Journal of Experimental Child Psychology 100 (2008) 252–263

Contents lists available at ScienceDirect

Journal of Experimental Child Psychology

journal homepage: www.elsevier.com/locate/jecp

Developmental trends in everyday conditional reasoning: The retrieval and inhibition interplay

Wim De Neys *, Deborah Everaerts

Experimental Psychology Lab, University of Leuven, Tiensestraat 102, Leuven 3000, Belgium

ARTICLE INFO

Article history:
Received 21 June 2007
Revised 4 March 2008
Available online 21 April 2008

Keywords:
Cognitive development
Decision making
Reasoning

ABSTRACT

Two experiments examined developmental patterns in children's conditional reasoning with everyday causal conditionals. In Experiment 1, a group of pre-, early, young, and late adolescents generated counterexamples for a set of conditionals to validate developmental claims about the counterexample retrieval capacity. In Experiment 2, participants in the same age range were presented with a conditional reasoning task with similar conditionals. Experiment 1 established that counterexample retrieval increased from preadolescence to late adolescence. Experiment 2 showed that acceptance rates of the invalid *affirmation of the consequent* inference gradually decreased in the same age range. Acceptance rates of the valid *modus ponens* inference showed a U-shaped pattern. After an initial drop from preadolescence to early adolescence, modus ponens acceptance ratings increased again after the onset of adolescence. Findings support the claim that the development of everyday conditional reasoning can be characterized as an interplay between the development of a counterexample retrieval and inhibition process.

Introduction

The ability to reason on the basis of a conditional "if–then" statement is considered one of the cornerstones of human cognition. However, the conditional inferences that we draw are not always logically appropriate (Evans & Over, 1996; Manktelow, 1999). Imagine, for example, that we are being told that "If Jane turns on the air-conditioner, she feels cool" and we receive the additional information that "Jane turned on the air-conditioner." In this case, standard logic tells us that we should always conclude that "Jane feels cool." This is the classic *modus ponens* (MP) inference. However, people

* Corresponding author. Fax: +32 16 326099.
E-mail address: wim.deneys@psy.kuleuven.be (W. De Neys).

0022-0965/$ - see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.jecp.2008.03.003

sometimes fail to draw this conclusion (e.g., Byrne, 1989; Cummins, 1995; De Neys, Schaeken, & d'Ydewalle, 2002; Markovits & Quinn, 2002; Thompson, 1994). At the same time, people sometimes accept invalid conclusions. For example, when being told that "If Jane turns on the air-conditioner, she feels cool" and "Jane feels cool," people might be tempted to accept that therefore "Jane turned on the air-conditioner." This conclusion, the so-called *affirmation of the consequent* (AC) inference (or fallacy), is not logically warranted. Research on conditional reasoning has been trying to identify the specific factors that mediate people's performance on these "if–then" problems (for an overview, see Markovits & Barrouillet, 2002). In the current article, we focus on specific developmental predictions with respect to the key factors that affect the conditional reasoning process.

One crucial factor that has become a major research issue is the retrieval of background knowledge about the conditional from long-term memory (e.g., Manktelow, 1999). Research has focused on the impact of two specific types of stored background knowledge or so-called counterexamples: alternative causes and disabling conditions. An alternative cause (alternative) is a possible cause that can also produce the effect mentioned in the conditional (e.g., taking off some clothes or turning on a fan in the preceding example can also make you feel cooler). A disabling condition (disabler) prevents the effect from occurring despite the presence of the cause (e.g., having a fever, having a broken air-conditioner).

The effects of retrieved counterexamples on adults' reasoning are well established. Numerous studies have shown, for example, that when people manage to retrieve an alternative, they no longer tend to commit the infamous AC inference (e.g., Byrne, 1989; Cummins, 1995; Quinn & Markovits, 1998; Thompson, 1994). That is, once we think of the possibility that Jane "might have had a cold beverage," we will be less inclined to conclude that it is necessarily the case that Jane turned on the air-conditioner when we are being told that she feels cool. Likewise, in case we retrieve a stored disabler, we will tend to reject the MP inference. We will be less inclined to conclude that Jane will necessarily feel cooler after she turned on the air-conditioner, for example, if we also consider the fact that "the air-conditioner may be broken" (e.g., Byrne, 1989; Cummins, 1995; De Neys, Schaeken, & d'Ydewalle, 2002, 2003a; Markovits and Quinn, 2002; Thompson, 1994). The outcome of the counterexample search directly determines the extent to which the inferences will be drawn. The more counterexamples that are retrieved, the less likely that the different conclusions will be accepted (De Neys, Schaeken, & d'Ydewalle, 2003b; Liu, Lo, & Wu, 1996).

Developmental studies have shown that the background knowledge mediation is also crucial in young children's reasoning. Janveau-Brennan and Markovits (1999), for example, found that the better elementary school children were at generating alternatives for a set of conditionals in a pretest, the less they tended to accept the AC inference in an actual reasoning task. Furthermore, because cognitive development will increase children's knowledge base and contribute to more efficient memory retrieval (e.g., Kail, 1992), successful counterexample retrieval should become more likely when children grow older. The studies with adults established that retrieval of alternatives and disablers decreases acceptance of the AC and MP inferences, respectively. Hence, if children's reasoning is indeed mediated by counterexample retrieval, the more efficient counterexample retrieval should result in an age-related decline in acceptance of the AC and MP inferences. A number of studies with preadolescents have clearly demonstrated this developmental pattern (e.g., Janveau-Brennan & Markovits, 1999; Markovits, 2000; Markovits, Fleury, Quinn, & Venet, 1998; Markovits et al., 1996); when elementary school children reason with familiar everyday conditionals, acceptance rates of the AC and MP inferences show a steady decline from Grade 1 to Grade 6 (i.e., 6–11 years of age).

Note that in standard logic, the AC inference is considered fallacious and should be rejected. The better alternative retrieval and resulting increased AC rejection for older elementary school children thereby contributes to an age-related increase in children's logical reasoning performance. However, standard logic also tells us that the MP inference is valid and should always be accepted. Hence, from a logical point of view, the more efficient disabler retrieval that accompanies cognitive development actually results in an age-related increase in reasoning errors on the basic MP problem.

Adult reasoners have been shown to sidestep this problem by selectively inhibiting disablers during everyday reasoning. At least those adults highest in cognitive capacity seem to adhere to a basic logical validity notion that will conflict with the disabler activation (e.g., De Neys, Schaeken, & d'Ydewalle, 2005a; Markovits & Doyon, 2004; Verschueren, Schaeken, & d'Ydewalle, 2005). This inhibition process is considered a second key component of the conditional reasoning process. An active inhibition of the tendency to take stored disablers into account thereby helps adults to reason in line with the logical standards. Consequently, adults' conditional reasoning has been characterized as a counterexample retrieval and inhibition interplay (e.g., De Neys et al., 2005a; Quinn & Markovits, 2002).

Markovits and Barrouillet (2002) suggested that the development of conditional reasoning can also be characterized as an interplay between the development of the retrieval and inhibition process. They expected that after the onset of adolescence, children's inhibitory capacities would become sufficiently strong to start overriding the impact of disablers. The development of the inhibition process should result in an increased MP acceptance for older adolescents. Interestingly, the few studies that have examined adolescents' reasoning with everyday conditionals seem to confirm this prediction. Contrary to the preadolescent trend, these studies tend to report an age-related increase in MP acceptance from early adolescence to late adolescence over the high school years (e.g., Barrouillet, Markovits, & Quinn, 2001; Klaczynski & Narasimham, 1998; Markovits, 1995; Markovits & Vachon, 1989).

However, there are a number of fundamental complications that prevent drawing clear conclusions. First, a simple comparison between the elementary and high school studies is not fully warranted given that different materials and tasks were used. As Markovits and Barrouillet (2002) indicated, the untested core prediction is that a direct comparison of elementary school children and adolescents would show a U-shaped MP acceptance trend from preadolescence to later adolescence; MP acceptance can be expected to be lowest for children at the start of adolescence (i.e., 12–13 years of age). On the one hand, these early adolescents probably will not yet be very successful at the disabler inhibition. On the other hand, disabler activation will be more likely for early adolescents than for younger preadolescents because they will have access to more disablers. Consequently, more disablers will be retrieved and MP acceptance initially should decrease from preadolescence to early adolescence. Although there will be even more disablers available for later adolescents, their inhibitory capacities should become increasingly sufficient to block the logically inappropriate disabler activation. Therefore, MP acceptance should start to increase again for the later adolescents.

Because retrieving stored alternatives does not conflict with the logical status of the AC inference, there is no need to inhibit the alternatives. Because the older age groups can be expected to be better at the alternative retrieval, one can predict a stable decreasing AC acceptance from preadolescence to later adolescence. In the current study, we tested these crucial trend predictions by directly comparing the performance of pre-, young, middle, and late adolescents.

A further problem for the retrieval and inhibition view is that the observed better logical performance for older adolescents is open to alternative explanations. This issue boils down to a wider controversy concerning the nature of the development of logical reasoning skills. One fundamental limitation of the previous studies with adolescents in this respect is that the availability of disablers for the selected conditionals was not always measured explicitly. Indeed, a closer examination of the method sections of these articles makes clear that the studies often adopted conditionals for which successful disabler retrieval is already quite hard for adults (e.g., "If one throws a ball against a wall, it will bounce"; "If wood is put in a fire, it will burn"). In the Klaczynski and Narasimham (1998) study, for example, adolescents' MP acceptance for the selected everyday conditionals and unfamiliar conditionals (e.g., "If a hydra is cut, then it will make tetmon" [i.e., nonsense conditionals for which people have, by definition, no access to disablers]) was quite similar. It is clear that if disabler retrieval for the everyday conditionals is unsuccessful anyway, inhibition is not required and a developmental increase in its efficiency cannot affect the inference acceptance. In this case, the better logical MP performance for later adolescents could be simply attributed to a more developed logical competence per se. In sum, the point is that the postulation of an additional inhibition process can be justified only when it can be shown that there are indeed possible disablers stored.

To validate the availability and retrieval assumptions, Experiment 1 of the current study presented pre-, young, middle, and late adolescents with a counterexample generation task in which they needed to retrieve as many alternatives or disablers as possible for a set of conditionals. The generation experiment will establish whether the different age groups manage to retrieve counterexamples for the adopted conditionals and whether there is indeed an age-related increase in the retrieval efficiency over the whole age range.

Experiment 2 tested the reasoning performance in the different age groups. The experiment also manipulated the counterexample availability as a further test of the role of the retrieval and inhibition components in the development of everyday reasoning; the reasoning task adopted conditionals for which the generation experiment indicated that age-matched participants could retrieve either many or few possible counterexamples. For all reasoners, it will be the case that the more counterexamples that are stored for a conditional, the more successful the retrieval can be. Hence, if adolescents' reasoning is mediated by alternative retrieval, not only should their AC acceptance show an age-related decrease, but also they should display lower AC acceptance for conditionals with many (vs. few) alternatives, just like younger reasoners.

Note that, in theory, the expected decreasing AC acceptance trend during adolescence could also be attributed to a more developed logical competence and not necessarily to a more efficient alternative retrieval process. Indeed, critics always could argue that young adolescents begin to master the formal concept of logical (in)validity and simply start using this knowledge to reason formally and reject AC. Hence, the access to counterexamples no longer would be critical for their inference evaluation. Contrary to our retrieval claim, such a "logical competence" explanation would predict that the impact of the availability of alternatives will tend to diminish in the older age groups; if adolescents no longer rely on alternative retrieval to reject AC, it will no longer matter whether many or few alternatives can be retrieved.

If adolescents start to inhibit the disabler activation, the inhibition can be expected to be more or less demanding depending on the disabler availability. It has been shown that, for conditionals with many disablers, the stored disablers will be automatically activated by a passive spreading of activation in the memory nodes (e.g., De Neys, Schaeken, & d'Ydewalle, 2002, 2005b; Quinn and Markovits, 2002); the disablers will automatically "pop up" in mind. Consequently, more disablers, and more strongly activated disablers (e.g., De Neys et al., 2003a), will need to be inhibited for conditionals with many (vs. few) disablers. One can expect that an increase in the inhibition demands will affect the success rate of the inhibition. For conditionals with many disablers, the inhibition will be far more demanding. Hence, fewer disablers will be filtered out successfully, and consequently MP should be less accepted than when only few disablers are available. Therefore, the inhibition hypothesis entails that although older adolescents' MP acceptance should increase overall, older adolescents should nevertheless show lower MP acceptance when there are many (vs. few) disablers available. If adolescents' reasoning would be determined purely by their logical knowledge of the argument structure, the mere availability of disablers should not affect their conclusions.

In sum, the current study attempts to validate the developmental characterization of everyday conditional reasoning as an interplay between the development of a retrieval and inhibition process. Experiment 1 examines the evolution of the counterexample generation performance for everyday conditionals from preadolescence to late adolescence, and Experiment 2 examines the evolution of the reasoning performance over this age range. The manipulation of the counterexample availability in Experiment 2 will thereby help to clarify the nature of the expected increased logical reasoning performance for older adolescents.

Experiment 1: Retrieval test

Method

Participants

A total of 48 preadolescents (mean age = 10.8 years, *SD* = 0.95), 26 young adolescents (mean age = 13.0 years, *SD* = 0.89),[1] 41 middle adolescents (mean age = 15.2 years, *SD* = 1.02), and 44 late adolescents (mean age = 17.3 years, *SD* = 0.81) volunteered to participate in the experiment. All children were recruited from the areas surrounding the University of Leuven, Belgium. Preadolescents attended elementary school, and all adolescents attended high school. All children were native Dutch speakers.

[1] Due to problems with participant recruitment in Grades 7 and 8, we managed to contact only a limited number of young adolescents. We opted to allocate the majority of these to Experiment 2.

Materials and procedure

Participants were requested to generate counterexamples for a set of four conditionals. Approximately half of the children in each age group were asked to generate disablers, whereas the other half were asked to generate alternatives. All conditionals expressed familiar causal relations (Table 1). Two conditionals were classified as having many possible alternatives and disablers in previous generation studies with adults (De Neys et al., 2002; Dieussaert, Schaeken, & d'Ydewalle, 2002), whereas the other two conditionals had few possible alternatives and disablers.

Item format and instructions were adopted from De Neys and colleagues (2002) and Cummins (1995). The following presents an example of the item format in the alternative generation task:

Rule: If the air conditioner is turned on, then you feel cool.
Fact: You feel cool, but the air conditioner was not turned on.
Please write down as many factors as you can that could make this situation possible.

Item format of the disabler generation task was similar except that under the heading "Fact:" would appear "The air-conditioner was turned on, but you don't feel cool." Items like these were constructed for each conditional; they were typed one to a page in a booklet.

All students were tested in small groups during a course break. Participants were given approximately 10 min to complete the test. The top sheet of the generation task booklet included the written instructions, which were read aloud to the participants. Participants were asked to keep trying to think of possible counterexamples for at least 1 min. The experimenter further clarified the instructions by discussing possible counterexamples for an example item (e.g., "If Jon lies in the sun, he will get sunburn") together with the children before the start of the actual generation test. Task instructions stressed the importance of producing items that were reasonably realistic and different from each other. Participants were instructed that simple variations of the same idea (e.g., for the earlier example: "taking off coat," "taking off shirt," "taking off sweater," "taking off tank top," "taking off cardigan," "taking off waistcoat," "taking off turtleneck") would be scored as a single item and needed to be avoided.

Coding and scores

The generation protocols were scored by a rater to identify unrealistic items and items that were variations of a single idea. The list of accepted counterexamples as judged by the two raters in the study of De Neys and colleagues (2002) was provided to clarify the rating task. Overall, 14% of the generated counterexamples were disallowed by the rater. Not surprisingly, there were fewer erroneous generations for the older participants (preadolescents = 18%, young adolescents = 17%, middle adolescents = 14%, late adolescents = 8%).

To get an indication of our rater's reliability, we asked a second rater to score the generation protocols of half of the participants in each age group. Interrater reliability reached .86.

Results and discussion

Approximately half of the participants in each age group generated disablers, whereas the other half generated alternatives. Half of the conditionals in the generation task were previously classified in studies with adults (Cummins, 1995; De Neys et al., 2002) as having many or few possible alternatives and disablers. To test whether the adult classification was valid for younger reasoners, the frequency factor (few or many) was entered as a within-participant factor in the analysis. This resulted in a 4 (Age Group) $\times$ 2 (Counterexample Type) $\times$ 2 (Counterexample Frequency) mixed-model analysis of variance (ANOVA).

Table 1
Conditionals for the generation task

If the brake is depressed, then the car slows down. (Many alternatives and disablers)
If water is heated to 100 °C, then it boils. (Few alternatives and disablers)
If John studies hard, then he does well on the test. (Many alternatives and disablers)
If the gong is struck, then it sounds. (Few alternatives and disablers)

As expected, there was a significant effect of age, $F(3, 151) = 39.60$, $MSE = 0.09$, $p < .0001$, $\eta_p^2 = .44$. A trend test showed that the number of generated counterexamples increased linearly across the different age groups, $F(1, 151) = 110.36$, $MSE = 0.94$, $p < .0001$, $\eta_p^2 = .42$. There was also a main effect of counterexample type, $F(1, 151) = 12.35$, $MSE = 0.94$, $p < .0001$, $\eta_p^2 = .08$, and a significant main effect of counterexample frequency, $F(1, 151) = 474.49$, $MSE = 0.42$, $p < .0001$, $\eta_p^2 = .76$, showing that participants generated more counterexamples for the many conditionals than for the few conditionals. These effects, however, were partially subsumed under a significant Counterexample Frequency × Counterexample Type interaction, $F(1, 151) = 26.41$, $MSE = 0.42$, $p < .0001$, $\eta_p^2 = .15$, and an Age Group × Counterexample Type interaction, $F(3, 151) = 20.17$, $MSE = 0.42$, $p < .0001$, $\eta_p^2 = .30$.

Simple effect tests of the Counterexample Frequency x Counterexample Type interaction indicated that although the effect was more pronounced on the alternatives, the counterexample frequency effect was significant for both disablers, $F(1, 305) = 108.10$, $MSE = 0.55$, $p < .0001$, $\eta_p^2 = .26$, and alternatives, $F(1, 305) = 227.62$, $MSE = 0.55$, $p < .0001$, $\eta_p^2 = .43$. Simple effect tests of the Age Group × Counterexample Type interaction showed that the effect of counterexample frequency was significant for all of the age groups: preadolescents, $F(1, 156) = 25.90$, $MSE = 0.49$, $p < .0001$, $\eta_p^2 = .14$; young adolescents, $F(1, 156) = 93.95$, $MSE = 0.49$, $p < .0001$, $\eta_p^2 = .38$; middle adolescents, $F(1, 156) = 141.07$, $MSE = 0.49$, $p < .0001$, $\eta_p^2 = .48$; and late adolescents, $F(1, 156) = 195.19$, $MSE = 0.49$, $p < .0001$, $\eta_p^2 = .57$. All generated more counterexamples for the many conditionals than for the few conditionals. A trend test revealed that the impact of the counterexample frequency increased linearly across the different age groups, $F(1, 156) = 37.15$, $MSE = 0.49$, $p < .0001$, $\eta_p^2 = .19$. Hence, the impact of the frequency factor was more pronounced in the older age groups, but all age groups showed the expected effect. The age, type of counterexample, and frequency factors did not interact, $F(3, 151) = 2.52$, $MSE = 0.42$.

Fig. 1 gives an overview of the findings. For completeness, the figure shows the impact of counterexample frequency on counterexample generation across the different age groups separately for alternatives and disablers. In sum, the findings establish that the counterexample retrieval efficiency increases with age and that the different age groups retrieve more counterexamples for conditionals for which adults have stored many counterexamples.

Experiment 2: Reasoning test

Method

Participants

A total of 47 preadolescents (mean age = 10.9 years, $SD = 0.82$), 42 young adolescents (mean age = 12.9 years, $SD = 0.91$), 45 middle adolescents (mean age = 15.1 years, $SD = 0.99$), and 45 late ado-

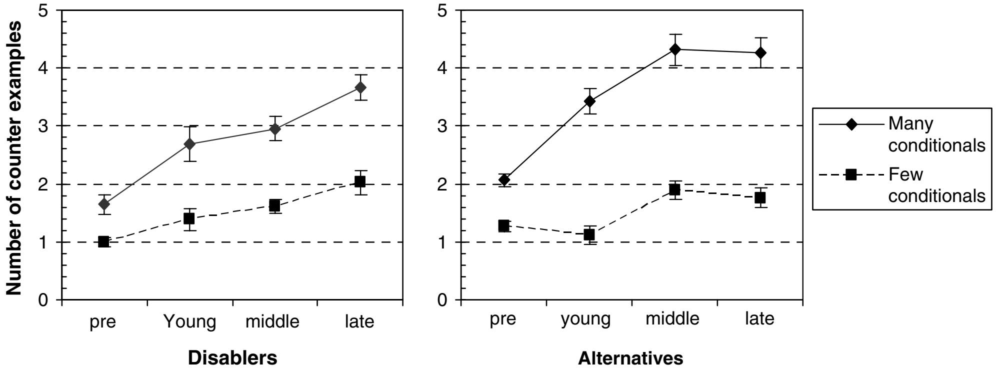

Fig. 1. Mean numbers of generated counterexamples (disablers or alternatives) in the different age groups for conditionals classified as having many or few possible counterexamples. Error bars are standard errors.

lescents (mean age = 17.3 years, *SD* = 0.82) volunteered to participate in the experiment. Participants were recruited from the same schools as the children in Experiment 1. All children were native Dutch speakers.

Materials and procedure

The four conditionals from Experiment 1 and four additional causal conditionals from the original generation study of De Neys and colleagues (2002) were selected for the reasoning task. The number of possible counterexamples of the selected conditionals varied systematically. The number of counterexamples constituted a 2 (few vs. many) $\times$ 2 (alternatives vs. disablers) design with two conditionals per cell. One conditional in each cell was embedded in the MP inference, whereas the other conditional was embedded in the AC inference. This produced a total of eight inferences for each participant to evaluate. The experiment was run on a computer. Each argument was presented on-screen together with a 7-point rating scale and accompanying statements (Cummins, 1995). This resulted in the following format:

Rule: If Jenny turns on the air-conditioner, then she feels cool.
Fact: Jenny turns on the air-conditioner.
Conclusion: Jenny feels cool.
Given this rule and this fact, give your evaluation of the conclusion:

------1------	------2------	------3------	------4------	------5------	------6------	------7------
Very	Sure	Somewhat	I	Somewhat	Sure	Very
Sure		Sure	I	Sure		Sure
			I			
That I CANNOT draw			I			That I CAN draw
this conclusion						this conclusion

Participants typed the number that best reflected their evaluation of the conclusion. The premises and conclusion were presented in yellow. The remaining text appeared in white on a black background.

Participants were tested in small groups in a computer room, with an individual booth for every participant. Instructions were presented verbally and on-screen. They showed an example item that explained the specific task format. Participants were told that the task was to decide whether or not they could accept the conclusions. Care was taken to make sure that all children understood the precise nature of the rating scale. Instructions stated that there were no time limits. The inferences were presented in random order. The experimental session was preceded by one practice trial.

It should be pointed out that the current instructions did not represent an attempt to rapidly school participants in the meaning of logical validity (Evans, 2002). Children were simply told that they could evaluate the conclusions by the criteria they personally judged to be relevant (e.g., Cummins, 1995; De Neys et al., 2002). Although the children were still situated in an experimental setting, this should have allowed them to reason as they would in everyday circumstances (Cummins, 1995; see also Galotti, 1989).

Results and discussion

Participants' mean inference acceptance ratings were subjected to a 4 (Age Group) $\times$ 2 (Inference Type) $\times$ 2 (Counterexample Frequency) mixed-model ANOVA with age group as a between-participants factor and with inference type and counterexample frequency as within-participant factors. Note that with respect to the frequency factor, we focus on the classic effect of the number of disablers on MP and the effect of the number of alternatives on AC. Thus, the frequency factor refers to disablers on MP and to alternatives on AC.

As expected, there were significant main effects of age group, $F(3, 175) = 5.80$, $MSE = 3.14$, $p < .01$, $\eta_p^2 = .09$, and counterexample frequency, $F(1, 175) = 76.83$, $MSE = 1.16$, $p < .0001$, $\eta_p^2 = .31$, but no effect of inference type, $F(1, 175) < 1$. There was also a significant Age Group × Inference Type interaction, $F(3, 175) = 5.43$, $MSE = 1.73$, $p < .005$, $\eta_p^2 = .09$, a significant Counterexample Frequency × Inference Type interaction, $F(1, 175) = 23.91$, $MSE = 1.01$, $p < .0001$, $\eta_p^2 = .12$, and a significant Age Group × Counterexample Frequency interaction, $F(3, 175) = 6.64$, $MSE = 1.16$, $p < .001$, $\eta_p^2 = .10$. All of these effects, however, were subsumed under a significant Age Group × Inference Type × Counterexample Frequency interaction, $F(3, 175) = 5.29$, $MSE = 1.01$, $p < .005$, $\eta_p^2 = .08$. Fig. 2 illustrates this finding.

Simple effect tests revealed that for MP, there was a marginally significant main effect of age, $F(3, 175) = 2.32$, $MSE = 2.69$, $p < .08$, $\eta_p^2 = .04$, a significant main effect of counterexample frequency, $F(1, 175) = 10.40$, $MSE = 0.99$, $p < .005$, $\eta_p^2 = .06$, but no interaction between age and counterexample frequency, $F(3, 175) = 1.59$, $MSE = 0.99$. As expected, all age groups exhibited less MP acceptance when many (vs. few) disablers were available. A trend test also established that the MP age trend was quadratic in nature, $F(1, 175) = 4.45$, $MSE = 2.69$, $p < .05$, $\eta_p^2 = .03$. As expected, after an initial drop, MP acceptance ratings increased again after the onset of adolescence.

For AC, simple effect tests revealed that there was a significant main effect of age, $F(3, 175) = 9.81$, $MSE = 2.18$, $p < .0001$, $\eta_p^2 = .14$, a significant main effect of counterexample frequency, $F(1, 175) = 87.14$, $MSE = 1.19$, $p < .0001$, $\eta_p^2 = .33$, and a significant interaction between age and counterexample frequency, $F(3, 175) = 9.70$, $MSE = 1.19$, $p < .0001$, $\eta_p^2 = .14$. Additional simple effect tests showed a significant effect of counterexample frequency for late adolescents, $F(1, 175) = 77.54$, $MSE = 1.19$, $p < .0001$, $\eta_p^2 = .31$, middle adolescents, $F(1, 175) = 21.13$, $MSE = 1.19$, $p < .0001$, $\eta_p^2 = .11$, and young adolescents, $F(1, 175) = 14.87$, $MSE = 1.19$, $p < .001$, $\eta_p^2 = .08$, but not for preadolescents, $F(1, 175) = 1.88$, $MSE = 1.19$. Hence, consistent with the "retrieval and inhibition" view of reasoning development, there is no indication whatsoever for the claim that the counterexample frequency effect disappears in the older age groups. All adolescent groups produced significantly lower AC acceptance ratings when many (vs. few) alternatives were available. Finally, trend tests established that the age effect on AC was linear in nature for the conditionals with many alternatives, $F(1, 175) = 46.62$, $MSE = 1.99$, $p < .0001$, $\eta_p^2 = .21$. AC acceptance decreased linearly as a function of age when many alternatives were available. When few alternatives were available, the linear trend was not significant, $F(1, 175) = 2.04$, $MSE = 1.37$, although a post hoc contrast suggested that the two oldest groups still tended to reject AC more than did the two youngest groups, $F(1, 175) = 2.81$, $MSE = 1.37$, $p < .10$, $\eta_p^2 = .02$.

General discussion

The current study characterized the development of everyday conditional reasoning as an interplay between the development of a counterexample retrieval and inhibition process (e.g., Markovits &

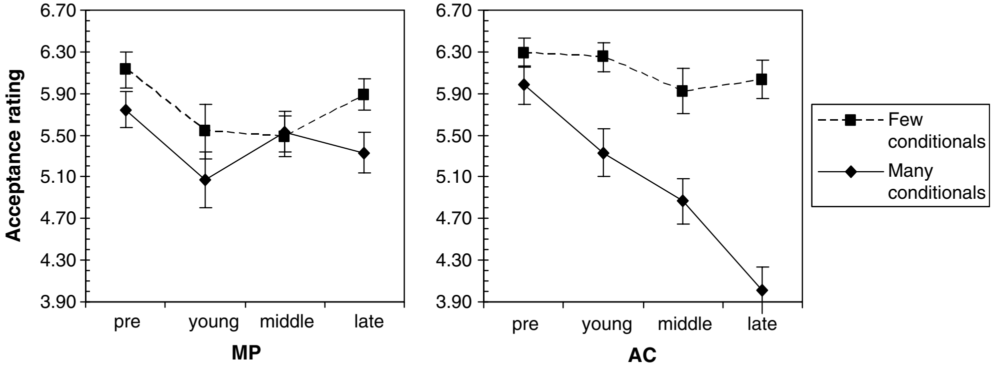

Fig. 2. Effect of the number of available counterexamples (few or many) on MP and AC acceptance in the different age groups. Error bars are standard errors.

Barrouillet, 2002). The results of the generation experiment (Experiment 1) established that the ease with which alternatives and disablers are retrieved gradually increases from preadolescence to late adolescence. Consistent with the age-related increased alternative retrieval, Experiment 2 showed that AC acceptance ratings gradually decreased over this age range. This effect was especially clear when many alternatives were available. The manipulation of the counterexample availability further established that inference acceptance in the individual age groups tended to depend on the ease with which alternatives could be retrieved.

Consistent with the age-related increased disabler availability observed in the generation experiment (Experiment 1), MP acceptance in Experiment 2 also decreased initially from preadolescence to early adolescence. Although there were even more disablers available for older adolescents, MP acceptance nevertheless increased for middle and late adolescents. This finding fits with the claim that after the onset of adolescence (i.e., 12–13 years of age), reasoners start inhibiting the logically inappropriate disabler activation. In addition, the counterexample availability manipulation showed that despite the overall superior performance, older adolescents' MP acceptance also decreased when many disablers were available. This further establishes that even for the most mature reasoners, MP acceptance depends on the efficiency of the inhibition process.

In the developmental literature, models that feature the key role of inhibitory processing in cognitive development have become increasingly popular (e.g., Dempster & Corkill, 1999; Dempster & Brainerd, 1995; Harnishfeger & Bjorklund, 1994). By now, there is abundant evidence that inhibitory processing is mediated by the (pre)frontal lobes of the brain (for a review, see Aron, Robbins, & Poldrack, 2004). The current U-shaped MP trend indicates that the disabler inhibition during everyday reasoning arises quite late and suddenly after the onset of adolescence. In this respect, it is interesting to note that the frontal lobes are known to be one of the last brain areas to develop, with full maturity not reached until young adulthood (e.g., Casey, Tottenham, Liston, & Durston, 2005). Much publicity has been given, for example, to the finding that the crucial myelination of the frontal neurons that optimize the impulse transfer is typically not complete until late adolescence. The spontaneous inhibition of salient biasing disablers in an everyday context is known to be quite difficult even for adults (e.g., De Neys, 2006; De Neys et al., 2005a). Given the demanding nature of the disabler inhibition, it makes sense that it would not start to show up before the frontal neural systems are sufficiently matured. Hence, it is possible that the late onset of the disabler inhibition can be attributed to specific neurological constraints (for studies on the role of the prefrontal cortex in belief inhibition during reasoning, see also De Neys, Vartanian, & Goel, in press; Goel and Dolan, 2003; Houdé, 2007).

At face value, the decreasing and increasing MP trend (with the shift around early adolescence) seems to fit with a more traditional Piagetian view that children do not enter the stage of formal reasoning until 12 years of age or older. Clearly, the claim that reasoners selectively inhibit disablers requires that they have some basic knowledge of logical validity; reasoners need to recognize the MP and AC argument structures and need to be aware of their different validity status. Otherwise, reasoners cannot tell whether they need to inhibit counterexamples or not. Hence, in stressing the role of retrieval and inhibition, we do not want to downplay the role of logical competence per se. Indeed, any inhibition account presupposes such competence. However, the current framework deviates from a more traditional view on at least two crucial points. First, it is not claimed that preadolescents do not yet acknowledge the validity of the MP inference (e.g., Morris, 2000). It is assumed that the "turning point" in the MP acceptance can be attributed to the development of inhibitory capacities rather than to a suddenly acquired logical competence. Second, the article argues against the view that adolescents reason on the basis of purely formal mechanisms whereby the increase in adolescents' performance would be attributed simply to the development of an abstract, decontextualized, logical knowledge skill. The results clearly showed that all age groups' inference acceptance ratings were affected by the number of available alternatives and disablers. If adolescents would start using an abstract logical "database" to reason, such content mediation no longer would be expected. The current study establishes that adolescents' everyday conditional reasoning is affected by the ease with which alternatives are retrieved and the ease with which disablers are inhibited. The findings thereby contradict any developmental model that would minimize the role of these factors.

De Neys, Schaeken, and d'Ydewalle (2003c) examined the impact of individual differences in working memory capacity on adults' everyday reasoning. Interestingly, the study reported capacity-related

MP and AC acceptance trends in adults' performance that were very similar to the currently observed age-related trends. AC acceptance tended to decrease linearly with increasing capacity, whereas MP acceptance showed a U-shaped trend. MP acceptance was lowest for adults with medium working memory spans. Thus, the current findings seem to fit with the general claim that developmental differences can become the individual variation of adulthood (e.g., Kuhn, Katz, & Dean, 2004; Kuhn, Weinstock, & Flaton, 1994). De Neys and colleagues (2003c) explained the U-shaped trend by arguing that only the adults highest in cognitive capacity would inhibit disablers spontaneously. The current study focused on the average performance in the different age groups. De Neys and colleagues' findings with adults suggest that not all adolescents will show the disabler inhibition and that there might be individual variation in the development of the inhibitory capacities. In this respect, it would be interesting to combine the current developmental approach with an individual differences approach (e.g., Handley, Capon, Beveridge, Dennis, & Evans, 2004; Simoneau & Markovits, 2003). Simoneau and Markovits (2003), for example, also showed that a task designed to measure the efficiency of inhibitory processing predicted whether or not MP was accepted when fourth and sixth graders reasoned with contrary-to-fact conditionals (e.g., "If a car has lots of gas, it will not run"). In sum, we do not exclude that for some individuals the inhibition may start at an earlier (or later) age or that others might never display it. The point is that the current findings indicate that the disabler inhibition for familiar conditionals is typically not displayed before the onset of adolescence.

It is clear that when evaluating the current findings, some limitations need to be taken into account. First, in the current study, we tried to assess children's reasoning performance in a natural everyday context. Therefore, children were not explicitly instructed to reason logically (e.g., Cummins, 1995; Evans, 2002; Galotti, 1989). We will point to some clear benefits of this procedure, but as one reviewer argued, it is also clear that more strict instructions might help to facilitate inhibitory processing for younger children. We agree that under these circumstances, disabler inhibition might show up before the onset of adolescence. Second, the current developmental findings are correlational in nature. It would be prudent to validate the results with a more direct experimental inhibition manipulation in future studies. For example, one straightforward prediction would be that manipulations that affect inhibitory processing efficiency, such as time pressure and dual task load (De Neys et al., 2005a; Evans & Curtis-Holmes, 2005), would have a differential impact on pre- and postadolescents' reasoning performance. In the meantime, the current correlational findings remain to be interpreted with caution. Third, although the current study established that counterexamples become more available with age, we did not pinpoint the exact factor that is responsible for the increase. We believe that the effect results both from an increase in semantic knowledge per se (older children have stored more counterexamples) and from an increase in retrieval efficiency (more processing resource available for retrieval). However, the current study did not allow us to disentangle the exact impact of each of the two factors.

Finally, we do want to point to some wider theoretical implication of the current findings. As argued earlier, the disabler inhibition phenomenon implies that a reasoner adheres to the validity of the MP inference. Previous demonstrations of children's ability to adhere to logical norms during reasoning have explicitly instructed or trained the participants (e.g., Barrouillet et al., 2001; Markovits & Vachon, 1989; Morris, 2000; but see also Klaczynski, 2001). Although this shows that adolescents and younger children are able to reason in accordance with normative standards and that they grasp the meaning of logical concepts when they are properly instructed, it does not show that this ability has something to do with children's spontaneous reasoning in everyday life (when the norm is not provided explicitly). This critique is not applicable to the current study. The children were not forced to reason logically, and none of them had received any specific training. The fact that 12- and 13-year-olds spontaneously start to respect the MP rule suggests that young adolescents might be far more logical then we have believed previously.

Acknowledgments

Wim De Neys is a postdoctoral fellow of the Fund for Scientific Research Flanders (FWO–Vlaanderen). The authors thank Ellen Gillard and Elke Van Gelder for their help in running the studies.

References

Aron, A. R., Robbins, T. W., & Poldrack, R. A. (2004). Inhibition and the right inferior frontal cortex. *Trends in Cognitive Sciences, 8*, 170–177.

Barrouillet, P., Markovits, H., & Quinn, S. (2001). Developmental and content effects in reasoning with causal conditionals. *Journal of Experimental Child Psychology, 81*, 235–248.

Byrne, R. M. J. (1989). Suppressing valid inferences with conditionals. *Cognition, 31*, 61–83.

Casey, B. J., Tottenham, N., Liston, C., & Durston, S. (2005). Imaging the developing brain: What have we learned about cognitive development? *Trends in Cognitive Sciences, 9*, 104–110.

Cummins, D. D. (1995). Naive theories and causal deduction. *Memory & Cognition, 23*, 646–658.

Dempster, F. N., & Brainerd, C. J. (1995). *Interference and inhibition in cognition*. San Diego: Academic Press.

Dempster, F. N., & Corkill, A. J. (1999). Individual differences in susceptibility to interference and general cognitive ability. *Acta Psychologica, 101*, 395–416.

De Neys, W. (2006). Dual processing in reasoning: Two systems but one reasoner. *Psychological Science, 17*, 428–433.

De Neys, W., Schaeken, W., & d'Ydewalle, G. (2002). Causal conditional reasoning and semantic memory retrieval: A test of the semantic memory framework. *Memory & Cognition, 30*, 908–920.

De Neys, W., Schaeken, W., & d'Ydewalle, G. (2003a). Causal conditional reasoning and strength of association: The disabling condition case. *European Journal of Cognitive Psychology, 15*, 161–176.

De Neys, W., Schaeken, W., & d'Ydewalle, G. (2003b). Inference suppression and semantic memory retrieval: Every counterexample counts. *Memory & Cognition, 31*, 581–595.

De Neys, W., Schaeken, W., & d'Ydewalle, G. (2003c). Working memory and everyday conditional reasoning: A trend analysis. *Proceedings of the Annual Conference of the Cognitive Science Society, 25*, 312–317.

De Neys, W., Schaeken, W., & d'Ydewalle, G. (2005a). Working memory and everyday conditional reasoning: Retrieval and inhibition of stored counterexamples. *Thinking & Reasoning, 11*, 349–381.

De Neys, W., Schaeken, W., & d'Ydewalle, G. (2005b). Working memory and counterexample retrieval for causal conditionals. *Thinking & Reasoning, 11*, 123–150.

De Neys, W., Vartanian, O., & Goel, V. (in press). Smarter than we think: When our brain detects we're biased. *Psychological Science*.

Dieussaert, K., Schaeken, W., & d'Ydewalle, G. (2002). The relative contribution of content and context factors on the interpretation of conditionals. *Experimental Psychology, 49*, 181–195.

Evans, J. St. B. T. (2002). Logic and human reasoning: An assessment of the deduction paradigm. *Psychological Bulletin, 128*, 978–996.

Evans, J. St. B. T., & Curtis-Holmes, J. (2005). Rapid responding increases belief bias: Evidence for the dual-process theory of reasoning. *Thinking & Reasoning, 11*, 382–389.

Evans, J. St. B. T., & Over, D. E. (1996). *Rationality and reasoning*. Hove, UK: Psychology Press.

Galotti, K. M. (1989). Approaches to studying formal and everyday reasoning. *Psychological Bulletin, 105*, 331–351.

Goel, V., & Dolan, R. J. (2003). Explaining modulation of reasoning by belief. *Cognition, 87*, B11–B22.

Handley, S. J., Capon, A., Beveridge, M., Dennis, I., & Evans, J. St. B. T. (2004). Working memory, inhibitory control, and the development of children's reasoning. *Thinking & Reasoning, 10*, 175–195.

Harnishfeger, K. K., & Bjorklund, D. F. (1994). A developmental perspective on individual differences in inhibition. *Learning and Individual Differences, 6*, 331–356.

Houdé, O. (2007). First insights on neuropedagogy of reasoning. *Thinking & Reasoning, 13*, 81–89.

Janveau-Brennan, G., & Markovits, H. (1999). The development of reasoning with causal conditionals. *Developmental Psychology, 35*, 904–911.

Kail, R. (1992). Processing speed, speech rate, and memory. *Developmental Psychology, 28*, 899–904.

Klaczynski, P. A. (2001). Analytic and heuristic processing influences on adolescent reasoning and decision making. *Child Development, 72*, 844–861.

Klaczynski, P. A., & Narasimham, G. (1998). Representations as mediators of adolescent deductive reasoning. *Developmental Psychology, 5*, 865–881.

Kuhn, D., Katz, J. B., & Dean, D. (2004). Developing reasoning. *Thinking & Reasoning, 10*, 197–219.

Kuhn, D., Weinstock, M., & Flaton, R. (1994). How well do jurors reason? Competence dimensions of individual variation in a juror reasoning task. *Psychological Science, 5*, 289–296.

Liu, I., Lo, K., & Wu, J. (1996). A probabilistic interpretation of "if–then.". *Quarterly Journal of Experimental Psychology A, 49*, 828–844.

Manktelow, K. I. (1999). *Reasoning and thinking*. Hove, UK: Psychology Press.

Markovits, H. (1995). Conditional reasoning with false premises: Fantasy and information retrieval. *British Journal of Developmental Psychology, 13*, 1–11.

Markovits, H. (2000). A mental model analysis of young children's conditional reasoning with meaningful premises. *Thinking & Reasoning, 6*, 335–347.

Markovits, H., & Barrouillet, P. (2002). The development of conditional reasoning: A mental model account. *Developmental Review, 22*, 5–36.

Markovits, H., & Doyon, C. (2004). Information processing and reasoning with premises that are empirically false: Interference, working memory, and processing speed. *Memory & Cognition, 32*, 592–601.

Markovits, H., Fleury, M., Quinn, S., & Venet, M. (1998). The development of conditional reasoning and the structure of semantic memory. *Child Development, 69*, 742–755.

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

Markovits, H., & Vachon, R. (1989). Reasoning with contrary-to-fact propositions. *Journal of Experimental Child Psychology, 47*, 398–412.

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.
Morris, A. K. (2000). Development of logical reasoning: Children's ability to verbally explain the nature of the distinction between logical and nonlogical forms of argument. *Developmental Psychology, 36*, 741–758.
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*, B93–B101.
Quinn, S., & Markovits, H. (2002). Conditional reasoning with causal premises: Evidence for a retrieval model. *Thinking & Reasoning, 8*, 179–191.
Simoneau, M., & Markovits, H. (2003). Reasoning with premises that are not empirically true: Evidence for the role of inhibition and retrieval. *Developmental Psychology, 39*, 964–975.
Thompson, V. A. (1994). Interpretational factors in conditional reasoning. *Memory & Cognition, 22*, 742–758.
Verschueren, N., Schaeken, W., & d'Ydewalle, G. (2005). Everyday conditional reasoning: A working memory-dependent tradeoff between counterexample and likelihood use. *Memory & Cognition, 33*, 107–119.