

The development of logical intuitions: Task-specific automatization in mathematical reasoning

Esther Boissin^{1*}, Laura Charbit², Elora Taieb², Grégoire Borst², Wim De Neys²

¹ Department of Psychology, Cornell University, Ithaca, USA

² Université Paris Cité, CNRS, LaPsyDÉ, Paris, France

*Corresponding author: Esther BOISSIN

boissinesther@gmail.com,

Department of Psychology, Cornell University, 211 Tower Rd, Ithaca, NY, United States, 14850

In press at Journal of Cognition and Development,

October 2026

Abstract

Recent reasoning theories propose that people can intuitively process elementary logico-mathematical principles. These controversial “logical intuitions” are believed to develop through a learning process where reasoners gradually automatize their application. To test this automatization hypothesis, we compared the reasoning performance of younger (5th-grade, ~10 years old) and older (7th-grade, ~12 years old) French students on mathematical reasoning tasks (Study 1: decimal comparison, 105 5th graders and 103 7th graders; Study 2: arithmetic word problems, 207 5th graders and 103 7th graders) that have been well-practiced by 7th graders. We used a two-response paradigm in which participants first provided an intuitive response under time pressure and/or cognitive load, followed by a more deliberate response. The results were mixed. In the decimal comparison task, 7th graders were more likely than 5th graders to produce correct intuitive responses without needing deliberative correction. However, this advantage did not extend to the arithmetic word problems, where 7th graders did not outperform 5th graders, either at the intuitive or deliberate stage. Overall, these findings provide partial support for the developmental automatization hypothesis and point to its potential task-dependent nature.

Keywords: logical intuition, numerical reasoning, adolescent

Introduction

Human reasoning is often described as biased. People frequently rely on quick, effortless impressions that can conflict with logical or probabilistic principles (Evans, 2008; Kahneman, 2011). Dual-process theories explain such errors by distinguishing fast, automatic Type 1 processes—grounded in subjective impressions—from slower, effortful Type 2 processes that implement logico-mathematical principles (Evans & Stanovich, 2013). In this view, accuracy typically requires deliberate correction of an initial, compelling response.

However, recent findings challenge this assumption and suggest that intuition is not inherently flawed. In that respect, the two-response paradigm (Thompson et al., 2011), which distinguishes between fast, intuitive answers and slower, deliberative ones, provides direct support for the effectiveness of intuition. In this paradigm, participants are asked to provide two answers to each problem: an initial response under time pressure and cognitive load (to constrain the engagement of deliberation), followed by a second, unconstrained response (to allow for deliberation, Bago & De Neys, 2017). Across a range of reasoning tasks, most correct final answers were already correct at the initial, intuitive stage (Bago & De Neys, 2017, 2019, 2020; Raelison, Keime, et al., 2021). These findings directly challenge the assumption that intuition is the main source of reasoning errors and that deliberation is required for correct responding. Instead, they show that intuitive processes can produce correct responses. This pattern invites a shift in emphasis: rather than treating intuition as the default source of error, it becomes necessary to specify when intuitive processes yield correct responses, and why. Therefore, these findings point to two central questions: under what conditions do correct logical intuitions emerge, and what mechanisms support their development?

One proposal is that accurate intuitions emerge through automatization. Stanovich (2011) described mindware as the rules, principles, and strategies that individuals acquire through learning and practice. Building on this idea, De Neys and Pennycook (2019) proposed that repeated engagement with logico-mathematical principles allows this mindware to become instantiated and, with sufficient practice, automatized. As a result, once-effortful reasoning principles can be retrieved and applied intuitively, without deliberate reflection. When the necessary knowledge has been well practiced, it can operate automatically and guide reasoning from the outset, yielding correct responses that are immediate and effortless.

Crucially, multiple intuitions can co-occur, accordingly different candidate responses may be activated in parallel, with the final output reflecting the one that is strongest or most accessible. When the relevant logico-mathematical knowledge is well entrenched, the logical intuition may be activated rapidly and with high accessibility, increasing its chances of dominating over competing incorrect intuitions. In contrast, when logico-mathematical

knowledge is less entrenched, incorrect responses should be more likely to prevail. Accordingly, the automatization account entails that correct intuitive responses should be more frequent when the underlying principles have been practiced extensively and can be retrieved automatically.

Preliminary evidence is consistent with this idea. Charbit et al. (2025) and Raoelison, Boissin, et al. (2021) found that 12th-grade students were more likely than 7th graders to produce correct intuitive answers on classic reasoning problems such as the bat-and-ball task (“A bat and a ball together cost \$1.20. The bat costs \$1 more than the ball. How much does the ball cost?”; Frederick, 2005). Together, these findings suggest that sound intuitive reasoning becomes more prevalent by the end of secondary education in late adolescence, potentially reflecting increased exposure and practice with the underlying logico-mathematical principles.

At the same time, these results leave open whether age differences truly reflect the effect of automatization or a broader developmental limitation. That is, the precise nature of the less prevalent sound intuiting among young adolescents is not clear. One possibility is that in general, younger reasoners simply have a less efficient “System 1” and are overall less able to generate accurate intuitive responses. Alternatively, the poorer performance may originate from the nature of the tasks typically used in prior dual-process research. Many of the classic reasoning tasks—such as the bat-and-ball problem—draw on principles that are only introduced and practiced during secondary education (e.g., solving an equation with two unknowns, Hoover & Healy, 2017). Younger students are therefore unlikely to have automatized these principles, making correct intuitive responses improbable. From this perspective, the developmental pattern may not indicate a lack of intuitive reasoning capacity but rather limited opportunity to acquire and automatize the relevant knowledge. Distinguishing between these explanations is crucial for understanding how logical intuitions develop. If automatization is the key mechanism, correct intuitions should appear even in younger students—provided the task relies on well-learned, highly practiced principles. Conversely, if a genuine developmental constraint limits the ability to generate correct intuitions, younger students should show poor intuitive performance regardless of how well the underlying principle has been practiced.

In a recent paper, Boissin, Charbit, Taieb, et al. (2026) provided evidence for the automatization account by manipulating task familiarity—defined as how well the underlying logico-mathematical principle has been practiced—as a proxy for automatization, while holding age constant. Across two studies with 7th graders (12-13 years old), they contrasted performance on numerical tasks that rely on logico-mathematical principles extensively practiced by the start of secondary school (Study 1: arithmetic word problems, Riley et al., 1983; Study 2: decimal comparison, Desmet et al., 2010) with a less-practiced task drawing on

principles typically not yet well instantiated at that age (the bat-and-ball problem). They used the two-response paradigm to differentiate intuitive from deliberative responses. The key pattern was that accuracy was substantially higher on the well-practiced tasks than on the less-practiced bat-and-ball task both for intuitive and deliberative responses, and—critically—correct responses on the well-practiced tasks predominantly appeared intuitively. In contrast, performance on the less-practiced bat-and-ball task remained near floor even after deliberation. This suggests that correct intuitive reasoning depends on prior knowledge instantiation and directly argues against a generic “System 1” developmental deficit.

The present study extends this logic one step further. If correct intuitive responding on the “well-practiced” tasks indeed reflects automatization, then it should also be graded as a function of exposure within the earlier developmental window in which these principles are acquired and consolidated. That is, younger elementary schoolers should have had less practice on the very tasks that were well instantiated for the 7th graders Boissin, Charbit, Taieb, et al. (2026) tested. Therefore, we compared 5th (10-11 years old) and 7th (12-13 years old) French graders on the same numerical tasks used in Boissin, Charbit, Taieb, et al. (2026), testing whether 5th graders show reduced—intuitive—accuracy on these tasks.

Study 1 focused on a decimal magnitude comparison task and Study 2 on arithmetic word problems. The decimal comparison task assessed students’ understanding of decimal magnitudes, a core topic typically introduced and practiced during elementary education (Desmet et al., 2010). A well-documented bias in this domain is the “whole number bias,” whereby students assume that numbers with more digits are larger. For example, they may judge 1.45 to be greater than 1.5 simply because 45 is numerically greater than 5 (Roell et al., 2017). The arithmetic word problems required students to apply basic operations and interpret relational statements in verbal form—skills introduced and reinforced throughout primary school (Lubin et al., 2015; Verschaffel, 1994). A typical error is to add rather than subtract when misled by the phrasing “more than.” For instance, in “Mary has 8 marbles. She has 5 more marbles than John. How many marbles does John have?”, students often respond 13 instead of 3 (Riley et al., 1983).

To distinguish intuitive from deliberate answers, we used a two-response paradigm in both studies. We predicted that 7th graders would show higher intuitive and deliberative accuracies than 5th graders, consistent with greater automatization of the underlying logico-mathematical principles.

Study 1

Methods

Pre-registration

The design and research questions were preregistered on the AsPredicted website (<https://aspredicted.org>) and stored on the Open Science Framework (https://osf.io/3m9zj/overview?view_only=f939f58c6a3148c19e6cdac29a36bf7e) where all data and material can be accessed. No specific statistical analyses were pre-registered.

Participants

In total, we recruited 208 French pupils; 105 5th graders (40 women and 10 neutral-gender, M age = 10.54 years, SEM = 0.05 years) and 103 7th graders (42 women and 5 neutral-gender, M age = 12.46 years, SEM = 0.06 years). We preregistered a target sample of at least 100 students per grade. However, because the experiment was conducted in classrooms, we tested and included all students who had signed up. A sensitivity analysis indicated that our sample size would allow us to detect small to medium size accuracy differences (effect size $d_z = .46$) between the grade groups with 95% power.

The 5th graders were recruited from elementary schools and the 7th graders from middle schools, as French students transition to middle school in Grade 6. All schools were public schools located in the Île-de-France region (greater Paris area). Socioeconomic status was not measured at the individual level; instead, school selection relied on the social position index (Indice de Position Sociale, IPS) published for each school by the French Ministry of Education, a school-level indicator of students' average socioeconomic background. Schools were selected so that this index was comparable across cohorts: none belonged to the national priority-education networks (REP/REP+), which identify schools serving socioeconomically disadvantaged areas. Note also that French middle schools follow a common national curriculum, with no ability-based streaming of course content. Testing took place in the students' regular classrooms, either with the full class (up to approximately 30 students) or in half-class groups (approximately 15), under the supervision of at least one teacher and one experimenter. All students who provided assent and whose parents provided written informed consent were included; the research objectives were explained to families, and confidentiality in data handling was assured.

Materials

Decimal comparison problems. Participants were presented with decimal comparison problems adapted from (Roell et al., 2019). On each trial, a pair of decimal numbers

was shown, and participants were asked to select the larger one. The smaller number always had more digits after the decimal point—for example, “0.326 vs. 0.6.” In such cases, the common heuristic error is to choose “0.326,” based on the incorrect belief that longer decimals represent larger values. However, the correct answer is “0.6.” Note that in the French curriculum, decimal numbers are introduced in Grade 4 (around age 9; Ministère de l'Éducation nationale, 2020).

The decimal numbers included one, two, or three digits after the decimal point. Task difficulty was systematically varied by manipulating the numerical distance between the first digits following the decimal point, with distances ranging from 1 to 6.

Half of the problems were featured in their standard “conflict” version in which the intuitively cued “heuristic” response cues an answer that conflicts with the correct answer (e.g., “0.6 vs 0.326”). To ensure that participants were engaged in the task, we also presented problems which were featured in their “no-conflict” version, and which were used as control problems. The control consisted of pairs of decimal numbers in which the largest decimal had the greatest number of digits after the decimal point (e.g., “0.644 vs. 0.3”). Here the correct response is “0.644.” Two sets of problems were used in order to counterbalance problem content: across participants, each item pair was presented in both its conflict and its no-conflict version, while no participant saw both versions of the same pair. The two sets had the same digit-length composition and were matched on numerical distance, and performance did not differ by set (Supplementary Material, Section S1). The presentation order of the conflict and no-conflict problems was randomized in each set. Participants were randomly assigned to one of the two sets for and were presented in total with 10 conflict and 10 no-conflict items. All the problems are presented in Supplementary Material.

Two-response format (and development). Participants responded to each problem using a two-response procedure, where they first provided a ‘fast’ answer, directly followed by a second ‘slow’ answer (Thompson et al., 2011). This method allowed us to capture both an initial “intuitive” response and then a final “deliberate” one. To minimize the possibility that deliberation was involved in producing the initial ‘fast’ response, participants had to provide their initial answer within a strict time limit (see Bago & De Neys, 2017, 2019, 2020; Raelison & De Neys, 2019). The deadline for the 7th graders was set at 1.5 seconds based on a pilot study with 33 7th-grade French pupils (M age = 12.19 years, SD = 0.47 years; 16 women, 4 neutral gender students). In this pilot, students gave correct responses with an average response time of 1.53 seconds (SEM = 0.10) on conflict items. We decided to set the absolute deadline threshold to 1.5 seconds (and not lower) to create time-pressure while keeping the number of missed trials acceptable (see Trial Exclusions). For 5th graders, the deadline was set at 2 seconds.

To ensure that 7th graders provided their most intuitive responses, we further constrained the initial stage by asking them to perform a concurrent cognitive load task, similar to the paradigm used for adults (Bago & De Neys, 2017, 2019, 2020; Raelison & De Neys, 2019). The load task was based on the dot memorization task (Miyake et al., 2001) given that it had been successfully used to burden executive resources during reasoning tasks (e.g., De Neys, 2006). Participants had to memorize a complex visual pattern (i.e., 4 crosses in a 3x3 grid) presented briefly before each reasoning problem. After their initial (intuitive) response to the problem, participants were shown four different patterns and had to identify the one that they had memorized (see Bago & De Neys, 2019, for more details).

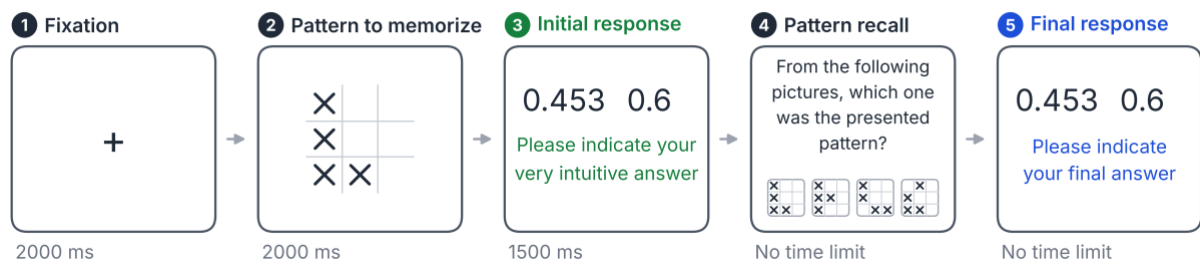
Note that pilot testing indicated that the concurrent load was too demanding for 5th graders, and we therefore omitted it in this group. Accordingly, only the time-pressure constraint was applied at the initial response stage for 5th graders, whereas both time pressure and concurrent load were applied at the initial response stage for 7th graders.

Procedure

The experiment was programmed in Qualtrics (Qualtrics, Provo, UT) and run individually on iPads. The participants were tested in their classrooms, in groups, under the supervision of at least one teacher. Participants were instructed that the experiment would take 15 minutes and that it demanded their full attention. Participants were instructed to determine which of the two decimal numbers was the largest. First, a general description of the task was presented in which participants were instructed that they would need to solve decimal comparison, for which they would have to provide two consecutive responses. They were told that we were interested in their very first, initial answer that came to mind and that—after providing their initial response—they could reflect on the item and take as much time as they needed to provide a final answer. The full task instructions are provided in the Supplementary Material Section S2. In order to familiarize themselves with the two-response procedure, they first solved two unrelated practice problems. Next, only the 7th graders familiarized themselves with the cognitive load procedure by solving two load trials and, finally, they solved two items which included both cognitive load and the two-response procedure. Figure 1 shows a typical trial for the decimal comparison task which started with a fixation cross for 2000ms, followed by the visual matrix for the cognitive-load task for 2000ms that is only presented to the 7th graders and followed by the pair of decimals (e.g., “0.453 vs. 0.6”) where participants had up to 1500 ms (7th graders) or 2000 ms (5th graders) to select their initial response. If they had not provided an answer before the time limit, they were given a reminder that it was important to provide an answer within the time limit on subsequent trials. Then, 7th graders were presented with four visual matrices and had to choose the one that they had previously memorized. They received

feedback as to whether their memory-response was correct. If the answer was not correct, they were reminded that it was important to perform well on the memory task on subsequent trials. Finally, the same reasoning problem was presented again, and participants were asked to provide a final deliberate answer (with no time limit). At the end of the study, participants were asked to complete a page with demographic questions.

7th graders



5th graders

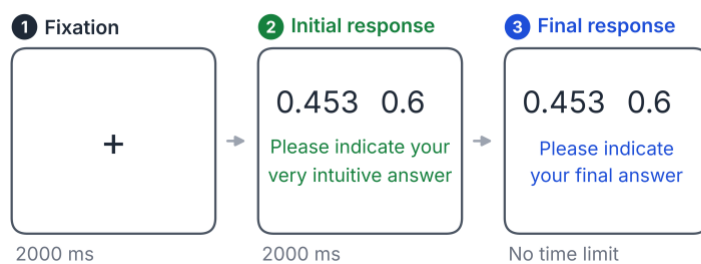


Figure 1. Time course of a complete two-response trial for the decimal comparison task in Study 1, for both 7th and 5th graders (who were presented with a less stringent two-response paradigm).

Trial exclusion

We discarded trials in which 5th graders failed to provide their initial response before the deadline (13.5% of all trials), and we analyzed the remaining 86.5%. For the 7th graders, we discarded trials in which participants either failed to provide their initial response before the deadline (24.1% of all trials) or failed to pick the correct matrix in the load task (18.0% of the remaining trials), resulting in the analysis of 62.2% of all decimal comparison trials.

On average, for the decimal comparison problems, each 5th grader contributed 8.55 (SEM = 0.15) conflict trials and 8.75 (SEM = 0.14) no-conflict trials out of 10. For the 7th graders, each participant contributed an average of 5.99 (SEM = 0.22) conflict trials and 6.5 (SEM = 0.22) no-conflict trials out of 10.

Results

Response accuracy. We first analyzed the effect of grade on initial and final conflict accuracies separately by comparing performance between 5th and 7th graders. For each participant and each response stage, we calculated the average proportion of correct responses for conflict problems for each group and then contrasted these averages using a t-test.

Figure 2 shows that initial intuitive accuracy was higher for the 7th graders ($M = 59.70\%$, $SEM = 3.49\%$) than for the 5th graders ($M = 38.87\%$, $SEM = 3.37\%$), $t(205) = 4.298$, $p < .001$, *Cohen's d* = 0.598. This suggests that 7th graders were better able to respond intuitively to the decimal comparison task.

Figure 2 also shows the same pattern for accuracy at the final deliberative response stage, which was higher for the 7th graders ($M = 69.71\%$, $SEM = 3.43\%$) than for the 5th graders ($M = 48.04\%$, $SEM = 3.88\%$), $t(205) = 4.176$, $p < .001$, *Cohen's d* = 0.581. This indicates that when able to engage in deliberation, 7th graders also outperform 5th graders.

As expected, for the no-conflict control problems, we observed that performance was high, both for the 5th graders (M initial response = 78.54% , $SEM = 2.78\%$, M final response = 82.19% , $SEM = 3.16\%$) and for the 7th graders (M initial response = 86.68% , $SEM = 2.35\%$, M final response = 89.34% , $SEM = 2.53\%$). This high performance indicates that participants managed to process the problems and were not randomly guessing.

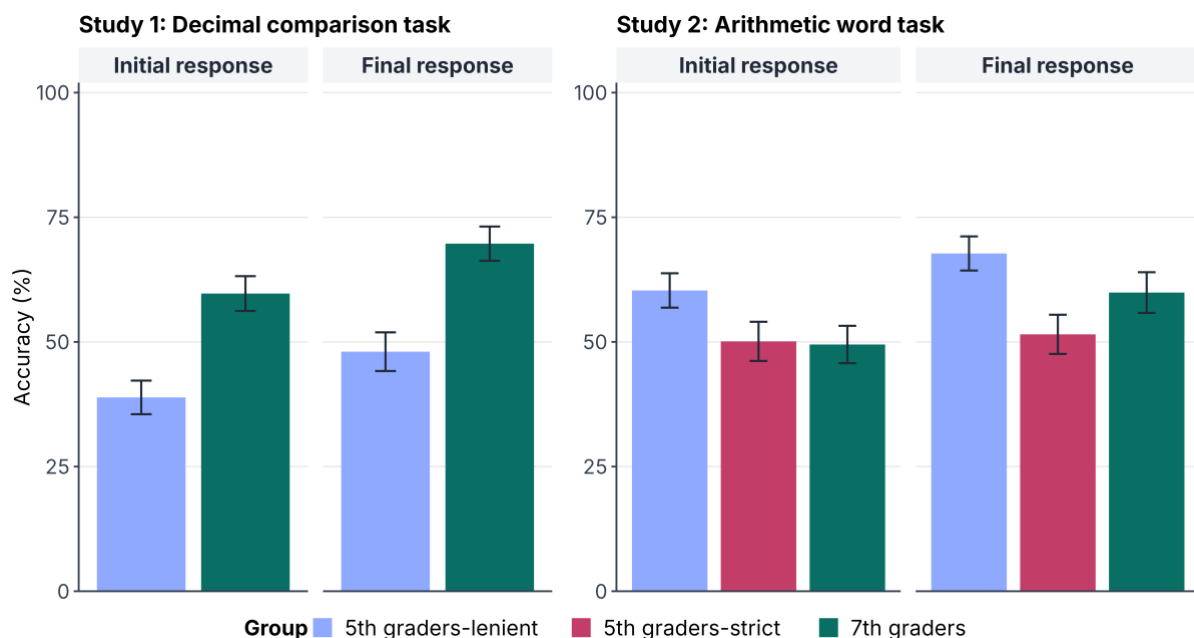


Figure 2. Average initial and final accuracy on conflict problems for the decimal comparison task in Study 1 for both 5th and 7th graders, and for the arithmetic word task in Study 2 for 5th graders-lenient, 5th graders-strict and 7th graders. Error bars represent standard error of the mean (SEM).

Direction of change. To better understand how participants changed (or did not change) their responses after deliberation, we performed a direction of change analysis for the conflict items (Bago & De Neys, 2017) for each grade. Especially, each trial is composed of two responses, the initial 'intuitive' one (with time and/or load constraints) and the final 'deliberate' one. Correct responses are labeled '1' and incorrect responses are labeled '0'. Hence, each trial can result in one of these four different patterns: "00" pattern, incorrect response at both response stages; "11" pattern, correct response at both response stages; "01" pattern, initial incorrect and final correct responses; "10" pattern, initial correct and final incorrect responses. Figure 3 shows the direction of change distribution for 5th and 7th graders.

As shown in Figure 3, overall, 7th graders predominantly provided correct responses, whether these responses were given immediately at the intuitive stage (i.e., "11" response patterns: $M = 50.88\%$, $SEM = 3.46\%$) or only at the deliberative stage (i.e., "01" response patterns: $M = 18.83\%$, $SEM = 2.18\%$), rather than biased "00" response patterns ($M = 21.47\%$, $SEM = 3.45\%$). Critically, when correct responses were given, the "11" response patterns were more frequent than the "01" ones, $t(101) = 6.91$, $p < .001$, *Cohen's d* = 0.68. This indicates that correct responses primarily arise at the intuitive stage and do not necessarily require further deliberation.

Similarly, when 5th graders provided a correct response, they mostly did so at the intuitive stage, as indicated by the higher proportion of "11" response patterns ($M = 33.30\%$, $SEM = 3.19\%$) compared to "01" response patterns ($M = 14.54\%$, $SEM = 1.68\%$), $t(104) = 5.71$, $p < .001$, *Cohen's d* = 0.56. However, 5th graders provided fewer "11" responses than 7th graders, $t(203.632) = -3.68$, $p < .001$, *Cohen's d* = -0.51, and gave a predominant proportion of biased responses ($M = 46.66\%$, $SEM = 4.00\%$). This indicates that although 5th graders can produce correct intuitive responses on the decimal comparison task, such responses were more frequent among 7th graders.

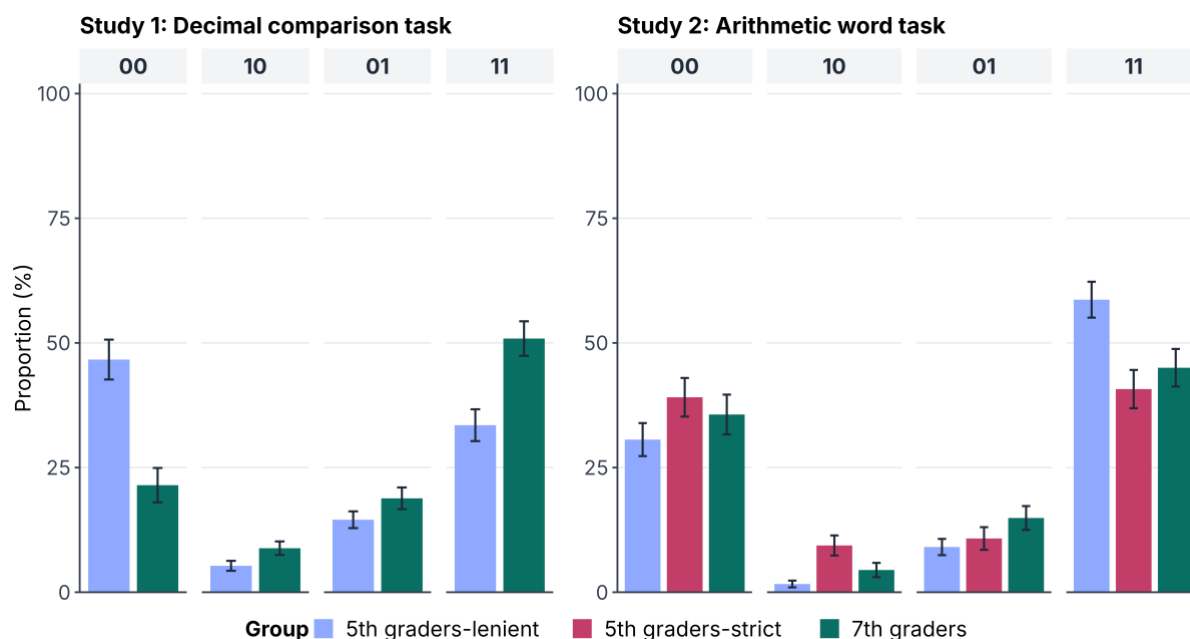


Figure 3. Proportion of each direction of change (i.e., 00, 10, 01, and 11 response patterns) for the conflict problems in the decimal comparison task in Study 1 between 5th and 7th graders, and for the arithmetic word task in Study 2 between the 5th graders-lenient, 5th graders-strict, and 7th graders. Error bars represent the standard error of the mean (SEM).

Response times. Response times were recorded at both response stages. Means and SEMs were computed on log-transformed values and back-transformed for readability. On conflict items, 5th graders responded within $M = 1.13$ s ($SEM = 0.03$) at the initial stage and $M = 0.91$ s ($SEM = 0.06$) at the final stage; 7th graders responded within $M = 1.18$ s ($SEM = 0.02$) and $M = 1.46$ s ($SEM = 0.04$), respectively. A detailed analysis of response times falls outside the scope of the present paper; response times broken down by direction of change pattern, together with exploratory latency comparisons, are reported in the Supplementary Material, Section S3 for the interested reader.

Study 2

In Study 2, we aimed to replicate and generalize the findings of Study 1 with a different set of logico-mathematical problems, focusing on arithmetic word problems. These problems measured the ability to apply mathematical operations, skills typically learned in primary school (Lubin et al., 2015; Verschaffel, 1994).

Pilot testing suggested that the full two-response paradigm for the arithmetic word problems could be overly demanding for some students. We therefore first administered a less stringent version for the 5th graders: participants completed the initial response under time pressure only (no concurrent memory load), and the deadline was less strict than for the 7th

graders. However, as reported below, the 5th graders showed very high accuracy—exceeding the 7th graders' performance.

To rule out the possibility that this discrepancy reflected the less stringent initial-stage constraints, we collected additional data from a separate sample of 5th graders using the full two-response paradigm and the same deadline as for 7th graders. For clarity, we report results from all samples together, referring to 5th graders tested under the less stringent paradigm as the *5th graders-lenient* and those tested under the full paradigm as the *5th graders-strict*.

Methods

Pre-registration

The design and research questions were preregistered on the AsPredicted website (<https://aspredicted.org>) and stored on the Open Science Framework (https://osf.io/3m9zj/overview?view_only=f939f58c6a3148c19e6cdac29a36bf7e) where all data and material can be accessed. No specific statistical analyses were pre-registered.

Participants

In total, we recruited 204 French pupils, 101 in 5th grade (51 women and 7 neutral-gender, M age = 9.76 years, SEM = 0.06 years), and 103 in 7th grade (61 women and 4 neutral-gender, M age = 12.21 years, SEM = 0.07 years). We also recruited 106 French 5th graders who were assigned to the strict condition (51 women and 4 neutral-gender, M age = 9.64 years, SEM = 0.07 years). We preregistered a target sample of at least 100 students per grade. However, because the experiment was conducted in classrooms, we tested and included all students who had signed up. A sensitivity analysis indicated that our sample size would allow us to detect small to medium size accuracy differences (effect size $d_z = .46$) between the groups (i.e., between the 7th graders and 5th graders-lenient, and between the 7th graders and the 5th graders-strict) with 95% power.

The 5th graders (including the 5th graders-strict) were recruited from public elementary schools and the 7th graders from public and private middle schools, all located in the Île-de-France region. School selection followed the same criteria as in Study 1: comparable school-level social position indices (IPS), no schools from the priority-education networks (REP/REP+). Testing took place in the students' regular classrooms, either with the full class (up to approximately 30 students) or in half-class groups (approximately 15), under the supervision of at least one teacher and one experimenter. All students who provided assent and whose parents provided written informed consent were included.

Materials

Arithmetic word problems. Participants were presented with French arithmetic word problems taken from Lubin et al. (2013). Each problem was presented sequentially as a set of three sentences: (a) the number of objects of one protagonist (e.g., "Mary has 30 marbles"), (b) the number of objects one protagonist had in excess relative to the other protagonist (e.g., "She has 5 more marbles than John") and (c) the problem's question (e.g., "How many marbles does John have?"). All the operations involved numbers that were all multiples of 5. Note that in the French curriculum, arithmetic word problems involving addition and subtraction are introduced from Grade 1 (around age 6; Ministère de l'Éducation nationale, 2020).

Participants had to select a response among four response choices which were composed of (1) the correct response (i.e., "25") obtained by subtracting the second quantity from the first, (2) the intuitively cued "heuristic" response (i.e., "35") that consists in the addition of both number of objects, (3) a foil option which was the sum of correct and heuristic responses (i.e., "60"), and (4) a second foil option which was the second greatest common divisor multiple of 5 (i.e., "5"). The four response choices appeared in a random order. For example, the arithmetic word problem stated:

- (a) *Mary has 30 marbles.*
- (b) *She has 5 more marbles than John.*
- (c) *How many marbles does John have?*
 - 25
 - 35
 - 60
 - 5

Half of the problems were featured in their standard "conflict" version in which the intuitively cued "heuristic" response cues an answer that conflicts with the correct answer. To ensure that participants were engaged in the task, we also presented problems which were featured in their "no-conflict" version, and which were used as control problems. In these control problems, we changed the protagonist for whom the participants must count the marbles. Thus, for the no-conflict items, applying an addition cues the correct answer (Lubin et al., 2013), for instance:

- (a) *Mary has 30 marbles.*
- (b) *John has 5 more marbles than her.*
- (c) *How many marbles does John have?*
 - 35
 - 20
 - 55
 - 10

In this case the intuitively cued "35" answer is correct. The four response options for the no-conflict items were: (1) the correct response ("35"), based on the addition of the two quantities; (2) a first foil ("20"), corresponding to half the correct response, rounded to the nearest multiple of 5; (3) a second foil ("55"), obtained by summing the correct response and the first foil; and (4) a third foil ("10"), corresponding to the largest multiple of 5 shared by foils 1 and 2, excluding their values.

Two sets of problems were used in order to counterbalance problem content. The presentation order of the conflict and no-conflict problems was randomized in each set. Participants were randomly assigned to one of the two sets for each block and were presented in total with five conflict and five no-conflict items.

Two-response format (and development). Similar to Study 1, students responded to the arithmetic word task using a two-response procedure. For both the 5th graders-strict and 7th graders, the deadline during the initial stage was set at 5 seconds as in Boissin, Charbit, Taieb, et al. (2026). This cutoff was based on the deadline used in the two-response paradigm for problems with similar sentence structure and word length (e.g., the bat-and-ball problem, Bago & De Neys, 2019). To put it in perspective, Lubin et al. (2013) showed that 9th graders solve arithmetic word problems in an average of 7.34 seconds, indicating that our 5-second deadline is sufficiently challenging for 7th graders. As we explained, for the 5th graders-lenient, the initial deadline during the initial response stage was set at 8 seconds to relax the two-response demands.

To further minimize the possibility that deliberation was involved in producing the initial 'fast' response, participants from the 5th graders-strict and 7th grade groups had to provide their initial answer within a strict time limit while performing a concurrent cognitive load task (see Bago & De Neys, 2017, 2019; Raoelison & De Neys, 2019). For the load task, participants had to memorize a complex visual pattern (i.e., 4 crosses in a 3x3 grid) presented briefly before each reasoning problem. After their initial (intuitive) response to the problem, participants were shown four different patterns and had to identify the one they had memorized (see Bago & De Neys, 2019, for more details).

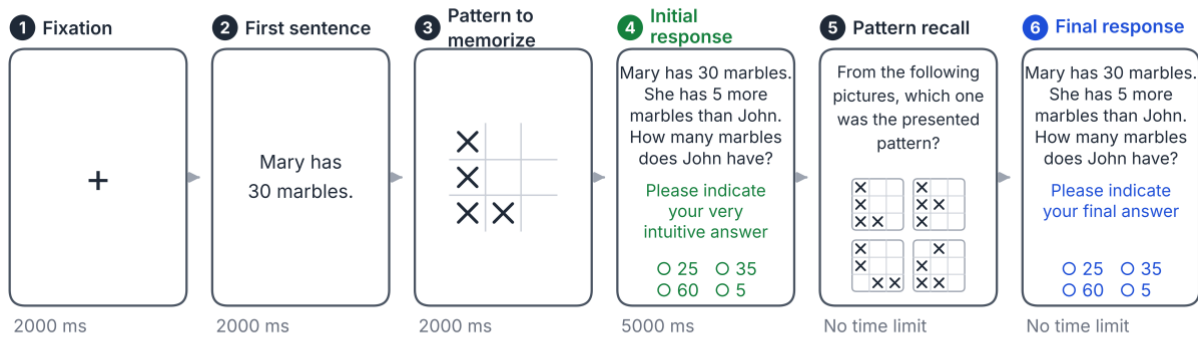
Finally, after completing the initial response stage, all participants were asked to respond to the same exact problem in the final response stage, but this time without a time limit or further constraints. The time limit and cognitive load were applied only for the initial response and not for the final one, where participants were allowed to deliberate (see below).

Procedure

The experiment was programmed in Qualtrics (Qualtrics, Provo, UT) and run individually on iPads. The participants were tested in their classrooms, in groups, under the supervision of at least one teacher. Participants were instructed that the experiment would take fifteen minutes and that it demanded their full attention.

First, a general description of the task was presented in which participants were instructed that they would need to solve reasoning problems, for which they would have to provide two consecutive responses. They were told that we were interested in their very first, initial answer that comes to mind and that—after providing their initial response—they could reflect on the problem and take as much time as they needed to provide a final answer. The full task instructions are provided in the Supplementary Material Section S2. In order to familiarize themselves with the two-response procedure, they first solved two unrelated practice problems. Next, 5th graders-strict and 7th graders familiarized themselves with the cognitive load procedure by solving two load trials and, finally, they solved two problems which included both cognitive load and the two-response procedure. Figure 4 shows a typical trial both for the less constraining two-response paradigm—presented to the 5th graders-lenient—and the full traditional one—presented to the 5th graders-strict and 7th graders. Both paradigms started with a fixation cross for 2000ms, followed by the first sentence of the problem (e.g., "Mary has 30 marbles.") for 2000ms. In the traditional two-response paradigm, the first sentence was followed by the visual matrix for the cognitive-load task for 2000ms. In both paradigms, the full problem was presented afterwards, at which point 5th graders-lenient had 8000ms and 5th graders-strict and 7th graders had 5000ms, to select their initial answer among the four choices. After 7000ms for 5th graders-lenient and 4000ms for the 5th graders-strict and 7th graders, the background of the screen turned yellow to warn participants that they only had a short amount of time left to answer. If they had not provided an answer before the time limit, they were given a reminder that it was important to provide an answer within the time limit on subsequent trials. Then, the 5th graders-strict and 7th graders were presented with four visual matrices and had to choose the one that they had previously memorized. They received feedback as to whether their memory-response was correct. If the answer was not correct, they were reminded that it was important to perform well on the memory task on subsequent trials. Finally, the same reasoning problem was presented again, and all participants were asked to provide a final deliberate answer (with no time limit). At the end of the study, participants were asked to complete a page with demographic questions.

7th graders and 5th graders-strict



5th graders-lenient

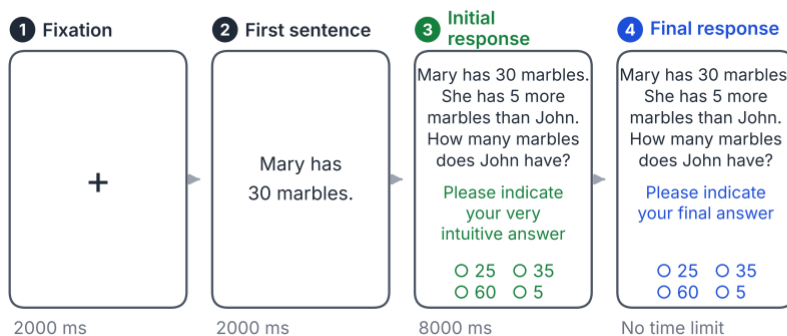


Figure 4. Time course of a complete two-response trial for the arithmetic word task in Study 2, for 7th, 5th graders-strict and 5th graders-lenient (who were presented with a less stringent two-response paradigm).

Trial exclusion

We discarded trials in which 5th graders-lenient failed to provide their initial response before the deadline (15.9% of all trials), analyzing the remaining 84.1%. For 7th graders, we discarded trials in which participants either failed to provide their initial response before the deadline (21.0% of all trials) or failed to pick the correct matrix in the load task (25.1% of the remaining trials), resulting in the analysis of 59.2% of all arithmetic word trials. For the 5th graders-strict, we discarded trials in which participants either failed to provide their initial response before the deadline (32.5% of all trials) or failed to pick the correct matrix in the load task (36.2% of the remaining trials), resulting in the analysis of 43.0% of all trials. We had pre-registered that if we removed more than 50% of all trials of the 5th graders-strict, we would not discard trials where children failed to pick the correct matrix in the load task. Consequently, we analyzed 67.5% of all trials for the 5th graders-strict.

On average, 5th graders-lenient contributed 4.11 (SEM = 0.11) conflict trials and 4.30 (SEM = 0.11) no-conflict trials out of 5. 7th graders contributed an average of 3.01 (SEM = 0.14) conflict trials and 3.16 (SEM = 0.15) no-conflict trials out of 5. 5th graders-strict

contributed an average of 3.58 (SEM = 0.14) conflict trials and 3.16 (SEM = 0.15) no-conflict trials out of 5.

Results

Response accuracy. We first analyzed the effect of grade on initial and final conflict accuracies separately by comparing performance first, between 5th graders-lenient vs 7th graders, second between 5th graders-strict vs 7th graders and third between 5th graders-strict and 5th graders-lenient. For each participant and each response stage, we calculated the average proportion of correct responses for conflict problems for each group and then contrasted these averages using a t-test.

Contrary to Study 1, the findings from Study 2 were less clear. As shown in Figure 2, for the critical initial intuitive responses, the 5th graders-lenient outperformed the 7th graders, $t(199) = 2.123$, $p = .035$, *Cohen's d* = 0.299, whereas the 5th graders-strict and the 7th graders did not differ, $t(202) = 0.117$, $p = .907$, *Cohen's d* = 0.016. The difference between the 5th graders-lenient and the 5th graders-strict approached significance, $t(199) = 1.947$, $p = .053$, *Cohen's d* = 0.275.

Accuracy at the final deliberative response stage was relatively high for the 5th graders-lenient (M = 67.74%, SEM = 3.42%), the 5th graders-strict (M = 51.52%, SEM = 3.93%), and the 7th graders (M = 59.90%, SEM = 4.08%). Final accuracy did not differ significantly between the 5th graders-lenient and the 7th graders, $t(199) = 1.469$, $p = .143$, *Cohen's d* = 0.207, nor between the 5th graders-strict and the 7th graders, $t(202) = 1.480$, $p = .140$, *Cohen's d* = 0.207. The final accuracy of the 5th graders-lenient was significantly higher than that of the 5th graders-strict, $t(199) = 3.110$, $p = .002$, *Cohen's d* = 0.439. Overall, these findings indicate that 7th graders were not better at solving arithmetic word problems than the 5th-grade groups, either at the intuitive or the deliberative stage.

Direction of change. Figure 3 plots the direction of change for Study 2. Overall, the 7th graders gave more “11” response patterns (M = 45.02%, SEM = 3.77%) than “01” response patterns (M = 14.89%, SEM = 2.39%), $t(101) = 6.260$, $p < .001$, *Cohen's d* = 0.620. The same pattern was observed for the 5th graders-lenient (M “11” = 58.67%, SEM = 3.60%; M “01” = 9.07%, SEM = 1.63%), $t(98) = 11.236$, $p < .001$, *Cohen's d* = 1.129, and for the 5th graders-strict (M “11” = 40.74%, SEM = 3.84%; M “01” = 10.78%, SEM = 2.27%), $t(101) = 6.069$, $p < .001$, *Cohen's d* = 0.601. This suggests that when children give correct responses to arithmetic problems, they mostly do so intuitively without the need for further deliberation.

Aligned with the results for accuracy, the group that gave the most “11” responses was the 5th graders-lenient (5th graders-lenient vs. 7th graders: $t(198.797) = 2.621$, $p = .009$, *Cohen's*

$d = 0.369$; 5th graders-lenient vs. 5th graders-strict: $t(198.515) = 3.411$, $p < .001$, *Cohen's d* = 0.481). The 7th graders did not give significantly more “11” responses than the 5th graders-strict, $t(201.938) = 0.796$, $p = .427$, *Cohen's d* = 0.111.1

Response times. On conflict items, 5th graders-lenient responded within $M = 4.55$ s ($SEM = 0.13$) at the initial stage and $M = 3.06$ s ($SEM = 0.36$) at the final stage; 5th graders-strict within $M = 3.50$ s ($SEM = 0.07$) and $M = 5.86$ s ($SEM = 0.40$); and 7th graders within $M = 3.22$ s ($SEM = 0.08$) and $M = 5.67$ s ($SEM = 0.28$), respectively. As in Study 1, response times broken down by direction of change pattern, together with exploratory latency comparisons, are reported in the Supplementary Material, Section S3.

Discussion

The present study tested whether automatization of logico-mathematical principles supports accurate intuitive reasoning. Using the two-response paradigm, we compared French 5th and 7th graders on two reasoning tasks—decimal comparison in Study 1 and arithmetic word problems in Study 2—both of which rely on logico-mathematical principles introduced and practiced in primary education. If automatization supports the emergence of correct intuitions, older students should be more likely to respond correctly at the intuitive stage.

Study 1 supported this prediction. Seventh graders outperformed 5th graders on both initial and final responses in the decimal comparison task, with the age difference particularly pronounced at the intuitive stage. This pattern is consistent with the view that, with additional exposure and practice, the relevant logico-mathematical principles become more accessible and can generate correct responses as soon as the intuitive stage. Importantly, this advantage emerged even though the initial stage was more constraining for 7th graders, who responded under both time pressure and concurrent cognitive load, whereas 5th graders responded under time pressure alone (and with a slightly longer deadline). If anything, these more lenient initial-stage constraints should have inflated 5th graders' intuitive accuracy—by leaving more room for ‘possible residual deliberation’—yet 5th graders still produced fewer correct initial responses. The procedural asymmetry therefore cannot account for the observed grade difference; rather, it implies that our estimate of the developmental effect is conservative.

Study 2, by contrast, did not yield a robust Grade 7 advantage on arithmetic word problems when initial-stage constraints were matched (i.e., relative to the 5th graders-strict), and the less constrained 5th graders-lenient even exceeded 7th-graders' performance. Moreover, 7th graders also did not outperform 5th graders-lenient at the final stage when they were allowed to deliberate. Under matched initial-stage constraints, Study 2 does not show a reliable advantage for 7th graders.

One speculative explanation for the unexpected Study 2 findings is that developmental gains on this class of arithmetic word problems largely occur before late primary school. Lubin et al. (2015) report relatively high accuracy across Grades 3–5 (about 78% overall), consistent with an early rise followed by attenuation of age-related gains. On this account, differences between 5th graders-lenient and 7th graders may be modest because the relevant principle is already substantially instantiated by Grade 5, leaving limited room for detectable grade-related change. It is noteworthy that arithmetic word problems are introduced earlier in the French curriculum than decimal numbers (i.e., Grade 1 vs. Grade 4), which tentatively suggests that our 5th graders had relatively more experience with word problems than with decimal comparisons (albeit less than 7th graders in absolute terms). Clearly, this does not imply that development is complete by Grade 5. Indeed, accuracy in Study 2 remained well below ceiling across groups and stages, leaving substantial room for further improvement. The more specific implication is that continued learning and automatization may not translate into a smooth, grade-by-grade increase in correct intuitive responding on this class of arithmetic word problems: developmental change can unfold in a non-linear way, with periods of slowed or near-stable performance before subsequent gains are observed at later points. This broader pattern is consistent with evidence that the development of arithmetic fluency is uneven across grades and operations (Gliksman et al., 2022), and with the more general observation that skill acquisition often proceeds through plateaus followed by later leaps (Gray & Lindstedt, 2017). Notably, such plateaus are not limited to computational fluency: periods of little grade-related change have also been documented for children's understanding of basic arithmetic principles (Robinson & Dubé, 2009). Extended periods of near-stable performance thus appear to be a recurrent feature of mathematical development. Within this perspective, the absence of a robust Grade 7 advantage in Study 2 is compatible with a temporary stabilization across this age window: automatization may still be progressing, but its behavioral expression may advance in steps rather than increasing monotonically with grade level. Obviously, this post hoc explanation remains speculative. Naturally, a follow-up would be to test younger children (below Grade 5), where exposure to these principles is lower and larger developmental differences may be more likely to emerge. Unfortunately, the current two-response design may prove too challenging at that age and this approach would call for the development and validation of an alternative constraints paradigm.

A complementary explanation is that accuracy on arithmetic word problems does not depend on the arithmetic principle alone. Word problem difficulty is co-determined by linguistic factors: solving these problems requires constructing an adequate representation of the situation described in the text, and performance varies with the phrasing (Daroczy et al., 2015) and format (Koedinger & Nathan, 2004) of the problem over and above the arithmetic it requires. Indeed, our conflict problems instantiate the classic “inconsistent language” structure,

in which the relational term (“more than”) cues the inappropriate operation (Lewis & Mayer, 1987). Accuracy on these problems therefore reflects not only the automatization of the arithmetic principle but also the comprehension processes required to apply it. If these comprehension processes do not substantially change between Grades 5 and 7—or develop at a different pace than the arithmetic principle itself—grade-related gains in automatization may fail to translate into detectable accuracy differences.

More generally, the divergence between Study 1 and Study 2 suggests that developmental change in correct intuitive responding may be task-dependent within primary-school mathematics. In the decimal comparison task, we observed a clear Grade 7 advantage at the initial stage, whereas in the arithmetic word problems this advantage was attenuated or absent. At minimum, this indicates that the developmental signature of automatization does not generalize uniformly across tasks. One possibility is that these tasks differ in when developmental gains emerge and whether performance shows a plateau within the Grade 5–7 window: for some problems, additional experience may still translate into measurable improvements in early correct responding, whereas for others, performance may stabilize over this period despite remaining below ceiling. Importantly, the present data do not allow strong claims about which task is “more automatized” overall; rather, they suggest that the timing and shape of developmental change (e.g., gradual gains vs. temporary plateaus) may differ across task types. A comparable unevenness is well documented in children's mathematical development: at a given age, children master some mathematical principles substantially better than others, and different principles progress at different rates (Robinson & Dubé, 2009). Clarifying what task properties predict these different automatization trajectories remains an open question.

Nevertheless, taken together, the present findings converge with the broader claim that correct intuitive responding can emerge with increased practice and knowledge instantiation. Replicating Boissin, Charbit, Taieb, et al. (2026), 7th graders produced correct responses already at the intuitive stage in both studies—something that previous research had not observed on less-practiced problems such as the bat-and-ball task (Charbit et al., 2025; Raelison, Boissin, et al., 2021). This contrast indicates that early adolescents are not inherently unable to generate correct intuitions; rather, intuitive accuracy depends on whether the relevant principles have become sufficiently automatized to guide responding rapidly. These results thus support the view that logical intuitions can develop through learning and automatization. Training evidence converges with this view. Boissin, Charbit, Raelison, et al. (2026) found that the benefit of an intuitive debiasing training on classic reasoning problems increased with age, and that this benefit was already visible at the initial, intuitive response stage.

Limitations

Finally, it remains important to consider individual variability in the ability to automatize logico-mathematical principles. Individual differences in mathematical development are well documented at two levels: children differ reliably in their conceptual grasp of mathematical principles and in their ability to carry out the corresponding computations, and the two do not always develop in step (Gilmore & Papadatou-Pastou, 2009). The present framework points to a third potential locus of individual differences: the degree to which acquired principles become automatized and can guide responding intuitively. Some individuals may have a greater propensity for automatization, and understanding what facilitates or moderates this process remains largely unexplored. While education undoubtedly plays a key role, we still lack a full understanding of how automatization develops over time. Boissin et al. (2024) demonstrated that even individuals from the Himba of Namibia, who lack formal education and urban-based informal learning, can be trained to reason effectively with logico-mathematical principles. This suggests that education—formal and informal—alone may not fully account for differences in reasoning performance and potentially in the automatization process.

Cognitive ability is one plausible contributor to these developmental differences. Charbit et al. (2025) report that cognitive ability predicts performance on the bat-and-ball task among older adolescents (12th graders), but not among younger students (7th graders). In adults, cognitive ability has also been reported to relate more strongly to intuitive accuracy on the same task (Raoelison et al., 2020). Together, these findings suggest that the association between cognitive ability and performance may become more apparent once the relevant principle is sufficiently automatized—an interpretation that remains to be tested directly.

Dispositional factors are another plausible contributor. Actively open-minded thinking, for instance, has been linked to higher intuitive accuracy and greater responsiveness to debiasing interventions (Boissin & Pennycook, 2025). Recent work suggests that individual-difference factors such as cognitive ability and thinking dispositions may matter in part because they are associated with the acquisition, instantiation, and eventual automatization of relevant mindware (Burič, 2023; Burič & Šrol, 2020; Erceg et al., 2023; Šrol & De Neys, 2021; Stanovich, 2018). Although the precise mechanisms remain unclear, these findings raise the possibility that intuitive accuracy reflects not only exposure and practice, but also stable individual differences that shape who benefits most from them. Future work should test this more directly by crossing variation in task automatization—for example, using grade level as a proxy for degree of automatization—with cognitive and dispositional measures.

From an educational perspective, the findings from Study 1 suggest that sustained practice with logico-mathematical principles can strengthen their automatization, leading to more accurate and efficient intuitive reasoning. This highlights the value of repeated exposure

and practice for developing intuitive problem-solving skills. At the same time, Study 2 cautions against assuming a smooth, grade-by-grade increase in spontaneous correct responding across all problem types. A key implication is therefore to treat the developmental trajectory of “intuitive” success as potentially uneven and task-dependent, and to specify which learning conditions lead to correct intuitive responses. Identifying those conditions remains an important goal for future work.

In the same vein, our two studies compared different principles embedded in different task formats. Because performance on the same mathematical content can vary substantially with its surface format (Koedinger & Nathan, 2004), and because what learners appear to know about a given principle depends on the context and type of assessment used (Prather & Alibali, 2009), differences between the decimal comparison task and the arithmetic word problems cannot be unambiguously attributed to the underlying principles themselves. Future studies should explore a broader range of tasks—ideally holding the principle constant while varying the surface format—to better understand the task-specific factors that influence the ease of automatization. Recognizing this task specificity cautions against overgeneralizing the automatization hypothesis across different reasoning tasks and raises critical questions about the boundaries of automatization. By examining a wider spectrum of tasks, future research could provide a more nuanced understanding of how practice and exposure shape intuitive reasoning performance across various problem types.

In conclusion, this study provides mixed evidence for the automatization hypothesis. While 7th graders demonstrated greater intuitive reasoning in decimal comparison tasks, this advantage did not extend to arithmetic word problems. These findings emphasize the potential task-dependent nature of automatization and highlight the need to consider the specificity of different reasoning problems when evaluating how intuitive responses emerge from practice.

References

- Bago, B., & De Neys, W. (2017). Fast logic?: Examining the time course assumption of dual process theory. *Cognition*, 158, 90–109. <https://doi.org/10.1016/j.cognition.2016.10.014>
- Bago, B., & De Neys, W. (2019). The Smart System 1: Evidence for the intuitive nature of correct responding on the bat-and-ball problem. *Thinking & Reasoning*, 25(3), 257–299. <https://doi.org/10.1080/13546783.2018.1507949>

- Bago, B., & De Neys, W. (2020). Advancing the specification of dual process models of higher cognition: A critical test of the hybrid model view. *Thinking & Reasoning*, 26(1), 1–30. <https://doi.org/10.1080/13546783.2018.1552194>
- Boissin, E., Charbit, L., Raoelison, M., Borst, G., Caparos, S., & De Neys, W. (2026). Reactivating logic? Age-dependent effects of intuitive debiasing training [In press]. *Thinking & Reasoning*. <https://doi.org/10.1080/13546783.2026.2727940>
- Boissin, E., Charbit, L., Taieb, E., Borst, G., & De Neys, W. (2026). Developing System 1: knowledge instantiation drives sound intuiting [In press]. *Trends in Neuroscience and Education*, 44, 100295. <https://doi.org/10.1016/j.tine.2026.100295>
- Boissin, E., Josserand, M., De Neys, W., & Caparos, S. (2024). Debiasing thinking among non-WEIRD reasoners. *Cognition*, 243, 105681. <https://doi.org/10.1016/j.cognition.2023.105681>
- Boissin, E., & Pennycook, G. (2025). Who benefits from debiasing? *Cognition*, 262, 106166. <https://doi.org/10.1016/j.cognition.2025.106166>
- Burič, R. (2023). Acquired Knowledge and Bias Susceptibility: Mindware Automatization Measured with A Two-Response Paradigm and Its Relationship with Conflict Detection. *Studia Psychologica*, 65(4), Article 4. <https://doi.org/10.31577/sp.2023.04.883>
- Burič, R., & Šrol, J. (2020). Individual differences in logical intuitions on reasoning problems presented under two-response paradigm. *Journal of Cognitive Psychology*, 32(4), 460–477. <https://doi.org/10.1080/20445911.2020.1766472>
- Charbit, L., Boissin, E., Raoelison, M., & De Neys, W. (2025). Emergence of the smart intuitor: How cognitive ability shapes adolescent reasoning. *Thinking & Reasoning*, 1–22. <https://doi.org/10.1080/13546783.2025.2527672>
- Daroczy, G., Wolska, M., Meurers, W. D., & Nuerk, H.-C. (2015). Word problems: A review of linguistic and numerical factors contributing to their difficulty. *Frontiers in Psychology*, 6, 348. <https://doi.org/10.3389/fpsyg.2015.00348>
- De Neys, W. (2006). Automatic–Heuristic and Executive–Analytic Processing during Reasoning: Chronometric and Dual-Task Considerations. *Quarterly Journal of*

Experimental Psychology, 59(6), 1070–1100.

<https://doi.org/10.1080/02724980543000123>

De Neys, W. (2021). On dual- and single-process models of thinking. *Perspectives on Psychological Science*, 16(6), 1412–1427. <https://doi.org/10.1177/1745691620964172>

De Neys, W., & Pennycook, G. (2019). Logic, Fast and Slow: Advances in Dual-Process Theorizing. *Current Directions in Psychological Science*, 28(5), 503–509. <https://doi.org/10.1177/0963721419855658>

Desmet, L., Grégoire, J., & Mussolin, C. (2010). Developmental changes in the comparison of decimal fractions. *Learning and Instruction*, 20(6), 521–532. <https://doi.org/10.1016/j.learninstruc.2009.07.004>

Erceg, N., Galić, Z., Bubić, A., & Jelić, D. (2023). Who detects and why: How do individual differences in cognitive characteristics underpin different types of responses to reasoning tasks? *Thinking & Reasoning*, 29(4), 594–642. <https://doi.org/10.1080/13546783.2022.2108897>

Evans, J. St. B. T. (2008). Dual-Processing Accounts of Reasoning, Judgment, and Social Cognition. *Annual Review of Psychology*, 59(1), 255–278. <https://doi.org/10.1146/annurev.psych.59.103006.093629>

Evans, J. St. B. T., & Stanovich, K. E. (2013). Dual-Process Theories of Higher Cognition: Advancing the Debate. *Perspectives on Psychological Science*, 8(3), 223–241. <https://doi.org/10.1177/1745691612460685>

Frederick, S. (2005). Cognitive Reflection and Decision Making. *Journal of Economic Perspectives*, 19(4), 25–42. <https://doi.org/10.1257/089533005775196732>

Gilmore, C. K., & Papadatou-Pastou, M. (2009). Patterns of individual differences in conceptual understanding and arithmetical skill: A meta-analysis. *Mathematical Thinking and Learning*, 11(1–2), 25–40. <https://doi.org/10.1080/10986060802583923>

Gliksman, Y., Berebbi, S., & Henik, A. (2022). Math Fluency during Primary School. *Brain Sciences*, 12(3), 371. <https://doi.org/10.3390/brainsci12030371>

- Gray, W. D., & Lindstedt, J. K. (2017). Plateaus, Dips, and Leaps: Where to Look for Inventions and Discoveries During Skilled Performance. *Cognitive Science*, 41(7), 1838–1870. <https://doi.org/10.1111/cogs.12412>
- Hoover, J., & Healy, A. (2017). Algebraic reasoning and bat-and-ball problem variants: Solving isomorphic algebra first facilitates problem solving later. *Psychonomic Bulletin & Review*, 24(6), 1922–1928. <https://doi.org/10.3758/s13423-017-1241-8>
- Kahneman, D. (2011). *Thinking, Fast and Slow*. Farrar, Straus and Giroux.
- Koedinger, K. R., & Nathan, M. J. (2004). The real story behind story problems: Effects of representations on quantitative reasoning. *Journal of the Learning Sciences*, 13(2), 129–164. https://doi.org/10.1207/s15327809jls1302_1
- Lewis, A. B., & Mayer, R. E. (1987). Students' miscomprehension of relational statements in arithmetic word problems. *Journal of Educational Psychology*, 79(4), 363–371. <https://doi.org/10.1037/0022-0663.79.4.363>
- Lubin, A., Houdé, O., & De Neys, W. (2015). Evidence for children's error sensitivity during arithmetic word problem solving. *Learning and Instruction*, 40, 1–8. <https://doi.org/10.1016/j.learninstruc.2015.07.005>
- Lubin, A., Vidal, J., Lanoë, C., Houdé, O., & Borst, G. (2013). Inhibitory control is needed for the resolution of arithmetic word problems: A developmental negative priming study. *Journal of Educational Psychology*, 105(3), 701–708. <https://doi.org/10.1037/a0032625>
- Ministère de l'Éducation nationale, de la Jeunesse et des Sports. (2020). *Programmes d'enseignement de l'école et du collège* (Bulletin officiel n° 31 du 30 juillet 2020).
- Miyake, A., Friedman, N. P., Rettinger, D. A., Shah, P., & Hegarty, M. (2001). How are visuospatial working memory, executive functioning, and spatial abilities related? A latent-variable analysis. *Journal of Experimental Psychology: General*, 130(4), 621–640. <https://doi.org/10.1037/0096-3445.130.4.621>

- Prather, R. W., & Alibali, M. W. (2009). The development of arithmetic principle knowledge: How do we know what learners know? *Developmental Review*, 29(4), 221–248.
<https://doi.org/10.1016/j.dr.2009.09.001>
- Raoelison, M., Boissin, E., Borst, G., & De Neys, W. (2021). From slow to fast logic: The development of logical intuitions. *Thinking & Reasoning*, 27(4), 599–622.
<https://doi.org/10.1080/13546783.2021.1885488>
- Raoelison, M., & De Neys, W. (2019). Do we de-bias ourselves?: The impact of repeated presentation on the bat-and-ball problem. *Judgment and Decision making*, 14(2), 170–178. <https://doi.org/10.1017/S1930297500003405>
- Raoelison, M., Keime, M., & De Neys, W. (2021). Think slow, then fast: Does repeated deliberation boost correct intuitive responding? *Memory & Cognition*, 49(5), 873–883.
<https://doi.org/10.3758/s13421-021-01140-x>
- Raoelison, M., Thompson, V. A., & De Neys, W. (2020). The smart intuitor: Cognitive capacity predicts intuitive rather than deliberate thinking. *Cognition*, 204, 104381.
<https://doi.org/10.1016/j.cognition.2020.104381>
- Riley, M. S., Greeno, J. G., & Heller, J. I. (1983). Development of children's problem-solving ability in arithmetic. In H. P. Ginsburg (Ed.), *The development of mathematical thinking* (pp. 153–196). Academic Press.
- Robinson, K. M., & Dubé, A. K. (2009). Children's understanding of addition and subtraction concepts. *Journal of Experimental Child Psychology*, 103(4), 532–545.
<https://doi.org/10.1016/j.jecp.2008.12.002>
- Roell, M., Viarouge, A., Houdé, O., & Borst, G. (2017). Inhibitory control and decimal number comparison in school-aged children. *PLOS ONE*, 12(11), e0188276.
<https://doi.org/10.1371/journal.pone.0188276>
- Roell, M., Viarouge, A., Houdé, O., & Borst, G. (2019). Inhibition of the whole number bias in decimal number comparison: A developmental negative priming study. *Journal of Experimental Child Psychology*, 177, 240–247.
<https://doi.org/10.1016/j.jecp.2018.08.010>

- Šrol, J., & De Neys, W. (2021). Predicting individual differences in conflict detection and bias susceptibility during reasoning. *Thinking & Reasoning*, 27(1), 38–68.
<https://doi.org/10.1080/13546783.2019.1708793>
- Stanovich, K. E. (2011). *Rationality and the Reflective Mind*. Oxford University Press, USA.
- Stanovich, K. E. (2018). Miserliness in human cognition: The interaction of detection, override and mindware. *Thinking & Reasoning*, 24(4), 423–444.
<https://doi.org/10.1080/13546783.2018.1459314>
- Thompson, V. A., Prowse Turner, J. A., & Pennycook, G. (2011). Intuition, reason, and metacognition. *Cognitive Psychology*, 63(3), 107–140.
<https://doi.org/10.1016/j.cogpsych.2011.06.001>
- Verschaffel, L. (1994). Using retelling data to study elementary school children's representations and solutions of compare problems. *Journal for Research in Mathematics Education*, 25(2), 141–165. <https://doi.org/10.2307/749506>

Supplementary Material

S1. Materials and Item-level analyses

Study 1: Decimal comparison task

Table S1. Decimal comparison problems from Study 1's practice.

Left number	Right number
0.2	0.789
0.975	0.83

Table S2. Decimal comparison problems from Study 1. Set A is shown in bold and Set B in plain text.

Item number	Conflict version (C)		No-conflict version (NC)	
	Left number	Right number	Left number	Right number
1	0.67	0.9	0.6	0.97
2	0.8	0.36	0.86	0.3
3	0.232	0.3	0.2	0.332
4	0.7	0.613	0.713	0.6
5	0.184	0.82	0.18	0.824
6	0.41	0.295	0.415	0.29
7	0.548	0.8	0.5	0.848
8	0.2	0.177	0.277	0.1
9	0.35	0.6	0.3	0.65

10	0.5	0.24	0.54	0.2
11	0.9	0.42	0.92	0.4
12	0.81	0.9	0.8	0.91
13	0.3	0.231	0.331	0.2
14	0.326	0.6	0.3	0.626
15	0.48	0.191	0.481	0.19
16	0.528	0.74	0.52	0.748
17	0.8	0.745	0.845	0.7
18	0.175	0.3	0.1	0.375
19	0.5	0.34	0.54	0.3
20	0.24	0.8	0.2	0.84

The 20 item pairs of Table S2 were divided into two halves (pairs 1–10 and pairs 11–20). Participants assigned to Set A solved the conflict versions of pairs 1–10 and the no-conflict versions of pairs 11–20; participants assigned to Set B solved the reverse. Across participants, each pair thus contributed both its conflict and its no-conflict version, while no participant saw both versions of the same pair.

The two halves had the same digit-length composition: each contained four 1-digit vs. 2-digit pairs, four 1-digit vs. 3-digit pairs, and two 2-digit vs. 3-digit pairs. They were also closely matched on numerical distance, both for the conflict versions (Set A: $M = 0.24$; Set B: $M = 0.23$; $t(18) = 0.06$, $p = .954$) and the no-conflict versions ($M = 0.29$ vs. 0.32 , $t(18) = 0.32$, $p = .753$). Note that, by construction, the no-conflict versions involve larger numerical distances than the conflict versions ($M = 0.30$ vs. 0.23 overall); since the no-conflict items serve as control items, and since all critical comparisons are computed on conflict items presented identically across grades, this asymmetry does not bear on the reported contrasts.

Conflict accuracy did not differ between the two sets in either grade, at either response stage (initial responses—5th graders: Set A $M = 42.2\%$, Set B $M = 35.8\%$, $t(103) = 0.94$, $p = .349$; 7th graders: $M = 58.3\%$ vs. 61.3% , $t(100) = 0.42$, $p = .676$; final responses—5th graders: $M = 54.2\%$ vs. 42.5% , $t(103) = 1.52$, $p = .132$; 7th graders: $M = 69.6\%$ vs. 69.8% , $t(100) = 0.03$, $p = .978$). The set factor was therefore not included in the models.

We also examined whether numerical distance affected accuracy on the conflict items. Within participants (median split on distance within each set), initial-stage accuracy did not differ between closer and farther pairs (5th graders: $M = 40.2\%$ vs. 37.9% , $t(104) = 0.87$, $p = .388$; 7th graders: $M = 59.7\%$ vs. 59.5% , $t(98) = 0.05$, $p = .962$). At the final stage, closer pairs were answered more accurately in both grades (5th graders: $M = 50.5\%$ vs. 45.0% , $t(104) = 2.20$, $p = .030$, Cohen's $d = 0.21$; 7th graders: $M = 73.2\%$ vs. 64.0% , $t(98) = 2.91$, $p = .004$, Cohen's $d = 0.29$). Distance effects were thus inconsistent in direction across response stages, as might be expected when pairs are constructed around digit-length incongruence rather than magnitude. Because all children within a set solved identical conflict items, these item-level effects are constant across grades and cannot contribute to the developmental comparisons.

Study 2: Arithmetic Word task

The items presented here are translated into English. For the original French version, please reach out.

Practice items

Marc has 10 euros in his pocket.

Marc gives 5 to his friend Thomas.

How much money does he have left in his pocket?

- 15
- 10
- 5
- 1

Pauline has 12 euros in her pocket.

Pauline gives 6 to her friend Emilie.

How much money does she have left in her pocket?

- 6
- 18
- 24
- 3

Table S3. Arithmetic word problems used in Study 2 (translated from French).

Set	Item	Conflict type	Problem
A	A1	Conflict	<p>Agathe has 30 books.</p> <p>She has 20 books more than Ines.</p> <p>How many books does Ines have?</p> <ul style="list-style-type: none"> - 10 - 50 - 60 - 5
	A2	Conflict	<p>Alice has 45 flowers.</p> <p>She has 25 flowers more than Chloe. How many flowers does Chloe have?</p> <ul style="list-style-type: none"> - 20 - 70 - 90 - 10
	A3	Conflict	<p>Antoine has 45 olives.</p> <p>He has 20 olives more than Victor.</p> <p>How many olives does Victor have?</p> <ul style="list-style-type: none"> - 25 - 65 - 90 - 5
	A4	Conflict	<p>Arthur has 35 boxes.</p> <p>He has 5 boxes more than Pierre.</p> <p>How many boxes does Pierre have?</p> <ul style="list-style-type: none"> - 30 - 40

Set	Item	Conflict type	Problem
			<ul style="list-style-type: none"> - 70 - 10
	A5	Conflict	<p>Axel has 30 stamps. He has 10 stamps more than Thomas. How many stamps does Thomas have?</p> <ul style="list-style-type: none"> - 20 - 40 - 60 - 10
	A6	No-conflict	<p>Lola has 40 markers. Emma has 10 markers more than Lola. How many markers does Emma have?</p> <ul style="list-style-type: none"> - 50 - 25 - 75 - 5
	A7	No-conflict	<p>Louis has 30 marbles. Charles has 15 marbles more than Louis. How many marbles does Charles have?</p> <ul style="list-style-type: none"> - 45 - 25 - 70 - 5
	A8	No-conflict	<p>Raphael has 45 coins. Maxime has 5 coins more than Raphael. How many coins does Maxime have?</p>

Set	Item	Conflict type	Problem
			<ul style="list-style-type: none"> - 50 - 25 - 75 - 5
	A9	No-conflict	<p>Lucas has 30 candies.</p> <p>Paul has 5 candies more than Lucas. How many candies does Paul have?</p> <ul style="list-style-type: none"> - 35 - 20 - 55 - 5
	A10	No-conflict	<p>Lucie has 40 pencils.</p> <p>Sarah has 25 pencils more than Lucie. How many pencils does Sarah have?</p> <ul style="list-style-type: none"> - 65 - 35 - 100 - 5
B	B1	Conflict	<p>Elise has 50 photos.</p> <p>She has 15 photos more than Jade. How many photos does Jade have?</p> <ul style="list-style-type: none"> - 35 - 65 - 100 - 5
	B2	Conflict	<p>Tom has 40 matches.</p> <p>He has 5 matches more than Alan. How many matches does Alan have?</p>

Set	Item	Conflict type	Problem
			<ul style="list-style-type: none"> - 35 - 45 - 80 - 5
	B3	Conflict	<p>Eva has 30 bottles.</p> <p>She has 25 bottles more than Lea.</p> <p>How many bottles does Lea have?</p> <ul style="list-style-type: none"> - 5 - 55 - 60 - 1
	B4	Conflict	<p>Robin has 50 newspapers.</p> <p>He has 20 newspapers more than Hugo. How many newspapers does Hugo have?</p> <ul style="list-style-type: none"> - 30 - 70 - 100 - 10
	B5	Conflict	<p>Jules has 50 notebooks.</p> <p>He has 10 notebooks more than Nathan.</p> <p>How many notebooks does Nathan have?</p> <ul style="list-style-type: none"> - 40 - 60 - 100 - 20
	B6	No-conflict	Zoe has 35 candles.

Set	Item	Conflict type	Problem
			<p>Manon has 25 candles more than Zoe. How many candles does Manon have?</p> <ul style="list-style-type: none"> - 60 - 30 - 90 - 15
	B7	No-conflict	<p>Julia has 40 apples. Marie has 20 apples more than Julia. How many apples does Marie have?</p> <ul style="list-style-type: none"> - 60 - 30 - 90 - 15
	B8	No-conflict	<p>Laura has 45 cards. Celia has 10 cards more than Laura. How many cards does Celia have?</p> <ul style="list-style-type: none"> - 55 - 30 - 85 - 5
	B9	No-conflict	<p>Lena has 35 plants. Julie has 10 plants more than Lena. How many plants does Julie have?</p> <ul style="list-style-type: none"> - 45 - 25 - 70 - 5
	B10	No-conflict	<p>Leon has 35 paintbrushes.</p>

Set	Item	Conflict type	Problem
			<p>Theo has 15 paintbrushes more than Leon.</p> <p>How many paintbrushes does Theo have?</p> <ul style="list-style-type: none"> - 50 - 25 - 75 - 5

Error patterns on conflict items (Study 2)

To verify that incorrect initial responses corresponded to the intuitively cued response, we examined their distribution across the three incorrect options. The heuristic response (the addition of the two quantities) accounted for 80.8%, 71.9%, and 86.2% of incorrect initial responses on conflict items for the 5th graders-lenient, 5th graders-strict, and 7th graders, respectively—well above the 33.3% expected if errors had been distributed evenly across the three incorrect options. The conflict problems thus triggered the intended competition between the cued addition and the required operation.

S2. Task instructions

Instructions were administered in French; English translations are provided below. Screens are listed in the order in which participants encountered them. Values in brackets differed between groups. Consent and demographic screens are omitted.

Study 1: Decimal comparison task

General instructions (all groups).

"Read these instructions carefully. The task that follows includes about twenty exercises and a practice session. We want to know your intuitive answer, that is, the first answer that comes to mind, and then how you respond after having spent more time thinking."

"So, you will first be asked to give your intuitive answer: you must respond with the first answer that comes to mind. No need to rack your brain! You just have to give the first answer as quickly as possible. Then, the problem will be presented again and you can take all the time you want

to think about the solution. Once you have decided, you can choose your final answer. Press the Next button to read the rest of the instructions."

"You are going to see 2 numbers. Each time, you must press the LARGER number. It is very important to give your intuitive answer as quickly as possible. You will then have more time to think and choose your final answer. Press Next to read the rest of the instructions."

Deadline instructions.

"To make sure that you respond as quickly as possible for the first answer, there will be a time limit of [2 seconds / one and a half seconds]. Press Next to see an example of a problem, which will be presented for [2 seconds / one and a half seconds]. In this example, you do not need to choose an answer. Simply let the time run out."

"After selecting your first answer within the [2 seconds / one and a half seconds], the problem will be presented again and you can take all the time you want to think about it. When you have made your choice, you can then select your final answer. Once you have made your choice and clicked on it, you will automatically move on to the next problem. We will now do a practice session. Press Next to begin the practice."

Load task instructions (7th graders only).

"During the exercises, there will also be a pattern to memorize. You will see a grid with crosses, and you will have to memorize their location. We will first train you on the memorization alone. You will see an example of a pattern for 2 seconds, and you will then have to select it among 4 different patterns. Press Next to begin."

Pattern recognition prompt: "Among the following patterns, which one was presented to you?"

Feedback: "That was the correct pattern!" / "That was not the correct pattern. Try to concentrate on the pattern next time."

"There is 1 pattern to memorize per exercise. It will be presented just before the numbers. The difficulty can vary from one pattern to another. Try to memorize as many crosses as possible. It is not always easy to remember the pattern and solve the exercises at the same time. Concentrate on the pattern as a priority. We will now do one last practice session, with 2 complete exercises this time. Press Next to begin."

Deadline feedback (displayed whenever no response was given before the deadline).

"You did not answer before the time limit. Try to respond faster on the next trials. It does not matter if you are not sure of your intuitive answer—indicate it anyway. You will have more time to think afterwards."

Transition to the final response (practice).

"You will now be able to think as much as you want before giving your final answer. The color of the text has also changed to indicate that this is the final answer. Press Next to continue."

"This was the first practice exercise. You will do a second exercise to make sure you have fully understood. Remember that you must first respond as quickly as possible with the first answer that comes to mind. You will then have the opportunity to think longer to give your final answer."

"You have finished the practice. Press Next to start the task."

Study 2: Arithmetic word task

General instructions (all groups).

"Read these instructions carefully. The task that follows includes about twenty exercises and a practice session. There will be several reasoning exercises. We want to know your intuitive answer, that is, the first answer that comes to mind, and then how you respond after having spent more time thinking."

"So, as soon as you have the complete problem statement, you will first be asked to give your intuitive answer: you must respond with the first answer that comes to mind. No need to rack your brain! You just have to give the first answer as quickly as possible. Then, the statement will be presented again and you can take all the time you want to think about the solution. Once you have decided, you can choose your final answer. Press the Next button to begin the first part. To sum up, it is very important to give your intuitive answer as quickly as possible. You will then have more time to think and choose your final answer."

Practice instructions.

"We will begin with a practice session. For all problems, a fixation cross will appear first. Then, the first sentence of the problem will be presented for 2 seconds, followed by the complete problem. As a reminder, we are interested in your intuitive answer, the very first one that comes to mind. You must first respond with the first answer that comes to mind, without thinking. You just have to give the first answer as quickly as possible. To make sure that you respond as quickly as possible for the first answer, there will be a time limit of [8 seconds / 5 seconds]. When only one second remains, the background will turn yellow to remind you of the time limit. Try to respond before this limit! Then, the statement will be presented again and you can take all the time you want to think about it. When you have made your choice, you can then select your final answer. Once you have made your choice and clicked on it, you will automatically move on to the next problem. Press Next when you are ready to begin the practice."

Load task instructions (5th graders-strict and 7th graders).

"During the exercises, there will also be a pattern to memorize. You will see a grid with crosses, and you will have to memorize their location. We will first train you on the memorization alone. You will see an example of a pattern for 2 seconds, and you will then have to select it among 4 different patterns. Press Next to begin."

Pattern recognition prompt: "Among the following patterns, which one was presented to you?"

Feedback: "That was the correct pattern!" / "That was not the correct pattern. Try to concentrate on the pattern next time."

"There is 1 pattern to memorize per exercise. It will be presented after the first sentence of each problem statement. The difficulty can vary from one pattern to another. Try to memorize as many crosses as possible. It is not always easy to remember the pattern and solve the exercises at the same time. Concentrate on the pattern as a priority. We will now do one last practice session, with 2 complete exercises this time. Press Next to begin."

Deadline feedback (displayed whenever no response was given before the deadline).

"You did not answer before the time limit. Try to respond faster on the next trials. It does not matter if you are not sure of your intuitive answer—indicate it anyway. You will have more time to think afterwards."

Transition to the final response (practice).

"You will now be able to think as much as you want before giving your final answer. The color of the text has also changed to indicate that this is the final answer. Press Next to continue."

"This was the first practice exercise. You will do a second exercise to make sure you have fully understood. Remember that you must first respond as quickly as possible with the first answer that comes to mind. You will then have the opportunity to think longer to give your final answer."

"You have finished the practice! Press Next to start the task."

S3. Response times

Response times were recorded at both response stages. Means and SEMs were computed on log-transformed values and back-transformed for readability. Three features of the design constrain the interpretation of these latencies. First, initial response times are bounded by the response deadline and are therefore truncated by design: slower response tendencies surface as missed deadlines rather than as slow responses, so initial latencies cannot serve as a free measure of the speed of intuitive processing. In addition, for the groups tested under concurrent load (the 7th graders in both studies, and the 5th graders-strict in

Study 2), responding was interleaved with maintaining the dot pattern in memory, so their latencies also reflect the cost of managing the secondary task and are not directly comparable to those of groups tested without load. Second, final response times are collected on a representation of a problem participants have already read and answered once, so their latency has no unambiguous referent (the final latency alone, or the sum of both stages). Third, because the deadlines were calibrated to each task's basic processing demands (1.5–2 s for the decimal comparisons, 5–8 s for the arithmetic word problem), latencies are not comparable across studies. The values below should therefore be interpreted with caution.

Table S4 reports mean response times per direction of change pattern. Descriptively, on trials where an initially incorrect response was corrected ("01"), final responses were slower than initial ones in all groups of both studies (e.g., Study 1, 5th graders: $M = 1.16$ s vs. 1.50 s), consistent with corrections taking additional time. We refrain from testing this contrast formally because, as noted above, the two latencies are not commensurable.

We next compared initial response times on correct versus incorrect initial responses (on conflict items, computed on log-transformed participant means). In Study 1, correct initial responses were slightly slower than incorrect ones for 5th graders ($M = 1.24$ s vs. $M = 1.07$ s, $t(171) = 2.01$, $p = .046$, $d = 0.31$), but not reliably so for 7th graders ($M = 1.21$ s vs. $M = 1.12$ s, $t(163) = 1.80$, $p = .074$). In Study 2, correct and incorrect initial responses did not differ in any group (5th graders-lenient: $t(158) = 1.30$, $p = .195$; 5th graders-strict: $t(147) = 0.75$, $p = .456$; 7th graders: $t(149) = 1.50$, $p = .135$). The variant of this comparison restricted to "11" versus "00" trials—as suggested by one of the reviewers—yields larger differences in Study 1 (5th graders: $M = 1.31$ s vs. $M = 1.06$ s, $t(142) = 4.50$, $p < .001$, $d = 0.75$; 7th graders: $M = 1.25$ s vs. $M = 1.13$ s, $t(120) = 4.20$, $p < .001$, $d = 0.81$) and no significant differences in Study 2 ($ps = .055$ – $.852$). Note, however, that this variant conditions initial latencies on the final response: it excludes initially incorrect responses that were later corrected ("01") and initially correct responses that were later abandoned ("10"), and it aggregates within-participant differences with between-participant differences in overall speed.

That closely related analytic choices yield diverging outcomes underscores that these latencies afford only limited process inferences. Two further considerations apply to the observed slowing of correct initial responses. First, there is no established latency threshold separating intuitive processing from fast deliberation; a slowing of this magnitude, within a 1.5–2 s deadline, is compatible with fast deliberation but does not demonstrate it. Second, neither the automatization hypothesis nor recent dual process models assume that all intuitive responses are generated at the same speed: intuitions vary in activation strength (De Neys, 2021), and the assumption that heuristic intuitions are necessarily generated faster than logical ones has been directly examined and challenged (Bago & De Neys, 2017). A correct initial response that is somewhat slower than an incorrect one is thus also compatible with a weaker,

not-yet-fully automatized logical intuition. Within this design, initial response latencies therefore do not support firm process inferences in either direction.

Table S4. Mean response times (in seconds) on conflict items, by direction of change pattern. M (SEM).

Study	Group	Pattern	Initial response	Final response
Study 1	5th	00	1.06 (0.04)	0.86 (0.09)
		01	1.16 (0.04)	1.50 (0.16)
		10	1.19 (0.09)	1.90 (0.60)
		11	1.31 (0.03)	0.95 (0.08)
	7th	00	1.13 (0.03)	1.41 (0.10)
		01	1.12 (0.03)	1.61 (0.09)
		10	1.09 (0.04)	1.66 (0.15)
		11	1.25 (0.02)	1.46 (0.05)
Study 2	5th graders-lenient	00	4.54 (0.24)	3.67 (0.94)
		01	4.57 (0.34)	7.43 (0.93)
		10	3.20 (0.67)	7.62 (4.07)
		11	4.79 (0.13)	2.69 (0.29)
	5th graders-strict	00	3.54 (0.11)	4.91 (0.36)
		01	3.20 (0.18)	8.69 (1.21)
		10	3.82 (0.17)	9.69 (1.32)
		11	3.48 (0.10)	5.15 (0.38)

7th	00	3.07 (0.11)	5.63 (0.79)
	01	3.27 (0.19)	8.64 (0.69)
	10	2.76 (0.22)	6.17 (1.08)
	11	3.34 (0.09)	5.50 (0.37)

References

- Bago, B., & De Neys, W. (2017). Fast logic?: Examining the time course assumption of dual process theory. *Cognition*, 158, 90–109. <https://doi.org/10.1016/j.cognition.2016.10.014>
- De Neys, W. (2021). On dual- and single-process models of thinking. *Perspectives on Psychological Science*, 16(6), 1412–1427. <https://doi.org/10.1177/1745691620964172>