



**Relational processing and working memory capacity in  
comprehension of relative clause sentences**

**Author**

Andrews, Glenda, Birney, Damian, Halford, Graeme S

**Published**

2006

**Journal Title**

Memory and Cognition

**Downloaded from**

<http://hdl.handle.net/10072/14212>

**Link to published version**

<https://doi.org/10.3758/BF03193275>

**Griffith Research Online**

<https://research-repository.griffith.edu.au>

Relational Processing and Working Memory Capacity in Comprehension of Relative Clause

Sentences

Glenda Andrews

School of Psychology

Griffith University, Gold Coast Campus, Queensland

Australia

&

Damian Birney

School of Psychology

University of Sydney

Australia

&

Graeme S. Halford

School of Psychology

Griffith University, Mt Gravatt Campus, Queensland

Australia

Running head: RELATIVE CLAUSE SENTENCES

Address reprint requests to Glenda Andrews, School of Psychology, Griffith

University, PMB 50, Gold Coast Mail Centre 9726, Australia. Telephone: 61 7 55528613.

Fax: 61 7 55528291. E-mail: g.andrews@griffith.edu.au.

## Abstract

Previous research indicates that the cognitive load imposed by tasks in various content domains increases with the complexity of the relational information processed. Sentence comprehension entails processing noun-verb relations to determine *who did what to whom*. The difficulty of object-extracted relative clause sentences might stem from the complex noun-verb relations they entail. Across three studies, participants read 16 types of object- and subject-extracted relative clause sentences at their own pace then responded to a comprehension question for each sentence. Relational processing was assessed using a premise integration task or a Latin Square task. These tasks predicted comprehension of object-relatives before and after controlling for subject-relatives. Working memory (WM) capacity was assessed using Reading Span or forward and backward digit span tests. WM tasks predicted comprehension of object-relatives before but not after controlling for subject-relatives. Comprehension of object-relatives relied more heavily on a domain-general capacity to process complex relations than on WM capacity.

**Key words:** Relative clause sentences; cleft sentences; comprehension; processing load; verbal working memory; relational complexity; integration cost; thematic role assignment; post-interpretative processing

## Relational Processing and Working Memory Capacity in Comprehension of Relative Clause Sentences

Numerous studies have demonstrated that object-extracted (object-relative) sentences are more difficult to understand than comparable subject-extracted (subject-relative) sentences (e.g., Blaubergs & Braine, 1974; Blumenthal, 1966; Fodor & Garrett, 1967; Ford, 1983; King & Just, 1991; Marks, 1968; Miller & Isard, 1964; Traxler, Morris, & Seely, 2002; Traxler, Williams, Blozis, & Morris, 2005) and that object-cleft sentences are more difficult than subject-cleft sentences (Caplan, Alpert, & Waters, 1999; Gordon, Hendrick, & Levine, 2002; Waters & Caplan, 2001). Examples are shown in Tables 1 and 2. The greater difficulty of object-relatives and object-clefts is usually attributed to their linguistic complexity, and the processing loads that flow from this. We report three experiments in which we used an individual differences approach to examine the nature of the cognitive demands imposed by object-relatives and object-clefts. Individuals' cognitive capacity was assessed using measures based on two approaches, working memory (Just & Carpenter, 1992) and relational complexity (Halford, Wilson, & Phillips, 1998). These measures were compared in terms of the extent to which they accounted for variance in comprehension of object-relatives and object-clefts.

Many potential sources of linguistic complexity have been identified. Frazier (1985) and Gibson (2000) provided reviews of sentence complexity metrics proposed during the 1960's. More recently, Gibson (2000) proposed Dependency Locality Theory (DLT), which includes metrics based on memory cost and integration cost. In DLT, working memory (WM) resources are required for storage of information about the structure that has already been processed, and for integration of the current word into the structure. Integration cost is imposed as discourse referents such as nouns and verbs are incorporated into the mental representation that is constructed during sentence processing. Integration cost is quantified in

terms of the number of new discourse referents (nouns and verbs) that intervene between the first occurrence and the point of integration. It varies across the sentence, but is usually highest at the verbs. Integration cost is higher for object-relatives than for subject-relatives because of the greater distance between the dependents in object-relatives.

Just and Carpenter and their colleagues (Carpenter, Miyake, & Just, 1994; Daneman & Carpenter, 1980; Daneman & Merickle, 1996; Just & Carpenter, 1992; King & Just, 1991) have investigated the association between sentence comprehension and WM capacity using an individual differences approach. In their approach, sentence comprehension performance depends on the relation between an individual's WM capacity and the load imposed by the sentences. WM capacity is conceptualized as a finite resource that mediates the simultaneous storage and computation of information. The WM constructs of Just, Carpenter and their colleagues and of Gibson (2000) are similar in that they each assume that there is a single pool of resources that can be flexibly allocated to storage and computational/integration functions. Just and Carpenter (1992) likened their WM construct to the central executive component of Baddeley's (1986) model, which also includes two slave systems, the Phonological Loop, and the Visuo-spatial Sketchpad. Just and Carpenter assess WM capacity using complex span tests such as Reading Span (Daneman & Carpenter, 1980) in which participants must retain a number of sentence-final words while reading a set of sentences

Carpenter, et al. (1994) argued that the WM load imposed by comprehension is directly related to sentence complexity. Although they did not provide a complexity metric, they identified five sources of complexity, which along with sentence length determine cognitive load. The first is the number of thematic roles associated with a single verb. Verbs with three roles (agent, patient, recipient) impose a higher computational demand than verbs with two roles (agent, patient). Complexity is higher in sentences with two verbs than in

those with a single verb. Complexity is higher if roles occur in non-canonical order (e.g., if the noun that occupies the patient role precedes the noun in the agent role). Complexity is higher if the first noun must be retained while other roles are computed. This occurs in sentences with an embedded structure (e.g., 3-, 4-, 5-role object-relatives in Table 1), in which the relative clause interrupts the main clause. It does not occur in sentences with a right-branching structure (e.g., subject-relatives in Table 1). Finally, complexity is greater if a noun plays different thematic roles in different clauses. Sheldon (1974) called this non-parallel function.

King and Just (1991) contrasted object-relatives such as (1) and subject-relatives such as (2) below. Object-relatives impose a higher WM demand because they entail two sources of complexity (non-parallel function, non-canonical order) that are not present in subject-relatives. Sentences were presented visually, one word at a time. On-line reading times were recorded, and end-of-sentence comprehension questions were presented. WM capacity was assessed using the Reading Span task and participants were classified into High Span and Low Span groups.

*The reporter that the senator attacked admitted the error.* (1)

*The reporter that attacked the senator admitted the error.* (2)

The results for the 1-sentence trials showed that High Span participants had better comprehension overall than Low Span participants, but the difference was greater for object-relatives. These data and the reading times were interpreted as showing that the object-relatives imposed a higher demand on verbal WM than subject-relatives (King & Just, 1991).

Traxler et al. (2005) monitored participants' eye-movements while they read reversible object- and subject-relative sentences similar to (1) and (2). WM capacity was assessed using a Reading Span task. In contrast to King and Just's (1991) results, Traxler et al's results provided no evidence that WM capacity moderated the magnitude of the

difference between object- and subject-relatives. However, there was some evidence that WM capacity moderated the effects of a semantic variable (animacy of the sentential subject) on the object-relative penalty. Specifically, participants with higher WM capacity were better able to make use of semantic information to reduce the difficulty of object-relatives.

Other researchers have also questioned the view that syntactic processing draws on the same verbal WM resource as other language tasks. For example, Lewis' (1996) model includes a limited capacity memory that is specialised for syntactic relations. Similarly, Caplan and Waters (1999; Waters & Caplan, 2004) proposed a WM subsystem dedicated to interpretive processing, described as first-pass, obligatory processing that operates at an unconscious level and that is used to extract initial meaning from the linguistic signal. Interpretive processing can be assessed using on-line measures such as self paced word-by-word reading or listening times, word and phoneme monitoring, and lexical decision times. Post-interpretive processing is a conscious, controlled type of processing that involves using meaning for other purposes such as reasoning or entering information into long-term memory. It is more likely to be tapped by off-line procedures such as end-of-sentence acceptability judgements and comprehension questions, which allow for re-analysis or review of the sentence prior to the response. Post-interpretive processing draws on a more general verbal working memory, such as that proposed by Just and Carpenter (1992).

Caplan, Waters and their colleagues have reported many studies relevant to the links between WM capacity, interpretive processing, and post-interpretive processing. Some studies have included patients with aphasia or dementia of the Alzheimer's type (DAT) whose WM capacities are impaired. For example, Rochon, Waters, and Caplan (2000) assessed patients with DAT and elderly controls on a WM battery that included tests of primary memory (simple span tests) and of the central executive (backward digit span,

counting span, tracking task) and two off-line sentence comprehension tests (sentence-picture matching and video-verification). Interpretive processing was operationalized in terms of syntactic complexity. Sentences in which the thematic roles were in non-canonical order were considered as more syntactically complex than sentences in which the order was canonical. Post-interpretive processing was operationalized in terms of number of propositions. The DAT patients had reduced spans and impaired central executive processing relative to controls, but they showed no impairment on comprehension of the more syntactically complex sentences. Central executive processing was however associated with poorer comprehension of sentences with two propositions. Rochon et al. concluded that the WM impairments of DAT patients were related to their ability to map sentence meanings onto events in the world (i.e., to post-interpretive processing) rather than to their ability to assign syntactic structure initially (i.e., interpretive processing). However they acknowledged several difficulties associated with the distinction between interpretive and post-interpretive processing. For example, post-interpretive processes can sometimes enter into the computation of thematic roles, and the number of propositions might depend on the mental models represented by individual listeners or readers. Thus the theoretical distinction between interpretive and post-interpretive processing is not clear-cut. Nor can it be assumed that off-line measures necessarily assess post-interpretive processes only, given that Rochon et al. used off-line tasks to assess both interpretive and post-interpretive processing.

Using an on-line procedure, Waters and Caplan (2002) also found no relation between WM capacity and interpretive processing in DAT patients and elderly controls. They used an auditory moving window presentation to assess syntactic processing in object- and subject-relatives and object- and subject-clefts. Listening times for each phrase in the sentences were recorded. There was no evidence for an association between WM capacity and the effects of syntactic complexity. Waters and Caplan interpret results such as these as

support for a separate WM subsystem that is dedicated to interpretive processing, but they allow a role for more general WM resources in post-interpretive processing. There is some common ground between the positions of Waters and Caplan and of Traxler et al., (2005) in that neither agrees with Just and Carpenter's (1992) claim that verbal WM is involved in syntactic processing, although they each allow a role for WM in other aspects of sentence processing.

An unresolved issue is the extent to which Just and Carpenter's definition of WM (the capacity for the simultaneous storage and computation of information) captures the demands imposed by object-relative sentences. According to Traxler et al., (2005), the source of the object-subject difference might lie in the binding of constituents to positions in the syntactic tree and the assignment of thematic roles, rather than the maintenance of information. In the absence of lexical or semantic cues, readers experience difficulty assigning thematic roles either because of competition between nouns for argument slots or because an initial incorrect assignment must be abandoned and the sentence reanalyzed. Reanalysis is more likely with object-relatives because readers initially treat the sentential subject as the subject of the relative clause verb. This yields the correct assignment of thematic roles for subject relatives, but it results in misanalysis of object-relatives. Traxler et al's position implies that object-relatives impose a greater demand for computational resources as opposed to storage resources. If so, tasks that assess computational processing might be better predictors of comprehension of object-relatives than complex span tasks (e.g., Reading Span), which require simultaneous storage and computation, and in which item difficulty reflects the increasing storage demands imposed as set size increases. Assessment of computational capacity more independently of storage capacity requires tasks that incorporate manipulations of computational complexity rather than storage load. Thus a metric of cognitive complexity is needed.

Another unresolved issue is whether syntactic processing draws on a specialised resource as proposed by Caplan and Waters (1999) and Lewis (1996) or more general purpose resources. Just and Carpenter's (1992) WM construct is more general, in that it is utilized by many language tasks, but they make no claims about its application to non-linguistic domains.

The involvement of domain general resources in sentence processing is consistent with Larkin and Burns' (1977) results. They demonstrated that difficulty with embedded structures is not confined to linguistic contexts. Participants heard a list of stimuli and then attempted to recall them in pairs (i.e., first item paired with last item, second item paired with next to last item, etc). This simulates one aspect of sentence comprehension, noun-verb pairing. Participants completed the pairing task with lists of four different lengths (4, 6, 8, and 10), which correspond to one, two, three and four levels of embedding. There were four different stimulus types. Three conditions (digits, digits and letters, nouns and verbs) entailed unravelling embedded structure, whereas the baseline (sentences) condition also required comprehension. That the effect of embedding was significant in all conditions argues against a purely linguistic account of the difficulty.

Brain imaging studies suggest that comprehension of complex sentences involves the cortical regions traditionally associated with language processing as well as more domain general regions. PET (Caplan, Alpert, & Waters, 1998, 1999; Stromswold, Caplan, Alpert, & Rauch, 1996; Waters, Caplan, Alpert, & Stanczak, 2003) and fMRI technologies (Constable, et al., 2004; Just, Carpenter, Keller, Eddy, & Thulborn, 1996) indicated that compared to subject-relatives and subject-clefts, object-relatives and object-clefts were associated with greater rCBF in Broca's area (Caplan et al., 1998; 1999; Stromswold et al., 1996; Waters, et al., 2003) as well as homologous regions in the right hemisphere and the dorsolateral prefrontal cortex (DLPFC) bilaterally (Just et al., 1996). The DLPFC is involved in tasks that

require working memory, planning, and executive control (Stuss & Levine, 2002). The DLPFC has also been shown to be important in integrating relations in transitive inference and matrix completion tasks that have non-linguistic content (Christoff et al., 2001; Kroger et al., 2002; Waltz, et al., 1999).

Thus, in the current research we employed a domain general approach to cognitive complexity (described next) as the basis for developing measures of computational capacity. These measures will be used (along with WM tasks) as predictors of comprehension of object-relative sentences.

*Relational Complexity (RC) theory* (Halford, Wilson, & Phillips, 1998)

RC theory proposes that many higher cognitive processes can be characterized as involving complex relations. Complexity is defined by the number of arguments or entities related in a single decision. Each argument corresponds to a dimension and number of dimensions corresponds to the number of interacting variables that constrain responses or decisions. A metric of relational complexity is defined. Unary relations have a single argument as in class membership, dog(fido). Binary relations have two arguments as in larger-than(elephant, mouse). Ternary relations have three arguments as in arithmetic-addition(2,3,5). Quaternary relations such as proportion have four interacting components as in  $2/3 = 6/9$ , whereas quinary relations entail five interacting components.

RC is related to processing load, which increases with the complexity of relations processed. Thus, quinary relations impose a higher load than quaternary relations, which impose a higher load than ternary relations, and so on. Processing load can be indexed by secondary task indicators (Maybery, Bain, & Halford, 1986) or by measures derived from brain imaging studies (Christoff et al., 2001; Kroger, et al., 2002). The effective relational complexity for a cognitive process is the least complex relation required to represent the process. This can be determined algorithmically by a decomposition and recomposition

technique (Halford et al., 1998, Section 3.4.3; Phillips & Nikki, 2003). Where tasks entail more than one step, the processing complexity of the task is the relation that must be represented to perform the most complex step involved in the task, using the least demanding strategy available to humans for that task (Halford et al., 1998, Section 2.1).

RC theory specifies human processing capacity limitations (Halford et al., 1998). On average, adult humans can process quaternary relations, that is, four variables can be related in a single cognitive representation, though a minority can process quinary relations under optimal conditions. Halford, Baker, McCredden and Bain (2005) provided empirical support for this limitation.

Many of the predictions derivable from RC theory depend on the way tasks can be decomposed. Processing of very complex concepts depends on reducing effective complexity so as to make optimal use of available processing capacity. RC theory proposes two strategies for complexity reduction; conceptual chunking and segmentation. Conceptual chunking is the recoding of concepts into fewer dimensions. For example, velocity defined as  $velocity = distance/time$ , entails a ternary relation but can be recoded into a unary relation as when speed is indicated by the position of a pointer on a dial. However the reduction in load comes at a cost. Access to the relations that make up the concept is lost. For example, if velocity is represented as a unary relation, changes in velocity as a function of time and/or distance cannot be computed. In general, variables can be chunked if the relation between them does not need to be considered. For example, to establish that an element, A differs from two others, B and C, B and C can be chunked because the relation between them is not relevant to the decision (Chalmers & Halford, 2003).

Complexity reduction can also be accomplished through segmentation of tasks into less complex steps, which can be processed serially (Halford et al., 1998). The ease of segmentation might underlie the greater difficulty of object-relative as compared to subject-

relative sentences, as explained next.

The ultimate goal of comprehension is meaning interpretation and this involves assigning nouns to thematic roles (e.g., agent, patient, recipient) of the verbs to determine *who did what to whom* (Caplan & Waters, 1999). Clearly, thematic role assignment is central to comprehension. It is a demanding process for humans (Linebarger, Schwartz, & Saffran, 1983) and it imposes a high computational load (Haarmann, Just, & Carpenter, 1997). Thematic role assignment can be characterized as processing the relations between nouns and verbs, and the complexity of these relations might be a factor in the difficulty of object-relatives. The higher computational load associated with object-relatives might occur because these noun-verb relations are difficult to segment.

Consider the 5-role sentences in Table 1. The meaning of the subject- and object-relative forms can be expressed as three propositions: like(actor, teacher); watch(teacher, clown); laugh(clown). There are three verbs and five roles to be filled (3 agents, 2 patients). Notice also that *teacher* and *clown* appear in two propositions. In the object-relative form, the three nouns occur prior to any of the verbs. Furthermore, the roles occur in non-canonical order. Patient nouns, *teacher* and *clown* occur before the agents of the verbs. Role assignment cannot be finalized until the verbs are encountered. When the first verb, *liked*, is encountered there are three nouns waiting to be assigned to their roles. Based on semantic content alone each would be equally acceptable as the agent or patient of *liked* and also of the subsequent verbs, *watched* and *laughed*, which occur immediately afterwards. This concentration of verbs toward the end of the sentence and the fact that two nouns are related to two verbs make segmentation, as defined in RC theory, difficult and create pressure for thematic role assignments to be considered in the same step. This imposes a high processing load.

By contrast, in subject-relatives, the nouns and verbs are distributed throughout the

sentence and the agent and patient roles occur in canonical order. This makes segmentation, as defined in RC theory, easier because the propositions can be processed one at a time. For example, *actor* and *teacher* can be assigned to the agent and patient roles of *liked* before the subsequent propositions, *watch(teacher, clown)* and *laugh(clown)* are encountered. The thematic relations between nouns and verbs can be processed as they are encountered, thereby avoiding concentration and the concomitant processing load. Thus in the RC approach, the load imposed by sentence comprehension depends on the extent to which the task of assigning nouns to thematic roles can be segmented. Segmentation is easier in subject-relatives and more difficult in object-relatives. Because object-relatives are resistant to segmentation, the complexity of the noun-verb relations is higher.

This focus on the thematic role assignment process is consistent with Carpenter et al. (1994) in that the five sources of sentence complexity they identified influence either the total number of role assignments or the ease with which they can be made. However, RC and WM approaches differ in their assumptions about the precise nature of the load imposed by object-relatives. As noted earlier, the WM approach assumes that the five sources of complexity impose demands for simultaneous storage and computation of information. RC theory is more consistent with Traxler et al's (2005) view that maintenance of information is not the source of the object-subject difference, but that computational demand is crucial. In DLT (Gibson, 2000), computational load (integration cost) fluctuates across the sentence but is maximal at the verbs. Integration cost depends on the distance between the noun and the point of integration. Integration cost in DLT and processing load in RC appear to have some commonality in that both entail processing and both tend to accumulate over object-relative sentences. RC attributes the accumulation to lack of opportunity for segmentation, whereas DLT attributes it to distance, but distance between noun and verb implies the requirement to process additional nouns and (possibly) verbs before the assignment of the first noun to the

appropriate role can be completed. Thus distance and lack of opportunity for segmentation appear to be referring to similar phenomena.

DLT and RC theory differ in terms of the phenomena that influenced their development. Whereas DLT is based on analyses of linguistic structure, RC was motivated by analyses of reasoning and its development. Consequently, RC theory incorporates processes such as conceptual chunking, segmentation, and analogical mapping that apply across many content domains. The RC metric applies to conscious controlled processing, rather than first-pass obligatory processes. In this sense, it resembles post-interpretive processing in Caplan and Waters' (1999, Waters & Caplan, 2001, 2004) approach. In terms of Baddeley's (1986) WM model, RC theory corresponds most closely to the computational functions of the central executive. As noted previously, Just and Carpenter's (1992) WM construct has also been likened to the central executive. However, Baddeley (1993) has claimed that complex span measures might not accurately assess the central executive because they reflect unspecified contributions of the Phonological Loop and aspects of long-term memory. While RC theory assumes that the slave systems contribute to cognitive processing, the capacity for which it offers a metric corresponds solely to the central executive.

The RC metric (Halford et al., 1998) has been successfully applied in recent years to topics in cognitive development (e.g., Andrews & Halford, 1998, 2002; Andrews, Halford, Bunch, Bowden, & Jones, 2003; Halford, 1993; Halford, Andrews, Dalton, Boag, & Zielinski, 2002a; Halford, Andrews, & Jensen, 2002b), reasoning in adults (Birney & Halford, 2002; Birney, Halford, & Andrews, under review; Zielinski, Goodwin, & Halford, under review), and applied areas such as mathematics education (English & Halford, 1995) and air traffic control (Boag, Härtel, & Halford, in press).

Andrews, Halford, and Prasad (1998) provided preliminary evidence for a link

between relational processing in non-linguistic domains and sentence comprehension. Object- and subject-relative sentences similar to 2- and 3-role sentences in Table 1 were presented to 4- to 8-year-old children. Each sentence was followed by a single comprehension question (e.g., *Who walked?*). Children's WM capacity was assessed using a Listening Span task, which is analogous to the Reading Span task. Their capacity to process complex relations was assessed using two tasks, hierarchical classification and transitivity, which have been shown to entail ternary relations (Andrews & Halford; 1998, 2002; Halford, et al., 1998). As expected, the object-relative sentences (especially those with three roles) were more difficult than subject-relatives and performance on all tasks improved with age. Regression analyses showed that the relational processing tasks accounted for variance in comprehension independently of age and Listening Span. These results suggest that comprehension of complex sentences and relational processing in non-linguistic domains involve common processes. The comprehension questions assessed understanding of the relations between the noun and verbs in the sentence, whereas hierarchical classification and transitivity tests involved relations of a different type, but of similar complexity. The current research will extend these preliminary findings by determining whether performance on tasks based on this domain general approach to complexity predicts adults' comprehension of complex object-relative and object-cleft sentences.

In Experiments 1, 2, and 3, comprehension of object- and subject-relatives and/or object- and subject-clefts was assessed using end-of-sentence comprehension questions. Participants also completed predictor tasks based on RC and WM approaches. The relational processing hypothesis (based on RC theory) was that comprehension of object-relatives and object-clefts would be predicted by performance on tasks from non-linguistic domains (*n*-term task in Experiments 1 and 2, Latin Square Task, LST in Experiment 3), which involve complex relations. Significant variance in comprehension should be accounted for even after

controlling for comprehension of subject-relatives and subject-clefts. The WM hypothesis (based on Just & Carpenter's, 1992 WM approach) was that comprehension of object-relatives and object-clefts would be predicted by WM capacity as indexed by Reading Span (Experiments 1 and 2) even after controlling for comprehension of subject-relatives and subject-clefts. In Experiment 3, this hypothesis was tested using digit span forward (FDS) and digit span backwards (BDS) as measures of short-term memory (STM) or WM.

### Experiment 1

Experiment 1 investigated whether comprehension of object-relatives would be predicted by performance on a relational processing task, *n*-term premise integration. The *n*-term task included items at three levels of complexity and is an extended version of a transitive inference task. Transitive inferences have the form if  $a R b$  and  $b R c$ , then  $a R c$ , where  $R$  is a transitive relation and  $a$ ,  $b$ , and  $c$  are the elements related. Previous research (Trabasso, 1975) showed that transitive inferences are made by constructing an ordered array of the elements,  $a R b R c$ . Once the array is constructed, the relation between  $a$  and  $c$  is apparent. Construction of the array involves integrating two binary relations,  $R(a,b)$  and  $R(b,c)$  into the ordered triple,  $R(a,b,c)$ . Halford, et al. (1998, Section 3.4.3) demonstrated using their decomposition and recomposition technique that the 3-term transitive inference task is ternary-relational and cannot be reduced to a series of binary relations. There is evidence that premise integration is the point of maximum cognitive load (Maybery, et al., 1986) and is capacity limited in children (Halford, Maybery, & Bain, 1986). The study by Waltz et al. (1999) suggested that successful integration of relations in a transitive inference task might depend on an intact DLPFC.

Expanding on this previous work, we designed a task in which participants constructed ordered series of three, four, and five elements, based on premise relations. We reasoned that if integrating two binary relations into an ordered triple to assign the elements

to three slots is equivalent to a ternary relation, then assigning elements to four (five) slots should approximate the complexity of a quaternary (quinary) relation, provided the premises are integrated in a single decision. The basic procedure involved constructing a sequence of  $n$  alphabetic letters, whose order was consistent with the premise information. Premises consisted of pairs of alphabetic letters with a  $>$  or  $<$  relation defined between the letters as shown in Figure 1. For example, two premises  $F > P$  and  $P > M$  are necessary and sufficient to define the ordered triple  $F > P > M$ . The 4-term (5-term) version would require that three (four) premise pairs be provided. If this basic procedure were used it would have been possible for participants to segment the 4- and 5-term items. For the 4-term items for example, three premises  $F > P$ ,  $P > M$ ,  $M > T$ , would be provided and participants would be required to construct the sequence  $F > P > M > T$ . This could be accomplished by first constructing the triple  $F > P > M$  and then concatenating the final element,  $T$  to extend the series. The 5-term series could be completed by first constructing a 3-term series and then concatenating the fourth and fifth elements, one at a time. Three modifications were introduced to minimise the likelihood of this type of segmentation. The premises were presented in random spatial order rather than in an order that corresponded to the correct final order of the elements. An additional (redundant) premise specifying the relation between two nonadjacent elements in the final sequence was included. A mixture of  $>$  and  $<$  relations was used in the premises. These modifications should make concatenation a less attractive strategy, thereby constraining participants toward considering more premises in the same decision. Extraneous storage demand was minimised by having premises continuously available. Thus the task should impose a computational, rather than storage load. If comprehension of object-relatives involves processing of complex relations, then  $n$ -term scores should account for significant variance in comprehension of object-relatives.

Participants completed a version of the Reading Span test (Daneman & Carpenter,

1980), which requires participants to read sets of sentences and then attempt to recall the final word of each sentence. The test reflects the view that computational processing (reading) and storage (maintaining the final words) draw on the same resource pool. Based on the WM approach and King and Just's (1991) results (described above), we expected that Reading Span scores would account for significant variance in comprehension of object-relatives, although the results of Traxler et al., (2005) suggest otherwise.

### *Method*

#### *Participants*

The initial sample consisted of 68 students (53 females, 15 males) enrolled in first year psychology who participated in return for course credit. The data of three participants whose first language was not English were excluded. Two participants did not complete the reading span test. For the *n*-term task, data of one participant were lost because of a computer malfunction.

#### *Apparatus and procedure*

Three IBM compatible Optima (2 x 386, 1 x 486) computers with SVGA colour monitors were used to administer sentence comprehension and *n*-term tasks. Testing was spread over two sessions with a total duration of approximately 1.5 hours, with session order counterbalanced. In one session, groups of three to ten participants completed the reading span test. In the other, they completed sentence comprehension and *n*-term tasks in counterbalanced order.

*Sentence comprehension task.* There were 96 semantically reversible sentences, divided into two sets of 48. Each set contained eight instances of object-relative and subject-relative sentences requiring three, four, and five role assignments, as shown in Table 1. Across the sets, each sentence content was used in both object- and subject-relative forms. Each comprehension question referred to a single noun-verb relation. There were five, six

and eight question types for the 3-, 4-, and 5-role sentences respectively. For example, for the 3-role sentences shown in Table 1 the five questions were, *Who touched? Who walked? Who was touched? What did the duck do? What did the monkey do?* The questions for each sentence were randomly selected for each participant from the available options.

The six sentence types were intermixed and presented in a different random order for each participant. Sentences were displayed, one at a time, on the upper half of the computer screen in red Times Roman lettering (font size 24) on a grey background. Participants read each sentence carefully at their own pace and pressed ENTER when they thought they understood the sentence. The sentence was then replaced by a comprehension question. Participants responded by typing a single noun or verb.

*N-term task.* Items at three levels of complexity were generated. Sequences of three, four, and five letters were formed by selecting letters at random without replacement. A > relation was imposed on adjacent elements such that the first letter was greater than the second, the second greater than the third, and so on. These sequences were the correct descending orders that participants were required to construct. A set of premise relations containing  $n-1$  adjacent pairs and one pair of non-adjacent letters was constructed for each sequence. A combination of < and > signs was used in the premises for each sequence. Examples of premises and the corresponding 3-, 4-, and 5-term sequences are shown in Figure 1.

The screen was divided into two sections by a white vertical line approximately 8 cm (3 in) from the left. Premise relations were displayed in white upper case letters (Times Roman, font size 24) on a grey background in the left section, in a different randomly determined vertical order for each participant. On the right side was a row of  $n$  boxes with white outlines, and white > signs between them. Participants' task was to mentally combine the premise relations to construct a descending sequence of letters of length  $n$ , and to enter

the sequence using the keyboard. At the outset of each item, the left-most box was highlighted in white. The first letter typed appeared in red in the highlighted box. It remained in view momentarily before being replaced by an asterisk. Subsequent boxes were highlighted only when a valid letter (one that appeared in the premises) was typed in the preceding box. Thus during construction of the sequence, a maximum of one letter was visible in the response boxes at any one time. When  $n$  letters had been entered, the entire sequence was displayed. Participants were advised to construct the entire sequence mentally before beginning to type, because they were unable to reorder the letters once they had been entered. No time limit was imposed. The items were blocked according to series length. Within each block, one practice item was followed by ten test items yielding maximum scores of 10 for each level and 30 for the three levels combined.

*Reading Span.* The 44 sentences were 11 to 16 words in length. The final words were one-syllable, high frequency, concrete nouns. Half the sentences were made nonsense by reversing the order of the last four to six pre-terminal words. Turner and Engle (1989) used this method. For example, *The possum took the apple from the sill and then disappeared into the night* became *The possum took the apple from the sill the into disappeared then and night*. Sentences were randomly assigned to three sets at each set size (2, 3, 4, 5) and an additional practice set at set size 2. Each set of sentences was printed in Times Roman lettering (font size 18) and copied onto a separate transparency. Response sheets corresponding to three different orders of set presentation were provided. The sheets had spaces for *make sense* judgements on one side, and *recall of final words* on the reverse side.

Instructions were read aloud by the experimenter. The practice set was administered to ensure that participants understood the procedure and could easily read the sentences that were presented on a screen using an overhead projector. Sentences were exposed one at a time for approximately 8 seconds through a cardboard window. Participants read the

sentence, recorded their *make sense* judgements by circling *Yes* or *No* on the response sheet, then looked up immediately. Subsequent sentences were presented as soon as participants were ready. When all sentences in the set had been presented, the experimenter said *Recall* and participants attempted to record the final words in any order on the sheet provided. Three 2-sentence sets were presented first, and set size was increased systematically thereafter. Reading Span scores were calculated by deducting the number of errors on *make sense* judgements from the number of final words correctly recalled. We used this continuous scaling rather than WM classifications (e.g., high span, low span) because the latter are known to be unstable (Waters & Caplan, 2003).

### *Results and Discussion*

Sixty-two participants completed sentence comprehension, *n*-term, and reading span tasks. After deletion of one outlier with a large standardized residual, sample size was 61. Table 3 shows the descriptive statistics and correlations among *n*-term, reading span, and comprehension of object- and subject-relative sentences (%'s correct averaged across number of roles). Consistent with predictions, *n*-term, and reading span were each significantly correlated with comprehension accuracy of object-relatives. A multiple regression analysis showed that Reading Span and *n*-term accounted for 46.1% of variance in comprehension of object-relatives, Multiple  $R = .68$ ,  $F(2, 58) = 24.84$ ,  $p < .001$ . *N*-term accounted for 37.85% of variance independently of reading span ( $p < .001$ ). Reading Span accounted for 3.65% of variance independent of *n*-term ( $p = .055$ ). There was little shared variance (2.87%).

Table 3 shows significant associations between comprehension of object- and subject-relatives, and marginally significant associations of subject-relatives with both *n*-term and Reading Span. A second analysis examined the extent to which *n*-term and Reading Span account for comprehension of object-relatives when these associations are taken into

account. Reading Span, *n*-term, and comprehension of subject-relatives accounted for 52.7% of variance in comprehension of object-relatives, Multiple  $R = .73$ ,  $F(3, 57) = 21.15$ ,  $p < .001$ . *N*-term accounted for 30.68% of variance independently ( $p < .001$ ). Subject-relatives accounted for 6.54% of variance independently ( $p < .05$ ) indicating the importance of expertise with relative clauses. Reading Span no longer contributed significant unique variance, but would have contributed to the shared variance (12.97%). This confirms that comprehension of object- and subject-relatives involves some similar demands, but there are additional demands associated with object-relatives. *N*-term captures some of this.

In summary, the regression analyses provided support for the relational processing hypothesis. *N*-term, which required processing of complex relations in a non-linguistic domain accounted for variance in comprehension of object-relatives even after controlling for comprehension of subject-relatives and WM capacity. There was less support for the WM hypothesis. Reading Span accounted for variance in comprehension of object-relatives after controlling for processing of complex relations. This is consistent with earlier findings (e.g., King & Just, 1991) that comprehension of object-relatives is related to individual differences in WM capacity. However the association with Reading Span disappeared when comprehension of subject-relatives was controlled. This also provides preliminary evidence for a dissociation of the two conceptualisations of capacity, one based on simultaneous storage and computational processing, and the RC account which focuses primarily on computational processing.

## Experiment 2

Experiment 2 broadened the range of sentence types to include object-cleft and subject-cleft sentences with three, four, and five role assignments (as shown in Table 2) in addition to restrictive relative clause sentences. Another purpose was to assess the replicability of the non-significant correlation between *n*-term and Reading Span. In

Experiment 1, the *n*-term task was computer-administered whereas Reading Span was administered to groups of participants using manual presentation. These method differences might have masked a significant association between the tasks. A computer-administered version of Reading Span was used in Experiment 2. We expected significant associations between *n*-term and comprehension of object-relatives/clefts (relational processing hypothesis) and between Reading Span and comprehension of object-relatives/clefts (WM hypothesis).

### *Method*

#### *Participants*

The participants were first-year psychology students who participated in return for course credit. All were native speakers of English. Analyses were based on the 68 participants (43 females, 25 males) who provided complete data.

#### *Apparatus and procedure*

Four 233 MHz Pentium II personal computers with 35.7cm UVGA (1024 × 768) colour monitors were used to administer the two sentence comprehension tasks (restrictives, clefts), *n*-term and Reading Span tasks. The tasks were administered in two sessions lasting approximately 50 minutes each. The clefts and restrictives were completed in different sessions.

Comprehension of restrictive relative clause and cleft sentences was assessed using the same procedure as in Experiment 1. The comprehension questions for the clefts were similar to those used for restrictives except that for the 3-role clefts, all questions required a noun response, because there is only one verb in the relevant part of the sentences. The *n*-term task was identical to that described for Experiment 1.

The Reading Span task was converted to computer-administered format using DMDX software (Forster & Forster, 1999). Instructions were displayed in green lettering on

a dark blue screen. The sentences were in white lettering. A practice set of two sentences was presented followed by feedback to ensure participants understood the requirements. Sentences were presented one at a time and participants recorded their "make sense" judgements by pressing keys designated as "Yes" and "No". Each sentence remained on the screen until a key press was registered or until 10 seconds had elapsed, whichever occurred first. After each set of sentences, a "Recall" signal appeared on the screen and participants attempted to write the final word of each sentence in the set on the sheet provided. Scoring was the same as in Experiment 1. A recent study ( $N = 69$ ) in our lab (Murphy & Andrews, in preparation) indicates that this version of the Reading Span test has a test-retest reliability of .70 over a 2-week interval. This value is comparable to those reported by Waters and Caplan (2003) for their simple (.73) and complex (.76) Sentence Span tasks and it meets Nunnally's (1978) criterion for minimum reliability adequacy.

### *Results and Discussion*

Table 4 shows the descriptive statistics and correlations among  $n$ -term, Reading Span, and comprehension accuracy for object-relatives/clefts, and subject-relatives/clefts (%'s correct averaged across roles). Consistent with the relational processing and WM hypotheses,  $n$ -term and Reading Span were each significantly correlated with object-relatives/clefts, and with subject-relatives/clefts. In Experiment 1, the latter associations were marginally significant.  $N$ -term and Reading Span tasks were not significantly correlated, despite the fact that both tasks were computer-administered to individual participants in Experiment 2.

A multiple regression analysis showed that Reading Span and  $n$ -term accounted for 28.4% of variance in comprehension of object-relatives/clefts, Multiple  $R = .53$ ,  $F(2, 65) = 12.89$ ,  $p < .001$ .  $N$ -term accounted for 20.6% of variance independently of Reading Span ( $p < .001$ ). Reading Span accounted for 5.7% of variance independent of  $n$ -term ( $p < .05$ ). In a

second analysis, Reading Span, *n*-term, and subject-relatives/clefts accounted for 42.9% of variance in object-relatives/clefts, Multiple  $R = .66$ ,  $F(3, 64) = 16.02$ ,  $p < .001$ . *N*-term accounted for 7.56% of variance independently ( $p < .01$ ). Subject-relatives/clefts accounted for 14.52% of variance independently ( $p < .001$ ). As in Experiment 1, Reading Span was no longer a significant unique predictor, but might have contributed to shared variance (20%).

### Experiment 3

The associations between comprehension of object-relatives/clefts and the *n*-term task observed in Experiments 1 and 2 are consistent with the interpretation that each task involves processing of complex relational information. A related explanation is that *n*-term predicts comprehension of object-relatives/clefts because the tasks involve mental re-ordering of task elements. Comprehension of object-relatives/clefts might entail re-ordering of the nouns from non-canonical to canonical order in the course of thematic role assignment. In the *n*-term task, re-ordering of the premises and the letters within the premises is required because these are presented in random order and a combination of > and < relations are used. It is important to note that re-ordering and relational complexity explanations are not independent. Non-canonical order contributes to the complexity of object-relatives and object-clefts by making segmentation more difficult and concentrating processing of the noun-verb relations. Similarly in the *n*-term task, premise relations are presented in non-canonical order and a mixture of > and < signs is used to ensure that complexity is not reduced through segmentation. In both cases, removing the reordering component from the tasks would drastically reduce their complexity. In Experiment 3, we attempted to distinguish between relational complexity and mental re-ordering explanations by using a different relational task, which does not involve re-ordering, as a predictor of comprehension.

The Latin square task was developed to assess the impact of relational complexity on

adult cognition (Birney, 2002; Birney, Halford, & Andrews, under review; Birney & Halford, 2000; 2001). In a typical problem, an incomplete  $4 \times 4$  matrix (Figure 2) is presented. The participants' task is to determine which of four elements should fill a target cell so that the conditions of the Latin square are satisfied, namely that only one of the four possible elements occurs in each row and column of the matrix. The relational complexity manipulation is based on an increasingly complex instantiation of this rule (Birney, 2002). The simplest problems require application of this rule in a single row or column. Figure 2A shows a binary-relational problem that can be solved by comparing the three elements already in column 3 with the specified set of four elements to determine the missing element. Using Birney and Halford's (2002) notation this can be represented as:

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

where R and C stand for row and column respectively and the superscripts are row and column numbers. The symbol “ $\rightarrow$ ” represents the higher order relation “IMPLIES”. The continuous underlining indicates those arguments that can be chunked without loss of information necessary to make the current decision. This is in accordance with the principle that: Where A is compared with B and C (e.g., red is different from blue and green), B and C can be chunked because the relation between them need not be processed (Chalmers & Halford, 2003). The relations among the known elements in column 3 do not need to be processed and therefore need not be represented separately. Furthermore, the elements that constrain the target cell are determined by being in the same column as the target, and this does not require further processing.

Figure 2B shows a ternary-relational problem. The value of the target cell is resolved by considering elements in the row and column that intersect the target cell. The problem would be represented as:

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

The two elements in column 2 can be chunked by the principle above, because relations between them do not need to be processed, and the constraint they exercise on the target cell is easily recognised by the fact that they are in the same column. However the square in  $R^2C^4$  cannot be chunked with the other terms because (by the Latin square defining principle) elements in row 2 are not independent of the elements in the intersecting columns. This means that the constraint exercised by the element in  $R^2C^4$  is not pre-identified by being in the same row or column, and requires additional processing. The cell that intersects with column 2 is the target cell, so elements in row 2 need to be considered to make the current decision. The relations among the existing elements in column 2 need not be considered *per se* and therefore can be chunked.

In the quaternary-relational problem in Figure 2C, the target cell cannot be determined by the binary and ternary strategies just described. These strategies result in knowing only that the target cell is not a cross. Solution 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. The problem can be represented as:

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

Thus applications of the defining principle require different levels of relational integration. This was the basis for the complexity manipulation. The relational processing hypothesis was that Latin Square scores would predict comprehension of object-relatives, before and after controlling for subject-relatives.

It is generally accepted that WM is involved in comprehension of complex sentences (Carpenter et al., 1994; Gibson, 1998). Reading Span is a widely used measure of WM capacity (Daneman & Merikle, 1996). Our version of Reading Span was based on Turner and Engle (1989) and seems to be consistent with the theoretical assumptions of the WM

approach. It is somewhat surprising then that stronger associations with comprehension of object-relatives/clefts were not observed. In Experiments 1 and 2, *n*-term, which does not involve sentence processing, was a stronger predictor of comprehension than Reading Span, which does.

In Experiment 3, we assessed WM using digit span forward (FDS) and digit span backward (BDS). FDS differs from Reading Span in being a simple rather than complex span test. Simple span tests do not require simultaneous storage and processing to the same extent that complex span tests do. In BDS, digits are presented in a particular order, but must be recalled in the reverse order. It could be argued that BDS requires simultaneous storage and processing of information (re-ordering of digits) and as such it constitutes a complex span task. Inclusion of BDS will also allow further exploration of the mental re-ordering hypothesis. If the re-ordering interpretation of previous findings is correct, then BDS should account for the greater difficulty of object- as compared to subject-relatives. The WM hypothesis was tested using FDS and BDS.

### *Method*

#### *Participants*

The participants were first-year psychology students who participated in return for course credit. A total of 167 participants (114 females, 53 males) completed sentence comprehension. For 153 of these participants, data for LST, FDS and BDS, were also available.

#### *Apparatus and Procedure*

IBM 486 computers with 14" SVGA color monitors were used to administer sentence comprehension, Latin square, FDS and BDS tasks, which were presented as part of a larger study (Birney, 2002).

*Sentence comprehension.* Each participant received 48 sentences, six instances each

of object- and subject-relatives with two, three, four and five role assignments, as shown in Table 1. Comprehension questions were identical to those used in Experiments 1 and 2. A self-paced procedure was used. Sentences appeared one at a time in the upper half of the screen in 15 to 20 mm yellow lettering on a grey background. Participants pressed the spacebar when they thought they understood the sentence. A question replaced the sentence in the upper half of the screen and response options were displayed in the lower half in 15 to 20mm red lettering. Participants responded by clicking on a response option with the left mouse button.

*Latin Square Task.* Participants were instructed to work through the problems as quickly and as accurately as possible and to do all working in their heads. Four practice trials of increasing complexity were presented. The first was a trivial example of a single row of three cells of which two were filled. The second was an incomplete  $3 \times 3$  Latin Square. The third and fourth were ternary and quaternary problems, respectively. Detailed feedback using row and column labels was provided for the practice problems.

The test phase consisted of 36 items (12 items at each complexity level) presented in a different random order to each participant. The incomplete Latin Square and the response options were displayed on the left and right sides of the screen, respectively as in Figure 2, except that the completed square was not provided. Participants indicated which element should fill the marked cell by clicking on a response option. No feedback was provided for test items. The number of correct responses at each complexity level was converted to a proportion and these were summed to yield a score out of 3.

*Forward Digit Span (FDS) and Backward Digit Span (BDS).* The FDS task took the traditional format but was presented on the computer. Digits were presented one at a time on the screen at 1000 ms intervals. The word “Go” was displayed after the final digit indicating that participants should enter the string of digits in the same order as they were presented

using the numeric keypad or the keys at the top of the keyboard. They were not permitted to change a digit once it had been entered. Two items were presented at each list length (2 to 9) starting at list length 2 and increasing systematically thereafter. The FDS score was the number of items (out of 16) recalled correctly. Due to a programming error FDS was not administered to 30 participants. These missing values were estimated from BDS scores using the regression approach. The BDS task took the same format as the FDS task except participants entered the digits in the reverse order to that in which they were presented.

### *Results and Discussion*

Table 5 shows the descriptive statistics and correlations among Latin square, FDS, BDS and comprehension accuracy for object- and subject-relatives (%'s correct averaged across roles) after exclusion of six participants who had large standardized residuals in a preliminary analysis. Latin Square, FDS, and BDS were significantly correlated with sentence comprehension. Unlike the previous experiments, the associations between working memory and RC measures were also significant.

FDS, BDS and Latin Square accounted for 22.8% of variance in comprehension of object-relatives, Multiple  $R = .48$ ,  $F(3, 143) = 14.11$ ,  $p < .001$ . Latin square accounted for 15.52% variance independently ( $p < .001$ ) of FDS and BDS. Neither FDS nor BDS accounted for unique variance. The remaining 5.68% variance was shared. In a second analysis, comprehension of subject-relatives was included as a predictor along with Latin Square, FDS, and BDS. A total of 37.6% of variance in object-relatives, Multiple  $R = .61$ ,  $F(4, 142) = 21.37$ ,  $p < .001$  was accounted for. Latin Square (4.41%,  $p < .01$ ) and subject-relatives (14.75%,  $p < .001$ ) each accounted for independent variance, but FDS and BDS did not. The remaining 16.64% variance was shared.

The patterns of unique and shared variance involving Latin Square in Experiment 3 parallel those observed for  $n$ -term in Experiments 1 and 2. Latin Square made a unique

contribution before and after comprehension of subject-relatives was included as a predictor of object-relatives. The smaller unique contribution of Latin Square as compared to *n*-term in the regressions might reflect the different ranges of item complexity. Item complexity ranged from binary to quaternary for Latin Square and from ternary to quinary for *n*-term. These results support the relational processing hypothesis and suggest that the difficulty of object-relatives stems (in part) from the complexity of the relations they entail.

There was less support for the WM hypothesis. Although the zero-order correlations suggested that STM and /or WM are involved in comprehension of object-relatives, the regression analyses showed that these processes are involved to a similar extent in comprehension of the subject-relatives. That is, the involvement of WM (as indexed by FDS and BDS) was not unique to object-relatives. These findings are similar to those involving Reading Span in Experiments 1 and 2. Entering FDS without BDS or BDS without FDS did not change the pattern of significance in any of the regression analyses. That is, the significant association between FDS and BDS did not mask the unique contribution of working memory.

Two aspects of the results argue against the re-ordering explanation outlined above. First, the correlation between Latin Square and object-relatives is similar in magnitude to the correlation of *n*-term with object-relatives in Experiments 1 and 2. If the re-ordering explanation were correct, a weaker association might have been expected in Experiment 3 because the Latin Square does not entail re-ordering. Second, the BDS was the only predictor task in Experiment 3 that required re-ordering of elements. If the re-ordering hypothesis were correct, BDS should contribute unique variance in the multiple regression analyses, but this did not occur.

### General Discussion

The research hypotheses were based on RC and WM explanations of individual

differences in comprehension of object-relative and object-cleft sentences. The hypothesis that sentence comprehension entails processing of complex relations received support from the regression analyses involving the relational processing tasks. The *n*-term task required integration of varying numbers of ordinal relations, and the procedure was intended to constrain participants toward integrating the relations in the same decision. Similarly the Latin Square task involved integration of relations of varying complexity. The contributions of *n*-term and Latin Square remained significant after controlling for comprehension of subject-relatives and subject-clefts. This suggests that the greater difficulty of the object-extracted sentences is due to their greater complexity and that tasks that involve processing of complex relations capture this additional complexity. That both *n*-term and Latin Square tasks predicted comprehension of object-relatives and object-clefts argues against an alternative re-ordering explanation because *n*-term involves re-ordering, whereas Latin Square does not.

It might be claimed that comprehension of the 5-role object-relative sentences involves problem solving rather than normal sentence comprehension processes, and that the correlations between sentence comprehension and our relational tasks might merely reflect a willingness or ability to develop strategies and to engage in difficult tasks (as indexed, for instance, in measures of fluid intelligence). If so, the correlations should be drastically reduced if the 5-role sentences (which were very difficult) were excluded. The multiple regression analyses in the three studies were repeated without the 5-role sentences. The variance accounted for by the predictors decreased slightly in Experiments 1 and 3, and increased slightly in Experiment 2. In all cases the pattern of significance was unchanged. The predictors that were significant (nonsignificant) when the 5-role sentences were included remained significant (nonsignificant) when they were excluded. Thus the correlations do not depend on inclusion of the 5-role sentences. It might be argued further

that the 4-role sentences are also quite difficult and that a different pattern of results might emerge if these too were excluded. The analyses were repeated without the 4-role and 5-role sentences. The total variance accounted for was reduced in all three experiments, but the patterns of significance were largely unchanged. The relational processing tasks accounted for significant unique variance in comprehension of the 3-role object-relatives and/or clefts (Experiments 1, 2) and the 2- and 3-role object-relatives (Exp. 3). Comprehension of subject relatives/clefts contributed significant unique variance in Experiments 2 and 3, but not in Experiment 1. The unique contribution of WM was not significant in any data set. That similar patterns of significance were observed when the analyses were restricted to the 2- and/or 3-role sentences suggests that our relational processing tasks are tapping into normal sentence comprehension processes rather than processes that are specific to the 4- and/or 5-role sentences, which are very complex.

There is a further reason to suspect that the cross task correlations do not simply reflect a general willingness to develop strategies to deal with difficult tasks. If this were the case, we might have expected the correlations between comprehension of complex sentences and the WM tasks to be as strong as those between comprehension and the relational tasks because the WM tasks were also quite difficult and the tasks seem amenable to different strategies. In fact, however the relational processing tasks were consistently stronger predictors of comprehension.

Our findings are consistent with the hypothesis that comprehension of object-relatives and object-clefts, the *n*-term task, and the Latin Square task all require processing of complex relations that are extremely difficult to decompose into less complex components. The *n*-term and Latin Square tasks were designed specifically to meet these criteria. We argue that this is also the case for object-relative sentences. As noted previously, understanding the thematic relations between nouns and verbs is critical to sentence

comprehension, and this was the focus of the comprehension questions.

It seems likely that our off-line procedure would allow participants to reanalyse the sentences and recompute the thematic roles. It is also likely that participants attempted to comprehend the entire sentence (rather than a part thereof) because they did not know which question would be presented until after the sentence had left the screen. Recomputation of the thematic roles is known to be more difficult for object- than for subject-relatives (Waters & Caplan, 2001), arguably because noun-verb relations in object-relatives are very difficult to segment. This imposes a constraint to represent the entire set of noun-verb relations in the same decision. For the 3-role object-relatives (see Table 1), this would mean assigning nouns to three roles of the verbs (two agent roles, one patient role), which entails a ternary relation. The 4-role and 5-role object-relatives involve additional thematic roles, and would entail quaternary and quinary relations, respectively. According to RC theory, adult humans can process up to four variables in a single decision, which implies a quaternary-relational limit. This is consistent with the extreme difficulty of the 5-role object-relatives (Table 1), and also with findings from non-linguistic domains (Halford et al., 2005)

The notion that reasoning and language processing are closely linked is consistent the work of other researchers. Polk and Newell (1995) proposed a deductive reasoning model based on linguistic mechanisms, and prefrontal regions, including the DLPFC, have been shown to be involved in relational processing (Waltz et al., 1999) as well as in comprehension of complex sentences (Just et al., 1996). Thus our cross-task results might reflect individual differences in the integrity of a common brain region (e.g., the DLPFC) that is recruited by tasks from multiple domains that involve complex relations.

An interesting question is whether the ability to process complex relations comes into play during sentence comprehension or afterwards. It seems clear that our self-paced procedure with end-of-sentence comprehension test would allow readers to reanalyse the

sentence prior to responding. This might mean that post-interpretive processing is involved, to use Caplan and Waters' (1999) terminology. However, it does not necessarily mean that comprehension performance would be insensitive to on-line processing. The RC approach takes account of fluctuations in processing demand, in that estimates of complexity and processing load reflect the peak demand imposed during a task (Halford, et. al, 1998, Section 2.1). Estimates of peak load based on the RC approach are highly correlated with maximal integration cost (Gibson, 2000) imposed during sentences. Thus, participants who can process complex relations will be better able to cope with the peak load imposed by thematic role assignments during the sentence. Having successfully assigned the nouns to their roles, they would be well placed to respond correctly to the end-of-sentence question. The use of on-line techniques in future work would allow a more localised examination of the associations with relational complexity and WM measures. For example, the associations between the relational processing measures and phrase by phrase reading times or eye movement variables during reading object- and subject-relatives could be examined and contrasted with the corresponding associations with WM measures.

Support for the WM hypothesis was more equivocal. The significant correlations between WM measures (Reading Span, FDS, BDS) and comprehension of object-relatives and object-clefts are consistent with the involvement of WM in sentence comprehension. However the claim that object-relatives and object-clefts impose higher WM demands than the corresponding subject-relatives and subject-clefts was not supported by the regression analyses. When comprehension of subject-relatives and subject-clefts was included as a predictor, the contributions of Reading Span (Experiments 1, 2) and FDS and BDS (Experiment 3) were no longer significant.

It remains possible that our WM measures were in some way inadequate. In recognition of this possibility, we switched from group administration of a manual version of

Reading Span to individual testing using a computer administered version (Experiments 1 to 2) and from Reading Span to FDS and BDS (Experiments 1 and 2 to 3). Despite these changes, the outcome was similar. The reliability of the WM measures might have affected the outcome of the regression analyses. While the test-retest reliability of the Reading Span test we used in Experiment 2 satisfies Nunnally's (1978) criterion for minimum reliability adequacy, it is not impressive. Thus the possibility remains that a different pattern of results might have been obtained if WM measures with higher reliability or a number of different WM measures had been used or if we had adopted a more on-line methodology to examine sentence processing.

Our comprehension results seem inconsistent with those of King and Just (1991) who reported a significant interaction of WM capacity with sentence complexity. The difference between high and low span participants was greater for object-relatives than for subject-relatives. Accordingly, we expected reading span to account for variance in object-relatives after controlling of subject-relatives. One possible explanation is that Reading Span is sensitive to the demands of sentences within a narrower range of complexity than was used in our experiments. However, this was not supported by the additional multiple regression analyses which excluded the 5-role sentences or the 4- and 5-role sentences. A more likely explanation of the discrepancy implicates the different procedures used to assess comprehension. Our procedures minimized storage demands by having entire sentences visible on the screen, with unlimited time for decoding. This ensured that the demand imposed by comprehension was primarily for computational capacity. This was also the case for the *n*-term and Latin Square tasks. King and Just (1991) presented sentences visually one word at a time. Sentences were never seen in their entirety so there would have been no opportunity for the regressive eye movements so frequent in normal reading (Martin & Romani, 1994; Ni & Shankweiler, 1995), and consequently, a greater reliance on initial

encoding processes. Comprehension failures in King and Just's procedure could be due to inadequate initial encoding, failure to maintain the sentence in memory, inadequate integration of discourse elements, or processing of the noun-verb relations. Our procedure would have minimized the first and second of these potential sources of failure. The lower than expected associations between Reading Span and comprehension of object-relatives and object-clefts might be due to the lower storage demands of our more ecologically valid procedure. If Reading Span is primarily a measure of storage capacity and our procedures imposed low storage demands, then the observed correlations are unsurprising.

Our findings have implications for Just and Carpenter's (1992) working memory model, which assumes that WM resources can be flexibly allocated to computational or storage demands or to some combination of the two. If so, then Reading Span should predict comprehension irrespective of whether it assesses mainly storage or mainly computational capacity, and irrespective of the mix of storage and computational demands imposed by the comprehension test. The current findings cast doubt on that assumption, and suggest that there is more independence between storage and computational resources (Halford, Maybery, O'Hare, & Grant, 1994; Halford, Phillips, & Wilson, 2001; Klapp, Marshburn, & Lester, 1983) than is assumed in the WM approach.

In conclusion, we suggest that relational complexity which has been found to be applicable to a wide range of other cognitive domains might also be applicable to sentence comprehension. If so, this opens the way to investigate factors that contribute to complexity in both linguistic and nonlinguistic domains. The fact that the relational complexity metric seems to perform similarly to the sophisticated DLT (Gibson, 2000) metric might also open up potential for a general complexity metric, applicable to both linguistic and nonlinguistic domains. Such a metric probably would not encompass every aspect of linguistic comprehension, because there are likely to be modular processes that are independent of

general cognitive complexity. However any steps towards an integrated approach to cognitive complexity are potentially useful.

## References

- Andrews, G., & Halford, G. S. (1998). Children's ability to make transitive inferences: The importance of premise integration and structural complexity. *Cognitive Development, 13*, 479-513.
- Andrews, G., & Halford, G. S. (2002). A cognitive complexity metric applied to cognitive development. *Cognitive Psychology, 45*, 153-219.
- Andrews, G., Halford, G. S., Bunch, K. M., Bowden, D., & Jones, T. J. (2003). Theory of mind and relational complexity. *Child Development, 74*, 1435-1458.
- Andrews, G., Halford, G. S., & Prasad, A. (1998). *Processing Load and Children's Comprehension of Relative Clause Sentences*. (ERIC Document Reproduction Service No. ED 420 091).
- Baddeley, A. D. (1986). *Working memory*. Oxford: Clarendon Press.
- Baddeley, A. (1993). Working memory or working attention? In A. Baddeley & L. Weiskrantz (Eds.) *Attention: Selection, Awareness, and Control. A Tribute to Donald Broadbent*. (pp. 152-170). Oxford: Clarendon Press.
- Birney, D. P. (2002). *The Measurement of Task Complexity and Cognitive Ability: Relational Complexity in Adult Reasoning*. Unpublished PhD Dissertation, University of Queensland, St Lucia, Brisbane.
- Birney, D. P., & Halford, G. S. (2000, August). *Methods for analysing complexity in reasoning tasks: Links to Fluid intelligence*. Paper presented at The Fourth International Conference on Thinking, University of Durham, UK.
- Birney, D. P., & Halford, G. S. (2001). *Understanding cognitive complexity: Evidence from cognitive psychology and individual differences*. Paper presented at the Third International Spearman Seminar, University of Sydney, Australia.
- Birney, D. P., & Halford, G. S. (2002). Cognitive complexity of suppositional reasoning: A

- application of the relational complexity metric to the knight-knave task. *Thinking and Reasoning*, 8, 109-134.
- Birney, D. P., Halford, G. S., & Andrews, G. (under review). *Measuring the Influence of Complexity on Relational Reasoning: The Development of the Latin Square Task*.
- Blaubergs, M. S., & Braine, M. D. S. (1974). Short-term memory limitations on decoding self-embedded sentences. *Journal of Experimental Psychology*, 102, 745 – 748.
- Blumenthal, A. L. (1966). Observations with self-embedded sentences. *Psychonomic Science*, 6 (10), 453-454.
- Boag, C. C., Härtel, C. E. J., & Halford, G. S. (in press). *An integrated model of situation awareness and decision making in air traffic control to explain performance errors*. Proceedings of the European Association for Aviation Psychology Conference, Crieff, Scotland.
- Caplan, D., Alpert, N., & Waters, G. S. (1999). PET studies of syntactic processing with auditory sentence presentation. *NeuroImage*, 9, 343-351.
- Caplan, D., Alpert, N., & Waters, G.S. (1998). Effects of syntactic structure and prepositional number on patterns of regional cerebral blood flow. *Journal of Cognitive Neuroscience*, 10(4), 541-552.
- Caplan, D., & Waters, G. S. (1999). Verbal working memory and sentence comprehension. *Behavioral and Brain Sciences*, 22, 77-94
- Carpenter, P. A., Miyake, A., & Just, M. A. (1994). Working memory constraints in comprehension: Evidence from individual differences, aphasia, and aging. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics*. (pp.1075-1122). San Diego, CA: Academic Press.
- Chalmers, K. A., & Halford, G. S. (2003). Young children's understanding of oddity: reducing complexity by simple oddity and "most different" strategies. *Cognitive*

*Development*, 18, 1-23.

Christoff, K., Prabhakaran, V., Dorfman, J., Zhao, Z., Kroger, J., Holyoak, K. J., & Gabrieli,

J. D. E. (2001). Rostrolateral prefrontal cortex involvement in relational integration during reasoning. *NeuroImage*, 14, 1136-1149.

Constable, R. T., Pugh, K. R., Berroya, E., Mencl, W. E., Westerveld, M., Ni, W., &

Shankweiler, D. (2004). Sentence complexity and input modality effects in sentence comprehension: an fMRI study. *NeuroImage*, 22, 11-21.

Daneman, M., & Carpenter, P. (1980). Individual differences in working memory and

reading. *Journal of Verbal Learning and Verbal Behavior*, 19, 405-438.

Daneman, M., & Merikle, P. M. (1996). Working memory and language comprehension: A

meta-analysis. *Psychonomic Bulletin and Review*, 3, 422-433.

English, L. D., & Halford, G. S. (1995). *Mathematics education: Models and processes*.

Hillsdale, NJ: Erlbaum.

Fodor, J. A., & Garrett, M. (1967). Some syntactic determinants of sentential complexity.

*Perception and Psychophysics*, 2, 289-296.

Ford, M. (1983). A method for obtaining measures of local parsing complexity throughout

sentences. *Journal of Verbal Learning and Verbal Behavior*, 22, 203–218.

Forster, K. I., & Forster, J. C. (1999). *DMDX Display System: Laboratory software for*

*mental chronometry*. University of Arizona, Arizona.

<http://www.u.arizona.edu/~kforster/dmastr/dmastr.htm>

Frazier, L. (1985). Syntactic complexity. In D. Dowty & L. Karttunen & A. Zwicky (Eds.),

*Natural language parsing*. Cambridge: Cambridge University Press

Gibson, E. (2000). The dependency locality theory: A distance-based theory of linguistic

complexity. In Y. Miyashita, A. Marantz, & W. O'Neil, (Eds.), *Image, language,*

*brain* (pp. 95-126), Cambridge, MA: MIT Press.

- Gibson, E. (1998). Linguistic complexity: locality of syntactic dependencies. *Cognition*, 68, 1-76.
- Gordon, P. C., Hendrick, R., & Levine, W. H. (2002). Memory-load interference in syntactic processing. *Psychological Science*, 13, 425-430.
- Haarmann, H. J., Just, M. A., & Carpenter, P. A. (1997). Aphasic sentence comprehension as a resource deficit: A computational approach. *Brain and Language*, 59, 76-120.
- Halford, G. S. (1993). *Children's, understanding: The development of mental models*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Halford, G. S., Andrews, G., Dalton, C., Boag, C., & Zielinski, T. (2002a). Young children's performance on the balance scale: The influence of relational complexity. *Journal of Experimental Child Psychology*, 81, 417-445.
- Halford, G. S., Andrews, G., & Jensen, I. J. (2002b). Integration of Category Induction and Hierarchical Classification: One Paradigm at Two Levels of Complexity. *Journal of Cognition and Development*, 3, 143 - 177.
- Halford, G. S., Baker, R., McCredden, J. E., & Bain, J. D. (2005). How many variables can humans process? *Psychological Science*, 16, 70-76.
- Halford, G. S., Maybery, M. T., & Bain, J. D. (1986). Capacity limitations in children's reasoning: A dual task approach. *Child Development*, 57, 616-627.
- Halford, G.S., Maybery M.T., O'Hare, A.W., & Grant, P. (1994). The development of memory and processing capacity. *Child Development*, 65, 1338-1356.
- Halford, G. S., Phillips, S., & Wilson, W. H. (2001). Processing capacity limits are not explained by storage limits. *Behavioral and Brain Sciences*, 24, 123-124.
- Halford, G. S., Wilson, W. H., & Phillips, S. (1998). Processing capacity defined by relational complexity: Implications for comparative, developmental, and cognitive psychology. *Behavioral and Brain Sciences*, 21, 803-831

- Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, *99*, 122-149.
- Just, M. A., Carpenter, P. A., Keller, T. A., Eddy, W. F., & Thulborn, K. R. (1996) Brain activation modulated by sentence comprehension. *Science*, *274*, 114-116.
- King, J., & Just, M. A., (1991). Individual differences in syntactic processing: The role of working memory. *Journal of Memory and Language*, *30*, 580-602.
- Klapp, S. T., Marshburn, E. A., & Lester, P. T. (1983). Short-term memory does not involve the "working memory" of information processing: The demise of a common assumption. *Journal of Experimental Psychology: General*, *112*, 240-264.
- Kroger, J., Sabb, F. W., Fales, C., Bookheimer, S. Y., Cohen, M. S., & Holyoak, K. (2002). Recruitment of anterior dorsolateral prefrontal cortex in human reasoning: A parametric study of relational complexity. *Cerebral Cortex*, *12*, 477-485.
- Larkin, W., & Burns, D. (1977). Sentence comprehension and memory for embedded structure. *Memory and Cognition*, *5*, 17-22.
- Lewis, R. L. (1996). A theory of grammatical but unacceptable embeddings. *Journal of Psycholinguistic Research*, *25*, 93-116.
- Linebarger, M. C., Schwartz, M. F., & Saffran, E. M. (1983). Sensitivity to grammatical structure in so-called agrammatic aphasics. *Cognition*, *13*, 361-379.
- Marks, L. E. (1968). Scaling of grammaticalness of self-embedded English sentences. *Journal of Verbal Learning and Verbal Behavior*, *7*, 965-967.
- Martin, R., & Romani, C. (1994). Verbal working memory and sentence comprehension: A multiple-components view. *Neuropsychology*, *8*, 506-523.
- Maybery, M. T., Bain, J. D., & Halford, G. S. (1986). Information processing demands of transitive inference. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *12*, 600-613.

- Miller, G. A., & Isard, S. (1964). Free recall of self-embedded English sentences. *Information and Control*, 7, 292-303.
- Murphy, K. M., & Andrews, G. (in preparation). *Test-retest reliability of short-term memory and working memory measures*.
- Ni, W., & Shankweiler, D. (1995, January). *Difficulties in on-line processing of relative clause sentences*. Paper presented at the LSA Conference, New Orleans.
- Nunnally, J. (1978). *Psychometric Theory*. New York: McGraw-Hill.
- Phillips, S., & Nikki, K. (2003). [Increased bilateral occipitoparietal activity during retention of binary versus unary indexed lists in pair recognition](#). *Neuroimage* 20, 1226-1235
- Polk, T. A., & Newell, A. (1995). Deduction as verbal reasoning. *Psychological Review*, 102, 533-566.
- Rochon, E., Waters, G. S., & Caplan, D. (2000). The relationship between measures of working memory and sentence comprehension in patients with Alzheimer's Disease. *Journal of Speech, Language, and Hearing Research*, 43, 395-413.
- Sheldon, A. (1974). The role of parallel function in the acquisition of relative clauses in English. *Journal of Verbal Learning and Verbal Behavior*, 13, 272 -281.
- Stromswold, K., Caplan, D., Alpert, N., & Rauch, S. (1996). Localization of syntactic comprehension by positron emission tomography. *Brain and Language*, 52, 452-473.
- Stuss, D. T., & Levine, B. (2002). Adult clinical neuropsychology: Lessons from the study of the frontal lobes. *Annual Review of Psychology*, 53, 401-433.
- Trabasso, T. (1975). *Representation, memory, and reasoning: How do we make transitive inferences?* In A. D. Pick (Ed.). *Minnesota symposia on child psychology* (Vol. 9, pp.135-172). Minneapolis: University of Minnesota Press.
- Traxler, M. J., Morris, R. K., & Seely, R. E. (2002). Processing subject and object relative clauses: Evidence from eye movements. *Journal of Memory and Language*, 47, 69-

90.

- Traxler, M. J., Williams, R. S., Blozis, S. A., & Morris, R. K. (2005). Working memory, animacy, and verb class in the processing of relative clauses. *Journal of Memory and Language, 53*, 204-224.
- Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language, 28*, 127-154.
- Waltz, J. A., Knowlton, B. J., Holyoak, K. J., Boone, K. B., Mishkin, F. S., de Menezes Santos, M., Thomas, C. R., & Miller, B. L. (1999). A system for relational reasoning in human prefrontal cortex. *Psychological Science, 10*(2), 119-125.
- Waters, G. S., & Caplan, D. (2001). Age, working memory, and on-line syntactic processing in sentence comprehension. *Psychology and Aging, 16*(1), 128-144.
- Waters, G., & Caplan, D. (2002). Working memory and online syntactic processing in Alzheimer's Disease: Studies with auditory moving window presentation. *Journal of Gerontology, 57B*(4), 298-311.
- Waters, G. S., & Caplan, D. (2003). The reliability and stability of verbal working memory measures. *Behavior Research Methods, Instruments, & Computers, 35*, 550-564.
- Waters, G. S., & Caplan, D. (2004). Verbal working memory and on-line syntactic processing: Evidence from self-paced listening. *Quarterly Journal of Experimental Psychology, 57A*, 129-163.
- Waters, G. S., Caplan, D., Alpert, N., & Stanczak, L. (2003). Individual differences in rCBF correlates of syntactic processing in sentence comprehension: effects of working memory and speed of processing. *NeuroImage, 19*, 101-112.
- Zielinski, T. A., Goodwin, G., & Halford, G. S. (under review). *Relational complexity and logic: Categorical syllogisms revisited.*

Author Notes

We thank Campbell Dickson and Jason Andalis for computer programming; Tina Dalton, Katie Bunch and Joanne Todd for research assistance; Lee Osterhout, Matt Traxler, Ted Gibson, and an anonymous reviewer for their insightful comments on previous drafts.

Table 1

*Examples of 2-, 3-, 4-, and 5-role Object-relative and Subject-relative Sentences used in Experiments 1, 2, and 3*

Form	Roles	Example sentences
Object	2	Sally saw the horse that the cow followed.
Subject	2	Sally saw the cow that followed the horse
Object	3	The duck that the monkey touched walked.
Subject	3	The monkey touched the duck that walked.
Object	4	The artist that the waiter warned the chef about talked.
Subject	4	The waiter warned the chef about the artist that talked.
Object	5	The clown that the teacher that the actor liked watched laughed.
Subject	5	The actor liked the teacher that watched the clown that laughed.

Table 2

*Examples of 2-, 3-, 4-, and 5-Role Object- and Subject-cleft Sentences in Experiment 2*

Form	Roles	Example sentences
Object	3	It was the cook that the king sent the man to
Subject	3	It was the king that sent the man to the cook
Object	4	It was the doctor that the farmer that the politician helped liked.
Subject	4	It was the farmer that liked the doctor and it was the politician that helped the farmer
Object	5	It was the bear that the ox pushed the horse that the pig bit onto
Subject	5	It was the ox that pushed the horse onto the bear and it was the pig that bit the horse

Table 3

*Simple Correlations and Descriptive Statistics for Comprehension of Object-relatives and Subject-relatives, Reading Span (RS), and N-term Tasks in Experiment 1*

	1	2	3	4
1. Object-relatives (%)	1.00			
2. Subject-relatives (%)	.43**	1.00		
3. Reading Span	.29*	.26*	1.00	
4. N-term	.65***	.23*	.16	1.00
<i>M</i>	75.61	91.12	32.13	23.71
<i>SD</i>	12.93	6.70	6.42	6.97
<i>N</i>	61	61	61	61

\*  $p < .05$ ; \*\*  $p < .01$ ; \*\*\*  $p < .001$

Table 4

*Simple Correlations and Descriptive Statistics for Comprehension of Object-relatives/clefts, Subject-relatives/clefts, Reading Span, and N-term Tasks in Experiment 2*

	1	2	3	4
1. Object-relatives/clefts (%)	1.00			
2. Subject-relatives/clefts (%)	.59***	1.00		
3. Reading Span	.28*	.34**	1.00	
4. N-term	.48***	.39**	.09	1.00
<i>M</i>	69.14	86.57	27.26	23.09
<i>SD</i>	14.15	7.65	7.44	6.68
<i>N</i>	68	68	68	68

\*  $p < .05$ ; \*\*  $p < .01$ ; \*\*\*  $p < .001$  (2-tailed)

Table 5

*Simple Correlations and Descriptive Statistics for Comprehension of Object-relatives, Subject-relatives, Forward Digit Span, Backward Digit Span, and Latin Square Tasks in Experiment 3*

	1	2	3	4	5
1. Object-relatives	1.00				
2. Subject-relatives	.57***	1.00			
3. Forward digit span	.24**	.26**	1.00		
4. Backward digit span	.25**	.37***	.61***	1.00	
5. Latin square	.46***	.48***	.25**	.39***	1.00
<i>M</i>	75.48	88.78	10.23	8.43	2.26
<i>SD</i>	14.17	9.91	2.03	2.62	0.44
<i>N</i>	147	147	147	147	147

\*  $p < .05$ ; \*\*  $p < .01$ ; \*\*\*  $p < .001$  (2-tailed)

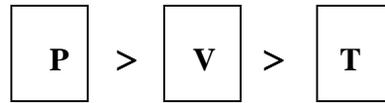
## Figure Captions

*Figure 1.* Example items at three levels of complexity in  $n$ -term task. Premise information is shown on the left and completed sequences on the right.

*Figure 2.* Example items at three levels of complexity in the Latin Square Task with problem squares on the left, response options in the middle, and completed square on the right. Integration in a single dimension (A), integration in two intersecting dimensions (B), and integration in multiple dimensions (C). Completed squares were not presented to participants. Participants chose one option to fill the target cell “?”

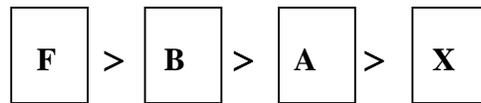
*Premises*

**T < P**  
**P > V**  
**T < V**

*Correct sequences***Ternary**

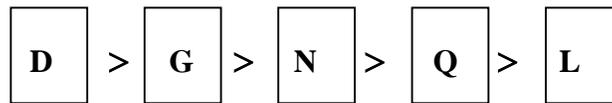
---

**B > A**  
**A < F**  
**F > B**  
**X < A**

**Quaternary**

---

**N < G**  
**D > G**  
**N > Q**  
**L < Q**  
**G > L**

**Quinary**

**A Problem Square**

		●	
●		?	■
▲		■	
		+	

**Option**

▲
■
+
●

**Completed Square**

+	■	●	▲
●	+	▲	■
▲	●	■	+
■	▲	+	●

**B Problem Square**

	▲		●
	?		■
			+
	●	+	

**Option**

▲
■
+
●

**Completed Square**

+	▲	■	●
●	+	▲	■
▲	■	●	+
■	●	+	▲

**C Problem Square**

▲			
	+		
		▲	
	?		+

**Option**

▲
■
+
●

**Completed Square**

▲	■	+	●
■	+	●	▲
+	●	▲	■
●	▲	■	+