CHILD DEVELOPMENT

Child Development, November/December 2016, Volume 87, Number 6, Pages 1856–1876

Coping With Logical Fallacies: A Developmental Training Program for Learning to Reason

Michael Christoforides
Ministry of Education

George Spanoudis
University of Cyprus

Andreas Demetriou
University of Nicosia

This study trained children to master logical fallacies and examined how learning is related to processing efficiency and fluid intelligence (gf). A total of one hundred and eighty 8- and 11-year-old children living in Cyprus were allocated to a control, a limited (LI), and a full instruction (FI) group. The LI group learned the notion of logical contradiction and the logical structure of the schemes involved. The FI group learned, additionally, to recognize other deductive reasoning principles. Reasoning improved proportionally to training. Awareness improved equally in LI and FI. Changes in reasoning and awareness changes were related to attention control and gf. Awareness mediated the influence of training on reasoning but not vice versa, suggesting that awareness is necessary for conditional reasoning. Implications are discussed.

Conditional reasoning is important for intelligent functioning, because it allows integration and evaluation of information (Johnson-Laird and Khemlani (2014). It is grounded on four logical schemes slowly mastered throughout childhood and adolescence (Moshman, 2011; Muller, Overton, & Reene, 2001): modus ponens (MP), modus tollens (MT), affirming the consequent (AC), and denying the antecedent (DA). Two of the schemes, MP and MT, are decidable and rather easy to grasp, because all information needed for a conclusion is present in the premises. In MP, if one accepts that "If A then B" and "A occurs," one must also accept that "B necessarily occurs." In MT, if B did not occur, it necessarily follows that A did not occur either. The other two, AC and DA are not decidable because the conclusion depends on information not given in the premises. Specifically, in AC, if B occurs, it does not follow that A would also occur, because a third, nonspecified factor, may be involved. In DA, it does not follow that B would not occur if A does not occur because a third factor may cause B. Thus, these two schemes are called "logical fallacies" because they may deceive the thinker to draw a conclusion that is not tenable. MP and MT are attained early in development, at 7–9 years, by practically everyone. The two fallacies are not mastered before the age of 11–12 years, and then no more than about one third of adults can handle them systematically (Gauffroy & Barrouillet, 2009; Johnson-Laird & Wason, 1970; Moshman, 2011; Markovits, 2014; Overton, 1990; Ricco, 2010; Wason & Evans, 1975).

AC and DA, the two logical fallacies, deceive the thinker that a conclusion can be drawn because they appear equivalent to MP and MT, respectively. Obviously, events and discourse in everyday life are often patterned according to them. A failure to recognize and resist them results in misinterpretations and wrong decisions. For this reason, the fallacies have been studied extensively in psychology and the cognitive sciences in search of their causes and of training methods that would enable thinkers to cope with them (Nisbett, Fong, Lehman, & Cheng, 1987; Ricco, 2010). The studies designed to train conditional reasoning have been met with limited success so far. This study implemented a

---

This article is based on the doctoral dissertation of the first author Michael Christoforides, which was conducted under the guidance of the other authors, and it was written by the third author Andreas Demetriou. Special thanks are due to Henry Markovits, Michael Shayer, and four anonymous reviewers for their constructive comments on earlier versions of this article.

Correspondence concerning this article should be addressed to Andreas Demetriou, University of Nicosia, Research Foundation, Makedonitissas Avenue, P.O. Box 24005, 1700 Nicosia, Cyprus. Electronic mail may be sent to ademetriou@ucy.ac.cy or demetriou.a@unic.ac.cy.

0009-3920/2016/8706-0017
DOI: 10.1111/cdev.12557

training program aiming to enable children to master the logical fallacies and specify the possible contribution of various aspects of mental processing, such as attention control and working memory (WM), and intelligence, such as inductive reasoning and cognitive flexibility. Below we will first review cognitive, developmental, and learning research on deductive reasoning and then state the predictions of the present study.

Cognitive Requirements

There is general agreement that intellectual functioning is based on information integration and evaluation processes. There is less agreement as to what these processes actually are. In psychometric theory, fluid intelligence (gf) is the main information integration engine. It involves processes abstracting similarities that enable the thinker to induce relations between objects and relations between relations (e.g., Athens is for Greece what Paris is for France). In principle, there is no limit to how far abstraction of similarities and induction of relations can go, spanning from simple perceptual similarities between actual objects to relations between relations linking mental constructs (Carroll, 1993; Hunt, 2011; Spearman, 1927). However, there is no agreement as to how gf is related to deductive reasoning. In Carroll (1993) 3-stratum theory, inductive and deductive reasoning reside in gf, and they interact through general intelligence (g). This interaction was never specified.

In cognitive theory, deductive reasoning is important for information integration and evaluation because it provides criteria for checking the truth and validity of relations between representations standing for reality. However, there is also no agreement about the relations between inductive and deductive reasoning. According to Rips (2001), they are clearly distinct, each activated when its satisfying conditions are met. According to Johnson-Laird and Khemlani (2014), they are closely related because they are both based on the construction of mental models that stand for the realities concerned, actual or mental. These models express possibilities, and the relations between the models involved are inferred on the basis of knowledge or information that ignited the inferential process.

Developmental Requirements

In developmental theory, there have been many models for conditional reasoning development. They ranged in emphasis from models assuming that an underlying general mental logic guides development to models assuming that inference is always domain-specific and thus reasoning is based on learning the constraints of different domains (Ricco, 2010). The current dominant view is that there is deductive reasoning competence emerging from simpler forms of reasoning, such as class and inductive reasoning, throughout childhood and adolescence (Overton, 1990). This emergence is crucially supported by developments in two important domains: first, attainment of metacognitive awareness about logical schemes and about truth and validity that allow systematic search for alternative possible implications of the representations involved in a chain of reasoning (Markovits, Thompson, & Brinsson, 2014; Moshman, 2011; Piaget, 1970; Piaget, 1976; Ricco & Overton, 2011). Second, changes in WM provide the necessary representational capacity for the representation of mental models or rules related to these possible implications (Barrouillet & Lecas, 1999; Rips, 2001). It is accepted that actual reasoning performance may often be less than optimal because reasoning competence, although increasingly powerful, may be compromised by various procedural factors, including content familiarity and relevance (Ricco, 2010).

Demetriou and colleagues (Demetriou et al., 2013; Demetriou, Spanoudis, & Shayer, 2014; Spanoudis, Demetriou, Kazi, Giorgala, & Zenonos, 2015; Zebec, Demetriou, & Topic, 2015) advanced a model aiming to explicate how the various processes are intertwined in development throughout childhood and adolescence. This model assumes that reasoning develops in four cycles, from birth to late adolescence. Growth through these cycles is associated with changes in both the nature of representation and their inferential interlinking. Changes in various aspects of mental efficiency, such as executive control and WM, also occur in liaison with representational and inferential changes. Here, we summarize the two last cycles, which are relevant to the present study.

The cycle of rule-based inference develops in two phases, first from 7 to 8, when rule-based inference emerges from the representations of the previous cycle, and then from 8 to 10, when rules are aligned and integrated with each other. Specifically, in the first phase, relations between semantic blocks defining generic concepts, such as object classes and number, come into focus, enabling children to grasp their underlying rules. In reasoning, MP is a first strong sign of mastering the thread running through representations of reality (the premises):

for example, birds fly; tagi is a bird; therefore, tagi flies. In the next phase, at 8–10 years, children can align rules, grasp their underlying relation, and check if arguments are consistent with each other vis-à-vis this relation. This is reflected in children's grasp of MT in this phase. MT requires one to invert the argument structure (e.g., birds fly; tagi does not fly; therefore, tagi is not a bird) and align it with the standard MP structure to check if they are consistent.

Thus, in this period, inference gradually shifts from representations to rules interrelating representations. This is reflected in awareness about mental processes interlinking representations, such as syntax in language (Olson & Astington, 2013) and logical necessity in reasoning (Miller, Custer, & Nassau, 2000). As a result, in this cycle, executive control becomes flexible enough to allow children to shift systematically between conceptual spaces (e.g., various object categories), activate space-specific instances, and interrelate them according to specific inferential or procedural constraints (e.g., Brydges, Reid, Fox, & Anderson, 2013; Demetriou & Christou, 2015; Demetriou, Spanoudis, & Shayer, 2014; Demetriou, Spanoudis, Shayer, van der Ven, et al., 2014).

In *the cycle of principles* interrelated rules are reduced to general principles that may be used to evaluate acceptable and nonacceptable relations between rules and representations, and generate predictions about their possible instantiations. For example, the integrated MP–MT inferential rule of the previous cycle is raised to the principle of conditional deductive inference when combined with the understanding that there may be other factors related to an effect in addition to those specified in the premises. When constructed, this principle renders the recognition of the logical fallacies possible (e.g., "birds fly; tagi flies; it is not decidable if tagi is a bird"), because it places premises in a context of stated and nonstated possibilities. In the first phase, from 11 to 13 years, fallacies may be grasped under familiar conditions feeding the inferential process with experience-based models showcasing the relations involved. In the second phase, from 13 to 17 years, these relations may be implemented into arbitrary or nonreal contexts, suggesting an explicit conception of the system involved (Demetriou et al., 2014).

The inferential process as such comes in focus in this cycle (Demetriou & Bakracevic, 2009; Demetriou, Efklides, & Platsidou, 1993; Demetriou & Kazi, 2006). Thus, in adolescence, executive control may explicitly be applied on inferential processes as such, allowing adolescents to choose between reasoning and/or heuristic processes according to the specificities of the problem at hand, evaluate relative truth and validity, and decide when a logical argument is nondecidable.

These cycles are marked by changes in the pattern of relations between reasoning and processes indexing processing efficiency, such as attention control and WM (Demetriou, Spanoudis, & Shayer, 2014; Demetriou, Spanoudis, Shayer, van der Ven, et al., 2014). Specifically, changes early in each cycle (i.e., at 6–8 years and 11–13 years) are predicted by changes in processing efficiency (e.g., attention control); changes late in the cycle (i.e., 4–6 years, 8–10 years, and 14–16 years) are predicted by changes in processes reflecting representational integration, such as WM. It was suggested that, at the beginning of cycles, changes in processing efficiency reflect inferential changes better than WM because they signify improvement in the command of recently acquired mental units. Later in the cycle, when these mental units are consolidated and they start to be interrelated, WM is a better index because alignment and interlinking of representations both requires and facilitates WM.

However, efficiency and WM index rather than cause transitions in reconceptualization. The critical causal factor for cognitive change is cognizance of mental processes and representations because it makes metarepresentation possible. That is, cognizance produces cognitive experiences about mental functioning (including self-evaluations and external feedback about its efficiency vis-à-vis a goal), and metarepresentation generates new representations and processing patterns by explicitly representing and encoding relations between earlier representations into new and more efficient forms (Demetriou, Spanoudis, Shayer, van der Ven, et al., 2014; Spanoudis et al., 2015).

The model summarized above suggests that logical schemes are special programs for information representation and integration. Therefore, mastering any logical scheme is a function of the representational and integrational demands of this scheme. The logical fallacies are actually simpler than they appear. They require grasp of a general principle integrating all four logical schemes above. This principle specifies that the four schemes are not symmetrical (e.g., MP does not imply AC and MT does not imply DA) and that two of them (AC and DA) are nondecidable, because there may be possibilities not specified in the representations given. To learn this principle requires (a) awareness of the four logical schemes, (b) the cognitive flexibility that is necessary to generate and align ensuing mental models, (c) the representational capacity required to represent these models to (d) induce their commonalities and differences and specify their implications, and (e) state (metarepresent) them as a general truth-validity principle. This principle will then guide future encounters with similar problems (Demetriou, Spanoudis, & Shayer, 2014). The present study aims to pinpoint how each of the factors contributes to the construction of this fundamental deductive reasoning principle.

Training Reasoning

Learning to reason has become the focus of considerable research because it may have important professional and life implications (Nisbett et al., 1987; Ricco, 2010). The focus of these studies varied depending upon their theory of reasoning. The studies assuming that logic is crucial in reasoning trained participants to master the truth tables associated with each of the various logical schemes. However, success was meager (e.g., Muller et al., 2001; Staudenmayer & Bourne, 1977), suggesting that focusing on the underlying logical relations is not enough. The studies assuming that reasoning emerges from pragmatic experiences that may direct children to grasp underlying logical relations organized training to provide examples and counterexamples related to each scheme. The assumption was that children would abstract the implications of each scheme and construct the necessary inferential patterns. This approach was successful among adolescents who were inferentially advanced enough to use the examples encountered to flesh out mental schemes they were already using. However, it was not successful with younger children with limited proficiency in conditional reasoning (O'Brien & Overton, 1980; O'Brien & Overton, 1982; Overton, Byrnes, & O'Brien, 1985). A third approach was based on Johnson-Laird's (2006) mental models theory. Training here aimed to enable children to envisage and handle the mental models necessary to represent the relations in each scheme and reason accordingly. This approach succeeded with participants who demonstrated a WM capacity high enough to enable them to represent the necessary relations involved in the critical mental models (Barrouillet, 1997; Simoneau & Markovits, 2003).

Rationale of the Study and Predictions

Obviously, none of the factors examined by the training studies above (logic, pragmatic examples, or mental models learning) were sufficient to generate the change wanted. This study capitalized on the successful aspects of these studies and the cognitive and developmental models about the organization and development of reasoning to design a learning program that would be more successful than the training studies available.

This study focused on two critical phases of conditional reasoning development: that is, the second phase of rule-based inference, from 8 to 10 years, when the two determinate schemes are mastered and the first phase of the principle-based inference, from 11 to 13 years, when fallacies come within reach. Third- and sixth-grade primary school children represented these two phases, respectively. Thus, we could test if a relatively short training program (about a month), simulating the grasp of awareness and related experience that presumably unfolds spontaneously over these two phases, would be enough to enable children to master the fallacies. In addition to a control group in each age group that did not received any training, there were two levels of training, limited (LI) and full instruction (FI; the terms "training" and "instruction" are used interchangeably). The LI group was explicitly induced into the logical structure of each of the four logical schemes and also into the notions of logical contradiction and consistency. It is assumed, based on the analysis above, that this is the minimal requirement for grasping the general principle integrating all four logical schemes into a system that specifies the logical implications of each scheme. The FI group learned, additionally, to recognize several crucial principles of deductive reasoning and practiced them by construction and use of related mental models. Moreover, to specify how, if at all, learning to reason depends on the various processing and intelligence processes discussed earlier, all of these processes (i.e., processing efficiency, WM, inductive reasoning, and cognitive flexibility) together with the four logical schemes and awareness about them were examined at pretest. This design allows testing the following predictions.

1. Both third and sixth graders would solve MP and MT tasks but not the AC and the DA tasks. However, sixth graders may outperform third graders in AC and DA tasks expressed through conventional relations, indicating that transition to principled reasoning is activated (Moshman, 2011; Muller et al., 2001). Thus, a significant effect of content is also expected.
2. Sixth graders would outperform third graders in awareness, indicating an emerging insight

into underlying inferential processes that are associated with the transition between the two phases.

3. In concern to structural relations, it is expected that all processes (i.e., executive efficiency, gf, deductive reasoning, and awareness) would emerge as distinct constructs. The strength of their relations may vary from low, according to some theories (e.g., Rips, 2001), to high, according to other theories (e.g., Demetriou, Spanoudis, Shayer, van der Ven, et al., 2014; Johnson-Laird & Khemlani, 2014).
4. The two training groups received the same training on the awareness of the four logical schemes. However, the FI received more training on various other aspects of deductive reasoning and related awareness. Naturally, (a) the FI group would gain more from training than the LI group in both reasoning and awareness. Also, a larger difference in reasoning than awareness may be expected, provided that they both received the same training on awareness, but the FI group practiced more. In concern to age, one might predict that (b) training would remove differences between the two age groups involved, canceling out any age differences in attainment after intervention.
5. The relations between learning in reasoning and learning in awareness and the other processes examined here would be different to reflect their differences in learning requirements. Specifically, (a) reasoning would depend more on gf rather than attention control because deductive reasoning capitalizes on general inferential processes to form scheme-specific reasoning patterns. However, (b) change in awareness would depend more on attention control and cognitive flexibility because these two dimensions of mental functioning reflect ability to focus and regulate mental processes.
6. In the literature, interaction between inferential processes in reasoning and awareness about them is not symmetrical. It is assumed that changes in awareness lead to changes in reasoning, but the inverse relation may not hold, because inference may run automatically (Ricco & Overton, 2011). A strong test of this assumption may be the role of each process as a mediator of training effects to the other process. Specifically, it may be predicted that the mediating role of awareness in reasoning change would be larger than the mediating role of reasoning in awareness change (Demetriou, Spanoudis, Shayer, van der Ven, et al., 2014; Spanoudis et al., 2015).

Method

Participants

A total of 180 children, equally drawn from third ($M_{\text{age}} = 9.18$, $SD = .45$, range = 8.13–10.33 years) and sixth grades ($M_{\text{age}} = 11.68$, $SD = .46$, range = 10.83–13.00 years) were examined. Therefore, the age range of children in each age group varied within the two successive developmental phases related to the hypotheses of the study. Males and females were almost equally represented in each grade (45 males and 45 females in third grade and 48 males and 42 females in sixth grade). These children were of urban origin, living in Nicosia, the capital of Cyprus. All children were Greek Cypriots; they were all Greek Orthodox and native speakers of Greek. They all came from middle-class families, with at least one parent having a university degree who worked as professionals, businessmen, or officers in public administration.

Through each child's school, an invitation for participation in the study was sent to children's families in several schools in Nicosia. The first 180 children whose parents permitted participation and satisfied criteria for inclusion were included in the study (i.e., typically developing children without any history of mental or physical health problems). Children came from 22 schools. The children of each grade were randomly allocated to the three experimental groups (i.e., control, LI, and FI, see below), according to the order of consent form return. That is, the first child in each grade was allocated to the FI, the second to the LI, and the third to the control group, etc. This composition of the sample frees results from any possible bias related to a specific classroom or school.

Tasks and Procedures

Four task batteries were used: the first addressed processing efficiency and WM, the second addressed gf and semantic/verbal fluency (S/Vf), the third addressed the four logical schemes of deductive reasoning specified above, and the fourth addressed awareness about them.

Processing Efficiency Batteries

Speed and control of processing battery. A series of computer administered Stroop-like tasks, first used by Demetriou, Christou, Spanoudis, and Platsidou (2002), were used to address speed and attention control. These tasks were similar to the standard Stroop task in that children read color words or recognized their ink color. That is, three Greek words, which have the same number of letters—κόκκινο (red), πράσινο (green), κίτρινο (yellow)—were used and participants were tested under two combinations of word meaning and ink color, that is, reading a color word printed in the same ink color and naming the ink color of a word denoting a different color. Participants were instructed to use the R, the G, and the Y keys for red, green, and yellow, respectively. A total of 32 (4 condition × 8 stimuli) stimulus arrangements were presented. To facilitate responding, a red, a green, and a yellow sticker were placed on the respective keys. Four response times were estimated: word compatible, word incompatible, color compatible, and color incompatible. This test was highly reliable (Cronbach's $\alpha = .82$).

WM battery. Three computer-administered tasks addressed WM. The geometrical figures task, first used by Demetriou, Mouyi, and Spanoudis (2008), addressed visuospatial short-term storage. In this task, triangles, squares, rectangles, hexagons, circles, open angles, and arcs were superimposed on each other to form configurations of increasing complexity. A total of 15 stimulus arrangements were presented. Nine of them involved 2, five involved 3, and one involved 4 superimposed figures, respectively. Two-figure arrangements were presented for 1 (3 items), 2 (3 items), 3 (2 items), or 4 (1 item) s; three-figure arrangements was presented for 1 (1 item), 2 (2 items), or 3 (2 items) s; the arrangement with four figures was presented for 4 s. The participant's task was to identify the stimulus arrangement among five alternatives presented immediately after the presentation of the stimulus arrangement.

The Corsi block task, first used by Case (1985), addressed visuospatial WM. A 5 × 5 squares layout was shown on screen; a cartoon stepped randomly in several of these squares. To test WM, children were asked to recall these squares in reverse order. The memory demand ranged from one to seven cartoon appearances (three sets for each level).

Daneman and Carpenter's task (1980) addressed phonological WM. Children read a sequence of sentences and their task was to store and recall, in order of presentation, the last word of each sentence. Levels of difficulty varied from 1 to 7, with two trials at each level. A point was given for each level if both the words and their presentation order were correct. A point was given for each correct word in each level and a point was taken away if the recall order of a level was wrong.

The score on each of the WM tasks was the sum of total correct responses given. Reliability was acceptable given the small number of tasks involved (Cronbach's alpha for the three WM tasks = .53).

Cognitive Batteries

gf battery. Naglieri's test of nonverbal ability addressed gf (Naglieri and Das, 1997). This test includes 38 items requiring spatial visualization, pattern completion, serial reasoning, and reasoning by analogy. This test correlates highly with Raven's standard progressive matrices (.62). One point was given for each matrix solved (total score varying between 3 and 36). This test was very reliable (Cronbach's $\alpha = .88$).

Sematic/verbal fluency battery. To test the possible relation between cognitive flexibility and reasoning learning (Markovits & Quinn, 2002; De Neys, Schaeken, & D'Ydewalle, 2005), a test of S/Vf was used. Children were asked to produce as many words as possible in 30 s, for each of six categories: animals, electrical appliances, fruits, toys, clothing, and terms denoting relatives. The score for each category was the number of category exemplars recalled in the time allocated and it varied between 3 and 9. Performance on this task obviously reflects both semantic fluency in accessing concepts from long-term memory and verbal ability in spelling them out. The time constraints imposed for recall also render this task an index of mental efficiency. The reliability of the test was good (Cronbach's $\alpha = .69$).

Reasoning and Awareness Batteries

Reasoning battery. A reasoning battery was developed for the purposes of the present study. This battery involved 24 tasks, six for each of the four logical arguments (i.e., MP, MT, AC, and DA). The tasks addressed to each argument involved conventional, arbitrary, and nonreal relations. Two tasks were used in each Argument × Content combination as follows: conventional content referred to standard entities and relations (e.g., "If the figure is a square then it has four angles"; "If the last digit of a number is 4, then this number is even"). Arbitrary content involved familiar entities, but their relations were arbitrary (e.g., "If it is a triangle then it is red"; "If the hidden number is a two-digit number then it has 12 cycles underneath"). Nonreal relations involved familiar entities declared to have an unreal property (e.g., "If it is a rapid then it has 8 legs"; "If the hidden figure is a circle, then it has right angles"). All but one (rapid) of concepts used in these tasks came from school mathematics and they were familiar to the children (i.e., square, circle, triangle, side, angle, digit, even, and rapid). These concepts were selected because they are not involved in pragmatic, causal, or permission relations that might interfere with the target conditional relations investigated here. Children were interviewed before taking the pretest reasoning battery for their knowledge about these terms. Very few children needed any explications about them.

The major premise of each argument above was presented together with a minor premise and three options for the conclusion (printed underneath). The minor premise defined the logical form of the argument (e.g., the minor premise of the first argument above stated that "the figure is a square," "the figure has four sides," "the figure does not have four sides," "the figure is not a square," for MP, MT, AC, and DA, respectively). Of the three options given for the conclusion, the first confirmed and the second denied either the major or the minor premise (according to the argument); the third option always declared the argument as undecidable (e.g., "based on the premises we cannot be sure if the figure has four sides or not"). Children were instructed to "take it for granted that the premises of each argument are true and choose, based on what these premises state, the right conclusion among the options provided."

Presentation order of the three content-specific sets was the same across participants, proceeding from supposedly easy to more difficult content: conventional, arbitrary, and nonreal relations. This presentation order may transfer experience across content types, facilitating participants to differentiate content from logical relations. This is a strict test of the hypotheses concerned with the effect of learning on content as it may show the possible resistance of difficult content to training. The presentation order of the eight tasks within each content set was as follows: MP, AC, MT, DA, MT, DA, MP, AC. This presentation order ensures a balanced interchange between easy and difficult schemes and efficiency in the examination processes Markovits et al., 1996; Quinn & Markovits, 1998; Simoneau & Markovits, 2003).

Performance on each task was scored on a 1 (*pass*)/0 (*fail*) basis. Cronbach's alphas for pretest and posttest were .68 and .82, respectively.

Awareness battery. The awareness battery was specifically designed for the present study and it included seven tasks. These tasks are based on similar tasks used in the past to examine changes in self-awareness about mental processes and specify their relations with development in cognitive ability (Demetriou & Kazi, 2006; Demetriou, Kazi, & Georgiou, 1999; Kazi, Demetriou, Spanoudis, Zhang, & Wang, 2012). These studies showed that awareness of cognitive processes recycles with intellectual development across developmental phases.

Each task involved a target conditional or transitive reasoning argument (two premises and the conclusion) and three arguments given as options. The format of the transitivity tasks was similar to the format of the conditional arguments. That is, there were two premises (e.g., George is heavier than Andy, Andy is heavier than Michael) and a conclusion stated as a question (i.e., "is George heavier or lighter than Michael?"). All awareness tasks involved conventional content which was different from the content used on any of the tasks included in the reasoning battery.

The participant was first asked to solve the target argument and then to choose which of the three options was similar to the target argument in "the way of thought" required to reach the conclusion. Before choosing one of the three options, the experimenter explicated "the way of thought" required by the target argument to children who failed to solve it, focusing on the model that is critical for the argument concerned. For example, the explication for a DA argument (e.g., "If one is ill one does not go to school; George is not ill; did George go or he didn't go to school?") focused on the nondecidable nature of the argument and referred to alternative causes that might have caused the answer (e.g., "someone who is not ill may not go to school for several reasons, including lack of transportation"). This manipulation was judged necessary to ensure that failure to solve the awareness tasks was not caused by lack of access to the logical scheme concerned. Therefore, the emphasis here was not on reasoning as such, but on metareasoning because this battery focused on judgments about the nature of the inferential processes involved in the various tasks rather than on the execution of these processes.

MT, AC, DA, and transitivity arguments were presented as target arguments. The choices involved combinations of all of these arguments, including the correct one. Transitivity was preferred over MP arguments for two reasons. First, MP are very easy for the children involved in the study and thus not appropriate to reflect changes in awareness possibly evolving with mastering the logical relations involved. Second, transitivity is clearly different from the other arguments, allowing uncovering any possible sensitivity to clear differences in logical structure. An example follows: Target argument: If it rains, the grass is wet. It rained earlier. Is the grass wet? It is explained why the grass will be wet. Then, children are asked to choose which of the arguments following requires the same line of thought: (a) If the car is out of petrol it doesn't start. The car started. Did it run out of petrol? (b) If it rains, there are clouds. There are clouds. Will it rain? (c) If Nic wins the lottery, he will buy a brand new car. Nic won the lottery. Will Nic buy a brand new car?

Awareness tasks were always presented after the reasoning tasks. This order was considered appropriate to ensure that inferential processes were "primed," in a way, so that awareness about them may emerge, if possible or available. Also, this order ensures that no transfer occurred from awareness to reasoning performance, at least at the pretest. Performance on each task was scored on a 1 (*pass*)/0 (*fail*) basis (alpha reliability = .75 and .79 at pre- and posttest, respectively).

Upon completion of the posttest, children were asked to justify their answers on 10 tasks (i.e., 1 MT, 4 AC, and 5 DA) that were representative of the Argument × Content (i.e., conventional, arbitrary, and nonreal relations). Obtaining these explanations might highlight how intervention influenced the representation of relations and the line of argument. Explanations were asked at the end of the experiment rather than at earlier phases because the reflections activated by explanations might unduly interfere with the training process.

Training

Training aimed to develop an analytic approach to propositions involved in an argument as contrasted to their everyday use, raise awareness about (a) the chain of relations between propositions leading to a conclusion, and (b) the four basic logical schemes and provide practice in the construction of mental models following from each scheme. Training always started with an argument given to the child to solve and evolved according to the plan related to the session concerned. There were six 20- to 30-min sessions, each focusing on a particular aspect of the inferential process and it was delivered individually. The aims of each session, instructions to children, and related examples are detailed in Table S1. It is noted that the order of presentation implements a systematic progression of training from general (e.g., everyday vs. analytic and formal approach to premises and conclusions) to more specific themes (e.g., contradiction and logical necessity) and culminating into the explicit representation of the four logical arguments. Children were introduced into each session's target concepts in reference to a specific problem and they were then asked to solve sample problems, receiving feedback about their answers.

The first session aimed to raise awareness about the analytical approach to logical arguments as contrasted to their "everyday" use in language. In the second session, children learned to differentiate between the stated and the possibly implied meaning of propositions and focus on the first. The third session focused on logical contradiction and truth. The fourth session focused on the notion of contradiction and consistency, aiming to enable children to recognize propositions which are consistent with a target proposition and propositions which are in conflict with it. The fifth session focused explicitly on the notions of logical necessity and sufficiency. The last session focused on the explicit recognition of the four logical schemes and the construction of alternative mental models implied by each. The content used in the training arguments was always different from the content used in pretest and posttest batteries.

Design and Procedure

All speeded, WM, and S/Vf tasks were addressed once, before training, because of both substantive and practical reasons. That is, performance on these tasks before the experiment was used to examine if individual differences in these dimensions influence learning as planned here. Addressing these batteries at posttest might be informative about the possible generalization of learning to these processes, but this would exceed the time constraints of the study. All reasoning and awareness tasks were given twice, before and after training to measure the possible impact of learning on the processes trained. There was no significant difference in performance on any battery before training. Children in the FI group participated in all six sessions. Children in the limited instruction group participated in Sessions 4 and 6.

Children were individually examined and trained on all tests by the same experimenter (the first author, who is a teacher by training) in a specially provided room at children's schools. That is, each child was first examined with all tests included in the pretest; reasoning and the processing efficiency batteries were given on a different day, in a randomized order. Following pretest, children were involved in the training program, depending upon their experimental group. Each training session was delivered on a different day, separated by an average of 3 days. The posttest reasoning and awareness batteries were identical to the pretest batteries and they were given at least one full week after and, on average, 15 days after the last training session. This delay was considered necessary to remove any possible recency effects of training on the posttest, rendering examination after training a delayed rather than an immediate posttest. Pretest took place from March through June 2007. Training and posttest took place from August 2007 through June 2008.

Results

Attainment

To specify possible differences in attainment as a function of the various factors involved in this study, several repeated measures analyses of variance were run. These analyses were considered appropriate to uncover possible training effects in their interactions with the other between- (age) and within-subject factors (logical relations, content, and awareness involved). No covariates were included in these analyses because interactions between processes were explored by various structural equation modeling (SEM) approaches to be described next. It is noted that preliminary analyses examined if there is any association between the school of origin of participants and pretest or posttest performance on the four logical schemes. No such association was found (p for all χ^2 for School × Logical Scheme association > .05). Also, preliminary analyses revealed no gender differences. Thus, both school of origin and gender were excluded from all analyses below.

The first analysis examined attainment in reasoning and it is related to the first of our predictions (see Table S2, for the raw mean scores used in this analysis). This was a 2 (grade) × 3 (training) × 2 (pretraining vs. posttraining) × Content (conventional, arbitrary, nonreal relations) by logical scheme (MP, MT, AC, DA) analysis with repeated measures on the last three factors. This analysis was applied on the sum scores of the six tasks addressed to each of the logical schemes (2 tasks × 3 content types—real, arbitrary, nonreal relations). Both the effect of grade, $F(1, 174) = 11.92$, $p < .001$, $\eta^2 = .06$, and the effect of instruction, $F(2, 174) = 15.39$, $p < .0001$, $\eta^2 = .15$, were highly significant and strong. Consistent with Prediction 1, the first effect reflected the fact that sixth graders performed better than third graders. Consistent with Prediction 4, trained children outperformed controls. Of the various within-subject effects, the comparison of pretraining with posttraining performance, $F(1, 174) = 175.33$, $p < .0001$, $\eta^2 = .50$, and its interaction with training, $F(2, 174) = 41.74$, $p < .0001$, $\eta^2 = .32$, were highly significant and very powerful, indicating that performance improved extensively as a result of instruction, proportionately to the degree of instruction received. Also, consistent with Prediction 1, the effect of logical scheme, $F(3, 172) = 362.82$, $p < .0001$, $\eta^2 = .86$, was extremely powerful, indicating that performance on MP and MT was much higher than performance on the fallacies. The effect of content, $F(2, 173) = 86.98$, $p < .0001$, $\eta^2 = .50$, was also highly significant and powerful, indicating that dealing with conventional relations was easier than dealing with nonreal relations ($M_{\text{difference}} = 0.853$, $SE = .09$, $p < .05$) or arbitrary relations ($M_{\text{difference}} = 1.13$, $SE = .09$, $p < .05$); the difference between nonreal and arbitrary relations was also significant, ($M_{\text{difference}} = 0.275$, $SE = .07$, $p < .05$). The interaction between these constructs was highly significant, $F(6, 169) = 21.42$, $p < .0001$, $\eta^2 = .43$. This effect reflected the fact that there was practically no difference between the three content types in performance on MP and MT tasks as it approached ceiling across all of them; the content effect noted above emerged in concern to both AC and DA. However, content interacted significantly with time and training, $F(4, 346) = 4.37$, $p < .002$, $\eta^2 = .05$. This interaction indicated that the training effect increased systematically with content: It was minimal for real relations (pre–posttest difference between LI and FI = .03 $SE = .04$, $p > .05$) and significant for arbitrary (pre–posttest difference between LI and FI = .22, $SE = .02$, $p < .05$) and nonreal relations (pre–posttest difference between LI and FI = .26, $SE = .04$, $p < .05$).

Therefore, success on the various tasks is consistent with the literature and expectations. Specifically, performance on MP and MT was close to ceiling at pretest among both third and sixth graders: 80% or more of all children solved all 12 MP and MT tasks already at pretest. Success on the fallacies was much lower. It can be seen that pretraining performance of all three experimental groups at both third (13%–16% on AC and 21%–24% on DA) and sixth grade (20%–29% on AC and 23–35% on DA) was very low. Posttraining performance rose considerably among the LI (43% and 49% on AC and DA, respectively) and the FI third graders (58% and 57% on AC and DA, respectively). Performance of LI (43% and 50% on AC and DA, respectively) and FI sixth graders (71% and 67% on AC and DA, respectively) was very similar. Thus, there was no change from pretraining to posttraining measures among controls; there was an average increase of about 29% among the LI group and an average increase of 40% among the FI group, regardless of grade.

To illuminate the influence of training on tasks of the three content types, average percentage success was calculated according to task category. This would highlight if learning liberated reasoning, so to speak, from its dependence on familiar content. It is clear that only FI was found to have this liberating effect. Specifically, no group of children mastered familiar fallacies before instruction (success on these tasks was always < 46%); however, all four groups, LI and FI at both grades, mastered them after instruction (success always > 58%). In concern to nonreal relations, LI did not enable third (mean success increased from 12% to 43%) or sixth graders (mean success increased from 21% to 46%) to reach a 50% success criterion that would credit them as an age group with the ability to deal with the fallacies. However, FI did enable both third- (mean success increased from 12% to 60%) and sixth-grade (mean success increased from 12% to 70%) children to solve this kind of task. Although the pattern was similar in concern to arbitrary relations, as expected (Prediction 1) they proved more difficult than nonreal relations as they were mastered only by FI sixth graders. Specifically, both LI third graders (mean success increased from 11% to 34%) and LI sixth graders (mean success increased from 7% to 34%) did not master the fallacies with arbitrary content. FI third graders did not (mean success increased from 13% to 43%) but FI sixth graders did master the arbitrary content (mean success increased from 8% to 61%). Obviously, training did not completely cancel age differences in deductive reasoning. That is, children of both age groups moved forward as a result of training but, contrary to Prediction 4ii, a distance between the two age groups was preserved.

The second analysis examined attainment in awareness and it is related to the second prediction (see Table S2, for the raw mean scores used in this analysis). This was a 2 (grade) × 3 (training) × 2 (pretraining vs. posttraining) × Logical Scheme (determinate vs. fallacies) analysis of variance with repeated measures on the last two factors. The sum of the scores on the four fallacies and the three determinate tasks were transformed into *z* scores to ensure comparability. Consistent with Prediction 2, both the effect of grade, $F(1, 174) = 48.78$, $p < .0001$, $\eta^2 = .22$, and training, $F(2, 174) = 10.50$, $p < .008$, $\eta^2 = .05$, were highly significant indicating that 11-year-olds outperformed 9-year-olds and trained children exceeded controls. The pretest–posttest comparison was marginally significant, $F(1, 174) = 3.67$, $p < .06$, $\eta^2 = .02$, but the Pre–Posttest × Training interaction was highly significant, $F(2, 174) = 6.87$, $p < .001$, $\eta^2 = .07$, indicating that children involved in training performed better than controls at posttest.

These results reflected (see Figure 1) the fact that an average of 20% and 38% of third graders solved the awareness tasks concerning the fallacies and the determinate logical tasks, respectively. The corresponding figures for sixth-grade controls were 47% and 62%, respectively. Of the LI third graders, 26% and 43% solved the fallacies and the determinate tasks, respectively. The corresponding figures for sixth-grade controls were 62% and 76%, respectively. Of the FI third graders, 43% and 66% solved the fallacies and the determinate tasks. The corresponding figures for sixth graders were 65% and 79%, respectively. This pattern suggests that awareness of determinate logical schemes emerges spontaneously at the age of 11 years, but it is within the reach of 8-year-olds if systematic training is available. LI, however, was not enough to make them aware of the fallacies. This, however, was sufficient to make 11-year-olds aware of them. These patterns are fully consistent with Prediction 2, but not with Prediction 4ii.

Structural Relations

To test predictions about relations between processes, several types of SEM were employed. These models aimed to map the effects of training and the possible involvement of processing efficiency, WM, intelligence, and conceptual/verbal fluency. SEM methods for modeling of training studies are preferable over more classical methods for two reasons. First, they allow the researcher to capture underlying latent constructs and specify how they relate with each other. Second, they allow one to separate effects at various levels of generality, freeing modeling from possible noise that is associated with possible covariation between observed measures (see Bentler, 2006; Demetriou et al., 2013; Joreskog, 1970).

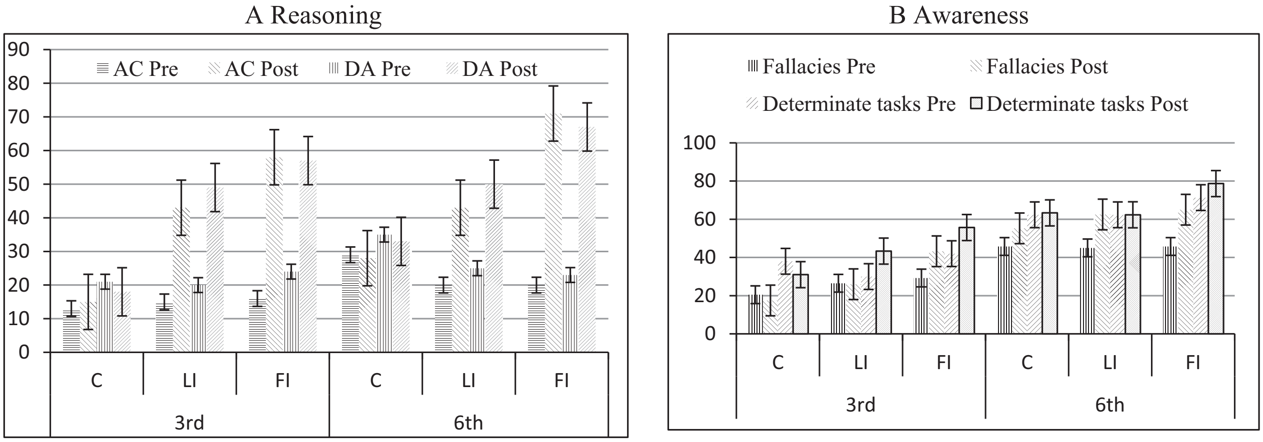

Figure 1. Mean percent success on reasoning and awareness tasks as a function of grade and experimental condition.
Note. AC = affirming the consequent; DA = denying the antecedent; Pre = pretest; Post = posttest.

In all models, the following factors were involved (see Figure 1). There was a factor for attention control, defined in reference to the incongruent measures (a processing speed factor defined in reference to the two congruent measures was found to be redundant to attention control and it was thus dropped). There was also a factor for WM related to the three corresponding tasks. To capture inductive reasoning, three sum scores were created to stand for performance on the Naglieri test (one for performance on spatial visualization and pattern completion, another for serial reasoning, and another for reasoning by analogy). These scores were related to a factor standing for gf. To capture S/Vf, three sum scores were created to stand for performance on the respective tasks (recalling animals and fruits, electrical appliances and toys, and clothing and relatives). These scores were related to another factor standing for S/Vf. To capture performance on the fallacies before and after intervention, two sum indicators were created for each testing, one for AC and one for DA, which involved all three content types (real relations, arbitrary relations, and nonreal relations). The two sum scores of each testing time were regressed on a different factor. These factors stand for deductive inference as such. It is noted that MP and MT were not included in these models. Limited variation because of ceiling performance strips these schemes of any informational value for the models. In the same fashion, two sum scores were created for logical awareness at each testing time. One of them involved awareness of determinate logical relations (i.e., transitivity and MT) and the other involved the awareness of the fallacies. The two scores at each testing wave were related to a different factor to stand for logical awareness before and after intervention. To facilitate model convergence, all scores above were transformed into *z* scores. The mean raw scores and standard deviations of the measures used in this analysis are presented in Table S3. The correlations between the *z* scores used in the various models are shown in Tables S4–S7. It can be seen that correlations between tasks addressed to the same construct were always significant and generally moderate to high. The maximum likelihood method was employed because scores conformed to a normal distribution. All participants were included in these models as there were no outliers to be excluded.

The WM, inductive reasoning, and S/Vf factors were regressed on a second-order factor reflecting general representational (WM), inferential power (inductive reasoning), and flexibility (S/Vf). This factor is a strong measure of gf. The pretraining factors for deductive inference and awareness were regressed on another second-order factor, standing for broad general deductive reasoning (gDR). This manipulation allows the dissociation of the effects of gf and gDR on postinstruction deductive inference and awareness from the effects of the specific processes represented by each of the specific factors (e.g., information storage, search and alignment, retrieval in gf and scheme use, and awareness in gDR).

Exploring the Overall Structure of Processes

These factors were used in two different types of models. A series of MIMIC models (models with multiple indicators and multiple causes; Bentler, 2006) were first used to explore the relations between processes and the possible impact of training and age. In this model, both grade (third and sixth) and training (control, LI, and FI) were specified as categorical variables which were then used in the various relations as specified below. The MIMIC approach was combined with the simplex approach (Joreskog, 1970) in order to precisely disentangle the effects of training and each of the various pretraining factors on each of the two posttraining factors. Specifically, the attention control factor was regressed on grade, gf was regressed on attention control, and gDR was regressed on gf. At a first step, each of the two posttraining factors was regressed on training, attention control, and the *residuals* of all other pretraining factors. In principle, this model assumes that performance after training varies as a function of training, a basic executive function indexed by attention control, and the *additional* inferential and representational processes represented by each of the other factors that are not accounted for by training and executive function. The fit of this model fell short of the standards for acceptability, $\chi^2(170) = 411.49$, comparative fit index (CFI) = .87, root mean square error of approximation (RMSEA) = .089, CI = .078, .10, model Akaike's information criterion (AIC) = 71.49. According to Iacobucci (2010), a good fit is indicated if the indices are as follows: $\chi^2/df < 2$, CFI > .9, and RMSEA < .08.

The relative misfit of the model was caused by the fact that some of the relations between pretraining and posttraining factors were very low. To purify the model, nonsignificant relations or relations with nonsignificant factor residuals were dropped from the model. This manipulation resulted in a significant improvement of the model fit, $\chi^2(173) = 383.45$, CFI = .88, RMSEA = .082, CI = .071, .093, model AIC = 37.41. Based on modification indices available, one extra relation was introduced into this model. Specifically, the factor standing for performance on the Naglieri test was regressed on grade, in addition to its regression on gf. This addition resulted in a large improvement of the model fit, rendering it perfectly acceptable, $\chi^2(172) = 315.17$, CFI = .92, RMSEA = .068 (CI = .056, .080; model AIC = −28.83).

The relations between processes examined at pretest were generally consistent with the third prediction. As expected, all processes emerged as distinct first-order factors. Also, WM, inductive reasoning, and semantic/verbal flexibility were related to one second-order factor standing for gf and deductive inference and awareness were related to another second-order factor standing for gDR. The relations between these general factors favored theories assuming that gf and deductive reasoning are strongly related. It can be seen in Figure 2 that the gf factor accounted for 61% (.78) of the variance of the deductive reasoning factor. Notably, attention control was highly related to gf (−.88), accounting for 77% of its variance. Thus, gf mediates strongly between conditional reasoning and executive control processes, in addition to its direct contributions specified above. That is, more than two thirds of the effect of gf on deductive reasoning (46% of 61%, .78 × −.88) came from the strong attention control–fluid intelligence relation.

The effect of training on the two posttest factors was consistent with the fourth prediction. That is, the effect of training on deductive inference (.51) was much larger than on awareness (.23). The relations between each of these two posttraining factors with the pretraining factors are shown in Table 1. It can be seen that these relations are generally consistent with the fifth prediction. Specifically, the reasoning factor, in addition to training, was significantly related not only to the pretraining deductive inference (.42), but also to attention control (−.18), gf (.16), and awareness (.22). The pattern of relations of posttest awareness was considerably different. In addition to training, this factor was highly related to attention control (−.68), pretest awareness (.60), gf (.32), and cognitive flexibility (.15).

Exploring the Influence of Learning on Relations Between Processes

The MIMIC models above showed how much of the postinstruction variance on the factors of interest came from instruction (in simpler words, how many individuals changed because of instruction). However, these models did not specify the precise size of this effect (i.e., how much these individuals changed) or the possible differentiation of the relations between factors as a result of instruction. To uncover these relations, a structured means analysis was implemented, involving the three experimental groups. This analysis was based on the final MIMIC model discussed above. Specifically, the same measurement–factor and factor–factor relations were involved in each group. Also, all measurement–factor relations and all pretraining factor–factor relations were constrained to be equal across the three groups, assuming that all measurements represented the measured construct equally well in the three groups and that the three groups were structurally identical before training. To compare the three groups with regard to change in reasoning

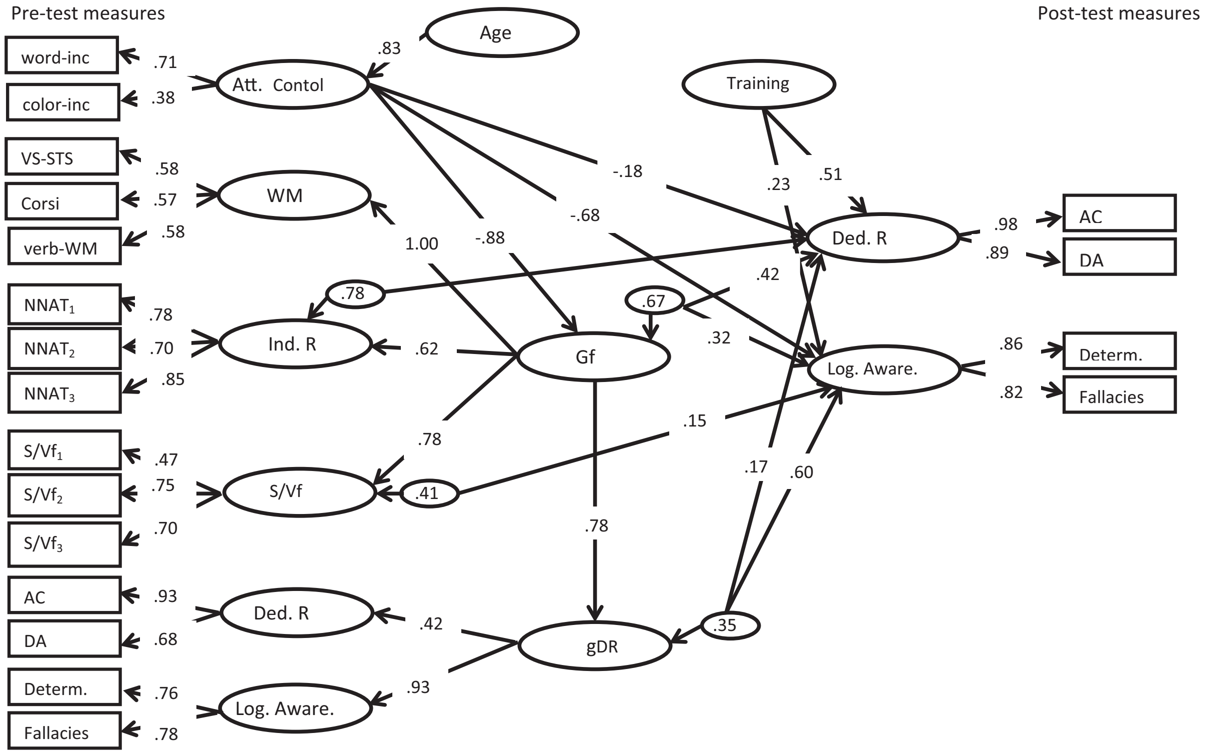

Figure 2. The models with multiple indicators and multiple causes model applied on pretraining and posttraining measures on the whole sample.

Note. All relations are statistically significant. Ellipses associated with factors depict residuals. Att. Control = attention control; WM = working memory; Ind. R = inductive reasoning; S/Vf = spatial/visual reasoning; Ded R = deductive reasoning; Log. aware = logical awareness; gf = fluid intelligence; gDR = general deductive reasoning; AC = affirming the consequent; DA = denying the antecedent; Determ. = determinate syllogisms; Fallacies = undecidable syllogisms.

Table 1

Structural Relations Between Processes and Intercepts Revealed by the MIMIC and the Structured Means Models

Process	AttC	gf	S/Vf	gDR	Reason	Aware	Instruction	Intercept	Var
Reasoning									
MIMIC	−.18	.16		.22	.42		.51	—	0.54
Structure means									
Control	−.38	*.14*			.66			0	0.60
LI	*−.19*	.37			.41			.72	0.34
FI	−.25	.43			.40			.92	0.41
Awareness									
Mimic	−.68	.32	.15	.60	—	—	.23	—	1.00
Structure means									
Control	−.37	.89				.26		0	1.00
LI	−.47	.52				.68		.35	0.95
FI	−.62	.48				.45		.36	0.82

Note. Fit indices for the MIMIC model: $\chi^2(172) = 315.17$, comparative fit index (CFI) = .92, root mean square error of approximation (RMSEA) = .068, CI = .056, .080, model Akaike's information criterion (AIC) = −28.83. Fit indices for the structural equation modeling: $\chi^2(476) = 597.35$, CFI = .93, RMSEA = .068, CI = .049, .083, AIC = –.354 = 65. Nonsignificant relations are shown in italics. MIMIC = models with multiple indicators and multiple causes; AttC = attention control; gf = fluid intelligence; S/Vf = spatial/visual intelligence; gDr = general deductive reasoning; Var = variance accounted for; LI = limited instruction; FI = full instruction.

and in awareness as a result of instruction, each of the observed variables and each of the five factors were also regressed on their intercept. The intercepts of the observed variables were specified as free parameters in each group, but each intercept was constrained to be equal across the three groups. The intercepts of all pretraining factors were fixed to 0 in all three groups to implement the premise that the three groups were identical in these abilities before instruction. This premise is also based on the finding that the three groups did not differ at pretest. However, the intercepts of the two postinstruction factors were fixed to 0 in the control group and they were specified as free parameters in the two instruction groups. Under these conditions, the intercepts of the variables may be taken as a kind of baseline for the variables. Differences in the means of the variables between the control group and each of the two instruction groups must, therefore, arise from our intervention, which, in the context of the study, was the only factor differentiating the three groups. These differences are expressed in the intercepts of the factors to which the variables are related. That is, significant factor intercepts in the group where the intercepts were set as free parameters indicate that this group has more (or less) of the ability represented by the factors than the other group because the intercepts in this later group were fixed to 0. Finally, it is noted that these intercepts may be taken as equivalent to effect sizes because they represent *z* rather than raw scores. The fit of this model was very good, $\chi^2(474) = 584.03$, CFI = .93, RMSEA = .064, CI = .045, .063, model AIC = −363.97.

The structural relations and the factor intercept values uncovered by this model are also shown in Table 1. It can be seen that all intercepts were significant, indicating that both levels of instruction were effective at changing both reasoning and awareness. However, the reasoning intercept of the FI group (.92) was significantly higher than the reasoning intercept of the LI group (.72; $z = 2.69$, $SE = .13$, $p < .025$). The awareness intercepts were practically identical (.34 and .36 in the LI and the FI group, respectively). Therefore, only change in reasoning varied as a function of the level of instruction. This pattern is close to Prediction 4i.

How did instruction impact relations between processes? It can be seen in Table 1 that these relations were generally consistent with Prediction 5. At a general level, consistent with Prediction 4i, both learning in reasoning and awareness were related with processing efficiency and gf and these relations varied with training. However, the profile of relations varied across groups and relations. Specifically, dependence of reasoning on attention control was similar in the three groups, varying in the low to moderate range (i.e., −.38, −.19, and −.25 for control, LI, and FI, respectively; negative relations reflect the fact that attention control scores are (RT) and thus decreasing with increasing speed). However, dependence of awareness on attention control increased systematically as a function of instruction (i.e., −.37, −.47, and −.62, respectively), indicating that training activates attention control which is reflected in enhanced awareness. Interestingly, however, dependence of reasoning on gf increased as a function of instruction; it was nonsignificant in the control group (.14) but moderate and significant in the LI (.37) and the FI group (.43), indicating that the more the effort required for learning the more investment is made in gf. Dependence of awareness on gf was always high but it was considerably higher in the control (.89) than the LI (.52) and the FI groups (.48). Self-regressions were significant and generally moderate to high for both reasoning (i.e., .66, .41, and .40 the three groups, respectively) and awareness (.26, .68, and .45, respectively). These results are consistent with Prediction 5, suggesting that awareness is closely related to executive control and cognitive flexibility in addition to its relation to gf reflecting general inferential power.

Exploring Learning × Age Interactions

To examine if the two age groups reacted differently to training, the structure means model above was tested in a two-group setup focusing on age. The first group included all third graders and the second group involved all sixth graders. The fit of this model was very good, $\chi^2(325) = 400.17$, CFI = .94, RMSEA = .051, CI = .031, .067, model AIC = −.249.83. There was a trend for the effect of training on reasoning to be stronger among sixth (.52) rather than third graders (.41); inversely, the effect of training on awareness was stronger among third (.27) rather than sixth graders (.17). Also, dependence of both, reasoning and awareness on attention control was stronger among third graders (−.29 and −.40, respectively) than sixth graders (−.12 and −.21, respectively). Dependence of reasoning (.20 vs. .21 for third and sixth graders, respectively) and awareness (.58 vs. .59 for third and sixth graders, respectively) on gf was practically identical as it was dependence on gDR (.13 vs. .19 and .62 vs. .73 for reasoning and awareness among third and sixth graders, respectively) and S/Vf (.17 vs. 22, respectively). Therefore, the only major difference between the two age groups was concerned with the relations between attention control and awareness.

Exploring How Processes Mediate Learning

How is the effect of training on each of the two posttest factors mediated by the other factor? A slight modification of the MIMIC model was employed to answer this question. Specifically, in this model, each of the posttest factors was regressed on both of pretest training factors, itself and the other one. To specify if reasoning carries any training effects on awareness, reasoning was regressed on awareness, and awareness was regressed, in addition to the two pretest factors, on posttest reasoning as well. To specify if awareness carries any training effects on reasoning, these two extra relations were inverted. That is, awareness was regressed, in addition to the two pretest factors, on posttest reasoning, and posttest reasoning was also regressed on posttest awareness. In these models, which fit the data well, $\chi^2(176) = 326.95$, CFI = .92, RMSEA = .068, CI = .056, .079, model AIC = −.30.05, the direct effects of training on reasoning (.52) and awareness (.24) were very similar to the effects obtained above. However, there was a vast difference in the indirect effects. Specifically, the indirect effect of training that reasoning carried on awareness, although significant, was low (.12). However, the indirect effect of training carried by awareness on reasoning was practically equal to the direct effect of training (.51). This is an impressive finding: Practically all of the effect of training on reasoning was mediated by awareness (i.e., 26% out of 27% of posttraining reasoning variance accounted for by training). However, only a very small amount of the effect of training on awareness was mediated by reasoning (i.e., only 1% out of 6% of posttraining awareness variance was accounted for by training. This huge difference in mediation is associated with the fact that training affected differently the relations between reasoning and awareness. Specifically, the relation between awareness before training with reasoning after training was very low (.07). This relation increased extensively and significantly after training (.39), $z = 2.84$, $p < .01$. The relation of reasoning before training with awareness after training was also very weak (.09). Training strengthened this relation (.20), but the increase was not significant, $z = .83$, $p > .05$. Also, the regression of reasoning on awareness at posttest (.39) was significantly higher than the regression of awareness on reasoning at posttest (.20), $z = 2.59$, $p < .01$.

Comparing cross-lagged relations estimated within each of the three experimental groups may shed light on the mechanism through which training exerts its impact. To estimate these relations, the structured means model presented above was modified so that posttraining reasoning and awareness were regressed on both pretraining factors. The cross-lagged relations do speak about the role of awareness as a conveyor of training effects. Specifically, the regression of posttraining reasoning on pretraining awareness increased systematically across the three experimental groups (i.e., −.05, .21, and .25 in the control, the LI, and the FI group, respectively). The difference between the control and the LI group was marginally significant ($z = 1.57$, $p < .06$), and the difference between the control and the FI group was significant ($z = 1.76$, $p < .05$). The regression of posttraining awareness on pretraining reasoning was always very low and nonsignificant (.09, .14, and −.09 in the three groups, respectively). Therefore, training improved the tuning between awareness and reasoning, resulting in improved reasoning performance. However, any improvement in reasoning per se as a result of training neither transferred to awareness nor it modified its relations with it. Obviously, these results are fully consistent with Prediction 6.

Qualitative Evidence

The differences between explanations given by children in the three experimental groups are consistent with the results above. First, many more trained children than control children invoked multiple models to explain their reasoning. Specifically, the majority of children in the control group advanced equivalence explanations (62% and 31% of third and sixth graders, respectively) rather than model-based explanations standing for an explicit MP–AC and MT–DA differentiation (0% and 7% of third and sixth graders, respectively). On the contrary, a relatively large number of children in the FI group (33% and 40% of third and sixth graders, respectively) advanced model-based explanations indicating this differentiation. Children in the LI group split between those who advanced equivalence explanations (17% and 27% of third and sixth graders, respectively) and those who advanced model-based explanations (14% and 13% of third and sixth graders, respectively). Second, there was an interaction between content and explanations. That is, even trained children who gave multiple models explanations to conventional tasks (i.e., 20%, 67%, and 67% of third graders and 63%, 60%, and 90% of sixth graders in the control, the LI, and the FI group, respectively) reverted to equivalence explanations of their responses to arbitrary tasks (i.e., 7%, 13%, and 30% of third graders and 7%, 23%, and 47% of sixth graders, for the three experimental groups, respectively).

Discussion

In short, the two determinate logical schemes, MP and MT, were fully mastered by third graders, but the two fallacies, AC and DA, were not mastered even by sixth graders. Our intervention program enabled both age groups to master the fallacies. The key to this success was awareness of the inferential identity of each scheme and the principle of logical consistency. Overall, awareness almost fully mediated the influence of training on deductive inference as such, although it was itself based on deductive inference more than on any aspect of control or awareness. However, awareness was highly dependent on attention control, and this relation strengthened with training. Several important cognitive, developmental, and educational implications emerge from these findings. These are discussed below.

Cognitive Implications

Cognitive and psychometric theories disagree about the composition of deductive reasoning and its relation with gf in general and inductive reasoning in particular. Our findings suggest, first, that inferential and awareness processes are integral constituents of deductive reasoning. The general factor accounting for pretest performance on reasoning, and awareness tasks substantiates this interpretation (see Figure 1). This is consistent with theories postulating that deductive reasoning is not possible without awareness (Demetriou & Kazi, 2006; Moshman, 2011; Ricco & Overton, 2011). Cognizance lifts inductive inference to deductive inference by imposing evaluative constraints and criteria on both the flow of inference and the conclusions possible. Cognizance may also moderate decisions about the allocation of mental resources.

Second, deductive reasoning is highly related to but is not identical with gf. The model in Figure 1 suggests that gf provides deductive reasoning with the representational possibilities needed to represent the premises involved (e.g., WM), the abstraction processes needed to integrate the premises along a relational theme (e.g., inductive inference), and the semantic/verbal flexibility needed to shift between conceptual spaces in order to envision the models implied by alternative conclusions (i.e., relational trials). The strong relation between attention control and gf in the model (Figure 1) suggests that gf also mediates between executive control processes and deductive reasoning. These findings favor theories claiming that inductive and deductive reasoning share inferential processes (Johnson-Laird & Khemlani, 2014) rather than theories claiming their independence (Rips, 2001).

Dual process theories of reasoning suggest that faultless automated inference would be preferable over conscious but effortful reasoning. The first is fast and errorless; the second is slow and error-prone, especially when many and unfamiliar factors need to be considered (Kahneman, 2011; Ricco & Overton, 2011). Thus, development or learning to reason would ideally shift the balance toward automated reasoning. Some findings in this study relate to this question. First, automating reasoning needs awareness at some point in time. Practically all of the change in deductive inference was mediated by awareness, but very little of training-related change in awareness was mediated by reasoning. Second, awareness change was heavily mediated by attention control and, to a lesser degree, by cognitive flexibility. This later finding suggests that the ability to focus processing on a target, the substance of attention control, reflects an awareness of both the process involved and the means needed to match this process to a stimulus and activate it accordingly. However, the participation of gf in deductive reasoning strengthened with training (.14, .37, and .43, across the three training groups, respectively), suggesting that fluent inferential processes gradually take control over reflective processes.

This pattern of results bears some implications for psychometric theories of intelligence. Specifically, the inferential component of deductive reasoning, in itself, is a weak index of general intelligence. This finding seems, superficially, consistent with literature suggesting that performance on conditional reasoning tasks is largely encapsulated into crystallized prior knowledge related to task content and language ability (Byrne & Tasso, 1999; Overton, Ward, Noveck, Black, & O'Brien, 1987). The very crystallization process, however, is a strong index of several important components of general intelligence, primarily executive control and awareness. These components mediate the construction of principles about inference rendering deductive reasoning possible (Demetriou & Kazi, 2006; Spanoudis et al., 2015). We turn to development to highlight how the various processes are interwoven with growth.

Developmental Implications

The developmental model outlined in the Introduction accommodates the pattern of acquisition of the four logical schemes and awareness about them. According to this model, MP is grasped in pragmatic contexts already at 5–6 years of age; pragmatic MP at this phase is basically an induction that locks two representations ("A occurs" and "B occurs") together into an inductive rule (i.e., "When A occurs, B also occurs"). Transition to the cycle of rule-based inference at 6–7 years lifts the representational alignments of the previous cycle into a rule-based representational imperative (A and B, A, therefore B). MT is grasped in the second phase of the cycle of rule-based inference, at 8–9 years, indicating that the representational imperative is fluent enough to be read both ways (A and B, not B, therefore not A).

The integration of MP and MT into a fluent inferential ensemble suggests that inductive imperatives are transformed into deductive necessities when the rules underlying their relations are explicitly metarepresented into a system specifying how different inferential spaces are interrelated. The rules are as follows: (a) Different representational spaces may have different inferential constraints (e.g., birds fly, mammals walk, fish swim, etc.) yielding different inductive implications about individual elements in each space (e.g., blackbirds fly, elephants walk, sharks swim, etc., respectively). (b) Moving across representational spaces is possible; however, shifting across spaces (e.g., imagining that "elephants are birds") implies transfer to the constraints of the new space (i.e., "elephants must fly"). (c) The primary premise defines the constraints of the space; the secondary premise only specifies an application domain of this space. Therefore, actual properties (e.g., elephants are mammals) are overwritten once they conform to the deductive rule "A & B, A → B," which cuts across spaces. Grasping the fallacies entails only one further metarepresentational step. (d) This is the suppositional stance that brings disparate representational spaces back into the deductive rule as a deductive moderator "$\text{A}_{(\text{but probably also C, D, E, . . .})}$ & B." When A vis-à-vis B is represented as one option among others the MP–AC and the MT–DA equivalence necessarily breaks because asserting B (AC) or denying A (DA) hints to the options beyond A. Obviously, grasping and integrating these rules into a smoothly running metalogical system is a major developmental construction that takes place throughout the last two cycles of development.

Pretest performance suggested that spontaneously grasping the last two steps is difficult even for sixth graders. Posttest performance suggested that this is feasible by systematic learning focusing on the necessary awareness and skills: that is, training freed children from the fallacy of the MP–AC and MT–DA equivalence, enabling them to envisage the alternative models that might be implied by the lack of information needed for a necessary conclusion. However, this fallacy is strong enough to surface any time the alternative models must be constructed without any aid from previous knowledge or experience. Performance across the three content domains and related explanations shed light on the process of overcoming the equivalence fallacy. That is, familiar tasks are easier because they are more likely to be associated with experience-based mental models than both arbitrary and nonreal tasks; nonreal tasks are easier than arbitrary tasks because nonreal tasks may be more easily supported by reality-geared counterexample models (e.g., we know that all rabbits have four legs) than arbitrary relations, which must be checked one by one vis-à-vis both the deductive rule and the deductive moderator. Therefore, building the *suppositional stance* is an important component of any reasoning training program. This stance must be intertwined with an explicit representation of a grid of implications associated with the deductive moderator above and practice to dominate over contextual fallacies. Below we explicate how the training program attained this target.

Learning and Educational Implications

It was noted in the Introduction that training using logic as such, counterexamples, or mental models met with limited success. This is to be contrasted with the success of the present intervention program. Admittedly, this program involved components of all three of the factors above: Although truth tables were never explicitly taught here, children were trained to recognize the four logical schemes and several other principles underlying conditional reasoning. Also, training in taking an analytical approach to premises as contrasted to their literal meaning facilitated children to adopt the suppositional stance to problems. They also practiced in envisioning mental models related to each scheme or principle, using them as examples implementing or countering each. Comparing and evaluating models for each scheme demonstrates the value of the suppositional stance.

These findings suggest that the key to the success of the program was explicit awareness of the four schemes and the notion of logical consistency that allowed their integration into the moderated deductive rule specified above. In terms of spontaneous developmental time, this short training program pulled children up by an almost full developmental phase, preserving a distance between ages. That is, trained third graders became able to cope with the fallacies if aided by context and trained sixth graders became able to cope with fallacies regardless of content and context. Interestingly, the awareness gained was commensurate to reasoning gains.

Building cognizance is a demanding process in several ways. First, to build up, cognizance requires mental resources and skills for handling them that would allow the thinker to assimilate information about inferential and representational processes into online reasoning tasks. The higher the demand for cognizance the more these resources and skills are needed. This is suggested by the fact that awareness was strongly related to attention control and that this relation increased systematically with increasing training (−.37, −.47, and −.62, for control, LI, and FI, respectively). Therefore, building and fluently using awareness to identify logical schemes and implement them accordingly requires systematic capitalization on mental resources available.

Second, mental resources vary with developmental phase, suggesting that building cognizance may also vary across phases. It is reminded that both awareness and reasoning were more closely related with attention control among third (−.40 and −.29, respectively) rather than sixth graders (−.21 and −.12, respectively). This suggests that the main driver of change in reasoning and awareness at the early steps of constructing the rules of inference is the efficient handling of mental resources. Later, other processes may be involved. Thus, awareness training in the cycle of rule-based inference may generate insight into the logical implications of the various schemes, but this insight is not yet crystallized into the metalogical rules specified earlier. The qualitative results summarized above suggest that this crystallization occurs in the cycle of principle-based inference. Thus, with development and/or relevant training deductive reasoning is freed from content or relational constraints, availing itself as a tool of informational, relational, and decisional evaluation. Admittedly, the drop of relations between attention control and awareness in this cycle suggests that the crystallization of these metalogical rules draws upon processes that escaped the present program. This assumption may be related to the finding that the training program changed deductive inference more than awareness. That is, one might argue that lifting inference into awareness and causing an explicit metarepresentation of this awareness into metalogical schemes that can readily be called upon is a cumbersome process that needs a more systematic ad hoc training program than was provided by this experiment. Further research is needed to resolve this issue.

The present study bears some important implications for education (see Demetriou & Christou, 2015). First, a program aiming to strengthen reasoning would have to build awareness about the logical schemes of interest that are appropriate for the developmental cycle of interest. To attain this aim, children must differentiate between general types of reasoning, such as inductive, analogical, and deductive reasoning and associate current tasks with the appropriate type. Also, they must acquire awareness of similarities and differences between logical arguments within each type. Second, training must highlight the line of inference related to each logical scheme and allow children to practice with exemplars varying in content from familiar to unfamiliar and arbitrary and unrealistic. Children must be able to use self-initiated mental models, varying in content and context, to check and validate conclusions, such as MP, MT, AC, and DA. Similarities and differences between exemplars should be pondered and explicitly represented. The aim would be to lift reasoning from a search of relations between mental models to a grasp of underlying logical relations (see Table S1, for examples). Third, training auxiliary processes, such as information search (attention), representation and storage (WM), and mental scanning (flexibility) would also help. The aim would be to enable children to focus attention on currently relevant information and familiarize them with general and personal representational limits. For instance, recalling digits of an increasing number would enable children to realize that there is a cut-off point between what can and what cannot be stored. Letting attention flow to irrelevant sounds may have the same negative effect. Thus, one must be careful to keep inference within the limits of what is personally feasible. Finally, training must be tuned to developmental phase. At preschool, children must realize that the information in the premises is connected by inference. At primary school, directed comparisons across the various arguments would enable children to differentiate form from content and understand that logic constrains inference. In adolescence, children must grasp the conditional and suppositional nature of reasoning and the role of form in constraining inference.

Limitations and Future Research

There are several directions in which future research might extend this study. First, extending this study in age phases not covered here might show if the present findings apply to other developmental phases as well. For example, it is important to know if building awareness related to the inferential possibilities associated with the cycle of representation-based inference (i.e., from 2 to 6 years of age) would accelerate transitions within and across this cycle in the fashion found to work here. It is also important to examine if these transitions relate to other aspects of mental functioning, such as executive control and gf in the way found here. Second, it is equally important to examine if the gains attained here in deductive reasoning are preserved over time and generalize to other aspects of reasoning not examined here, such as causal reasoning, and the various aspects of mental processing examined here only at the pretest. Third, the voluntary participation in the study may have positively biased the results because of relevant motivation and interest. Therefore, the present findings would have to be validated under conditions free of this possible bias. Finally, the current rise of interest in the relations between brain and cognitive development renders it highly interesting to examine the cognitive changes caused by the present intervention relate to brain changes (see Demetriou, Spanoudis, & Shayer, submitted; Mackey, Miller Singley, & Bunge, 2013).

References

Barrouillet, P. (1997). Modifying the representation of *if . . .then* sentences in adolescents by inducing a structure mapping strategy. *Current Psychology of Cognition*, *16*, 609–637.

Barrouillet, P., & Lecas, J. F. (1999). Mental models in conditional reasoning and working memory. *Thinking and Reasoning*, *5*, 289–302.

Bentler, P. M. (2006). *EQS 6 structural equations program manual*. Encino, CA: Multivariate Software.

Brydges, C. R., Reid, C. L., Fox, A. M., & Anderson, M. (2013). A unitary executive function predicts intelligence in children. *Intelligence*, *40*, 458–469. doi:10.1016/j.intell.2012.05.006

Byrne, R. J., & Tasso, A. (1999). Deductive reasoning with factual, possible, and counterfactual conditionals. *Memory and Cognition*, *27*, 726–740 doi:10.3758/BF03211565.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. New York, NY: Cambridge University Press.

Case, R. (1985). *Intellectual development: Birth to adulthood*. New York, NY: Academic Press.

Daneman, M., & Carpender, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, *19*, 450–466. doi:10.1016/S0022-5371(80)90312-6

De Neys, W., Schaeken, W., & D'Ydewalle, G. (2005). Working memory and counterexamples retrieval for causal conditionals. *Thinking & Reasoning*, *11*, 123–150. doi:10.1080/13546780442000105

Demetriou, A., & Christou, C. (2015). *Understanding and facilitating the development of intellect*. Educational Practices Series-26. Geneva, Switzerland: International Bureau of Education.

Demetriou, A., Christou, C., Spanoudis, G., & Platsidou, M. (2002). The development of mental processing: Efficiency, working memory, and thinking. *Monographs of the Society of Research in Child Development*, *67*, Serial Number 268. http://www.jstor.org/stable/3181583

Demetriou, A., Efklides, A., & Platsidou, M. (1993). The architecture and dynamics of developing mind: Experienstructuralism as a frame for unifying cognitive developmental theories. *Monographs of the Society for Research in Child Development*, *58*, (Serial No. 234).

Demetriou, A., & Kazi, S. (2006). Self-awareness in g (with processing efficiency and reasoning). *Intelligence*, *34*, 297–317. doi:10.1016/j.intell.2005.10.002

Demetriou, A., Kazi, S., & Georgiou, S. (1999). The emerging self: The convergence of mind, self, and thinking styles. *Developmental Science*, *2*, 387–409.

Demetriou, A., & Bacracevic, K. (2009). Cognitive development from adolescence to middle age: From environment-oriented reasoning to social understanding and self-awareness. *Learning and Individual Differences*, *19*, 181–194.

Demetriou, A., Mouyi, A., & Spanoudis, G. (2008). Modeling the structure and development of g. *Intelligence*, *5*, 437–454. doi:10.1016/j.intell.2007.10.002

Demetriou, A., Spanoudis, G., & Shayer, M. (2014). Inference, reconceptualization, insight, and efficiency along intellectual growth: A general theory. *Enfence*, *46*, 365–396.DOI: /10.4074/S0013754514003097

Demetriou, A., Spanoudis, G., Shayer, M., Mouyi, A., Kazi, S., & Platsidou, M. (2013). Cycles in speed-working memory-G relations: Towards a developmental-differential theory of mind. *Intelligence*, *41*, 34–50. doi:10.1016/j.intell.2012.10.010

Demetriou, A., Spanoudis, G., Shayer, M., van der Ven, S., Brydges, C. R., Kroesbergen, E., . . . Swanson, L. H. (2014). Relations between speed, working memory, and intelligence from preschool to adulthood: Structural equation modeling of 14 studies. *Intelligence*, *46*, 107–121. doi:10.1016/j.intell.2014.05.013

Gauffroy, C., & Barrouillet, P. (2009). Heuristic and analytic processes in mental models for conditionals: An integrative developmental theory. *Developmental Review*, *29*, 249–282. doi:10.1016/j.dr.2009.09.002

Hunt, E. (2011). *Human intelligence*. Cambridge, UK: Cambridge University Press.

Iacobucci, D. (2010). Structural equations modeling: Fit Indices, sample size, and advanced topics. *Journal of Consumer Psychology*, *20*, 90–98. doi:10.1016/j.jcps.2009.09.003

Johnson-Laird, P. N. (2006). *How we reason*. New York: Oxford University Press.

Johnson-Laird, P. N., & Khemlani, S. S. (2014). Toward a unified theory of reasoning. *The Psychology of Learning and Motivation*, *59*, 1–42.

Johnson-Laird, P. N., & Wason, P. C. (1970). A theoretical analysis of insight into a reasoning task. *Cognitive Psychology*, *1*, 134–148. doi:10.1016/0010-0285(70)90009-5

Joreskog, K. G. (1970). *Estimation and testing of simplex models*. Princeton, NJ: Educational Testing Service.

Kahneman, D. (2011). *Thinking: Fast and slow*. New York, NY: Penguin.

Kazi, S., Demetriou, A., Spanoudis, G., Zhang, X. K., & Wang, Y. (2012). Mind–culture interactions: How writing molds mental fluidity. *Intelligence*, *40*, 622–637. doi:10.1016/j.intell.2012.07.001

Mackey, A. P., Miller Singley, A. T., & Bunge, S. A. (2013). Intensive reasoning training alters patterns of brain connectivity at rest. *The Journal of Neuroscience*, *33*, 4796–4803.

Markovits, H. (2014). On the road toward formal reasoning: Reasoning with factual causal and contrary-to-fact causal premises during early adolescence. *Journal of Experimental Child Psychology*, *128*, 37–51. doi:10.1016/j.jecp.2014.07.001

Markovits, H., & Quinn, S. (2002). Efficiency of retrieval correlates with "logical" reasoning from causal conditional premises. *Memory & Cognition*, *30*, 696–706. doi:10.3758/BF03196426

Markovits, H., Thompson, V. A., & Brinsson, J. (2014). Metacognition and abstract reasoning. *Memory and Cognition*, *42*, 681–693. doi:10.3758/s13421-014-0488-9

Markovits, H., Venet, M., Janveau-Brennan, G., Malfait, N., Pion, N., & Vadeboncoeur, I. (1996). Reasoning in young children: Fantasy and information retrieval. *Child Development*, *67*, 2857–2872. doi:10.1111/j.1467-8624.1996.tb01892.x

Miller, S. A., Custer, W. L., & Nassau, G. (2000). Children's understanding of the necessity of logically necessary truths. *Cognitive Development*, *15*, 383–403. doi:10.1016/S0885-2014(00)00034-4

Moshman, D. (2011). *Adolescent rationality and development: Cognition, morality, and identity*. New York, NY: Psychology Press.

Muller, U., Overton, W. F., & Reene, K. (2001). Development of conditional reasoning: A longitudinal study. *Journal of Cognition and Development*, *2*, 27–49. doi:10.1207/S15327647JCD0201_2

Nisbett, R. E., Fong, G. T., Lehman, D. R., & Cheng, P. (1987). Teaching reasoning. *Science*, *238*, 625–631.

O'Brien, D., & Overton, W. F. (1980). Conditional reasoning following contradictory evidence: A developmental analysis. *Journal of Experimental Child Psychology*, *30*, 44–60. doi:10.1016/0022-0965(80)90074-0

O'Brien, D., & Overton, W. F. (1982). Conditional reasoning and the competence-performance issue: A developmental analysis of a training task. *Journal of Experimental Child Psychology*, *34*, 274–290. doi:10.1016/0022-0965(82)90046-7

Olson, D., & Astington, J. W. (2013). Preschool children conflate pragmatic agreement and semantic truth. *First Language*, *33*, 617–627. doi:10.1177/0142723713508869

Overton, W. F. (1990). Competence and procedures: Constraints on the development of logical reasoning. In W. F. Overton (Ed.), *Reasoning, necessity, and logic* (pp. 1–30). Hillsdale, NJ: Erlbaum.

Overton, W. E., Byrnes, J. P., & O'Brien, D. P. (1985). Developmental and individual differences in conditional reasoning: The role of contradiction training and cognitive style. *DevelopmentaI Psychology*, *21*, 692–701.

Overton, W. F., Ward, S. L., Noveck, I. A., Black, J., & O'Brien, D. P. (1987). Form and content in the development of deductive reasoning. *Developmental Psychology*, *23*, 22–30.

Piaget, J. (1970). Piaget's theory. In P. H. Mussen (Ed.), *Carmichael's handbook of child development* (pp. 703–732). New York, NY: Wiley.

Piaget, J. (1976). *The grasp of consciousness*. Cambridge, MA: Harvard University Press.

Pillow, B. L. (2008). Development of children's understanding of cognitive activities. *The Journal of Genetic Psychology*, *169*, 297–321. doi:10.3200/GNTP.169.4.297-321

Quinn, S., & Markovits, H. (1998). Conditional reasoning, causality, and the structure of semantic memory: Strength of association as a predictive factor for content effects. *Cognition*, *68*, 93–101. doi:10.1016/S0010-0277(98)00053-5

Ricco, R. (2010). The development of deductive reasoning across the life span. In W. F. Overton & R. M. Lerner (Eds.), *The handbook of life-span development: Cognition, biology and methods* (Vol. 1, pp. 391–430). Hoboken, NJ: Wiley.

Ricco, R. B., & Overton, W. F. (2011). Dual systems competence ←-→ procedural processing: A relational developmental systems approach to reasoning. *Developmental Review*, *31*, 119–150. doi:10.1016/j.dr.2011.07.005

Rips, L. J. (2001). Two kinds of reasoning. *Psychological Science*, *12*, 129–134.

Simoneau, M., & Markovits, H. (2003). Reasoning with premises that are not empirically true: Evidence for the role of inhibition and retrieval. *Developmental Psychology*, *3*, 964–975. doi:10.1037/0012-1649.39.6.964

Spanoudis, G., Demetriou, A., Kazi, S., Giorgala, K., & Zenonos, V. (2015). Embedding cognizance in intellectual development. *Journal of Experimental Child Psychology*, *132*, 32–50.

Spearman, C. (1927). *The abilities of man*. London, UK: MacMillan.

Staudenmayer, H., & Bourne, L. E., Jr. (1977). Learning to interpret conditional sentences: A developmental study. *Developmental Psychology*, *13*, 616–623. doi:10.1037/0012-1649.13.6.616

Wason, P. C., & Evans, J. S. B. T. (1975). Dual processes in reasoning? *Cognition*, *3*, 141–154.

Zebec, M., Demetriou, A., & Topic, M. (2015). Changing expressions of general intelligence in development: A 2-wave longitudinal study from 7 to 18 years of age. *Intelligence*, *49*, 94–109. doi:10.1016/j.intell.2015.01.004