix, separating RMT conditions and tasks. Contrary to Redick et al.'s (2012) recommendation, using OSPAN capacity rather than OSPAN letters generally increased the OSPAN's correlations across the board, though it remained the weakest predictor of Gf ($r = .25$). As per Redick et al.'s suggestion, the following regression analyses will continue to use OSPAN letters. However, regardless of whether OSPAN letters or OSPAN capacity is used as the predictor, the outcomes of the analyses do not change.

More important for our research question was to determine whether the RMT's relationship to Gf was demanding processes similar to those required by the classic WM measures (*n*-back and OSPAN), or whether it was indeed contributing its predicted variance over and above these typical WM measures. We conducted a multiple linear regression predicting the Gf factor, with the first model containing the two classic WM measures and the second adding the RMT. As can be seen in Table 5, the classic WM measures predicted a considerable 31% of the variance in Gf, mainly driven by the *n*-back ($sr^2 = .28$, $p < .001$). The OSPAN also predicted a significant, though small, unique portion ($sr^2 = .03$, $p < .05$). Importantly, once we added in the RMT, the predicted variance increased to 48%, a significant change, $\Delta R^2 = .166$, $p < .001$. With the RMT in the model, the OSPAN now provided nothing unique, with its predicted variance of Gf being subsumed by either the *n*-back or RMT. The *n*-back maintained some unique predictive variance of Gf, $sr^2 = .09$, $p < .001$, though the RMT had the highest unique component predicting Gf, $sr^2 = .17$, $p < .001$.

Although the LST has been used as a Gf measure (Birney et al., 2012), it was primarily designed to tap relational integration (Birney et al., 2006). Thus, it is possible that the strong relationship between the RMT and our Gf factor³ is primarily a result of the LST being included in the Gf factor. To demonstrate that the relationship still holds even without the LST, we reconducted the prior regression, this time predicting the common factor formed only from APM and letter series, using the same extraction

Table 2. Descriptives for RMT (split by interference conditions)

RMT Condition	Mean	SD
RMT Grand Total	.67	.13
Interference 0	.65	.16
Same Int 0	.84	.16
Asc Int 0	.61	.22
Diff Int 0	.50	.25
Interference 1	.67	.14
Same Int 1	.87	.12
Asc Int 1	.63	.20
Diff Int 1	.51	.23
Interference 2	.66	.14
Same Int 2	.87	.11
Asc Int 2	.63	.22
Diff Int 2	.46	.24

² The LST reliability here is derived through three complexity subscales (2D/3D/4D) averaged across the basic and DC LST variants. This produces a lower-bound estimate of the total scale α but is comparable to the LST total $\alpha = .79$, reported by Birney et al. (2012) in a population of managers.

³ To the best of our knowledge, this is the highest correlation between the RMT and Gf yet observed in published research.

Fig. 4 Scatterplot with RMT total score (raw, out of 108) on the *x*-axis and Gf factor on the *y*-axis

method. This two-task Gf factor accounted for 70.9% of the variance in the two component measures, with an eigenvalue of 1.42. The results were largely unchanged from the two-task Gf regression (Model 1 $R^2 = .334$, Model 2 $R^2 = .460$). The only substantial change was that the OSPAN remained a significant unique predictor in the second model ($sr^2 = .03$, $p = .02$), though it was still the lowest of the three tasks (RMT $sr^2 = .15$, $p < .001$; *n*-back $sr^2 = .08$, $p = .001$). Thus, the strong relationship between the RMT and Gf observed here does not appear to be inflated simply by the inclusion of the LST in the Gf factor. Given the largely identical outcomes between the two-task and three-task Gf factors, we proceed with the remaining analyses using the three-task Gf factor.

RMT prediction of Gf: Experimental manipulations

The first regression analysis made it clear that the RMT does indeed have an impressive relationship to Gf, accounting for 16.6% of Gf variability over and above the classic WM measures. Although this is substantially higher than Chuderski's (2014) finding of 5.9%, it should be noted that he included additional WM measures. Our next regressions (which are the novel component of our experiment) aimed to uncover the parameters involved in the RMT that are substantive to this relationship. These include complexity (match type), inhibition (interference), and attentional control (string preservation).

Table 3. Correlation between RMT, WM measures, and Gf factor

	RMT	OSPAN	<i>N</i> -Back
Gf Factor	.61**	.17	.53**
RMT Total	—	.13	.44**
OSPAN (Letters)		—	-.02

For match complexity, we regressed (in order) *same*, then *ascending*, then *different*. The first model, containing just *same*, accounted for 24% of the variance in Gf, $R^2 = .244$, $p < .001$. Adding *ascending* increased this to 33%, $\Delta R^2 = .086$, $p = .001$. Adding *different* then further increased this to 38%, $\Delta R^2 = .044$, $p = .012$. In this final model, all three predictors made small but significant unique contributions (*same*, $sr^2 = .04$, $p < .05$; *ascending*, $sr^2 = .04$, $p < .05$; *different*, $sr^2 = .04$, $p < .05$), while still leaving the majority, $R^2 = .26$, as shared variance.

For interference, we conducted similar analyses, iterating on the regression model as the task increased in interference. It is worth reiterating that no mean differences were found between the interference conditions. The following results are thus particularly interesting. The first model, including only no-interference trials, accounted for 25% of the variance in Gf, $R^2 = .249$, $p < .001$. The second model added low-interference trials and increased the explained variance in Gf to 49%, $\Delta R^2 = .240$, $p < .001$. However, the third model, adding high-interference trials, did not increase the variance explained in Gf significantly, $\Delta R^2 = .006$, $p > .05$. In the final model, only low interference provided a unique contribution, $sr^2 = .21$, $p < .001$.

Our final regression model considered the string-preservation parameter. Again, it is worth keeping in mind that string preservation also had no impact on the mean scores. The first model consisted solely of *string-preserve* trials (which theoretically minimized attentional-control demands) and accounted for a significant 26% of the variance in Gf, $R^2 = .259$, $p < .001$. Adding the *string-replace* trials (which theoretically translated to higher attentional demands) increased this accounted-for variance to 39%, $\Delta R^2 = .13$, $p < .001$. In the final model, only *string-replace* trials had a significant unique contribution, $sr^2 = .13$, $p < .001$.

Table 4. Full task-level correlation matrix

	RMT (T)	RMT-Pres	RMT-Repl	RMT-Same	RMT-Asc	RMT-Diff	LST	APM	L-Series	N-Back	OSPAN
RMT (Total)	—										
RMT-Preserve	.93**	—									
RMT-Replace	.93**	.73**	—								
RMT-Same	.75**	.65**	.74**	—							
RMT-Ascending	.81**	.78**	.74**	.47**	—						
RMT-Different	.86**	.80**	.79**	.54**	.47**	—					
LST	.51**	.51**	.43**	.39**	.43**	.42**	—				
APM	.47**	.51**	.36**	.35**	.37**	.41**	.54**	—			
L-Series	.53**	.51**	.47**	.50**	.45**	.38**	.47**	.42**	—		
N-back	.44**	.44**	.37**	.50**	.26**	.38**	.47**	.32**	.51**	—	
OSPAN (Letters)	.13	.18	.06	.04	.09	.15	.06	.24*	.15	-.02	—

* Significant at $< .05$; ** Significant at $< .01$. RMT = relation-monitoring task, LST = Latin square task, APM = Raven's Advanced Progressive Matrices, L-Series = letter series, OSPAN = operation span

Discussion

The aim of the present study was to experimentally manipulate the relation-monitoring task as a measure of relational integration by demanding different levels of relational complexity, attentional control, and inhibition, to determine what task features are essential for the task to produce its impressive prediction of Gf. Overall, our results were consistent with prior research demonstrating a strong relationship between the RMT and Gf (Chuderski, 2014; Krumm et al., 2009; Oberauer et al., 2008). In fact, our RMT showed an even stronger correlation ($r = .61$) than prior findings had (in the $r = .3$ – $.5$ range). It is worth reiterating how remarkable such a powerful relationship is in individual differences research (Cohen, 1988; Gignac & Szodorai, 2016), particularly when considering the apparent simplicity of the RMT, which requires no explicit storage over time or advanced mental manipulation. This seemingly simple task can predict as much as 37% of variance in a latent Gf factor composed of advanced, abstract series completion tasks such as Raven's, letter series, and the LST. Theorizing surrounding the RMT seems to indicate that this result simply comes about due to the purity of the task in measuring a most fundamental aspect of WM:

relational integration (Bateman et al., 2017; Halford et al., 1998; Oberauer, 2009). Our novel experimental manipulations illustrated that—in line with the relational integration hypothesis—all versions of the task could predict Gf. Although the majority of the predicted variance in Gf was shared among the different RMT conditions, we did identify potential further components of the RMT that appear to be related to increases in relational complexity and additional attentional control and inhibition demands.

Our three RMT match conditions (*same*, *ascending*, and *different*) appeared to tap similar demands in WM—which is theorized to be relational integration—a conclusion emerging with high reliability and with a large amount of shared variance in Gf accounted for between the complexity conditions (the match regression findings suggested that over two-thirds of the variance was shared between conditions). Beyond this shared variance, and contrary to expectations, all three levels of complexity provided something unique in predicting Gf. This was consistent with the mean differences (in that each match was more difficult than the last), but our hypothesis was that *ascending* would offer nothing unique in predicting Gf over and above *same*, because both match rules have the same theoretical binary complexity (i.e., the first two elements in the series can be systematically chunked), unlike in a *different* match. Although *same* and *different* did indeed each have independent components related to Gf, so did *ascending*. This indicates that some unique demand might be related to the ability to apply systematicity (Halford et al., 1998) to the relational integration process for elements that are different in appearance but can be systematically chunked down. For instance, once the relation between the digits 4 and 5 has been verified as ascending, they can be systematically chunked down into a single binding (4,5), because an ascending relation only requires knowing that the next digit 6 follows in the

Table 5. Multiple linear regression with the two classic WM measures predicting Gf (Model 1) then adding RMT (Model 2)

Model	Measure	β	sr^2	R^2	ΔR^2
1: Classic WM measures	OSPAN (letters)	.185	.03*	.310	.310**
	N-back	.529	.28**		
2: Add RMT	OSPAN (letters)	.122	.01	.476	.166**
	N-back	.327	.09**		
	RMT	.459	.17**		

order. We hypothesized that *ascending* would require no additional demand over *same* because they would both require sequential instances of binary relational integration (as opposed to *different*, which requires ternary relational integration). Although this still might be the case, our results indicated that the added challenge of applying systematic chunking to two visually distinct digits (4, 5) might constitute a demand related to both RMT performance and Gf. It is also possible that this unique *ascending* demand came about through a restriction on scanning: Because the ascending matches were always consecutive and sequential, the matches were most easily checked by scanning from left to right and top to bottom. Although they could be scanned in the opposite directions, this would require reversing the match to be checked, to *descending*. Conversely, for *same* and *different* matches, participants could scan from right to left or bottom to top, with only the one rule.

For the interference manipulation, again, the majority of the variance explained in Gf was shared among the three conditions (Int-0, Int-1, Int-2). The interference levels were virtually indistinguishable on a mean difference level; however, low interference (Int-1) provided a considerable unique component in predicting Gf, which was hypothesized to be the demand of dealing with the additional attentional interference of multiple duplicated digits. Our actual hypothesis related to mean differences, in that high interference (Int-2) might have cued participants to the target match, whereas low interference represented a “sweet spot” of interfering but not cueing. This sweet spot still did not make an apparent difference in task difficulty (Chuderski, 2014), but it might have tapped a unique demand independent from relational integration. Although such a demand would exist independently of the relational integration hypothesis, it could be explained by a visual search strategy in which participants purposely allocated no attention to potential distractors (Lu et al., 2017)—the non-string-ending digits. The finding that this strategy could also relate uniquely to Gf is preliminary but plausible, given that tasks such as Raven’s often involve many distinct visual elements, which must be considered independently across rows and columns (Verguts & De Boeck, 2002) and then ruled out as irrelevant (i.e., inhibited) or maintained for further consideration, as appropriate (Carpenter, Just, & Shell, 1990).

Our final manipulation, string preservation, was perhaps the most important. It is a parameter often taken for granted, yet one with potentially critical implications concerning the role of attentional control in the RMT. Prior work with the RMT had included string preservation (Chuderski, 2014; Krumm et al., 2009) in order to minimize the amount of new scanning required, thus maximizing the relational integration demands while minimizing the attentional-control demands. Our results indicated that string preservation (like interference) had no impact on overall task performance but did significantly change the relationship with Gf. That is, *string-*

replace trials offered a unique contribution that substantially increased the relationship to Gf (accounting for exactly one-third of the variance in the RMT, when compared to *string-preserve* trials). This means that, in line with the relational integration hypothesis, the task functions perfectly well as a relatively pure predictor of Gf with string preservation, but the relationship to Gf can be enhanced further by adding the incremental demands reflected in string replacement, in which rapid, flexible binding and unbinding is relevant.

It is also worth reiterating that the RMT surpassed the classic WM measures, predicting substantial variance over and above the unique and shared variation accounted for by complex span and *n*-back. It has frequently been theorized (Krumm et al., 2009; Oberauer et al., 2008) that this is because the RMT taps a fundamental aspect of WM: relational integration, which is also captured—albeit impurely—in these traditional WM measures (Oberauer et al., 2008), which might instead more strongly reflect the passive-storage or updating components of WM (Bateman & Birney, 2019). Like Chuderski (2014), we contributed further evidence to the relational integration hypothesis—the suggestion that Gf can be most fundamentally captured by measuring relational integration—by finding that all experimental variations of the RMT tapped similar demands, consistent with the ability to rapidly establish bindings between independent elements.

Some suggestions should be considered for future research. On the topic of string preservation, there is scope to further assess its impact on task demands. In our study, we replicated Chuderski’s (2014) methodology of preserving one to four strings at random and contrasting this to our novel manipulation of replacing all the strings (i.e., preserving none of them). However, in Oberauer et al. (2008) and Krumm et al. (2009), only a single string was preserved between trials, but the task updated at a faster rate (every 2 s). In contrast to this eight-string preservation, our manipulation seems minor, yet it still made a significant unique contribution in the relationship with Gf after controlling for WM, with one- to four-string preservation accounting for about one-third of the effect. That a seemingly minor manipulation had such an impact indicates that comparing a wider range of string preservation values (i.e., up to eight strings, rather than four) could elucidate a further substantive demarcation of relational integration and attentional-control demands. It is possible that an increase in the strings preserved (up to eight) might further minimize attentional-control demands, but the 2-s response window might counteract this. A future task analysis could thus consider both string preservation and response window independently.

It should be cautioned that although our experimental manipulations played out almost exactly as hypothesized (with regard to predicting Gf), the nature of regression analyses means that incrementing predictors based on the same construct (such as the various RMT conditions, all reflecting

relational integration) inevitably leads to a more reliable prediction of the dependent variable (Gf). Each subsequent predictor might capture some portion of Gf that prior predictors had failed to capture due to random noise or measurement error. Thus, although the pattern of unique contributions is aligned with the additional theorized demands (such as attentional control), we cannot be completely certain. This does not, however, take away from the remarkably strong overall correlation between the RMT and Gf.

Conclusion

In this experiment, we found encouraging results for the RMT as an assessment of relational integration and predictor of fluid intelligence. Theoretically, the RMT is a task demanding relational integration, but functionally it appears to be a powerful, reliable predictor of Gf. This is perhaps the most important implication of our results. Our battery of abbreviated Gf tasks took approximately 60–75 min to administer. A full battery typical of recruitment assessment can take 4–8 h (Chuderski, 2014), or even several days (Robertson, Gratton, & Sharpley, 1987). Yet the RMT, which takes only about 20 min to administer, predicts as much as 37% of variance in Gf—a correlation so high that it is only seen in 2%–3% of individual differences studies (Gignac & Szodorai, 2016).

In summary, we replicated prior research demonstrating a powerful relationship between the RMT and more theoretically complex Gf measures. The RMT is an insightful task because it requires no explicit storage over time and no advanced mental manipulation, instead primarily measuring relational integration. We continued Chuderski's (2014) breakdown of the task, supporting the notion that the task is a measure of the ability to rapidly establish bindings between multiple elements for relational integration. For the first time, we have also demonstrated that the task appears to have some attentional control demands associated with it, though these are not crucial to its relationship with Gf. Our results are thus strong evidence for the relational integration hypothesis (Bateman & Birney, 2019; Bateman et al., 2017; Chuderski, 2014; Halford et al., 1998; Oberauer et al., 2008) but may also coincide with a more attentionally oriented perspective (Kane et al. 2001; Shipstead, Harrison, & Engle, 2016). Our findings support both theories but suggest that each may serve a different purpose in the prediction of Gf. Maintaining focus during a complex task (such as Raven's) and orienting attention toward the goals of the item are helpful but represent a fundamentally different demand to the crucial ability to integrate a relation by binding elements in a mental workspace such as working memory. Ultimately, no matter how focused and well-oriented one is to the goals of the task, the capacity for relational integration can prove to be a cognitive obstacle only overcome with the capacity to strategically and systematically chunk

(Halford et al., 1998). Abstract reasoning and Gf are certainly complex constructs, with prototypical tasks that tap a wide range of theoretically elusive cognitive demands. Latent-variable analysis is the current gold standard for unravelling this constellation of demands, but theoretically driven experimental manipulations are key to determining what cognitive demands are most essential for interindividual variation. The importance of Gf tasks in applied settings such as recruitment and aptitude highlights needs both to understand these cognitive demands and to consider how we can assess them in a way that is both cost- and time-effective. The RMT appears to be a task that can answer the theoretical questions on the source of demands and can provide a pragmatic substitute for large-scale Gf batteries.

Author note The data and materials for this study can be accessed by contacting the corresponding author. None of the experiments were preregistered.

References

- Ackerman, P. L., Beier, M. E., & Boyle, M. O. (2005). Working memory and intelligence: The same or different constructs? *Psychological Bulletin*, 131, 30–60. <https://doi.org/10.1037/0033-2909.131.1.30>
- Bateman, J. E., & Birney, D. P. (2019). *The link between working memory and fluid intelligence is dependent on flexible bindings, not passive or systematic retention*. Manuscript under revision.
- Bateman, J. E., Birney, D. P., & Loh, V. (2017). Exploring functions of working memory related to fluid intelligence: Coordination, relational integration, and access. Paper presented at the 39th Annual Meeting of the Cognitive Science Society, London.
- Birney, D. P., & Bowman, D. B. (2009). An experimental–differential investigation of cognitive complexity. *Psychology Science Quarterly*, 51, 449–469.
- Birney, D. P., Bowman, D. B., Beckmann, J. F., & Seah, Y. Z. (2012). Assessment of processing capacity: Reasoning in Latin Square Tasks in a population of managers. *European Journal of Psychological Assessment*, 28, 216–226.
- Birney, D. P., Halford, G. S., & Andrews, G. (2006). Measuring the influence of complexity on relational reasoning: The development of the Latin Square Task. *Educational and Psychological Measurement*, 66, 146–171.
- Buehner, M., Krumm, S., Ziegler, M., & Pluecken, T. (2006). Cognitive abilities and their interplay: Reasoning, crystallized intelligence, working memory components, and sustained attention. *Journal of Individual Differences*, 27, 57–72.
- Bui, M., & Birney, D. P. (2014). Learning and individual differences in Gf processes and Raven's. *Learning and Individual Differences*, 32, 104–113.
- Carpenter, P. A., Just, M. A., & Shell, P. (1990). What one intelligence test measures: A theoretical account of the processing in the Raven Progressive Matrices Test. *Psychological Review*, 97, 404–431. <https://doi.org/10.1037/0033-295X.97.3.404>
- Chuderski, A. (2014). The relational integration task explains fluid reasoning above and beyond other working memory tasks. *Memory & Cognition*, 42, 448–463. <https://doi.org/10.3758/s13421-013-0366-x>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd). Hillsdale: Erlbaum.

- Engle, R. W. (1996). Working memory and retrieval: An inhibition-resource approach. In J. Richardson, R. W. Engle, L. Hasher, R. H. Logie, E. Stolzhus, & R. Zacks (Eds.), *Working memory and human cognition* (pp. 89–119). New York: Oxford University Press.
- Engle, R. W., & Kane, M. J. (2004). Executive attention, working memory capacity, and a two-factor theory of cognitive control. In B. H. Ross (Ed.), *The psychology of learning and motivation* (Vol. 44), pp. 145–199. San Diego: Elsevier Academic Press.
- Gignac, G. E., & Szodorai, E. T. (2016). Effect size guidelines for individual differences research. *Personality and Individual Differences*, 102, 74–78.
- Halford, G. S., Wilson, W. H., & Philips, S. (1998). Processing capacity defined by relational complexity: Implications for comparative, developmental and cognitive psychology. *Behavioral and Brain Sciences*, 21, 803–865.
- Hasher, L., & Zacks, R. T. (1988). Working memory, comprehension, and aging: A review and a new view. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 22), pp. 193–225. San Diego: Academic Press.
- Horn, J. L., & Cattell, R. B. (1967). Age differences in fluid and crystallized intelligence. *Acta Psychologica*, 26, 107–129.
- Kane, M. J., Bleckley, M. K., Conway, A. R. A., & Engle, R. W. (2001). A controlled-attention view of working memory capacity. *Journal of Experimental Psychology: General*, 130, 169–183. <https://doi.org/10.1037/0096-3445.130.2.169>
- Kane, M. J., & Engle, R. W. (2003). Working-memory capacity and the control of attention: The contributions of goal neglect, response competition, and task set to Stroop interference. *Journal of Experimental Psychology: General*, 132, 47–70. <https://doi.org/10.1037/0096-3445.132.1.47>
- Krumm, S., Schmidt-Atzert, L., Buehner, M., Ziegler, M., Michalczyk, K., & Arrow, K. (2009). Storage and non-storage components of working memory predicting reasoning: A simultaneous examination of a wider range of ability factors. *Intelligence*, 37, 347–364.
- Lu, J., Tian, L., Zhang, J., Wang, J., Ye, C., & Liu, Q. (2017). Strategic inhibition of distractors with visual working memory contents after involuntary attention capture. *Scientific Reports*, 7, 16314. <https://doi.org/10.1038/s41598-017-16305-5>
- Oberauer, K. (2009). Design for a working memory. In B. H. Ross (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 51), pp. 45–100. San Diego: Elsevier Academic Press. [https://doi.org/10.1016/S0079-7421\(09\)51002-X](https://doi.org/10.1016/S0079-7421(09)51002-X)
- Oberauer, K., Süß, H. M., Wilhelm, O., & Wittman, W. W. (2008). Which working memory functions predict intelligence? *Intelligence*, 36, 641–652. <https://doi.org/10.1016/j.intell.2008.01.007>
- Oberauer, K., Süß, H.-M., Wilhelm, O., & Wittman, W. W. (2003). The multiple faces of working memory: Storage, processing, supervision, and coordination. *Intelligence*, 31, 167–193.
- Raven, J. (1989). The Raven Progressive Matrices: A review of national norming studies and ethnic and socioeconomic variation within the United States. *Journal of Educational Measurement*, 26, 1–16.
- 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, 164–171.
- Redick, T. S., & Lindsey, D. R. B. (2013). Complex span and *n*-back measures of working memory: A meta-analysis. *Psychonomic Bulletin & Review*, 20, 1102–1113. <https://doi.org/10.3758/s13423-013-0453-9>
- Robertson, I., Gratton, L., & Sharpley, D. (1987). The psychometric properties and design of managerial assessment centres: Dimensions into exercises won't go. *Journal of Occupational Psychology*, 60, 187–195.
- Shipstead, Z., Harrison, T. L., & Engle, R. W. (2016). Working memory capacity and fluid intelligence: Maintenance and disengagement. *Perspectives on Psychological Science*, 11, 771–799. <https://doi.org/10.1177/1745691616650647>
- Stankov, L. (2000). Complexity, metacognition, and fluid intelligence. *Intelligence*, 28, 121–143.
- Sternberg, R. J. (1977). Component processes in analogical reasoning. *Psychological Review*, 84, 353–378. <https://doi.org/10.1037/0033-295X.84.4.353>
- Verguts, T., & De Boeck, P. (2002). The induction of solution rules in Raven's Progressive Matrices Test. *European Journal of Cognitive Psychology*, 14, 521–547.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.