Reactive and proactive cognitive control as underlying processes of number processing in children

Anne-Merel Meijer[1]*, Bart Aben[1], Bert Reynvoet[1,2], & Eva Van den Bussche[1,3]

1 Brain & Cognition, KU Leuven, Leuven, Belgium

2 Faculty of Psychology and Educational Sciences, KU Leuven Kulak, Kortrijk, Belgium

3 Department of Psychology, Vrije Universiteit Brussel, Brussels, Belgium

*Corresponding author:

Anne-Merel Meijer

Brain & Cognition – Experimental Psychology

KU Leuven

Tiensestraat 102 - bus 3711

3000 Leuven

Belgium

E-mail: annemerel.meijer@kuleuven.be

Manuscript Category: Original Report

Declaration of Interest: The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

ABSTRACT

Cognitive control is crucial to resolve conflict in tasks such as the flanker task. Reactive control is used when conflict is rare, while proactive control is more efficient in situations where conflict is frequent. Macizo and Herrera (2012) found that these two control processes can also underlie two-digit number comparison in adults. Specifically, they observed that the unit-decade compatibility effect decreased in a block containing many conflict trials as compared to a block containing few conflict trials (i.e., a list-wide proportion congruency effect). In the present study we assessed whether this finding also applies to children (7-, 9-, and 11-year-olds). Participants performed a flanker and a two-digit number comparison task. In both tasks the proportion of conflict was manipulated (80% vs 20%). Results from the Flanker task showed a typical list-wide proportion congruency effect in reaction times, in all participating age groups. In the number comparison task we observed list-wide proportion congruency effects in both reaction times and error rates, which did not interact with age. Our findings support the assumption that children as young as seven years old can effectively use proactive and reactive control strategies. We showed that this effect is not limited to standardized, artificial laboratory tasks, such as the flanker task, but also underlies more daily life tasks, such as the processing of Arabic numbers.

Keywords: reactive control, proactive control, number comparison, children

Introduction

Cognitive control is the goal-directed regulation of thoughts and actions (Chevalier, Dauvier, & Blaye, 2018). It allows us to coordinate goals and actions to reach intrinsic behavioral goals (Braver, 2012). Cognitive control is needed to focus on relevant information and inhibit inappropriate or automatic responses to achieve our goals. Without it, almost no daily activities would be possible in our complex world. It is a set of mental processes that comes into play when information cues are contradictory or when routine behavior is not sufficient (Friedman, Nessler, Cycowicz, & Horton, 2009). For example, if you live in a country where people drive on the left side of the road and then visit a country where people drive on the right side, you have to actively inhibit your routine driving behavior. In the lab, conflict tasks such as the *arrow flanker task* (Eriksen & Eriksen, 1974), are often used to study cognitive control. In this specific task, participants have to indicate the direction of a central arrow which is flanked by two arrows on each side. Trials can either be congruent (<<<<<) or incongruent (<<><<). On incongruent trials, where a conflict is present between the central arrow and the flankers, cognitive control is needed to inhibit the irrelevant flankers in order to provide the correct response to the relevant central arrow. Typically, incongruent trials are responded to slower and less accurately compared to congruent trials, which is defined as the congruency (or flanker) effect.

Several theories have suggested that there are different mechanisms of cognitive control that operate on different temporal scales (e.g., Aben, Verguts, & Van den Bussche, 2017; Braver, 2012; Braver, Paxton, Locke, & Barch, 2009; Egner & Hirsch, 2005). One prominent theory differentiating time scales of cognitive control is the Dual Mechanisms of Control theory (Braver, 2012). This theory proposes two cognitive control modes: a sustained proactive control mode and a more transient reactive control mode (Braver, Reynolds, & Donaldson, 2003). Proactive control allows you to keep a task goal continuously active in working memory and to anticipate upcoming conflict. In the flanker task, this implies that the flankers are actively inhibited before the stimulus is presented. This proactive mechanism is effort-demanding because it requires sustained control across multiple trials but it also increases performance on incongruent trials. Reactive control detects and resolves conflict only after it has occurred. In the flanker task, this implies that cognitive control is deployed to resolve conflict once an incongruent trial is presented, leading to relatively long reaction times on incongruent trials. Importantly, if a task contains many incongruent trials, this frequent conflict can be anticipated, and proactive control is the most efficient control mode. Contrarily, in situations where conflict is scarce or unexpected, an effortful proactive mode would be inefficient as most trials do not require increased control, and hence a reactive control mode is preferred (Braver, 2012). Empirically, this leads to the observation that the congruency effect is smaller in blocks with mostly incongruent trials (where conflict is continuously anticipated), compared to blocks where most trials are congruent (where conflict is only resolved after conflict detection). This effect is known as the list-wide proportion congruency effect (PCE; Bugg, McDaniel, Scullin, & Braver, 2011; Logan & Zbrodoff, 1979). The PCE has been found in typical conflict tasks that assess cognitive control, such as the flanker task and the Stroop task (e.g., Abrahamse, Duthoo, Notebaert, & Risko, 2013; Lehle, & Hübner, 2008). Brain imaging studies using fMRI manipulated the proportion of congruency and observed sustained activation in lateral PFC (De Pisapia & Braver, 2006) or right middle frontal gyrus (Marini et al., 2016) when conflict is frequent, but only transient activation in these brain regions when conflict was rare. Sustained activity on blocks with mainly incongruent trials has even been found in fronto-parietal areas *prior* to stimulus-onset, which reflects proactive suppressing of distracting information. Contrarily, fronto-parietal areas only showed on-trial transient activity when conflict was scarce, which reflects that the conflict is solved only after stimulus-onset (Aben et al., 2019). This line of neuroimaging research therefore implies that a proportion congruency manipulation is able to trigger differential control modes with differential neural patterns (for a more extensive overview, see Aben et al., 2019).

Although these cognitive control processes are mainly studied in artificial laboratory tasks, they are constantly used in our daily life activities as well. All tasks that exceed routine or automatic behavior rely on cognitive control, including the processing of two-digit numbers (e.g., "39"). Dealing with (multi-digit) numbers is one of the most important abilities that children learn at school (Nuerk, Moeller, Klein, Willmes, & Fischer, 2011). It is often crucial to compare two-digit numbers to a target number in order to complete the task at hand in a school context. When two-digit numbers are compared, a unit-decade-compatibility effect (UDCE) is observed (Nuerk, Weger, & Willmes, 2001), which is driven by the place-value structure of the two to-be-compared numbers. A number pair is termed unit-decade compatible whenever separate decade and unit digit comparisons lead to the same decision (e.g., when comparing 42 and 57, the decade digit 4 of "42" is smaller than 5 of "57" and the unit digit 2 of "42" is also smaller than 7 of "57"). A number pair is incompatible when unit and decade comparisons lead to different decisions (e.g., 47 vs. 62; $4 < 6$, but $7 > 2$). Thus, as is the case for the flanker task, incompatible trials contain a conflict, whereas compatible trials do not. Several studies in children and adults have shown that these incompatible number pairs are processed slower and less accurately than compatible pairs (i.e., UDCE; see Nuerk et al., 2011 for a review). This suggests that comparing numbers when a conflict is present, requires increased cognitive control.

Crucially, it has been shown by Macizo and Herrera (2013) that cognitive control exerted by adults when comparing two-digit numbers can also operate on different time-scales, as is the case for often studied laboratory conflict tasks (e.g., flanker or Stroop tasks). Specifically, the UDCE is modulated by the proportion of incompatible trials: the effect decreases as the proportion of incompatible trials increases (Macizo & Herrera, 2013). This closely mimics the PCE described above: in both cases the effect driven by the presence of conflict (i.e., the flanker effect and the UDCE, respectively) is modulated by the proportion of conflict, leading to decreased effects when conflict can be anticipated. This indicates that also in number processing, a skill we constantly use in daily life, proactive and reactive control modes can be utilized. In the current study, we aimed to investigate whether modulations of the UDCE can also be observed in children. Number processing is a crucial step in the development of arithmetic skills that are usually mastered during the first grades in elementary school, such as the place-value system or written calculations (Landerl & Kölle, 2009; Landerl, 2013). The speed of digit comparison is found to be a unique predictor of variations in arithmetic skills in children (Durand, Hulme, Larkin, & Snowling, 2005). Furthermore, a larger UDCE indicates problems with assigning the single digits their place-value position, which has been shown to be a reliable early precursor of arithmetic capabilities later on (Lambert & Moeller, 2019; Moeller, Pixner, Zuber, Kaufmann, & Nuerk, 2011). Because the processing of numbers is such an important skill in children's development, it is key to investigate its underlying cognitive control mechanisms.

During childhood, cognitive control becomes more efficient as children grow older (Diamond, 2013). Archibald and Kerns (1999) showed that cognitive control ability increases during middle childhood (i.e., from 7 to 12 years). The age at which children reach adult levels of cognitive control efficiency appears to depend on the type of conflict task (Ambrosi, Lemaire, & Blaye, 2016). For the flanker task, 4- to 6-year-olds show larger congruency effects than young adults, while 7- to 9- and 10- to 13-year-olds already show congruency effects that are comparable to adults (Checa, Castellanos, Abundis-Gutiérrez & Rueda, 2014). Furthermore, the specific cognitive control mechanisms that are dominant also develop across childhood. Early in childhood, the preferred cognitive control mode is reactive control, but a shift takes place over the years towards a more proactive mode of control (Andrews-Hanna et al., 2011; Munakata, Snyder, & Chatham, 2012; Waxer & Morton, 2011). While very young children (3.5 years old) rely predominantly on reactive control, 6- and 8-year-olds are already able to engage in proactive cognitive control strategies as well (Chatham, Frank, & Munakata, 2009; Chevalier, Martis, Curran, & Munakata, 2015). PCEs, which indicate a deceased congruency effect in situations where proactive control is most efficient compared to situations where reactive control is stimulated, have been observed already in 5- to 7-year-olds (Ambrosi, Lemaire & Blaye, 2016), as well as in 9- to 12-year-olds (Wilk & Morton, 2012). However, the efficiency in coordinating both control strategies seems to further increase from age 7 onwards (Chevalier, James, Wiebe, Nelson, & Espy, 2014) and the consistency in proactive control engagement across contexts still seems relatively low in children aged 7 to 11 (Kubota et al., 2020). A recent study by Niebaum et al. (2020), for example, investigated the awareness of pro- and reactive control demands in children. Participants performed a card sorting task with two decks of cards, one enabling proactive control and the other enabling reactive control. They found that the majority of 5-year-olds were not aware of the difference between decks and selected the decks randomly, although a subsample preferred to play the reactive deck. Ten-year-olds also selected the decks at chance, but did report awareness of the difference. Adults reported awareness and showed a preference for the proactive deck (Niebaum, Chevalier, Guild, & Munakata, 2020). These results illustrate that children, next to using reactive control, can already engage proactive control from the age of five, but the efficiency and consistency of their proactive control engagement might continue to develop during childhood.

With regards to the use of cognitive control in number processing, studies in children found that the UDCE is already present in 7- to 8-year-olds and that the magnitude of the effect increases with age (Mann, Moeller, Pixner, Kaufmann, & Nuerk, 2012; Nuerk, Kaufmann, Zoppoth, & Willmes, 2004). However, whether the UDCE is also modulated by list-wide proportion congruency in children remains unclear. A recent study by Surrey et al. (2019) assessed list-wide proportion congruency effects in 9- to 12-year-olds and young adults using number stimuli. They used a one-digit number comparison task, where the two to-be-compared numbers differed in physical size (i.e., size congruity manipulation), while the participants had to indicate the numerically larger number. Participants received mostly congruent and mostly incongruent blocks, depending on the proportion of congruent and incongruent trials (20% versus 80%). Findings indicated a PCE in all participating age groups. Although this study already provides a first indication that reactive and proactive control might also underlie number processing in children, the conflict in this one-digit number comparison task was artificially induced by manipulating the size of the numbers. Hence, the number comparison itself in this task did not contain conflict. Furthermore, in the version of this task where participants only have to respond to the physical size of the numbers (Henik & Tzelgov, 1982), the number processing is an automatic process that does not correlate with mathematical achievement (Bugden

& Ansari, 2011; Landerl, 2013). Contrarily, in the two-digit comparison task (e.g., Nuerk, Weger, & Willmes, 2001), the conflict is inherently present in the number comparison itself (cf. UDCE). Furthermore, in this task the numbers have to be processed intentionally instead of automatically. intentional number processing, in digit or number comparison, is found to predict variability in children's performance on standardized tests of mathematical achievement (Bugden & Ansari, 2011; Landerl, 2013). Finally, this task resembles daily life tasks more than one-digit number comparison with a size congruity manipulation. Hence, a study using a two-digit comparison task would allow us to thoroughly assess modulations of cognitive control engagement in children during number processing.

The goal of the present study was twofold. First, we aimed to confirm previous studies showing that children aged 7 or older are already able to exert both reactive and proactive control using a list-wide proportion congruency manipulation in a flanker task. We differentiated reactive and proactive control based on a manipulation of the list-wide proportion congruency in different blocks of trials, creating a block that primarily triggers a reactive control strategy (i.e., a mainly congruent or MC block) and a block that primarily triggers a proactive control strategy (i.e., a mainly incongruent or MI block). This approach is based on previous studies that also used a conflict task with a list-wide proportion congruency manipulation in an fMRI set-up to create contexts inducing reactive and proactive control (e.g., Aben et al., 2019; De Pisapia & Braver, 2006; Marini et al., 2016). For example, Aben et al. (2019) showed that in the MC context, on-trial transient activity (incongruent – congruent trials) was increased in fronto-parietal areas, compared to the MI context. In the MI context, sustained activity in similar fronto-parietal areas during the intertrial interval was increased, compared to the MC context. This shows that this manipulation of list-wide proportion congruency can induce reactive and proactive task contexts. Second, and crucially, we aimed to establish whether these cognitive control mechanisms can also be observed in children's natural numerical processing, by studying modulations of the UDCE triggered by a list-wide proportion congruency manipulation in a two-digit number comparison task. We hypothesized that children used reactive control when the proportion of conflict in both the flanker and the two-digit number comparison tasks was small, leading to significant flanker effects and UDCEs, respectively. We expected that children would use proactive control when the proportion of conflict in both the flanker and the two-digit number comparison tasks was large, leading to reduced flanker effects and UDCEs (i.e., PCEs), respectively. Given that previous research has shown that children use proactive control from age 7, we expected that these effects would not be modulated by age.

Method

Participants

A total of 144 children aged between 6 and 13 and 46 university students participated in this study. All participants, and the parents of the children, provided written informed consent. Four participants were excluded from analysis because of dyscalculia, and one participant because of the inability to remain seated during the experiment. Participants who made more than 20% errors on the flanker task (N=3) and/or more than 30% errors on the, more difficult, number comparison task (N=4) were also excluded. Finally, participants responding more than 2.5 *SD*s above their age group mean on the flanker and/or the number comparison task were excluded from analysis (N=10). Thus, a total of 168 participants were included for analysis; 38 7-year-olds (2nd grade, age range [6-8], M_{age} = 7.3, sd_{age} = 0.57), 42 9-year-olds (4th grade, age range [9-10], M_{age} = 9.2, SD_{age} = 0.38), 47 11-year-olds (6th grade, age range [11-13], M_{age} = 11.3, sd_{age} = 0.51) and 41 university students (age range [17-23], M_{age} = 19.8, sd_{age} = 1.88). In the Flemish school system, processing numbers up until 20 is taught in the first grade, while in the second grade the range of numbers goes up to 100 and the unit-decade system is explained. All of our participants were at least in second grade and therefore able to process two-digit numbers. This study was part of a larger research project approved by the Ethical Committee of the KU Leuven (G-2017 10 951) and was performed in accordance with the guidelines and regulations of the KU Leuven.

Apparatus and materials

All computer tasks were designed and presented with E-prime software version 2.0 (Psychology Software Tools, Pittsburgh, PA). The students were seated in dimly lit private cubicles in front of a 17-inch PC monitor (60 Hz, spatial resolution = 1280 x 1024) located approximately 75 cm from the subject. The children were tested in their classrooms, one grade at a time, with one laptop for each participant.

Tasks

The main tasks used in this experiment were the arrow flanker task for the university students (Eriksen & Eriksen, 1974), the child flanker task for the children (Christ, Kester, Bodner, & Miles, 2011) and a number comparison task (Macizo & Herrera, 2013) for all participants. As this study was part of a larger research project, the students also performed the TTR, the Cognitive Developmental skills in aRithmetics (CDR, Desoete & Roeyers, 2006), and the Abbreviated Math Anxiety Scale (AMAS; Hopko, Mahadevan, Bare, & Hunt, 2003). For counterbalancing, half of the students performed the tests in the order described here, while the other half performed the AMAS first, followed by the TTR and CDR. The children also performed the Processing Speed test (Vos, Sasanguie, Gevers, & Reynvoet, 2017), the Implicit Association Test (IAT; Greenwald, McGhee, & Schwartz, 1998), the Ordering Ability test (Vos et al., 2017), the Dot Comparison task (Gebuis & Reynvoet, 2011), the Tempo Test Rekenen (TTR; De Vos, 1992), and the Revised Child Math Anxiety Questionnaire (CMAQ-R; Ramirez, Chang, Maloney, Levine, & Beilock, 2016). Since these tasks were not part of the current study, they will not be discussed here.

Flanker tasks. In the *arrow flanker task* (Eriksen & Eriksen, 1974), which was used for the university students, participants had to indicate the direction of the central arrow which was flanked by two arrows on each side. Instructions were presented on the screen and encouraged participants to react as quickly and as accurately as possible. Trials could either be congruent (<<<<< and >>>>>) or incongruent (<<><< and >><>>). A trial consisted of a fixation cross (500ms), followed by a blank screen (500ms), and the flanker stimulus (presented until response). All stimuli were presented in white, in the center of a black background (font Courier New, size 24). In the *child flanker task*, arrows were replaced by images of fish swimming either to the left or to the right (see Figure 1). Every trial started with a fixation cross (1000ms) presented in black in the center of a light-grey screen. This was followed by the flanker stimulus (size = 792 x 612 pixels), which was presented until response. Participants had to press “a” for indicating a left response and “p” for a right response using an AZERTY keyboard. For the children, stickers depicting a left- and right-swimming fish were taped on these two keys. During the practice trials of the students, feedback was provided by presenting the messages “correct”, “incorrect”, or “too slow” (if a response limit of 2000ms was exceeded) for 1000ms. For the children, a thumbs-up or thumbs-down image was shown for 1500ms after correct and incorrect trials, respectively. In the subsequent experimental trials feedback was no longer provided.

The ratio of congruent versus incongruent trials was manipulated between blocks; in the mostly congruent (MC) block, 80% of the trials were congruent and 20% were incongruent; in the mostly incongruent (MI) block, 80% of the trials were incongruent and 20% were congruent. The order of MI and MC blocks was counterbalanced across participants. The university students performed 16 practice trials, and 160 experimental trials per block (i.e., a total of 320 experimental trials). For the children, each block consisted of only 100 trials (i.e., a total of 200 experimental trials), with the possibility to take a break after every 25th trial to limit fatigue and maintain concentration. The children performed eight practice trials.

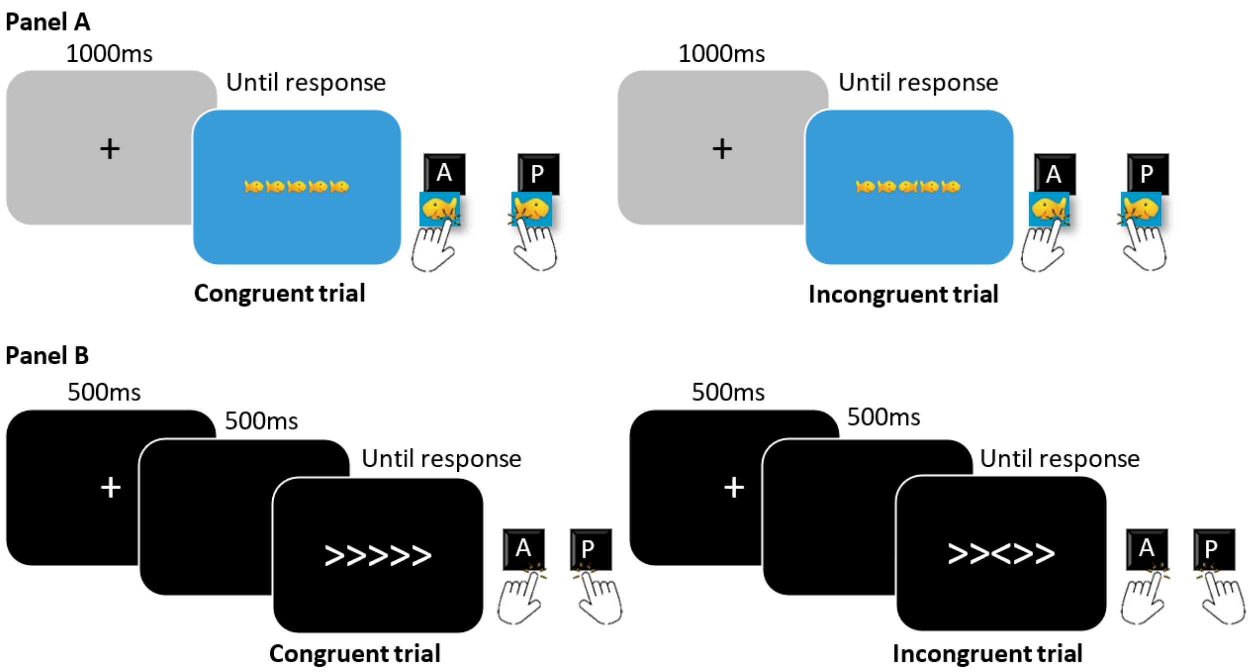

Figure 1. Schematic representation of the child flanker task (panel A) and the student flanker task (panel B).

Number comparison task. The stimuli and procedure for the *number comparison task* were based on Macizo and Herrera (2013), who kindly provided us with their stimulus lists. Their baseline condition (50% congruent and 50% incongruent trials) was not included in the present study, to limit the testing time for the children. In the number comparison task, two two-digit numbers were presented on each trial, one in the upper half of the screen and the other in the lower half (see Figure 2). Participants had to identify the numerically larger of the two numbers as fast and as accurately as possible by pressing J for the upper number or N for the lower, using their dominant hand on an Azerty keyboard. The two-digit numbers that were used ranged from 20 to 90. Trials could either be compatible (e.g., 24 vs 57) or incompatible (e.g., 27 vs 54). Each trial started with a central fixation, existing of two by two #-signs, presented for 480ms, which was followed by the two number stimuli. This stimuli remained on the screen until the participant's response. During the practice trials, feedback was provided for 1500ms either in the form of the messages "correct" or "incorrect" for the students or as a thumbs-up or thumbs-down picture for the children. This feedback was absent in the experimental trials. Between trials an inter-trial interval of 1000ms was present, during which a blank screen was shown.

As in the flanker task, the ratio of compatible versus incompatible trials was manipulated between blocks; in the mostly compatible (MC) block, 80% of the trials were compatible and 20% were incompatible (i.e. 27 vs 54); in the mostly incompatible (MI) block, 80% of the trials were incompatible and 20% were compatible. The order of these two blocks was counterbalanced across participants. The compatible and incompatible trials in the two conditions were equated in absolute distance, decade distance, unit distance and problem size (mean value of the two numbers in a given number pair), as detailed in Macizo and Herrera (2013). All participants performed eight practice trials and 100 experimental trials in each block (i.e., a total of 200 experimental trials), with breaks after every 50th trial for the students and every 25th trial for the children.

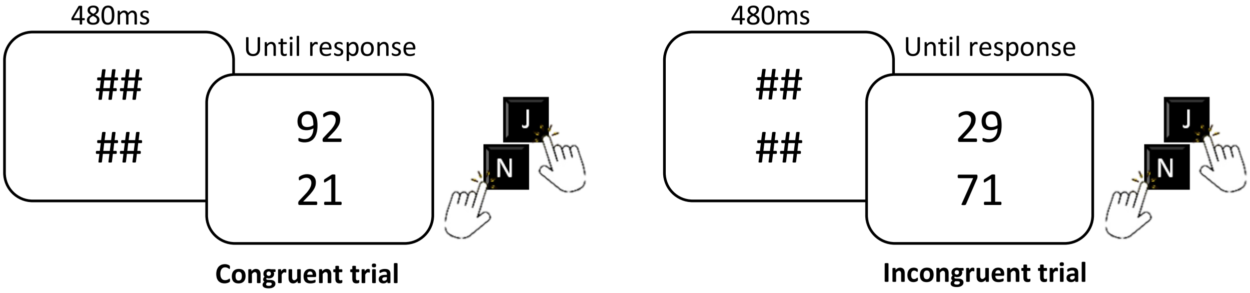

Figure 2. Schematic representation of the two-digit number comparison task.

Procedure

Testing took place at the Vrije Universiteit Brussel for the students and in the classrooms of the children during school hours. Before the testing, the parents of the children received an information letter to invite their children to participate. Only children of parents who provided written informed consent were included in this study. The university students were recruited via the university's participant pool. The students performed both the flanker and the number comparison task in one session and the order of these tasks was counterbalanced. After these tasks, they performed some other tasks mentioned above not included in the present study. For children, testing was divided in two sessions. Within the first session of the day, participants performed (among other tasks) the flanker task. In the second session, participants performed (among other tasks) the number comparison task. Each of the tasks took 10 to 15 minutes to complete, leading to a total duration around 20-30 minutes.

Statistical analysis

The flanker task contained only four different stimuli and two possible responses. Given the list-wide proportion congruency manipulation, this resulted in 33% (MI block) or 34% (MC block) of complete stimulus-response repetitions for the frequent trial type and 3.8% (MI block) or 4.2% (MC block) complete stimulus-response repetitions in the infrequent trial type. Complete repetition trials (e.g., >>>>> followed by >>>>>) can be responded to without engaging cognitive control, but by relying on low-level stimulus features (Mayr, Awh, & Laurey, 2003; Schmidt, 2013). Therefore, all complete repetition trials were excluded from statistical analysis in the flanker task (in total 27.50% of the trials). After excluding these complete repetition trials and the subsequent removal of incorrect trials (only for the RT analyses) and fast and slow responses exceeding the mean by 2.5 *SD* within each subject and per condition or faster than 150ms (see below), at least 15 observations remained in each condition for all participants for both the RT and error rate analyses. By design, there were no complete repetitions in the number comparison task.

For the RT analyses, trials including incorrect responses (i.e., 3.22% in the flanker task and 6.58% in the number comparison task), fast or slow responses exceeding the mean by 2.5 *SD* within each subject and per condition (i.e., 2.95% in the flanker task and 2.77% in the number comparison task) and fast responses below 150ms (i.e., 0.03% in the flanker task and 0.10% in the number comparison task), and were excluded from the RT analyses. For the flanker task, these trials were excluded after exclusion of the complete repetitions. The raw reaction times were transformed by a logarithmic scale, to optimally meet the linear mixed model assumptions (linearity, homoscedasticity, normally distributed residuals, absence of influential points, normally distributed random intercepts and slopes, and multicollinearity). For the error rate analyses, fast or slow responses exceeding the mean by 2.5 *SD* within each subject and per condition (i.e., 2.66% in the flanker task and 2.79% in the number comparison task) and fast responses below 150ms (i.e., 0.05% in the flanker task and 0.18% in the number comparison task) were excluded from the error rate analyses. For the flanker task, these trials were excluded after exclusion of the complete repetitions.

Linear mixed models were used to analyze reaction time (RT) data for both the flanker and the number comparison tasks separately. The DF approximation method was used, all fixed effects were included in all models and the random effects structure was modeled stepwise. The parameter estimates of the optimal model were statistically tested by using Satterthwaite's approximation of the degrees of freedom. The accuracy data were analyzed similarly, with a logistic link function from the lme4 package (Bates, Maechler, Bolker, & Walker, 2015). Since no adjustments of degrees of freedom are currently available for binary data, Wald's $\chi 2$ is reported instead of an *F*-value for the accuracy data. For each task, full models were constructed including all fixed factors and interactions. Fixed factors were Proportion Congruency, from here referred to as PC (MI versus MC), Congruency (congruent versus incongruent), and Age Group (7-, 9-, 11-year-olds, or university students). Sum coding was used, which compares the mean of each level of a predictor variable to the overall mean of all levels of the predictor. The random effects structure was modelled stepwise, corresponding to the DF approximation method. The first constructed model included a random intercept for each subject. The second and third models included random slopes for Congruency and PC, respectively. The fourth model included both Congruency and PC as random factors, and in the final fifth full factor model a random slope for the interaction between both random factors was additionally included. Models were fitted using linear mixed models with the maximum likelihood procedure in the lme4 package (Bates et al., 2015) for R (R Core Team, 2018). To find the model with the best fitting random effect structure, each augmented model was statistically compared to the previous one by using the likelihood ratio ($\chi 2$). Akaike information criterion (AIC; Akaike, 1974) is reported as a measure of model fit, with lower values indicating a better fit. The fixed effects of the optimal model were analyzed using Type III analysis of variance (ANOVA). Satterthwaite's adjustments were applied to obtain *F*- and *t*-statistics with approximate degrees of freedom. Post hoc comparisons were performed in case of significant Age Group or interaction effects and to examine congruency effects in the MC block, to assess our a priori hypothesis with regards to reactive control. The estimated marginal means of the slopes for Congruency, PC, and interaction effects were compared between the age groups, using Tukey's corrections for multiple comparisons. The anonymized, raw data that support the findings of this study are available in the Open Science Framework (OSF, https://osf.io/f4u6c/) with the identifier "https ://doi.org/10.17605 /OSF.IO/ F4U6C".

Results

Flanker task: Reaction times[1]

Model comparison of the linear mixed models for the transformed reaction time scores showed that the model with the full random effect structure was the best fit for our flanker task data (see Table 1).

[1] Analyzing the data while including the complete repetitions, yielded similar results in terms of the best fitting model and observed main and interaction effects in the ANOVAs. However, when including the complete repetitions, the PCE for the 9- and 11-year-olds no longer reached significance in the RT analysis. Overall, the interaction between block and congruency in the RT analyses (indexing the PCE) became less strong in the analysis including the complete repetitions (*F*-value of the block x congruency effect decreased from 107.43 to 83.55).

Table 1. Comparison of the models with different random effect structures for the flanker task

	Model	Random factor	*df*	AIC	log lik.	Test	χ^2	*p*
RT	1.	Subject (intercept)	18	4841	-2403			
	2.	Subject (intercept), Congruency (slope)	20	4503	-2231	2 vs. 1	342.80	<.001
	3.	Subject (intercept), PC (slope)	20	3962	-1961	3 vs. 1	833.11	<.001
	4.	Subject (intercept), Congruency (slope), PC (slope)	23	3853	-1903	4 vs. 2	655.91	<.001
						4 vs. 3	115.60	<.001
	5.	**Subject (intercept), Congruency (slope), PC (slope), Congruency × PC (slope)**	27	3835	-1890	5 vs. 4	25.96	<.001
Error rates	1.	Subject (intercept)	17	7311	-3639			
	2.	Subject (intercept), Congruency (slope)	19	7314	-3638	2 vs. 1	1.50	.47
	3.	Subject (intercept), PC (slope)	19	7312	-3637	3 vs. 1	3.71	.16
	4.	Subject (intercept), Congruency (slope), PC (slope)	22	7312	-3634	4 vs. 2	8.08	.04
						4 vs. 3	5.87	.12
	5.	**Subject (intercept), Congruency (slope), PC (slope), Congruency × PC (slope)**	26	7299	-3624	5 vs. 1	29.98	<.001
						5 vs. 2	28.48	<.001
						5 vs. 3	26.28	<.001
						5 vs. 4	20.40	<.001

Note. *df* = degrees of freedom, AIC = Akaike Information Criterion, log lik. = log likelihood, χ^2 = Chi Square, PC = Proportion Congruency.

The fixed effects of the full factor model were analyzed using Type III analysis of variance (ANOVA) with the Satterthwaite approximation of the degrees of freedom, represented in Table 2. Means (*SE*) for each condition and observed congruency effects are reported in Table 3. Results showed a significant main effect of Congruency,: participants responded faster to congruent trials ($\overline{X}_{\text{congruent}}$ = 6.39, $SE_{\text{congruent}}$ = 0.01) compared to incongruent trials ($\overline{X}_{\text{incongruent}}$ = 6.49, $SE_{\text{incongruent}}$ = 0.02), $F(1, 174.77) = 326.98$, $p < .001$. The significant main effect of Age Group indicated that RTs decreased with increasing age group, $F(3, 163.69) = 82.21$, $p < .001$. Post-hoc pairwise comparisons between age groups were all significant (all *t*'s > 4.06, *p*'s <.001), except for the 11-year-olds versus university students, $t(163) = -1.24$, $p = .60$. Mean values of the log transformed RTs were 6.82 (*SE* = 0.03) for the 7-year-olds, 6.44 (*SE* = 0.03) for the 9-year-olds, 6.27 (*SE* = 0.03) for the 11-year-olds, and 6.23 (*SE* = 0.03) for the university students. A significant Congruency by PC interaction was found, indicating a PCE, $F(1, 160.07) = 107.43$, $p < .001$. A larger congruency effect was found in the MC block (on average 0.149, *SE* = 0.0080) than in the MI block (on average 0.059, *SE* = 0.0063). The congruency effect was significant in the MC block, $t(169) = 18.56$, $p < .001$, suggesting that primarily reactive control was exerted on the incongruent trials in this block where conflict was rare. Congruency also interacted with Age Group, $F(3, 174.08) = 77.35$, $p < .001$. Post-hoc tests showed that the congruency effect was clearly larger for the university students (with Tukey's correction, every group differed significantly from the university students, all *t*'s > 10.90, all *p*'s < .001), while the differences between the child groups did not reach significance, all *t*'s < 1.74, all *p*'s > .30). However, congruency effects were still significant within each age group (all *t*'s > 3.52, *p*'s < .001). Mean congruency effects were 0.07 (*SE* = 0.01) for the 7-year-olds, 0.04 (*SE* = 0.01) for the 9-year-olds, 0.05 (*SE* = 0.01) for the 11-year-olds, and 0.25 (*SE* = 0.01) for the university students. The three-way interaction between Congruency, PC, and Age Group was also significant, $F(3, 157.59) = 5.62$, $p = .001$, indicating that the PCE differed across age groups. Post hoc analyses showed a significant PCE in all age groups. For the 7-year-olds the mean PCE was 0.13 (*SE* = 0.02, $t(178) = -7.00$, $p < .001$). For the 9-year-olds the mean PCE was 0.04 (*SE* = 0.02, $t(179) = -2.44$, $p = .016$). For the 11-year-olds the mean PCE was 0.07 (*SE* = 0.02, $t(177) = -3.96$, $p < .001$). For the university students the mean PCE was 0.12 (*SE* = 0.02, $t(113) = -7.42$, $p < .001$). Post-hoc pairwise comparisons between age groups with Tukey's correction showed that the 7-year-olds showed a significantly stronger PCE than the 9-year-olds ($t(179) = -3.39$, $p = .005$)

and the university students showed a significantly stronger PCE than the 9-year-olds ($t(144) = 3.15$, $p = .011$) (for mean values, see Table 3). Furthermore, when only taking into account the MC block to assess reactive control, a congruency effect was observed in that block for all age groups (congruency effects were on average 0.14, $SE = 0.02$, $t(179) = 7.96$, $p < .001$ for the 7-year-olds, 0.06, $SE = 0.02$, $t(180) = 3.86$, $p < .001$ for the 9-year-olds, 0.09, $SE = 0.02$, $t(179) = 5.69$, $p < .001$ for the 11-year-olds, and 0.31, $SE = 0.02$, $t(141) = 20.03$, $p < .001$ for the university students). The three-way interaction is also shown on Figure 3. None of the other effects reached significance. These results showed that all participating age groups primarily relied on reactive control when conflict was scarce (i.e., MC block) and all age groups showed a significant PCE, indicating proactive control engagement when conflict was frequent (i.e., MI block).

Table 2. Type III ANOVA of the fixed effects of the linear mixed model on RTs in the flanker task

Fixed effect	Sum Sq	Mean Sq	NumDF	DenDF	*F*	*p*
Congruency	21.05	21.05	1	174.77	326.98	<.001
PC	0.02	0.02	1	162.39	0.24	.63
Age Group	15.88	5.29	3	163.69	82.21	<.001
Congruency x PC	6.92	6.92	1	160.07	107.43	<.001
Congruency x Age Group	14.94	4.98	3	174.08	77.35	<.001
PC x Age Group	0.20	0.07	3	162.45	1.05	.37
Congruency × PC x Age Group	1.08	0.36	3	157.59	5.62	.001

Table 3. Means (*SE*) for the RTs and observed congruency effects (CE: incongruent - congruent) (in log(ms) and raw ms) in the different conditions of the flanker task

PC	Congruency	7-year-olds	9-year-olds	11-year-olds	University students
Mostly congruent	Congruent	6.76 (0.03) 867 (26.8)	6.41 (0.03) 609 (17.9)	6.23 (0.03) 506 (14.1)	6.06 (0.03) 430 (12.7)
	Incongruent	6.90 (0.03) 993 (33.9)	6.47 (0.03) 648 (21.1)	6.31 (0.03) 552 (17.0)	6.37 (0.03) 586 (19.1)
	CE	0.14 126	0.06 39	0.08 46	0.31 152
Mostly incongruent	Congruent	6.80 (0.03) 900 (29.3)	6.42 (0.03) 614 (19.0)	6.27 (0.03) 528 (15.5)	6.14 (0.03) 462 (12.7)
	Incongruent	6.81 (0.03) 904 (30.1)	6.44 (0.03) 626 (19.8)	6.29 (0.03) 539 (16.2)	6.33 (0.03) 560 (17.9)
	CE	0.01 4	0.02 12	0.02 11	0.19 98

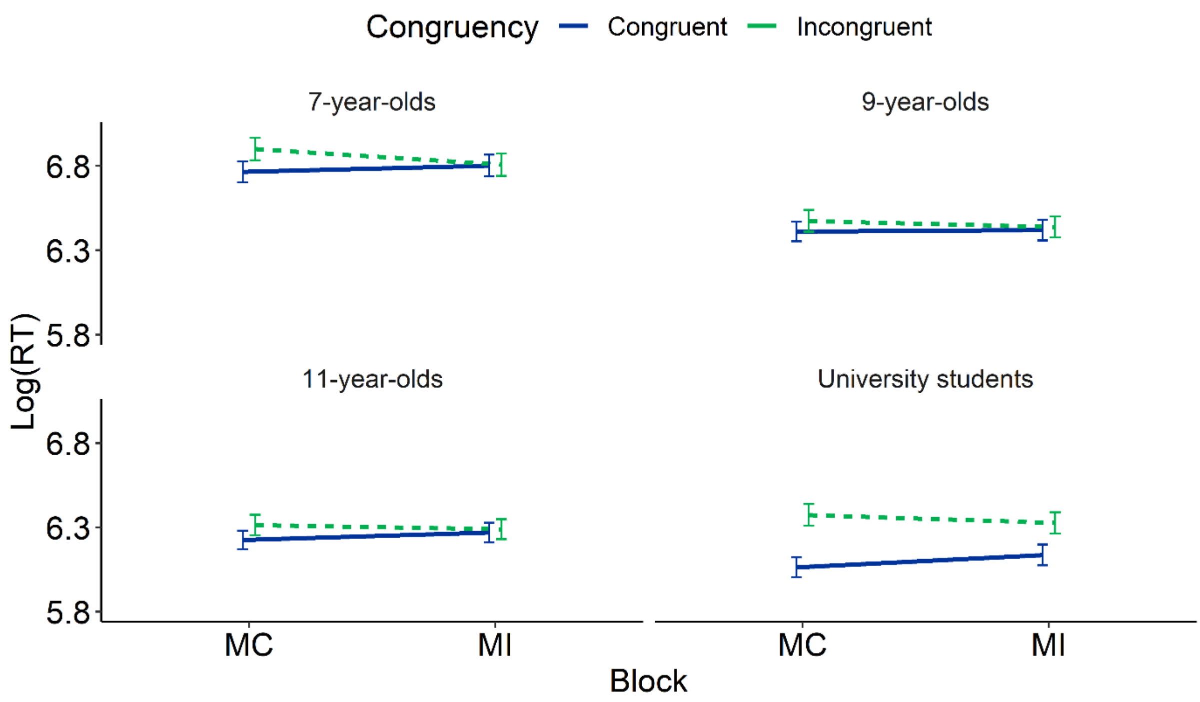

Figure 3. Flanker task: mean log-transformed reaction times per PC (MC and MI), Congruency (Congruent and Incongruent), and Age Group. Error bars represent confidence intervals.

Flanker task: Error rates

Model comparison of the linear mixed models for error rates showed that the full factor model had a significantly better fit than the other models (see Table 1). The fixed effects of this logistic linear mixed model were analyzed by a type III Wald chi-square test (see Table 4). Means (*SE*) for each condition and observed congruency effects are reported in Table 5. Results showed a significant main effect of Congruency, indicating a congruency effect: more errors were made on the incongruent trials than on the congruent trials ($\overline{X}_{\text{congruent}}$ = 0.95, $SE_{\text{congruent}}$ = 0.17, $\overline{X}_{\text{incongruent}}$ = 3.81, $SE_{\text{incongruent}}$ = 0.27, $\chi^2(1) = 55.89$, $p < .001$). A main effect of PC indicated that more errors were made in the mostly congruent (MC) block than in the mostly incongruent (MI) block ($\overline{X}_{\text{MC}}$ = 2.65, SE_{MC} = 0.22, $\overline{X}_{\text{MI}}$ = 1.37, SE_{MI} = 0.24, $\chi^2(1) = 12.31$, $p < .001$). A main effect of age group indicated that error rates differed across age groups, $\chi^2(3) = 10.54$, $p = .015$. Post hoc comparisons between groups with Tukey's correction showed only a significant difference between the 9-year-olds and the university students (odds ratio was 0.494, $z = -3.191$, $p = .008$). Mean error rates were 2.15 (*SE* = 0.34) for the 7-year-olds, 2.51 (*SE* = 0.38) for the 9-year-olds, 1.96 (*SE* = 0.30) for the 11-year-olds, and 1.26 (*SE* = 0.24) for the university students. Congruency interacted with Age Group ($\chi^2(3) = 46.68$, $p < .001$). Odds ratios (Incongruent/Congruent) for the congruency effects were 2.20 (*SE* = 0.58, $z = 3.01$, $p = .003$) for the 7-year-olds, 2.30 (*SE* = 0.60, $z = 3.20$, $p = .001$) for the 9-year-olds, 2.68 (*SE* = 0.70, $z = 3.76$, $p < .001$) for the 11-year-olds, and 21.62 (*SE* = 7.36, $z = 9.02$, $p < .001$) for the university students (see Table 5 for mean values). Post-hoc tests with Tukey's correction for comparing the congruency effect between age groups showed that the university students had a clearly larger congruency effect than the other age groups (*z*'s > 5.63, *p*'s < .001) while the child groups did not differ from each other in congruency effects (*z*'s < 0.15, *p*'s > .92). No significant interaction was found between Congruency and PC, which means that the PCE was not reflected in the overall error rates (see Figure 4 for a graphical representation). When only looking at the MC block, a significant congruency effect was observed ($\overline{X}_{\text{congruent}}$ = 1.48, $SE_{\text{congruent}}$ = 0.17, $\overline{X}_{\text{incongruent}}$ = 4.71, $SE_{\text{incongruent}}$ = 0.46, Odds ratio (Incongruent/Congruent) = 3.29, *SE* = 0.48, $z = 8.26$, $p < .001$), which was only significant for the university students. Odds ratios (Incongruent/Congruent) were 1.52 (*SE* = 0.40, *z* = 1.57, *p* = .397) for the 7-year-olds, 1.60 (*SE* = 0.36, *z* = 2.11, *p* = .151) for the 9-year-olds, 1.68 (*SE* = 0.38, *z* = 2.27, *p* = .105) for the 11-year-olds, and 28.89 (*SE* = 8.35, *z* = 11.64, *p* < .001) for the university students. None of the other effects reached significance. These results showed that all participating age groups relied primarily on reactive control when conflict was scarce (i.e., MC block). No significant PCE was observed, indicating that proactive control engagement when conflict was frequent (i.e., MI block) was not reflected in the error rates.

Table 4. Type III ANOVA of the fixed effects of the logistic linear mixed model on error rates in the flanker task

Fixed effect	Chi Sq	df	Pr (>Chi Sq)
Intercept	1445.87	1	<.001
Congruency	55.89	1	<.001
PC	12.31	1	<.001
Age Group	10.54	3	.015
Congruency x PC	1.41	1	.235
Congruency x Age Group	46.68	3	<.001
PC x Age Group	2.73	3	.436
Congruency × PC x Age Group	4.54	3	.209

Table 5. Mean (*SE*) and observed congruency effects (CE: incongruent - congruent) for the error rates (in %) in the flanker task

PC	Congruency	7-year-olds	9-year-olds	11-year-olds	University students
Mostly congruent	Congruent	2.10 (0.41)	2.68 (0.46)	2.15 (0.36)	0.39 (0.11)
	Incongruent	3.15 (0.69)	4.22 (0.78)	3.55 (0.66)	10.15 (0.12)
	CE	1.05	1.54	1.40	9.76
Mostly incongruent	Congruent	1.01 (0.44)	1.04 (0.48)	0.67 (0.31)	0.19 (0.12)
	Incongruent	3.14 (0.51)	3.35 (0.52)	2.81 (0.43)	3.01 (0.45)
	CE	2.13	2.31	2.14	2.82

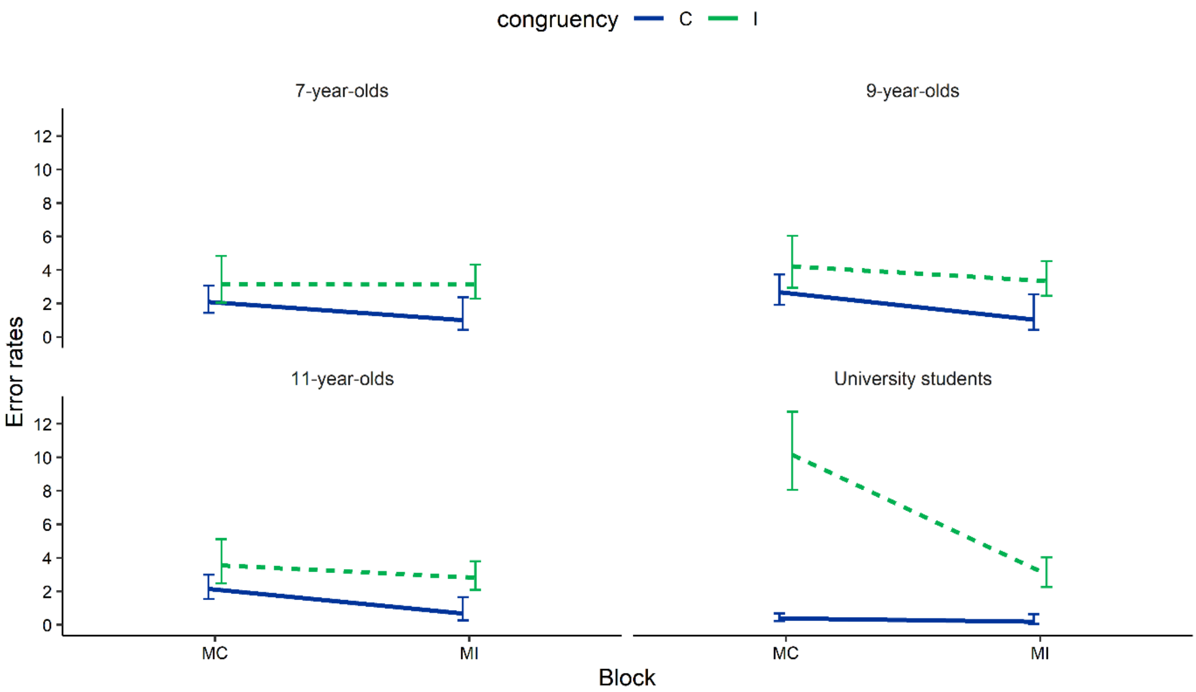

Figure 4. Flanker task: error rates in % per PC (MC and MI) and Congruency (Congruent and Incongruent) for the flanker task. Error bars represent confidence intervals.

Number comparison task: Reaction times

Model comparison of the linear mixed models for RTs showed that the model with random slopes for Congruency and PC showed the optimal fit (see Table 6).

Table 6. Comparison of the models with different random effect structures for the number comparison task

	Model	Random factor	*Df*	AIC	log lik.	Test	χ^2	*p*
RT	1.	Subject (intercept)	18	10221	-5093			
	2.	Subject (intercept), Congruency (slope)	20	10008	-4984	2 vs. 1	216.46	<.001
	3.	Subject (intercept), PC (slope)	20	9568	-4764	3 vs. 1	656.93	<.001
	4.	**Subject (intercept), Congruency (slope), PC (slope)**	23	9566	-4760	4 vs. 2	448.72	<.001
						4 vs. 3	8.25	.042
	5.	Subject (intercept), Congruency (slope), PC (slope), Congruency × PC (slope)	27	9572	-4759.1	5 vs. 4	1.54	.82
Error rates	1.	Subject (intercept)	17	15115	-7541			
	2.	**Subject (intercept), Congruency (slope)**	19	15111	-7536	2 vs. 1	8.513	.014
	3.	Subject (intercept), PC (slope)	19	15116	-7539	3 vs. 1	2.74	.25
						3 vs. 2	5.77	<.001
	4.	Subject (intercept), Congruency (slope), PC (slope)	22	15122	-7538	4 vs. 2	0	1
						4 vs. 3	1.00	.80
	5.	Subject (intercept), Congruency (slope), PC (slope), Congruency × PC (slope)	26	15122	-7535	5 vs. 4	7.89	.096

Note. df = degrees of freedom, AIC = Akaike Information Criterion, log lik. = log likelihood, χ^2 = Chi Square, PC = Proportion Congruency.

The fixed effects of the optimal model were analyzed using Type III ANOVA with the Satterthwaite approximation of the degrees of freedom (see Table 7). Means (*SE*) for each condition and observed congruency effects are reported in Table 8. Results showed a significant main effect of Congruency, indicating an UDCE: participants responded significantly slower on the incongruent trials than the congruent trials ($\overline{X}_{\text{congruent}}$ = 6.71, $SE_{\text{congruent}}$ = 0.01, $\overline{X}_{\text{incongruent}}$ = 6.74, $SE_{\text{incongruent}}$ = 0.01), $F(1, 191.42)$ = 43.53, $p < .001$. The significant main effect of Age Group indicated that younger participants responded slower than older participants, $F(3, 163.77)$ = 147.40, $p < .001$. Post-hoc pairwise comparisons between age groups were all significant (all *t*'s > 3.75, *p*'s <.001). Mean values of the log transformed RTs were 7.17 (*SE* 0.03) for the 7-year-olds, 6.75 (*SE* 0.03) for the 9-year-olds, 6.61 (*SE* 0.03) for the 11-year-olds, and 6.38 (*SE* 0.03) for the university students. A significant Congruency by PC interaction was found, indicating a PCE, $F(1, 875.85)$ = 3.93, p = .048 (see also Figure 5). A larger UDCE was found in the MC block (on average 0.036, *SE* = 0.0060) than in the MI block (on average 0.020, *SE* = 0.0058). The UDCE was significant in the MC block ($\overline{X}_{\text{congruent}}$ = 6.71, $SE_{\text{congruent}}$ = 0.01, $\overline{X}$-$_{\text{incongruent}}$ = 6.75, $SE_{\text{incongruent}}$ = 0.01, $t(588)$ = 6.02, $p < .001$), suggesting that primarily reactive control was exerted on the incongruent trials in this block where conflict was rare. This was the case in all age groups, except for the 7-year-olds (congruency effects were on average 0.01, *SE* = 0.01, $t(610)$ = 0.70, p = .90 for the 7-year-olds, 0.04, *SE* = 0.01, $t(615)$ = 3.80, $p < .001$ for the 9-year-olds, 0.04, *SE* = 0.01, $t(585)$ = 4.03, $p < .001$ for the 11-year-olds, and 0.05, *SE* = 0.01, $t(543)$ = 3.96, $p < .001$ for the university students). None of the other effects reached significance. These results showed that all participating age groups, except for the 7-year-olds, relied primarily on reactive control when conflict was scarce (i.e., MC block). A significant PCE was found, indicating proactive control engagement when conflict was frequent (i.e., MI block), but this effect did not interact with Age Group.

Table 7. Type III ANOVA of the fixed effects of the linear mixed model on RTs in the number comparison task

Fixed effect	Sum Sq	Mean Sq	NumDF	DenDF	*F*	*p*
Congruency	3.37	3.37	1	191.42	43.53	<.001
PC	0.00	0.00	1	160.32	0.02	.89
Age Group	34.19	11.40	3	163.77	147.40	<.001
Congruency x PC	0.30	0.30	1	875.85	3.93	.048
Congruency x Age Group	0.51	0.17	3	191.26	2.21	.089
PC x Age Group	0.13	0.04	3	160.30	0.55	.65
Congruency × PC x Age Group	0.13	0.04	3	875.05	0.54	.65

Table 8. Means (*SE*) for the RTs and observed congruency effects (CE: incongruent - congruent) (in log(ms) and raw ms) in the different conditions of the number comparison task

PC	Congruency	7-year-olds	9-year-olds	11-year-olds	University students
Mostly congruent	Congruent	7.16 (0.03) 1285 (36.9)	6.72 (0.03) 833 (22.7)	6.60 (0.03) 736 (19.0)	6.35 (0.03) 574 (15.9)
	Incongruent	7.17 (0.03) 1297 (39.8)	6.77 (0.03) 871 (25.5)	6.65 (0.03) 769 (21.2)	6.40 (0.03) 601 (17.7)
	CE	0.01 12	0.05 38	0.05 33	0.05 27
Mostly incongruent	Congruent	7.17 (0.03) 1297 (38.5)	6.74 (0.03) 844 (23.8)	6.60 (0.03) 733 (19.5)	6.37 (0.03) 585 (16.7)
	Incongruent	7.18 (0.03) 1309 (39.8)	6.77 (0.03) 871 (25.2)	6.62 (0.03) 746 (20.4)	6.39 (0.03) 597 (17.5)
	CE	0.01 12	0.03 27	0.02 13	0.02 12

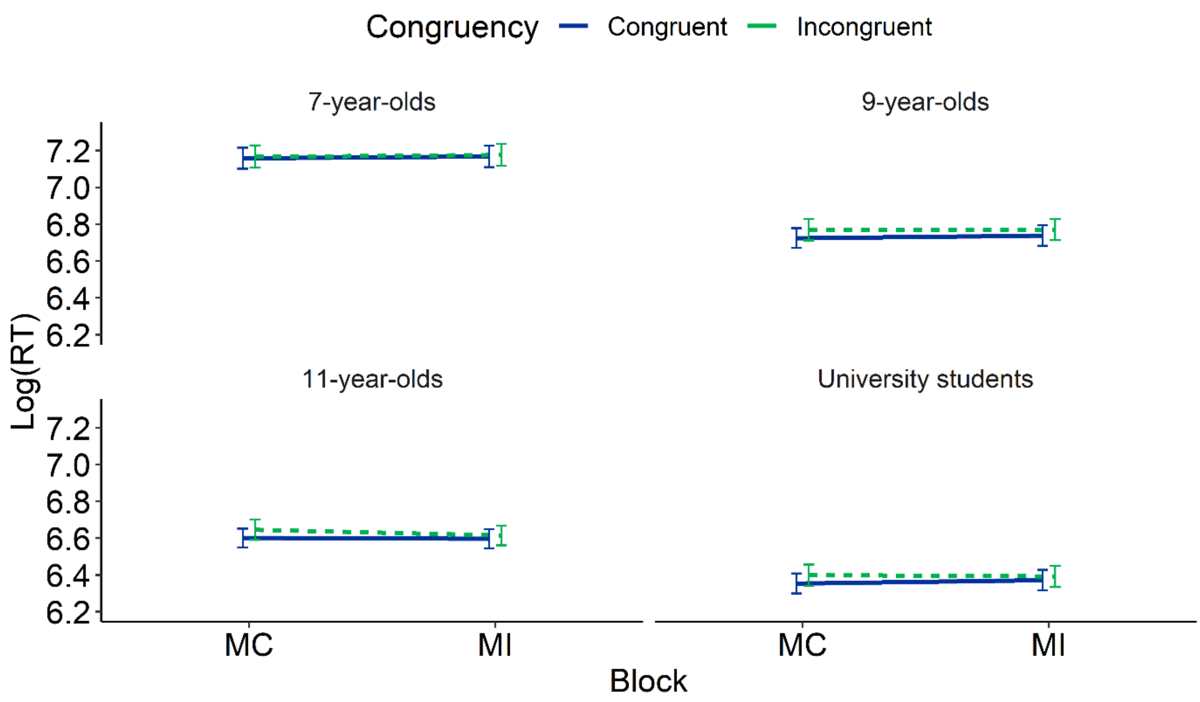

Figure 5. Number comparison task: log-transformed reaction times per PC (MC and MI) and Congruency (Congruent and Incongruent). Error bars represent confidence intervals.

Number comparison task: Error rates

Model comparison of the generalized linear mixed models or the error rates showed that the model with only a random slope for Congruency showed the best fit for our number comparison task data (see Table 6). The fixed effects of this logistic linear mixed model were analyzed by a type III Wald chi-square test (see Table 9). Means (*SE*) for each condition and observed congruency effects are reported in Table 10. Results showed a main effect of Congruency: more errors were made on the incongruent trials than on the congruent trials ($\overline{X}_{\text{congruent}}$ = 4.42, $SE_{\text{congruent}}$ = 0.30, $\overline{X}_{\text{incongruent}}$ = 7.14, $SE_{\text{incongruent}}$ = 0.42, $\chi^2(1) = 55.65$, $p < .001$). A main effect of PC showed that more errors were made in the mostly congruent (MC) block than in the mostly incongruent (MI) block ($\overline{X}_{\text{MC}}$ = 6.00, SE_{MC} = 0.35, $\overline{X}_{\text{MI}}$ = 5.28, SE_{MI} = 0.34, $\chi^2(1) = 5.66$, $p = .017$). The significant main effect of Age Group indicated differences in error rates between age groups ($\chi^2(3) = 22.50$, $p < .001$). Post hoc pairwise comparisons with Tukey's correction method showed that the university students made significantly fewer errors than the three other age groups (all *z*'s < -3.20, *p*'s < .01). The child age groups did not significantly differ from each other in terms of error rates. Mean values of the error rates were 6.54 (*SE* = 0.72) for the 7-year-olds, 7.23 (*SE* = 0.75) for the 9-year-olds, 5.89 (*SE* = 0.59) for the 11-year-olds, and 3.58 (*SE* = 0.43) for the university students. A significant Congruency by PC interaction was found, indicating a PCE ($\chi^2(1) = 15.13$, $p < .001$, see also Figure 6). Larger congruency effects were found in the MC block (odds ratio Incongruent/Congruent = 2.07, $z = 8.90$, $p < .001$) than in the MI block (odds ratio Incongruent/Congruent = 1.33, $z = 2.99$, $p = .003$), indicating a PCE and hence the exertion of proactive control in the MI block. The UDCE was significant in the MC block ($\overline{X}_{\text{congruent}}$ = 4.25, $SE_{\text{congruent}}$ = .28, $\overline{X}_{\text{incongruent}}$ = 8.42, $SE_{\text{incongruent}}$ = .62, Odds ratio (Incongruent/Congruent) = 2.07, *SE* = 0.17, $z = 8.90$, $p < .001$), suggesting that primarily reactive control was exerted on the incongruent trials in this block where conflict was rare. A significant Congruency by Age Group interaction indicated differences between age groups in terms of the congruency effect ($\chi^2(3) = 8.37$, $p = .039$). Post hoc tests showed that the congruency effect did not reach significance in the youngest age group, while in all three other age groups, more mistakes were made on incongruent trials in comparison to congruent trials. Odds ratio Incongruent/Congruent for the congruency effects were 1.56 (*SE* = 0.24, *z* = 2.83, *p* = .024) for the 7-year-olds, 1.94 (*SE* = 0.27, *z* = 4.70, *p* < .001) for the 9-year-olds, 2.52 (*SE* = 0.37, *z* = 6.30, *p* < .001) for the 11-year-olds, and 2.44 (*SE* = 0.45, *z* = 4.82, *p* < .001) for the university students. The three-way interaction effect did not reach significance, which indicates that the PCE did not significantly differ between age groups. Furthermore, when only taking into account the MC block to assess reactive control, a congruency effect was observed in that block for all age groups (Odds ratios (Incongruent/Congruent) were 1.55 (*SE* = 0.26, *z* = 2.66, *p* = .008) for the 7-year-olds, 1.95 (*SE* = 0.29, *z* = 4.48, *p* < .001) for the 9-year-olds, 2.56 (*SE* = 0.39, *z* = 6.17, *p* < .001) for the 11-year-olds, and 2.42 (*SE* = 0.46, *z* = 4.66, *p* < .001) for the university students). These results showed that all participating age groups primarily relied on reactive control when conflict was scarce (i.e., MC block). A significant PCE was found, indicating proactive control engagement when conflict was frequent (i.e., MI block), but this effect did not interact with Age Group.

Table 9. Type III ANOVA of the fixed effects of the logistic linear mixed model on error rates in the number comparison task

Fixed effect	Chi Sq	df	Pr (>Chi Sq)
Intercept	2352.51	1	<.001
Congruency	55.65	1	<.001
PC	5.66	1	.017
Age Group	22.50	3	<.001
Congruency x PC	15.13	1	<.001
Congruency x Age Group	8.37	3	.039
PC x Age Group	3.92	3	.270
Congruency × PC x Age Group	6.32	3	.097

Table 10. Mean (*SE*) and observed congruency effects (CE: incongruent - congruent) for the error rates in the number comparison task

PC	Congruency	7-year-olds	9-year-olds	11-year-olds	University students
Mostly congruent	Congruent	5.84 (0.72)	5.90 (0.69)	3.74 (0.46)	2.49 (0.36)
	Incongruent	8.80 (1.33)	10.84 (1.45)	8.92 (1.20)	5.86 (0.97)
	CE	2.96	4.94	5.18	3.37
Mostly incongruent	Congruent	5.85 (1.00)	5.47 (0.91)	5.88 (0.90)	2.37 (0.55)
	Incongruent	6.04 (0.75)	7.72 (0.86)	6.06 (0.67)	4.70 (0.59)
	CE	0.19	2.25	0.18	2.33

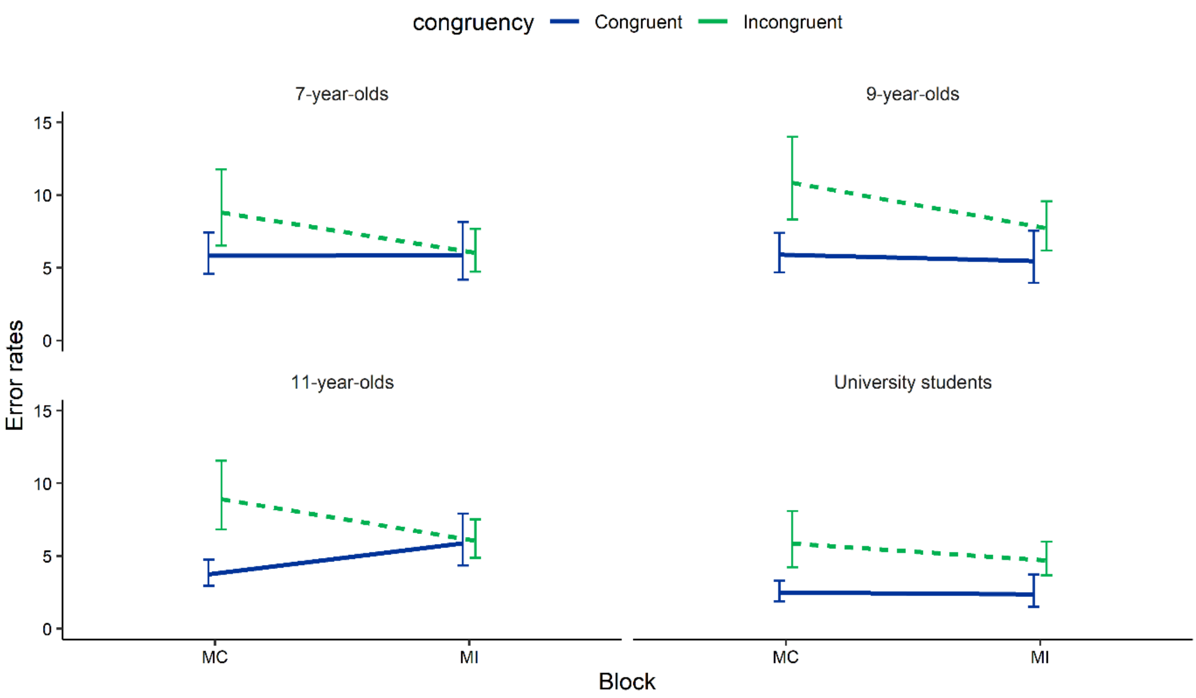

Figure 6. Number comparison task: error rates in % per PC (MC and MI) and Congruency (Congruent and Incongruent) for the number comparison task. Error bars represent confidence intervals.

Relation between the flanker and number comparison tasks

Since the present study included two different tasks, both with a list-wide proportion congruency manipulation, we can assess whether participants showed consistency across tasks in engaging proactive control mechanisms and how this changed with age. To this end, we calculated a proportion congruency effect score for each participant by calculating the congruency effects (incongruent – congruent RTs or error rates) for both blocks (MC and MI) and then subtracting the congruency effect in the MC block from that in the MI block. We did this for both tasks. Pearson correlations were then calculated between the PCEs of the two tasks, separately for the reaction times and the error rates. We did this for the complete sample of participants, as well as separately for each age group. The only correlations between the PCEs in the two tasks that reached significance were observed for the reaction times for the 11-year-olds (r = -.422, p = .003) and for the error rates for the 9-year-olds (r = -.339, p = .028). However, both these correlations were negative, indicating that in these cases a larger PCE in one task was related to a smaller PCE in the other task. Therefore, our data do not seem to support consistency in proactive control engagement between tasks.

Discussion

In this study, we assessed whether proactive and reactive control mechanisms are involved in the performance of laboratory conflict tasks and in two-digit number comparison in children (7-, 9-, and 11-year-olds) and young adults. Participants performed a number comparison task and a flanker task, with a list-wide proportion congruency manipulation in both tasks. We assessed the use of reactive control by studying the presence of congruency effects in the MC block where conflict was scarce and hence a reactive control strategy is expected. We studied the use of proactive control by examining the presence of a PCE, indicating a decreased congruency effect in the MI block where conflict is frequent and hence a proactive control strategy is expected.

Reactive control. In the flanker task, all participating age groups showed significant congruency effects in the reaction times, while only the university students showed these effects for error rate data for the MC block. These congruency effects indicate reliance on reactive control mechanisms. In the number comparison task we found significant congruency effects for all participating age groups in both the reaction time and the error rate data, except for the reaction time data of the 7-year-olds. These findings generally show that primarily reactive control strategies are exerted in the processing of flanker and two digit numbers in children aged 7 and older. We note however, that these conclusions with regards to reactive control are rather indirect, reliant on the assumption that our MC block indeed triggered primarily reactive control, as is typically assumed. Although it seems unlikely that participants used a proactive strategy in the MC block, given that there is hardly any conflict to proactively anticipate (and which could also not explain our observed

PCEs), it is still a possibility. Therefore, future research could for example include a neutral block as well, containing 50% congruent and 50% incongruent trials, which would allow a more direct assessment of whether participants used a reactive control strategy in the scarce conflict context as compared to a more neutral context.

Proactive control in the flanker task. Results from the flanker task showed a typical PCE in the reaction time data for all participating age groups, and this PCE was significantly stronger in the 7-year-olds than in the 9-year-olds as well as significantly stronger in university students than in the 9-year-olds. The PCE was not reflected in the error rate data of the flanker task. However, in the number comparison task we observed PCEs in both the reaction time and the error rate data, which did not interact with age group. This generally suggests that proactive control strategies can be exerted even by young children, although especially in the flanker task this seemed less stable. Our findings from the flanker task confirm that cognitive control can operate on different time scales, depending on the frequency of conflict. Although we found PCEs in all participating age groups, which is in line with a recent study where PCEs were observed in children aged 9 or older (Surrey et al., 2019), this PCE was stronger in the 7-year-olds than in the 9- and 11-year-olds. It could be that even though young children are able to use both reactive and proactive control strategies, coordinating both strategies might still be unstable in terms of efficiency. This could explain why we observed a stronger PCE for the 7-year-olds than for the older children and it could also explain why different studies report inconsistent results with regards to a PCE in children. This is also in line with Niebaum et al. (2020) who showed that although 10-year-olds were aware that one card deck required proactive control and the other required reactive control, their actual choices for which card deck to play with were still random, whereas adults preferred the proactive deck. Taking this into account, we would argue that proactive and reactive control mechanisms can be used by children from age 7 onwards, but the efficiency to deploy them can still vary during childhood. A longitudinal study within the same group of children could shed further light on the stability of these control mechanisms across age and at which age they are matured.

In the current study, we assumed that the effects we observed in the flanker task stem from reliance on cognitive control mechanisms while performing the task. However, several alternative accounts have been formulated, which could potentially also explain our findings of the flanker task (Schmidt, 2013). These alternative underlying mechanisms could be feature repetitions (Mayr, Awh, & Laurey, 2003), contingency learning (Schmidt & Besner, 2008) or temporal learning accounts (Cohen-Shikora, Suh & Bugg, 2019; Schmidt, 2013; Spinelli, Perry & Lupker, 2019). As these alternative accounts only relate to our flanker task and not to the number comparison task which was the central part of this study, we discuss them in more detail in Appendix B. To avoid reliance on low-level features as an alternative explanation for our flanker results as much as possible, we excluded all complete repetitions from our flanker data. Interestingly, a comparison between the analyses with and without the complete repetitions in the flanker task, revealed that the interaction between block and congruency (indexing the PCE) became slightly stronger in the analysis without the complete repetitions. At first sight, this seems counterintuitive: if low-level features are an alternative explanation for findings typically attributed to cognitive control exertion, we do not expect list-wide proportion congruency effects to increase when these features are removed. However, scrutinizing the data further showed that how the interaction is impacted by the removal of the complete repetitions, seemed to differ between children and adults. Specifically, in all child groups, PCEs were larger when repetitions were removed, whereas for the young adults, the PCE decreased. Although speculative, perhaps children do not rely yet on low-level features whenever they can, and the presence of these features might therefore rather introduce noise to the data instead of being helpful for task performance. This would explain why the removal of repetitions actually increased the observed PCE for children. Contrarily, young adults do seem to rely on these low-level features when possible (as expected), as evidenced by a decrease of the PCE when repetitions are removed. Importantly, even when all complete trial repetitions are removed from the analysis, partial response repetitions are still present in the data (e.g., >>>>> followed by <<><< both requiring the same response). These partial repetitions could have exerted an influence, as participants might use them to generate their response, instead of actually exerting cognitive control. In our study, it was not possible to also exclude these partial repetitions, as insufficient data points would have remained (i.e., only 3 observations in some cells). In future studies, the presence of partial and complete repetitions should be avoided by designing a task with more than two possible responses and a larger variety of stimuli, for example a number flanker task (e.g., Aben et al., 2019). Another potential limitation is that children and adults performed different versions of the flanker task. The adults performed a classic flanker task with arrows, while the children performed an adjusted version where the arrows were replaced by images of fish swimming to the left or right. Rueda et al. (2004) tested both the arrow and the fish flanker tasks in children (age 10) and adults to compare performance on the two different stimulus types. They showed that an arrow version of the flanker task is more difficult than a fish version. They also showed significantly smaller congruency effects in the fish version in comparison to the arrow version. It is therefore very well possible that these differences also apply to our study. This would imply that the adults generally would have had relatively slower reaction times and lower accuracy due to their task being more difficult than the children's task and that these task differences potentially led to smaller congruency effects for the children. However, all child participants performed the same task, so this does not impact the comparison between age groups in the child participants. Furthermore, this study was primarily focused on assessing the PCE in a numerical context, where the flanker task served as a baseline task.

Proactive control in the number comparison task. Our findings in the number comparison task confirm that cognitive control is involved in two-digit number comparison and that this cognitive control also operates on different timescales during this numerical process. Here, we looked at the unit-decade-compatibility effect (UDCE) and assessed whether this effect was also susceptible to the proportion of conflicting trials. As expected, the university students showed the most pronounced UDCE. The children also showed a significant UDCE, with the exception of the reaction time data of the 7-year-olds. This generally confirms earlier findings, in which an UDCE was found in 7- to 8-year-olds and the magnitude of the effect increased with age (Mann et al., 2012; Nuerk et al., 2004).

Processing of two-digit numbers is taught from the 2nd grade in Belgium, so all the participating age groups had experience with two-digit numbers, although the level of experience naturally varied between age groups. This could explain the more robust UDCE findings in the older age groups in comparison to the 7-year-olds. Importantly, we observed that this UDCE was modulated by the proportion of conflict present in the task, constituting a PCE in both reaction times and error rates. This replicates the findings by Macizo and Herrera (2013) in children. Importantly, the PCEs we observed were not influenced by age group, suggesting that even in children two-digit number comparison involves proactive control processes.

Previous studies have shown that cognitive control abilities might underlie numerical processes (Cragg & Gilmore, 2014; Cragg et al., 2017). Proactive control is an important ability, that allows us to actively maintain goal-relevant information in order to suppress distracting information and avoid cognitively demanding conflict resolution as much as possible (Braver, 2012). Our study empirically shows that proactive control does not only underlie performance on typical artificial conflict tasks, such as the flanker task, but also tasks that are omnipresent in children's education and our daily lives, such as comparing two-digit numbers. If we can pinpoint the age on which children start to use proactive control strategies, we can stimulate this switch from mainly reactive to mainly proactive control use in the classroom. Furthermore, problems with this switch can be detected early and perhaps remediated. Chevalier et al. (2015; see also Chevalier et al., 2020) for example has already shown that the use of proactive control can be triggered in young children by manipulating the onset and duration of a task cue in a sorting task (either by shape or color) paradigm. They implemented three different conditions in their task cues: proactive impossible (cue and target were presented at the same time, making it impossible to prepare the task), proactive possible (the cue was presented before the target, but stayed on the screen during target presentation, making it possible to either prepare the task or wait until target presentation) and proactive encouraged (the cue was presented and disappeared before target onset, thereby forcing the participant to process the cue before the target was presented). Results showed that 10-year-olds engaged proactive control when possible (in the proactive possible and proactive encouraged conditions), while 5-year-olds only engaged proactive control when reactive control was made more difficult (in the proactive encouraged condition). It is important to investigate cognitive control processes such as these, that (partially) underly number processing to develop strategies for improving number processing skills and arithmetic in children's education. As our study shows that proactive strategies can also underlie number processing, and others have already shown that these strategies can be encouraged, this could potentially benefit teaching and learning situations. For example, in explaining two-digit number comparison, a teacher could highlight the decade digit to encourage children to prioritize that digit over the unit digit, in a proactive way. A second possible explanation for why we observe proactive control strategies in this age group for number comparison, might be related to the effect of schooling, experience and increasing expertise with numbers (see also previous comment). If that is the case, then expediting the acquaintance with numbers in school might shift the onset of the use of proactive control strategies to a younger age. Studies from other countries have already shown that preschool children in America from 3 to 6 years old can already successfully compare two-digit numbers (Yuan, Prather, Mix, & Smith, 2019) and that experience with number processing at a young age can give children a head start in mathematical achievement in Asian countries (Gerofsky, 2015). A study in which two-digit number addition in Singaporean preschoolers was compared to Japanese preschoolers (Marcruz et al. 2020) showed that the strategies of solving the math problems were culture-dependent (either by using a base-10 decomposition method, standard algorithms or basic counting). These results show that there are different strategies to approach two-digit number processing in children and that certain strategies can be encouraged by parents or teachers. A study where proactive control strategies are compared between countries where two-digit numbers are introduced at different ages could shed further light on this, as well as interventional studies where two-digit numbers are either introduced earlier in a group of preschool children or not

Task Comparison. Although our findings show an interaction between block and congruency in both the flanker and the number comparison task, this interaction was found to interact with age only in the flanker task. Previous studies have found a block by congruency interaction in 7- to 11-year-olds as well, but not always consistently and the efficacy of coordinating both control strategies is still developing in that age range (Ambrosi, Lemaire & Blaye, 2016; Chatham, Frank, & Munakata, 2009; Chevalier et al., 2014; Chevalier et al., 2015; Kubota et al., 2020). Next to the fact that proactive control mechanisms might still be in full development in children with this age range leading to sometimes inconsistent results, we could also speculate that differences in the use of proactive control might be due to the nature of the tasks. Children aged 7 to 11 years have already received training with numbers in school and are already familiar with all the different digits. Comparing two-digit numbers is a skill that is also already prevalent in their arithmetic education in Flanders. Hence, this experience with two-digit numbers might trigger more proactive strategies. The flanker task on the other hand is a completely new task for the children which they have never encountered. Perhaps this novelty of the task might hamper the use of proactive strategies to some degree in this task for children. Still, this is purely speculative, and further research should address the role that schooling and experience play in the use of proactive control in children. Our data also showed that the observed PCEs in the flanker task and in the number comparison task did not positively correlate. This could also explain the inconsistency between our results of the number comparison task and the findings of Surrey et al. (2019) who did observe that the PCE in number processing depends on age. However, as argued in the introduction, they used a one-digit number comparison task where the conflict was induced by an additional, irrelevant parameter (i.e., physical size of the numbers). This might make their design more comparable to a flanker task, for which we also observed age-related modulations of the PCE. Furthermore, this size manipulation is mainly used to study automatic number processing (Bugden & Ansari, 2011; Landerl, 2013), which is a skill that shows no correlation with arithmetic abilities. Moreover, their size congruity manipulation does not resemble an everyday life task or situation where we have to deal with numbers. Contrarily, in our two-digit number comparison task the conflict was induced within the numbers itself, by the UDCE. Our task reflects the frequent occurrence of number comparison in daily life tasks and the inherent conflict that might be present there. In that sense, our study adds to our understanding of how and why adults and children exert cognitive control in the everyday processing of Arabic numbers, a skill that is found to be a unique predictor of variations in arithmetic skills in children.

Conclusion

In the current study we assessed children's and adults' reactive and proactive control strategies in a flanker and a number comparison task by using a list-wide proportion congruency manipulation. We observed that young children are already able to use both reactive and proactive control strategies in the artificial flanker task. Crucially, we found that in two-digit number processing, a numerical skill that we use constantly in daily life as well, children and adults also use reactive and proactive control strategies. We found no indication that these control strategies used during two-digit number comparison were modulated by age.

References

Aben, B., Verguts, T., & Van den Bussche, E. (2017). Beyond trial-by-trial adaptation: A quantification of the time scale of cognitive control. *Journal of Experimental Psychology: Human Perception and Performance, 43*(3), 509.

Aben, B., Calderon, C. B., Van der Cruyssen, L., Picksak, D., Van den Bussche, E., & Verguts, T. (2019). Context-dependent modulation of cognitive control involves different temporal profiles of fronto-parietal activity. *NeuroImage, 189*, 755-762.

Abrahamse, E. L., Duthoo, W., Notebaert, W., & Risko, E. F. (2013). Attention Modulation by Proportion Congruency: The Asymmetrical List Shifting Effect. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 39*(5), 1552-1562.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE transactions on automatic control*, *19*(6), 716-723.

Ambrosi, S., Lemaire, P., & Blaye, A. (2016). Do Young Children Modulate Their Cognitive Control? Experimental Psychology, 63(2), 117-126.

Andrews-Hanna, J. R., Mackiewicz Seghete, K. L., Claus, E. D., Burgess, G. C., Ruzic, L., & Banich, M. T. (2011). Cognitive control in adolescence: Neural underpinnings and relation to self-report behaviors. PloS one, 6, e21598. doi:10.1371/journal.pone.0021598

Archibald, S., & Kerns, K. (1999). Identification and Description of New Tests of Executive Functioning in Children. Child Neuropsychology, 5(2), 115-129.

Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., ... & Bolker, M. B. (2015). Package 'lme4'. *Convergence*, *12*(1), 2.

Braver, T. S., Reynolds, J. R., & Donaldson, D. I. (2003). Neural mechanisms of transient and sustained cognitive control during task switching. *Neuron, 39*(4), 713-726.

Braver, T. S., Paxton, J. L., Locke, H. S., & Barch, D. M. (2009). Flexible neural mechanisms of cognitive control within human prefrontal cortex. *Proceedings of the National Academy of Sciences, 106*(18), 7351-7356.

Braver, T. (2012). The variable nature of cognitive control: A dual mechanisms framework. *Trends in Cognitive Sciences, 16*(2), 106-113.

Bugden, S., & Ansari, D. (2011). Individual differences in children's mathematical competence are related to the intentional but not automatic processing of Arabic numerals. *Cognition, 118*(1), 35-47.

Bugg, J. M., McDaniel, M. A., Scullin, M. K., & Braver, T. S. (2011). Revealing list-level control in the Stroop task by uncovering its benefits and a cost. *Journal of Experimental Psychology: Human Perception and Performance, 37*(5), 1595.

Chatham, C.H., Frank, M.J., & Munakata, Y. (2009). Pupillometric and behavioral markers of a developmental shift in the temporal dynamics of cognitive control. *Proceedings of the National Academy of Sciences, 106*(14), 5529-5533.

Checa, P., Castellanos, M., Abundis-Gutiérrez, A., & Rosario Rueda, M. (2014). Development of neural mechanisms of conflict and error processing during childhood: Implications for self-regulation. Frontiers in Psychology, 5, 326.

Chevalier, N., James, T., Wiebe, S., Nelson, J., & Espy, K. (2014). Contribution of Reactive and Proactive Control to Children's Working Memory Performance: Insight From Item Recall Durations in Response Sequence Planning. Developmental Psychology, 50(7), 1999-2008.

Chevalier, N., Martis, S., Curran, T., & Munakata, Y. (2015). Metacognitive processes in executive control development: The case of reactive and proactive control. 27(6), 1125-36.

Chevalier, N., Dauvier, B., & Blaye, A. (2018). From prioritizing objects to prioritizing cues: A developmental shift for cognitive control. Developmental Science, 21(2), N/a.

Chevalier, N., Meaney, J. A., Traut, H. J., & Munakata, Y. (2020). Adaptiveness in proactive control engagement in children and adults. *Developmental cognitive neuroscience, 46*, 100870.

Christ, S., Kester, L., Bodner, K., & Miles, J. (2011). Evidence for Selective Inhibitory Impairment in Individuals With Autism Spectrum Disorder. Neuropsychology, 25(6), 690-701.

Cohen-Shikora, E., Suh, J., & Bugg, J. M. (2019). Assessing the temporal learning account of the list-wide proportion congruence effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 45*(9), 1703-1723.

Desoete, A., & Roeyers, H. (2006). *Cognitieve Deelvaardigheden Rekenen Handleiding*. Herentals: Vlaamse Vereniging voor Logopedisten.

Cragg, L., & Gilmore, C. (2014). Skills underlying mathematics: The role of executive function in the development of mathematics proficiency. *Trends in neuroscience and education*, *3*(2), 63-68.

Cragg, L., Keeble, S., Richardson, S., Roome, H. E., & Gilmore, C. (2017). Direct and indirect influences of executive functions on mathematics achievement. *Cognition, 162*, 12-26.

De Pisapia, N., & Braver, T. S. (2006). A model of dual control mechanisms through anterior cingulate and prefrontal cortex interactions. *Neurocomputing, 69*(10-12), 1322-1326.

De Vos, T. (1992). Tempo Test Rekenen (TTR) [Arithmetic Number Fact Test]. *Nijmegen: Berkhout*.

Diamond, A. (2013). Executive Functions. *The Annual Review of Psychology, 64*, 135–168.

Durand, M., Hulme, C., Larkin, R., & Snowling, M., (2005). The cognitive foundations of reading and arithmetic skills in 7- to 10-year-olds. *Journal of Experimental Child Psychology, 91(2)*, 113-136.

Egner, T., & Hirsch, J. (2005). Cognitive control mechanisms resolve conflict through cortical amplification of task-relevant information. *Nature neuroscience, 8*(12), 1784-1790.

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & psychophysics*, *16*(1), 143-149.

Friedman, David, Nessler, Doreen, Cycowicz, Yael M, & Horton, Cort. (2009). Development of and Change in Cognitive Control: A Comparison of Children, Young and Older Adults. *Cognitive, Affective, & Behavioral Neuroscience, 9*(1), 91-102.

Gebuis, T., & Reynvoet, B. (2011). Generating nonsymbolic number stimuli. *Behavior research methods*, *43*(4), 981-986.

Gerofsky, P. R. (2015). Why Asian Preschool Children Mathematically Outperform Preschool Children from Other Countries. *Western Undergraduate Psychology Journal, 3*(1).

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. (1998). Measuring individual differences in implicit cognition: the implicit association test. *Journal of personality and social psychology*, *74*(6), 1464.

Henik, A., & Tzelgov, J. (1982). Is three greater than five: The relation between physical and semantic size in comparison tasks. *Memory & cognition, 10*(4), 389-395.

Hopko, D. R., Mahadevan, R., Bare, R. L., & Hunt, M. K. (2003). The abbreviated math anxiety scale (AMAS) construction, validity, and reliability. *Assessment, 10*(2), 178-182.

Kubota, M., Hadley, L. V., Schaeffner, S., Könen, T., Meaney, J. A., Auyeung, B., Morey, C. C., Karbach, J., & Chevalier, N. (2020). Consistent use of proactive control and relation with academic achievement in childhood. *Cognition, 203*, 104329.

Lambert, K., & Moeller, K. (2019). Place-value computation in children with mathematics difficulties. *Journal of experimental child psychology, 178*, 214-225.

Landerl, K, & Kölle, C. (2009). Typical and atypical development of basic numerical skills in elementary school. *Journal of Experimental Child Psychology, 103(4)*, 546-565.

Landerl, K. (2013). Development of numerical processing in children with typical and dyscalculic arithmetic skills-a longitudinal study. *Frontiers in Psychology, 4*, 459.

Lehle, C., & Hübner, R. (2008). On-the-fly adaptation of selectivity in the flanker task. *Psychonomic Bulletin & Review, 15*(4), 814-818.

Logan, G. D., & Zbrodoff, N. J. (1979). When it helps to be misled: Facilitative effects of increasing the frequency of conflicting stimuli in a Stroop-like task. *Memory & cognition*, *7*(3), 166-174.

Macizo, P., & Herrera, A. (2013). The processing of Arabic numbers is under cognitive control. *Psychological Research, 77*(5), 651-658.

Mann, A., Moeller, K., Pixner, S., Kaufmann, L., & Nuerk, H. (2012). On the development of Arabic three-digit number processing in primary school children. Journal of Experimental Child Psychology, 113(4), 594-601.

Marcruz, O. Y. L., Carrie, H. K. L., Manabu, K., Mayumi, T., & Kumpei, M. (2020). Understanding of base-10 concept and its application: a cross-cultural comparison between Japan and Singapore. *International Journal of Early Years Education*, 1-15.

Marini, F., Demeter, E., Roberts, K. C., Chelazzi, L., & Woldorff, M. G. (2016). Orchestrating proactive and reactive mechanisms for filtering distracting information: Brain-behavior relationships revealed by a mixed-design fMRI study. *Journal of Neuroscience, 36*(3), 988-1000.

Mayr, U., Awh, E., & Laurey, P. (2003). Conflict adaptation effects in the absence of executive control. *Nature neuroscience, 6*(5), 450-452.

Moeller, K., Pixner, S., Zuber, J., Kaufmann, L., & Nuerk, H. C. (2011). Early place-value understanding as a precursor for later arithmetic performance—A longitudinal study on numerical development. *Research in developmental disabilities, 32*(5), 1837-1851.

Munakata, Y., Snyder, H. R., & Chatham, C. H. (2012). Developing cognitive control: Three key transitions. *Current directions in psychological science*, *21*(2), 71-77.

Niebaum, J., Chevalier, N., Guild, R. M., & Munakata, Y. (2020). Developing Adaptive Control: Age-related Differences in Task Choices and Awareness of Proactive and Reactive Control Demands. *Cognitive, Affective, & Behavioral Neuroscience*, 2020-10-02.

Nuerk, H., Weger, U., & Willmes, K. (2001). Decade breaks in the mental number line? Putting the tens and units back in different bins. *Cognition*, 82(1), B25-B33.

Nuerk, H., Kaufmann, L., Zoppoth, S., & Willmes, K. (2004). On the Development of the Mental Number Line: More, Less, or Never Holistic With Increasing Age? *Developmental Psychology, 40(6)*, 1199-1211.

Nuerk, H.-C., Moeller, K., Klein, E., Willmes, K., & Fischer, M. H. (2011). Extending the Mental Number Line. Zeitschrift Für Psychologie, 219(1), 3-22.

Ramirez, G., Chang, H., Maloney, E. A., Levine, S. C., & Beilock, S. L. (2016). On the relationship between math anxiety and math achievement in early elementary school: The role of problem solving strategies. *Journal of Experimental Child Psychology, 141, 83–100*.

Rueda, M. R., Fan, J., McCandliss, B. D., Halparin, J. D., Gruber, D. B., Lercari, L. P., & Posner, M. I. (2004). Development of attentional networks in childhood. *Neuropsychologia, 42*(8), 1029-1040.

Schmidt, J. R. (2013). Questioning conflict adaptation: Proportion congruent and Gratton effects reconsidered. *Psychonomic bulletin & review, 20*(4), 615-630.

Schmidt, J. R., & Besner, D. (2008). The Stroop effect: why proportion congruent has nothing to do with congruency and everything to do with contingency. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*(3), 514.

Spinelli, Giacomo, Perry, Jason R, & Lupker, Stephen J. (2019). Adaptation to Conflict Frequency Without Contingency and Temporal Learning: Evidence From the Picture-Word Interference Task. *Journal of Experimental Psychology. Human Perception and Performance, 45*(8), 995-1014.

Surrey, C., Kretschmer-Trendowicz, A., Altgassen, M., & Fischer, R. (2019). Contextual recruitment of cognitive control in preadolescent children and young adults. *Journal of experimental child psychology, 183,* 189-207.

Vos, H., Sasanguie, D., Gevers, W., & Reynvoet, B. (2017). The role of general and number-specific order processing in adults' arithmetic performance. *Journal of Cognitive Psychology*, 29(4), 469-482.

Waxer, M., & Morton, J. B. (2011). The development of future-oriented control: An electrophysiological investigation. *NeuroImage*, 56, 1648– 1654.

Wilk, H.A., & Morton, J.B. (2012). Developmental changes in patterns of brain activity associated with moment-to-moment adjustments in control. *NeuroImage (Orlando, Fla.), 63*(1), 475-484.

Yuan, L., Prather, R. W., Mix, K. S., & Smith, L. B. (2019). Preschoolers and multi-digit numbers: A path to mathematics through the symbols themselves. *Cognition, 189,* 89-104.

Appendix A. Stimuli that were used in the number comparison task

Block	Congruency	Above Number	Below Number	Correct response
MI	Congruent	42	89	n
MI	Congruent	51	98	n
MI	Congruent	67	21	j
MI	Congruent	23	79	n
MI	Congruent	23	89	n
MI	Congruent	98	47	j
MI	Congruent	84	23	j
MI	Congruent	31	92	n
MI	Congruent	97	36	j
MI	Congruent	96	25	j
MI	Congruent	32	59	n
MI	Congruent	58	41	j
MI	Congruent	68	51	j
MI	Congruent	89	62	j
MI	Congruent	31	47	n
MI	Congruent	62	93	n
MI	Congruent	37	25	j
MI	Congruent	36	48	n
MI	Congruent	98	53	j
MI	Congruent	52	98	n
MI	Incongruent	71	29	j
MI	Incongruent	81	29	j
MI	Incongruent	39	81	n
MI	Incongruent	28	91	n
MI	Incongruent	91	38	j
MI	Incongruent	39	82	n
MI	Incongruent	91	48	j
MI	Incongruent	91	27	j
MI	Incongruent	29	73	n
MI	Incongruent	29	83	n
MI	Incongruent	37	81	n
MI	Incongruent	91	37	j
MI	Incongruent	26	91	n
MI	Incongruent	28	93	n
MI	Incongruent	74	29	j
MI	Incongruent	84	29	j
MI	Incongruent	81	36	j
MI	Incongruent	91	36	j
MI	Incongruent	37	82	n
MI	Incongruent	37	92	n
MI	Incongruent	39	84	n
MI	Incongruent	91	46	j
MI	Incongruent	47	92	n
MI	Incongruent	48	93	n

MI	Incongruent	92	38	j
MI	Incongruent	47	91	n
MI	Incongruent	26	81	n
MI	Incongruent	25	91	n
MI	Incongruent	91	23	j
MI	Incongruent	27	95	n
MI	Incongruent	83	75	j
MI	Incongruent	87	95	n
MI	Incongruent	35	28	j
MI	Incongruent	49	56	n
MI	Incongruent	85	92	n
MI	Incongruent	78	94	n
MI	Incongruent	58	72	n
MI	Incongruent	74	59	j
MI	Incongruent	48	63	n
MI	Incongruent	72	56	j
MI	Incongruent	29	45	n
MI	Incongruent	52	36	j
MI	Incongruent	28	67	n
MI	Incongruent	74	35	j
MI	Incongruent	91	52	j
MI	Incongruent	24	73	n
MI	Incongruent	81	32	j
MI	Incongruent	43	92	n
MI	Incongruent	26	85	n
MI	Incongruent	96	37	j
MI	Incongruent	94	25	j
MI	Incongruent	28	97	n
MI	Incongruent	61	23	j
MI	Incongruent	87	49	j
MI	Incongruent	61	39	j
MI	Incongruent	49	61	n
MI	Incongruent	81	59	j
MI	Incongruent	28	41	n
MI	Incongruent	38	61	n
MI	Incongruent	52	39	j
MI	Incongruent	48	71	n
MI	Incongruent	72	59	j
MI	Incongruent	68	91	n
MI	Incongruent	27	51	n
MI	Incongruent	29	43	n
MI	Incongruent	53	29	j
MI	Incongruent	34	25	j
MI	Incongruent	41	32	j
MI	Incongruent	67	58	j
MI	Incongruent	76	85	n
MI	Incongruent	46	27	j
MI	Incongruent	39	58	n

MI	Incongruent	78	59	j
MI	Incongruent	74	93	n
MI	Incongruent	64	35	j
MI	Incongruent	83	54	j
MI	Incongruent	97	68	j
MI	Incongruent	61	53	j
MI	Incongruent	71	26	j
MI	Incongruent	25	81	n
MC	Congruent	69	21	j
MC	Congruent	79	21	j
MC	Congruent	31	79	n
MC	Congruent	78	21	j
MC	Congruent	79	32	j
MC	Congruent	42	89	n
MC	Congruent	98	51	j
MC	Congruent	21	67	n
MC	Congruent	23	79	n
MC	Congruent	89	23	j
MC	Congruent	78	32	j
MC	Congruent	87	41	j
MC	Congruent	52	98	n
MC	Congruent	96	31	j
MC	Congruent	32	87	n
MC	Congruent	79	34	j
MC	Congruent	41	86	n
MC	Congruent	42	87	n
MC	Congruent	51	96	n
MC	Congruent	53	98	n
MC	Congruent	21	65	n
MC	Congruent	21	75	n
MC	Congruent	23	64	n
MC	Congruent	34	75	n
MC	Congruent	51	92	n
MC	Congruent	24	75	n
MC	Congruent	92	41	j
MC	Congruent	98	47	j
MC	Congruent	23	84	n
MC	Congruent	31	92	n
MC	Congruent	36	97	n
MC	Congruent	25	96	n
MC	Congruent	27	98	n
MC	Congruent	27	69	n
MC	Congruent	83	41	j
MC	Congruent	94	52	j
MC	Congruent	73	21	j
MC	Congruent	39	21	j
MC	Congruent	41	59	n
MC	Congruent	79	51	j

MC	Congruent	38	21	j
MC	Congruent	31	58	n
MC	Congruent	32	59	n
MC	Congruent	41	58	n
MC	Congruent	68	51	j
MC	Congruent	89	62	j
MC	Congruent	47	31	j
MC	Congruent	32	58	n
MC	Congruent	42	58	n
MC	Congruent	79	53	j
MC	Congruent	23	48	n
MC	Congruent	59	34	j
MC	Congruent	96	71	j
MC	Congruent	25	49	n
MC	Congruent	58	34	j
MC	Congruent	43	67	n
MC	Congruent	24	35	n
MC	Congruent	31	42	n
MC	Congruent	58	47	j
MC	Congruent	83	94	n
MC	Congruent	59	38	j
MC	Congruent	74	53	j
MC	Congruent	96	75	j
MC	Congruent	26	57	n
MC	Congruent	72	41	j
MC	Congruent	93	62	j
MC	Congruent	25	37	n
MC	Congruent	48	36	j
MC	Congruent	43	21	j
MC	Congruent	54	32	j
MC	Congruent	98	76	j
MC	Congruent	59	27	j
MC	Congruent	75	43	j
MC	Congruent	65	97	n
MC	Congruent	26	49	n
MC	Congruent	45	68	n
MC	Congruent	87	64	j
MC	Congruent	56	23	j
MC	Congruent	34	67	n
MC	Congruent	95	62	j
MC	Incongruent	29	71	n
MC	Incongruent	81	29	j
MC	Incongruent	81	39	j
MC	Incongruent	28	91	n
MC	Incongruent	91	38	j
MC	Incongruent	28	67	n
MC	Incongruent	74	35	j
MC	Incongruent	52	91	n

MC	Incongruent	73	24	j
MC	Incongruent	32	81	n
MC	Incongruent	61	39	j
MC	Incongruent	61	49	j
MC	Incongruent	59	81	n
MC	Incongruent	28	41	n
MC	Incongruent	38	61	n
MC	Incongruent	34	25	j
MC	Incongruent	32	41	n
MC	Incongruent	67	58	j
MC	Incongruent	76	85	n
MC	Incongruent	46	27	j

Appendix B. Overview of low-level learning mechanisms that can alternatively explain the results

Feature repetitions (Mayr, Awh, & Laurey, 2003). Our flanker task only existed of 4 different stimuli (2 congruent and 2 incongruent) and only 2 responses could be given (left or right), a substantial part of the trials were complete stimulus-response repetitions (e.g., <<><< followed again by <<><<). This implies that reliance on low-level features as an alternative explanation for our results would have been likely. For this reason, we have excluded all complete repetitions from the analyses, thereby decreasing the role that low-level features played in the remaining data. Even when excluding these complete repetitions, significant congruency and proportion congruency effects were observed in the reaction time data for all age groups. This finding is in contrast to the work of Mayr, Awh, and Laurey (2003) who argued that conflict adaptation effects were mainly driven by stimulus-response repetition effects. However, even when all complete trial repetitions are removed from the analysis, partial response repetitions are still present in the data (e.g., >>>>> followed by <<><< both requiring the same response). In the example above, the response on the previous trial might speed up the response on the next trial if this response is identical, without the need to exert cognitive control on that second (incongruent) trial.

Contingency learning (Schmidt & Besner, 2008). Furthermore, another alternative mechanism which could account for conflict adaptation effects, is contingency learning. This account stems from item-specific-proportion congruency, where the proportion congruency of a single stimulus is manipulated throughout a list. For example, if in a Stroop task the word GREEN is presented in red 75% of the time, the word GREEN will be predictive of a red response. Over time, participants will respond faster to these incongruent trials, purely based on this learned contingency, which will decrease the congruency effect in mostly incongruent conditions. Of course, the formation of such S-R mappings is only possible when a stimuli is frequently mapped onto a certain response more than once (Schmidt, 2013). Contingency learning could therefore not have played a role in our number comparison task, given that each number pair was presented once. However, in our flanker task, contingency biases might have played a role. In the mostly incongruent conditions, 80% of the trials with flankers that point to the right will require a response to the left. Therefore, participants could form reversed stimulus-response mappings where they learn to press the button mapped to the response option opposite to the direction of the flankers. This strategy would be efficient, as it is rewarding in 80% of the trials. However, if this would be the unique low-level strategy that participants used, we would expect similar (and not reduced) congruency effects in MC and MI conditions, as we would expect that this strategy would also lead to slower reaction times and higher error rates for the congruent trials in the MI condition (given the learned reversed response mapping). Our findings don't show this pattern. Therefore, we would argue that contingency effects cannot fully explain our findings.

Temporal learning processes (Schmidt, 2013). A third potential alternative account could be temporal learning processes. When most trials in a list of trials are congruent, most trials can be quickly resolved because there is no conflict between the target and flankers. In this case, participants will have a rapid pace of responding until an incongruent trial is presented which cannot be resolved as easily. When the rapid response time threshold is exceeded on these incongruent trials, participants take their time to resolve the incongruent trial, leading to relatively slow response times. If most trials in a list are incongruent, the pace of responding will overall be slower because most of the trials require a conflict resolution. On the infrequent congruent trials, participants can react in a more relaxed manner because they can easily stay within the response time threshold. This temporal learning process is low-level because it is the response speed, and not the congruency per se, that determines the temporal expectancy. Participants are assumed to adapt to the time-on-task, rather than adapt to the list-wide proportion of conflict. However, the temporal learning hypothesis still involves a component for detection and resolution of conflict and therefore requires cognitive control (Spinelli, Perry & Lupker, 2019). Furthermore, research shows that the temporal learning effects are presumably caused by an inverse transformation of RT data (Cohen-Shikora, Suh & Bugg, 2019). The authors suggest analyzing data with generalized linear mixed-effect models which do not assume a normally distributed dependent variable. They showed that this approach to data analysis did not replicate decreased congruency effects as a result of longer RTs (Cohen-Shikora, Suh & Bugg, 2019). Other studies that control for temporal learning expectancies did not designate temporal learning account as a credible explanation for PC effects either (Spinelli, Perry & Lupker, 2019). Also note that this alterative account is based on speed of responding, and hence on RTs. It would have difficulty explaining why in our number comparison task, which is the main focus of this study, the PCEs are mainly observed in the error rate analysis. We cannot rule out the temporal learning account as an explanation for our results in both the flanker and the number comparison tasks, because our design does not facilitate controlling for temporal learning expectancies. However, recent literature suggest that conflict adaptation, rather than temporal learning, remains a credible explanation for list-wide proportion congruency effects (Cohen-Shikora, Suh & Bugg, 2019; Spinelli, Perry & Lupker, 2019).