



**HAL**  
open science

## The development of relational integration during childhood.

Bruno Dauvier, Christine Bailleux, Patrick Perret

► **To cite this version:**

Bruno Dauvier, Christine Bailleux, Patrick Perret. The development of relational integration during childhood.. *Developmental Psychology*, 2014, 50 (6), pp.1687 - 1697. 10.1037/a0036655. hal-01772111

**HAL Id: hal-01772111**

**<https://amu.hal.science/hal-01772111v1>**

Submitted on 26 Apr 2018

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

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

**Submitted to Developmental Psychology - May 31, 2013**

**Title**

The Development of Relational Integration During Childhood

**Running Title**

The development of relational integration

Dauvier, Bruno

Bailleux, Christine

Perret, Patrick

Aix-Marseille Université

## **Abstract**

*Relational integration* refers to the process whereby several variables are integrated within a single cognitive representation. In cognitive and developmental science, it is regarded as a central function of working memory that may contribute to the development of higher-order processes such as reasoning and fluid intelligence. In the present experiment, we administered the Relational Integration Level Assessment Task (RILAT), a novel experimental task designed to assess children's level of relational integration, to participants aged 5-12 years, along with a measure of fluid intelligence. Results yielded age norms for the development of successive levels of relational complexity, and indicated a smooth progression rather than abrupt, stage-like transitions. Relational integration was found to be closely associated with fluid intelligence, as measured by Raven's Progressive Matrices. Analyses of age-related changes in this association revealed that the relationship was strongest during the stabilization periods for each level of relational complexity.

*Keywords:* relational complexity theory, fluid intelligence, cognitive development, Relational Integration Level Assessment Task, varying-coefficient model

## **The Development of Relational Integration During Childhood**

### **Relational integration in cognitive and developmental science**

The idea that human intelligence is grounded in the ability to discover, represent and process relations has its roots in Spearman's seminal conceptualization of the general factor of intelligence. Spearman (1927) tied this *g* factor to a central cognitive mechanism that allows the relationship between two items to be abstracted and applied to another context. Modern-day theories echo this emphasis on the role of relational processing, with contemporary advocates arguing that many facets of higher cognition, such as reasoning, language, categorization, and planning, are relational in nature, insofar as they require the dynamic binding of elements into structured schemas, in addition to the elementary processing of isolated features or associations (Halford, Wilson, & Phillips, 2010). Analogical reasoning, for example, involves identifying the structural similarity between two situations, on the basis of relations between objects rather than attributes. The creative compositionality of human thought is regarded as a byproduct of our ability to bind "content elements such as words, objects, or events to places in a cognitive coordinate system or to variables in a schema" (Oberauer, 2010, p. 47). In cognitive science, significant progress has been made in designing plausible connectionist architectures that encode relational knowledge (Doumas, Hummel, & Sandhofer, 2008). Advances in neuroscience have also made it possible to identify the neurobiological correlates of relational integration. In their groundbreaking study, Waltz et al. (1999) showed that patients with frontotemporal dementia who had serious frontal lobe damage encountered particular difficulty with highly demanding relational tasks. More recent fMRI studies have confirmed that relational integration reliably recruits the left rostral prefrontal cortex. This has been demonstrated using tasks that involve the

processing of semantic relations (Bunge, Wendelken, Badre, & Wagner, 2005) or require visuospatial relational processing (Bunge, Helskog & Wendelken, 2009).

Contemporary research is also directed toward gaining a better understanding of how relational links are established in working memory (WM), which Halford and colleagues refer to as “the workspace where relational representations are constructed” (2010, p. 499). Oberauer (2009) adopted a functional perspective to this issue, theorizing a model in which WM’s main function—above and beyond actively maintaining information—is to support the dynamic binding of elements into coordinate systems, that is, the formation of relational structures. This perspective is described as *functional* in the sense that it is based on the notion that WM is an adaptive system, which “has certainly not evolved to help us remember telephone numbers” (p. 46). More crucially, WM is the mental workspace where new or complex problems are solved and, as such, is oriented toward representing the variables involved in those problems, as well as their interactions. According to the structural correspondence principle (Halford, Wilson, & Phillips, 1998), the relations in the resulting representation must correspond to relations in the world or in the problem statement. This assumption is in line with Johnson-Laird’s influential proposition that human reasoning, particularly relational reasoning, relies heavily on the formation of mental models (Goodwin & Johnson-Laird, 2005; Johnson-Laird, 1983). Like most researchers since the publication of Miller’s study on information-processing capacity in 1956, Oberauer views WM as having a limited capacity. However, in contrast to many researchers investigating WM, he contends that the most influential limiting factor is not the ability to retain information per se, but the ability to integrate several variables into the representation of a common coordinate system. This is the ability that is referred to as *relational integration* (RI). Empirical evidence in favor of this view comes from studies exploring the links between WM and fluid intelligence. A

well replicated finding in the literature is that individual variations in WM efficiency are closely correlated with measures of a broad range of higher cognitive abilities, such as reasoning or text comprehension (Kane, Hambrick & Conway, 2005). Oberauer, Süb, Wilhelm and Wittmann (2008) showed that (i) tasks involving RI are better predictors of performance on reasoning tasks than more traditional WM span tasks, and (ii) this predictive power prevails even when the RI tasks do not require any form of information storage.

Oberauer's model of WM is not the only one to place considerable importance on RI. In the field of cognitive development, the relational complexity theory originally proposed by Halford back in 1983 and regularly updated since then (Halford, Wilson, & Phillips, 1998, 2010) shares many of its core assumptions about the nature of WM limits and the constraints they place on fluid intelligence. However, Halford goes one theoretical step further than Oberauer, by defining a metric by which to identify a task's precise relational processing requirements. This relational complexity metric applies both to the problem's structural features and to the individual's processing capacity. *Relational complexity* (RC) is defined as the number of variables that must be related within the same coordinate system. Unary relations are based on a single variable, binary relations on two, ternary relations on three, and so on. RC theory posits that quaternary relations constitute the upper limit of human processing capacity. However, there are two mechanisms that can help individuals overcome this processing barrier: segmentation and chunking. Segmentation consists in breaking excessively complex tasks down into several steps, while chunking consists in compressing two or more variables into one. RC theory also makes strong predictions regarding age-related changes in relational processing capacity, and their causal role in cognitive development. Developmental changes are described as roughly occurring as follows: children become able to process unary relations at a median age of one year, binary relations at two

years, ternary relations at five years, and quaternary relations at 11 years. This gradual growth of processing capacity is thought to increase the number of variables that children can link together in their mental models, which in turn broadens the scope of their conceptual development and reasoning abilities. Halford and colleagues estimated its approximate chronological sequence mainly on the basis of performances on preexisting developmental tasks (e.g., class inclusion, theory of mind, transitivity, sentence comprehension, balance), to which analyses of complexity were applied (e.g., Andrews & Halford, 2002). Cognitive aging researchers have described a symmetrical decline in relational processing capacity, in that it seems to follow the developmental trajectory in reverse (Viskontas, Holyoak, & Knowlton, 2005). RI therefore lies at the heart of many debates over the mechanisms responsible for both cognitive growth and decline. However, surprisingly few efforts have been made to design specific tasks to assess this dimension of WM capacity, compared with the plethora of span measures dedicated to the assessment of its storage component (Conway et al., 2005). We now briefly review the few attempts that have been made in this direction.

Bunge et al. (2009) expanded on a task originally designed by Christoff, Ream, Geddes, and Gabrieli (2003) to create the Relational Matching Task (RMT). The stimulus display features four shapes with varying textures. Two types of trials are instantiated by specific questions that orient the participants' attention toward the detection of a feature similarity between two objects versus the detection of a *relational* similarity between two *pairs* of objects. This task, which nicely embodies the higher-order nature of relational processing, was designed to study the neurological correlates of basic RI processes, rather than individual differences in RI. In order to explore the respective contributions of storage and integration processes to fluid intelligence, Oberauer, Süß, Wilhelm, and Wittmann (2003, 2008) designed a series of RI tasks. In a verbal task, participants had to detect whenever

rhyiming words occurred next to each other in the same row of a matrix. In a numerical task, participants had to watch out for rows of numbers with the same last digit. Finally, in a spatial coordination task, they had to monitor the direction of aeroplanes in a flight control display, taking their relative speed and trajectory into account. Despite variations in the nature of the information that had to be processed (verbal, numerical, or spatial), all these tasks were designed to recruit the participants' ability to construct links between different items of information and mentally integrate them. This use of multiple tasks with varying contents fits well with the research approach of Oberauer and colleagues, which is based on structural equation modeling, that is, aggregating the scores on several tasks in order to suppress specific variance and reinforce the variance of the composite variable. However, this type of design is difficult to use in developmental research, which crucially requires that all tasks be equally applicable to a broad age range. Furthermore, the tasks administered by Oberauer and colleagues did not clearly specify or manipulate the level of RC to which participants were exposed.

The study by Birney, Halford, and Andrews (2006) represented a major step forward in this respect, as they developed a new experimental task that was explicitly derived from RC theory, namely the Latin Square Task (LST). The task is based on a matrix of 16 cells that can each be filled with one of four different geometric shapes. It works on the same principle as Sudoku problems, in that each shape must appear only once in each row or column. Participants are shown an incomplete matrix and asked to determine which of the shapes should be placed in a target cell. The LST had two important qualities with regard to RC theory requirements. First, the task minimizes the information that has to be held in memory, and consequently emphasizes the role of RI as opposed to storage. Second, it relies on one simple rule (suitable for a broad range of ages and abilities) that can be applied to items of



varying complexity. The RC of LST items is manipulated by controlling the numbers of rows and columns that need to be simultaneously considered in order to choose the right shape for the target cell. Although the LST has been administered to participants from various age groups (Birney et al., 2006; Zeuch, Holling, & Kuhn, 2011; Zhang, Xin, Lin, & Li, 2009) using it to measure RI in children raises specific difficulties. More recent study (Perret, Bailleux, & Dauvier, 2011) indicated that the influence of RC on children's performances is mediated by the differential use of specific deductive strategies. As the LST is a deductive task, it is particularly sensitive to variations in reasoning pathways or strategic approaches that may confound the effect of relational processing constraints through segmentation or chunking. Mindful of the sound principles on which the LST was built, we therefore designed a new experimental task that avoids previous pitfalls in the measurement of RI, in order to study its development and its relationship with fluid intelligence.

### **The Relational Integration Level Assessment Task (RILAT) and the present experiment**

The Relational Integration Level Assessment Task (RILAT) shares a number of features with the LST: a) freedom from the influence of prior knowledge; b) an emphasis on processing, as opposed to storage; and c) reliance on a single rule. The task features 12 items with increasing RC levels (binary, ternary, and quaternary). Each item contains one or several geometrical figures, each with up to four features (shape, color, texture, and number of stars). Furthermore, each feature has three modalities (shape: circle, square, or triangle; color: red, green, or blue; texture: filled, empty, or striped; and up to three stars). The task consists in finding a new figure to replace a question mark linked to one or more other figures by respecting a single rule: the response must share one and only one feature with each of the other figures. In the binary example in Figure 1 (red circle), a red square or a blue circle

would be correct but a red circle (two shared features) or a blue square (no shared feature) would be incorrect. This item is regarded as binary because two features (shape and color) have to be considered in interaction in order for the problem to be solved.

*Insert Figure 1 about her*

To limit the possibility of segmentation and chunking, which would reduce the integration demand, the problem state space needed to have an interaction structure. Table 1 shows the nine possible responses for the red circle in the binary example, indicating the number of shared features for each one. The correct responses are those that only share one feature with the red circle. The solution could have any one of the three colors or any one of the three shapes, but shape and color interact, such that the color of the response depends on its shape, and vice versa. Table 1 is made up of four areas and can be likened to the exclusive or (XOR) truth table. The XOR problem is known not to be linearly separable and reflects a kind of interaction between two dimensions. The correct and incorrect responses cannot be separated by a straight line, it is two-dimensional and cannot be simplified to a lower level of complexity. To learn to solve this problem, an artificial neural network needs one more layer than it does for AND or inclusive OR logical operations (Beale & Jackson, 1990). Returning to the RILAT, within a given feature (shape or color) some modalities can be chunked, as shown in Table 2. This chunking does not reduce the complexity level, but shows that the same complexity could be achieved with only two modalities per feature. The critical point in determining the dimensionality of the problem is neither the number of figures nor the number of modalities within a feature, but the number of features. A binary item need have only one visible figure, and providing more than three colors or shapes does not increase RC.

*Insert Tables 1 and 2 about here*

Ternary items comprise two figures with three features (shape, color, and texture), but

in order to solve these items, participants have to take the features of three figures into account: the two figures that are already visible, and the figure that is under construction. Table 3 sets out the three features of these three figures for the ternary example in Figure 1. Participants also have to ensure that the figure under construction has one link with each of the two target figures. These links are underlined in Table 3.

*Insert Table 3 about here*

The ternary items are intended to create an interaction pattern between the three features (shape, color, and texture) that make up the three dimensions of the interaction in the problem space. Table 4 shows all the possible responses for the ternary example (Fig. 1), indicating the numbers of features shared with each of the two figures. Out of the 27 ( $3^3$ ) possible responses, six are correct (i.e., they share one and only one feature with each figure) and are highlighted in bold type in the table. Each layer of the three-dimensional array contains at least one solution, and none of the rows or columns contains three solutions. This means that in the resolution process, no modality can be definitively rejected or retained independently of the other features. Correct responses are not a linearly separable subset of the three-dimensional array. No plane can intersect with the cube to separate the right and wrong responses, as one would expect in a three-way interaction pattern. Another way to figure out the interaction pattern is to have a look at the set of correct responses (see Table 5). When choosing the texture, both shape and color have to be taken into account. If the ongoing answer is green, it can be either filled or striped, depending on its shape (circle or triangle). If it is a green circle, then it must be filled, but if it is green but not a circle, it cannot be filled. Within the set of correct responses for a given item, the same two modalities of different features (e.g., circle and green) never appear together more than once. This means that an association between, say, a shape and a color may or may not be correct, depending

on the texture, and vice versa. This configuration corresponds to a three-way interaction between shape, color, and texture. The problem is thus three-dimensional and cannot be reduced to a lower level. Quaternary items are built on the same principle, by adding a third figure to each item and a new feature (one, two or three stars).

*Insert Tables 4 and 5 about here*

The main objective of the present study was to investigate the development of RI during childhood by means of this new task. A set of twelve items arranged in order of complexity level was created and administered to children aged 5-12 years. We specifically wanted to identify the two transitional periods that correspond to the emergence of the ability to deal with ternary and quaternary relations. As RI seems to play a key role in reasoning and fluid intelligence, children also completed Raven's Standard Progressive Matrices (SPM; Raven, 1938). The dataset allowed us to investigate the development of RI and its relation with fluid intelligence.

## **Method**

### **Participants**

Participants included 184 children aged 5-12 years ( $M = 98.4$  months,  $SD = 25.2.3$  months, range = 59-146 months), of whom 54% were girls and 46% were boys. These children were recruited from a preschool, an elementary school and a middle school located in a small town in southeastern France. Most participants were Caucasian and came from middle- to upper-class backgrounds, although data on ethnicity and socioeconomic status were not collected. Parental consent was obtained for all participants. Participants were tested individually in a quiet room at their school.

### **Materials and Procedure**

The children completed the RILAT and Sets B, C and D of the SPM. We believed that the

SPM were more suitable than Raven's Colored Progressive Matrices, as they would avoid the probable ceiling and floor effects arising from the sample's broad age range. As we nonetheless expected to see little variance for the youngest children, only the children in Grades 3-6 performed the SPM.

### **Material**

The RILAT comprised 12 items arranged in order of complexity, with four items per level of complexity. They were built by changing the features (shape, color, texture, and number of stars) on the basis of several frameworks that were checked for the numbers of correct responses they allowed. The number of correct responses, like the number of possible responses, varied according to the item's level of complexity. Binary items had a chance level of 0.44, but they were not intended to be discriminant, given the children's age range. The chance level for ternary and quaternary items never exceeded 0.25. For each item, participants were reminded of the three possible shapes, colors and, for the ternary and quaternary items, textures and numbers of stars, between which they had to choose.

### **Procedure**

The RILAT was administered individually, with each item printed on a separate sheet of paper. The test phase was preceded by a detailed instruction phase. First, a binary example (red triangle) was shown to the children and they were told that they had to find a figure to replace the question mark, abiding by one important rule: "Your response must only have one point in common with this figure." Their attention was drawn to the different features and to the modalities of each feature: "In this example, there are two possible points in common: shape or color. To replace the question mark, you must choose a shape (either a circle, a triangle, or a square) and a color (red, green, or blue)." The experimenter then gave them an example of an incorrect response, explaining why it broke the rule, and an example of a

correct response, explaining why it fitted. The same procedure was repeated with a ternary and a quaternary example. Finally, the children were given a binary example and asked to find the solution themselves. They had to supply their responses verbally and were prompted to confirm their “final offer” before proceeding to the next item. Self-corrections were accepted until the response was validated. If there was no response within two minutes, an item was deemed to have failed.

## Results

Univariate and bivariate descriptive statistics of correct response frequency by complexity level are provided in Table 6. For the whole sample, theoretical RC had a considerable effect on success rates. The mean correct response rate was 3.76 (out of 4) for binary items, 2.23 for ternary items and 1.29 for quaternary items. For the binary items, low variance and skewness reflected a ceiling effect that reduced the correlations with the other variables. The RILAT’s internal consistency was good (Cronbach’s alpha = 0.79) and the first factor of a principal components analysis based on the variance-covariance matrix explained over 40% of the total variance. The high correlation between Age and the total score ( $0.738, p < 0.01$ ) told us that RI, as measured by the RILAT, was closely linked to age between 5 and 12 years. Both ternary ( $r = 0.677, p < .01$ ) and quaternary items ( $r = 0.675, p < 0.01$ ) contributed to the overall correlation, but the binary items were less sensitive for this age group ( $r = 0.272, p < 0.01$ ). The correlation between the RILAT and SPM scores was also high ( $r = 0.671, p < 0.01$ ), consistent with the link between the SPM scores and age ( $r = .68, p < 0.01$ ). Table 7 indicates mean accuracy by age and level of complexity. Descriptively, a clear developmental effect appeared regarding both ternary and quaternary items. Expectile intervals (Jones, 1994; cf. note in Table 6) indicated a high level of individual differences. This variability was stable across age. The 80% interval had a width of around 0.40 between

6 and 10 years for the ternary items, and between 7 and 12 years for quaternary items.

*Insert Tables 6 and 7 about here*

### **Developmental Trend in Relational Integration**

To achieve a more fine-grained analysis of developmental changes in RI, we adopted a generalized linear mixed effect model (GLMM) approach using the lme4 package (Bates & Sarkar, 2009) in R (R Development Core Team, 2009). This allowed us to model response accuracy in relation to age and theoretical complexity level. The model used a binomial distribution with a logistic link function and a random intercept by participant. Age was standardized before being introduced in the model. Estimated parameters are provided in Table 8. Ternary items were taken as the intercept and used as a benchmark. Binary items appeared significantly easier than ternary ones (Binary-Ternary = 3.05,  $p < .01$ ), and quaternary items were significantly more difficult than ternary ones (Ternary-Quaternary = 1.51,  $p < .01$ ). The effect of age was significant at all three complexity levels. Comparisons with simpler, constrained models revealed that the slopes for ternary and quaternary items did not differ significantly from each other, while binary items were less sensitive to age, probably due to a ceiling effect.

A graphic representation of the fixed part of the model is given in Figure 2. The improvement in performance was clearly linked to age, but was relatively smooth at the group level. Individual differences in performance at a given age may have contributed to the smoothness of the developmental trends observed at the group level. A look at individual profiles revealed that for more than 70% of the children, there was a difference of fewer than two points between their performances on ternary and quaternary items. Very few children (14.6%) were highly successful at the ternary level (scores of 3 or 4), and completely failed the quaternary level at the same time (scores of 0 or 1), suggesting a smooth transition even

at the individual level.

*Insert Table 8 about here*

We divided the developmental trends into three phases for ease of presentation. For ternary and quaternary items, the chance level of success was slightly below 0.25 (e.g., six of the 27 possible responses for the ternary item in Figure 1 were deemed to be correct). Rates below this level were therefore considered to reflect random responding. In the early success phase, accuracy ranging from 0.25 to 0.5 reflected responding above chance level, but still less than half the time. Accuracy above 0.75 corresponded to mastery of the RI level. There was also a transitional phase, where accuracy ranged from 0.5 to 0.75, possibly corresponding to process stabilization.

The probability of success on the binary items was very high, even for the youngest children, meaning that all the children were in the mastery phase for these items. Regarding ternary items, children found the correct response in one in four trials ( $p = 0.25$ ) at around 72 months (6 years). After 6 years, children thus appeared to perform above chance level. A 0.50 probability of success was reached at around 92 months (7.6 years). After 110 months (9.15 years) the probability of success was greater than 0.75, meaning that most of the children resolved these items most of the time. The slopes of the ternary and quaternary items were roughly similar, such that the improvement in performance on quaternary items occurred at the same pace, but 30 months (2.5 years) later. Items were solved in one in four trials at 100 months (8.3 years). A 0.50 probability was reached at 120 months (10 years) and probability of 0.75 at 140 months (11.6 years). For ternary as for quaternary items, it took around 40 months (3.33 years) to go from the early signs of success to complete mastery. The two transitional phases occurred between 92 and 110 months (7.6-9.15 years) for ternary items, and between 120 and 140 months (10-11.6 years) for quaternary items.



*Insert Figure 2 about here*

### **Relational Integration and its Relationship with Fluid Intelligence**

The RILAT and SPM scores were closely correlated with age in our sample (0.74 and 0.68, respectively). The relationship between scores on ternary and quaternary items and SPM scores was analyzed after controlling for the effect of age with varying-coefficient models. Varying-coefficient models are multiple regression models where some coefficients can vary across time. In our case, the dependent variable was SPM scores and the explanatory variables were age and scores on the ternary and quaternary items. Binary items were removed from this analysis due to the ceiling effect mentioned earlier. As age was one of the explanatory variables, the estimated influence of ternary and quaternary scores was conditional upon age. Moreover, the regression coefficients of these variables were allowed to change as a function of age. In technical terms, two interaction terms between age (expressed in the form of its first natural quadratic splines) and scores on ternary and quaternary items were introduced in the linear model to allow for smooth nonlinear variations in the coefficients. The first step was to determine the optimum amount of smoothing. To this end, several varying-coefficient models with increasing degrees of freedom of the natural quadratic spline were fitted to the data. The AIC model selection criterion pointed to a model with three degrees of freedom (M3 exhibited the lowest AIC; cf. Table 9) that explained 71% of the variance of the SPM scores.

*Insert Table 9 about here*

*Insert Figure 3 about here*

The estimated standardized regression coefficient of age on SPM scores was .44 ( $p < .01$ ). Figure 3 shows changes in the links between scores on the ternary and quaternary items and SPM scores as a function of age. All the variables were standardized before being

introduced in the model, so that the scale of the y-axis corresponded to standardized coefficients ( $\beta$ ). The contribution of ternary items peaked at around 100 months (8.3 years,  $\beta = 0.70$ ) and strongly decreased after 110 months. It was slightly negative after 120 months, probably reflecting a methodological artifact related to the quadratic spline smoothing, which cannot follow an asymptotic trajectory to zero with a reasonable degree of freedom.

Quaternary items were influential throughout this period, but a little more so at around 130 months (10.8 years,  $\beta = 0.39$ ), when ternary items were at their lowest level. When Figures 2 and 3 are combined, we see that the two peaks were held when the probabilities of a success were around 0.65 for ternary and for quaternary items (i.e., during their transitional phases). When the mean levels and individual differences are considered together, two sensitive periods emerge. The first one presumably corresponds to the ternary transition and lasts from 90 (7.5 years) to 110 months (9.15 years), when the probability of success ranges from 0.5 to 0.75. The second period presumably corresponds to the quaternary transition, and lasts from 120 months (10 years) to 140 months (11.6 years).

*Insert Figure 4 about here*

Figure 4 illustrates the changes in the R-squared value of the varying-coefficient model with age. Two rough estimates of the R-squared conditional upon age were computed using different approaches. The solid line is the sum of the squared  $\beta$  of age and the squared  $\beta$  values of Figure 3. The dots' coordinates were computed by sliding a window across the data sorted by age. For the first dot, the x value was the mean age of the 19 youngest children and the y value was the squared correlation between the SPM scores and the fitted value according to M3 for the same children. The window was slid across the whole dataset to obtain the scatterplot. The two approaches were in good agreement with each other. At around 100 months (8.3 years), SPM performances were strongly correlated RI ( $R^2 = 0.75$ ),

as assessed by the ternary items. The influence of RI decreased after this peak, but started to increase again at around 120 months (10 years), when the probability of success for quaternary items reached 0.5. A second, smaller peak ( $R^2 = 0.42$ ) occurred at around 130 months (10.8 years), corresponding to the strongest influence of the quaternary items.

## Discussion

The aim of the present study was twofold. First, we wanted to explore developmental changes in RI, as measured by a task we designed specifically to tap children's ability to mentally coordinate several variables. Second, we wanted to investigate both the extent to which this ability was related to fluid intelligence, and age-related changes in this relationship.

The findings of our experiment broadly support the developmental sequence predicted by Halford's RC theory. First of all, item difficulty in the RILAT task systematically varied with RC: the mean proportion of correct responses decreased as a function of the number of variables. As expected, binary items did not pose any difficulty for the children in our sample: we did not expect to see any further development in the processing of binary relations across the age range we tested. We observed early successes on ternary items among the 5-6-year-olds, consistent with previous studies indicating that children start to grasp ternary relations at around this age. Andrews and Halford (2002), for instance, found that tasks from different domains (transitivity, class inclusion, or relative clause-sentence comprehension) that had all been deemed to have ternary relations were successfully processed at a median age of 5 years. Similar results have been reported with theory of mind tasks (Andrews, Halford, Bunch, Bowden, & Jones, 2003). More surprising, the earliest

successes on RILAT's quaternary items were observed among children aged 8-9 years in our sample. Relatively few studies have explored children's processing of quaternary relations. In their influential article setting out the core assumptions of RC theory, Halford and colleagues (1998) mentioned two domains of particular relevance for the study of quaternary relations: proportions and the balance scale. Proportions are described as quaternary because they involve the processing of relations that entail four arguments ( $a/b = c/d$ ). Although Inhelder and Piaget (1958) regarded the processing of proportions as a late acquisition (not before 11 years of age), more recent research suggests that under favorable conditions, children can exhibit valid proportional reasoning somewhat earlier (Boyer, Levine, & Huttenlocher, 2008). Regarding the complexity of the balance scale, Halford and colleagues (Andrews, Halford, Murphy, & Knox, 2008; Halford, Andrews, Dalton, Boag, & Zielinski, 2002) argued that problems in which both weight and distance vary on each side entail quaternary relations. The authors then noted that these types of balance task items are rarely solved before the age of 11 years. However, solving them requires not only the processing of quaternary relations, but also the comprehension and appropriate application of the torque rule (i.e., comparing the product of the weight and distance on the left side with the product of the weight and distance on the right side). This additional requirement could explain the developmental lag between the early success we observed on the RILAT's quaternary items and correct responses on the balance scale.

In addition to yielding these new data documenting the age at onset for each level of RI, our study highlighted two important developmental phenomena. First, the age norms suggested by our data should not disguise the breadth of individual variability. As indicated by the 80% expectile intervals in Table 7, there was considerable individual variability in the pace at which the children reached the successive levels of RI. Second, as illustrated in

Figure 2, the development of RI is not a series of abrupt, stage-like transitions. Rather, at both the group and individual levels, there is a smooth and gradual increase in the probability of success. For quaternary and ternary relations, comparable 40-month periods separated the early signs of success and complete mastery. Thus, even if transitions from one level to the next mark important developmental milestones, the development of RI also relies substantially on the child's ability to allocate optimum processing capacity to several different items in a sustained fashion.

The present experiment also adds to the growing literature on the influence that WM constraints exert on the development of fluid intelligence (Tillman, Nyberg, & Bohlin, 2008). As in the adult literature, however, the role of RI is often neglected on the developmental front: alternative views tend to focus on the storage (Hornung, Brunner, Reuter, & Martin, 2011) or executive (Engel de Abreu, Conway, & Gathercole, 2010) components of WM. Our results indicate that the relationship between RI and fluid intelligence is relatively close across childhood, but does vary according to age. Figures 3 and 4 show two peaks corresponding to two periods of stronger links between RI and SPM scores, the first one being the strongest. The first period lies in what we identified as the transitional phase for ternary items (90-110 months, around 8 years), where regression coefficients between performances on ternary items and SPM scores reached around 0.70. A few months later, during the transitional phase for quaternary items (120-140 months, around 10 years), the link between SPM and performances on quaternary items became stronger again, reaching values of around 0.40.

One possible interpretation for this is that RI and fluid intelligence enjoy an asymmetrical relationship. With reference to Pascual-Leone's theory, de Ribaupierre and Lecerf (2006),

suggested that the M factor (WM capacity) is necessary but not sufficient to resolve complex tasks such as Raven's progressive matrices. De Ribaupierre and colleagues highlighted a relation of implication between reasoning tasks and WM tasks, but concluded that the importance of WM declines with practice, in favor of other factors. The M factor can therefore be regarded as a developmental factor, and the others as individual factors. Similarly, some levels of RI may be necessary but not sufficient to resolve some SPM items. This implicative relation could explain why the closest correlation between RILAT and SPM scores appeared during the two transitional phases for ternary and quaternary items, rather than during the early success phases. The interindividual variance in the RILAT scores for a given level was of the same magnitude in both these phases. The transitional phases corresponded to mean success rates of 0.50-0.75 at the group level, but at the individual level, the 80% expectile intervals provided in Table 7 indicate that individual success rates ranged from around 0.40 to around 0.90. Thus, during these periods, while some children failed on more than half the items, others successfully resolved most of them. If mastery of ternary relations were a necessary condition for successfully performing some of the SPM items, the transitional phases would exhibit higher SPM-RI correlations, while some children would have fulfilled this necessary condition, others would still be a long way away from doing so. The smaller contribution of quaternary items, compared with ternary ones, could be explained by the growing influence of individual differences in some of the other processes tapped by the SPM, such as executive control, which is known to undergo critical changes at around 10 years.

In a recent study, Wiley, Jarosz, Cushen, and Colflesh (2011) emphasized the potential role of executive functioning to explain the correlation between performances on Raven's Advanced Progressive Matrices (RAPM) and scores on operation span (OSpan) and reading

span tasks. They showed that this correlation was considerably stronger for RAPM items containing rule combinations that appeared for the first time in the task, than for items that recycled previously discovered rules. According to the authors, high WM participants displayed a greater ability to inhibit previously used rules and avoid perseverative errors. Individual differences on WM tasks such as reading span and OSpan can be interpreted as stemming from the need to rely on executive abilities to maintain and retrieve task-relevant information in the presence of interference arising from the multiple activities (e.g., resolving the operation and learning a list of words in OSpan) that these tasks impose on participants (Unsworth & Engle, 2007). Executive functioning is probably also involved in resolving RILAT items, but our main hypothesis is that RI, rather than executive functioning, is responsible for the relationship we observed between RILAT and SPM scores. One possibility is that a degree of RI is needed to handle the multiple rules involved in resolving SPM items. This rule/capacity hypothesis predicts stronger correlations between RI assessments and SPM items involving more rules, but this prediction has been refuted by empirical results (Unsworth & Engle, 2005; Wiley et al., 2011). According to Wiley et al., the relevant feature is the novelty of the rule or combination of rules, rather than the sheer number of rules. As an alternative to Wiley et al.'s executive functioning hypothesis, we therefore postulate that the discovery of a new rule is more demanding because it requires simultaneously considering all the dimensions of the problem in order to identify the relevant relationships. Once the rules have been identified, segmentation and chunking become possible, thereby reducing the demand for RI. This hypothesis is also consistent with Wiley et al.'s finding that it is rule novelty, not rule change from one item to the next that also taps executive functioning, explaining the magnitude of the relationship between WM and reasoning.

Our findings are also compatible with ideas developed by Van der Maas (Van der Maas et al., 2006) within their dynamic model of the positive manifold of intelligence by mutualism. This model hypothesizes that the general factor of intelligence emerges as a byproduct of beneficial interactions between cognitive processes in the course of development. One example of such an interaction is the way in which better short-term memory helps us to develop better cognitive strategies, which reciprocally enhance short-term memory. This kind of relationship probably exists between RI and other processes involved in reasoning tasks, such as executive function or mental model formation. A higher level of RI would allow for the emergence of better mental models which, in return, would reduce the task's WM demands. Within this framework, the changes in the correlation between RILAT and SPM scores across childhood would reflect stronger interactions during the transitional phases, producing specific individual differences where some children become capable of building higher-level representations of SPM items and others do not. Our idea is that individual differences in fluid intelligence may be determined by varying aptitudes, depending on the point the child has reached in the developmental course. This fits in well with Van der Maas' model of intelligence through mutualism.



## References

- Andrews, G., & Halford, G. S. (2002). A cognitive complexity metric applied to cognitive development. *Cognitive Psychology, 45*, 153-219.
- Andrews, G., Halford, G. S., Murphy, K., & Knox, K. (2008). Integration of weight and distance information in young children: The role of relational complexity. *Cognitive Development, 24*, 49-60.
- Bates, D., & Sarkar, D. (2009). *lme4: Linear mixed-effects models using S4 classes*. Retrieved from <http://CRAN.R-project.org/>
- Beale, R., & Jackson, T. (1990). *Neural computing - An introduction*. Bristol: Adam Hilger.
- Birney, D. P., Halford, G. S., & Andrews, G. (2006). Measuring the influence of complexity on relational reasoning: The development of the Latin Square Task. *Educational and Psychological Measurement, 66*, 146-171.
- Bunge, S. A., Helskog, E. H., & Wendelken, C. (2009). Left, but not right, rostrolateral prefrontal cortex meets a stringent test of the relational integration hypothesis. *NeuroImage, 46*, 338-342.
- Bunge, S. A., Wendelken, C., Badre, D., & Wagner, A. D. (2005). Analogical reasoning and prefrontal cortex: Evidence for separable retrieval and integration mechanisms. *Cerebral Cortex, 15*, 239-249.
- Cattell, R. B. (1950). *Culture fair intelligence test: Scale 1*. Champaign, IL: Institute for Personality and Ability Testing.
- Christoff, K., Ream, J. M., Geddes, L. P., & Gabrieli, J. D. (2003). Evaluating self-generated information: Anterior prefrontal contributions to human cognition. *Behavioral Neuroscience, 117*, 1161-1168.
- Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R.

- W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review*, *12*, 769-786.
- De Ribaupierre, Lecerf
- Doumas, L. A. A., Hummel, J. E., & Sandhofer, C. M. (2008). A theory of the discovery and predication of relational concepts. *Psychological Review*, *115*, 1-43.
- Goodwin, G. P., & Johnson-Laird, P. N. (2005). Reasoning about relations. *Psychological Review*, *112*, 468-493.
- Halford, G. S., & Andrews, G. (2006). Reasoning and problem solving. In W. Damon, R. Lerner, D. Kuhn, & R. S. Siegler (Eds.) *Handbook of child psychology: Vol 2. Cognition, perception, and language* (pp. 557-608). Hoboken, NJ: Wiley.
- Halford, G. S. (1999). The development of intelligence includes the capacity to process relations of greater complexity. In M. Anderson (Ed.), *The development of intelligence* (pp. 193-213). Hove: Psychology Press.
- Halford, G. S., & Andrews, G. (2004). The development of deductive reasoning: How important is complexity? *Thinking and Reasoning*, *10*, 123-145.
- Halford, G. S., Cowan, N., & Andrews, G. (2007). Separating cognitive capacity from knowledge: A new hypothesis. *Trends in Cognitive Sciences*, *11*, 236-242.
- Halford, G. S., Wilson, W. H., & Phillips, S. (1998a). Processing capacity defined by relational complexity: Implications for comparative, developmental, and cognitive psychology. *Behavioral and Brain Sciences*, *21*, 803-865.
- Johnson-Laird, P. N. (1983). *Mental models*. Cambridge: Cambridge University Press.
- Jones, M. C. (1994). Expectiles and M-quantiles are quantiles. *Statistics and Probability Letters*, *20*, 149-153.
- Kane, M. J., Hambrick, D. Z., & Conway, A. R. A. (2005). Working memory capacity and

- fluid intelligence are strongly related constructs: Comment on Ackerman, Beier, and Boyle (2004). *Psychological Bulletin*, *131*, 66–71.
- Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, *63*, 81-97.
- Oberauer, K. (2009). Design for a working memory. *Psychology of Learning and Motivation*, *51*, 45-100.
- Oberauer, K., Süß, H.-M., Wilhelm, O., & Wittmann, W. W. (2003). The multiple faces of working memory—Storage, processing, supervision, and coordination. *Intelligence*, *31*, 167–193.
- Oberauer, K., Süß, H.-M., Wilhelm, O., & Wittmann, W. W. (2008). Which working memory functions predict intelligence? *Intelligence*, *36*, 641-652.
- R Development Core Team. (2009). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. Retrieved from <http://www.R-project.org>
- Raven, J. C. (1938). Standard Progressive Matrices (1956 revision). Issy-les-Moulineaux: Editions Scientifiques et Psychotechniques.
- Spearman, C. (1927). *The nature of "intelligence" and the principles of cognition*. London: MacMillan.
- Sweller, J. (1998). Can we measure working memory without contamination from knowledge held in long- term memory? *Behavioral and Brain Science*, *24*, 845-846.
- Unsworth, N., & Engle, R. W. (2005). Working memory capacity and fluid abilities: Examining the correlation between Operation Span and Raven. *Intelligence*, *33*(1), 67–81.  
doi:10.1016/j.intell.2004.08.003
- Unsworth, N., & Engle, R. W. (2007). The nature of individual differences in working

memory capacity: Active maintenance in primary memory and controlled search from secondary memory. *Psychological Review*, *114*(1), 104–132. doi:10.1037/0033-295X.

114.1.103

Van der Maas, H. L. J., Dolan, C. V., Grasman, R. P. P. P., Wicherts, J. M., Huizenga, H. M., & Raijmakers, M. E. J. (2006). A dynamical model of general intelligence: The positive manifold of intelligence by mutualism. *Psychological Review*, *113*(4), 842–861. doi:

10.1037/0033-295X.113.4.842

Viskontas, I. V., Holyoak, K. J., & Knowlton, B. J. (2005) Relational integration in older adults. *Thinking & Reasoning*, *11*, 390-410.

Waltz, J. A. (2006). Impairments of memory and reasoning in patients with neuropsychiatric illness: Disruptions of dynamic cognitive binding. In R. Engle, G. Sedek, U. von Hecker, & D. McIntosh (Eds.) *Cognitive limitations in aging and psychopathology* (pp. 346-369). Cambridge: Cambridge University Press.

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*, 119-125.

Wiley, J., Jarosz, A. F., Cushen, P. J., & Colflesh, G. J. H. (2011). New rule use drives the relation between working memory capacity and Raven's Advanced Progressive Matrices. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(1), 256–263.

doi:10.1037/a002161

Wilkinson, G. N., & Rogers, C. E. (1973). Symbolic descriptions of factorial models for analysis of variance. *Applied Statistics*, *22*, 392–9.

Zhang, L., Xin, Z., Lin, C., & Li, H. (2009). The complexity of the Latin Square Task and its influence on children's performance. *Chinese Science Bulletin*, *54*, 766-775.

## Footnotes

Footnote 1.

Expressed in Wilkinson and Rogers' (1973) notation used in R, the model equation was *SPM*

$\sim \text{Age} + \text{Ternary} + \text{Quaternary} + \text{ns}(\text{Age}, \text{df} = 0-4) : (\text{Ternary} + \text{Quaternary})$

The ns() function was used to compute the first natural splines of age.

Table 1. *Number of Features that Each of the Nine Possible Responses Shares with the Target*

*Figure (Red Circle) in the Binary Example in Figure 1*

	Red	Green	Blue
Circle	2	<b>1</b>	<b>1</b>
Square	<b>1</b>	0	0
Triangl e	<b>1</b>	0	0

Table 2. *Sets of Correct Responses for the Binary Example in Figure 1*

	Shape	Color
Solution 1	Circle	Green Blue
Solution 2	Square Triangle	Red

Table 3. *Relevant Information for Solving the Ternary Example in Figure 1*

	Figure 1	Figure 2	Response
Shape	Square	<u>Circle</u>	<u>Circle</u>
Color	<u>Green</u>	Red	<u>Green</u>
Texture	Empty	Striped	Filled



Table 4. Number of Features Shared by the 27 Responses with the two Target Figures (Green Empty Square and Red Striped Circle) in the Ternary Example in Figure 1

	Empty			Striped			Filled		
	Red	Green	Blue	Red	Green	Blue	Red	Green	Blue
Circle	1-2	2-1	<b><u>1-1</u></b>	0-3	1-2	0-2	0-2	<b><u>1-1</u></b>	0-1
Square	2-1	3-0	2-0	1-2	2-1	<b><u>1-1</u></b>	<b><u>1-1</u></b>	2-0	1-0
Triangle	<b><u>1-1</u></b>	2-0	1-0	0-2	<b><u>1-1</u></b>	0-1	0-1	1-0	0-0

Table 5. *Correct responses for the Ternary Example in Figure 1*

	Shape	Color	Texture
Solution 1	Circle	Green	Filled
Solution 2	Circle	Blue	Empty
Solution 3	Square	Red	Filled
Solution 4	Square	Blue	Striped
Solution 5	Triangle	Red	Empty
Solution 6	Triangle	Green	Striped

Table 6. *Descriptive Statistics of the Numbers of Correct Responses by Complexity Level and Bivariate Correlations*

	Mean	SD	Skewness	Age	Binary	Ternary	Quaternary	RILAT Total
Age	98.42	25.23	0.29					
Binary	3.76	0.58	-20.82	0.272**				
Ternary	2.23	1.48	-0.25	0.677**	0.217**			
Quaternary	1.29	1.31	0.65	0.675**	0.274**	0.681**		
RILAT Total	7.28	2.77	0.06	0.738**	0.455**	0.902**	0.894**	
SPM	36.64	8.94	-0.16	0.68**	0.361**	0.526**	0.613**	0.671**

Note. \*\*:  $p < 0.01$

Table 7. Mean Accuracy by Age and Complexity Level

Age in years	Age in months	<i>n</i>	Binary	Ternary	Quaternary
5	66.03	31	0.87 (0.61, 0.97)	0.25 (0.05, 0.56)	0.04 (0.00, 0.15)
6	77.22	37	0.91 (0.71, 0.98)	0.22 (0.04, 0.50)	0.11 (0.01, 0.29)
7	89.77	26	0.93 (0.78, 0.99)	0.59 (0.37, 0.79)	0.25 (0.07, 0.42)
8	102.58	24	0.97 (0.88, 0.99)	0.69 (0.50, 0.88)	0.42 (0.16, 0.66)
9	115.04	24	0.94 (0.82, 0.99)	0.70 (0.37, 0.93)	0.44 (0.19, 0.70)
10	126.87	15	1.00 (1.00, 1.00)	0.80 (0.51, 0.96)	0.48 (0.15, 0.80)
11	136.95	21	0.98 (0.92, 0.99)	0.91 (0.73, 0.98)	0.72 (0.50, 0.92)
12	145.00	6	1.00 (1.00, 1.00)	0.95 (0.83, 0.99)	0.77 (0.36, 0.90)

*Note:* The intervals shown in brackets are 80% expectile intervals. Expectiles (Jones, 1994) are to the mean what quantiles are to the median. They provide a sensitive measure of individual differences in contexts where distributions are asymmetrical with numerous equal values.

Table 8. *GLMM Estimates of Accuracy by Age and Complexity*

Generalized Linear Mixed effect Model				
Accuracy ~ Complexity + Age: Complexity + (1 Participant)				
	Estimate	<i>SE</i>	<i>z</i> value	Pr(>   <i>z</i>  )
Ternary (Intercept)	0.40	0.12	3.29	< .01
Binary-Ternary	3.05	0.24	12.68	< .01
Ternary- Quaternary	1.51	0.14	10.76	< .01
Binary:Age	0.91	0.24	3.85	< .01
Ternary:Age	1.49	0.13	11.05	< .01
Quaternary:Age	1.37	0.13	10.32	< .01
Random effect	Variance = 0.98			

Table 9. *Goodness-of-Fit Statistics for the Varying-Coefficient Models*

Varying-coefficient model				
	Model	Model <i>df</i>	AIC	$R^2$
M0	Multiple regression	4	219.67	0.62
M1	Varying-coefficient linear	6	205.47	0.67
M2	Varying-coefficient quadratic spline ( <i>df</i> = 2)	8	205.98	0.68
<b>M3</b>	<b>Varying-coefficient quadratic spline (<i>df</i> = 3)</b>	<b>10</b>	<b>199.71</b>	<b>0.71</b>
M4	Varying-coefficient quadratic spline ( <i>df</i> = 4)	12	202.11	0.72

*Figure 1:* Examples of RILAT items testing the three levels of relational complexity.

*Figure 2.* Probability of success as a function of age by complexity level, according to the fixed part of the GLMM



*Figure 3.* Changes in the standardized regression coefficients of scores on ternary and quaternary items as a function of age, with SPM scores as the dependent variable

*Figure 4.* Estimated R-squared of the varying-coefficient model as a function of age.

Note: The moving window estimation was obtained by sorting the participants' scores by age and computing the squared correlation between their SPM scores and the fitted values given by the model on a window of 19 children. We slid this window through the data and plotted the R-squared values against the mean age of the subsample. The approximate R-squared is the sum of all the estimated  $\beta^2$  values as a function of age.