Structure and implementation of novel task rules:
A cross-sectional developmental study

Frederick Verbruggen[1,2], Rossy McLaren[2], Maayan Pereg[3], Nachshon Meiran[3]

1. Department of Experimental Psychology, Ghent University
2. School of Psychology, University of Exeter
3. Department of Psychology, Ben-Gurion University of the Negev

Acknowledgements

This work was supported by an ERC starting grant to FV (No. 312445).

Abstract

Rule-based performance improves remarkably throughout childhood. The present study examined how children and adolescents structured tasks and implemented rules when novel task instructions were presented in a child-friendly version of a novel instruction-learning paradigm. Each mini-block started with the presentation of the new stimulus-response mappings for a GO task. Prior to implementing this mapping, responses were required to advance through screens during a preparatory (NEXT) phase. Children (4-11 years) and late adolescents (17-19 years) responded more slowly during the NEXT phase when the NEXT response was incompatible with the instructed stimulus-response mapping. This instruction-based interference effect was more pronounced in young children than in older children. We argue that these findings are most consistent with age-related differences in rule structuring. We discuss the implications of our findings for theories of rule-based performance, instruction-based learning, and development.

People often have to perform novel tasks or actions. The present study examined two critical aspects of novel-task performance, namely the abilities to follow instructions and to structure tasks hierarchically. These two issues are related when novel task instructions have to be deferred. For example, when you are about to travel to the United Kingdom for the first time, a friend may tell you that you have to look to the left when crossing a street. However, you should only follow her instructions once you have reached your destination, and failures to do so could have serious negative consequences. Here we tested how children and late adolescents performed in such novel-task situations.

From instructions to rule-based behavior

When instructions are presented, a task ‘model' or 'set' has to be created. This involves selecting and gating information from the perceptual and motor systems (Cole, Laurent, & Stoko, 2013) and chunking relevant task components (Bhandari & Duncan, 2014). Such cognitive structures allow flexible and rule-based behavior in complex environments (Bhandari, Badre, & Frank, 2017).

Once task structures are created, they have to be implemented. Instructed rules have powerful effects on behavior when they are implemented or maintained for future use (Meiran, Liefooghe, & De Houwer, 2017). Indeed, even if their execution is deferred (like in the example above), rules can influence ongoing performance. In a recent study, subjects were presented with novel instructions at the beginning of each miniblock (Meiran, Pereg, Kessler, Cole, & Braver, 2015a). These instructions described the stimulus-response (S-R) mapping for the GO phase of the block (e.g., ‘© = left, £ = right’). Before subjects could apply these instructions, they had to advance through a NEXT phase. In this phase, stimuli were presented but their identity could be ignored and subjects simply had to press the same NEXT key on each trial (which was either the left or right key). Even though the S-R rules had never been applied before, subjects were slower to respond to NEXT stimuli when the NEXT response and the GO response were incompatible (‘£’ requiring a left response in the

NEXT phase but a right response in the GO phase) compared to when they were compatible ('©' requiring a left response in both phases). This instruction-based interference effect shows that instructions enable 'automatic' task performance (Meiran et al., 2017).

Several lines of research suggest that interference during the task-implementation or execution phases can be reduced by creating hierarchical task structures (Cole, Meiran, & Braver, 2017). In a hierarchical task structure, a task cue or context determines the relevant response rules. Such hierarchical information can shield ongoing tasks (e.g. traveling to the airport) from pending instructions (e.g. walking in London), thereby reducing instruction- or rule-based interference.

The development of structuring and implementing rules

Rule-based behavior improves remarkably from infancy through childhood and adolescence (Bunge & Crone, 2009; Diamond, 2013). Such developmental improvements might be due to the ability to create and use hierarchical task structures (Bunge & Zelazo, 2006). For example, Amso, Haas, McShane, & Badre (2014) manipulated hierarchical structure (number of subtasks/branches) and number of competing alternatives within a branch independently. Age-related performance differences were primarily influenced by task structure, rather than competition between choice alternatives (see also Unger, Ackerman, Chatham, Amso, & Badre, 2016). In other words, the ability to structure rules improved throughout childhood.

Other studies also found age-related differences in the implementation phase. For example, Zelazzo, Frye, and Rapus (1996) observed a dissociation between knowing and doing in 3-year olds. In a simple rule-switching paradigm, 3-year olds kept doing the task they started with, even when instructed to perform the other task instead. Importantly, when the children were asked what the task rules were, they could accurately recall them, suggesting they experienced difficulties with implementing (but not remembering) the appropriate rules. The 'proactive control' literature also suggests that young children are less likely to implement or maintain rules than older children (Munakata, Snyder, & Chatham, 2012). This could be due to increased costs associated with advance rule implementation. For example, Blackwell and Munakata, (2012) showed that adding a secondary task to a card-sorting task particularly impaired performance of young children who tried to maintain task-related information over time (compared with children who did not maintain the rules). Thus, for young children, implementing rules in advance comes with challenges and can produce behavioral costs.

The present study

To date, most developmental studies focused on rule-based performance in situations in which children alternated between well-practiced tasks. This research largely ignores the early stages in which the novel instructions are presented and implemented for the first time (i.e., the first trials or blocks are usually practice and not further analyzed). However, task structures created in the beginning of the experiment determine future task performance (Bhandari & Duncan, 2014). In other words, these early phases are crucial.

The present study examined age-related differences in the task-formation and early implementation stages when novel task instructions were presented. We developed a child-friendly version of the NEXT paradigm (Meiran et al., 2015a). This task combines two elements that are usually studied separately, namely the ability to follow or implement instructions and the use of hierarchical structures to shield pending instructions. At the beginning of each miniblock, we showed the children two cartoon images of their ‘friends’ (task-instruction phase; Figure 1). New images were used for each miniblock. Some of their friends lived on the left side of the street, and some of them lived on the right side. In the evening (GO phase), they had to bring their friends home by pressing the appropriate left/right key (task-implementation phase). However, in the morning, before they could go home, all friends had to go to school first (NEXT phase), which was located on the left side of the screen for half of the subjects, and the right side for the other half. The GO and NEXT phase were indicated by the morning/evening background. Children (4-11 year) and late adolescents (17-19 year) performed this task.

Hierarchical control is needed in this task, since the NEXT and GO phases create two different contexts. As discussed above, the ability to contextualize behavior and structure tasks develops in young childhood. This ability would reduce interference from one context (GO) to the other (NEXT). Therefore, the ‘hierarchical-control’ account predicts that instruction-based interference effects (i.e. slower responding when the NEXT response is incompatible with the instructed GO response) should be *more pronounced* for younger children than for older children.

To observe an instruction-based interference effect, task rules have to be implemented or maintained in a highly accessible state (Meiran et al., 2017). Theoretical analyses link automatic effects of instruction to proactive control (Cole et al, 2017), but as noted above, young children are less likely to implement rules in advance. Therefore, the ‘advance-implementation’ account predicts impaired GO performance, but *less pronounced* interference effects in the NEXT phase for younger children than for older children (contrasting with practice-based interference effects that are typically larger for younger children; e.g. Huizinga, Dolan, & van der Molen, 2006).

Experiment

Method

Participants. 178 children (4–11 years) from two local schools in Devon (UK) and 30 late adolescents (17–19 years) from two local colleges (also in Devon) participated in this experiment (Table 1). We excluded 5 children because they did not complete the experiment, and 7 children because accuracy in the GO phase was below 60%. In Supplementary Materials, we show that excluding these subjects or certain trial types (see below) did not alter the main findings.

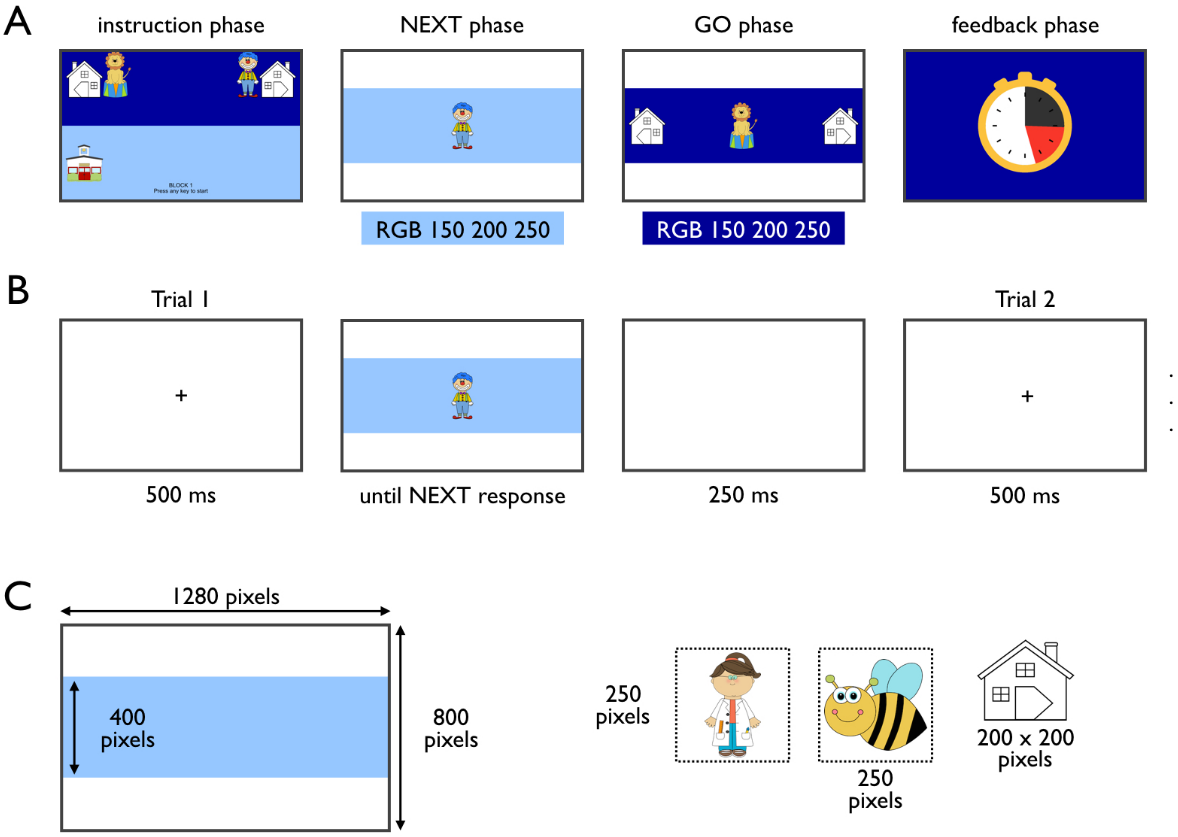


Figure 1: A depiction of the course of a miniblock, with four different phases **(A)**, and the trial course for NEXT trials **(B)**. The trial course for GO trials was very similar to the course for NEXT trials, except that the stimulus disappeared as soon as a response key (correct or incorrect) was pressed. **C** shows the size of the screen and the stimuli.

We aimed to recruit as many children and adolescents as possible. Therefore, we contacted two local primary schools, and all children for whom we obtained parental consent were invited to participate. Because we did not know in advance how many parental consents we would obtain, we could not determine the exact target sample in advance. The decision to stop testing was not influenced by the analyses of the data.

The children received a small prize (a sticker of a cartoon character of their choice and a certificate). The adolescents received monetary compensation (£2.50). The experiment was approved by the local research ethics committee. For the children and underage adolescents, parental informed consent and the subjects' assent were obtained. We obtained written informed consent from the other adolescents.

Table 1: Number of subjects and gender for each age group.

	4	5	6	7	8	9	10	11	17-19
Total Number	8	22	26	20	29	30	15	16	30
Number of Females	6	12	13	11	17	8	6	9	16

Procedure. The experiment took place in a quiet room at school (the children) or college (the adolescents), and was run on a 13-inch Macbook Pro using Psychtoolbox (Brainard, 1997). We tested one subject at a time. Stimuli consisted of cartoon images of various animals, (imaginary) creatures, and people. We used different stimuli in each miniblock, and they were easily distinguishable from each other. The 'a' and 'l' keys of the keyboard were the response keys, and we put arrow stickers on them as a reminder. Both keys were used in the GO phase. For half of the subjects, the 'a' (left) key was the NEXT response; for the others, the 'l' (right) key was the NEXT response.

Each miniblock consisted of four phases: an instruction phase, a NEXT phase, a GO phase, and a feedback phase (Figure 1). In the instruction phase, we presented the novel S-R mappings for the GO phase, and a response reminder for the NEXT phase (i.e., a school building on the left or right of the screen, depending on the counterbalancing of the NEXT response). The GO information appeared on the top of the screen, against a dark blue background; the NEXT reminder appeared on the bottom, against a light blue background. The instructions remained on the screen until subjects had pressed a key and at least three seconds had elapsed.

The trial course of the NEXT phase, indicated by a light blue background, is depicted in Figure 1. After an intertrial and fixation interval, a stimulus appeared and remained on screen until the correct NEXT key was pressed. Thus, if subjects pressed the incorrect key first (e.g. 'l' when the NEXT response was 'a'), the stimulus would remain on the screen; it would only disappear once they had pressed the NEXT key. The number of NEXT trials differed between blocks (see below).

The GO phase, indicated by a dark blue rectangle (‘evening’), always consisted of two trials. The trial course was the same as in the NEXT phase, except that the stimulus disappeared as soon as a response key (correct or incorrect) was pressed.

In the feedback phase, we presented a ‘clock’ (Figure 1). The dark grey area depicted the total response latency for the two GO trials. For each incorrect GO response, we added a time penalty (indicated in red). We also played a sound during the feedback phase: if subjects did not make GO errors, we presented ‘yihaa’ (if they had responded faster than in the preceding miniblock) or ‘ok’ (if they had responded slower); we presented ‘oops’ if they had made a GO error. The feedback remained on the screen for 1.5 seconds, after which the following miniblock started.

The experiment consisted of a practice phase and an experimental phase. The practice phase consisted of two parts. First, we explained the main task (see the Appendix for the main instructions), and subjects could practice the NEXT and GO responses. Then we presented three miniblocks that consisted of the instruction, NEXT, GO, and feedback phases. The practice miniblocks consisted of 0, 1, or 2 NEXT trials (each number of NEXT trials occurred once, and the order was randomized).

The experimental phase consisted of 48 miniblocks. 24 miniblocks consisted of 1 NEXT trial, 16 consisted of 2 NEXT trials, and 4 consisted of 3 NEXT trials; in 4 miniblocks, the GO phase started immediately (so there were no NEXT trials). We used this trial distribution to make the start of the GO phase unpredictable and to encourage preparation. The order of the mini-blocks was further pseudo-randomized: two of the first 10 miniblocks were 0-NEXT blocks. Again, this was done to encourage preparation. Subjects received a break after every 12 miniblocks; they could determine the duration of the break themselves. The whole experiment lasted 10-15 minutes (although the youngest children sometimes took a little longer).

Dependent variables and analyses

All data processing and analyses were completed using R (R Development Core Team, 2016). Anonymized data files, R scripts, and experiment documentation are deposited on OSF (osf.io/am4yk).

For the NEXT analyses, we focused on the first NEXT (NEXT-1) trial because the instruction-based interference effect is largest on the first trial (Meiran et al., 2015a) and performance on later NEXT trials could already be modulated by stimulus-specific practice effects. We decided on this before data collection had started. We excluded miniblocks in which subjects made GO errors, as these could indicate that subjects did not process the instructions (resulting in a data loss of 17%). We focused on three dependent variables. First, we analyzed the probability of a correct NEXT-1 trial. Second, we analyzed the latency of the NEXT response with all (correct and incorrect) NEXT-1 trials included. This RT analysis was included in order to make the results comparable to Meiran et al. (2015a) who did not examine NEXT errors. Furthermore, this measure might be most sensitive as it combines all trials in which traces of inappropriate motor activity (Everaert, Theeuwes, Liefooghe, & De Houwer 2014; Meiran, Pereg, Kessler, Cole, & Braver, 2015b) cause interference, or in case the activity is high enough, an incorrect response. Third, we recalculated RT after exclusion of incorrect NEXT-1 trials. For both RT analyses, we used a trimming procedure: we excluded trials for which RT was < 100 ms or > 10 seconds; then we calculated the mean and standard deviation, and we excluded RTs which were 2.5 standard deviations above the mean. This trimming was done for each subject and condition separately. This resulted in an additional data loss of 3%. Table 2 shows the average number of trials for each condition and age group.

For the GO analyses, we focused on two dependent variables: accuracy and RT. For the RTs, we excluded incorrect GO trials and used the same trimming procedure as the one used for the NEXT analyses (combined, this resulted in a data loss of 15%).

Table 2: Average number of trials in the NEXT analysis each condition and age group.

	4	5	6	7	8	9	10	11	17-19
Compatible trials	15	18	18	19	18	18	18	19	20
Incompatible trials	14	15	16	16	17	16	16	18	18

For all variables, we analyzed performance using the ezANOVA function (Lawrence, 2016) in R with age (in years) as a continuous between-subjects variable and compatibility (the NEXT analyses) or trial number (first or second trial in the GO analyses) as categorical within-subjects variables. This analysis is very similar to a multiple regression with an interaction term or a standard ANCOVA (except that the continuous variable is typically considered a 'nuisance' variable in an ANCOVA, whereas the continuous variable is of main interest in the present study). We performed two sets of analyses. First, we performed the analyses with all subjects included. We grouped all adolescents together and used the same age value for all of them (i.e., 18). Table 3 provides an overview of these analyses. Second, we repeated the analyses without the adolescents in case this 'extreme' group had an unduly influence on inferential statistics. Table 4 provides an overview of these analyses. Note that the main outcomes of the two sets of analyses were similar.

In a pilot study with adults (N = 29; Supplementary Materials), we found medium to large instruction-based interference effects (Cohen's d_z: 0.65–1.00). Therefore, we also examined the main effect of compatibility for the different age groups. To increase power and reduce the number of significance tests, we combined the data of the 4-5, 6-7, 8-9, 10-11, and 17-19 year olds, resulting in 5 groups. Table 5 provides an overview of these analyses.

In the main analysis, we focused on the raw RT data. In Supplementary Materials, we report an analysis of proportional instruction-based interference scores. The main numerical trends were similar as in the analysis reported below.

Results

NEXT phase. We found large interference effects in all analyses: subjects made more errors and responded slower on incompatible trials than on compatible trials (Figure 2). This conclusion is supported by the inferential statistics (Tables 3 & 4). Furthermore, the RT analyses revealed general age-related differences. Most importantly, the RT analyses which included NEXT responses that came after erroneously pressing the wrong key, also revealed significant interactions between age and compatibility: the intention-based interference effect decreased over age, which is consistent with the 'hierarchical-control' account but inconsistent with the 'advance-implementation' account. This decrease can also be seen in Figure 2D, which shows how the intention-based interference effect is influenced by age and overall response speed. For the RT analysis that only included correct NEXT responses, the interaction was not significant ($p = .051$) when adolescents were included, but it was significant ($p = .002$) without them (i.e. when the 'extreme' group was excluded; see above). The interaction was not significant in both accuracy analyses (*p's* > .14). Table 5 shows that the instruction-based interference effect was significant for all measures and age groups.

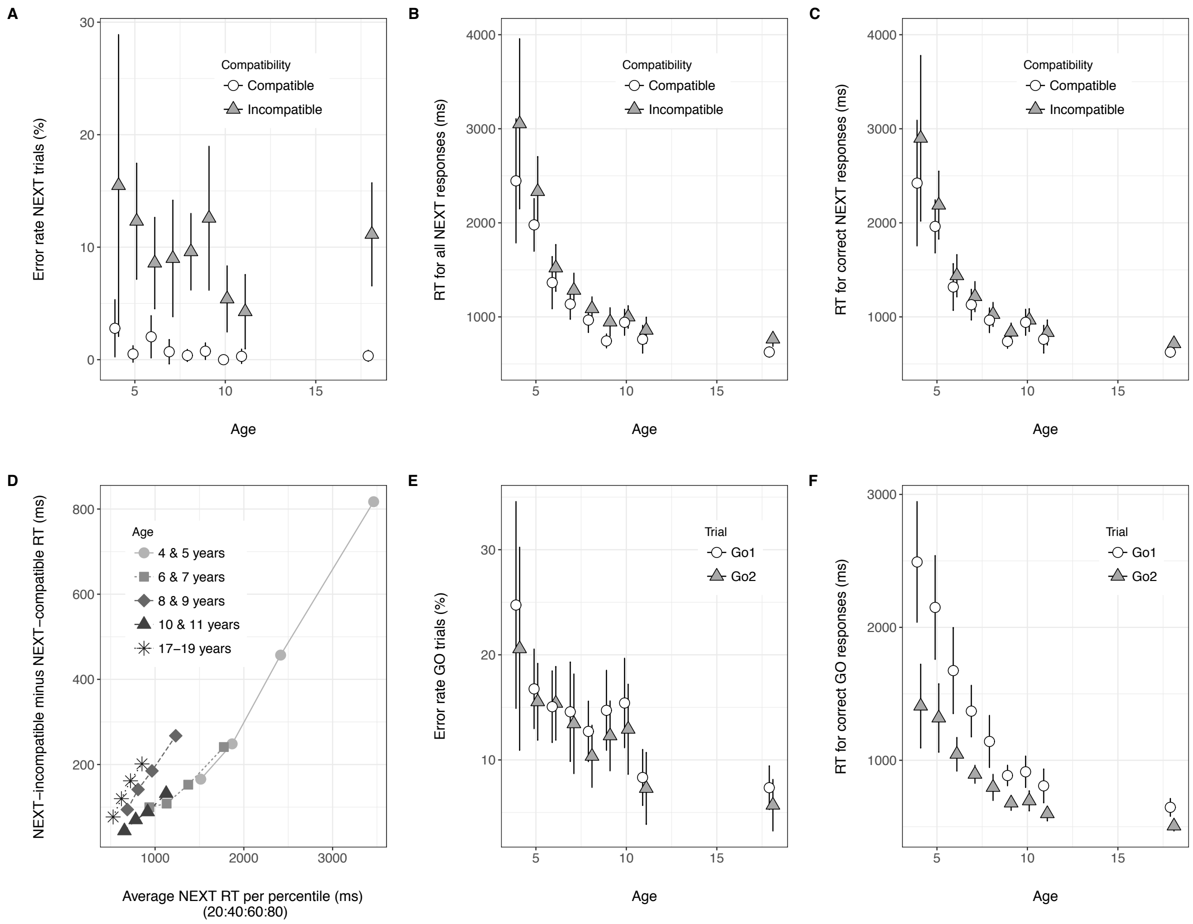


Figure 2: Overview of the NEXT and GO data (see Method section for a discussion of the different dependent variables). Panel D shows the NEXT-compatibility effect for the 20th, 40th, 60th, and 80th percentiles; this is for the analysis with all RT trials included.

GO phase. The GO analyses revealed that error rate and RT decreased over age, and that performance was generally worse on the first GO trial than on the second GO trial. The latter presumably reflects a task-switch cost (for reviews, see Kiesel et al., 2010; Vandierendonck, Liefooghe, & Verbruggen, 2010). The RT cost was larger for the younger children than for the older children and late adolescents, which is consistent with the previous literature (Chevalier & Blaye, 2009; Huizinga et al., 2006).

Table 3: Overview of the ANOVAs used to explore the effect of age, compatibility (NEXT) and trial number (GO1 or GO2) on performance. Age was a continuous numerical variable; thus, df1 = 1.

	Df	***Sum of squares effect***	***Sum of squares error***	***F***	***p***	***generalized η^2***
NEXT accuracy						
Age	1,194	0.003	1.439	0.463	.497	.001
Compatibility	1,194	0.815	1.372	115.281	**< .001**	.225
Age x Compatibility	1,194	0.000	1.372	0.015	.903	.000
NEXT RT (all NEXT responses included)						
Age	1,194	57461009	125243644	89.006	**< .001**	.297
Compatibility	1,194	3245765	10725756	58.707	**< .001**	.023
Age x Compatibility	1,194	278590	10725756	5.039	**.026**	.002
NEXT RT (correct NEXT responses only)						
Age	1,194	52994412	116227799	88.455	**< .001**	.300
Compatibility	1,194	1301115	7295666	34.598	**< .001**	.010
Age x Compatibility	1,194	145125	7295666	3.859	.051	.001
GO accuracy						
Age	1,194	0.435	2.531	33.334	**< .001**	0.138
Trial Number	1,194	0.026	0.189	26.967	**< .001**	0.010
Age x Trial Number	1,194	0.000	0.189	0.116	.734	0.000
GO RT						
Age	1,194	40974659	74749414	106.343	**< .001**	0.320
Trial Number	1,194	16296012	12334139	256.315	**< .001**	0.158
Age x Trial Number	1,194	3960352	12334139	62.291	**< .001**	0.043

Table 4: Overview of the ANOVAs used to explore the effect of age, compatibility (NEXT) and trial number (GO1 or GO2) on performance. Late adolescents and young adults were excluded from these analyses. Age was a continuous numerical variable; thus, df1 = 1.

	Df	***Sum of squares effect***	***Sum of squares error***	***F***	***p***	***generalized η^2***
NEXT accuracy						
Age	1,164	0.042	1.190	5.794	**.017**	.018
Compatibility	1,164	0.645	1.120	94.449	**< .001**	.218
Age x Compatibility	1,164	0.014	1.120	2.122	.147	.006
NEXT RT (all NEXT responses included)						
Age	1,164	73886993	89366724	135.593	**< .001**	.426
Compatibility	1,164	2995665	10084257	48.718	**< .001**	.029
Age x Compatibility	1,164	668873	10084257	10.878	**.001**	.007
NEXT RT (correct NEXT responses only)						
Age	1,164	67849107	83501518	133.258	**< .001**	.429
Compatibility	1,164	1190868	6852862	28.499	**< .001**	.013
Age x Compatibility	1,164	419350	6852862	10.036	**.002**	.005
GO accuracy						
Age	1,164	0.195	2.309	13.832	**< .001**	.073
Trial Number	1,164	0.022	0.169	21.404	**< .001**	.009
Age x Trial Number	1,164	0.000	0.169	0.449	.504	< .001
GO RT						
Age	1,164	42987226	56421758	124.950	**< .001**	0.393
Trial Number	1,164	17268650	9932973	285.117	**< .001**	0.207
Age x Trial Number	1,164	4930514	9932973	81.406	**< .001**	0.069

Table 5: Overview of planned comparisons to explore the NEXT-compatibility effect.

	diff	*lower CI*	*upper CI*	*df*	*t*	*p*	*BF*	g_{av}
NEXT effect accuracy								
4/5 olds	0.12	0.072	0.169	29	5.042	**< .001**	1003.94	1.582
6/7 olds	0.073	0.041	0.106	45	4.519	**< .001**	490.01	1.021
8/9 olds	0.105	0.071	0.14	58	6.119	**< .001**	158677.97	1.361
10/11 olds	0.047	0.025	0.068	30	4.394	**< .001**	205.99	1.402
17–19 olds	0.108	0.061	0.155	29	4.676	**< .001**	397.43	1.571
NEXT effect RT (all NEXT included)								
4/5 olds	422	192	651	29	3.761	**.001**	42.13	0.507
6/7 olds	152	54	249	45	3.131	**.003**	10.85	0.272
8/9 olds	162	99	224	58	5.197	**< .001**	6142.80	0.477
10/11 olds	77	41	113	30	4.362	**< .001**	189.97	0.291
17–19 olds	138	92	183	29	6.174	**< .001**	17831.86	0.772
NEXT effect RT (correct NEXT only)								
4/5 olds	292	86	499	29	2.890	**.007**	5.94	0.356
6/7 olds	105	29	182	45	2.771	**.008**	4.65	0.208
8/9 olds	81	44	118	58	4.389	**< .001**	423.35	0.265
10/11 olds	48	9	87	30	2.529	**.017**	2.86	0.185
17–19 olds	90	51	129	29	4.733	**< .001**	459.03	0.55

Note Table 5: Reported p values are uncorrected, but all t-tests were still significant after a Holm-Bonferroni correction. See Schönbrodt and Wagenmakers (2017) for a classification scheme for the interpretation of Bayes factors. We calculated the Bayes factors with the BayesFactor package in R, using the default prior (0.707). Hedges's average g (g_{av}) is the reported effect-size measure.

Exploratory analyses. We also ran an unplanned analysis to explore how the NEXT effect evolved throughout the experiment. The S-R mappings changed in every mini-block, so subjects could not practice the mappings. However, they could learn and practice the application of the overall task structure throughout the experiment. Both 'fast' and 'slow' learning mechanisms could produce such task- or structure-learning effects (Verbruggen, McLaren,& Chambers, 2014). Therefore, we repeated all NEXT analyses with experiment half (first 24 miniblocks vs. 24 last miniblocks) as an additional within-subjects variable. Because the number of trials was halved, we had to exclude some extra subjects from the RT analyses due to missing cells after data trimming (number of excluded subjects in the NEXT all-RT analysis: 1; in the NEXT correct-RT analysis: 6).

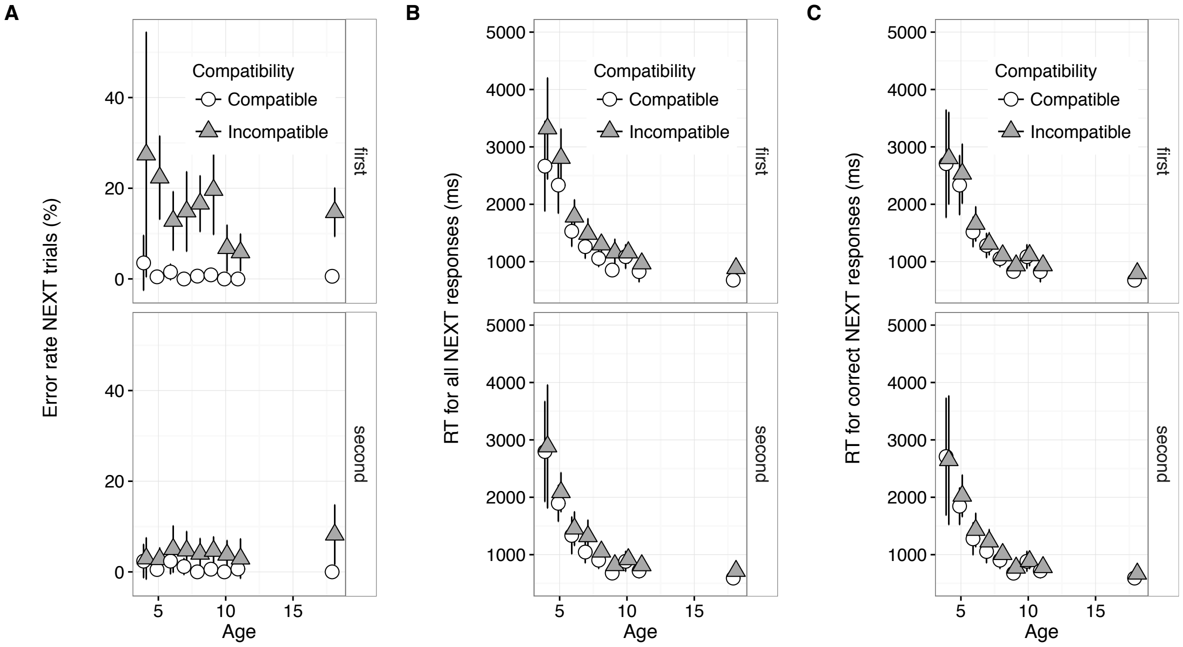


Figure 3: Overview of the NEXT data for the first and second half of the experiment (see Method section for a discussion of the different dependent variables).

The main RT analysis with all NEXT responses included (middle panels of Figure 3), revealed that the instruction-based interference effect decreased substantially throughout the experiment (NEXT effect first half: 267 ms; second half: 141 ms; $p = .007$, Table 6). A decrease was observed for all age groups, and the three-way interaction was non-significant, $p = .292$). The correct-RT analyses did not reveal any significant interactions between the interference effect and experiment half.

The accuracy analyses also showed that the interference effect decreased during the experiment (left panels of Figure 3; $p < .001$). Interestingly, significant three-way interactions were observed in the analyses with and without adolescents (Tables 6 & 7). Figure 3 shows that that the interference effect decreased more for younger children than for older children. This is consistent with the idea that young children have difficulties with the use of a hierarchical structure, but that this improves with some practice. However, it also shows that in the second part of the experiment, the effect was numerically largest for the late adolescents. It seems unlikely that this was due to a floor effect or a speed/accuracy trade-off (e.g. error rates were lower for the 11-year olds than for the late adolescents, yet their

NEXT RTs were comparable). Instead, this finding could reflect the costs of increased proactive control for the late adolescents. Indeed, GO performance was (numerically) better for the late adolescents. Thus, a possible explanation for these age-related differences is that late adolescents biased the GO task to a larger extent than the older children, leading to better GO performance but larger costs in the NEXT phase. Throughout the experiment, we used the feedback screens to encourage fast and correct GO performance, without mentioning NEXT performance. This could have induced a GO bias, and therefore, higher error rates in the NEXT phase. This highlights that 'proactive control' or rule implementation can come with certain costs, even in late adolescents.

Table 6: Overview of the ANOVAs used to explore the effect of age, compatibility, and experiment half on NEXT performance. Age was a continuous numerical variable; thus, df1 = 1.

	Df	***Sum of squares effect***	***Sum of squares error***	***F***	***p***	***generalized η^2***
NEXT accuracy						
Age	1,194	0.007	2.903	0.485	.487	.001
Compatibility	1,194	1.723	2.723	122.709	**< .001**	.153
Half	1,194	0.579	2.101	53.420	**< .001**	.057
Age x Compatibility	1,194	0.000	2.723	0.013	.909	.000
Age x Half	1,194	0.044	2.101	4.077	**.045**	.005
Compatibility x Half	1,194	0.582	1.788	63.156	**< .001**	.058
A x C x H	1,194	0.059	1.788	6.400	**.012**	.006
NEXT RT (all NEXT included)						
Age	1,193	132682582	280640291	91.248	**< .001**	.269
Compatibility	1,193	8124198	24107078	65.042	**< .001**	.022
Half	1,193	11404564	34832143	63.191	**< .001**	.031
Age x Compatibility	1,193	333362	24107078	2.669	.104	.001
Age x Half	1,193	1012000	34832143	5.607	**.019**	.003
Compatibility x Half	1,193	779212	20262113	7.422	**.007**	.002
A x C x H	1,193	117153	20262113	1.116	.292	.000
NEXT RT (correct NEXT only)						
Age	1,188	112052500	250087632	84.234	**< .001**	.267
Compatibility	1,188	1967512	14622138	25.297	**< .001**	.006
Half	1,188	6973384	26498290	49.475	**< .001**	.022
Age x Compatibility	1,188	28488	14622138	0.366	.546	.000
Age x Half	1,188	634154	26498290	4.499	**.035**	.002
Compatibility x Half	1,188	188	15825461	0.002	.962	.000
A x C x H	1,188	16862	15825461	0.200	.655	.000

Table 7: Overview of the ANOVAs used to explore the effect of age, compatibility, and experiment half on NEXT performance. Late adolescents and young adults were excluded from these analyses. Age was a continuous numerical variable; thus, df1 = 1.

	Df	*Sum of squares effect*	*Sum of squares error*	*F*	*p*	*generalized* η^2
NEXT accuracy						
Age	1,164	0.093	2.407	6.327	**.013**	.012
Compatibility	1,164	1.36	2.23	100.016	**< .001**	.148
Half	1,164	0.554	1.747	51.981	**< .001**	.066
Age x Compatibility	1,164	0.035	2.23	2.55	.112	.004
Age x Half	1,164	0.061	1.747	5.735	**.018**	.008
Compatibility x Half	1,164	0.578	1.427	66.402	**< .001**	.069
A x C x H	1,164	0.062	1.427	7.127	**.008**	.008
NEXT RT (all NEXT included)						
Age	1,163	171404286	197300316	141.606	**< .001**	.385
Compatibility	1,163	7358573	23003775	52.141	**< .001**	.026
Half	1,163	11360111	34047834	54.385	**< .001**	.04
Age x Compatibility	1,163	702690	23003775	4.979	**.027**	.003
Age x Half	1,163	830297	34047834	3.975	**.048**	.003
Compatibility x Half	1,163	744969	19654374	6.178	**.014**	.003
A x C x H	1,163	261020	19654374	2.165	.143	.001
NEXT RT (correct NEXT only)						
Age	1,158	141691900	181675437	123.227	**< .001**	.374
Compatibility	1,158	1674591	14136173	18.717	**< .001**	.007
Half	1,158	6879870	25890490	41.985	**< .001**	.028
Age x Compatibility	1,158	123479	14136173	1.38	.242	.001
Age x Half	1,158	622945	25890490	3.802	.053	.003
Compatibility x Half	1,158	4715	15657150	0.048	.828	.000
A x C x H	1,158	43	15657150	0	.983	.000

General Discussion

We examined structuring and implementing novel task instructions in children and late adolescents. We found that subjects' ability to prepare novel tasks improved with age, as seen in GO performance. However, this did not result in an age-related increase in intention-based interference effects: we found interference effects on NEXT-1 trials for all age groups, but these tended to be largest for the youngest children (4-5 year olds).

These results are consistent with the 'hierarchical-control' account. Situations in which multiple rules can be relevant (in our case, the NEXT and GO rules) require a hierarchical structure to determine the correct response and to reduce interference between competing task elements. Young children face difficulties with creating or using such structures (Amso et al., 2014; Unger et al., 2016). This could explain the larger instruction-based interference effects for the youngest children. The ‘hierarchical-structure’ account also receives support from another recent NEXT study (Meiran, Pereg, Givon, Danieli, & Shahar, 2016), which demonstrated that adults who were less successful in the GO phase, had poorer fluid intelligence, or were generally slower, also had a larger NEXT effect (i.e., adults with poorer working memory might also experience more problems with hierarchical or complex task sets, somewhat similar to children, than adults with better working memory). Meiran’s findings are also consistent with research on goal neglect, which suggests associations between fluid intelligence and the ability to ‘chunk’ task knowledge (Bhandari & Duncan, 2014).

Our results did not provide much support for the ‘advance-implementation’ account as described in the Introduction. Previous developmental work suggests that young children are less likely to implement task rules in advance than older children, adolescents, and young adults. Therefore, the ‘advance-implementation’ account predicted that GO performance would be impaired but the instruction-based interference effect in the NEXT phase should be absent (or at least be smaller) for the younger children. Instead, we observed the largest interference effects for the youngest children. The presence of the interference effects and decent GO performance indicate that even the youngest children of our sample could implement novel S-R rules in advance. This conclusion is consistent with a study showing that young children engaged in proactive control (i.e. they prepared rules in advance) when the task was more difficult (Chevalier, Martis, Curran, & Munakata, 2015). Here we used novel S-R mappings in each miniblock. This prevented stimulus-specific practice and the consequent formation of long-term memory traces, and could have encouraged the implementation of the rules during the instruction phase. However, consistent with the results of Blackwell and Munakata (2012), implementing these rules came with a substantial cost in young children (i.e. large interference effects during the NEXT phase).

The exploratory analyses revealed that the instruction-based interference effects (in the accuracy and main RT analyses) decreased throughout the experiment. In the accuracy analyses, this effect was most pronounced for the youngest children. The decrease is consistent with findings in adults (Meiran et al., 2015a). In NEXT experiments, subjects cannot learn specific S-R associations. However, they may gradually get better at 'separating' the GO phase (indicated by the dark blue background) from the NEXT phase (indicated by the light blue background). In other words, we speculate that hierarchical structures (with the context cue modulating the choice options) and their usage further evolved throughout practice, reducing interference between the GO and NEXT components of the task. This idea is consistent with other findings in the task-learning literature (Bhandari et al., 2017).

By contrasting the hierarchical-structure and advance-implementation accounts, readers may get the incorrect impression that the task-formation and -implementation phases are independent. But when people create an inefficient non-hierarchical structure or when they have difficulties managing the contingencies within the structure, more competition between the various choice options occurs (producing larger instruction-based interference effects). Thus, task structure will have knock-on effects on the implementation stage. Interestingly, goal neglect (i.e. the dissociation between knowing and doing) has also been associated with the formation of inefficient task structures (Bhandari & Duncan, 2014). This raises the intriguing possibility that failing to implement or execute a task (i.e. goal neglect; a negative 'symptom') and applying the rules when not required (i.e. instruction-based interference; a positive 'symptom') both arise from a failure to create an efficient task structure. Future research is needed to test how these phenomena are related.

To conclude, we observed intention-based interference effects in all groups, indicating that even the younger children of our sample implemented novel rules at the beginning of each miniblock. We attribute the (numerically) larger RT costs to age-related differences in the creation of hierarchical task structures. Furthermore, we propose that the NEXT paradigm might be a useful tool to study structuring and implementation of instructions in different age groups, and more generally, the powerful effects instructions and intentions can have on behavior.

References

Amso, D., Haas, S., McShane, L., & Badre, D. (2014). Working memory updating and the development of rule-guided behavior. *Cognition, 133*, 201–10. https://doi.org/10.1016/j.cognition.2014.06.012

Bhandari, A., Badre, D., & Frank, M. J. (2017). Learning Cognitive Control. In T. Egner (Ed.), *The Wiley Handbook of Cognitive Control* (pp. 127–142). Chichester, West Sussex, UK: John Wiley & Sons.

Bhandari, A., & Duncan, J. (2014). Goal neglect and knowledge chunking in the construction of novel behaviour. *Cognition, 130*, 11–30. https://doi.org/10.1016/j.cognition.2013.08.013

Bunge, S. A., & Crone, E. A. (2009). Neural correlates of the development of cognitive control. In J. M. Rumsey & M. Ernst (Eds.), *Neuroimaging in developmental clinical neuroscience* (pp. 22–37). New York City, NY: Cambridge University Press.

Bunge, S. A., & Zelazo, P. D. (2006). A brain-based account of the development of rule use in childhood. *Current Directions in Psychological Science, 15(3)*, 118–121. https://doi.org/10.1111/j.0963-7214.2006.00419.x

Chevalier, N., Martis, S. B., Curran, T., & Munakata, Y. (2015). Metacognitive Processes in Executive Control Development: The Case of Reactive and Proactive Control. *Journal of Cognitive Neuroscience, 27*, 1125-1136. https://doi.org/10.1162/jocn_a_00782

Chevalier, N., & Blaye, A. (2009). Setting goals to switch between tasks: Effect of cue transparency on children’s cognitive flexibility. *Developmental Psychology, 45(3)*, 782–797. https://doi.org/10.1037/a0015409

Cole, M. W., Laurent, P., & Stocco, A. (2013). Rapid instructed task learning: A new window into the human brain’s unique capacity for flexible cognitive control. *Cognitive Affective & Behavioral Neuroscience, 13*, 1–22. https://doi.org/10.3758/s13415-012-0125-7

Cole, M. W., Meiran, N., & Braver, T. S. (2017). The task novelty paradox: Flexible control of inflexible neural pathways during rapid instructed task learning. *Neuroscience & Biobehavioral Reviews*. https://doi.org/10.1016/j.neubiorev.2017.02.009

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

Everaert, T., Theeuwes, M., Liefooghe, B., & De Houwer, J. (2014). Automatic motor activation by mere instruction. *Cognitive, Affective, & Behavioral Neuroscience, 14*, 1300–1309. https://doi.org/10.3758/s13415-014-0294-7

Huizinga, M., Dolan, C. V., & van der Molen, M. W. (2006). Age-related change in executive function: developmental trends and a latent variable analysis. *Neuropsychologia, 44(11)*, 2017–36. https://doi.org/10.1016/j.neuropsychologia.2006.01.010

Kiesel, A., Steinhauser, M., Wendt, M., Falkenstein, M., Jost, K., Philipp, A. M., & Koch, I. (2010). Control and interference in task switching-A review. *Psychological Bulletin, 136(5)*, 849–74. https://doi.org/10.1037/a0019842

Meiran, N., Liefooghe, B., & De Houwer, J. (2017). Powerful instructions: Automaticity without practice. Current Directions in Psychological Science. Retrieved from https://doi.org/10.1177/0963721417711638

Meiran, N., Pereg, M., Givon, E., Danieli, G., & Shahar, N. (2016). The role of working memory in rapid instructed task learning and intention-based reflexivity: An individual differences examination. *Neuropsychologia, 90*. 180-189. https://doi.org/10.1016/j.neuropsychologia.2016.06.037

Meiran, N., Pereg, M., Kessler, Y., Cole, M. W., & Braver, T. S. (2015a). The power of instructions: Proactive configuration of stimulus–response translation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 41(3)*, 768–786. https://doi.org/10.1037/xlm0000063

Meiran, N., Pereg, M., Kessler, Y., Cole, M. W., & Braver, T. S. (2015b). Reflexive activation of newly instructed stimulus–response rules: evidence from lateralized readiness potentials in no-go trials. *Cognitive, Affective, & Behavioral Neuroscience, 15*, 365–373. https://doi.org/10.3758/s13415-014-0321-8

Munakata, Y., Snyder, H. R., & Chatham, C. H. (2012). Developing Cognitive Control: Three Key Transitions. *Current Directions in Psychological Science, 21(2)*, 71–77. https://doi.org/10.1177/0963721412436807

Schönbrodt, F. D., & Wagenmakers, E.-J. (2017). Bayes factor design analysis: Planning for compelling evidence. *Psychonomic Bulletin & Review*. https://doi.org/10.3758/s13423-017-1230-y

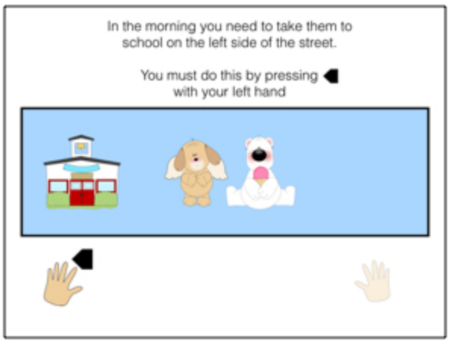
Unger, K., Ackerman, L., Chatham, C. H., Amso, D., & Badre, D. (2016). Working memory gating mechanisms explain developmental change in rule-guided behavior. *Cognition, 155*, 8–22. https://doi.org/10.1016/j.cognition.2016.05.020

Vandierendonck, A., Liefooghe, B., & Verbruggen, F. (2010). Task switching: Interplay of reconfiguration and interference control. *Psychological Bulletin, 136(4)*, 601–26. https://doi.org/10.1037/a0019791

Verbruggen, F., McLaren, I. P. ., & Chambers, C. D. (2014). Banishing the Control Homunculi in Studies of Action Control and Behavior Change. *Perspectives on Psychological Science, 9(5)*, 497–524. https://doi.org/10.1177/1745691614526414

Zelazo, P. D., Frye, D., & Rapus, T. (1996). An age-related dissociation between knowing rules and using them. *Cognitive Development, 11(1)*, 37–63. https://doi.org/10.1016/S0885-2014(96)90027-1

Appendix: Task instructions.



In the morning you need to take them to school on the left side of the street.

You must do this by pressing ◀ with your left hand

After school, your friends need to get home as soon as possible.
If it is too dark outside, they may lose their way and their mum will become really worried!

Some live on the left side of the street, so you need to press ◀

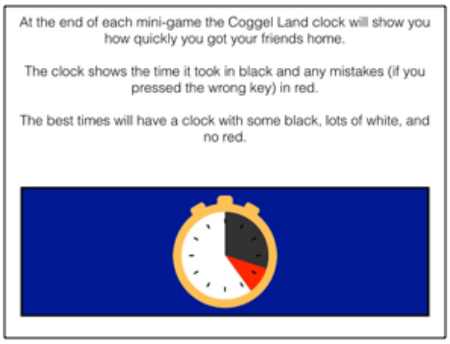
Some live on the right side of the street, so you need to press ▶

This picture shows you which side of the street two of your friends live on.
Dusty Dog lives on the left side.
Barnaby Bear lives on the right side.

So you need to press ◀ with your left finger when you see Dusty Dog
and ▶ with your right finger when you see Barnaby Bear

There are lots of friends in Coggel Land.
At the beginning of each mini-game we will show you which friend lives on the left side of the street (◀) and which friend lives on the right side of the street (▶).

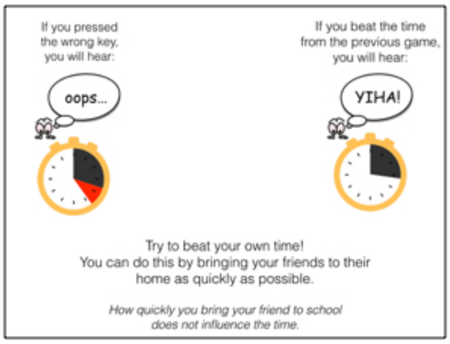
For each mini-game you must remember really well where your friends live and make sure they get home as quickly as possible.
Otherwise, they may lose their way.



At the end of each mini-game the Coggel Land clock will show you how quickly you got your friends home.

The clock shows the time it took in black and any mistakes (if you pressed the wrong key) in red.

The best times will have a clock with some black, lots of white, and no red.

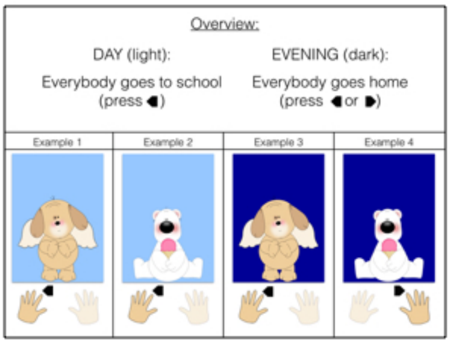


If you pressed the wrong key, you will hear:

If you beat the time from the previous game, you will hear:

Try to beat your own time!
You can do this by bringing your friends to their home as quickly as possible.

How quickly you bring your friend to school does not influence the time.



Overview:

DAY (light):
Everybody goes to school (press ◀)

EVENING (dark):
Everybody goes home (press ◀ or ▶)

Overview:

We will show the instructions at the beginning of each mini-game.

You can start each game by pressing a key (but you have to wait at least three seconds).

We will start with three short 'practice' games, followed by 48 games for real.

Good luck!