Title:

Adaptiveness of fluctuations in intra-individual variability of performance is process-dependent in middle childhood

Running title:

Adaptiveness of intra-individual variability

Authors:

Roser Cañigueral[1], Keertana Ganesan[1], Claire R. Smid[1],

Abigail Thompson[1], Nico U. F. Dosenbach[2] & Nikolaus Steinbeis[1]

Affiliations:

[1]Department of Clinical, Educational and Health Psychology, University College London, 26 Bedford Way, London WC1H 0AP, United Kingdom

[2] Departments of Neurology, Pediatrics, Radiology and Biomedical Engineering, Washington University in St. Louis School of Medicine, 660 South Euclid Avenue, St. Louis, MO 63110, United Stated of America

Corresponding author:

Nikolaus Steinbeis

Department of Clinical, Educational and Health Psychology, University College London, 26 Bedford Way, London WC1H 0AP, United Kingdom

n.steinbeis@ucl.ac.uk

Word count: 6050

Abstract

There is ongoing debate on the relationship between intra-individual variability (IIV) of cognitive processes and task performance. While psychological research has traditionally assumed that lower intra-individual variability (IIV) aids consistent task performance, some studies suggest that greater IIV can also be adaptive, especially when flexible responding is required. Here we selectively manipulate two cognitive processes with differing task demands, response speed (Going; simple task) and inhibitory control (Stopping; complex task), by means of a training paradigm and assess how this impacts IIV and its relationship to task performance. A group of 208 6-13-year-old children were randomly allocated to an 8-week training targeting Going (control group) or Stopping (experimental group). The stop signal task was administered before and after training. Training Going led to adaptive *reductions* in Going IIV, which allows more consistent and efficient Going performance. In contrast, training Stopping led to adaptive *increases* in Stopping IIV, where greater flexibility in cognitive processing is required to meet higher task demands. These findings provide systematic and causal evidence of the process-dependent relationship of IIV and task performance in the context of Going and Stopping, suggesting a more nuanced perspective on IIV with implications for developmental, ageing and intervention studies.

Keywords: intra-individual variability; cognitive training; inhibitory control; response speed; middle childhood

Introduction

Fluctuations in performance are a hallmark of cognitive processing over the lifespan (Shalev, Bauer, and Nobre 2019), and previous research shows that intra-individual variability (IIV) measures of performance are more sensitive to developmental differences than conventional mean measures (Tamnes et al. 2012). This has resulted in a wealth of investigations using IIV measures of cognitive performance as markers of development-, ageing- and training-related changes (Tamnes et al. 2012; MacDonald, Hultsch, and Dixon 2003; Cherbuin, Sachdev, and Anstey 2010; Cubillo et al. 2022; Ram et al. 2005). The prevailing assumption is that *reductions* in IIV are adaptive because cognitive performance becomes more consistent and optimised (MacDonald, Nyberg, and Bäckman 2006; Williams et al. 2005; Cubillo et al. 2022; Ram et al. 2005). However, some studies also report an adaptive role of IIV *increases*, suggesting that IIV may not consistently reflect the same phenomenon across cognitive domains or even within the same individual (Allaire and Marsiske 2005). Such a disparity in findings may be accounted for by differing task demands, where improving performance in simple tasks (e.g. tasks with predictable or deterministic responses) might require consistent responses, but becoming better at complex tasks (e.g. tasks with unpredictable responses) might instead require flexible cognitive processing (Li et al. 2001; MacDonald, Hultsch, and Dixon 2003). Importantly, training studies provide valuable insight into how IIV of different cognitive processes fluctuates when performance is causally manipulated to improve (von Bastian and Oberauer 2014). Here we use a training design and leverage the differing task demands of response speed (i.e. Going; simple task) and inhibitory control (i.e. Stopping; complex task) to investigate how IIV in these two cognitive processes is modulated in middle childhood when performance is causally improved.

Despite a longstanding tradition in psychology research of relying on mean levels of performance as the main outcome measure, there has been increased interest in IIV measures of accuracy and reaction times (Williams et al. 2005; Nesselroade 1991b; 1991a; Shalev, Bauer, and Nobre 2019; Thompson, Schel, and Steinbeis 2021). Hultsch and colleagues (Hultsch and MacDonald 2004; Hultsch, MacDonald, and Dixon 2002) describe two types of IIV: *dispersion*, which refers to within-person variability across different tasks at a single timepoint, and *inconsistency*, which refers to within-person fluctuations across trials or sessions of the same task. The latter form of IIV, particularly in relation to reaction times (RT), will be the focus of the present study. IIV is increasingly recognised as a complementary source of information to mean measures rather than a source of noise or random error (Williams et al. 2005; Nesselroade 1991a). Moreover, IIV shows higher sensitivity than mean levels of performance as a marker of development (Tamnes et al. 2012), ageing (MacDonald, Hultsch, and Dixon 2003; Cherbuin, Sachdev, and Anstey 2010), and brain disorders (MacDonald, Nyberg, and Bäckman 2006). Changes in IIV have also been associated with changes in prefrontal brain structure and function, white matter integrity, and dopaminergic neuromodulation (MacDonald, Li, and Bäckman 2009; Tamnes et al. 2012; van Belle et al. 2015). However, although there has been much progress in unravelling the functional significance of IIV, it is not yet well understood whether IIV modulations reflect adaptive or maladaptive cognitive processing.

It is generally assumed that a *reduction* in behavioural IIV reflects more efficient cognitive performance and therefore is adaptive (Unsworth 2015), whereas increased IIV reflects lapses of attention and failure to maintain cognitive control (MacDonald, Li, and Bäckman 2009; West et al. 2002). This notion is supported by studies showing that IIV follows a U-shaped function over the lifespan, where IIV decreases from childhood into young adulthood (reflecting optimisation of cognitive processing) and increases again in the elderly (reflecting a decline in cognitive function) (Williams et al. 2005; MacDonald, Nyberg, and Bäckman 2006). A recent cognitive training study also shows that, after a working memory training, children show better accuracy and reduced IIV in working memory and selective attention tasks (Cubillo et al. 2022), consistent with the idea that training improves efficiency and stability in cognitive processing (von Bastian and Oberauer 2014). Similar findings have been reported in older adults, where repeated practice of memory speed tasks results in IIV reductions (Ram et al. 2005). Moreover, correlational studies show that lower IIV is associated with better task performance (MacDonald, Hultsch, and Dixon 2003; Rabbitt et al. 2001), supporting the idea that reductions in IIV are adaptive.

However, some studies suggest that *increases* in IIV can also be adaptive, as they could reflect flexible cognitive processing in response to changes in the environment (Li, Huxhold, and Schmiedek 2004). For instance, increased IIV during childhood is key for learning, as it allows the testing and acquisition of new strategies that eventually lead to positive development (Allaire and Marsiske 2005; Siegler 1994; Nussenbaum and Hartley 2019). In line with this, it was found that children show especially variable behaviour on trials immediately before discovering a new strategy, as well as on the trial where the new strategy is discovered (Siegler and Jenkins 1989). Increased IIV is also observed when performing tasks with a higher level of cognitive demand or tasks that allow room for improvement: Garrett and colleagues (Garrett, McIntosh, and Grady 2014) used a face-matching task and found that as task difficulty gradually increased so did IIV levels, likely reflecting that participants were testing new strategies to overcome increased task demands. Further, correlational analyses show that, particularly for tasks where strategy use plays a central role (e.g. spatial memory tasks with increasing difficulty levels), increased IIV is associated with better task performance (Li et al. 2001).

Altogether, the studies reported above shed some light onto the question of when IIV is adaptive or maladaptive. However, they also point out the complexity of defining the functional role of IIV at different lifespan stages and within different cognitive domains. In this sense, training studies hold the potential to provide rich insight into how IIV of different cognitive processes is modulated when performance is causally manipulated to become more efficient and stable (von Bastian and Oberauer 2014; Cubillo et al. 2022). Here, we aimed to address this question in the context of response speed and inhibitory control during middle childhood. Response speed, or Going, refers to the cognitive ability of promptly responding to a stimulus, and reflects the speed in which individuals can sense, perceive, understand and respond to new information (Silva and Lee 2021). Inhibitory control, or Stopping, refers to the cognitive ability of suppressing impulsive or habituated responses to achieve long term goals (Diamond 2013). Importantly, while Going can be considered a simple task (i.e. requires a simple motor response), Stopping is arguably more complex since it requires the initiation of a response followed by its subsequent inhibition. In the context of the stop signal task, the level of task complexity during Going and Stopping is further defined by their predictability: while participants are always presented with a go signal (predictable), they are randomly presented with a stop signal in a subset of trials (unpredictable). These two cognitive processes thus offer the possibility to contrast how IIV is modulated by training across domains with different levels of cognitive demand. Further, Going abilities during childhood have been related to positive academic outcomes (Geary 2010), while Stopping abilities predict positive cognitive and socio-emotional development (Moffitt et al. 2011). Moreover, both Going and Stopping abilities show protracted development, with marked qualitative and quantitative improvements during childhood (Luna, Padmanabhan, and O'Hearn 2010; Durston et al. 2002; Geary 2010; Kail 1991). The potential for malleability, together with their positive impact on wellbeing, makes Going and Stopping excellent candidates to investigate training effects during childhood.

Previous studies measuring IIV in RTs often rely on conventional variability measures that assume a Gaussian distribution in RTs (e.g. standard deviation or coefficient of variation). However, because RTs are positively skewed, they are more closely fitted by ex-Gaussian distributions: by combining parameters from the Gaussian and exponential distribution, ex-Gaussian distributions offer a much finer level of analysis with greater interpretative power than conventional measures (Luce 1986; McAuley et al. 2006; Matzke et al. 2013; Matzke, Love, and Heathcote 2017). In particular, they generate three parameters of interest: the mu parameter (mean of the Gaussian distribution) reflects average processing speed; the sigma parameter (standard deviation of the Gaussian distribution) reflects variability in processing speed; the tau parameter (mean and standard deviation of the exponential distribution, i.e. tail of the distribution) reflects the degree and variability of occasional extremely slow responses (i.e. extremely slow processing speed), and has been linked to attentional lapses and transient periods of inefficient task performance (Hervey et al. 2006; Karalunas et al. 2014; West et al. 2002). Here we employed ex-Gaussian parameters from the go and stop RT distributions to examine how IIV in Going and Stopping responses is modulated by training.

The present study aimed to study how training modulates IIV of two cognitive processes with differing task demands (Going and Stopping) in middle childhood. Six- to thirteen-year-old children underwent an 8-week inhibitory control (experimental group; stop signal task) or response speed (control group; reaction time task) training, and additionally completed the stop signal task before training (T0), immediately after training (T1), and one-year after training (T2). An ex-Gaussian approach was used to generate mean (mu) and IIV (sigma, tau) measures of Stopping and Going responses at T0, T1 and T2; Gaussian mean and standard deviation (SD) measures were generated for the training data. To establish how task accuracy and response features (mean processing speed and IIV) are associated during Going and Stopping, we first tested the relation between these measures at T0 separately for each process of interest. In line with the general assumption that reductions in behavioural IIV reflect more efficient cognitive performance, we expected that greater Going accuracy (probability of hit, pHit) would be related to faster and less variable go responses. However, because the inhibitory control training is more cognitively demanding, we predicted that greater Stopping accuracy (probability of correctly Stopping, pStop) would be related to faster and more variable stop responses. In line with these hypotheses, we also expected that, after the response speed training, the control group would show more accurate, faster and *less* variable Going performance, reflecting more efficient cognitive processing; instead, after the inhibitory control training, we hypothesised the experimental group would show more accurate, faster and *more* variable Stopping performance, reflecting more flexible cognitive processing. Note also that we did not have specific predictions about how Going performance would be modulated by the inhibitory control training in the experimental group, or about how Stopping performance would be modulated by the response speed training in the control group. Finally, we hypothesised that training-related changes would be maintained at T2, and that they would be further supported by similar modulations of Going and Stopping responses over the training weeks.

Results

Correlations at T0 between accuracy and RT parameters

We first tested how response accuracy is related to RT parameters (mu, sigma, tau) at T0 for Going and Stopping separately. When testing effects on Going measures (GoRT), there was a negative correlation between mu and pHit (b = -.338, $t(187.1) = 4.058$, $p < .001$) (Figure 1A), a negative correlation between sigma and pHit (b = -.104, $t(187.6) = 4.612$, $p < .001$) (Figure 1B), and no correlation between tau and pHit (b = -.008, $t(171.6) = .619$, $p = .536$) (Figure 1C), indicating that greater Going accuracy was related to faster and *less* variable Going performance. When testing effects on Stopping measures (SSRT), there was a negative correlation between mu and pStop (b = -.042, $t(198.4) = 3.970$, $p < .001$) (Figure 1D), a positive correlation between sigma and pStop (b = .006, $t(189.0) = 3.345$, $p < .001$) (Figure 1E), and no correlation between tau and pStop (b = -.003, $t(196.9) = 1.647$, $p = .101$) (Figure 1F), indicating that greater Stopping accuracy was related to faster and *more* variable Stopping performance.

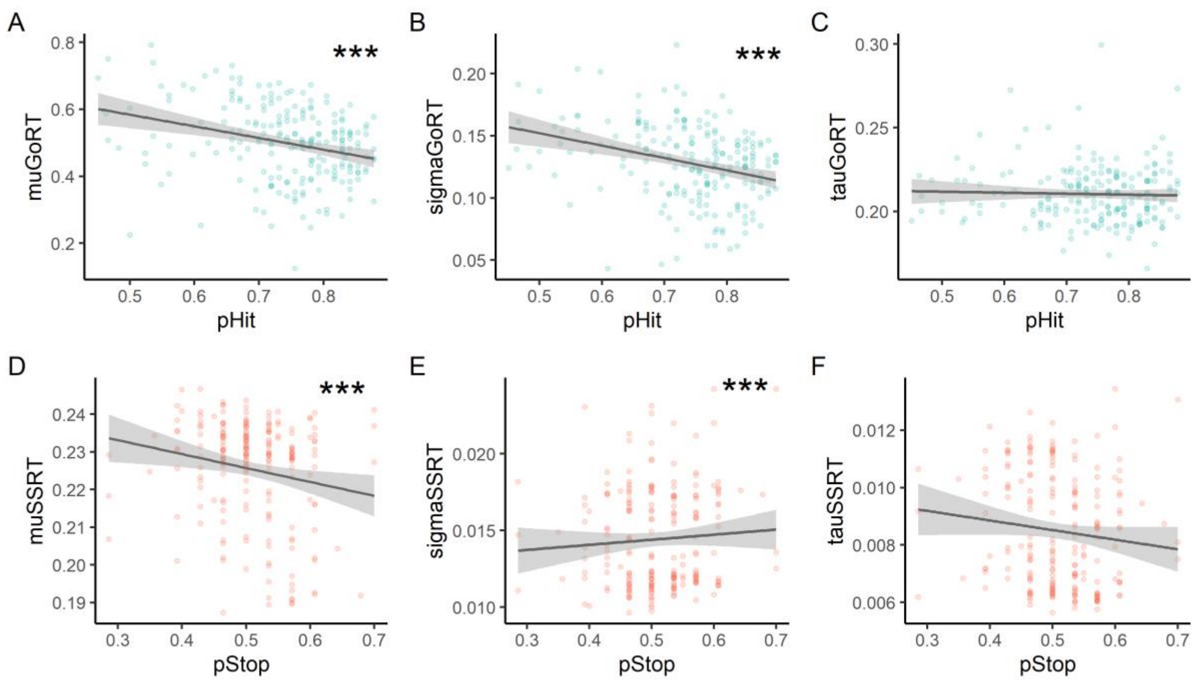

Figure 1. Scatterplots for correlations between accuracy and RT parameters for Going (blue; top row) and Stopping (red; bottom row) at T0 (before training). A) muGoRT and pHit. B)

sigmaGoRT and pHit. C) tauGoRT and pHit. D) muSSRT and pStop. E) sigmaSSRT and pStop. F) tauSSRT and pStop.

Training effects on Going

We tested the effects of training on pHit and GoRT parameters (mu, sigma and tau) (see Supplementary Materials S1, Table S1-1).

pHit (accuracy)

For pHit (Figure 2A), there was no main effect of Timepoint ($F(2,412) = 2.332$, $p = .098$), no main effect of Group ($F(1,206) = 3.292$, $p = .071$), but there was an interaction effect between Timepoint and Group ($F(2,412) = 4.716$, $p = .009$). Post-hoc pairwise comparisons showed that, in the control group, there were no differences across timepoints (T0-T1: $t(412) = .743$, $p = 1.000$, $d_z = .105$; T0-T2: $t(412) = .409$, $p = 1.000$, $d_z = .058$; T1-T2: $t(412) = .334$, $p = 1.000$, $d_z = .093$); in the experimental group, children were less accurate at T1 compared to T0 ($t(412) = 3.311$, $p = .015$, $d_z = .453$) and T2 ($t(412) = 3.147$, $p = .026$, $d_z = .430$), and there were no differences between T0 and T2 ($t(412) = .164$, $p = 1.000$, $d_z = .022$). Moreover, at T0 and T2 there were no differences between control and experimental groups (T0: $t(593) = .224$, $p = 1.000$, $d_z = .034$; T2: $t(593) = .309$, $p = 1.000$, $d_z = .046$), while at T1 the control group was more accurate than the experimental group ($t(593) = 3.487$, $p = .008$, $d_z = .524$).

Mu (mean processing speed)

For mu (Figure 2B), there was a main effect of Timepoint ($F(2,412) = 84.11$, $p < .001$), a main effect of Group ($F(1,206) = 183.8$, $p < .001$), and an interaction effect between Timepoint and Group ($F(2,412) = 69.13$, $p < .001$). Post-hoc pairwise comparisons showed that, in the control group, performance was faster at T1 compared to T0 ($t(412) = 4.511$, $p < .001$, $d_z = .635$) and

T2 ($t(412) = 8.965$, $p < .001$, $d_z = 1.262$), and slower at T2 compared to T0 ($t(412) = 4.454$, $p < .001$, $d_z = .627$); in the experimental group children performed slower at T1 compared to T0 ($t(412) = 12.19$, $p < .001$, $d_z = 1.668$), and at T2 compared to T0 ($t(412) = 13.92$, $p < .001$, $d_z = 1.904$). There were no differences between T1 and T2 ($t(412) = 1.725$, $p = 1.000$, $d_z = .236$). Moreover, at T0 there were no differences between control and experimental groups ($t(532) = 2.537$, $p = .172$, $d_z = .416$), while at T1 and T2 the control group performed faster than the experimental group (T1: $t(532) = 16.57$, $p < .001$, $d_z = 2.719$; T2: $t(532) = 10.32$, $p < .001$, $d_z = 1.693$).

Sigma (variability in processing speed)

For sigma (Figure 2C), there was no main effect of Timepoint ($F(2,412) = .354$, $p = .702$), but there was a main effect of Group ($F(1,206) = 168.8$, $p < .001$) and an interaction effect between Timepoint and Group ($F(2,412) = 68.88$, $p < .001$). Post-hoc pairwise comparisons showed that, in the control group, performance was less variable at T1 compared to T0 ($t(412) = 8.331$, $p < .001$, $d_z = 1.172$) and T2 ($t(412) = 5.304$, $p < .001$, $d_z = .746$), and less variable at T2 compared to T0 ($t(412) = 3.027$, $p = .039$, $d_z = .426$); in the experimental group performance was more variable at T1 compared to T0 ($t(412) = 8.197$, $p < .001$, $d_z = 1.121$) and T2 ($t(412) = 4.277$, $p < .001$, $d_z = .585$), and at T2 compared to T0 ($t(412) = 3.920$, $p = .002$, $d_z = .536$). Moreover, at T0 there were no differences between control and experimental groups ($t(586) = 1.496$, $p = 1.000$, $d_z = .227$), while at T1 and T2 performance for the control group was less variable than for the experimental group (T1: $t(586) = 16.61$, $p < .001$, $d_z = 2.520$; T2: $t(586) = 7.834$, $p < .001$, $d_z = 1.189$).

<u>Tau (degree and variability of occasional extremely slow responses)</u>

For tau (Figure 2D), there was a main effect of Timepoint ($F(2,412) = 952.8$, $p < .001$), a main effect of Group ($F(1,206) = 2500.2$, $p < .001$), and an interaction effect between Timepoint and Group ($F(2,412) = 453.4$, $p < .001$). Post-hoc pairwise comparisons showed that, in the control group, there were no differences between T0 and T1 ($t(412) = 2.117$, $p = .522$, $d_z = .298$), but the degree and variability in extremely slow responses was lower at T2 compared to T0 ($t(412) = 12.44$, $p < .001$, $d_z = 1.751$) and T1 ($t(412) = 10.32$, $p < .001$, $d_z = 1.453$); in the experimental group the degree and variability in extremely slow responses was lower at T1 compared to T0 ($t(412) = 41.57$, $p < .001$, $d_z = 5.683$), and lower at T2 compared to T0 ($t(412) = 47.89$, $p < .001$, $d_z = 6.548$) and T1 ($t(412) = 6.325$, $p < .001$, $d_z = .865$). Moreover, at all timepoints the degree and variability in extremely slow responses was greater for the control group than for the experimental group (T0: $t(615) = 6.516$, $p < .001$, $d_z = .929$; T1: $t(615) = 44.31$, $p < .001$, $d_z = 6.314$; T2: $t(615) = 40.18$, $p < .001$, $d_z = 5.725$).

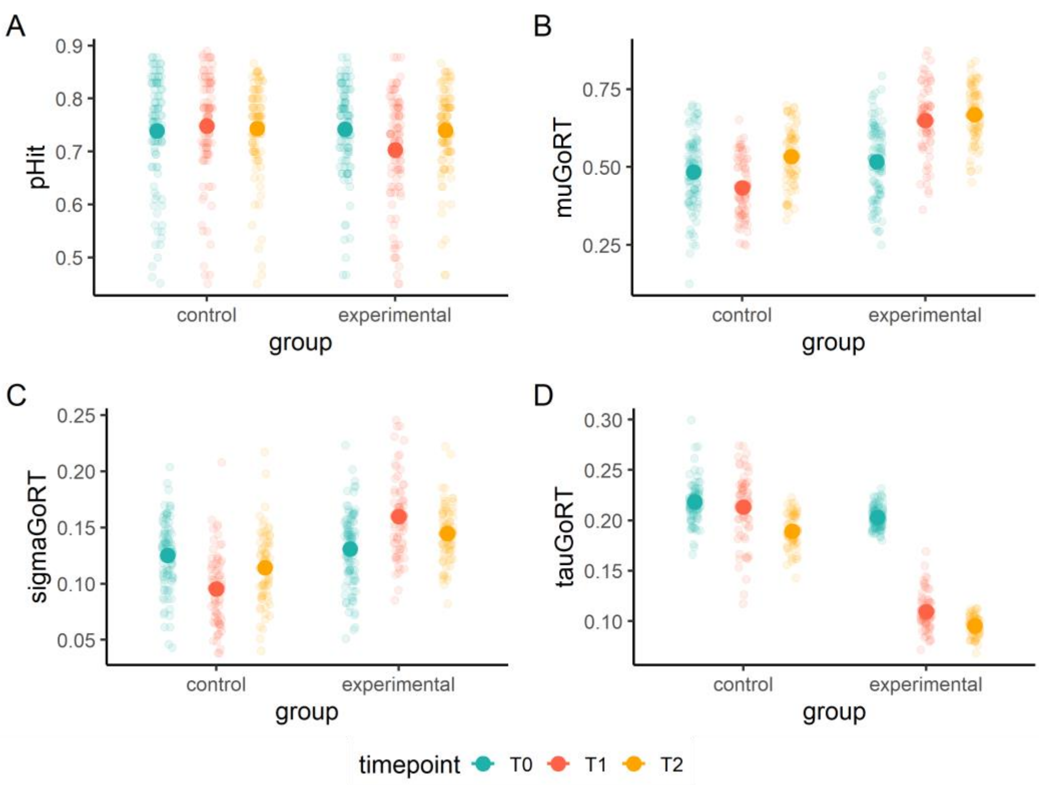

Figure 2. Plots for go accuracy and GoRT measures at T0 (before training), T1 (after training) and T2 (one-year follow-up) for the control and experimental groups: estimated marginal mean (filled circle) and raw individual datapoints (shaded circles). A) pHit. B) Mu. C) Sigma. D) Tau.

Training effects on Stopping

We tested the effects of training on pStop and SSRT parameters (mu, sigma and tau) (see Supplementary Materials S1, Table S1-2).

<u>pStop (accuracy)</u>

For pStop (Figure 3A), there was a main effect of Timepoint ($F(2,412) = 6.030$, $p = .003$), a main effect of Group ($F(1, 206) = 30.59$, $p < .001$), and an interaction effect between Timepoint and Group ($F(2,412) = 9.438$, $p < .001$). Post-hoc pairwise comparisons showed that, in the control group, there were no differences across timepoints (T0-T1: $t(412) = 1.423$, $p = 1.000$, $d_z = .200$; T0-T2: $t(412) = 1.123$, $p = 1.000$, $d_z = .158$; T1-T2: $t(412) = 2.546$, $p = 1.000$, $d_z = .358$); in the experimental group, children were more accurate at T1 compared to T0 ($t(412) = 4.740$, $p < .001$, $d_z = .648$), at T2 compared to T0 ($t(412) = 3.735$, $p = .003$, $d_z = .510$), and there were no differences between T1 and T2 ($t(412) = 1.005$, $p = 1.000$, $d_z = .137$). Moreover, at T0 there were no differences between control and experimental groups ($t(604) = .800$, $p = 1.000$, $d_z = .118$), while at T1 and T2 the experimental group was more accurate than the control group (T1: $t(604) = 6.572$, $p < .001$, $d_z = .966$; T2: $t(604) = 3.200$, $p = .022$, $d_z = .470$).

<u>Mu (mean processing speed)</u>

For mu (Figure 3B), there was a main effect of Timepoint ($F(2,412) = 88.19$, $p < .001$), no main effect of Group ($F(1,206) = .608$, $p = .436$), and an interaction effect between Timepoint and

Group ($F(2,412) = 195.4$, $p < .001$). Post-hoc pairwise comparisons showed that, in the control group, children performed slower at T1 compared to T0 ($t(412) = 12.12$, $p < .001$, $d_z = 1.706$) and at T2 compared to T0 ($t(412) = 11.14$, $p < .001$, $d_z = 1.568$), but there were no differences between T1 and T2 ($t(412) = .984$, $p = 1.000$, $d_z = .138$); in the experimental group children performed faster at T1 compared to T0 ($t(412) = 15.06$, $p < .001$, $d_z = 2.058$) and T2 ($t(412) = 18.65$, $p < .001$, $d_z = 2.550$), and slower at T2 compared to T0 ($t(412) = 3.598$, $p = .005$, $d_z = .492$). Moreover, at T0 and T2 the experimental group performed slower than the control group (T0: $t(613) = 10.77$, $p < .001$, $d_z = 1.545$; T2: $t(613) = 3.270$, $p = .017$, $d_z = .469$), while at T1 the experimental group performed faster than the control group ($t(613) = 15.48$ $p < .001$, $d_z = 2.219$).

Sigma (variability in processing speed)

For sigma (Figure 3C), there was a main effect of Timepoint ($F(2,412) = 217.7$, $p < .001$), a main effect of Group ($F(1,206) = 10.03$, $p = .002$), and an interaction effect between Timepoint and Group ($F(2,412) = 574.8$, $p < .001$). Post-hoc pairwise comparisons showed that, in the control group, performance was less variable at T1 compared to T0 ($t(412) = 9.757$, $p < .001$, $d_z = 1.373$) and T2 ($t(412) = 9.179$, $p < .001$, $d_z = 1.292$), but there were no differences between T0 and T2 ($t(412) = .577$, $p = 1.000$, $d_z = .081$); in the experimental group performance was more variable at T1 compared to T0 ($t(412) = 38.60$, $p < .001$, $d_z = 5.277$) and T2 ($t(412) = 15.92$, $p < .001$, $d_z = 2.177$), and at T2 compared to T0 ($t(412) = 22.68$, $p < .001$, $d_z = 3.100$). Moreover, at T0 performance for the experimental group was less variable than for the control group ($t(604) = 24.33$, $p < .001$, $d_z = 3.573$), at T1 performance for the experimental group was more variable than for the control group ($t(604) = 20.95$, $p < .001$, $d_z = 3.077$), and at T2 there were no differences between groups ($t(604) = 2.668$, $p = .118$, $d_z = .392$).

Tau (degree and variability of occasional extremely slow responses)

For tau (Figure 3D), there was a main effect of Timepoint (F(2,412) = 436.5, p < .001), a main effect of Group (F(1,206) = 3.905, p = .049), and an interaction effect between Timepoint and Group (F(2,412) = 532.5, p < .001). Post-hoc pairwise comparisons showed that, in the control group, the degree and variability in extremely slow responses was lower at T1 compared to T0 (t(412) = 3.384, p = .012, d_z = .476) and T2 (t(412) = 7.754, p < .001, d_z = 1.091), and greater at T2 compared to T0 (t(412) = 4.370, p < .001, d_z = .615); in the experimental group the degree and variability in extremely slow responses was greater at T1 compared to T0 (t(412) = 43.35, p < .001, d_z = 5.927) and T2 (t(412) = 15.40, p < .001, d_z = 2.106), and at T2 compared to T0 (t(412) = 27.95, p < .001, d_z = 3.821). Moreover, at T0 the degree and variability in extremely slow responses was lower for the experimental group than for the control group (t(608) = 20.75, p < .001, d_z = 3.022), at T1 it was greater for the experimental group than for the control group (t(608) = 23.21, p < .001, d_z = 3.381), and at T2 there were no differences between groups (t(608) = 1.262, p = 1.000, d_z = .184).

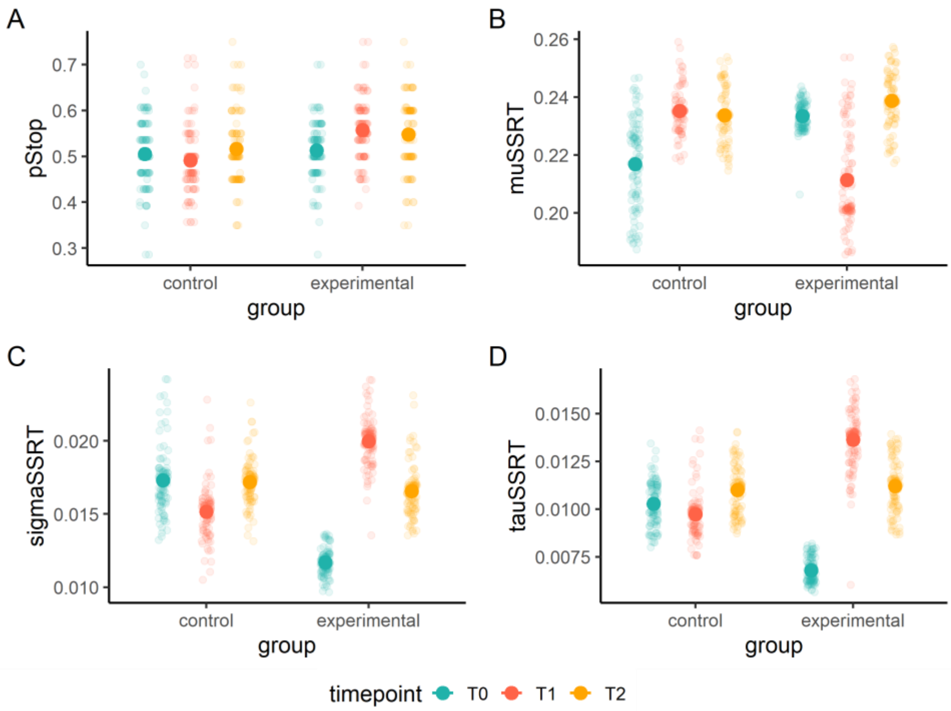

Figure 3. Plots for stop accuracy and SSRT measures at T0 (before training), T1 (after training) and T2 (one-year follow-up) for the control and experimental groups: estimated marginal mean (filled circle) and raw individual datapoints (shaded circles). A) pStop. B) Mu. C) Sigma. D) Tau.

Changes in Going and Stopping over training weeks

We also tested how accuracy, mean processing speed and IIV in processing speed changed over the training weeks for each group (see Supplementary Materials S1, Table S1-3). Note that the control group measures were extracted from the reaction time task in the response speed training, so measures were related to Going performance and included pHit, mean-GoRT and SD-GoRT; instead, the experimental group measures were extracted from the stop signal task in the inhibitory control training, so measures were related to Stopping performance and included pStop, mean-SSRT and SD-SSRT.

For accuracy in the control group (pHit) (Figure 4A), there was a main effect of Week, where accuracy decreased over weeks (b = -.016, $t(353.5) = 11.11$, $p < .001$). For accuracy in the experimental group (pStop) (Figure 4B), there was a main effect of Week, where accuracy increased over weeks (b = .002, $t(849.9) = 4.789$, $p < .001$).

For mean processing speed in the control group (mean-GoRT) (Figure 4C), there was a main effect of Week, where mean processing speed decreased over weeks (b = -17.75, $t(531.3) = 10.57$, $p < .001$). For mean processing speed in the experimental group (mean-SSRT) (Figure 4D), there was also a main effect of Week, where mean processing speed again decreased over weeks (b = -13.19, $t(443.4) = 8.275$, $p < .001$).

For IIV in processing speed in the control group (SD-GoRT) (Figure 4E), there was a main effect of Week, where SD in processing speed decreased over weeks (b = -2.159, $t(478.8) = 2.077$, $p = .038$). For IIV in processing speed in the experimental group (SD-SSRT) (Figure 4F), there was a main effect of Week, where SD in processing speed increased over weeks (b = 2.540, $t(353.0) = 2.353$, $p = .019$).

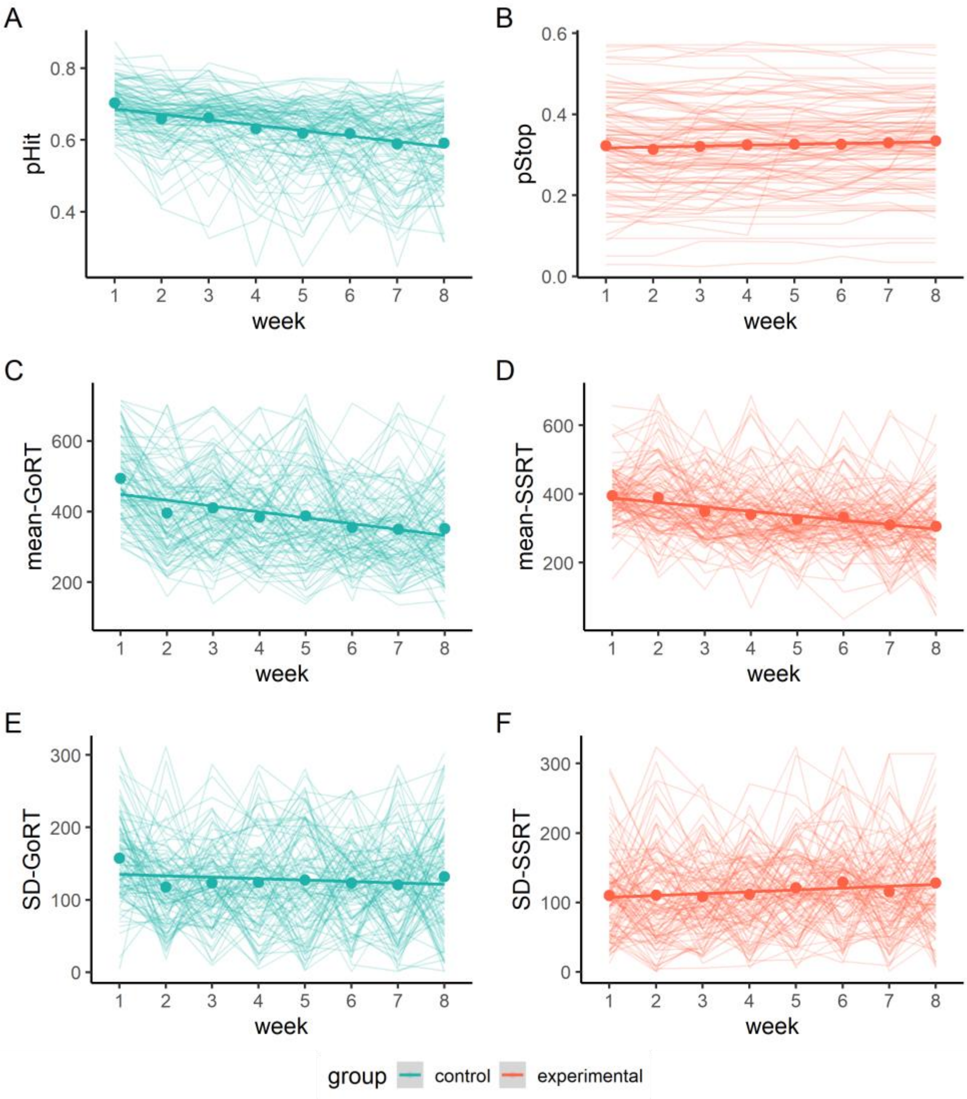

Figure 4. Plots for measures over training weeks for control (response speed training; left panels) and experimental (inhibitory control training; right panels) groups: estimated marginal mean for each week (filled circle), estimated slope across weeks (thick line), and raw individual datapoints (thin lines). A) Accuracy for control group (pHit). B) Accuracy for experimental group (pStop). C) Mean processing speed for control group (mean-GoRT). D) Mean processing speed for experimental group (mean-SSRT). E) SD in processing speed for control group (SD-GoRT). F) SD in processing speed for experimental group (SD-SSRT).

Discussion

The present study investigated how training modulates IIV of two cognitive processes with differing task demands, i.e. response speed (i.e. Going) and inhibitory control (i.e. Stopping), during middle childhood. A group of 208 six- to thirteen-year-old children underwent an 8-week inhibitory control (experimental group) or response speed (control group) training, and completed the stop signal task before (T0), immediately after (T1), and one-year after (T2) the training. We found that, at T0, higher Going accuracy was related to faster and *less* variable go responses; instead, higher Stopping accuracy was related to faster and *more* variable stop responses. In line with this, we found that, while the control group's Going performance became faster and less variable after the training, the experimental group's Stopping performance became more accurate, faster and *more* variable. Importantly, these patterns were further supported by modulations in Going and Stopping responses over the training weeks. Overall, these findings demonstrate that Going and Stopping IIV are differently modulated by training during middle childhood, which in turn suggests distinct functional roles of IIV across these two cognitive processes.

To establish how task accuracy and response features (mean processing speed and IIV) are associated during Going and Stopping, we first examined how these measures were related at T0, that is, before participants completed any training. For Going, there was a negative association between accuracy and processing speed, as well as between accuracy and IIV as measured by sigma (i.e. variability in processing speed), indicating that better accuracy is related to faster and *less* variable performance. Importantly, these findings are in line with previous studies showing similar correlations (MacDonald, Hultsch, and Dixon 2003; Rabbitt et al. 2001), as well as with the prevalent assumption that reductions in behavioural IIV indicate more efficient cognitive performance and therefore are adaptive (Unsworth

2015). For Stopping, there was a negative association between accuracy and processing speed, and a positive association between accuracy and IIV as measured by sigma. Crucially, these findings indicate that better Stopping performance is related to faster and *more* variable response inhibition, suggesting that increases in Stopping IIV are adaptive (Li, Huxhold, and Schmiedek 2004). Stopping entails a higher level of cognitive demand, that is, it requires the initiation of a response and its subsequent inhibition, and it is more unpredictable than Going in the context of the stop signal task. Therefore, it may possibly require more flexible cognitive processing to improve performance, for instance to efficiently switch from a go response to a stop response. Note that we did not find a relation between Going or Stopping accuracy IIV as measured by tau, suggesting that response accuracy is not associated with the degree and variability of occasional extremely slow responses for either Going or Stopping.

Next, we looked at training effects on both Going and Stopping. For Going, the control group showed no immediate or long-term training effects on response accuracy, and interestingly accuracy levels also dropped over the training weeks. However, responses gradually became faster over the training weeks, and showed immediate and long-term reductions after the training. Given that the aim of the training is to causally improve performance (von Bastian and Oberauer 2014), the lack of training effects on response accuracy is unexpected. One possibility is that, although children were instructed to correctly respond to the go signal while responding as fast as possible, they prioritised being fast over being accurate, thus leading to improvements in processing speed but not on accuracy (Heitz 2014). Importantly, Going IIV showed a gradual reduction over the training weeks, as well as marked reductions after training: sigma was reduced immediately after training (although note these effects did not show long-term maintenance), and tau showed a long-term reduction. Thus, consistent with the correlations at T0 and previous studies (Cubillo et al. 2022; Ram et al. 2005) the response speed training led to improvements in Going efficiency (i.e. more consistent (reduction in sigma) and less degree and variability in occasional extremely slow responses (reduction in tau)), supporting the notion that reductions in Going IIV are adaptive (Unsworth 2015; MacDonald, Li, and Bäckman 2009; West et al. 2002).

Although we did not have specific hypotheses about how Going performance would be modulated by the inhibitory control training, we found interesting changes in the experimental group. In contrast to the control group, the experimental group showed an immediate decrease in Going accuracy after the inhibitory control training, as well as an immediate and long-term slowing of mean processing speed. In the context of the stop signal task such slowing of go responses can be interpreted as a form of proactive control linked to the inhibitory control training, where participants learn to strategically slow down go responses in order to increase the probability of correctly Stopping if the stop signal is presented (Verbruggen and Logan 2008). This interpretation is further supported by the finding showing that the experimental group improved their Stopping performance after the training (discussed below). It is likely that such slowing of go responses contributed to their poor Going accuracy levels, since they might have often responded too late to the go signal, and therefore recorded more missed go responses. The experimental group also showed an immediate increase in IIV as measured by sigma, which further supports the idea that increases in Going IIV are maladaptive as they contribute to low accuracy (Williams et al. 2005; MacDonald, Nyberg, and Bäckman 2006; Unsworth 2015). However, there was a drastic immediate and long-term reduction in IIV as measured by tau, indicating that participants showed a lower degree and variability in occasional extremely slow responses. A possible interpretation is that, because participants show a general slowing of go responses, the GoRT distribution is less skewed and therefore tau is reduced. Another possibility is that, if tau reflects attentional lapses (Hervey et al. 2006; Karalunas et al. 2014; West et al. 2002), a reduction in tau indicates that participants increase their attention toward the task in order to improve their Stopping performance.

For Stopping, the experimental group showed a gradual improvement in response accuracy over the training weeks, which resulted in better Stopping accuracy both immediately and one-year after the inhibitory control training. Moreover, Stopping responses also became faster over the training weeks, resulting in faster performance after the training (although these effects were not maintained at T2). Together, these findings provide evidence that our inhibitory control training program is effective in improving inhibitory control abilities, and further show the potential of long-term maintenance of these improvements. Crucially, Stopping IIV showed a gradual increase over the training weeks, which resulted in a marked increase of sigma and tau immediately and one-year after training (although note long-term effects were slightly reduced). These findings are in line with previous studies suggesting that increases in IIV are adaptive when the environment requires greater flexible cognitive processing, for instance during learning periods or in tasks with higher cognitive demand (Li, Huxhold, and Schmiedek 2004; Siegler 1994; Siegler and Jenkins 1989; Allaire and Marsiske 2005; Garrett, McIntosh, and Grady 2014). A key aspect of Stopping is that it is arguably a complex task, because it requires efficient switching from the initiation of the motor response to its subsequent inhibition, and it is more unpredictable than Going. Thus, it is likely that during the inhibitory control training, the experimental group acquired greater flexibility to improve their Stopping performance, consequently leading to increases in Stopping IIV during and after the training.

Despite not having specific predictions about how the response speed training would modulate Stopping performance, we also looked at training-related changes in the control group. In contrast to the experimental group, the control group showed no change in Stopping accuracy after the response speed training, although stop responses became slower. Such slowing in stop responses may be related to the fact that, since the control group has been trained in responding faster to the go signal, they will incorrectly respond during stop trials more often: in line with the horse race model (Logan and Cowan 1984; Verbruggen and Logan 2008), and given that the training is adaptive, this will lead to a decrease in the stop signal delay (i.e. the stop signal is presented earlier) and the stop response distribution will become overall slower. Moreover, the control group also showed a decrease in Stopping IIV as measured by sigma and tau (which was not maintained at T2). Together with the fact that after the training they show no improvement in stop accuracy and slower stop responses, these findings support the notion that, in the context of Stopping, increases in IIV are adaptive. Importantly, the response speed training involves a reaction time task where participants just need to respond to the go signal as fast as possible: because this task has low cognitive demand, it is likely that it did not require greater cognitive flexibility in order to improve performance, but rather greater consistency in the responses. Thus, the control group training might have led to an overall decrease in IIV, in turn leading to decreases in Stopping IIV.

Overall, these findings support the notion that IIV may reflect distinct functional roles across cognitive domains (Allaire and Marsiske 2005). In particular, our findings show that causally manipulating improvements in Going performance by means of training (control group) leads to reductions in Going IIV, while the absence of such training (experimental group) leads to both worse Going performance and increased Going IIV. This suggests that for simple cognitive processes with a low level of cognitive demand and predictable responses (i.e. Going), *lower* IIV has an adaptive function by contributing to the optimisation of behaviour. Instead, causally manipulating improvements in Stopping performance (experimental group) leads to increases in Stopping IIV, whereas the lack of Stopping training (control group) leads to both worse Stopping performance and reduced Stopping IIV. Therefore, for complex cognitive processes with a high level of cognitive demand and unpredictable responses (i.e. Stopping), *greater* IIV has an adaptive function by contributing to the diversification of responses. These findings thus provide causal evidence of the process-dependent association between IIV and task performance and offer a more nuanced interpretation of the functional significance of IIV across different cognitive domains.

An important question that arises from these findings is whether these patterns result from the features of the Going or Stopping process itself (e.g. Stopping is more cognitively demanding), or from the features of the tasks used to train Going and Stopping (e.g. if the task is adaptive to participant performance). Although our study was not designed to thoroughly distinguish between these possibilities, we suggest that in the context of our training program it is likely that the former played a stronger role. In fact, while the response speed training and inhibitory control training differed in the cognitive process that was targeted, they were highly similar in terms of task features (e.g. both were adaptive, both showed the same go and stop stimuli). Future studies that systematically modulate task features across different cognitive processes will be needed to clarify this question. Furthermore, our findings have important implications for developmental, ageing and intervention studies that rely on IIV as a marker of age- and training-related changes. Contrary to the prevalent assumption that lower IIV levels indicate better outcomes, our results support the claim that IIV does not consistently reflect the same phenomenon across cognitive domains (Allaire and Marsiske 2005), and therefore its functional significance should be interpreted in relation to the specific cognitive process under study.

To conclude, the present study shows that Going IIV and Stopping IIV are differently modulated by training during middle childhood, in turn reflecting distinct functional roles of IIV across these two cognitive processes. In particular, we find that a response speed training leads to adaptive *reductions* in Going IIV, which allow more consistent and efficient Going performance when task demands are low; instead, an inhibitory control training leads to adaptive *increases* in Stopping IIV, where greater flexibility in cognitive performance is required to meet the higher cognitive demands of inhibiting a response. Overall, these findings challenge our current understanding of IIV in cognitive processing during childhood, with implications for developmental, ageing and intervention studies.

Materials and Methods

Participants

A group of 262 children from schools in the Greater London area enrolled in the study and were randomly allocated to either the control or experimental group. Fifty-four participants were excluded because they were either missing information on training group allocation, they did not complete any training sessions, and/or they did not complete any pre-post assessments for the stop signal task. Thus, the final sample consisted of 208 children, with 101 children in the control group and 107 children in the experimental group (see Supplementary Materials S2 for a flowchart describing sample sizes throughout the study). Demographics information for the full group and each of the training groups is summarized in Table 1; note there are no differences in age or SES across groups. Formal consent was obtained from parents, and participants were compensated for their participation in the study. The study was granted ethical approval by the local Research Ethics Committee.

Table 1. Participant demographics

	Total	Control group	Experimental group	Group differences
N	208	101	107	-
Gender	110 F, 98 M	57 F, 44 M	53 F, 54 M	-
Age: Mean (*SD*)	8.95 (1.57)	8.94 (1.53)	8.96 (1.62)	p = .932
Age range: min - max	6.025 - 13.32	6.064 - 12.61	6.025 - 13.32	-
SES†: Mean (*SD*)	1.64 (.607)	1.56 (.537)	1.70 (.659)	p = .113

†Socio-economic status (SES) calculation was adapted from Hollingshead 1975 (1 = highest SES score, 5 = lowest SES score); 9 children were missing SES scores.
SD: standard deviation; F: female; M: male.

Training program

The training program consisted of an 8-week intervention where participants completed 4 training sessions per week, with each session lasting 15 minutes. Within each training week, 1 session took place at the children's school and was supervised by the experimenters, whereas for the remaining 3 sessions participants were encouraged to take them at home supervised by the parents (note that, for children who enrolled in the study after the outbreak of the COVID-19 pandemic in March 2020, all training sessions took place at home). The training was computerised and happened in a gamified context, where children were instructed to earn as many points as they could through the games (i.e. the training tasks) (Figure 5A). Moreover, the training was adaptive to each child's performance to avoid ceiling and floor effects, as well as to keep children motivated throughout the sessions. There were 7 training games which were randomly assigned to the sessions, so that participants would play a different set of games in each session (around 3 games per session). The games happened in different settings (e.g. forest, desert, snow, mountain), and required participants to gain points by collecting treasures, gems or coins whilst avoiding a perpetrator (e.g. dragon, monster, ghost).

Figure 5. A) General narrative of the training program, children were instructed to earn as many points as they could through the games. B) Sample training game where the go signal is the pile of treasure, the stop signal is the dragon, the reward for successful Stopping is a gem.

While the training games were presented in the same manner across both groups, the instructions given to each group varied according to the abilities being trained. The experimental group underwent an inhibitory control training, where the stop signal task was implemented in the context of the training games, and different stimuli were used as go and stop signals depending on the game (Figure 5B). Briefly, participants were instructed to press or release a key as quickly as possible after the go signal appeared: 5 games required a spacebar keypress, 1 game required either a left or down arrow keypress depending on the go signal, and 1 game required releasing the spacebar key. However, on stop trials (26-47%

of total trials depending on game, mean = 32%) a stop signal would immediately appear after the go signal, and in this case participants were instructed not to respond to the go signal, thus requiring them to inhibit the go signal response. The stop signal delay (SSD; i.e. delay between the presentation of the go signal and the stop signal) was initially set at 200 ms, and was adjusted to participants' performance using an adaptive staircase procedure: if participants successfully inhibited their response then the SSD was increased by 50 ms to make the task more difficult, however if participants did not inhibit their response then the SSD was decreased by 50 ms to make the task easier. This ensured that the training was adaptive and avoided floor or ceiling effects.

The control group underwent a response speed training, which used the same games played by the experimental group, but participants were instructed to correctly respond to all go signals as quickly as possible (regardless of the stop signal). To ensure the training was adaptive, a rolling average of the reaction time across the previous 10 trials plus 2 standard deviations was used as threshold: if the response time for a given trial was faster than the threshold, the duration of the go signal was decreased by 50 ms to make the task harder; if the response time was slower than the threshold, the duration of the go signal was increased by 50 ms to make the task easier.

Pre-post assessments: stop signal task

Before and after the training there were 3 assessment timepoints that took place onsite at the author's laboratory: before the training (T0), after the training (T1), and one-year follow-up (T2). Note that, due to the outbreak of the COVID-19 pandemic in March 2020, some participants completed one or more assessment timepoints online from home. The assessment battery included the stop signal task measuring inhibitory control, several tasks measuring other executive functions, structural and functional imaging measurements, as well as questionnaires measuring IQ, socio-economic status, creativity, mental health and academic performance. For the scope of the present study, we will focus on the stop signal task.

Participants completed a child-friendly version of the stop signal task, which differed from the training games in that it was not implemented in a gamified context. For participants with assessment timepoints happening before the COVID-19 outbreak in March 2020, the task was programmed in E-Prime software (Psychology Software Tools, Pittsburgh, PA) and completed locally. For participants with assessment timepoints happening after the COVID-19 outbreak in March 2020, the task was designed using PsychoPy3 (Peirce et al. 2019), and was made available online via Pavlovia (www.pavlovia.org). Participants first practiced the task over 10 trials and then completed a total of 80 trials as part of the main task. Each trial started with the presentation of a fixation cross for 1250 ms, followed by a honey pot (go signal) that appeared either on the left side or right side of the screen (Figure 7). Participants were instructed to respond as fast as possible according to the side where the honey pot appeared: if the stimulus appeared on the left, participants were instructed to press the left arrow key, and if the stimulus appeared on the right, participants were instructed to press the down arrow key. On Go trials (75% of the total trials), the honey pot disappeared when participants responded or after 1000 ms (Figure 6). On Stop trials (25% of the total trials), the go signal was immediately followed by a stop signal, which corresponded to a picture of bees and was displayed for 300 ms (Figure 6). In the presence of a stop signal, participants were instructed not to respond to the go signal, thus requiring them to inhibit the go signal response. The delay between the presentation of the go signal and the stop signal (i.e. stop signal delay, SSD) was adjusted to participants' performance using an adaptive staircase procedure: at the beginning of the task the SSD was set at 200 ms; if participants successfully inhibited their response, the SSD was increased by 50 ms to make the task more difficult; if participants did not inhibit their response, the SSD was decreased by 50 ms to make the task easier. This adjustment is meant to avoid floor or ceiling effects and ensure the task is adaptive.

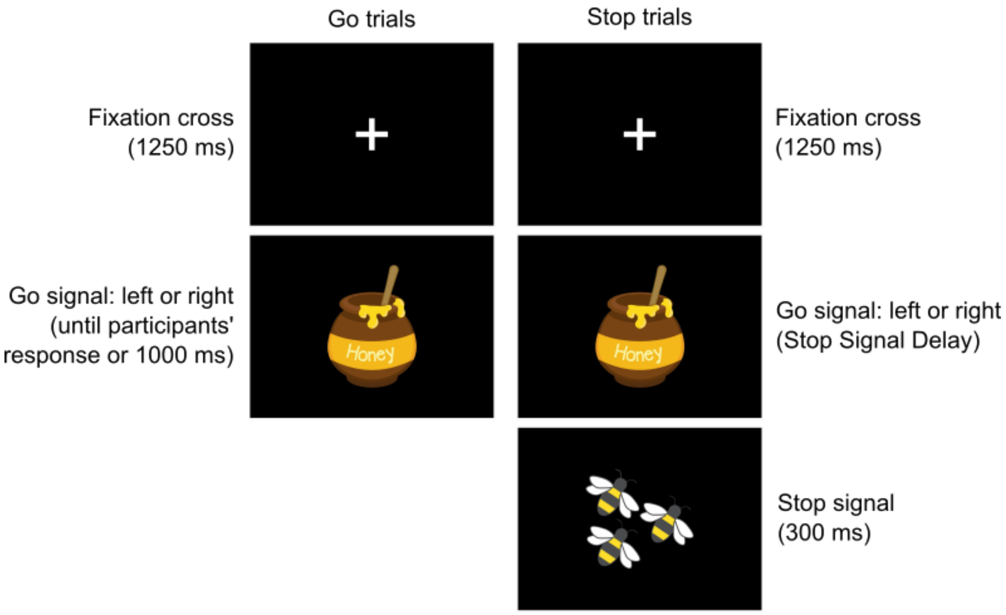

Figure 6. Sample go and stop trials for the child-friendly version of the stop signal task.

Measures

<u>Training data</u>

Because each training game included a small number of trials, within each session we pooled together trials from games that required the same time of key response (spacebar keypress, arrows keypress or key release). Trials with reaction times below 100 ms were excluded, and if a set of pooled games did not reach a minimum of 50 trials (Verbruggen et al. 2019) it was excluded from further analyses.

For the experimental group, we calculated the Stop Signal Reaction Time (SSRT) for each set of pooled games, according to the horse-race model of Stopping (Logan and Cowan

1984) and the integration method (i.e. with replacement of go omissions) (Verbruggen et al. 2019). Following this procedure, we first determined the maximum reaction time for correct go responses and replaced go omission trials with this value. Next, we rank-ordered all reaction times for go responses and determined the percentage of failed inhibitions: the go reaction time (GoRT) that corresponded to this percentage was determined (nth GoRT). Finally, we computed the SSRT as the difference between the nth GoRT and the mean SSD. A set of pooled games was excluded from a session if the SSRT was negative, if the mean RT for go successful trials was smaller than the mean RT for stop unsuccessful trials, if the probability of false alarm was lower than 25% or greater than 75%, or if the probability of correct go responses was lower than 50% (Verbruggen et al. 2019). For each session, we averaged the SSRT values across the sets of pooled games, resulting in an SSRT value for each participant and session. Finally, we computed the mean and SD of the SSRT across all sessions happening within the same training week, resulting in a mean and SD SSRT value for each participant and week. We also computed accuracy levels in stop responses, where the probability of correctly Stopping (pStop) was computed as the proportion of correct stop responses relative to the total stop trials.

For the control group, we calculated the mean Go Reaction Time (GoRT) for each set of pooled games. A set of pooled games was excluded from a session if the probability of correct go responses was lower than 50%. For each session, we averaged the GoRT values across the sets of pooled games, resulting in a GoRT value for each participant and session. Finally, we computed the mean and SD of the GoRT across all sessions happening within the same training week, resulting in a mean and SD GoRT value for each participant and week. We also computed accuracy in go responses, where the probability of hit (pHit) was computed as the proportion of correct go responses relative to the total go trials.

<u>Pre-post assessment data</u>

For both groups, we excluded trials with reaction times below 100 ms, and calculated the SSRT according to the horse-race model of Stopping (Logan and Cowan 1984) to aid in our exclusion criteria. Participants were excluded if the SSRT was negative, if the mean RT for go successful trials was smaller than the mean RT for stop unsuccessful trials, or if the probability of correct go responses was lower than 30% (Verbruggen et al. 2019). Note this more lenient criteria was used for pre-post assessments due to the smaller amount of data available per timepoint and participant.

Ex-Gaussian measures for SSRTs and GoRTs were estimated using a hierarchical Bayesian Parametric Approach (BPA) implemented with the Dynamic Models of Choice software (Heathcote et al. 2019; Matzke et al. 2013). The BPA assumes that SSRTs and GoRTs form an ex-Gaussian distribution and uses Markov Chain Monte Carlo (MCMC) sampling of the observed participant stop signal task data in order to estimate the three parameters that describe the SSRT and GoRT distributions: mu, sigma and tau (Matzke et al. 2013).

Finally, we also computed measures of accuracy in go and stop responses. For go responses, the probability of hit (pHit) was computed as the proportion of correct go responses relative to the total go trials. For stop responses, the probability of correctly Stopping (pStop) was computed as the proportion of correct stop responses relative to the total stop trials.

Statistical analyses

Task data was cleaned using MATLAB (R2021a, MathWorks), and analysed with R (R Core Team 2017), using the lme4 and lmerTest packages (Bates et al. 2015; Kuznetsova, Brockhoff, and Christensen 2017).

Training data

Outliers were excluded based on the 1.5*IQR criterion, and missing data (8%) were imputed (Jeličić, Phelps, and Lerner 2009) with the Multivariate Imputation by Chained Equations (mice) R package (Buuren and Groothuis-Oudshoorn 2011). The variables included as predictors in the multiple imputation specification were group, week, age, pHit, pStop, mean and SD, and all variables but group, week and age were imputed. One hundred multiple imputed datasets were created and pooled for statistical analyses (see Supplementary Materials S3, Figure S3.1, for density plots of the complete case dataset versus pooled imputed datasets). Linear mixed models with pHit (control group), pStop (experimental group), and mean or SD (of GoRT and SSRT for the control and experimental group, respectively) as dependent variable and week (1-8) as covariate were fitted. Note that, for each dependent variable, two additional models with age as a nuisance covariate or with a 2-way interaction between week and age (to test age-dependent training effects) were also fitted, but their goodness-of-fit (Akaike Information Criteria) was lower than for the model excluding age; moreover, the pattern of results was the same across all three models. All analyses were also run with the complete case dataset: the pattern of results was generally the same, so only the results from the pooled imputed datasets are reported in the main text, and results from the complete case dataset are reported in Supplementary Materials S4.

Pre-post assessment data

Outliers were excluded based on the 1.5*IQR criterion, and missing data (4.79%) were imputed (Jeličić, Phelps, and Lerner 2009) with the Multivariate Imputation by Chained Equations (mice) R package (Buuren and Groothuis-Oudshoorn 2011). The variables included as predictors in the multiple imputation specification were group, timepoint, age, pHit, pStop, muSSRT, sigmaSSRT, tauSSRT, muGoRT, sigmaGoRT and tauGoRT, and all variables but group, timepoint and age were imputed. One hundred multiple imputed datasets were created and pooled for statistical analyses (see Supplementary Materials S3, Figure S3.2, for density plots of the complete case dataset versus pooled imputed datasets). Linear models with the T0 pooled data were run between GoRT measures (mu, sigma, tau) and pHit, as well as between SSRT measures (mu, sigma, tau) and pStop, with group as a nuisance covariate. Linear mixed models with pHit, pStop, and mu, sigma or tau (of GoRT and SSRT) as dependent variable, group (control, experimental) as between-subject factor, timepoint (T0, T1, T2) as within-subject factor were fitted. Note that, for each dependent variable, two additional models with age as a nuisance covariate or with a 2-way interaction between week and age (to test age-dependent training effects) were also fitted, but their goodness-of-fit (Akaike Information Criteria) was lower than for the model excluding age; moreover, the pattern of results was the same across all three models. Because it is not yet possible to run linear mixed model post-hoc tests with pooled datasets, we re-ran all linear mixed models with post-hoc pairwise comparisons using Bonferroni's adjustment on a single imputed dataset, selected as the first of the 100 imputed datasets (see Supplementary Materials S3, Figure S3.3, for density plots of the complete case dataset versus single imputed dataset). Results for main and interaction effects showed the same pattern when using the pooled imputed datasets or the single imputed dataset, so only the results from the single imputed dataset are reported for linear mixed models. All analyses were also run with the complete case dataset: the pattern of results was generally the same, so only the results from the pooled/single imputed datasets are reported in the main text, and results from the complete case dataset are reported in Supplementary Materials S4.

Acknowledgements

The authors gratefully acknowledge the participating families in this study.

Competing interests

The authors declare no competing interests.

References

Allaire, Jason C., and Michael Marsiske. 2005. “Intraindividual Variability May Not Always Indicate Vulnerability in Elders’ Cognitive Performance.” *Psychology and Aging* 20 (3): 390–401. https://doi.org/10.1037/0882-7974.20.3.390.

Bastian, Claudia C. von, and Klaus Oberauer. 2014. “Effects and Mechanisms of Working Memory Training: A Review.” *Psychological Research* 78 (6): 803–20. https://doi.org/10.1007/s00426-013-0524-6.

Bates, Douglas, Martin Mächler, Ben Bolker, and Steve Walker. 2015. “Fitting Linear Mixed-Effects Models Using Lme4.” *Journal of Statistical Software* 67 (1). https://doi.org/10.18637/jss.v067.i01.

Belle, Janna van, Tamar van Raalten, Dienke J. Bos, Bram B. Zandbelt, Bob Oranje, and Sarah Durston. 2015. “Capturing the Dynamics of Response Variability in the Brain in ADHD.” *NeuroImage: Clinical* 7: 132–41. https://doi.org/10.1016/j.nicl.2014.11.014.

Buuren, Stef van, and Karin Groothuis-Oudshoorn. 2011. “Mice: Multivariate Imputation by Chained Equations in R.” *Journal of Statistical Software* 45 (3): 1–67. https://doi.org/10.18637/jss.v045.i03.

Cherbuin, Nicolas, Perminder Sachdev, and Kaarin J. Anstey. 2010. “Neuropsychological Predictors of Transition From Healthy Cognitive Aging to Mild Cognitive Impairment:

The PATH Through Life Study." *The American Journal of Geriatric Psychiatry* 18 (8): 723–33. https://doi.org/10.1097/JGP.0b013e3181cdecf1.

Cubillo, Ana, Henning Hermes, Eva Berger, Kirsten Winkel, Daniel Schunk, Ernst Fehr, and Todd A. Hare. 2022. "Intra-individual Variability in Task Performance after Cognitive Training Is Associated with Long-term Outcomes in Children." *Developmental Science*, no. February 2021 (March): 1–16. https://doi.org/10.1111/desc.13252.

Diamond, Adele. 2013. "Executive Functions." *Annual Review of Psychology* 64 (1): 135–68. https://doi.org/10.1146/annurev-psych-113011-143750.

Durston, Sarah, Kathleen M. Thomas, Yihong Yang, Aziz M. Ulug, Robert D. Zimmerman, and B.J. Casey. 2002. "A Neural Basis for the Development of Inhibitory Control." *Developmental Science* 5 (4): F9–16. https://doi.org/10.1111/1467-7687.00235.

Garrett, Douglas D., Anthony R. McIntosh, and Cheryl L. Grady. 2014. "Brain Signal Variability Is Parametrically Modifiable." *Cerebral Cortex* 24 (11): 2931–40. https://doi.org/10.1093/cercor/bht150.

Geary, David C. 2010. "Mathematical Learning Disabilities." In *Advances in Child Development and Behavior*, 39:45–77. https://doi.org/10.1016/B978-0-12-374748-8.00002-0.

Heathcote, Andrew, Yi-Shin Lin, Angus Reynolds, Luke Strickland, Matthew Gretton, and Dora Matzke. 2019. "Dynamic Models of Choice." *Behavior Research Methods* 51 (2): 961–85. https://doi.org/10.3758/s13428-018-1067-y.

Heitz, Richard P. 2014. "The Speed-Accuracy Tradeoff: History, Physiology, Methodology, and Behavior." *Frontiers in Neuroscience* 8 (8 JUN): 1–19. https://doi.org/10.3389/fnins.2014.00150.

Hervey, Aaron S., Jeffery N. Epstein, John F. Curry, Simon Tonev, L. Eugene Arnold, C. Keith

Conners, Stephen P. Hinshaw, James M. Swanson, and Lily Hechtman. 2006. “Reaction Time Distribution Analysis of Neuropsychological Performance in an ADHD Sample.” *Child Neuropsychology* 12 (2): 125–40. https://doi.org/10.1080/09297040500499081.

Hollingshead, A. B. 1975. *Four Factor Index of Social Status. Unpublished Manuscript*. New Haven, CT: Yale University.

Hultsch, David F., S. W. S. MacDonald, and Roger A. Dixon. 2002. “Variability in Reaction Time Performance of Younger and Older Adults.” *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences* 57 (2): P101–15. https://doi.org/10.1093/geronb/57.2.P101.

Hultsch, David F., and Stuart W. S. MacDonald. 2004. “Intraindividual Variability in Performance as a Theoretical Window onto Cognitive Aging.” In *New Frontiers in Cognitive Aging*, 65–88. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198525691.003.0004.

Jeličić, Helena, Erin Phelps, and Richard M. Lerner. 2009. “Use of Missing Data Methods in Longitudinal Studies: The Persistence of Bad Practices in Developmental Psychology.” *Developmental Psychology* 45 (4): 1195–99. https://doi.org/10.1037/a0015665.

Kail, Robert. 1991. “Developmental Change in Speed of Processing during Childhood and Adolescence.” *Psychological Bulletin* 109 (3): 490–501. https://doi.org/10.1037/0033-2909.109.3.490.

Karalunas, Sarah L., Hilde M. Geurts, Kerstin Konrad, Stephan Bender, and Joel T. Nigg. 2014. “Annual Research Review: Reaction Time Variability in ADHD and Autism Spectrum Disorders: Measurement and Mechanisms of a Proposed Trans-Diagnostic Phenotype.” *Journal of Child Psychology and Psychiatry* 55 (6): 685–710. https://doi.org/10.1111/jcpp.12217.

Kuznetsova, Alexandra, Per B. Brockhoff, and Rune H. B. Christensen. 2017. “LmerTest Package: Tests in Linear Mixed Effects Models.” *Journal of Statistical Software* 82 (13). https://doi.org/10.18637/jss.v082.i13.

Li, Shu-Chen, Steven H. Aggen, John R. Nesselroade, and Paul B. Baltes. 2001. “Short-Term Fluctuations in Elderly People’s Sensorimotor Functioning Predict Text and Spatial Memory Performance: The MacArthur Successful Aging Studies.” *Gerontology* 47 (2): 100–116. https://doi.org/10.1159/000052782.

Li, Shu-Chen, Oliver Huxhold, and Florian Schmiedek. 2004. “Aging and Attenuated Processing Robustness.” *Gerontology* 50 (1): 28–34. https://doi.org/10.1159/000074386.

Logan, Gordon D., and William B. Cowan. 1984. “On the Ability to Inhibit Thought and Action: A Theory of an Act of Control.” *Psychological Review* 91 (3): 295–327. https://doi.org/10.1037/0033-295X.91.3.295.

Luce, R. Duncan. 1986. *Response Times: Their Role in Inferring Elementary Mental Organization*. Oxford, UK: Oxford University Press.

Luna, Beatriz, Aarthi Padmanabhan, and Kirsten O’Hearn. 2010. “What Has FMRI Told Us about the Development of Cognitive Control through Adolescence?” *Brain and Cognition* 72 (1): 101–13. https://doi.org/10.1016/j.bandc.2009.08.005.

MacDonald, Stuart W. S., David F. Hultsch, and Roger A. Dixon. 2003. “Performance Variability Is Related to Change in Cognition: Evidence from the Victoria Longitudinal Study.” *Psychology and Aging* 18 (3): 510–23. https://doi.org/10.1037/0882-7974.18.3.510.

MacDonald, Stuart W. S., Shu-Chen Li, and Lars Bäckman. 2009. “Neural Underpinnings of Within-Person Variability in Cognitive Functioning.” *Psychology and Aging* 24 (4): 792–

808. https://doi.org/10.1037/a0017798.

MacDonald, Stuart W.S., Lars Nyberg, and Lars Bäckman. 2006. “Intra-Individual Variability in Behavior: Links to Brain Structure, Neurotransmission and Neuronal Activity.” *Trends in Neurosciences* 29 (8): 474–80. https://doi.org/10.1016/j.tins.2006.06.011.

Matzke, Dora, Conor V. Dolan, Gordon D. Logan, Scott D. Brown, and Eric-Jan Wagenmakers. 2013. “Bayesian Parametric Estimation of Stop-Signal Reaction Time Distributions.” *Journal of Experimental Psychology: General* 142 (4): 1047–73. https://doi.org/10.1037/a0030543.

Matzke, Dora, Jonathon Love, and Andrew Heathcote. 2017. “A Bayesian Approach for Estimating the Probability of Trigger Failures in the Stop-Signal Paradigm.” *Behavior Research Methods* 49 (1): 267–81. https://doi.org/10.3758/s13428-015-0695-8.

McAuley, Tara, Melvin Yap, Shawn E. Christ, and Desiree A. White. 2006. “Revisiting Inhibitory Control Across the Life Span: Insights From the Ex-Gaussian Distribution.” *Developmental Neuropsychology* 29 (3): 447–58. https://doi.org/10.1207/s15326942dn2903_4.

Moffitt, Terrie E., Louise Arseneault, Daniel Belsky, Nigel Dickson, Robert J. Hancox, H. Harrington, Renate Houts, et al. 2011. “A Gradient of Childhood Self-Control Predicts Health, Wealth, and Public Safety.” *Proceedings of the National Academy of Sciences* 108 (7): 2693–98. https://doi.org/10.1073/pnas.1010076108.

Nesselroade, John R. 1991a. “Interindividual Differences in Intraindividual Change.” In *Best Methods for the Analysis of Change*, edited by L. M. Collins and J. L. Horn, 92–105. Washington, DC: American Psychological Association.

———. 1991b. “The Warp and Woof of the Developmental Fabric.” In *Visions of Aesthetics, the Environment, and Development: The Legacy of Joachim F. Wohlwill*, edited by R.

Downs, L. Libel, and D. S. Palermo, 213–40. Hillsdale, NJ: Erlbaum. https://doi.org/10.2307/3033464.

Nussenbaum, Kate, and Catherine A. Hartley. 2019. "Reinforcement Learning across Development: What Insights Can We Draw from a Decade of Research?" *Developmental Cognitive Neuroscience* 40 (December): 100733. https://doi.org/10.1016/j.dcn.2019.100733.

Peirce, Jonathan, Jeremy R. Gray, Sol Simpson, Michael MacAskill, Richard Höchenberger, Hiroyuki Sogo, Erik Kastman, and Jonas Kristoffer Lindeløv. 2019. "PsychoPy2: Experiments in Behavior Made Easy." *Behavior Research Methods* 51 (1): 195–203. https://doi.org/10.3758/s13428-018-01193-y.

R Core Team. 2017. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.r-project.org/.

Rabbitt, Patrick, Paul Osman, Belinda Moore, and Brian Stollery. 2001. "There Are Stable Individual Differences in Performance Variability, Both from Moment to Moment and from Day to Day." *The Quarterly Journal of Experimental Psychology Section A* 54 (4): 981–1003. https://doi.org/10.1080/713756013.

Ram, Nilam, Patrick Rabbitt, Brian Stollery, and John R. Nesselroade. 2005. "Cognitive Performance Inconsistency: Intraindividual Change and Variability." *Psychology and Aging* 20 (4): 623–33. https://doi.org/10.1037/0882-7974.20.4.623.

Shalev, Nir, Anna-Katharina R Bauer, and Anna C. Nobre. 2019. "The Tempos of Performance." *Current Opinion in Psychology* 29 (October): 254–60. https://doi.org/10.1016/j.copsyc.2019.06.003.

Siegler, Robert S. 1994. "Cognitive Variability: A Key to Understanding Cognitive Development." *Current Directions in Psychological Science* 3 (1): 1–5.

https://doi.org/10.1111/1467-8721.ep10769817.

Siegler, Robert S., and E. Jenkins. 1989. “How Children Discover New Strategies.” In . Lawrence Erlbaum Associates, Inc.

Silva, Marc A., and Jaylene M. Lee. 2021. “Neurocognitive Testing.” In *Reference Module in Neuroscience and Biobehavioral Psychology*, Second Edi, 1–6. Elsevier. https://doi.org/10.1016/B978-0-12-822963-7.00047-5.

Tamnes, Christian K., Andres M. Fjell, Lars T. Westlye, Y. Ostby, and Kristine B. Walhovd. 2012. “Becoming Consistent: Developmental Reductions in Intraindividual Variability in Reaction Time Are Related to White Matter Integrity.” *Journal of Neuroscience* 32 (3): 972–82. https://doi.org/10.1523/JNEUROSCI.4779-11.2012.

Thompson, Abigail, Margot A. Schel, and Nikolaus Steinbeis. 2021. “Changes in BOLD Variability Are Linked to the Development of Variable Response Inhibition.” *NeuroImage* 228 (July 2020): 117691. https://doi.org/10.1016/j.neuroimage.2020.117691.

Unsworth, Nash. 2015. “Consistency of Attentional Control as an Important Cognitive Trait: A Latent Variable Analysis.” *Intelligence* 49 (March): 110–28. https://doi.org/10.1016/j.intell.2015.01.005.

Verbruggen, Frederick, Adam R. Aron, Guido P.H. Band, Christian Beste, Patrick G. Bissett, Adam T. Brockett, Joshua W. Brown, et al. 2019. “A Consensus Guide to Capturing the Ability to Inhibit Actions and Impulsive Behaviors in the Stop-Signal Task.” *ELife* 8 (April): 1–26. https://doi.org/10.7554/eLife.46323.

Verbruggen, Frederick, and Gordon D. Logan. 2008. “Response Inhibition in the Stop-Signal Paradigm.” *Trends in Cognitive Sciences* 12 (11): 418–24. https://doi.org/10.1016/j.tics.2008.07.005.

West, Robert, Kelly J. Murphy, Maria L. Armilio, Fergus I.M. Craik, and Donald T. Stuss. 2002. "Lapses of Intention and Performance Variability Reveal Age-Related Increases in Fluctuations of Executive Control." *Brain and Cognition* 49 (3): 402–19. https://doi.org/10.1006/brcg.2001.1507.

Williams, Benjamin R., David F. Hultsch, Esther H. Strauss, Michael A. Hunter, and Rosemary Tannock. 2005. "Inconsistency in Reaction Time Across the Life Span." *Neuropsychology* 19 (1): 88–96. https://doi.org/10.1037/0894-4105.19.1.88.