

Development of Relational Processing in Hot and Cool Tasks

Katie M. Bunch and Glenda Andrews*

Applied Cognitive Neuroscience Research Unit, Behavioural Basis of Health Program,

Griffith Health Institute

&

School of Applied Psychology

Griffith University, Gold Coast Campus, Queensland, Australia

*Corresponding author: Glenda Andrews, School of Psychology, Griffith University, Gold Coast Campus, 4222, Queensland, Australia.

Email: g.andrews@griffith.gu.edu.au

This research was supported by an Australian Postgraduate Award (K. M. Bunch) and an Australian Research Council Discovery Grant (DP 0452547)

Running Head: RELATIONAL PROCESSING IN HOT AND COOL TASKS

Abstract

The research investigated the role of complexity and the hot-cool distinction in cognitive development. The 120, 3- to 6-year-old children completed four hot tasks, which involved an affective component and three cool tasks, which did not. All tasks included binary- and ternary-relational items. Complexity was a major source of difficulty on all tasks, especially for younger children. Consistent with a hot-cool distinction, ternary-relational processing emerged earlier and more 4- and 5-year-olds mastered ternary-relational items in hot than cool tasks. Overall performance was better in hot than cool tasks at 4 years but this pattern was reversed at 6 years.

Development of Relational Processing in Hot and Cool Tasks

The human prefrontal cortex (PFC) has been implicated in higher cognitive processes including reasoning (Goel, 2007), intelligence (Shaw et al., 2006), rule use (Bunge & Zelazo, 2006), and relational processing (Waltz et al., 1999). Maturation of the PFC is thought to play a role in the development of these processes during childhood (Johnson & de Haan, 2011). The behavioral study reported here investigated two proposals relevant to PFC function in cognitive development namely, cognitive complexity and the hot-cool distinction. Previous research (e.g., Andrews & Halford, 2002; Frye, Zelazo & Palfait, 1995) has demonstrated the importance of cognitive complexity. Recent research has attempted to distinguish between developmental changes in so-called hot tasks, which include a significant affective component and so-called cool tasks, which do not. We investigated both factors in the same sample of 3- to 6-year-old children.

Two complexity theories that have been applied to cognitive development are cognitive complexity and control-revised (CCC-R) theory (Zelazo, Müller, Frye, & Marcovitch, 2003) and relational complexity (RC) theory (Halford, 1993; Halford, Wilson & Phillips, 1998). In CCC-R theory (Zelazo et al., 2003) there are four rule types of increasing complexity. The simplest stimulus-reward associations are observed in children as young as 30 months (Overman, Bachevalier, Schuhmann & Ryan, 1996). Conditional rules are more complex and emerge later. Children can use a pair of univalent rules (e.g., if red light, then stop; if green light, then go) from 3 years of age. In bivalent rule pairs, each stimulus is associated with two responses. Higher-order rules integrate pairs of bivalent rules into rule hierarchies and facilitate selection among task sets. For example, in the dimensional change card sorting (DCCS) task, a red boat is sorted with red flower in the color game, but with blue boat in the shape game. Children can use higher-order rules to integrate incompatible bivalent rule pairs from around 5 years (Zelazo, Jacques, Burack & Frye, 2002). Construction of more complex rules involves reprocessing of (and reflection on) less complex rules, and the recruitment of additional PFC subregions. Stimulus-reward associations involve the medial regions including the orbitofrontal cortex (OFC). Univalent and bivalent

conditional rules involve the ventrolateral (VLPFC) and the dorsolateral prefrontal cortex (DLPFC). Higher-order rules recruit the rostrolateral prefrontal cortex (Bunge & Zelazo, 2006). This mapping concurs with evidence about frontal lobe maturation. For example, gray matter maturation occurs earliest in the OFC, later in the VLPFC and later still in the DLPFC (Gogtay et al., 2004).

In RC theory (Halford, 1993; Halford, et al., 1998) complexity corresponds to the number of arguments in a relation or the number of interacting variables that constrain responses. Unary relations have a single argument as in class membership, $\text{dog}(\text{fido})$, binary relations have two arguments as in larger-than (elephant, mouse), ternary relations have three arguments as in addition ($2, 3, 5$), quaternary relations have four interacting components as in $2/3 = 6/9$, and so on. Processing load increases with complexity. The complexity of relations that children can process increases with age. The transition from binary to ternary relations, which is of most interest here, occurs at around 5 years of age (Andrews & Halford, 2002; Andrews, Halford, Bunch, Bowden & Jones, 2003; Halford, Andrews, Dalton, Boag & Zielinski, 2002a; Halford, Andrews & Jensen, 2002b). RC manipulations have been incorporated into studies of brain function. Adults with DLPFC lesions performed poorly on ternary-relational tasks requiring relational integration, but succeeded on comparable binary-relational tasks (Waltz et al., 1999). Brain imaging studies have implicated the DLPFC and RLPFC in relational integration. As complexity increased more anterior regions were recruited (Christoff et al., 2001; Christoff & Owen, 2006; Kroger et al., 2002)

Task analyses based on rule complexity and on relational complexity are translatable to some extent. For example, the DCCS requires higher order rules in CCC-R theory and is ternary-relational in RC theory (Andrews et al., 2003). The theories differ in at least three respects. Whereas CCC-R theory applies to hierarchically structured tasks, RC theory also applies to tasks with nonhierarchical structures. CCC-R theory has been applied to cognitive development mainly during the preschool period, whereas RC theory has been applied across the life span (e.g., Andrews, Halford, Murphy and Knox, 2009; Andrews, Birney & Halford, 2006; Andrews & Todd, 2008; Birney & Halford, 2002; Birney, Halford, & Andrews, 2006; Viskontas, Morrison, Holyoak,

Hummel, & Knowlton, 2004). Both theories incorporate a means for estimating complexity. RC theory also specifies a set of principles for applying the RC metric (Andrews & Halford, 2011). CCC-R emphasizes the role of reflection in developmental transitions whereas RC theory also acknowledges the importance of other mechanisms (Halford & Andrews, 2006). Our current purpose is to examine complexity effects rather than to decide between complexity approaches.

Traditional tests of PFC function are relatively abstract and decontextualized. Because they involve no obvious motivational/affective components, such tasks have been characterized as eliciting cool executive function (EF) (Zelazo & Müller, 2011). DCCS, class inclusion and transitive inference are cool tasks suitable for children. Tasks that entail a significant affective component are characterized as eliciting hot EF because the affective significance of rewards and/or punishments must be considered and/or updated. In object reversal tasks selection of one of two objects is rewarded, then the contingencies are reversed and selection of the previously non-rewarded object is rewarded. Other tasks that elicit hot EF are gambling (Bechara, Damasio, Damasio & Anderson, 1994), theory of mind (Zelazo & Müller, 2011), delay of gratification (Mischel, Shoda & Rodriguez, 1989) and delay discounting (McClure, Laibson, Loewenstein & Cohen, 2004). Both delay of gratification and delay discounting involve a trade-off between immediate and delayed rewards. Tasks that elicit hot or cool EF are distinguished in a relative rather than absolute manner. While it is probably not possible to identify pure measures of either hot or cool EF, it is generally accepted that some tasks elicit hot EF to a greater extent than others.

Tasks that elicit mainly cool EF have been linked to lateral PFC. For example, imaging studies have implicated lateral PFC in the DCCS (Crone, Donohue, Honomichl, Wendelken & Bunge, 2006; Moriguchi & Hiraki, 2009). Five PET and fMRI studies of explicit transitive inference (reviewed by Goel, 2007) reported increased DLPFC activation relative to baseline. Tasks that elicit mainly hot EF have been associated with ventral and medial prefrontal regions (VMPFC). Such tasks involve determining, updating, and integrating the reinforcement values of stimuli. These are primary functions of ventral and medial PFC regions (Elliot, Dolan & Frith, 2000; Rolls,

2004). In reversal learning tasks, human patients (Rolls, Hornoak, Wade, & McGrath, 1994) with medial PFC damage tend to perseverate on the initial discrimination and fail to reverse their responses. Patients with VMPFC lesions show impairments on gambling tasks (Bechara et al., 1994) in which participants select cards from four decks, with the aim of winning money. This involves integrating information about gains and losses to identify the advantageous (gains outweigh losses) and disadvantageous (losses outweigh gains) decks. Patients continue to select from disadvantageous decks, even when they can describe the contingencies (Bechara et al., 1994).

Theory of mind (ToM) refers to a set of concepts for representing mental states of ourselves and others. ToM tasks require flexible problem representation and reversal of responding in a social context (Zelazo & Müller, 2011). Several fMRI studies (reviewed by Johnson & de Haan, 2011) showed consistent activation of medial PFC in children during such tasks. Liu, Sabbagh, Gehring, and Wellman's (2009) ERP study revealed a late slow wave at left, medial, and right frontal sites in children who passed ToM tasks, but not in those who failed.

In delay of gratification choice tasks, children choose between a smaller immediate reward and a larger delayed reward. Such tasks are characterized as hot because they involve the arousing, affective qualities of rewards and the desired objects are hotly displayed (Metcalf & Mischel, 1999). These tasks also involve episodic future thinking (Atance & O'Neill, 2001) which is a function of the cool system (Metcalf & Mischel), raising the possibility that they have hot and cool components. Consistent with this, McClure et al.'s (2004) functional imaging study showed that valuation of immediate rewards during a delay discounting task activated medial PFC regions in adults whereas valuation of delayed rewards activated more lateral regions.

Medial and lateral PFC regions receive their blood supplies via different cerebral arteries and have different patterns of connectivity with other regions; however, they are interconnected (Johnson & de Haan, 2011). Neuropsychological research suggests that hot and cool EF are dissociable. Impairments in hot EF have been observed in adults with intact cool EF and vice versa (Bechara, Damasio, Tranel & Anderson, 1998). Research has begun to examine whether a similar

dissociation is present earlier in the life span. The dissociation might be evidenced by different patterns of age-related change, or different ages of mastery on hot versus cool tasks of the same complexity, or if hot and cool tasks loaded on separate factors in exploratory factor analyses.

Hongwanishkul, Happaney, Lee, and Zelazo (2005) investigated hot and cool EF in normally developing 3- to 5-year-old children, using two cool tasks, self-ordered pointing and DCCS and two hot tasks, delay of gratification and children's gambling task (CGT). The CGT (Kerr & Zelazo, 2004) is a simplified 2-deck version of the IOWA gambling Task (Bechara et al., 1994). The aim is to win as many rewards as possible. The advantageous deck offers low gains and minimal losses (net gain over trials). The disadvantageous deck offers high gains but higher losses (net loss over trials). Age-related improvements were observed in all tasks, but the patterns of age-related change did not distinguish hot from cool tasks. For three tasks, mastery was inferred from children's performance in relation to chance, assessed at the individual or group level. The majority of 4- and 5-year-olds demonstrated mastery on the DCCS. Pass rates were not reported for self-ordered pointing. For delay of gratification, group-based analyses revealed above chance performance by 4- and 5-year-olds, but not 3-year-olds. For the CGT, 4- and 5-year-olds performed at chance level, and 3-year olds performed below chance. Thus the pattern of mastery was not consistent for the hot tasks. The cool tasks were positively correlated, whereas the hot tasks were negatively correlated when age was partialled. The latter finding is difficult to interpret because of the poor CGT performance of the older children, relative to other studies (Kerr & Zelazo, 2004; Bunch, Andrews, & Halford, 2007) in which 5-year-olds performed above chance level. The DCCS and CGT are comparable in complexity. They require integration under a higher order rule in CCC-R theory, (Frye et al., 1995; Kerr & Zelazo, 2004) and are ternary-relational in RC theory (Bunch, et al., 2007; Halford, Bunch & McCredden, 2007). That 5-year-olds demonstrated mastery on the DCCS but not the CGT might be interpreted as evidence for a hot-cool dissociation whereby children master cool tasks earlier than they master hot tasks of similar complexity. However, this interpretation does not hold when the other hot task is considered. Delay of gratification is also

ternary-relational according to RC theory. Given that both hot tasks are comparable in complexity, mastery should be observed at about the same age, but this was not the case. The cool tasks were correlated with general intelligence and with two aspects of temperament, but the hot tasks were not. These differential associations with intelligence and temperament provided the only support for the hot-cool distinction in Hongwanishkul et al.'s study.

Prencipe, Kesek, Cohen, Lamm, and Zelazo (2011) investigated the hot-cool distinction in 8- to 15-year-olds using two cool tasks (Stroop colour word, backward digit span) and two hot tasks, (IOWA gambling, delay discounting). Age-related improvements were observed in all tasks. Consistent with a hot-cool dissociation, the improvements occurred from an earlier age in the cool tasks. Mastery was reported only for the gambling task in which only the 15-year-olds performed above chance on the final trial block. The inter-correlations between tasks did not distinguish the hot and cool tasks and an exploratory factor analysis yielded a single factor solution.

Other studies have examined the hot-cool distinction by manipulating the motivational significance of the stimuli within a single task. Carlson, Davis and Leach (2005) found that motivationally significant stimuli disrupted 3-year-olds' performance on the less-is-more task, in which children must point to the smaller quantity in order to get the larger quantity of rewards. The 3-year-olds performed at chance when the rewards were real candies and above chance when reward quantities were represented symbolically. The findings were interpreted as showing symbolic representation reduced the "hotness" of the rewards and this facilitated task performance.

Prencipe and Zelazo (2005) varied motivational significance in a delay of gratification task. Children in the self-gratification condition made decisions for themselves; those in the other-gratification condition made decisions for the researcher. Involvement of hot EF was assumed to be greater in the self-gratification condition. Three-year-olds in the other-gratification condition chose the delayed option at above chance level whereas those in the self-gratification condition did so at below chance level. These findings are consistent with a hot-cool dissociation at 3 years, however this interpretation would be stronger if the same children had served in both conditions. The hot-

cool was not evident among 4-year-olds who performed at chance level in both conditions.

The current research examined the hot-cool distinction in 3- to 6-year-old children in a more comprehensive manner. All children completed three cool and four hot tasks, and complexity was taken into account. The three cool tasks were DCCS, transitive inference, and class inclusion. The four hot tasks were ToM, CGT, reversal learning-conditional discrimination and delay of gratification. For all tasks except the last two mentioned, items at binary- and ternary-relational levels of complexity were developed in previous studies, and cognitive complexity was shown to impact significantly on young children's performance (Andrews & Halford, 2002; Andrews et al., 2003; Bunch et al., 2007; Frye et al., 1995; Halford et al., 2007). RC analyses showed that reversal learning (henceforth RL) is binary-relational, and more complex ternary-relational conditional discrimination (henceforth CD) problems were developed. Standard delay of gratification (henceforth DoG) items are ternary-relational according to RC analysis, and less complex binary-relational items were included. Thus all cool and hot tasks included items at two levels of complexity (binary-and ternary-relational) which involved very similar stimuli and procedures. Thus if children succeed on binary-relational items but perform poorly on ternary-relational items, their difficulty on the latter is likely due to complexity rather than to other aspects of the tasks.

Based on CCC-R and RC theories and previous findings, ternary-relational items were expected to be more difficult than binary-relational items in each of five tasks (DCCS, transitive inference, class inclusion, CGT, ToM). Stronger age effects were expected on ternary-relational than binary-relational items. By 5 years, at least 50% of children should demonstrate mastery on ternary-relational items. The majority of children at each age should succeed on binary-relational items.

Complexity effects in RL/CD and DoG tasks were also examined. As noted, in RL participants learn to discriminate between two objects. Selection of one object is rewarded; selection of the other object is not. Once the discrimination is learned, the contingencies are reversed so that the previously unrewarded object is rewarded. Initial learning and reversal phases are presented in succession in RL problems, whereas in CD problems they are presented simultaneously. The

choice between two objects is conditional on a third variable. For example, triangle might be rewarded (square not rewarded) when the background is white, but when the background is striped, square is rewarded and triangle is not. CD is ternary-relational because there are three interacting variables (background, object, outcome). RL is binary-relational because there are two variables (object, outcome). Consistent with this analysis, Rudy, Keith and Georgen (1993) found that children under 4.5 years succeeded on initial learning and reversal learning phases of their three-phase procedure, but were unable to reach criterion in the final CD phase. CD was always attempted last therefore fatigue might explain children's difficulty in that study.

In DoG, children choose between a smaller immediate reward and a larger delayed reward. Our RC analysis shows that DoG involves three variables (magnitude a, magnitude b, delay) and is ternary-relational. Simpler items that involve the choice between two immediate rewards (magnitude a, magnitude b) are binary-relational because delay is held constant. Consistent with this analysis, Thompson, Barresi, and Moore (1997) found that children chose the larger quantity more often on no delay than delay trials. It is also consistent with Lemmon and Moore's (2001, cited in Lemmon & Moore, 2007), finding that 3- and 4-year-olds succeeded on simple future choices (e.g., one or two stickers at the end of the game) where delay was constant. Existing studies provide only indirect evidence for complexity effects in DoG and RL/CD. RC theory predicts similar complexity and age effects and comparable mastery levels for these and the other five tasks.

A hot-cool dissociation would be evidenced by different patterns of age-related change, different mastery levels on hot versus cool tasks, or if hot and cool tasks loaded on separate factors. The earlier emergence of ternary-relational processing in hot than cool tasks would be consistent with the earlier maturation of medial than lateral PFC (Gogtay et al., 2004). However, others have proposed that hot tasks might be more difficult than cool tasks of the same complexity because overcoming affective pull in hot tasks requires additional reflection on the rule structure (Zelazo & Cunningham, 2007; Zelazo & Müller, 2011; Zelazo, Qu, & Kesek, 2010). If so, the opposite pattern would be expected.

Method

Participants

One hundred and twenty-six children from two day-care centers and one primary school in the same suburb of Gold Coast, Australia participated. The day-care centre directors and the school principal gave permission for the research to be conducted on their premises. To recruit the child participants, teachers distributed copies of the information sheet outlining the purpose and nature of the research and the parental consent form. These were distributed directly to parents of children in their classes, or to the children who took them home to their parents. The Index of Community Socio-Educational Advantage for the primary school was 988. While this value is close to the Australian average of 1000, inspection of the distribution across quartiles indicates that the population in this suburb is best described as lower to middle class (www.myschool.edu.au).

Six children did not complete all tasks. Three children with pre-diagnosed disorders (a 3-year-old girl with an unspecified developmental delay, a 4-year-old boy with cerebral palsy, and a 5-year-old boy with attention deficit disorder) completed a subset of the tasks to ensure they received rewards, as other children did. It was decided prior to testing that the data of these children would not be included in the analyses. A further three children failed to complete all tasks due to refusal to continue (2) and relocation to a different area (1). Inspection of the responses of these children yielded no evidence of atypical performance on the tasks. The final sample consisted of fifteen boys and fifteen girls in each of four age groups ($N = 120$). The ages (months) were: 3-year-olds ($M = 42.60$, $SD = 3.56$, range = 36-47), 4-year-olds ($M = 52.87$, $SD = 2.50$, range = 49-57), 5-year-olds ($M = 65.90$, $SD = 3.57$, range = 60-71), and 6-year-olds ($M = 77.10$, $SD = 2.76$, range = 72-83).

Children completed the sub-tests of the NEPSY executive function and attention battery (Korkman, Kirk, & Kemp., 1998) appropriate to their ages. These were statue and visual attention (3- and 4-year-olds), and visual attention, tower, and auditory attention (5- and 6-year-olds). Scaled subtest scores were combined to produce executive function and attention domain scores which were converted to their percentile equivalents. The mean percentiles (standard deviations) for the

3-, 4-, 5-, and 6-year-olds were 72.86 (15.69), 70.10 (17.44), 75.00 (20.03), and 83.87 (15.62) respectively. An analysis of variance (ANOVA) revealed a significant effect of age, $F(3, 116) = 3.56, p = .017$. Scheffe tests showed that the 6-year-olds' mean percentile was significantly higher than the 4-year-olds ($p = .027$) and no other difference approached significance. The results of the analyses reported below did not change substantively when NEPSY percentiles were controlled statistically.

General procedure

Recruitment and testing procedures received institutional ethical approval. Parents gave written consent. Children gave verbal consent. Children were tested individually by a female researcher in a quiet area of their child care center or school. Testing durations ranged from approximately 1 hour 30 minutes (older children) to 2 hours 45 minutes (younger children). Testing was spread over two to four sessions depending on the school schedule and children's willingness to continue. The interval between the first and last testing session ranged from 1 day to 10 days ($M = 6$). Seven or eight children in each age group received the seven tasks in each four different presentation orders. In transitivity and class inclusion tasks, binary-relational items always preceded ternary-relational items. In ToM, RL-CD, CGT and DCCS tasks, binary- and ternary-relational items were presented in different sessions in counterbalanced order. For DoG, the binary- and ternary-relational items were intermixed and presented in quasi-random order in the same session. As noted further below, practice or demonstration trials were included in most relational processing tasks. On the rare occasions when children responded incorrectly or the researcher suspected that the child had not fully understood the requirements, these trials were repeated. No children responded incorrectly when practice trials were repeated.

Cool Tasks

Transitivity.

Transitive inference involves deducing an inference $A R C$ from premises $A R B$ and $B R C$, where R is a transitive relation and A, B, and C are the elements being related. Determining the

relation between A and C requires that the premises be integrated to form an ordered triple, $A R B R C$, from which the inference $A R C$ can be made.

The task was identical to the 5-term task used by Andrews and Halford (2002). On each trial children were shown a different premise display consisting of four pairs of colored squares in which one color was higher than the other. For example, the premises blue above purple, red above blue, yellow above green, green above red, together define the unique top-down order, yellow–green–red–blue–purple, more generally $A > B > C > D > E$, where A is top position and E is bottom. The task was to build towers using a separate set of colored squares. The importance of placing squares in the correct order was emphasized. In each of the two practice trials, children constructed a 5-square tower and received feedback on their responses. On completion of each tower, the researcher engaged the child in discussion about the relative positions of the colored squares in the tower and their correspondence to the premises. Care was taken to ensure that they realized that ‘higher than’ referred to squares in adjacent (e.g., B and C) and nonadjacent positions (e.g., B and D), and that a square could be inserted above, between, or below those already in place. Throughout the procedure, children were reminded to consult the premises (clues) and that they might need to look at more than one clue to determine the correct order.

In the binary-relational items, children constructed two 5-square towers, beginning with an internal pair, either BC or CD. They were given the two colors which they attempted to order. In the example, BC corresponds to green–red. The premise, green above red was relevant. Adding each subsequent square required consideration of a single premise. To place blue, the relevant premise is red above blue. Placing blue below red, yielded the order green–red–blue. This is a concatenation strategy, which entails processing one binary relation at a time. One point was awarded for each correctly ordered initial pair and subsequent square (max = 8). In the ternary-relational items, two squares, B and D were placed first (i.e., B above D). Using the previous example, the two relevant premises (green above red; red above blue) must be integrated to form the triple, (green–red–blue), from which green–blue can be deduced. As a check on guessing, square

C (red) was placed after B and D. If children integrated BC and CD to place B above D, the correct position of C between B and D would be apparent. Eight items were administered. Credit was given for correct placement of B, D, and C (max = 8). Standardized alpha = .74 (Andrews & Halford, 2002).

Class Inclusion.

As in Andrews and Halford (2002), there were six stimulus displays each consisting of colored geometric shapes that comprised an inclusion hierarchy. One display, which will be used to illustrate the procedures contained three yellow squares and two blue squares. In this display, the superordinate class is squares, the major subclass is yellow things, the minor subclass is blue things. The researcher first identified the colors and shapes by pointing to the appropriate elements while labeling the classes. For each display there were two preliminary questions to ensure that children could identify the relevant color-shape combinations, *Are all the yellow things square?* Three test questions followed. Subclass comparison questions, *Are there more yellow things or more blue things?* are binary-relational because two entities are compared. Binary-relational scores (max = 6) reflect correct responses to subclass comparison questions. Superordinate-major subclass comparison questions, *Are there more squares or more yellow things?* are ternary-relational because recognition that the superordinate class includes more elements than the major subclass requires all three classes to be considered. Superordinate-minor subclass comparison questions, *Are there more squares or more blue things?* were included because errors on these questions provide an estimate of guessing (see Hodkin, 1987). Ternary-relational scores (max = 6) reflect correct responses to superordinate-major subclass questions minus errors on superordinate-minor subclass comparison questions. Standardized alpha = .81 (Andrews & Halford, 2002).

Dimensional Change Card Sort Task.

The materials and procedures for the ternary-relational items were identical to Andrews et al, (Experiment 2, 2003). Target cards differed in both shape and color (e.g., red triangle, green circle). On each of four trials, the researcher first demonstrated how to sort two test cards (e.g., red

circle, green triangle) according to one dimension (either color or shape). The target cards were then rearranged and children attempted to sort the same two test cards so that they matched the target cards on the dimension that had been demonstrated (either color or shape). The DCCS involves two levels of a hierarchical rule structure in CCC-R theory. In RC theory the task is difficult to decompose into subtasks and is ternary-relational. Halford et al. (2007) designed less complex binary-relational items. The target cards were as described for the ternary-relational items. The test cards differed because the currently irrelevant dimension was held constant. On color trials, the objects on the test cards were the same shape (e.g., red circle, green circle). On shape trials, the objects were the same color (e.g., red triangle, red circle). Children received four color and four shape trials (max = 8) at each level of complexity. Two presentation orders (color-shape-shape-color and shape-color-color-shape) were used with equal frequency in each age group.

Hot Tasks

Children's Gambling Task.

The ternary- and binary-relational versions (binary-gain and binary-loss) were the same as Kerr and Zelazo (2004) and Bunch et al. (2007) respectively, with the exception that four (not five) trial blocks were presented. Rewards were M&M sweets or stickers depending on parents' preferences. There was one set of laminated cards for each version. Each set contained two decks of 40 cards. The front sides of all cards displayed happy and sad faces, corresponding to the number of rewards won and lost. In the ternary-relational version, cards in the advantageous deck yielded a gain of one and a loss of zero or one reward. Cards in the disadvantageous deck yielded a gain of two and a loss of zero, four, five or six rewards. In the binary-relational versions, values on one variable (either losses or gains) were held constant. In the binary-relational (gain) version the values were: gain of two, loss of zero or one (advantageous deck) and gain of one, loss of zero or one (disadvantageous deck). In the binary-relational (loss) version they were: gain of one, loss of zero or one cards (advantageous deck) and gain of one, loss of zero or five rewards (disadvantageous deck). Children received a reward to motivate them to play, then an initial stake of 10 rewards

(M&Ms or stickers). The three versions were presented using comparable procedures and instructions. Prior to the test trials, there were six demonstration trials. Children watched as the researcher selected a card from one deck and then received and forfeited the rewards as indicated on the card. On demonstration trials, the researcher selected three cards from each deck. The dependent measure was number of choices from the advantageous deck (max = 40).

Delay of Gratification.

Each child received 24 items: four binary- and four ternary-relational items of each of three trial type (shared, self, other gratification). For each child, there were two sticker books (one each for the child and researcher), approximately 30 stickers, and two envelopes (one each for the child and researcher). The researcher explained that she and the child would each receive stickers to place in their books. Children were told that they would be asked to make choices for themselves, the researcher, or both themselves and the researcher. They were told that they could choose how many stickers each would get and whether they received the stickers immediately or later. If the delayed options were chosen, stickers would be placed in the envelopes and saved until the end of the game, when they and the researcher would each take their stickers home. The options always consisted of between one and four stickers. The researcher responded in a uniform, mildly positive manner to children's responses. Responses were scored as correct if the larger quantity was chosen (max = 12 at each level). This was the delayed option on ternary-relational items.

Reversal Learning and Conditional Discrimination (RL/CD).

The RL (binary-relational) and CD (ternary-relational) problems were closely matched in terms of stimuli, task instructions and procedures. The problems were presented on a Dell latitude laptop computer (30 × 25 cm screen). Each stimulus slide contained two white objects with black outlines (4 cm in length or diameter) presented side by side. The left-right position of the objects was counterbalanced across slides. Feedback in the form of a happy face (correct response) and sad face (incorrect response) was provided after each selection. On each CD trial, the two objects were presented on a colored background (e.g., red) for 50% of the trials and different colored background

(e.g., blue) for the remaining trials. Children were told that they should take notice of the background color and the shape of the objects and try to choose the object that made a happy face appear. They pressed the right or left shift keys to select the objects on the right or left side of the screen, respectively. There were four 12-trial blocks. The RL items were presented in a similar manner; however the objects were presented on a constant background color (gray). The learning and reversal phases each consisted of two, 12-trial blocks. The RL and CD test trials were preceded by two information trials with feedback provided after each. The dependent variable for RL and CD was number of correct responses (max = 48).

Theory of Mind.

The tasks, materials and procedures are described in detail in Andrews et al. (Experiment 3, 2003). Eight binary-relational items were presented. Four of these required children to indicate whether objects in the environment were perceptible to themselves and to others. For example, was a drum placed on the child's side of a barrier visible to the child? (yes) Was it visible to the researcher? (no). Four items examined children's understanding of the transformations used in false belief and appearance reality tasks. For example, a yellow plastic filter was held up for the children to look through, and they were asked what color things appear through the plastic. This was then repeated using a purple filter. Each binary-relational item was scored as correct if the component questions were both correct (max = 8). Ternary-relational items were three appearance-reality (fish-filter, skewer in glass, crayon box) and three false-belief (unexpected contents, perceived color, chocolate hiding) tasks that are widely used to assess theory of mind in young children (e.g., Frye, et al. 1995; Wellman, Cross & Watson, 2001). Each included a reality question and an appearance/belief question. Ternary-relational items were scored as correct if responses to both component questions were correct (max = 6). Previous research in our lab indicates acceptable reliability levels for the binary-relational (Cronbach's $\alpha = .69$) and the ternary-relational tasks (Cronbach's $\alpha = .75$) (Andrews et al.).

Results

Preliminary analyses revealed no significant effects of presentation order, so this factor was not included in further analyses. There were no significant differences between the binary-relational (gain) and binary-relational (loss) versions of the CGT, therefore scores on these versions were averaged. All 5- and 6-year-olds were error-free on the binary-relational (but not ternary-relational) ToM items, therefore non-parametric analyses were conducted for this task.

Descriptive statistics for each age group for binary- and ternary-relational items for the seven tasks are shown in Table 1. ANOVAs with complexity as a within-subject variable and age as a between-subject variable were conducted on six tasks (transitivity, class inclusion, DCCS, CGT, DoG, RL/CD). The results are summarized in Table 2. Complexity had a significant and strong effect on all tasks. Binary-relational items were significantly easier than ternary-relational items. Performance on all tasks improved with age. The Complexity \times Age group interactions reflected stronger complexity effects in the younger than older age groups, and stronger age effects for ternary-relational than binary-relational items.

Additional factors were analyzed in some tasks. In the CGT, there was a significant Trial Block \times Complexity \times Age group interaction, $F(9, 336) = 5.17, p < .001, \eta^2 = .122$. On the ternary-relational CGT, selections from the advantageous deck decreased across blocks for 3-year-olds, increased across blocks for 5- and 6-year-olds, and did not change across blocks for 4-year-olds. On the binary-relational CGT, selections from the advantageous deck increased across blocks for all age groups. In the DoG task, there was a significant Complexity \times Item type interaction, $F(2, 224) = 9.58, p = .000, \eta^2 = .079$. The complexity effect was significant for all items types, but it was stronger for self gratification ($\eta^2 = .534$) and for shared gratification ($\eta^2 = .515$) than for other gratification ($\eta^2 = .297$). The effect was due to binary-relational self- and shared-gratification items, which were easier than the other-gratification items.

For TOM, a Wilcoxon signed-rank test ($N = 120$) showed a significant effect of complexity, $Z = -7.49, p = .000$. Accuracy was higher for the binary- than ternary-relational items. Complexity effects were significant for 3-year-olds, $Z = -4.79, p = .000$, 4-year-olds, $Z = -4.55, p = .000$, and 5-

year-olds, $Z = -3.24$, $p = .000$, but not for 6-year-olds, $Z = -1.0084$, $p = .125$. Kruskal-Wallis tests revealed significant improvements in accuracy as age increased on binary-relational, $\chi^2(3, N = 120) = 65.76$, $p = .000$ and ternary-relational items, $\chi^2(3, N = 120) = 83.82$, $p = .000$.

For the analyses based on individuals, mastery criteria were set at the level of performance that was significantly above chance ($p = .05$) according to the binomial distribution. Table 3 shows that the majority of children in all age-groups performed above chance on binary-relational items of all tasks, with one exception. For class inclusion, 33.33% of 3-year-olds performed above chance. When the criterion is relaxed to allow one error (5 out of 6) the majority in all age groups performed above chance. A minority of 3- and 4-year-olds performed significantly above chance on ternary-relational items, whereas at least 50% of 5- and 6-year-olds did so.

The hot-cool distinction was examined in relation to patterns of age-related change, mastery levels, overall performance, and the factor structure. The patterns of age-related change on ternary-relational items in hot and cool tasks were examined using Scheffe tests. These analyses were restricted to children who demonstrated mastery on the simpler binary-relational items in each task. Means, standard deviations and sample sizes are reported next only for the tasks and age groups where some children did not demonstrate mastery on the binary-relational items. Where all children demonstrated mastery of the binary-relational items, the descriptive statistics are the same as in Table 1. For three hot tasks (DoG, CD, ToM), there was a consistent pattern of age-related change in that ternary-relational processing improved significantly between 3 years (DoG, $M = 18.40$, $SD = 15.54$, $n = 24$) and 4 years (DoG, $M = 39.58$, $SD = 27.74$, $n = 28$; ToM, $M = 45.40$, $SD = 29.17$, $n = 29$), (largest $p = .005$), between 4 and 5 years (DoG, $M = 73.21$, $SE = 20.45$, $n = 28$) (largest $p = .001$), but not between 5 and 6 years (smallest $p = .316$). For CGT, selections from the advantageous deck did not change significantly between 3 ($M = 37.59$, $SD = 25.52$, $n = 28$) and 4 years ($M = 50.26$, $SD = 28.86$, $n = 29$) ($p = .266$). When only the final trial block was examined, the pattern is more similar to the other hot tasks. The increase between 3 ($M = 34.64$, $SD = 38.82$, $n = 28$) and 4 years ($M = 54.48$, $SD = 40.93$, $n = 29$) was significant by one-tailed test ($p = .048$).

There was a further significant increase from 4 to 5 years ($p < .001$), but not between 5 and 6 years ($p = .777$).

For the three cool tasks (transitivity, class inclusion, DCCS), ternary-relational processing did not improve significantly from 3 (transitivity: $M = 2.34$, $SD = 9.38$, $n = 16$; class inclusion, $M = 12.96$, $SD = 46.29$, $n = 18$; DCCS: $M = 67.24$, $SD = 16.17$, $n = 29$) to 4 years (transitivity: $M = 1.79$, $SD = 5.60$, $n = 28$; class inclusion, $M = 0.00$, $SD = 28.18$, $n = 29$) (smallest $p = .569$). Significant improvements were observed on all three tasks from 4 to 5 years (largest $p = .001$). There was no significant improvement between 5 and 6 years on class inclusion ($p = .412$) or DCCS ($p = .793$), but 6-year-olds performed significantly better than 5-year-olds ($p = .001$) on transitive inference. Thus the hot and cool tasks are best distinguished by significant improvements between 3 and 4 years on the hot tasks and their absence on the cool tasks.

The difference between the hot and cool tasks in the percentages of children who demonstrated on the ternary-relational items was examined using Mann-Whitney tests. The unit of analysis was task (7 cases) and separate tests were conducted for each age group. Significantly more 4-year-olds performed above chance level on hot (mean rank = 5.38) than cool tasks (mean rank = 2.17), $Z = -1.98$, $p = .048$. The corresponding comparisons for 3-, 5-, and 6-year-olds were not significant, smallest $p = .154$. The proportions of hot and cool tasks on which individual children performed above chance on the ternary-relational items were compared. The 4-year-olds passed a higher proportion of hot than cool tasks (.33 vs .07), $F(1, 29) = 21.90$, $p < .001$, $\eta^2 = .43$ as did the 5-year-olds (.88 vs .73), $F(1, 29) = 11.52$, $p = .002$, $\eta^2 = .284$. These proportions did not differ for the 3-year-olds and 6-year-olds. The analyses of mastery levels showed that the distinction between ternary-relational processing in hot and cool tasks was evident at 4 and 5 years of age, but it was more robust at 4 years.

A further test of the hot-cool distinction was based on total scores (binary + ternary) for each task, which were converted to z -scores. Composite scores were computed by averaging the relevant z -scores. Hot Z_{comp} (standardized $\alpha = .883$) and cool Z_{comp} (standardized $\alpha = .877$) were

subjected to a mixed 2(hot-cool) \times 4(age group) ANOVA which yielded a significant main effect of age, $F(3, 116) = 243.89, p < .001, \eta^2 = .863$ and a significant Hot-cool \times Age interaction, $F(3, 116) = 3.71, p = .014, \eta^2 = .087$, which is shown in Figure 1. For the 4-year-olds, mean hot Zcomp was significantly higher than mean cool Zcomp, $F(1, 29) = 5.98, p = .021, \eta^2 = .171$. For the 6-year-olds, mean cool Zcomp was significantly higher than mean hot Zcomp, $F(1, 29) = 6.02, p = .020, \eta^2 = .172$. For the 3- and 5-year-olds, there were no significant hot-cool differences. This analysis is consistent with the preceding analyses in that it provides evidence for a hot-cool distinction at 4-years of age, when children performed better on the hot than cool tasks. It suggests further that this advantage was reversed by 6 years, when children performed better on cool than hot tasks.

Finally, exploratory factor analyses using the maximum likelihood extraction method were conducted to determine whether separate factors would distinguish the hot and cool tasks. The first analysis was based on the full sample ($N = 120$) and the variables were total (binary + ternary) scores on the seven tasks. The analysis yielded a single factor solution (eigenvalue = 4.73) which accounted for 68.07% of the variance. The factor loadings were .71 (CGT), .78 (DoG), .84 (RL/CD) and .91 (ToM) for the hot tasks and .77 (DCCS), .86 (transitivity) and .87 (class inclusion) for the cool tasks. Two further analyses were conducted based on the ternary-relational scores only and a reduced sample that excluded the 6-year-olds. These analyses also yielded single factor solutions and factor loadings which were comparable to those reported above. The factor analyses did not support the hot-cool distinction.

Discussion

Relational processing was investigated in normally-developing 3- to 6-year-old children, using three cool and four hot tasks and a within-subjects design. The first hypothesis referred to complexity and age effects. Complexity was a major source of difficulty on all hot and cool tasks. Ternary-relational items were more difficult than the corresponding binary-relational items. The complexity effects are unlikely to be due to specific knowledge or task procedures because the binary- and ternary-relational items were closely matched in these respects, and children performed

well on binary-relational items. Complexity effects have previously been reported in many cool tasks (e.g., Andrews & Halford, 1998; 2002, Frye et al., 1995 Zelazo et al., 2003), and in two hot tasks, CGT (Bunch et al., 2007) and ToM (Andrews et al., 2003). The complexity effect for RL/CD is consistent with findings from Rudy et al.'s (1993) 3-phase procedure in which RL was assessed in phases 1 and 2 and CD in phase 3. In the current study, RL and CD were assessed independently. The complexity effect for DoG is consistent with Thompson et al. (1997), although they did not interpret their findings in terms of complexity theories. Complexity had a greater effect on younger than older children. By 5 years of age, the majority of children had mastered ternary-relational items, whereas binary-relational items were generally mastered by 3 years of age. These median ages are consistent with predicted norms (Halford, 1993) and previous research (Andrews et al., 2003; Andrews & Halford, 2002; Bunch et al., 2007; Halford et al., 2002a, 2002b; Rudy et al., 1993). Thus complexity is as important in hot tasks as it is in cool tasks.

The complexity manipulations were derived from RC theory (Halford et al. 1998) and the findings support that theory. The complexity effects on ToM, DCCS, and CGT are also interpretable in terms of CCC-R theory (Zelazo et al., 2003). CCC-R theory has not been applied to transitive inference, class inclusion, DoG, or RL/CD, however the hierarchical structure of class inclusion suggests that this task would be amenable to analysis using CCC-R theory. The findings broaden the scope of complexity theories to encompass a wider range of cognitive developmental phenomena.

Complexity analyses in RC theory are based on information that constrains the current decision, rather than information being stored for future processing. Therefore processing load in RC theory differs from conceptions of working memory load that also include storage demands (e.g., Daneman & Carpenter, 1980). The relational processing tasks used here had minimal storage loads and comparable stimuli and procedures were used at different complexity levels. Therefore the complexity effects seem unlikely to reflect differential storage load.

Several analyses addressed the hot-cool distinction. In children who mastered the binary-

relational items, different patterns of age-related change were observed on ternary relational items in hot versus cool tasks. The main difference was that significant improvement was observed between 3 and 4 years of age on hot tasks but not on cool tasks. Perhaps because of the earlier onset of improvement, the 4-year-olds (and to a lesser extent 5-year-olds) demonstrated higher rates of mastery of ternary-relational processing in hot than cool tasks. The composite scores reflecting overall performance also showed that 4-year-olds performed better on hot than cool tasks. That the hot-cool difference was most apparent in 4-year-olds is consistent with the previous findings that ternary-relational processing begins to emerge between 4 and 5 years of age (Andrews & Halford, 2002). The advantage for hot tasks at 4 years and the reversal to an advantage for cool tasks by 6 years might reflect the different maturation rates of medial versus lateral regions of PFC. Earlier maturation of the medial regions might allow younger children to succeed on hot tasks that rely more on these regions. Older children are able to recruit the later maturing lateral regions and this allows them to succeed on the cool tasks. The 6-year-olds' better performance on cool than hot tasks could be considered consistent with Prencipe et al.'s (2011) recent finding that performance on cool tasks developed earlier in 8- to 15-year-olds than performance on hot tasks did. However, this comparison must remain tentative because of the different tasks and age ranges in the studies.

The exploratory factor analysis provided no support for the hot-cool distinction. This parallels Prencipe et al.'s (2011) research in which a single factor accounted for variance in performance on hot and cool tasks. Such findings might reflect the age-sensitive nature of the hot-cool dissociation which was demonstrated in the current study. Further research with larger samples at each age would be required to detect whether the factor structure changes as a function of age.

The hot-cool classification of tasks was based on task characteristics, prior use by other researchers and evidence for the involvement of medial versus lateral PFC. In relation to the first criterion, other features of the hot tasks might have facilitated the earlier emergence of ternary-relational processing. One such feature is the opportunity to use feedback from earlier trials to improve performance on subsequent trials. This was possible in two hot tasks, CGT and CD.

However, neither ToM nor DoG involved feedback, so this factor cannot fully explain the hot-cool difference. Nor can it explain the robust performance on the binary-relational items. One way to control for differences between tasks is to manipulate the stimulus content within the same task. The current study incorporated this approach. DoG included three item types (self-, other-, shared-gratification) which might differentially recruit hot and cool systems (Qu & Zelazo, 2007), but there was no evidence for hot-cool distinction when the items types were contrasted. In relation to the third criterion, we relied on studies of brain function in children. Where no such studies exist, we drew on adult studies. Inference from studies of adult brain function assumes comparable functional specialization and localization in adult and child brains, an assumption that might not be warranted (Johnson, 2001).

Of the published research examining hot-cool distinction, Hongwanishkul et al.'s (2005) study of 3- to 5-year-olds is most relevant to the current work. They included the CGT, DoG, DCCS tasks, but without the binary-relational items. Five-year-olds performed above chance on the CGT in the current study but not in Hongwanishkul et al.'s. In the current study, CGT and DoG were positively correlated before ($r = .68, p = .000$) and after ($r = .31, p = .001$) age was partialled. The negative age-partialled correlation in Hongwanishkul et al.'s sample might be a spurious finding stemming from the poor CGT performance of their 5-year-olds.

Our findings for individual tasks are broadly consistent with predictions however one discrepancy deserves comment. Whereas more than one-third of 4-year-olds demonstrated ternary-relational processing in CGT, CD, and ToM, just 16.67% did so on DoG. Thus while DoG was similar to other hot tasks in the pattern of age-related change, it was more similar to the cool tasks in terms of mastery levels. One possible explanation is that self- and other-gratification DoG items elicited hot and cool systems to varying degrees (Prencipe & Zelazo, 2005). However, analyses in which the item types were scored separately did not support this. Another possibility is that DoG involves both hot and cool systems and recruits both medial and lateral prefrontal regions, as was found for delay discounting in adults (McClure et al., 2004).

The hot-cool differences evident in the current research are consistent with the view that the hot tasks rely more heavily on the earlier maturing medial regions, while the cool tasks rely more heavily on the later maturing lateral PFC regions. Whereas ternary-relational processing in cool tasks must await maturation of the lateral PFC, the earlier maturing medial regions might help to kick start ternary-relational processing in hot tasks and promote the formation of connections between the medial and lateral regions.

One practical implication of the findings is that by including both hot and cool tasks in their assessment batteries and by taking complexity into account, clinicians might gain a more nuanced picture of young children's level of cognitive development. There is scope for further research integrating brain function and theory-based manipulations of complexity. The current findings suggest that studies investigating the brain systems underlying the hot-cool distinction should ensure that task complexity is taken into account. Similarly, investigation of the brain regions recruited as complexity increases should avoid confounding complexity with hot-cool status of tasks, especially if 4-year-olds are included. The current findings for normally developing children open the way for further research investigating the hot-cool distinction in atypically developing children, such as those with ADHD, who perform poorly on executive function tasks (Yang et al., 2011). The tasks described here take account of complexity so they would allow a more rigorous investigation of the hot-cool distinction and its potential links to atypical development than has been possible to date.

References

- Andrews, G., Birney, D. P., & Halford, G. S. (2006). Relational processing and working memory capacity in comprehension of relative clause sentences, *Memory and Cognition*, *34*, 1325–1340.
- 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. (2011). Recent advances in relational complexity theory & its application to cognitive development. In P. Barrouillet & V. Gaillard (Eds). In *Cognitive Development and Working Memory: A dialogue between Neo-Piagetian and Cognitive Approaches*. (pp. 47-68). Hove, England: Psychology Press.
- Andrews, G., & Todd, J. M. (2008). Two sources of age-related decline in comprehension of complex relative clause sentences. In N. B. Johansen (Ed.) *New Research on Short term Memory*. (pp. 93 - 123). NovaScience Publishers.
- Andrews, G., Halford, G. S., Bunch, K. M., Bowden, D., & Jones, T. (2003). Theory of mind and relational complexity. *Child Development*, *74*, 1476-1499.
- Andrews, G., Halford, G. S., Murphy, K., & Knox, K. (2009). Integration of weight and distance information in young children: The role of relational complexity. *Cognitive Development*, *24*, 49-60.
- Atance, C. M., & O'Neill, D. K. (2001). Episodic future thinking. *Trends in Cognitive Sciences*, *5*, 533-539.
- Bechara, A., Damasio, A. R., Damasio, H., & Anderson, S. W. (1994). Insensitivity to future consequences following damage to human prefrontal cortex. *Cognition*, *50*, 7-15.
- Bechara, A., Damasio, H., & Tranel, D., & Anderson, S. W. (1998). Dissociation of working memory from decision-making within the human prefrontal cortex. *The Journal of Neuroscience*, *18*, 428-437.

- 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. (2006). Measuring the influence of complexity on relational reasoning: The development of the Latin Square Task. *Educational and Psychological Measurement, 66*, 146-171.
- Bunch, K. M., Andrews, G., & Halford, G. S. (2007). Complexity effects on the children's gambling task. *Cognitive Development, 22*, 376-383.
- Bunge, S. A., & Zelazo, P. D. (2006). A brain-based account of the development of rule use in childhood. *Current Directions in Psychological Science, 15*, 118-121.
- Carlson, S. M., Davis, A. C., & Leach, J. G. (2005). Less is more: Executive function and symbolic representation in preschool children. *Psychological Science, 16*, 609-616.
- Christoff, K., & Owen, A. M. (2006). Improving reverse neuroimaging inference: Cognitive domain versus cognitive complexity. *Trends in Cognitive Sciences, 10*, 352-353.
- 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.
- Crone, E. A. Donohue, S. E., Honomichl, R., Wendelken, C., & Bunge S. A. (2006). Brain regions mediating flexible rule use during development. *Journal of Neuroscience, 26*, 11239-11247.
- Daneman, M., & Carpenter, P.A. (1980) Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior, 19*. 450-466.
- Elliot, R., Dolan, R. J., & Frith, C. D. (2000). Dissociable functions in the medial and lateral orbitofrontal cortex: Evidence from human neuroimaging studies. *Cerebral Cortex, 10*, 308-317.
- Frye, D., Zelazo, P. D., & Palfai, T. (1995). Theory of mind and rule-based reasoning. *Cognitive Development, 10*, 483-527.
- Goel, V. (2007). Anatomy of deductive reasoning. *Trends in Cognitive Sciences, 11*, 435-441.

- Gogtay, N., Giedd, J. N., Lusk, L., Hayashi, K. M., Greenstein, P., Vaituzis, A. C., ... Thompson, P. M. (2004). Dynamic mapping of human cortical development through early adulthood. *Proceedings of the National Academy of Sciences of the United States of America*, *101*, 8174-8179.
- Halford, G. S. (1993). *Children's understanding: The development of mental models*. Hillsdale, NJ: Erlbaum.
- Halford, G.S., & Andrews, G. (2006). Reasoning and problem solving. In D. Kuhn and R. Siegler (Eds). *Handbook of Child Psychology: Volume 2, Cognitive, Language and Perceptual Development*, (pp. 557-608) 6th edition. Wiley.
- 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. (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., Bunch, K., & McCredden, J. E. (2007). Problem decomposability as a factor in complexity of the dimensional change card sort task. *Cognitive Development*, *22*, 384-391.
- Halford, G. S., Wilson, W. H., & Phillips, S. (1998). Processing capacity defined by relational complexity: Implications for comparative, developmental, and cognitive psychology. *Behavioural Brain Sciences*, *21*, 803-831.
- Hodkin, B. (1987). Performance model analysis in class inclusion: An illustration with two language conditions. *Developmental Psychology*, *23*, 683-689.
- Hongwanishkul, D., Happaney, K. R., Lee, W. S. C., & Zelazo, P. D. (2005). Assessment of hot and cool executive function in young children: Age-related changes and individual differences. *Developmental Neuropsychology*, *28*, 617-644.
- Johnson, M. H. (2001). Functional brain development in humans. *Nature Reviews Neuroscience*, *2*, 475-483.

- Johnson, M. H., & de Haan, M. (2011). *Developmental Cognitive Neuroscience. An Introduction* (3rd edition). Oxford, England: Wiley-Blackwell.
- Kerr, A., & Zelazo, P. D. (2004). Development of “hot” executive function: The children’s gambling task. *Brain and Cognition*, *55*, 148-157.
- Korkman, M., Kirk, U., & Kemp, S. L. (1998). *NEPSY. A developmental neuropsychological assessment*. San Antonio, TX: Harcourt Brace.
- Kroger, J. K., Sabb, F. W., Fales, C. L., Bookheimer, S. Y., Cohen, M. S., & Holyoak, K. J. (2002). Recruitment of anterior dorsolateral prefrontal cortex in human reasoning a parametric study of relational complexity. *Cerebral Cortex*, *12*, 477-485.
- Lemmon, K., & Moore, C. (2007). The development of prudence in the face of varying future rewards. *Developmental Science*, *10*, 502-511.
- Liu, D., Sabbagh, M. A., Gehring, W. J., Wellman, H. M. (2009). Neural correlates of children’s theory of mind. *Child Development*, *80*, 318-326.
- McClure, S. M., Laibson, D. I., Loewenstein, G., & Cohen, J. D. (2004). Separate neural systems value immediate and delayed monetary rewards. *Science*, *306*, 503-507.
- Metcalf, J., & Mischel, W. (1999). A hot/cool-system analysis of delay of gratification: Dynamics of willpower. *Psychological Research*, *106*, 3-19.
- Mischel, W., Shoda, Y., & Rodriguez, M. L. (1989). Delay of gratification in children. *Science*, *244*, 933-938.
- Moriguchi, Y., & Hiraki, K. (2009). Neural origin of cognitive shifting in young children. *Proceedings of the National Academy of Sciences of the United States of America*, *106* (14), 6017-6021.
- Overman, W. H., Bachevalier, J., Schuhmann, E., & Ryan, P. (1996). Cognitive gender differences in very young children parallel biologically based cognitive gender differences in monkeys. *Behavioural Neuroscience*, *110*, 673-684.
- Prencipe, A., Kesek, A., Cohen, J., Lamm, C., Lewis, M. D., & Zelazo, P. D. (2011). Development of hot and cool executive function during the transition to adolescence. *Journal of Experimental*

Child Psychology, 108, 621-637.

- Prencipe, A., & Zelazo, P. D. (2005). Development of affective decision making for self and other. *Psychological Science*, 16, 501-505.
- Qu, L., & Zelazo, P. D. (2007). The facilitative effect of positive stimuli on 3-year-olds flexible rule use. *Cognitive Development*, 22, 456-473.
- Rolls, E. T. (2004). The functions of the orbitofrontal cortex. *Brain and Cognition*, 55, 11-29.
- Rolls, E. T., Hornak, J., Wade, D., & McGrath, J. (1994). Emotion-related learning in patients with social and emotional changes associated with frontal lobe damage. *Journal of Neurology, Neurosurgery and Psychiatry*, 57, 1518-1524.
- Rudy, J. W., Keith, J. R., & Georgan, K. (1993). The effect of age on children's learning of problems that require a configural solution. *Developmental Psychobiology*, 26, 171-184.
- Shaw, P., Greenstein, D., Lerch, J., Clasen, L., Lenroot, R., Gogtay, N., ... Giedd, J. (2006). Intellectual ability and cortical development in children and adolescents. *Nature*, 440, 676-679.
- Thompson, C., Barresi, J., & Moore, C. (1997). The development of future-oriented prudence and altruism in preschoolers. *Cognitive Development*, 12, 199-212.
- Viskontas, I.V., Morrison, R.G., Holyoak, K.J., Hummel, J.E., & Knowlton, B.J. (2004). Relational integration, inhibition, and analogical reasoning in older adults. *Psychology and Aging*, 19, 581-591.
- Waltz, J. A., Knowlton, B. J., Holyoak, K. J., Boone, K. B., Mishkin, F. S., Santos, M. M., ... Miller, B. L. (1999). A system for relational reasoning in human prefrontal cortex. *Psychological Science*, 10, 119-125.
- Wellman, H.M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: The truth about false belief. *Child Development*, 72, 655-684.
- Yang, B.-R., Chan, R. C. K., Gracia, N., Cao, X.-Y., Zou, X.-B., Jing, J., ... Shum, D. (Published online June 1, 2011). Cool and hot executive functions in medication-naive attention deficit hyperactivity disorder. *Psychological Medicine*. doi:10.1017/S0033291711000869.
- Zelazo, P. D., & Cunningham, W. A. (2007). Executive function: Mechanisms underlying emotion regulation. . In J. J. Gross (Ed.), *Handbook of emotion regulation* (pp. 135-158). New York: The

Guilford Press.

- Zelazo, P. D., Jacques, S., Burrack, J. A., & Frye, D. (2002). The relation between theory of mind and rule use: Evidence from persons with autism-spectrum disorders. *Infant and Child Development, 11*, 171-195.
- Zelazo, P. D., & Müller, U. (2011). Executive function in typical and atypical development. In U. Goswami (Ed.), *The Wiley-Blackwell Handbook of Childhood Cognitive Development* (pp. 574-603) 2nd edition. Oxford: Blackwell.
- Zelazo, P. D., Müller, U., Frye, D., & Marcovitch, S. (2003). The development of executive function. *Monographs of the Society for Research in Child Development, 68*, 1-155.
- Zelazo, P. D., Qu, L., & Kesek, A. K. (2010). Hot executive function. Emotion and the development of cognitive control. In S. D. Calkins & M. A. Bell (Eds.), *Child development at the intersection of emotion and cognition* (pp. 97-111). Washington, DC: American Psychological Association.

Table 1

Mean Percentage Correct (SDs) for Binary- and Ternary-relational Items by Task and Age Group (N = 120)

		Age Group			
<i>Cool tasks</i>	Complexity	3-years	4-years	5-years	6-years
Transitivity	Binary	80.00* (14.53)	93.75* (9.14)	97.50* (5.09)	99.17* (3.17)
	Ternary	1.67 (7.14)	2.08 (5.76)	45.83* (24.42)	65.83* (20.22)
Class Inclusion	Binary	80.00* (18.26)	93.33* (9.39)	99.44* (3.04)	99.44* (3.04)
	Ternary	-18.33 (43.84)	0.00 (27.68)	66.11* (25.70)	79.44* (24.25)
Dimensional Change Card Sort	Binary	94.58* (7.10)	98.33* (4.32)	99.58* (2.28)	99.58* (2.28)
	Ternary	67.50* (15.94)	71.25* (17.72)	96.25* (8.14)	99.58* (2.28)
<i>Hot tasks</i>					
Children's Gambling Task	Binary	85.25* (14.36)	87.42* (9.98)	90.92* (5.14)	87.83* (6.31)
	Ternary	36.50 (25.58)	50.33 (28.36)	80.00* (18.23)	85.83* (7.86)
Delay of Gratification	Binary	83.06* (12.47)	86.11* (10.57)	88.06* (10.87)	90.56* (7.81)
	Ternary	17.50 (14.90)	39.17 (28.21)	73.33* (19.74)	80.28* (13.04)
Reversal Learning/ Cond. Discrim ^a	Binary	82.50* (8.34)	90.42* (8.63)	95.00* (4.66)	96.04* (4.90)
	Ternary	51.88 (9.64)	64.03* (17.30)	85.28* (11.40)	87.99* (11.10)
Theory of Mind	Binary	82.92* (11.60)	92.92* (12.14)	100.00* (0.00)	100.00* (0.00)
	Ternary	15.00 (19.75)	45.00* (28.75)	85.56* (22.63)	96.11* (11.32)

* Indicates means that are significantly above chance level.

^a Cond. Discrim: Conditional Discrimination

Table 2

Analyses of Complexity and Age Effects and their Interaction by Task

	Complexity	Age	Complexity × Age
<i>Cool tasks</i>			
Transitivity	$F(1, 116) = 1432.95, p < .001, \eta^2 = .925$	$F(3, 116) = 124.85, p < .001, \eta^2 = .764$	$F(3, 116) = 60.62, p < .001, \eta^2 = .611$
Class Inclusion	$F(1, 116) = 459.99, p < .001, \eta^2 = .799$	$F(3, 116) = 79.04, p < .001, \eta^2 = .671$	$F(3, 116) = 49.92, p < .001, \eta^2 = .564$
DCCS	$F(1, 116) = 161.20, p < .001, \eta^2 = .523$	$F(3, 116) = 50.34, p < .001, \eta^2 = .566$	$F(3, 116) = 42.36, p < .001, \eta^2 = .523$
<i>Hot tasks</i>			
CGT	$F(1, 116) = 134.38, p < .001, \eta^2 = .537$	$F(3, 116) = 34.18, p < .001, \eta^2 = .469$	$F(3, 116) = 26.40, p < .001, \eta^2 = .406$
DoG	$F(1, 116) = 294.91, p < .001, \eta^2 = .718$	$F(3, 116) = 59.86, p < .001, \eta^2 = .608$	$F(3, 116) = 43.62, p < .001, \eta^2 = .530$
CD/RL	$F(1, 116) = 223.94, p < .001, \eta^2 = .659$	$F(3, 116) = 70.93, p < .001, \eta^2 = .647$	$F(3, 116) = 21.10, p < .001, \eta^2 = .353$

Table 3

Percentages of Children with Significantly Above-chance Performance on Binary- and Ternary-Relational Items by Task and Age Group (N = 120)

		Complexity (pass criterion)		Age Group			
		Binary	Ternary	3-years	4-years	5-years	6-years
<i>Cool tasks</i>							
Transitivity	Binary (7 out of 8)			53.33	93.33	100.00	100.00
	Ternary (4 out of 8)			0.00	0.00	50.00	90.00
Class Inclusion	Binary (6 out of 6)			33.33	63.33	96.67	96.67
	Ternary (4 out of 6)			3.33	3.33	80.00	90.00
DCCS	Binary (7 out of 8)			96.67	100.00	100.00	100.00
	Ternary (7 out of 8)			10.00	16.67	83.33	100.00
<i>Hot tasks</i>							
CGT	Binary (25 out of 40)			93.33	96.67	100.00	100.00
	Ternary (25 out of 40)			13.33	36.67	96.67	100.00
DoG	Binary (9 out of 12)			80.00	90.00	93.33	100.00
	Ternary (9 out of 12)			0.00	16.67	66.67	76.67
CD/RL	Binary (30 out of 48)			100.00	100.00	100.00	100.00
	Ternary (30 out of 48)			6.67	43.33	96.67	96.67
TOM	Binary (5 out of 8)			100.00	96.67	100.00	100.00
	Ternary (4 out of 6)			3.33	36.67	90.00	96.67

Figure Captions

Figure 1. Mean Z composite scores for hot and cool tasks by age group.

