Developmental Psychology

When Less Is Not Always More: Stereotype Knowledge and Reasoning Development

Wim De Neys, and Karolien Vanderputte

Online First Publication, December 13, 2010. doi: 10.1037/a0021313

Developmental Psychology

0012-1649/10/$12.00 DOI: 10.1037/a0021313

When Less Is Not Always More: Stereotype Knowledge and Reasoning Development

Wim De Neys
Centre National de la Recherche Scientifique and Université de Toulouse

Karolien Vanderputte
University of Leuven

Developmental studies on heuristics and biases have reported controversial findings suggesting that children sometimes reason more logically than do adults. We addressed the controversy by testing the impact of children's knowledge of the heuristic stereotypes that are typically cued in these studies. Five-year-old preschoolers and 8-year-old children were tested with a card game version of the classic base-rate task. Problems were based on stereotypes that were familiar or unfamiliar for preschoolers. We also manipulated whether the cued stereotypical response was consistent (no-conflict problems) or inconsistent (conflict problems) with the correct analytic response that was cued in the problem. Results showed that an age-related performance decrease on the conflict problems was accompanied by an age-related performance increase on the no-conflict problems. These age effects were most pronounced for problems that adopted stereotypes that were unfamiliar for the 5-year-old preschoolers. When preschoolers were familiar with the stereotypes, their performance also started being affected. Findings support the claim that previously reported age-related performance decreases on classic reasoning tasks need to be attributed to the increased need to deal with tempting heuristics and not to a decrease in analytic thinking skills per se.

Keywords: reasoning, decision making, dual-process theory, heuristics

One of the main findings of reasoning and decision-making research over the last decades is that human judgment is frequently biased: In a wide range of reasoning tasks, people often do not give the answer that is correct according to logic or probability theory (e.g., Evans, 2003; Kahneman, Slovic, & Tversky, 1982). Influential dual process theories of thinking have explained this "rational thinking failure" by positing two different human reasoning systems (e.g., Epstein, 1994; Evans, 2008; Kahneman, 2002; Sloman, 1996; Stanovich & West, 2000). The common failure to provide the correct answer on classic reasoning tasks has been attributed to the pervasiveness of the so-called heuristic system. It is argued that human thinking typically relies on the operation of intuitive feelings and stereotypical beliefs instead of a deliberate, controlled reasoning process. Whereas the fast and undemanding heuristics can provide us with useful responses in many situations, they may also bias reasoning in tasks that require more elaborate, analytic processing. That is, both systems will sometimes cue different responses. In these cases, the logical, analytic system will need to override the intuitive belief-based response generated by the heuristic system (Houdé, 1997; Stanovich & West, 2000). Because the analytic operations heavily burden our limited working memory resources, the analytic override will frequently fail and the heuristic system will dominate our thinking.

Over the last few years, dual process studies with adult populations have provided substantial evidence for the framework (e.g., De Neys, 2006; Goel & Dolan, 2003; Houdé & Tzourio-Mazoyer, 2003; Stanovich & West, 2000). Attempts to extend the work with adults to children, however, have not always revealed the expected developmental trends. One of the central dual process claims is that analytic thinking is demanding and burdens our working memory, whereas heuristic thinking is effortless and operates automatically. Hence, one would expect that age-related increases in working memory will boost analytic thinking capacity and facilitate the possibility that erroneous heuristics are overridden. Consequently, dual process theories share with many developmental theories the general assumption that children's reasoning becomes more analytic and less heuristic with age (e.g., Kokis, Macpherson, Toplak, West, & Stanovich, 2002).

Empirical findings, however, have been mixed. Although some studies found the predicted pattern and showed that children's reasoning performance increased with age, other studies showed the exact opposite trend and reported that younger children behaved more logically than did adults (e.g., Davidson, 1995; De Neys, 2007; Jacobs & Potenza, 1991; Morsanyi & Handley, 2008; Reyna & Ellis, 1994). Jacobs and Potenza (1991), for example,

Wim De Neys, Centre National de la Recherche Scientifique (CNRS), Toulouse, France, and Laboratoire Cognition, Langues, Langage, Ergonomic (CLLE), Université de Toulouse, Toulouse, France; Karolien Vanderputte, Department of Psychology, University of Leuven, Leuven, Belgium.

This research was supported by grants from the Fund for Scientific Research—Flanders (Fonds Wetenschappelijk Onderzoek—Vlaanderen). Parts of this article were written during a Fonds Wetenschappelijk Onderzoek–funded stay of Wim De Neys at the University of California, Santa Barbara. We thank Russ Revlin for his hospitality and valuable comments on an earlier version of this article.

Correspondence concerning this article should be addressed to Wim De Neys, Laboratoire CLLE, Université de Toulouse, Maison de la Recherche, 5 allée A. Machado, 31058 Toulouse Cedex 9, France. E-mail: wim.deneys@univ-tlse2.fr

presented children with a version of Kahneman and Tversky's (1973) notorious base-rate neglect problems. In the base-rate task, a heuristic response based on a stereotypical personality description is pitted against more reliable statistical base rate information. In the problems people first get information about the composition of a sample (e.g., a sample with 90 lawyers and 10 engineers). People are also told that short personality descriptions are made of all the participants and they will get to see one description that was drawn randomly from the sample. Consider the following example:

> A psychologist wrote thumbnail descriptions of a sample of 100 participants consisting of 10 engineers and 90 lawyers. The description below was chosen at random from the 100 available descriptions.
>
> Jack is a 45-year-old man. He is married and has four children. He is generally conservative, careful, and ambitious. He shows no interest in political and social issues and spends most of his free time on his many hobbies, which include home carpentry, sailing, and mathematical puzzles.
>
> Which one of the following two statements is most likely?
>
> a. Jack is an engineer.
>
> b. Jack is a lawyer.

Logically speaking, given the size of the two groups in the sample, it will be more likely that a randomly drawn individual will be a lawyer. However, intuitively, many people will be tempted to respond that the individual will be an engineer based on heuristic beliefs cued by the description.

Kahneman and Tversky (1973), and numerous others since (e.g., De Neys, Vartanian, & Goel, 2008), observed that most adults neglect the base-rate information and tend to give the intuitive response that is cued by the description. When Jacobs and Potenza (1991) presented similar problems to 7-year-old first-graders, however, they found that the youngsters preferred the base-rate response much more frequently and easily outperformed adults and older children. Hence, participants seemed to be getting more biased and less analytic with increasing age.

The counterintuitive developmental findings have been interpreted as a severe blow for the standard dual process view (e.g., Klaczynski, 2000; Reyna, 2000). Nevertheless, it has been argued that the trends can be explained if one takes into account that heuristic thinking also needs to develop (e.g., Kokis et al., 2002; Reyna, Mills, Estrada, & Brainerd, 2006). Although dual process theories have primarily focused on the role of the analytic system, they do characterize thinking as an interplay of both analytic and heuristic processes. Kokis et al. (2002) suggested that the mere finding that younger children err less frequently on the classic decision-making problems is not necessarily problematic because the crucial heuristic knowledge will often be less developed for children. Because children lack knowledge of many stereotypes, they will be less biased by the very heuristic beliefs that are impeding adults' reasoning. Reyna, Brainerd, Mills, and colleagues (e.g., Reyna & Brainerd, 1994; Reyna & Mills, 2007) have since long stressed this same point in their developmental fuzzy-trace framework (see further). Hence, the basic idea is that children are benefitting from a "less is more" effect. Bluntly put, if you do not know what an engineer is, your background beliefs cannot bias your decisions either.

There is some evidence for this claim in the original Jacobs and Potenza (1991) study. Consider the actual adaptation of the base-rate problem that Jacobs and Potenza presented to their first-graders:

> In Juanita's class 10 girls are trying out to be cheerleaders, and 20 are trying out for the band. Juanita is very popular and very pretty. She is always telling jokes and loves to be around people. Do you think Juanita is trying out to be a cheerleader or for the band?

It is indeed not unreasonable to suggest that first-graders are less familiar with the cued stereotype (e.g., popular and pretty cheerleader vs. unpopular band geek) than are older participants. It is interesting that Jacobs and Potenza also observed that on related problems that did not seem to invoke stereotype knowledge (i.e., judgments with respect to objects rather than social classes) the 7-year olds no longer outperformed the older age groups. In sum, although participants' familiarity with the adopted stereotypes was not explicitly tested or manipulated, there are good reasons to believe that stereotype knowledge or familiarity is a major determinant of children's reasoning performance. This suggests that a fair comparison of analytic reasoning skills calls for a situation in which different age groups need to deal with biasing heuristics. In the present study, we present a systematic examination of these claims to reconcile the conflicting developmental findings in the field.

In our study, we presented base-rate problems to a group of 5-year old preschoolers and 8-year-old third graders. In a pretest, we first looked for stereotypes that both age groups were familiar with and stereotypes that were unfamiliar for the preschoolers but familiar for the third graders (i.e., unfamiliar problems).

Together with the stereotype knowledge factor (i.e., familiar or unfamiliar for the youngest age group) we also manipulated a crucial factor with respect to the conflict between the cued heuristic and analytic problem solutions. Reasoning and decision-making studies have focused on the biasing impact of heuristic thinking. However, heuristic thinking is not always wrong. In daily life, the two systems will often cue the same response. Hence, the default heuristic response will not conflict with analytic considerations, and people can rely on mere heuristic thinking without any need to engage in further demanding analytic processing. Indeed, dual process theories explicitly state that the analytic inhibition of the heuristic system will only be triggered in case of a conflict between the two systems.

The classic reasoning and decision-making tasks have capitalized on the need for analytic thinking and are typically constructed in such a way that the two systems cue conflicting responses. In the present study, however, we presented both traditional conflict and newly constructed no-conflict versions of the base-rate problems. In the no-conflict problems, we simply switched the base-rates around (e.g., the description describes an engineer that was drawn from a sample with 90 engineers and 10 lawyers). This allows a crucial test of the developmental claims. On unfamiliar conflict problems the preschoolers should outperform the older age group because the youngsters will not be biased by heuristic knowledge yet. However, if the reason for the better performance is the lack of knowledge, we should observe the reversed pattern on the no-conflict problems. Indeed, the older children will have little trouble in solving these problems because they can rely on mere heuristic thinking. It is well established that heuristic processing is cognitively less demanding than analytic thinking (De Neys, 2006; Stanovich & West, 2000). Because the younger children lack this heuristic knowledge, they will be forced to engage in more demanding analytic thinking and should have more difficulties solving the problem. Hence, older reasoners should outperform younger ones on the no-conflict versions of problems with unfamiliar stereotypes, whereas the reversed trend is expected for the conflict versions. Consequently, the key prediction is an interaction between conflict status and stereotype familiarity. Furthermore, one may expect that on problems with familiar stereotypes both age groups should perform well on no-conflict problems and should be biased on the conflict versions. Hence, age effects should be less pronounced on the familiar problems.

Note that our predictions gain some further credence from the perspective of fuzzy trace theory (e.g., Brainerd & Reyna, 2001; Reyna, 2004). As in many dual process theories, in the fuzzy trace framework, thinking is conceived as an interplay between qualitative-heuristic and quantitative-analytic processing. The qualitative-heuristic process is assumed to operate on the gist of a problem, whereas quantitative-analytic processing is more detail-oriented and operates on verbatim representations. Although fuzzy trace theory, as standard dual process theories, entails that heuristic and analytic processing modes are available at all points in development, the theory stipulates that with increasing age reasoners will show an increasing tendency to rely more on heuristic thinking because of the computational advantages (e.g., fewer cognitive demands) of gist-based heuristic processing. It is this core idea of increased reliance on heuristic processing with age that sets fuzzy-trace theory apart from more classic developmental theories. Reyna and Ellis (1994), for example, already used this framework to explain why children are less susceptible to framing effects than adults. If one assumes that when solving base-rate problems, basing one's decision on the cued stereotypical description relies on gist-based processing, whereas basing one's decision on the base-rates relies on verbatim, quantitative processing (see Reyna & Brainerd, 1994, and, Reyna & Mills, 2007, for an overview of studies that are consistent with this claim), the fuzzy trace framework fits with the present predictions. On the unfamiliar conflict problems, one can expect that the youngest children will not manage to extract the gist of the problem. Therefore, they are bound to rely more on the base-rate information than are older children, for whom the stereotypes are familiar and gist-based processing will dominate their thinking. However, on the unfamiliar no-conflict problems the absence of gist-based processing will force children to engage in more demanding, error-prone quantitative processing, thereby giving an advantage to the older children who can rely on gist-based processing. Both age groups should manage to engage in gist-based processing for familiar material, hence minimizing any age effects on the familiar problems.

A final issue concerns the adopted task format. It is clear that Jacobs and Potenza (1991) needed to present more basic, stripped-down versions of the original Kahneman and Tversky (1973) problems that were better suited for a younger population. In their quest for simplification, however, they left out a crucial task component. Contrary to the original problem, Jacobs and Potenza's versions did not tell the children that the character was randomly selected (see example above). From a normative point of view, the base-rates only matter when the description is drawn randomly from the sample. If this condition is not met, it is perfectly fine to base one's decision on the descriptive information and select the stereotype response (e.g., Gigerenzer, Hell, & Blank, 1988). It follows that the counterintuitive findings might be nothing more than an artifact. Without random sampling instruction the task becomes a mere stereotype matching task and no longer requires analytic thinking. Hence, the more frequent selection of the stereotype response in the older age groups would be the appropriate normative pattern.

Gigerenzer et al. (1988) already demonstrated that misconceptions about the random sampling assumption could be avoided when participants could observe the drawing themselves. In the present study, we adopted this procedure to design a child-friendly version of the base-rate task that sidestepped the methodological shortcomings of the previous studies. Children were familiarized with the basic composition of the sample by showing them cards that depicted the two groups. For example, nine cards depicted a boy and one card depicted a girl. On the back of the cards, we presented the stereotypical descriptive information. In this case, for example, children would be told that on the back of the cards they would find the child's favorite toy (e.g., a toy truck or a doll). Next, children could observe how the experimenter shuffled the cards, put them in a bag, and randomly drew one card from the bag. The experimenter showed children the back side of the drawn card (e.g., a truck) and then asked them whether there would be a boy or girl on the front. This format maintained the crucial characteristics of the original base-rate problems while remaining appropriate for testing (very) young children.

Pretest

Although studies on stereotype use by younger children are sparse, there is evidence that even preschoolers are familiar with specific stereotypes concerning, for example, gender and body weight (e.g., Cramer & Steinwert, 1998; Williams, Bennett, & Best, 1975). On the basis of these literature findings and the expert opinion of kindergarten teachers, we constructed a questionnaire consisting of 21 short questions related to possible stereotype knowledge (e.g., "What toys do boys/girls like to play with most?" "What do thin/fat children like to eat most?" "What are daddies/mommies doing at home," "What sport are rich/poor kids playing?" "Where do Black/White people live?" etc.). The questionnaire was orally presented to a sample of 15 preschoolers (mean age = 5.56 years, $SD = 0.33$) and 28 third graders (mean age = 8.45, $SD = 0.41$). We used children's answers to select items for which both age groups gave a similar dominant response (i.e., familiar stereotypes) and items for which the youngest age group had not yet developed such a dominant response (i.e., unfamiliar stereotypes). Figure 1 gives an overview of the most appropriate material that was selected for the actual experiment.

Note that our choice for the specific age range (i.e., preschool vs. third grade) was motivated by our goal to test the impact of stereotype familiarity as early as possible (i.e., preschoolers) and to keep the task and material engaging enough for the older contrast group. Bluntly put, although one might reasonable assume that a card game about the favorite toys of little boys and girls might still be interesting and engaging for 8-year olds, the task might be less suited to keep a test group of young teenagers or adults motivated. Within the present age range, however, the selected material and

Familiar Stereotypes

Question: What do kids/grannies like to do most?
Dominant response: Kids like to play games (preschool: 93%; third grade: 89%), and grannies like to knit (preschool: 47%; third grade: 50%).

Question: What toys do boys/girls like to play with most?
Dominant response: Boys play with cars (preschool: 55%; third grade: 32%), and girls play with dolls (preschool: 87%; third grade: 61%).

Question: What do thin/fat children like to eat most?
Dominant response: Thin kids eat fruit (preschool: 67%; third grade: 82%), and fat kids eat candy (preschool: 80%; third grade: 68%).

Question: What do daddies/mommies do at home?
Dominant response: Daddies mow the lawn (preschool: 67%; third grade: 46%), and mommies clean and iron (preschool: 60%; third grade: 86%).

Unfamiliar Stereotypes

Question: What do Dutch people/Italians like to eat most?
Dominant response: Dutch people like cheese (preschool: 0%; third grade: 57%), and Italians like pizza/pasta (preschool: 13%; third grade: 100%).

Question: What do principals/workmen drink most?
Dominant response: Principals drink coffee (preschool: 13%; third grade: 64%), and workmen drink beer (preschool: 0%; third grade: 50%).

Figure 1. Pretest material used to construct base-rate problems in Experiment 1. Preschool indicates the percentage of preschoolers who gave the response. Third grade indicates the percentage of third graders who gave the response.

task format should minimize such a confound and help to maximize the ecological validity.

Experiment

Method

Participants. Fifty-four preschoolers (mean age = 5.64 years, *SD* = 0.26) and 46 third graders (mean age = 8.75, *SD* = 0.30) of a suburban kindergarten and associated elementary school participated in the study.

Material. The children were presented with plastic cards (6 cm × 7.5 cm) that had an image of a cartoon character on the front and an image of an object on the back side. Figure 2 shows an example.

In each problem, children were presented with 10 cards. The characters on the front sides belonged to one of two groups (e.g., girl or boy). The base-rate in each problem was nine to one. The object on the back side of the card was associated with a stereotypical characteristic of the group in question (e.g., doll or toy truck). The selected groups and objects were based on the pilot findings.

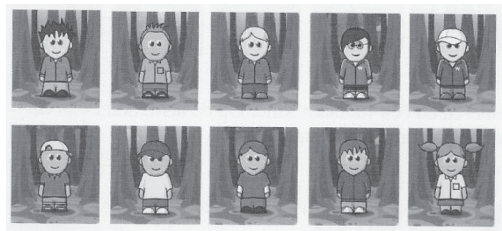
Figure 2. Front and back of one of the game cards.

For each problem the experimenter started by laying out the 10 cards in front of the child with the front sides up. Children were familiarized with the task content and observed how the experimenter shuffled the cards, put them in a bag, and randomly drew one card from the bag. Next, the experimenter showed children the back side of the drawn card with the stereotypical object (e.g., a truck) and then asked them to which one of the two groups the character on the other side would belong. Figure 3 presents a schematic overview of the instructions and procedure.

Children solved four problems with familiar stereotypes (i.e., familiar problems) and two problems based on stereotypes that were only familiar for the third graders in the study (i.e., unfamiliar problems).[1] Half of the familiar and unfamiliar problems were conflict items in which the shown object depicted a stereotypical characteristic of the smallest group in the sample. Hence, the analytic response, based on the sample size, and the heuristic response, based on the stereotype knowledge, conflicted. The other half of the problems were no-conflict items in which the object

[1] We had a hard time finding appropriate unfamiliar stereotypes in the pilot study. Most stereotypes seemed to be familiar or unfamiliar for both age groups. Instead of including less appropriate material in the unfamiliar condition or discarding appropriate familiar material, we decided to stick to an oversampling of familiar material.

<< The experimenter puts the 10 cards in front of the child >>



"On these cards you can see nine boys and one girl. On the back side of each card there is a picture of each kid's favorite toy."

"OK. As you can see, I'm shuffling the cards really well now, and then I put them in this bag. OK. Now, I'm going to draw one card from the bag without looking."

"See, this is the back side of the card that I drew:"

<< The experimenter shows the back side >>

"Now what do you think is most likely; will there be a boy or a girl on the front?"

Figure 3. Schematic overview of the instructions and task format.

depicted a stereotypical characteristic of the largest group. Hence, both analytic sample size considerations and heuristic knowledge cued the same response. Note that in order to manipulate the conflict nature of the problem, all 10 cards actually had the same object on the back side.

After children had solved the six selected stereotype problems, they were presented with a final abstract control problem. In this problem the cards did not depict a character or object but were simply colored yellow or blue. There were nine yellow cards and one blue card. The back sides of the cards were white. The experimenter showed the white back side after drawing it from the bag and asked children what color the other side would have. This control problem allowed us to check whether preschoolers had mastered the basic probabilistic skills to select the base-rate response.

On the conflict and control problems, responses that were in line with the base-rates (e.g., "boy" or "yellow") were scored as correct responses. On the no-conflict problems the base-rates and stereotypical knowledge cued the same response and selection of this response was scored as correct.

After children selected a response they were also asked for a verbal justification ("Can you tell me why you think so?") and an estimation of the group sizes in the problem ("Can you tell me how many *boys* and *girls* there were on the cards?). We included the estimation question as a basic manipulation check to make sure that our participants were paying sufficient attention and processed the preambles. We simply recorded participant's responses without any further cueing or probing by the experimenter. Note that work in the fuzzy trace tradition (e.g., Brainerd & Reyna, 1992a) has used recall-based measures to differentiate gist-based and verbatim processing and to study the relation between memory and reasoning performance. The present open-ended, unstructured estimation question was not designed to address these issues.

Procedure. All participants were tested individually. They were told that they would be playing a game of cards and that they would need to answer a couple of questions. The complete session lasted about 10 min, and it was videotaped for subsequent scoring. The problems were presented in a fixed order. The four problems with familiar stereotypes were presented first. Next, the two problems with the unfamiliar stereotypes were presented. Finally, the abstract control problem was presented. We alternated the conflict nature of the problems. Hence, a conflict problem was always followed by a no-conflict problem (and vice versa). Half the participants in each age group started with a conflict version, the other half started with a no-conflict version. The content of the conflict and no-conflict versions was fully crossed. Problems that were presented in a conflict version to half the participants were presented as no-conflict problems (by switching the base-rates around) to the other half of the participants in every age group.

Results

Reasoning accuracy. For each participant, we calculated the average performance on the conflict and no-conflict problems and subjected these to a 2 (age, between subjects) × 2 (conflict, within subjects) × 2 (stereotype familiarity, within subjects) analysis of variance (ANOVA). Figure 4 gives an overview of the results.

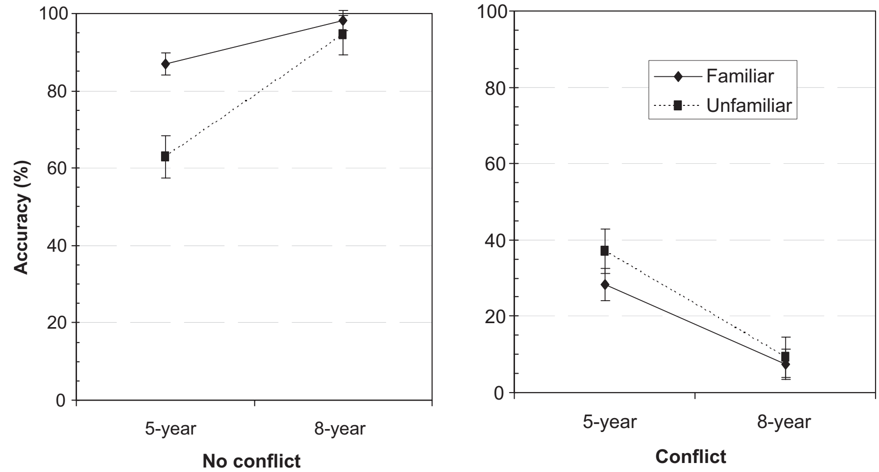


Figure 4. Mean accuracy for familiar and unfamiliar conflict and no-conflict problems in the two age groups. Error bars indicate standard errors.

There was a main effect of conflict, $F(1, 98) = 449.55$, $p < .0001$, $\eta_p^2 = .82$. As Figure 4 shows, the conflict problems were overall harder than the no-conflict problems. The main effect of age was not significant, $F(1, 98) > 1$, but age did interact with the conflict status of the problems, $F(1, 98) = 54.94$, $p < .001$, $\eta_p^2 = .36$. As expected, reasoning performance on the conflict problems decreased with age, whereas the older children outperformed the younger ones when heuristic thinking cued the correct response on the no-conflict problems. There was a also a significant two-way interaction between Conflict and Stereotype Familiarity factors, $F(1, 98) = 9.86$, $p < .0001$, $\eta_p^2 = .09$, and a significant three-way interaction among Age, Conflict, and Stereotype Familiarity factors, $F(1, 98) = 4.96$, $p < .05$, $\eta_p^2 = .05$. As Figure 4 indicates, the age effects were stronger for unfamiliar than for familiar problems. As expected, the developmental trends were less pronounced when 5-year-olds were familiar with the cued stereotype and heuristic thinking could help or bias their performance. Other factors and interactions were not significant.

Verbal justifications. After the children had selected a response, we asked them for a verbal justification. We were especially interested in justifications that referred to the cued stereotype (e.g., "because daddies usually mow the lawn," "because boys play with trucks") to validate the accuracy findings. We coded the number of times children in the two age groups gave such stereotype justifications on familiar and unfamiliar problems and subjected these to a 2 (age, between subjects) × 2 (stereotype familiarity, within subjects) ANOVA. If older children indeed start to rely more on stereotype knowledge when solving the tasks, we expected the number of stereotype responses to increase with age. Furthermore, if our classification of familiar and unfamiliar material was valid, we expected to see more stereotype justifications for familiar than for unfamiliar problems. As Figure 5 shows, these predictions were indeed confirmed. With increasing age there were more stereotype justifications, $F(1, 98) = 56.97$, $p < .0001$, $\eta_p^2 = .37$, and stereotype justifications were also more frequent for familiar problems, $F(1, 98) = 5.33$, $p < .025$, $\eta_p^2 =$

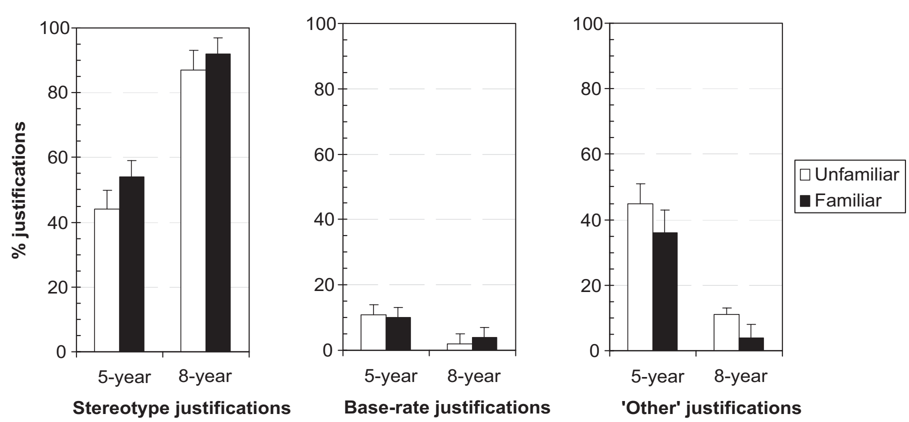


Figure 5. Mean percentage of stereotype (left panel), base-rate (middle panel), and other (right panel) justifications for familiar and unfamiliar problems in the two age groups. Error bars indicate standard errors.

.05. The interaction between the two factors was not significant, $F(1, 98) < 1$.

In an additional analysis, we looked for justifications in which children referred to the base-rates in the sample (e.g., "because there were more mommies than daddies"). For completeness, the average number of these base-rate responses was also entered in a 2 (age, between subjects) × 2 (stereotype familiarity, within subjects) ANOVA. None of the factors in the design reached significance. As Figure 5 shows, it is clear that both age groups rarely referred to the base-rates in their justifications. By itself, this is not surprising, given that even adults hardly explicitly refer to the base-rates in verbal protocols (e.g., De Neys & Glumicic, 2008).

It should be clear that we are well aware that asking for a verbal response justification is quite hard for young children and can be problematic. We observed, for example, that (especially with the preschoolers) there were a lot of justifications that could not be labeled as stereotype or base-rate responses. These are presented in Figure 5 as "other" responses.[2] The vast majority of these concerned trials in which children simply failed to give a proper verbal justification (e.g., "because I think so" or "I don't know"). In addition, the data might be affected by a competence-performance confound. That is, it might be that younger children are more familiar with the stereotypes than their verbal responses show and simply struggle to express their knowledge verbally. Hence, although the justification findings present some support for our claims, it is clear that the data need to be interpreted with caution.

Last, one can also use the verbal justification data to control for possible individual familiarity differences in our data. Although we validated the overall classification of the familiar and unfamiliar stereotypes, it is clear that for some individuals the classification might not hold. Obviously, if a 5-year-old would be familiar with stereotypes deemed to be unfamiliar or unfamiliar with the stereotypes deemed familiar, this would distort the reasoning findings. As one reviewer suggested, one can use the justifications as a strict filter to reduce the resulting noise. Hence, one can discard all responses on the reasoning task for which a participant's justification did not match the pilot classification (i.e., referring or not referring to the cued stereotype is taken as an index of familiarity). Although such an analysis is bound to be affected by the above mentioned limitations of verbal justification data, we included it for completeness. More specifically, we discarded all reasoning trials on familiar problems for which participants did not refer to the cued stereotype in their justification. In the group of 5-year-olds, we discarded all reasoning trials on unfamiliar problems for which participants did refer to the cued stereotype. Similarly, reasoning trials on unfamiliar problems for which 8-year olds did not mention the cued stereotype in their justification were also discarded. This filtering resulted in a discarding of a total of 26% of the reasoning trials. Table 1 shows the filtered results.

As Table 1 indicates, the age trends were magnified (e.g., 5-year olds' accuracy on the unfamiliar conflict problems reached 71%) but consistent with the overall analysis. This presents some further support for the validity of our pilot classifications.

Estimation question. After children had tried to give a verbal justification, we asked them to recall the group sizes in the problem. If children were simply not paying attention while the experimenter presented the problem, the children would fail to solve the base-rate problem, but this would not necessarily point to a lack of analytic or heuristic reasoning skills per se (e.g., Brainerd & Reyna, 1992a; De Neys & Van Gelder, 2009). The recall question allowed us to control for such a confound. If children were paying sufficient attention and processed the preambles, they should have little difficulty correctly estimating the base-rate information immediately after they solved the problem. We recorded whether children remembered the exact number information (e.g., "nine boys and one girl"), whether they at least correctly estimated which group was the largest (e.g., "more boys" or "eight boys and two girls"),[3] or whether they could not even correctly identify the largest group. The individual problem distributions within each age group were very similar, and consequently, we focused on the general pattern over the different problem types. Table 2 shows the results.

Table 1
Reasoning Accuracy After Individual Justification-Based Familiarity Control

	No conflict				Conflict			
	5 year		8 year		5 year		8 year	
Familiarity	*M*	*SE*	*M*	*SE*	*M*	*SE*	*M*	*SE*
Familiar %	100		100		3.2	3.2	0	
Unfamiliar %	49.3	9.3	100		70.8	9.5	0	

Table 2 clearly indicates that the problem information was properly processed. Children hardly ever failed to recall which one of the two groups was the largest. Eight-year-olds were better at recalling the exact number information, $t(98) = 3.63$, $p < .001$, $d = 0.67$, but even five-year-olds managed to give the exact number information for the vast majority of trials.

Control problem. On the abstract control problem, heuristic thinking could not bias or help analytic thinking. Solving the problem relies on mere analytic thinking about the group sizes. Thereby, the problem allowed us to check whether preschoolers had mastered the necessary analytic knowledge about the impact of group size on probability estimates. This was clearly the case. The vast majority of participants in both age groups managed to solve the problem correctly. Eighty percent ($SD = 40\%$) of the 5-year-olds solved the abstract problem correctly, whereas the 8-year-olds scored 76% ($SD = 43\%$). The performance difference between the two groups was not significant, $t(98) = 0.54$, $p = .60$, but the overall performance was clearly above chance level, $t(99) = 6.73$, $p < .0001$, $d = 0.68$. These data further establish that the bad performance on conflict problems should not be attributed

[2] The classification of the justifications into the "stereotype," "base-rate," or "other" category was very clear, however. When a second rater coded the justifications, interrater reliability reached .92.

[3] The criterion for the largest group response was any response that indicated 6 of 10 or better.

Table 2
Overview of the Base-Rate Estimation Performance in the Two Age Groups

	5 year		8 year	
Base-rate estimation	*M*	*SE*	*M*	*SE*
Exact number correct %	69.6	5.3	91.7	3
Largest group correct %	28.6	5.5	8.3	3
Error %	1.8	0.93	0	

to a lack of analytic knowledge per se but rather to the biasing impact of developing heuristic knowledge.[4]

An interesting additional observation in the present study is the contrast between the performance on the abstract problems and the unfamiliar no-conflict problems in our group of preschoolers. We noted above that 80% of the preschoolers solved the abstract problem correctly. As Figure 4 indicates, performance on the unfamiliar no-conflict problems reached only 63% (*SD* = 46%). This difference was significant, $F(1, 45) = 4.29$, $p < .05$, $\eta_p^2 = .09$. We argued that because the unfamiliar no-conflict problem is unfamiliar for preschoolers, they will need to rely on demanding analytic reasoning to solve the problem, whereas the older reasoners can rely on heuristic thinking. This helps to explain why older reasoners outperform the preschoolers. However, it is noteworthy that on the abstract problem, in which computing the correct base-rate response is also based on analytic reasoning, preschoolers' performance increased from 62% to 80%. Hence, preschoolers seem to underperform on the unfamiliar no-conflict problem (i.e., in theory, they could also score 80%). One possible explanation for this performance gap is that younger children's analytic reasoning is hindered by the mere presence of a context. For example, children might manage to reason analytically with abstract material such as yellow and blue cards, but once realistic objects (e.g., "cheese," "pizza," "principals") are mentioned, it might become harder to engage this analytic processing. Note that although preschoolers are not familiar with the crucial stereotypes (e.g., "Italians like pizza") they do know what, for example, pizza and Italians are, of course. It might be the case that reasoners are generally distracted by the superficial problem content of the more realistic material that we adopted (e.g., see Morris, 2000, for a related suggestion). Arguably, this will especially hamper performance when heuristic processing is not available (i.e., on the unfamiliar problems for preschoolers). Consequently, one could argue that preschoolers pay a double price on the unfamiliar no-conflict problems. On one hand, they cannot rely on heuristic processing and will be impaired when contrasted with older reasoners. On the other hand, engaging their analytic skills will be hindered by the presence of realistic content, resulting in an underperformance in comparison with the abstract problems. Clearly, these suggestions are speculative and remain to be validated. The abstract problem in the present study was only included to have a raw indication of participants' knowledge about the impact of group size on probability estimates. The high accuracy rates on the abstract problem do allow us to conclude that this principle is sufficiently mastered.

General Discussion

The present study focused on the conflicting developmental findings that have been reported in reasoning and decision-making studies. We designed a card game version of the infamous base-rate task that minimized methodological complications in previous studies. Results indicated that reasoning performance on the classic conflict problems indeed decreased with age: Five-year-old preschoolers selected the appropriate base-rate response more frequently than did 8-year-olds. However, as expected, results for the no-conflict problems showed the reversed age trend: Older children outperformed the younger ones when heuristic and analytic base-rate thinking cued the same response. These age effects were most pronounced for problems that adopted stereotypes that were unfamiliar for the 5-year-old preschoolers. When the preschoolers were familiar with the stereotypes, they also started being biased by heuristic thinking on the conflict problems or benefitting from it on the no-conflict problems. These findings clearly illustrate the role of development of the heuristic system on the reasoning performance. Consistent with previous claims (Kokis et al., 2002; Reyna & Brainerd, 1994), this suggests that the observed age related performance decrease on traditional conflicting reasoning problems needs to be attributed to the increased need to deal with tempting heuristics and not to a decrease in analytic thinking skills per se.

As we noted, our findings fit within the fuzzy-trace framework. The observed age trends are consistent with the core fuzzy-trace prediction concerning the age-related increased reliance on heuristic processing. One of the key advantages of the fuzzy trace framework over more classic dual process frameworks is that its proponents have explicitly tried to provide a processing specification of what heuristic and analytic processing entails (i.e., how heuristic and analytic processing works, e.g., Brainerd & Reyna, 2001). For example, fuzzy trace theory allows one to specify that the age-related increased need to deal with tempting heuristics that we refer to can be conceptualized as an increased reliance on gist-based processing. However, although the findings are consistent with fuzzy trace predictions, the present study was not designed to directly pit fuzzy trace theory against the more classic dual process theories. Such a comparison would require more specific operational measures of gist-based and verbatim processing (i.e., see Reyna, 2004, for examples). Combining these measures with the present design might help to provide more direct validation of the dual process assumptions and test more fine-grained predictions in future studies.

When considering the present results it is important to keep in mind that people's knowledge state is not purely dichotomous in nature. The distinction we made between familiar and unfamiliar problems is not absolute, of course. One might haven noted, for example, that although the age effect was less pronounced, reasoning performance on familiar conflict problems still decreased for the older age group. Likewise, reasoning performance on unfamiliar conflict problems in the youngest age group (in which stereotype knowledge was not supposed to bias reasoning) was

[4] In an additional control analysis, we only included participants who managed to solve the abstract problem correctly in the 2 (Age) × 2 (Conflict) × 2 (Familiarity) ANOVA on the reasoning data. However, the pattern of results was completely consistent with the overall analysis.

still far from perfect (whereas abstract problems were almost perfectly solved). These findings make sense if one takes into account that a problem will never be completely familiar or unfamiliar for a specific age group: Even preschoolers have some knowledge about the unfamiliar problems, whereas the familiar problems will not yet be as familiar as they are for 8-year-olds. Hence, on familiar conflict problems, older reasoners will still experience more heuristic bias than the younger ones, for example. Consequently, it is not surprising that solving these problems remains slightly harder for them than for the younger reasoners. The crucial point of the present familiarity manipulations is that they demonstrate that the extent of these age effects is indeed mediated by age-related differences in stereotype familiarity.

From a methodological point of view, our findings nicely illustrate the importance of taking heuristic development into account for reasoning and decision-making studies. As we noted, this factor has often been overlooked in previous developmental work, and this is bound to bias researchers' conclusions. To take but one specific case, recently Morsanyi and Handley (2008) observed that on a number of classic reasoning and decision-making tasks, performance decreased with children's cognitive capacity and age. Children who scored higher on classic cognitive (working memory) capacity tests were typically also more biased in the reasoning tasks (but see Brainerd & Reyna, 1992b, for a review of the distinction between correlations based on individual differences measures and those based on actual working memory capacity). Morsanyi and Handley suggested that heuristic search operations (e.g., the retrieval and accessing of stored stereotypes or background knowledge) that might be fully automated for adults would be demanding and require considerable working memory resources for younger children. Hence, their point was that only the most gifted children would manage to recruit stereotypical information, which would also make them more biased. Although we are quite sympathetic to this view (and made related suggestions, e.g., De Neys, 2007; De Neys & Everaerts, 2008), we want to clarify that there is an important confound in Morsanyi and Handley's work. Like most authors in the field, Morsanyi and Handley did not explicitly test the stereotype familiarity of their problems. Hence, it is possible that children with higher working memory capacity (who might live in culturally more enriched environments) simply have a larger or more developed knowledge base. Hence, the fact that children with larger working memory spans are also more biased does not necessarily imply that the heuristic operations require working memory capacity. It might simply be the case that children with larger spans are more familiar with the cued stereotypes or heuristic intuitions. Bluntly put, the smarter kids might simply know more. If less gifted children have not stored certain stereotypes yet, they can also not retrieve them, of course.

The present study is one of the first to systematically manipulate stereotype familiarity in developmental reasoning research. Although, the findings highlight the importance of stereotype knowledge development in children's reasoning, it is clear that the results will need to be further generalized in future studies. For example, we focused on a quite small age range in the present study (5-year-olds vs. 8-year-olds). We noted that the choice for this specific age range was motivated by our goal to test the impact of stereotype familiarity as early as possible. Now that the present study has demonstrated the impact of stereotype familiarity in the youngest age range, it would be interesting to contrast the performance of older age groups (e.g., adolescents vs. adults) to generalize the findings. Note that studies in these age ranges will benefit from the fact that older participants can be tested with the simple classic paper-and-pencil versions of the base-rate task.

At a more theoretical level, the present findings clearly argue against a unitary view on the development of reasoning and decision-making. Such a unitary view reflects the "illusion of replacement" (e.g., Brainerd & Reyna, 2001): The idea that with increasing age, one type of reasoning is replaced by another. This idea has dominated traditional reasoning and decision-making research in which cognitive development was implicitly conceived as a process whereby children's reasoning becomes less heuristic and more in line with logical standards. The developmental studies that pointed toward an age related increase in heuristic bias started cutting the ground under this view (e.g., Davidson, 1995; De Neys, 2007; Jacobs & Potenza, 1991; Morsanyi & Handley, 2008; Reyna & Ellis, 1994). However, it should be clear that these studies should not be interpreted as arguing for a reversed unitary view in which reasoning development is conceived as a process whereby analytic thinking is replaced by more contextualized, heuristic thinking. The question is not whether children's reasoning is becoming more analytic or more heuristic with age. As suggested previously (e.g., Davidson, 1995; Jacobs & Klaczynski, 2002; Reyna & Ellis, 1994), the present findings indicate that even at the preschool age, there is evidence for both the impact of more analytic and heuristic operations on the reasoning process. This underscores the idea that at every age, thinking should be conceived as an interaction between more intuitive and analytic processing (Kokis et al., 2002; Reyna, 2004). Not taking into account this dynamism is bound to bias any conclusion with respect to the development of the human reasoning engine.

References

Brainerd, C. J., & Reyna, V. F. (1992a). Explaining "memory free" reasoning. *Psychological Science, 3,* 332–339. doi:10.1111/j.1467-9280.1992.tb00042.x

Brainerd, C. J., & Reyna, V. F. (1992b). The memory independence effect: What do the data show? What do the theories claim? *Developmental Review, 12,* 164–186. doi:10.1016/0273-2297(92)90007-O

Brainerd, C. J., & Reyna, V. F. (2001). Fuzzy-trace theory: Dual processes in memory, reasoning, and cognitive neuroscience. *Advances in Child Development and Behavior, 28,* 49–100.

Cramer, P., & Steinwert, T. (1998). Thin is good, fat is bad: How early does it begin? *Journal of Applied Developmental Psychology, 19,* 429–451. doi:10.1016/S0193-3973(99)80049-5

Davidson, D. (1995). The representativeness heuristic and the conjunction fallacy effect in children's decision making. *Merrill-Palmer Quarterly, 41,* 328–346.

De Neys, W. (2006). Dual processing in reasoning: Two systems but one reasoner. *Psychological Science, 17,* 428–433. doi:10.1111/j.1467-9280.2006.01723.x

De Neys, W. (2007). Developmental trends in decision making: The case of the Monty Hall Dilemma. In J. A. Elsworth (Ed.), *Psychology of decision making in education, behavior, and high risk situations* (pp. 271–281). Hauppauge, NY: Nova Science.

De Neys, W., & Everaerts, D. (2008). Developmental trends in everyday conditional reasoning: The retrieval and inhibition interplay. *Journal of Experimental Child Psychology, 100,* 252–263. doi:10.1016/j.jecp.2008.03.003

De Neys, W., & Glumicic, T. (2008). Conflict monitoring in dual process theories of reasoning. *Cognition, 106,* 1248–1299. doi:10.1016/j.cognition.2007.06.002

De Neys, W., & Van Gelder, E. (2009). Logic and belief across the life span: The rise and fall of belief inhibition during syllogistic reasoning. *Developmental Science, 12,* 123–130. doi:10.1111/j.1467-7687.2008.00746.x

De Neys, W., Vartanian, O., & Goel, V. (2008). Smarter than we think: When our brains detect that we are biased. *Psychological Science, 19,* 483–489. doi:10.1111/j.1467-9280.2008.02113.x

Epstein, S. (1994). Integration of the cognitive and psychodynamic unconscious. *American Psychologist, 49,* 709–724. doi:10.1037/0003-066X.49.8.709

Evans, J. St. B. T. (2003). In two minds: Dual-process accounts of reasoning. *Trends in Cognitive Sciences, 7,* 454–459. doi:10.1016/j.tics.2003.08.012

Evans, J. St. B. T. (2008). Dual-processing accounts of reasoning, judgment and social cognition. *Annual Review of Psychology, 59,* 255–278. doi:10.1146/annurev.psych.59.103006.093629

Gigerenzer, G., Hell, W., & Blank, H. (1988). Presentation and content: The use of base-rates as a continuous variable. *Journal of Experimental Psychology: Human Perception and Performance, 14,* 513–525. doi:10.1037/0096-1523.14.3.513

Goel, V., & Dolan, R. J. (2003). Explaining modulation of reasoning by belief. *Cognition, 87,* B11–B22. doi:10.1016/S0010-0277(02)00185-3

Houdé, O. (1997). Rationality in reasoning: The problem of deductive competence and the inhibitory control of cognition. *Cahiers de Psychologie Cognitive/Current Psychology of Cognition, 16,* 108–113.

Houdé, O., & Tzourio-Mazoyer, N. (2003). Neural foundations of logical and mathematical cognition. *Nature Reviews Neuroscience, 4,* 507–514. doi:10.1038/nrn1117

Jacobs, J. E., & Klaczynski, P. A. (2002). The development of judgment and decision making during childhood and adolescence. *Current Directions in Psychological Science, 11,* 145–149.

Jacobs, J. E., & Potenza, M. (1991). The use of judgment heuristics to make social and object decisions: A developmental perspective. *Child Development, 62,* 166–178.

Kahneman, D. (2002, December). *Maps of bounded rationality: A perspective on intuitive judgment and choice. Nobel Prize lecture.* Retrieved from http://nobelprize.org/nobel_prizes/economics/laureates/2002/kahnemann-lecture.pdf

Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgment under uncertainty: Heuristics and biases.* Cambridge, MA: Cambridge University Press.

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review, 80,* 237–251. doi:10.1037/h0034747

Klaczynski, P. A. (2000). Is rationality really "bounded" by information processing constraints? *Behavioral and Brain Sciences, 23,* 683–684. doi:10.1017/S0140525X00413439

Kokis, J. V., Macpherson, R., Toplak, M. E., West, R. F., & Stanovich, K. E. (2002). Heuristic and analytic processing: Age trends and associations with cognitive ability and cognitive styles. *Journal of Experimental Child Psychology, 83,* 26–52. doi:10.1016/S0022-0965(02)00121-2

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. doi:10.1037/0012-1649.36.6.741

Morsanyi, K., & Handley, S. J. (2008). How smart do you need to be to get it wrong? The role of cognitive capacity in the development of heuristic-based judgment. *Journal of Experimental Child Psychology, 99,* 18–36. doi:10.1016/j.jecp.2007.08.003

Reyna, V. F. (2000). Data, development, and dual processes in rationality. *Behavioral and Brain Sciences, 23,* 694–695. doi:10.1017/S0140525X0054343X

Reyna, V. F. (2004). How people make decisions that involve risk: A dual processes approach. *Psychological Science, 13,* 60–66.

Reyna, V. F., & Brainerd, C. J. (1994). The origins of probability judgment: A review of data and theories. In G. Wright & P. Ayton (Eds.), *Subjective probability* (pp. 239–272). New York, NY: Wiley.

Reyna, V. F., & Ellis, S. (1994). Fuzzy-trace theory and framing effects in children's risky decision making. *Psychological Science, 5,* 275–279. doi:10.1111/j.1467-9280.1994.tb00625.x

Reyna, V. F., & Mills, B. A. (2007). Converging evidence supports fuzzy-trace theory's nested sets hypothesis (but not the frequency hypothesis). *Behavioral and Brain Sciences, 30,* 278–280. doi:10.1017/S0140525X07001872

Reyna, V. F., Mills, B. A., Estrada, S. M., & Brainerd, C. J. (2006). False memory in children: Data, theory, and legal implications. In M. P. Toglia, J. D. Read, D. F. Ross, & R. C. L. Lindsay (Eds.), *The handbook of eyewitness psychology: Memory for events* (pp. 473–510). Mahwah, NJ: Erlbaum.

Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin, 119,* 3–22. doi:10.1037/0033-2909.119.1.3

Stanovich, K. E., & West, R. F. (2000). Individual differences in reasoning: Implications for the rationality debate. *Behavioral and Brain Sciences, 23,* 645–665. doi:10.1017/S0140525X00003435

Williams, J. E., Bennett, S. M., & Best, D. L. (1975). Awareness and expression of sex stereotypes in young children. *Developmental Psychology, 11,* 635–642. doi:10.1037/0012-1649.11.5.635

Received July 28, 2009
Revision received July 5, 2010
Accepted August 13, 2010 ■