Psychology Science Quarterly, Volume 51, 2009 (4), pp. 449-469

An experimental-differential investigation of cognitive complexity

DAMIAN P. BIRNEY[1] & DAVID B. BOWMAN[2]

Abstract

Cognitive complexity as defined by differential and experimental traditions was explored to investigate the theoretical advantage and utility of relational complexity (RC) theory as a common framework for studying fluid cognitive functions. RC theory provides a domain general account of processing demand as a function of task complexity. In total, 142 participants completed two tasks in which RC was manipulated, and two tasks entailing manipulations of complexity derived from the differential psychology literature. A series of analyses indicated that, as expected, task manipulations influenced item difficulty. However, comparable changes in a psychometric index of complexity were not consistently observed. Active maintenance of information across multiple steps of the problem solving process, which entails strategic coordination of storage and processing that cannot be modelled under the RC framework was found to be an important component of cognitive complexity.

Key words: Complexity; relational complexity; fluid intelligence

[1] *Correspondence concerning this article should be addressed to:* Damian P. Birney, PhD, AGSM Building (G27), Accelerated Learning Laboratory, University of New South Wales, NSW, Australia, 2052; email: d.birney@unsw.edu.au

[2] Accelerated Learning Laboratory, University of New South Wales, Australia

During the last 20 years, considerable research effort has been devoted to merging differential and experimental traditions to study cognitive function (e.g., Colom, Flores-Mendoza & Rebollo, 2003; Conway, Kane & Engle, 2003; Kyllonen & Christal, 1990; Mackintosh & Bennett, 2003; Unsworth & Engle, 2007). The original hope was that integrative, cross-paradigm research would lead to new insights about human performance that could not be achieved from either tradition alone (Cronbach, 1957). As might be expected, this research has tended to draw from well established concepts – working-memory (WM) from the experimental paradigm and fluid intelligence (Gf) from the differential approach (Ackerman, Beier & Boyle, 2005). While there is promising theoretical progress being made in some areas, for example in the identification of controlled attention as central to the association between Gf and WM, (Engle, Tuholski, Laughlin & Conway, 1999; Kane & Engle, 2003), it is becoming increasingly clear that the earlier hope that a combination of approaches would fill a gap in theory has not eventuated to the extent that was expected (see, Deary, 2001; Lohman & Ippel, 1993). In this paper we investigate *complexity* as defined psychometrically within differential psychology, as a means to test the utility of relational complexity theory as an analytic framework for investigating fluid cognitive functions.

The most recent ventures into a merger of experimental and differential paradigms was spurred in no small part by the recasting of Cronbach's (1957) call for unity by Kyllonen and Chrystal (1990). In a series of four studies, Kyllonen and Chrystal argued on empirical grounds (correlations in excess of +.80) that reasoning ability or 'g', was nothing more than WM capacity. This finding was considered to be significant because it was the first time that the study of WM capacity was framed within the differential paradigm of latent variable analyses (Conway, et al., 2003). It could be argued that in spite of some shortcomings, the Kyllonen and Chrystal approach has been successful, at least in terms of citations (over 500, Web of Science, October, 2009), not so much because of what it achieved in theoretical terms, but because endorsement of their findings requires little if any theoretical movement between the experimental and differential paradigms. The result was that cognitive psychologists were able to continue exploring the components of WM and differential psychologists were able to continue exploring the correlates of Gf, each without relinquishing any theoretical ground to the other, and importantly, without any apparent necessity to investigate the potential collinearity of their separate endeavours. In practical terms, this has meant two things. First, evidence gathered in support of one construct could be used to bolster arguments for the validity of the other without the need to build a theoretical link between the two. Second, such views have lead to what we believe is a worrying trend in the conceptualisation of the relationships between different fluid cognitive functions. This is epitomised, for example, by Blair (2006, p. 4) who claims terms such as fluid intelligence, executive function, attention, effortful control, and WM "…each essentially describes the same overarching construct". In response, Birney, Bowman, and Pallier (2006) argued that the theoretical synonymity of these concepts is not self-evident, and that merging concepts from different research paradigms (e.g., cognitive-experimental, differential, neuropsychological) need to be done far more cautiously. Our proposition here is that a cross-paradigm treatment of complexity is one way that this can be achieved.

Cognitive complexity from differential and experimental perspectives

Psychometric complexity

Jensen (1987) has argued that the most undisputed fact about general intelligence, ‘g’, is that loadings of tasks on this factor are an increasing monotonic function of the tasks’ complexity. That is, performance on a more complex task is a better indicator of general intelligence. This understanding has also been applied to validate complexity manipulations within the same task. For instance, Arend et al. (2003) define cognitive complexity of linear-syllogism items in terms of the item’s loading on the first un-rotated factor in a factor-analysis. The core criterion of complexity within differential psychology is, therefore, that correlations with (or loadings on) measures of general intelligence (Gf specifically) should increase with task complexity but all else being equal, not with increases in difficulty generated by other task features (Spilsbury, Stankov & Roberts, 1990; Stankov, 2000). We refer to this as *psychometric complexity* to differentiate these empirical effects from more process-oriented accounts of task complexity that we describe shortly.

Psychometric complexity is useful for internal validation of purported manipulations of task complexity, but it is, by definition, agnostic to underlying processes at the source of complexity. Thus, while there is a growing body of research exploring fluid cognitive functions using this basic conceptualisation and criterion (e.g., Schweizer, 1996, 2001; Stankov, 2000), psychometric complexity does not provide a clear conception of precisely what it is that makes a task complex. As such, researchers have often been left little option but to adopt a rather eclectic approach to defining the cognitive complexity of a task (cf Stankov, 2000). This is appropriate if one’s goal is simply to develop tasks that are good measures of intelligence. However, a greater emphasis on theory development is needed to understand why these tasks “work”.

Relational Complexity (RC)

Research emerging from the working-memory paradigm goes part of the way to addressing the shortcomings of the atheoretical psychometric complexity criterion. Relational Complexity (RC) theory offers a formal specification of how task features contribute to cognitive complexity (Halford, Wilson & Phillips, 1998a). RC theory was proposed as a conceptualization of WM that focuses on processing capacity. From this view, it is the complexity of the relations *between* elements that have to be combined and processed within a given task that determines cognitive demand. Halford, Cowan, and Andrews (2007) argue that evidence to date suggests that rather than being a process theory about particular tasks, RC theory is domain independent.

Specification of relational complexity theory: An *N*-ary relation, $R(A_1, A_2, .. A_N)$, is a psychologically substantive binding between *N* elements in WM that when instantiated, constitutes a decision. The more complex the relation, the greater the demand on WM. Binary relations have two arguments; ternary-relations have three arguments, and so on. The concept of relational reasoning and its link with RC theory can be derived from a small number of premises (Birney, Halford & Andrews, 2006; Halford, et al., 1998a). First, reasoning entails representing and processing relations between task entities. Second, processing internally represented relations generates non-trivial cognitive demand. Third, the complexity of the internally represented relation can be used to quantify the characteristics of the processes used in performing the task – this is relational complexity. And fourth, processing capacity is a function of the *peak* RC of the representations that an individual can process, and therefore, processing capacity can be quantified using the same metric.

According to RC theory, the cognitive demand of a complex task can be reduced through either segmentation or chunking. *Chunking* entails the recoding of a relation into a lower level relation with the result being a loss of access to relations between elements that are chunked (Halford, et al., 2007). Thus chunking is only a useful strategy when the full complexity of a relation need not be considered for successful completion of a task. *Segmentation* on the other hand entails the decomposition of a relation into a series of lower-level relations that are integrated in series. Relations are only defined between variables that are in the same segment (i.e., step) and relations between variables in different segments are inaccessible to the current reasoning process (Halford, et al., 1998a). We describe application of RC analyses in the tasks used here in subsequent sections.

Aims and objectives

We argue that there remains a gap in research on fluid cognitive functions at the intersection between experimental and differential paradigms. We aim to investigate how understanding the determinants of cognitive complexity might further bridge this gap. In the current work complexity is conceptualised in two ways: (1) psychometric complexity, as determined by the Gf-complexity effect (Stankov, 2000), and (2) as a formal model of task demand, as specified by RC theory (Halford, et al., 2007).

We explore four tasks in which cognitive complexity has been manipulated. Two tasks have manipulations based on RC theory and have been reported previously in the cognitive-experimental literature: the Latin Square task (LST) and the Sentence Comprehension Task (SCT). The other two tasks come predominantly from differential psychology (Stankov, 2000). The Swaps test (SWAPS) manipulates complexity in terms of the number of required mental permutations; and the complexity manipulation of the Triplet Numbers test (TRIP) entails increasing the nature of conjunctive and/or disjunctive statements in rule validation. An RC task analysis is reported for all tasks.

Two indicators of increases in cognitive complexity are considered. The first is the difficulty effect – task solution is expected to become more difficult as complexity is increased. The second indicator is the psychometric complexity effect described previously. That is, manipulations of cognitive complexity are expected to result in a monotonic increase in the association between Gf and task performance concomitant with the complexity manipulation. The criterion measure of Gf is the Raven's Progressive Matrices test.

Method

Participants

In total 191 students (127 female) with a mean age = 20.76 (*SD* = 6.07) participated in the study. The results are based on 142 students (96 female) with a mean age = 20.19 (SD = 5.02) who had complete data for all but the SCT (8 of the 142 participants did not complete the SCT). Missing data was either a result of the participant failing to attend the second testing session, and/or insufficient time to complete all tasks. The students were enrolled in a first-year undergraduate subject at the University of Queensland and received course credit for their participation. Testing was conducted in groups of no more than 20 students.

Materials

Latin Square Task (LST) – RC manipulation

The Latin Square Task (LST) was developed specifically to assess the impact of RC on adult cognition. Full details of the task and justifications of the RC analyses are reported in Birney, Halford, and Andrews (2006). In a typical problem, an incomplete 4 × 4 matrix (Figure 1) is presented. The participants' task is to determine which of four elements should fill a target cell so that the defining principle of the Latin Square is satisfied, namely that each of the four possible elements occurs in every row and column of the matrix only once. As described below, the RC manipulation in the LST is based on an increasingly complex instantiation of this rule.

Binary (RC2) processing in the LST. As a rule, binary LST items require integration of elements within either a single column or row but not across both. In the example binary problem in Figure 1A, three elements in column 3 are given. The fourth can be determined by comparing the elements that are present in column 3 with the four elements known to complete the set. There is no need to consider any other cell in the square. Using simple conjunction (AND) and implication (→) relations this can be represented as:

$$\mathrm{AND}(\underline{R^1C^3(\text{circle}), R^3C^3(\text{square}), R^4C^3(\text{cross})}) \rightarrow R^2C^3(\underline{\text{triangle}})$$

which is read, "row1column3 is a circle, AND row3column3 is a square AND row4column3 is a cross, together IMPLIES row2colum3 is a triangle". The continuous underlining indicates those arguments that can be chunked without loss of information necessary to make the current decision. To summarize the analysis, the crux of the problem entails two sets of elements – the complete known possible set {circle,square,cross,triangle} and the given set {circle,square,cross}. The comparison of these two groups of elements entails a binary relation. The relations between elements within the column 3 chunk do not need to be considered to make the current decision. We need only know that element {triangle} is different from all of {circle,square,cross} and whatever relations between {circle}, {square} and {cross} that might exist is not relevant for solution (which is why they can be chunked).

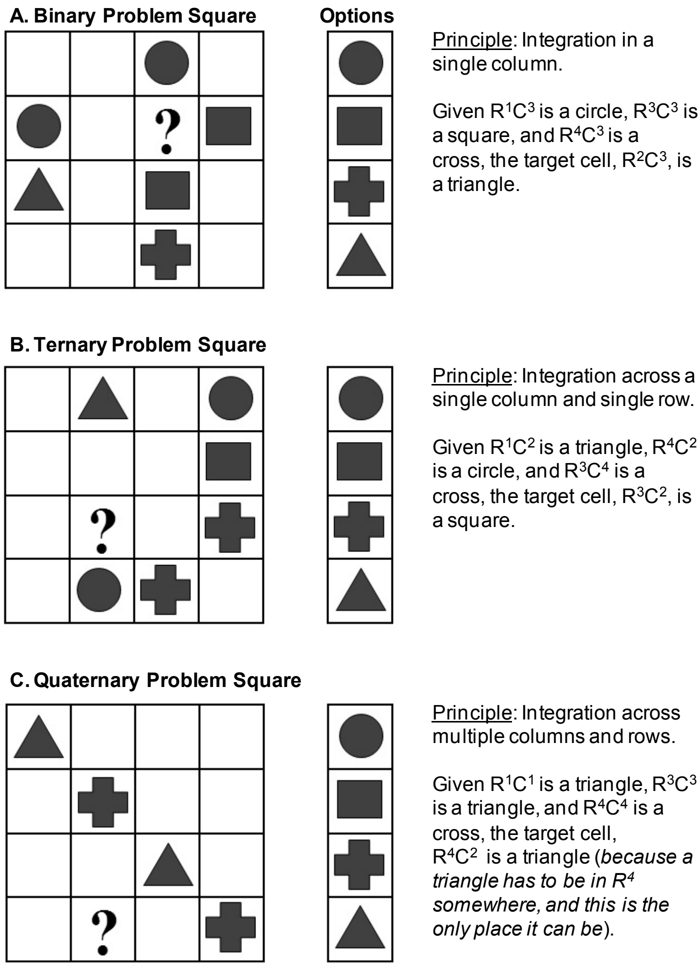

Figure 1:
Example Latin Square problems and verbal description of solutions for A). Binary; B). Ternary; and C). Quaternary items

RC analysis of ternary (RC3) LST items: Figure 1B shows a ternary-relational problem. The solution is achieved through the integration of information from column 2 and row 3. By the defining principle, the intersection, R^3C^2, must not contain an element that is present or can be determined to be present in the other cells of column 2 or row 3, and therefore the target response is element {square}. The problem solution is represented as:

$$\text{AND}(\underline{R^1C^2(\text{triangle}), R^4C^2(\text{circle})}, \underline{R^3C^4(\text{cross})}) \rightarrow \underline{R^3C^2(\text{square})}$$

The two elements in column 2 {triangle,circle} can be chunked because a) relations between them do not need to be processed, and b) the constraint they exercise on the target cell is clear given they are in the same column. However, the {cross} in R^3C^4 cannot be chunked with the other terms because (by the Latin Square defining principle) elements in row 3 are not independent of the elements in the intersecting columns (Birney, Halford, et al., 2006). Elements in row 3 need to be considered to make the current decision.

RC analysis of quaternary (RC4) LST items: In quaternary items as in Figure 1C, the target cell cannot be determined by the binary and ternary strategies just described. Applying these strategies to this problem results in knowing only that the target cell is not a {cross}. Solution (that the target cell is {triangle}) depends on integrating elements across multiple rows and columns, rather than a simple intersection. By an extension of the principles stated above, the three elements that constrain the target cell cannot be chunked, and must be processed separately (see Figure 1C). The problem can be represented as:

$$\text{AND}(\underline{R^1C^1(\text{triangle})}, \underline{R^3C^3(\text{triangle})}, \underline{R^4C^4(\text{cross})}) \rightarrow \underline{R^4C^2(\text{triangle})}$$

In summary, binary (RC2) problems require application within a single row or column; ternary (RC3) problems require application across a single row and single column; and quaternary (RC4) problems require application across multiple rows and columns.

A pool of 36 LST items consisting of 12 items at each level of RC were developed. Half the problems entailed a single processing step and half included an additional processing step of equal or lower RC level. The procedure is identical to that used by Birney, Halford, et al. (2006). So that the RC and Steps manipulation could be evaluated independent of presentation order, test items were administered in a different random order to each subject following an instruction, practice and feedback phase.

Sentence Comprehension Task (SCT) – RC manipulation

The core manipulation in the SCT is in terms of comprehensibility of centre-embedded sentences. The essence of the RC analyses is described briefly here but is reported in detail elsewhere (Andrews, Birney & Halford, 2006; Andrews & Halford, 2002). Comprehension of objective-relative (also called centre-embedded) sentences imposes simultaneous demand in the Halford et al. (1998a) sense because the assignment of the nouns to the thematic roles of the verbs entails processing relations between nouns and verbs in the sentence. The greater the amount of information that needs to be integrated in order to determine noun-verb roles, the greater the RC of the item. That is, in object-relative sentences such as, "*The duck that the monkey touched walked*", no assignment of roles is possible until the verb *touched* is encountered. To make sense of the sentence, *monkey* is assigned to the agent role, and *duck* must be assigned to the patient role. In this case, the assignment needs to be done in the same decision. We then know that *the monkey touched the duck,* and that *the duck was touched by the monkey*. The sentence is further complicated because the second verb, *walked*, needs to be considered as well. Andrews and Halford (2002) argue that the close proximity of the two verbs means that a temporal overlap in processing is likely, further increasing the demands of processing resources. The RC analysis of this sentence is represented by Halford et al. (1998a, Section 6.1.4) as a ternary relation using propositional format as follows:

TOUCH(monkey, duck) and WALK(duck). Processing is ternary here because even though the output is broken into two parts, all three roles need to be resolved for complete comprehension.

In comparison, in subject-relative (also called right-branching) sentences the assignment of noun-verb roles can be determined in series as the sentence is read, hence cognitive load is effectively mitigated online. A subject-relative example of the previous sentence would be: "The monkey touched the duck that walked". Subject- and object-relative sentences differ in that comprehension of the latter can be characterized by identification and decomposition of the separate elements of the presented sentence (e.g., agent and patient), and then a recombination of these elements in their appropriate noun-verb relations. This decomposition process is circumvented in subject-relative sentences. Again, full details of the RC analysis of the SCT and theoretical justifications are reported in Andrews, et al. (2006; also see Andrews & Halford, 2002).

Participants were presented 6 object-relative and 6 subject-relative sentences from each of 4 complexity levels (RC2 to 5; coinciding with number of roles). The sentences were randomly drawn from a total pool of 64 items such that no more than four items from one level were presented consecutively. Table 1 lists example object-relative items at each complexity level (subject-relative sentences were included primarily to facilitate motivation and are not analysed). Participants were permitted to read the sentence as many times as they needed to ensure that they understood it and pressed the spacebar when ready to proceed. To evaluate sentence comprehension, a probe question referring to a noun-verb relation in the sentence was presented. The total number of correctly answered object-relative sentences for a given RC level was the accuracy score. The time from when the sentence was displayed to when the participant pressed the spacebar to see the probe question was used as the response-time measure (i.e., a comprehension time measure).

Table 1:
Example 2-, 3-, 4- and 5-role object-relative sentences with example probe question types

RC	Example Sentences	Probe Questions	Response
SCT2	Sally saw the woman that the man helped.	2-1. Who helped? 2-2. Who was helped?	Noun Noun
SCT3	The duck that the monkey touched walked.	3-1. Who touched? 3-5. What did the monkey do?	Noun Verb
SCT4	The artist that the waiter warned the chef about talked.	4-1. Who warned? 4-6. What did the waiter do?	Noun Verb
SCT5	The clown that the teacher that the actor liked watched laughed.	5-5. Who was watched? 5-8. What did the actor do?	Noun Verb

Swaps Test – Mental permutations manipulation

In this task participants are presented with a letter triplet (e.g., JKL) and are instructed to mentally rearrange or 'swap' the positions of letters according to a series of simple instructions (e.g., Swap 1 and 2; then Swap 3 and 2). The task requirement is to provide the final order of elements. Thus, the Swaps test (SWAPS) entails a manipulation of increasing the number of mental permutations (of the same type) required for solution. Participants must mentally rearrange the order of elements according to the first rule provided, maintain the intermediate solution in memory, perform a further permutation according to the second rule, and so on until all 'swap rules' have been implemented and the final iteration of elements is reached. These manipulations have been shown to lead to concomitant increases in difficulty and psychometric complexity (Stankov, 2000; Stankov & Crawford, 1993). Four sets of 12 items were generated entailing one, two, three, and four swaps, respectively. The 48 items were presented in a different random order for each subject.

RC analysis of the Swaps Test: An RC analysis entails consideration of the cognitive processes needed for solution. As far as we are aware, the SWAPS has not been analysed in this way. Our RC analysis is supported by the previous work of Stankov and Crawford (1993).

The SWAPS entails what Stankov and Crawford (1993) referred to as a working-memory placekeeper (WMP). The core feature of the SWAPS is a manipulation of a multi-element ordering held in WM. This is different to the LST where a single element is integrated with either elements that can be maintained outside WM (i.e., cell elements displayed on the screen), or resolved elements that are already part of WM through segmentation (i.e., in two step problems). In the case of the SWAPS, a binding between three elements *already* stored in WM needs to be broken and then re-established to satisfy the new swap rule. This unpacking of an established binding is proposed to generate cognitive load over and above that required for a traditional instantiation of a binary relation. To model this requirement of the SWAPS, we expand slightly on the standard RC analysis to demarcate an explicit *manipulated* WMP component (designated by + alongside the RC classification). This component is not a separate segment (or step) and, according to RC theory, does not in and of itself increase the RC of the task. We believe it does, however, contribute to cognitive load. This is described in the following example analysis:

Example
Stimulus: J K L
Swap Instructions: Swap 1 with 3; Swap 2 with 3; Swap 1 with 2

RC Analysis

SWAP$\underline{((1,3), JKL)}$ → $\underline{WMP1(LKJ)}$	RC = 2	Step [1]
SWAP$\underline{((2,3),WMP1(LKJ))}$ → $\underline{WMP2(LJK)}$	RC = 2+	Step [2]
SWAP$\underline{((1,2), WMP2(LJK))}$ → $\underline{JLK}$	RC = 2+	Step [3]

Step [1] requires at least three component processes, (a) encoding and binding of the stimulus (J K and L) to their respective slots, (b) manipulation of the displayed elements (JKL) consistent with the first instruction (swap 1 with 3) to generate a new set of bindings (LKJ), and (c) storage of the outcome (WMP1). As encoding per se is outside of RC theory

(Halford et al, 1998), only parts (b) and (c) contribute to the modelled decision making process, and hence the entire Step [1] segment entails a binary relation. Step [2] entails the manipulation of the elements *already stored in working memory* (WMP1) to break the existing bindings and to generate new ones consistent with the second rule (swap 2 with 3). The outcome of this manipulation (WMP2) then replaces WMP1 as the contents of working-memory. Step [3] entails the same type of manipulation, however because this is the last step, the response is reported. Hence, while all three steps entail a binary relation, Step [1] is conceptually different from Steps [2] and [3] because the cognitive load of the Step [1] bindings can be mitigated by drawing on an external memory aid – the original stimulus ordering displayed on the screen. We hypothesise that the latter steps entail greater WM load because of the necessary manipulation of stored information.

In summary, the analysis of the SWAPS suggests that the maximum cognitive load in terms of RC theory is binary, regardless of level. That is, the more levels, the more binary processes (swaps) are required. However, closer analysis suggests that there are specific task features that are predicted to cause a higher cognitive demand (WMPs) that RC theory cannot accommodate without some slight modification. We return to this in the General Discussion.

Triplet Numbers Test: Conjunctive and/or disjunctive embedding

The Triplets Number Test (TRIP) requires participants to verify whether a triplet of digits fit a given rule. Participants are required to attempt as many items as possible within an undisclosed time limit. There are four rules that increase in complexity primarily as a function of the level of embedding of conjunctive and/or disjunctive statements (Stankov, 2000; Stankov & Crawford, 1993).

TRIP 1 - Search Triplets: If a 3 is present within the triplet then press yes, otherwise press no (time limit = 2 min).
TRIP 2 - Half-rule Triplets: If the second digit in the triplet is the largest then press yes, otherwise press no (time limit = 3 min).
TRIP 3 - One-rule Triplets: If the second digit is the largest AND the third digit is the smallest, then press yes, otherwise press no (time limit = 6 min).
TRIP 4 - Two-rules Triplets: If the first digit is the largest AND the second digit is the smallest, OR, if the third digit is the largest and the first digit is the smallest, then press yes, otherwise press no (time limit = 6 min).

Two hundred triplets were randomly generated at each level such that the three elements of the triplet were unique (i.e., no repeated digits in the triplet) and that the rule for the given level was correct approximately 50% of the time (level1 = 52%, level2 = 50%, level3 = 50%, level4 = 53%). A random ordering of these 200 items was generated and this order was presented to all subjects. At the start of each level, subjects were presented a trial of four triplets as practice with correct/incorrect and response time feedback. The current rule was displayed at the top left of the screen during practice trials but not test trials. The levels were presented in order of increasing difficulty (TRIP 1 through 4). Because of restriction in range on accuracy data, the proportion of correct responses per minute was used as the measure for this task.

RC analysis of the Triplet Numbers Test: We begin our representation of the RC analysis of the TRIP by considering a plausible strategy to deal with the most complex level of the task. TRIP4 entails a representation of two conjunctions embedded in a disjunction. Using the propositional notation[3], the triplet (a, b, c) might be represented as follows:

Triplet: a b c
OR(AND(>(a, bc), <(b, ac)), AND(>(c, ab), <(a, bc)))

A *prima' facie* RC analysis suggests that this might require quaternary level processing (i.e., one argument for each conjunct). However it is quite straightforward to segment the task into a series of binary relations that are pieced together in series. For instance, consider the first component of the first conjunctive statement "the first digit is the largest" which would be represented as >(a, bc). The b and c terms are chunked because there is no need to consider the relationship between these terms explicitly to make the current decision. It is simply a comparison of the first digit with "the other two". If the veracity of this relation is confirmed then we move on to the next component of the first conjunction ("the 2nd digit is the smallest") which is represented as <(b, ac). If both of these relations are confirmed (entailing another binary relation) then the entire TRIP4 rule is confirmed and the participant should respond "yes". Otherwise the participant would move the focus of reasoning onto the next conjunction. Effectively what this means is that solution in the most complex level of the TRIP requires a series of binary processes, and the specific digits in the triplet determines the number of binary processes that would need to be instantiated. TRIP2 and 3 also have a peak binary load (RC2), and as was the case for TRIP4, the number of binary processes entailed within a level depends on the digits making up the triplet. Following Halford, et al. (1998a) TRIP1 ("Is there a 3 present?") entails a unary relation.

Although the RC analyses of the SWAPS and the TRIP suggest that the peak cognitive demand of both tasks is binary, there is at least one salient difference that should be considered. As described earlier, the SWAPS requires decomposition and re-composition of bindings held in WM. This is not a requirement of the TRIP.

Raven's progressive matrices

A composite test of 32 items from the Raven's Progressive Matrices test was developed. Set I - Advanced items were used as practice but were not analysed. The 12 items from Set E of the Standard Progressive Matrices followed by a random selection of 20 items from the Advanced Progressive matrices comprised the test items to be scored. No time-limit was imposed.

[3] The propositional logic is read from the centre out from left to right. In the case of the example, this would be "a is > bc AND b < ac OR c > ab AND a < bc"

General procedure

Test administration was spread over two sessions of up to 2 hours each, separated by no more than two weeks. Each participant received the battery of tasks in a different order[4].

Results and discussion

Descriptive statistics for all tasks are provided in Table 2. The analyses that follow are broken into two main sections. We explore difficulty and psychometric complexity effects as a function of the complexity manipulation in each task, first with accuracy measures, and second with response times. Table 3 reports the correlations between variables analysed. Correlations between experimental task levels and RPM are reported values in column 1 of Table 3. The extent to which the correlations with the RPM increase as a function of the task manipulation is indicative of psychometric complexity effects which are tested for statistical significance in the following analyses.

Difficulty vs psychometric complexity

Following Stankov and Crawford (1993), the relationship between RPM (as an indictor of Gf) and accuracy as a function of complexity is investigated using repeated-measures analysis of covariance (ANCOVA) with RPM being added as a covariate. For each task we consider a two-step test of the difficulty and complexity effects. The difficulty effect is evaluated by testing the main effects of the complexity manipulation on performance. The test of psychometric complexity is the linear contrast of the complexity level × RPM interaction effect, which, if statistically significant, is indicative of a monotonic increasing relation with RPM. As expected, RPM was a significant predictor of performance in all four tasks (partial η^2 range = .21 to .30)[5].

The Latin Square Task

The manipulation in the LST was in terms of RC and number of processing steps.

Difficulty: A level of RC (RC2, RC3, RC4) × Steps (1S, 2S) × RPM (covariate) repeated-measures ANCOVA was conducted on the composite accuracy performance scores (interactions are plotted in Figure 2A). Controlling for RPM score[6], the main-effect for complexity

---

[4] The data reported here is part of a larger individual differences study containing a battery of 11 cognitive tasks. Aspects of the data on the LST and SCT have been published in Andrews, Birney et al. (2006). The analyses considered in that paper focused on the overall LST score as a measure of relational processing capacity in predicting overall sentence comprehension. The analyses did not consider differential levels of complexity in the LST or the SCT, which is the focus of the current analyses.

[5] Throughout the Results, η^2 refers to the effect size estimate, partial eta squared.

[6] In this and all subsequent analyses, the main-effects are evaluated at the mean level of RPM performance = 64.61% correct.

Table 2:
Descriptive statistics for accuracy and correct response times

Variable		Accuracy Mean	SD	α	Correct Response Times Mean (s)	SD
(1)	RPM	0.65	(0.17)	.85	12.25	(6.59)
(2)	LST2-1s	0.96	(0.12)	.64	7.89	(2.46)
(3)	LST2-2s	0.82	(0.23)	.64	14.32	(5.83)
(4)	LST2	0.89	(0.15)	.72	10.85	(3.13)
(5)	LST3-1s	0.89	(0.16)	.45	20.35	(12.47)
(6)	LST3-2s	0.77	(0.22)	.48	22.06	(10.66)
(7)	LST3	0.83	(0.16)	.61	21.00	(10.14)
(8)	LST4-1s	0.64	(0.30)	.71	24.31	(18.48)
(9)	LST4-2s	0.44	(0.27)	.53	51.01	(36.15)
(10)	LST4	0.54	(0.23)	.70	35.25	(24.67)
(11)	SCT1	0.83	(0.17)	.66[a]	6.46	(3.00)
(12)	SCT2	0.84	(0.19)	.66[a]	6.41	(2.43)
(13)	SCT3	0.75	(0.21)	.66[a]	14.13	(6.15)
(14)	SCT4	0.57	(0.27)	.66[a]	24.95	(11.57)
(15)	Swaps1	0.92	(0.15)	.82	5.83	(1.52)
(16)	Swaps2	0.86	(0.21)	.83	12.48	(6.17)
(17)	Swaps3	0.78	(0.24)	.83	18.69	(5.56)
(18)	Swaps4	0.72	(0.27)	.85	24.70	(8.01)
(19)	Trip1	30.50	(2.78)	.78[b]	0.89	(0.22)
(20)	Trip2	27.13	(3.23)	.78[b]	1.09	(0.24)
(21)	Trip3	23.39	(2.85)	.78[b]	1.41	(0.31)
(22)	Trip4	16.99	(3.28)	.78[b]	2.09	(0.69)
(23)	Trip1%	0.99	(0.01)			
(24)	Trip2%	0.97	(0.06)			
(25)	Trip3%	0.96	(0.04)			
(26)	Trip4%	0.88	(0.14)			

LST total scale reliability = 0.84; SWAP total scale reliability = 0.94; a = SCT reliability between-level = .66; b = triplet between-level reliability = .78; % = Mean proportion correct score.

level was significant, $F(2, 280) = 299.04$, $MSe = 9.80$, $p < .001$, $\eta^2 = .68$. As expected, collapsing over number of steps, RC2 items were significantly easier than RC3 items, which were significantly easier than RC4 items (all p's < .001 with Bonferroni correction). There was also a statistically significant effect for number of steps, $F(1, 140) = 158.22$, $MSe = 5.00$, $p < .001$, $\eta^2 = .53$, one-step items were significantly easier than two-step items. The two-way interaction between RC and Steps was significant, $F(2, 280) = 3.63$, $MSe = 0.12$, $p = .028$, $\eta^2 = .025$, though the effect size is small. Further examination of the interaction showed that while the general main-effect trends described above remained, the multivariate simple effects indicate the main effect of complexity level on performance is statistically greater for 2-step problems (Pillai's Trace = .717, $F(2, 139) = 176.44$, $p < .001$, $\eta^2 = .717$)

Table 3:
Inter-test correlation matrix of task accuracy data

Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)	(21)
(1) RPM	-																				
(2) LST2-1s	.37	-																			
(3) LST2-2s	.46	.46	-																		
(4) LST2	.49	.74	.94	-																	
(5) LST3-1s	.40	.48	.37	.47	-																
(6) LST3-2s	.42	.36	.49	.51	.40	-															
(7) LST3	.49	.48	.52	.58	.78	.89	-														
(8) LST4-1s	.26	.13	.33	.30	.23	.47	.43	-													
(9) LST4-2s	.45	.38	.45	.49	.38	.54	.56	.33	-												
(10) LST4	.43	.30	.48	.48	.37	.61	.60	.84	.79	-											
(11) SCT1	.28	.23	.30	.32	.31	.23	.31	.12	.28	.24	-										
(12) SCT2	.34	.05	.19	.17	.25	.20	.26	.07	.25	.19	.37	-									
(13) SCT3	.46	.27	.31	.34	.35	.19	.31	.16	.30	.28	.25	.39	-								
(14) SCT4	.37	.15	.20	.21	.28	.20	.28	.14	.25	.23	.32	.26	.43	-							
(15) Swaps1	.33	.22	.39	.38	.18	.33	.32	.23	.22	.28	.18	.18	.30	.17	-						
(16) Swaps2	.40	.26	.51	.49	.21	.27	.29	.19	.33	.31	.22	.20	.31	.19	.66	-					
(17) Swaps3	.42	.24	.48	.46	.28	.34	.37	.29	.38	.41	.21	.25	.42	.32	.49	.77	-				
(18) Swaps4	.48	.28	.49	.48	.30	.31	.36	.30	.34	.39	.17	.23	.45	.27	.55	.78	.82	-			
(19) Trip1	.15	-.02	.05	.03	.04	.13	.11	.16	.05	.13	.08	.09	.15	.09	.05	.11	.15	.18	-		
(20) Trip2	.39	.43	.17	.30	.22	.28	.30	.14	.26	.24	.08	.10	.27	.09	.22	.20	.19	.23	.51	-	
(21) Trip3	.40	.14	.19	.20	.12	.23	.21	.14	.23	.22	.07	.18	.25	.13	.30	.32	.41	.38	.34	.69	-
(22) Trip4	.45	.20	.28	.29	.20	.09	.16	.08	.25	.19	.16	.20	.38	.24	.26	.35	.43	.40	.15	.44	.67

Notes: r's > .17, .22, and .28, are significant at $p < .05$, .01, and .001 respectively (N=134); Correlations between SCT measures and other variables are based on N = 134 (all other correlations are based on N = 142) than 1-step problems (Pillai's Trace = .524, F(2, 139) = 76.44, p < .001, η^2 = .524). Thus, there is empirical evidence to indicate that both increasing relational complexity and adding an additional processing step in the LST contribute somewhat independently to the difficulty of the task.

Psychometric complexity: As previously indicated, higher levels of RPM were associated with higher accuracy on the LST (see Table 4). However, contrary to expectations, the relationship with RPM and performance was not moderated by level of complexity, thus there is no evidence for a psychometric complexity effect in the RC manipulation of the LST. Neither the full-interaction nor the linear contrast interaction effect was statistically significant; $F(2, 280) = 1.75$, $MSe = 0.06$, $p = .176$, $\eta^2 = .01$; and $F(1, 140) = 2.33$, $MSe = 0.10$, $p = .129$, $\eta^2 = .02$, respectively. (The three-way interaction between level, steps, and RPM was not significant, $F < 1$).

Given number of processing steps is theoretically and empirically associated with an increase in difficulty, the interaction effect between RPM and Steps is also of interest. The full interaction was statistically significant, $F(1, 140) = 13.15$, $MSe = 0.42$, $p < .001$, $\eta^2 = .09$. As this is a 1-df test, the test of the linear contrast is identical; hence there is evidence for psychometric complexity (where the "complexity" manipulation is now the number of processing steps). Closer examination of the interaction at each level of complexity suggests that the statistical significance of this finding is driven predominantly by binary level items. That is, the association between RPM and performance is significantly lower for 1-step than 2-step binary items, $F(1, 140) = 13.68$, $MSe = .256$, $p < .001$, $\eta^2 = .089$, but this effect did not reach statistical significance for ternary, $F(1, 140) = 2.66$, $MSe = .060$, $p = .105$, $\eta^2 = .019$, or quaternary items $F(1, 140) = 2.43$, $MSe = .133$, $p =.122$, $\eta^2 = .017$. The main theoretical implication of this is that the need to *maintain* information across two processing steps may be a significant contributor to psychometric complexity, at least for items of lower RC.

The Sentence Comprehension Task

The SCT manipulation was in terms of the RC of the centre-embedded of object-relative sentences.

Difficulty: A Level (RC2 to RC5) × RPM (covariate) repeated-measures ANCOVA analysis was conducted on the composite accuracy performance scores (interactions plotted in Figure 2B). Controlling for RPM score, the main-effect for SCT level was significant, $F(3, 396) = 69.52$, $MSe = 2.10$, $p < .001$, $\eta^2 = 0.35$. While level RC2 was not significantly easier than level RC3 ($p > .10$), each successive level of the SCT (RC3 to RC5) was significantly more difficult than the preceding level (all p's $\geq .001$ with Bonferroni correction). Consistent with Andrews and Halford (2002), this provides empirical evidence to indicate that increasing the level of center-embededness in object-relative sentences makes task performance difficult.

Psychometric complexity: The full-interaction between SCT level and RPM was statistically significant; $F(3, 396) = 2.85$, $MSe = 0.09$, $p = .037$, $\eta^2 = .02$. The linear contrast effect was also statistically significant, $F(1, 132) = 6.39$, $MSe = 0.24$, $p = .013$, $\eta^2 = .05$, though the effect size is relatively small. Consistent with expectations of the psychometric complexity

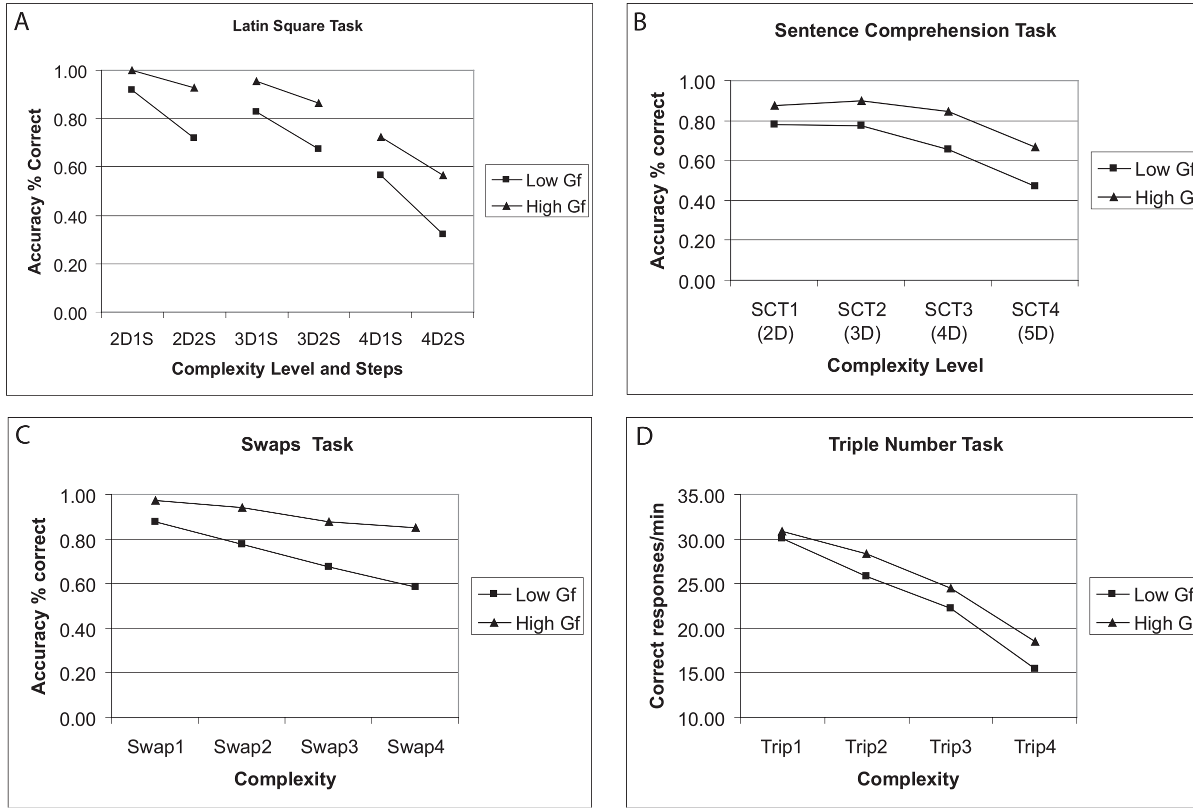

Figure 2:
Accuracy by Gf (RPM) ability for A). Latin Square Task; B). Sentence Comprehension Task; C). Swaps test; and D). Triplet Numbers Test. Note: Low and High Gf curves are plotted at 1 standard deviation below and above the mean, respectively

effect, the relationship between RPM and SCT performance is moderated by level. With each increase in the center-embeddedness of the sentence, the magnitude of the relationship with RPM tended to increase. This is supporting evidence that the RC manipulation in the SCT is consistent with psychometric complexity.

The Swaps Test

The manipulation in the SWAPS was in terms of the number of required mental permutations.

Difficulty: A Swaps (level 1 to 4) × RPM (covariate) repeated-measures ANCOVA was conducted on the composite accuracy performance scores (interactions plotted in Figure 2C). Controlling for RPM score, the main-effect for SWAPS was significant, $F(3, 420) = 75.11$, $MSe = 1.17$, $p < .001$, $\eta^2 = 0.35$. As expected, each successive level of SWAPS (level 1 to 4) was significantly more difficult than the preceding level (all p's < .001 with Bonferroni correction). Consistent with previous research (Stankov, 2000), increasing the number of required mental permutations in the Swaps tasks makes solution progressively more difficult.

Psychometric complexity: The full-interaction between SWAPS and RPM was statistically significant; $F(3, 420) = 10.78$, $MSe = 0.17$, $p < .001$, $\eta^2 = .07$. The linear contrast effect was also statistically significant, $F(1, 140) = 19.65$, $MSe = 0.50$, $p < .001$, $\eta^2 = .12$, and is the strongest psychometric complexity effect observed in the four experimental tasks. Each increase in number of mental permutations was associated with a concomitant increase in the magnitude of the association with RPM.

The Triplets Number Test

The manipulation in the TRIP was in terms of the complexity of the conjunctive/disjunctive rule to be verified.

Difficulty: A TRIP (1 to 4) × RPM (covariate) repeated-measures ANCOVA was conducted on the composite accuracy performance scores (number of correct responses per minute). The interaction is plotted in Figure 3B. Controlling for RPM score, the main-effect for Trip level was significant, $F(3, 420) = 998.90$, $MSe = 4761.30$, $p < .001$, $\eta^2 = 0.88$. As expected, each successive triplet level was significantly more difficult than the preceding one (all p's < .001 with Bonferroni correction). Thus, increasing the level of complexity in rule verification increases the difficulty of the task.

Psychometric complexity: The full-interaction between RPM and Trip level was statistically significant; $F(3, 420) = 6.27$, $MSe = 29.86$, $p < .001$, $\eta^2 = .04$. The linear contrast effect was also statistically significant, $F(1, 140) = 8.21$, $MSe = 65.90$, $p = .005$, $\eta^2 = .06$. Closer inspection (e.g., Figure 2D) indicates that the psychometric complexity effect is driven, at least in part, by association between TRIP1 and RPM being significantly weaker than for the remaining levels. Hence, statistically there is evidence in support of a psychometric complexity effect in the triplets task. However, in practical terms, this effect appears to be weaker in the present study than has been reported previously (e.g., Stankov, 2000). The significant linear TRIP × RPM interaction was primarily driven by small differences between high and low RPM scores for levels TRIP1 to TRIP3 and a large deviation for TRIP4.

Correct response time analyses

Response time has frequently been implicated in cognitive functioning. Hence it is of theoretical interest to investigate the pattern of difficulty and psychometric complexity effects using these measures. ANCOVA analyses were repeated for each task using mean correct response time (correctRT) as the dependent variable. Given that there is substantial overlap with the accuracy analyses, only a summary of the correctRT results are reported.

A statistically significant main effect for manipulation level was observed in each task. The level main-effects for the LST ($F(2, 240)=136.59$; $MSe=43394.82$; $p< .001$; $\eta^2=.53$), SCT ($F(3, 384)=15.118$; $MSe=47.73$; $p< .001$; $\eta^2=.11$), SWAPS ($F(3, 405)=532.97$; $MSe=9275.62$; $p< .001$; $\eta^2=.80$) and TRIP ($F(3, 420)=313.39$; $MSe=39.1$; $p< .001$; $\eta^2=.69$) mirrored those for accuracy in that higher complexity levels were associated with longer times to respond correctly. Where differences were observed between levels for accuracy, there were also differences between levels for correctRT, lending further support for a significant difficulty effect in each experimental task. In the LST, the main effect for number of processing steps was also significant ($F(1, 120)=128.21$; $MSe=23609.39$; $p< .001$; $\eta^2=.52$): 1-step problems were answered correctly significantly more quickly than 2-step problems.

The pattern of main-effects for RPM (covariate) across the four tasks did however differ from the accuracy data. In the SWAPS, RPM did not significantly predict overall correctRT ($F(1, 135)=0.21$; $MSe=10.58$; $p=.648$; $\eta^2 \approx.00$) and the test for psychometric complexity (level × RPM linear contrast) was not significant ($F(1, 135)=2.375$; $MSe=80.43$; $p=.126$; $\eta^2=0.02$).That is, speed of response was not a source of individual differences related to the RPM. In the TRIP, RPM did significantly predict overall correctRT ($p=.008$; $\eta^2=.05$) but this effect did not differ as a function of complexity (ie., the level × RPM linear contrast was not significant: $F(1, 140)=0.852$; $MSe=0.198$; $p=.358$; $\eta^2=.01$). The psychometric complexity effect was only observed in the LST (for both RC: $F(1, 120)=9.7$; $MSe=4886.09$; $p=.002$; $\eta^2=.08$, and STEPS: $F(1, 120)=12.78$; $MSe=2352.65$; $p<.001$; $\eta^2=.10$) and the SCT ($F(1, 128)=9.20$; $MSe=533.68$; $p=.003$; $\eta^2=.07$).

In sum, the correctRT analyses reported provide further evidence for a manipulation of difficulty across all tasks. However, support for psychometric complexity effects based on correctRT were only observed in the LST and SCT.

General discussion

The results in the present study are on the whole straightforward. Consistent with previous research, manipulations in all four tasks designed to increase the cognitive demand imposed on Gf (RPM) ability were successful. That is, regardless of whether the manipulation was based on relational complexity theory in terms of element integration in a Latin square (Birney, Halford, et al., 2006) or the center-embeddedness of sentences (Andrews & Halford, 2002), or on the number of mental permutations or the complexity of verifying conjunctive and/or disjunctive statements (Stankov, 2000), all the tasks became more difficult as demand was systematically increased. The next question was whether these manipulations made performance difficult because they required a greater investment of Gf abilities – as measured by the Raven's progressive matrices test – or because of some other factor. Again the results are straightforward. First, in all but one case (SWAPS correctRT only), performance in all tasks was significantly related to Gf abilities. However, as demonstrated in Figure 2, it was not the case that all the manipulations conformed to the psychometric complexity effect as predicted. Furthermore, being founded on RC theory – a formal theory of complexity – did not seem to provide any advantage to a task in terms of whether the psychometric complexity effect would be found or not. In the remainder of this paper, we explore why this may be the case drawing on insights from the RC analyses.

Tasks, processes, and complexity

RC theory has provided a formalised framework which has helped highlight task specific differences in cognitive demand. However, considering the details of our RC analyses along with the empirical findings reported, the interpretation we advocate here is that it is not storage (e.g., number of steps) or even processing (e.g., relational complexity demand) that is the theoretical crux of what is common in these tasks. Rather we conclude that it is the need to actively maintain task relevant information across the problem space which is common to the psychometric complexity effect. That is, our data suggests that *active* maintenance (rather than passive storage) is a source of individual differences in Gf.

A number of findings in the current work lead us to this conclusion. First, our RC analysis of the SWAPS suggests that the maintenance and manipulation of intermediate bindings of elements to an ordinal position (i.e., the RC2+ segments in our analysis) is an important aspect of task performance over and above the apparent peak binary RC load. This is in marked contrast to the TRIP, which according to our RC analysis did not require the same degree of *active maintenance* in WM in spite of also requiring binary processes. In the TRIP, each of the binary processes can be resolved and their solutions are only required (if at all) to keep track of the progression of rule verification processing. These task differences were reflected in the data: Relatively strong evidence for a psychometric complexity effect was observed in the SWAPS but was considerably weaker for the TRIP.

An active-maintenance interpretation was observed empirically in the LST. While psychometric complexity was not observed across RC levels in the LST, it was observed in the Step manipulation. Two-step problems which require an interim solution from the first step to be maintained and used in WM (Birney, Halford, et al., 2006), were not only more difficult than one-step problems, they were psychometrically more complex.

The SCT findings can also be interpreted in terms of a significant active maintenance load. While this interpretation of task performance is less faithful to the standard RC framework, we believe it provides a plausible account of the present findings. As shown earlier, the RC analysis of a sentence such as “*The duck that the monkey touched walked*” is represented by multiple relations: TOUCH(monkey, duck) and WALK(duck). While Halford, et al. (1998a) argue that it is necessary to consider both relations at the outset to arrive at this representation (i.e, comprehension), ultimately there are two relations that need to be maintained and manipulated in WM. This is consistent with validation work on the RC theory conducted independently of Halford’s laboratory. Waltz, et al. (1999) showed that the *number* of relations (rather than the peak complexity of the relation, as specified by RC theory) was a significant source of difficulty in RPM-like problems. It seems that coordinating resources to link, stream, and maintain the outcome of a number of (different) steps together for processing is a significant source of cognitive demand.

Implications and conclusion: Studying complexity is useful

RC theory cannot accommodate all sources of cognitive demand in the tasks we have investigated. This is not fatal to the RC approach as it was never intended that it should account for all cognitive facets of all problems (Halford, Wilson & Phillips, 1998b). However, we believe our results do demonstrate the benefit of a common, formalised complexity framework. Theoretically, our findings support Kane et al.'s (2004) conclusion that the processing component responsible for the relation between WM and Gf may be something akin to controlled attention. We have achieved this insight not from correlational analyses of simple- and complex-span WM tasks, as Kane et al. have tended to do, but from considered analyses of cognitive tasks in which complexity is systematically manipulated. In spite of the complexity in these tasks being defined in a number of different ways, we have shown that through application of RC theory, tasks can be meaningfully compared so that similarities and differences and the relations they hold with Gf can be better understood.

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.

Andrews, G., Birney, D. P., & Halford, G. S. (2006). Relational processing and working memory in the comprehension of complex relative clause sentences. *Memory & Cognition, 34*, 1325-1340.

Andrews, G., & Halford, G. S. (2002). A cognitive complexity metric applied to cognitive development. *Cognitive Psychology, 45*, 153-219.

Arend, I., Colom, R., Botella, J., Contreras, M. J., Rubio, V., & Santacreu, J. (2003). Quantifying cognitive complexity: Evidence from a reasoning task. *Personality and Individual Differences, 35*, 659-669.

Birney, D. P., Bowman, D. B., & Pallier, G. (2006). Prior to paradigm integration, the task is to resolve construct definitions of gF and WM. *Behavioral and Brain Science, 29*, 127.

Birney, D. P., Halford, G. S., & Andrews, G. (2006). The Influence of Relational Complexity on Reasoning: The Development of the Latin Square Task. *Educational and Psychological Measurement, 66*, 146-171.

Blair, C. (2006). How similar are fluid cognition and general intelligence? A developmental neuroscience perspective on fluid cognition as an aspect of human cognitive ability. *Behavioral and Brain Sciences, 29*, 109-160.

Colom, R., Flores-Mendoza, C., & Rebollo, I. (2003). Working memory and intelligence. *Personality & Individual Differences, 34*, 33-39.

Conway, A. R., Kane, M. J., & Engle, R. W. (2003). Working memory capacity and its relation to general intelligence. *Trends in Cognitive Sciences, 7*, 547-552.

Cronbach, L. J. (1957). The two disciplines of scientific psychology. *American Psychologist, 12*, 671-684.

Deary, I. J. (2001). Human intelligence differences: Towards a combined experimental-differential approach. *Trends in Cognitive Sciences, 5*, 164-170.

Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. (1999). Working memory, short-term memory, and general fluid intelligence: A latent-variable approach. *Journal of Experimental Psychology: General, 128*, 309-331.

Halford, G. S., Cowan, N., & Andrews, G. (2007). Separating cognitive capacity from knowledge: A new hypothesis. *Trends in Cognitive Sciences, 11*, 236-242.
Halford, G. S., Wilson, W. H., & Phillips, S. (1998a). Processing capacity defined by relational complexity: Implications for comparative, developmental, and cognitive psychology. *Behavioral and Brain Sciences, 21*, 803-831.
Halford, G. S., Wilson, W. H., & Phillips, S. (1998b). Relational complexity metric is effective when assessments are based on actual processes. *Behavioral and Brain Science, 21*, 848-864.
Jensen, A. (1987). The g beyond factor analysis. In R. R. Ronning, J. A. Glover, J. C. Conoley & J. C. Witt (Eds.), *The Influence of Cognitive Psychology on Testing* (Vol. 3, pp. 87-142). Hillsdale, NJ: Lawrence Erlbaum Associates.
Kane, M. J., & Engle, R. W. (2003). Working-memory capacity and the control of attention: The contributions of goal neglect, response competition, and task set to Stroop interference. *Journal of Experimental Psychology: General, 132*, 47-70.
Kane, M. J., Hambrick, D. Z., Tuholiski, S. W., Wilhelm, O., Payne, T., & Engle, R. W. (2004). The generality of working memory capacity: A latent-variable approach to verbal and visuospatial memory span and reasoning. *Journal of Experimental Psychology: General, 133*, 189-217.
Kyllonen, P. C., & Christal, R. E. (1990). Reasoning ability is (little more than) working-memory capacity?! *Intelligence, 14*, 389-433.
Lohman, D. F., & Ippel, M. J. (1993). Cognitive diagnosis: From statistically based assessment toward theory-based assessment. In N. Frederiksen, R. J. Mislevy & I. I. Bejar (Eds.), *Test theory for a new generation of tests* (pp. 41-70). Hillsdale, NJ: Lawrence Erlbaum Associates.
Mackintosh, N. J., & Bennett, E. S. (2003). The fractionation of working memory maps onto different components of intelligence. *Intelligence, 31*, 519-531.
Schweizer, K. (1996). The speed-accuracy transition due to task complexity. *Intelligence, 22*, 115-128.
Schweizer, K. (2001). The role of mechanisms under high processing complexity. *European Psychologist, 6*, 133-143.
Spilsbury, G., Stankov, L., & Roberts, R. (1990). The effect of a test's difficulty on its correlation with intelligence. *Personality and Individual Differences, 11*, 1069-1077.
Stankov, L. (2000). Complexity, metacognition and fluid intelligence. *Intelligence, 28*, 121-143.
Stankov, L., & Crawford, J. D. (1993). Ingredients of complexity in fluid intelligence. *Learning and Individual Differences, 5*, 73-111.
Unsworth, N., & Engle, R. W. (2007). On the division of short-term and working memory: An examination of simple and complex span and their relation to higher order abilities. *Psychological Bulletin, 133*, 1038-1066.
Waltz, J. A., Knowlton, B. J., Holyoak, K. J., Boone, K. B., Mishkin, F. S., de Menezes Santos, M., et al. (1999). A system for relational reasoning in human prefrontal cortex. *Psychological Science, 10*, 119-125.