Developmental Science 8:4 (2005), pp 360–371

PAPER

Factors responsible for performance on the day–night task: response set or semantics?

Andrew Simpson and Kevin J. Riggs

Department of Psychology, London Metropolitan University, UK

Abstract

In a recent study Diamond, Kirkham and Amso (2002) obtained evidence consistent with the claim that the day–night task requires inhibition because the picture and its corresponding conflicting response are semantically related. In their study children responded more accurately in a dog–pig condition (see /day picture/ say 'dog'; see /night picture/ say 'pig') than the standard day–night condition (see /day picture/ say 'night'; see /night picture/ say 'day'). However, there is another effect that may have made the day–night condition harder than the dog–pig condition: the response set effect. In the day–night condition the names of the two stimuli ('day' and 'night') and the two corresponding conflicting responses ('night' and 'day') are from the same response set: both 'day' and 'night'. In the dog–pig condition the names of the stimuli ('day', 'night') and the corresponding responses ('dog', 'pig') are from a different response set. In two experiments (Experiment 1 with 4-year-olds (n = 25); Experiment 2 with $3\frac{1}{2}$, 4-, 5-, 7- and 11-year-olds (n = 81)) children were tested on four experimental conditions that enabled the effects of semantics and response set to be separated. Overall, our data suggest that response set is a major factor in creating the inhibitory demands of the day–night task in children of all ages. Results are discussed in relation to other inhibitory tasks.

Introduction

A broad range of tasks have been used to assess inhibitory control in young children (see Nigg, 2000, for a recent review). Children's performance on many of these tasks improves dramatically between 3 and 5 years. For example, performance on the day–night (Gerstadt, Hong & Diamond, 1994), bear–dragon (Jones, Rothbart & Posner, 2003), card sort (Fyre, Zelazo & Palfai, 1995), windows (Russell, Mauthner, Sharpe & Tidswell, 1991) and false belief (Carlson, Moses & Hix, 1998) task is very poor in most 3-year-olds but often error-free by 5 years. However, a major problem for researchers investigating inhibitory development is that in many of the above tasks it is unclear whether inhibitory development is responsible for improved performance (see for example, the card sort (Kloo & Perner, 2003) and windows (Simpson & Riggs, 2004) tasks).

Perhaps the best evidence that inhibition develops between 3 and 5 years comes from the day–night task (Gerstadt *et al.*, 1994; Diamond *et al.*, 2002; Simpson & Riggs, in press). In the original day–night version of this task children were presented with either a day picture of the sun, or a night picture of the moon (and some stars) and were instructed to say 'night' to the day picture and 'day' to the night picture (Gerstadt *et al.*, 1994). The 'prepotent' response of naming the stimulus must be suppressed in order to produce the 'conflicting' response required by the rule. Therefore the task requires *both* inhibition of a prepotent response and the generation of a conflicting response from working memory. It is therefore possible that performance might improve in the day–night task because: (i) the ability to inhibit the prepotent response improves (either in isolation or in conjunction with the working memory demands of the task) or, (ii) the ability to generate the conflicting response from working memory improves. Gerstadt *et al.* (1994) aimed to separate these two factors with an ingenious control condition. In the control condition the child had to remember specific responses to two abstract pictures. The requirement to generate a response from working memory was the same in both conditions, but because there were no names associated with the abstract pictures there were no prepotent responses to inhibit. By comparing performance in the inhibitory condition and the control condition, Gerstadt *et al.* (1994) obtained

Address for correspondence: Kevin J. Riggs, Department of Psychology, London Metropolitan University, Old Castle Street, London E1 7NT, UK; e-mail: k.riggs@londonmet.ac.uk

convincing evidence to suggest that it is the ability to inhibit the prepotent response (either in isolation or in conjunction with the working memory demands) rather than the sole ability to generate the conflicting response, which develops in children aged $3\frac{1}{2}$ to 5 years old.

In a more recent study Diamond, Kirkham and Amso (2002) investigated why the day–night task requires inhibition. In particular they focused on the relation between each stimulus and its conflicting response. In their study, 4-year-old children responded more accurately in a dog–pig condition (see |day picture| say 'dog', see |night picture| say 'pig') than the standard day–night condition. On the basis of these results they suggested that the day–night condition required inhibition because each stimulus was semantically related to (and the direct opposite of) its corresponding conflicting response. The day picture was semantically related to the response 'night' and the night picture was semantically related to the response 'day'. In the dog–pig condition the same stimuli (|day|, |night|) were not semantically related to their corresponding conflicting responses ('dog', 'pig'). The Diamond *et al.* (2002) proposal is compatible with a wealth of adult data on other versions of the Stroop task that provide robust evidence for a strong semantic effect (see MacLeod, 1991, for a review).

However, there is another effect that may have made the Diamond *et al.* (2002) day–night condition harder than the dog–pig condition: the response set effect. In the standard colour-conflict Stroop, the response set effect is that colour naming is harder (takes longer) if all the ink colours and words used in the experiment are the same (e.g. words: *red*, *blue*, *brown*, *orange*; and ink colours: *red*, *blue*, *brown*, *orange*) than if they are different (e.g. words: *red*, *blue*, *brown*, *orange*; and ink colours: *green*, *yellow*, *black*, *white*; see for example, Proctor, 1978; Stirling, 1979, and La Heij, 1988). The better known effect is the semantic effect: naming the ink colour of a word on a particular trial is harder if that word is a (semantically related) colour word, than if it is a (semantically unrelated) non-colour word (e.g. Klein, 1964; Dalrymple-Alford, 1972). The semantic effect is expressed at a different level than the response set effect (see Kornulum, Hasbroucq & Osman, 1990, for a review). The semantic effect is only expressed on a trial-by-trial basis: it reflects the relation between a stimulus and its response on a particular trial. The response set effect in contrast is dependent on the relation between all the stimuli and all the responses in an experiment.

In light of the adult literature it seems reasonable to suggest that response set may be a factor in determining the need for inhibition in the day–night task. In the day–night condition the names of the two stimuli ('day' and 'night') and the two conflicting responses ('night' and 'day') are from the same response set. In the dog–pig condition the names of the stimuli ('day', 'night') and the conflicting responses ('dog', 'pig') are from different response sets. In fact the day–night methodology may promote a response set effect. The child is made explicitly aware of the response set before the experiment starts. When the experimenter explains the rules the child is told the stimuli and responses. This is different from the standard colour-conflict Stroop where response set is only revealed implicitly during the course of the task. Knowing the response from the outset may enable the child to 'pre-activate' the responses required by the rules. So for example, in the day–night condition the child may pre-activate the responses 'night' and 'day'. In addition, with only two rules to remember the child will always face the problem that the incorrect pre-activated response will be triggered when a stimulus is presented. When the child sees the night picture it will trigger the incorrect pre-activated response 'night' and when the child sees the day picture it will trigger the incorrect pre-activated response 'day'. Knowing the response set before the experiment starts and having only two rules to remember may enhance the response set effect in the day–night task.

Many putative inhibitory tasks, like the day–night task, contain a known response set based on two rules. For example, the grass–snow (Carlson & Moses, 2001), hand game (Hughes, 1998), card sort (Fyre *et al.*, 1995), tapping (Diamond & Taylor, 1996) and Simon (Diamond, 2003) task. It is notable that in all these tasks the stimuli and conflicting responses are from the same response set. For example, in the hand game task the child is told to point a finger when the experimenter makes a fist and make a fist when the experimenter points a finger. The stimuli (|make a fist|, |point a finger|) and the conflicting responses (point a finger, make a fist) are from the same response set. Response set effects may contribute to making responses prepotent in all these tasks.

The question why responses are prepotent in the day–night task is an interesting one. Despite a current focus on inhibitory development very few researchers have asked why responses are prepotent in inhibitory tasks. This information is essential if we are to have a rational basis for identifying inhibitory tasks. There is also the concern that while inhibitory demands are probably high in some 'contrived' developmental tasks, it is unclear how performance in these tasks relates to children's everyday behavior. It is only if children experience the circumstances that create prepotent responses, that their weak inhibitory control will influence their thinking and behaviour. This study therefore aims to contribute to an understanding of the circumstances that create prepotent responses.

Experiment 1

In our first study we investigated the role of semantics and response set in the day–night task. Diamond *et al.* (2002) tested a 4- and a $4\frac{1}{2}$-year-old age group. The most striking results were obtained with the 4-year-olds (presumably because of their poorer inhibitory control). We therefore tested 4-year-olds with an age range similar to those in the Diamond *et al.* (2002) study.

In the Diamond *et al.* (2002) study, semantics and response set were confounded. In the standard day–night condition each stimulus was semantically related to its conflicting response; it was also the case that the two stimuli and two responses were from the same response set. In the dog–pig condition the stimuli and conflicting responses were neither semantically related (e.g. |day picture| is not semantically related to 'dog') nor from the same response set (the responses 'day' and 'night' are different from 'pig' and 'dog'). In our first study we devised four experimental conditions that allowed us to separate out the factors of semantics and response set. We also administered a control condition similar to that used in Gerstadt *et al.* (1994).

We wished to attribute any differences between the four experimental conditions to differences in inhibitory demands. For this to be possible the working memory demands of the experimental conditions must not differ from each other. To check rule memory we asked children two memory questions at the end of each condition. A possible concern here, as Munakata, Morton and Yerys (2003) have suggested, is that a child's ability to demonstrate rule knowledge when tested one way does not prove that it is sufficient for good performance when tested another way. That is, even if children answered the memory questions at the end of the day–night task it does not prove that rule memory was sufficient to support accurate responding during that task. However, our concern here was simply to estimate the *relative* strengths of rule memory in the four experimental conditions. The memory questions used in Experiment 1 provided such an estimate.

Method

Participants

Twenty-five children took part in the study: 15 boys and 10 girls. The ages ranged from 3 years, 10 months to 4 years, 2 months (M = 4 years, 0 months). This matched the 4-year-old age groups employed in the Diamond *et al.* (2002) study. The children attended a nursery school in the inner city borough of Tower Hamlets in London, England. All spoke English as a first language and none had behavioural or educational problems. The group was ethnically mixed and from a predominantly working-class background.

Design

A repeated-measures design was used. The factors were Task condition (Experimental, Control) and for the experimental conditions two further factors of Response set (Same, Different) and Semantics (Related, Unrelated). The dependent variables were response accuracy (out of 16 trials) and rule memory for the experimental conditions (out of two questions).

Materials

Ten pictures were used (moon, sun, car, book, black square, white square, blue square, red square, two abstract patterns), two for each condition (see Figure 1). The laminated picture cards used to introduce the stimuli measured 22 cm by 15 cm. The practice and test stimuli were presented in a flipbook with one stimulus on each page.

Procedure

Children were tested on a control condition and four experimental conditions (see Table 1). In the control condition children responded by saying 'cup' or 'boat' to two abstract pictures. For the four experimental conditions,

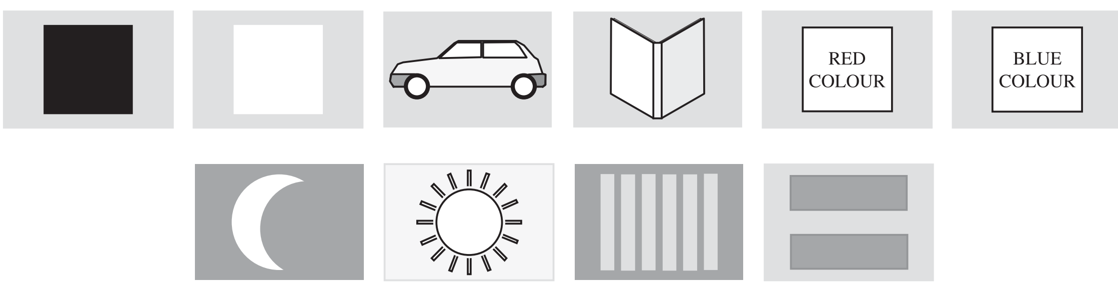

Figure 1 *Stimuli used in Experiment 1 (from top left: black square, white square, car, book, red square, blue square, moon, sun, abstract picture 1, abstract picture 2).*

Table 1 *The four experimental conditions used in Experiment 1 and Experiment 2*

Response set	Semantic relation between a stimulus and its conflicting response: Semantically related	Semantically unrelated
Stimuli and responses from the same response set	*same-response-set/semantically-related condition* see |red| say 'blue' see |blue| say 'red'	*same-response-set/semantically-unrelated condition* see |car| say 'book' see |book| say 'car'
Stimuli and responses from a different response set	*different-response-set/semantically-related condition* see |black| say 'yellow' see |white| say 'green'	*different-response-set/semantically-unrelated condition* see |sun| say 'pig' see |moon| say 'dog'

there were two conditions in which the stimuli and conflicting responses were from the same response set and two in which they were from a different response set. For the two conditions in which the stimuli and conflicting responses were from the same response set, there was one condition where each stimulus was semantically related to the corresponding conflicting response; the *same-response-set/semantically-related* condition (see |red| say 'blue', see |blue| say 'red') and one condition where they were semantically unrelated; the *same-response-set/semantically-unrelated* condition, (see |book| say 'car', see |car| say 'book'). For the two conditions in which the stimuli and conflicting responses were from a different response set, there was one condition where each stimulus was semantically related to the conflicting response; the *different-response-set/semantically-related* condition (see |black| say 'yellow', see |white| say 'green') and one condition where they were semantically unrelated; the *different-response-set/semantically-unrelated* condition (see |sun| say 'dog', see |moon| say 'pig').

Accuracy on the four experimental conditions was compared to accuracy on the control condition. It was assumed that if the four experimental conditions required inhibition then accuracy would be worse than in the control condition. If semantics determined inhibitory demands then accuracy would be worst in the two conditions in which each stimulus was semantically related to its corresponding conflicting response (red–blue/blue–red and black–yellow/white–green). If response set determined inhibitory demands then accuracy would be worst in the two conditions with stimuli and responses from the same response set (red–blue/blue–red and car–book/book–car). If a mixture of the two factors determined inhibitory demands then performance should be worst on the *same-response-set/semantically-related* condition (red–blue/blue–red) and best on the *different-response-set/semantically-unrelated* condition (sun–dog/moon–pig), with performance on the other two conditions somewhere in between.

Children were tested across three sessions over a period of not more than five days. The order of condition presentation was counterbalanced. Five different orders were used based on the order *same-response-set/semantically-related*, *different-response-set/semantically-unrelated*, *same-response-set/semantically-unrelated*, *different-response-set/semantically-related*, control. This order was used so no child received two same response set or two semantically related conditions consecutively.

All children were individually tested in a quiet area within the class or in an adjoining room or corridor. The child and experimenter sat next to each other at a table. The experimenter explained that they were going to play a 'silly game' in which they would see two pictures. For each condition the child was sequentially introduced to two pictures on laminated cards. The child was told how to respond to the picture, and asked to repeat back the response. All children were able to repeat the responses. For experimental conditions children were explicitly told not to give the veridical response.

The child was reminded of the rules before the pre-test session began. In the pre-test session the child was then given six practice trials with feedback (order ABABAB). If, for example, the child said 'pig' to the moon picture the experimenter confirmed that this was the correct response. If, however, the child responded 'moon' he or she was told that this was wrong because the correct response was to say 'pig' to moon. In the test session the child saw 16 cards, bound in a flip-book, in the same pseudo-random order (ABBABAABBABAABAB). No feedback was given if the child responded incorrectly in the test session. At the end of each experimental condition the child was asked two memory questions to check they could remember the rules, an example of one of these questions being, 'So . . . what did you have to say to the black square?' It was not possible to test rule memory for the control condition because the abstract pictures could not be verbally labelled.

Results and discussion

Accuracy

Accuracy in the four experimental and control conditions is shown in Figure 2. Performance was best in the control condition at 91%. Accuracy in the two conditions in which the stimuli and responses were from the same response set (red–blue/ blue–red and car–book /book–car) was identical at 78%. Accuracy was at 86% in the *different-response-set/semantically-unrelated* condition (sun–dog/ moon–pig) and 70% in the *different-response-set/ semantically-related* condition (black–yellow/white–green). Accuracy in the four experimental conditions was analyzed in a repeated-measures ANOVA with Response set (Same, Different) and Semantics (Related, Unrelated) as factors. There was a significant main effect for Semantics ($F(1, 24) = 6.61$, $p = .017$) but not Response set ($F(1, 24) = .05$, $p = .832$) and also a significant interaction ($F(1, 24) = 6.39$, $p = .019$). Planned comparisons revealed that this interaction was due to 4-year-olds being more accurate in the *different-response-set/semantically-unrelated* condition (sun–dog/moon–pig) than the other three experimental conditions (*same-response-set/semantically-related*, $t(24) = 2.08$, $p = .048$; *same-response-set/ semantically-unrelated*, $t(24) = 2.42$, $p = .024$; *different-response-set/semantically-unrelated*, $t(24) = 3.63$, $p = .001$). There were no significant effects involving gender or presentation order of condition.

To test the inhibitory demands of each experimental condition we compared accuracy in that condition to accuracy in the control condition (see Figure 2). Statistical analysis revealed that three of the experimental conditions were harder than the control (*same-response-set/semantically-related* condition: $t(24) = 4.04$, $p < .001$; *same-response-set/semantically-unrelated* condition: $t(24) = 3.32$, $p = .003$; *different-response-set/semantically-related* condition: $t(24) = 4.64$, $p < .001$). Performance in the *different-response-set/semantically-unrelated* condition was indistinguishable from the control ($t(24) = 0.738$, $p = .467$).

Taken together these data suggest, and support the Diamond *et al.* (2002) proposal, that *any* relation between the stimulus and its corresponding response will suffice to create inhibitory demands in the day–night task for 4-year-olds. They do not, however, support the specific proposal that the day–night task requires inhibition only if the stimulus and its conflicting response are semantically related. Accuracy on the *same-response-set/ semantically-unrelated* condition (car–book/book–car) was significantly worse than the control. This condition seemed to require inhibition even though each stimulus was not semantically related to its conflicting response.

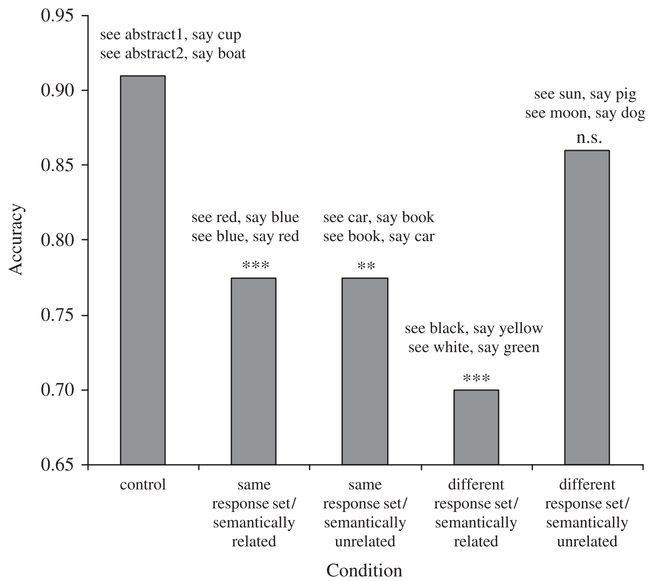

Figure 2 *Accuracy of 4-year-olds on the four experimental conditions and the control condition in Experiment 1. Significance values compare each experimental condition to the control (** is* $p < .01$, **** is* $p < .001$*).*

Neither do these data support the specific proposal that the day–night task requires inhibition only if the stimuli and responses belong to the same response set. Accuracy on the *different-response-set/semantically-related* condition (black–yellow/green–white) was significantly worse than the control and worse than the other three experimental conditions. This condition seemed to require inhibition even though 'yellow' and 'green' were from a different response set than black and white.

Memory questions

It could be argued that the rules used in some of the experimental conditions here were harder to remember than others. The two conditions with different response sets (black–yellow/white–green and sun–dog/ moon–pig) required memory for four items (e.g. black, yellow, white, green) whereas the two conditions with the same response set (red–blue/blue–red and car–book/book–car) required memory for only two items (e.g. red, blue).

Children's memory question performance is shown in Figure 3. Memory for the two conditions in which the stimuli and responses were from the same response set (red–blue/blue–red and car–book/book–car) was good at around 90%. Memory for the *different-response-set/ semantically-unrelated* condition (sun–dog/moon–pig) was a little poorer at 84%. Memory for the *different-response-set/semantically-related* condition (black–yellow/white–green) was worst at 68%. Statistical analysis revealed

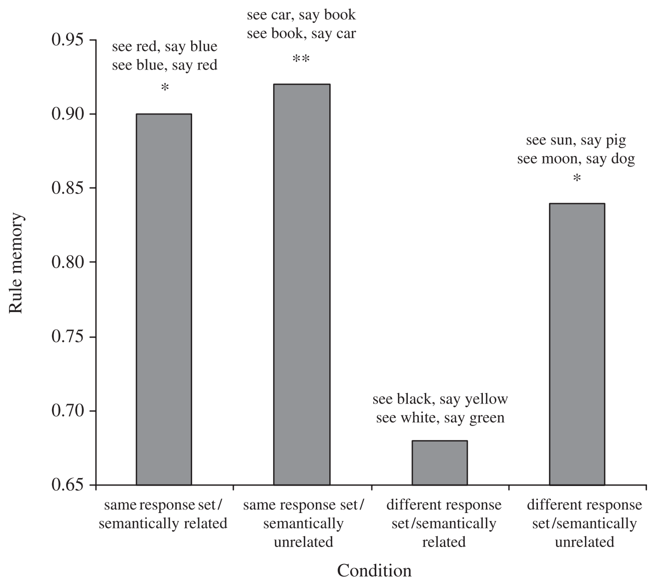

Figure 3 *Accuracy of 4-year-olds on the memory questions in Experiment 1. Significance values compare each condition to the* different-response-set/semantically-related *condition (* is* p < *.05, ** is* p < *.01).*

that memory for the task rules in the *different-response-set/semantically-related* condition was significantly worse than the other three conditions (Wilcoxon signed ranks test: *same-response-set/semantically-related* condition, $Z = 2.18$, $p = .029$; *same-response-set/semantically-unrelated* condition, $Z = 2.51$, $p = .012$; *different-response-set/semantically-unrelated* condition, $Z = 1.99$, $p = .052$).

These results suggest that children had particular problems remembering the rules say 'yellow' to |black| and 'green' to |white| in the *different-response-set/semantically-related* condition. This in turn raises the possibility that accuracy on this condition was poor because of higher memory demands rather than higher inhibitory demands. Why might children have more difficulty remembering the rules in this condition compared to the others, especially the four items in the *different-response-set/semantically-unrelated* condition? One possibility is that these children found it more difficult to remember the pairing of four items that were semantically related than the four items that were semantically unrelated.

Use of different stimuli in different conditions

Finally, we respond to a point made by one of the reviewers. A potential problem with this design is that different stimuli were used in different conditions – some stimuli were colours and some were objects. It was, however, impossible to use identical stimuli in all the conditions. The stimuli in the *same-response-set/semantically-related* condition had to be semantically related to each other, and the stimuli in the *same-response-set/semantically-unrelated* condition had to be semantically unrelated to each other. This cannot be done with the same stimuli. Moreover, previous research suggests that day–night task performance is unaffected by changes in the stimuli. For example, Simpson and Riggs (in press) observed that children found inhibiting an associate, as in the original day–night condition (e.g. not saying 'day' to a picture of a sun), a colour name (e.g. not saying 'black' to a black square) or an object name (e.g. not saying 'sun' to a picture of a sun) equally difficult.

Experiment 2

The data from the first study provided evidence for both a response set effect and a semantic effect in 4-year-olds' performance on the day–night task. However, the analysis was complicated because one of the four experimental conditions seemed to have higher working memory demands than the others. In our second study we sought to overcome the problems caused by working memory demands so that the role of response set and semantics could be examined more easily.

In the second study we modified the day–night task. Children were given feedback and reminded of the task rules if they made an error. For example, if the child said 'black' to a black square then the experimenter explained that they should have said 'white'. Our rationale was that reminding children of the rules would help them to remember those rules and so reduce the working memory demands of the task. Giving feedback may also have benefited performance in other ways such as by increasing their motivation to persevere with the task. Children of $3\frac{1}{2}$, 4 (replicating Experiment 1), 5, 7 and 11 years were tested. We wished to investigate how response set and semantics affected performance across childhood. Previous data suggest that *accuracy* on the day–night task is near perfect by 5 years (Gerstadt *et al.*, 1994) and that inhibition cannot be assessed using this measure at this age or above. We therefore used a reaction time measure to assess inhibition in this study.

Method

Participants

Eighty-one children took part in the study (see Table 2). For data analysis involving age this sample was split into five groups of $3\frac{1}{2}$-, 4-, 5-, 7- and 11-year-olds. The children attended a nursery or the adjoining primary school in the inner city borough of Tower Hamlets in London,

Table 2 *Details of children participating in Experiment 2*

Age group	*N*	Gender M/F	Mean (yrs)	SD (yrs)	Range (Y;M)	Number excluded due to failure of practice trials
$3^1/_2$	20	11/9	3.5	0.37	3;3–3;9	5
4	16	6/10	4.0	0.15	3;10–4;2	1
5	15	8/7	5.1	0.19	5;0–5;4	0
7	15	7/8	7.1	0.17	7;0–7;4	0
11	15	7/8	11.2	0.13	11;0–11;5	0

England. All spoke English as a first language and none had behavioural or educational problems. The group was ethnically mixed and from a predominantly working-class background.

Design

Design was mixed with a between-participants factor of Age ($3\frac{1}{2}$, 4, 5, 7, 11). The within-participants factors were Task condition (Experimental, Control) and for the experimental conditions two further factors of Response set (Same, Different) and Semantics (Related, Unrelated). The dependent variables were reaction time (seconds) and response accuracy.

Materials

The stimuli were identical to those used in Experiment 1 except that practice and test stimuli were presented on a laptop PC which had been programmed using Microsoft Visual Basic. An acoustic analysis programme (Sound Forge) was used to analyze the Wav sound files.

Procedure

Participants were individually tested in a quiet room. Three-and-a-half and 4-year-olds were tested across three sessions between three and five days apart. Five-year-olds were tested across two sessions on the same day, and 7- and 11-year-olds were tested in a single session. The order of condition presentation was counterbalanced as in Experiment 1.

The procedures used in Experiment 1 were modified in a number of ways. First, the practice and test stimuli were presented on a laptop PC that opened Wav sound files as each stimulus was presented. This enabled reaction times (from the onset of the stimulus to the start of the child's response) to be measured, as well as accuracy. Each picture was displayed for a maximum of 8 seconds. The experimenter could terminate the trial once the participant had responded, or could use this time to correct errors. The inter-trial interval was 2 seconds. Second, children received up to six practice trials. Only children who answered two consecutive trials correctly went on to perform the test trials. Any child who did not pass the practice trials on every condition was excluded from the experiment so that data were analysed only for children who passed all the experimental conditions (see Table 2). Third, during the test session feedback was given in the following manner. An incorrect response was corrected by the experimenter on both the pre-test and test trials. The child was told that he or she had given the wrong response and reminded of the appropriate rule. For example, if the child said 'car' to the car picture the experimenter said 'Remember in this game you have to say "book" to the car.' As each picture could be displayed for up to 8 seconds there was sufficient time to correct errors.

Results

Reaction time

Reaction times were used if they met three criteria: (i) the response was audible, (ii) the response was correct, and (iii) the reaction time was not more than two standard deviations longer than a child's mean reaction time for the condition. Each child's mean reaction time was then calculated for each condition. The number of trials used to calculate this mean was never less than 8 trials (i.e. 50%) for any child on any condition.

Reaction times in the experimental and control conditions are shown in Figure 4. Older children responded faster than the younger children in all conditions with an improvement from $3\frac{1}{2}$ to 11 years of between 1.33 and 1.82 seconds in the experimental conditions and 0.67 seconds in the control condition. Reaction times in the two conditions with the same response set (red–blue/blue–red and car–book/book–car) were longer than the control condition at all ages. For the 7- and 11-year-olds, reaction times in the two conditions with different response sets (black–yellow/white–green and sun–dog/moon–pig) were virtually identical to the control condition. For the $3\frac{1}{2}$- to 5-year-olds the reaction times in the

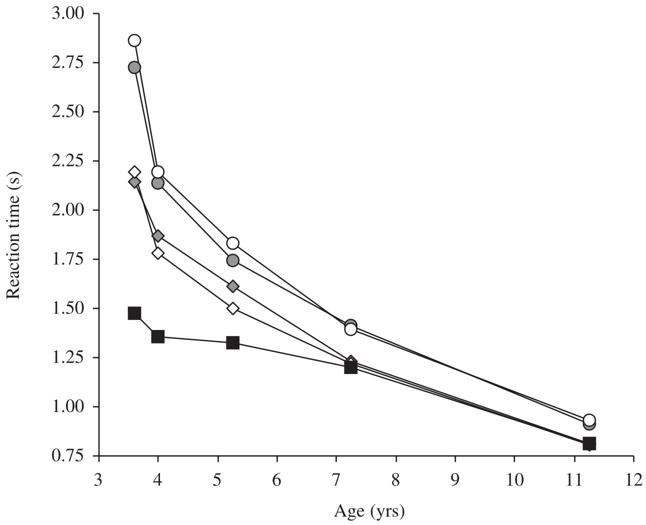

Figure 4 *Reaction time on the four experimental conditions and the control condition in Experiment 2.* Same-response-set/semantically-related *condition (red–blue/blue–red, ●);* same-response-set/semantically-unrelated *condition (car–book/book–car, ○);* different-response-set/semantically-related *condition (black–yellow/white–green, ◆);* different-response-set/semantically-unrelated *condition (sun–pig/moon–dog, ◇); control condition (abstract1–cup/abstract2–boat, ■).*

conditions with different response sets were intermediate between the conditions with the same response set and the control condition.

Reaction times were analysed in a mixed ANOVA with Age ($3\frac{1}{2}$, 4, 5, 7 and 11) as the between-participants factor and Response set (Same, Different) and Semantics (Related, Unrelated) as the within-participants factors. There was a significant main effect for Age ($F(4, 70) = 38.83$, $p < .001$) and Response set ($F(1, 70) = 38.24$, $p < .001$). There was no significant main effect or interactions involving semantics. The data were collapsed across this variable in the following analysis.

To test the inhibitory demands of the two conditions with different response sets and the two conditions with the same response set, we compared reaction times in these conditions to the reaction times of the control condition. This revealed that the conditions with the same response set (red–blue/blue–red and car–book/book–car) were harder than the control at all ages ($3\frac{1}{2}$ years $t(14) = 5.65$, $p < .001$; 4 years $t(14) = 4.10$, $p = .001$; 5 years $t(14) = 4.00$, $p = .001$; 7 years $t(14) = 3.90$, $p = .002$; 11 years $t(14) = 5.91$, $p < .001$). In contrast the conditions with different response sets (black–yellow/white–green and sun–pig/moon–dog) were only harder than the control at $3\frac{1}{2}$ years ($t(14) = 6.55$, $p < .001$), 4 years ($t(14) = 5.81$, $p < .001$) and 5 years ($t(14) = 2.55$, $p = .023$). At 7 years ($t(15) = 1.85$, $p = .084$) and 11 years ($t(14) = 1.04$, $p = .318$) performance in the conditions with different response sets was indistinguishable from the control condition. Thus comparison with the control condition suggested that the conditions with the same response set required inhibition at all ages, whereas the conditions with different response sets required inhibition at $3\frac{1}{2}$ to 5 years but not 7 to 11 years. There were no significant effects involving gender or order of condition.

Accuracy

With feedback reminding children of the task rules, all age groups achieved high levels of accuracy in all conditions. For the $3\frac{1}{2}$-year-olds accuracy in the experimental conditions ranged from 85% to 88% and was 94% in the control condition. There was an age-trend for improvement in all the conditions with 11-year-olds achieving between 96% and 99% in the experimental conditions and 98% in the control condition. Comparing accuracy in the experimental conditions to the control condition did not produce significant differences with any of the age groups.

Discussion

In the second study we sought to determine whether semantics or response set made the day–night task difficult for children aged between $3\frac{1}{2}$ and 11 years. The data were clear. A strong response set effect was obtained but no semantic effect. Comparison with the control condition also proved informative. Conditions with the same response set (red–blue/blue–red and car–book/book–car) produced longer reaction times than the control condition at all ages. This suggests that inhibition was required to stop picture naming in these conditions at all ages. In contrast, longer reaction times were observed in the conditions with different response sets (black–yellow/white–green and sun–dog/moon–pig) compared to the control condition in only $3\frac{1}{2}$-, 4- and 5-year-olds. At 7 and 11 years of age the reactions times in these conditions were indistinguishable from the control condition. This suggests that for younger children the conditions with different response sets do require inhibition, but in older children they do not. In children of 7 years and older the response set effect is so strong that they are – in effect – able to ignore the pictures in the conditions with different response sets.

Comparing the 4-year-olds' accuracy in Experiment 1 to the 4-year-olds' reaction times in Experiment 2, revealed differences in performance in the conditions with different response sets. In Experiment 1, children found the *different-response-set/semantically-related* condition (black–yellow/white–green) significantly more difficult (i.e. they were less accurate) than the *different-response-set/semantically-unrelated* condition (sun–dog/moon–pig). In Experiment 2 both conditions were moderately difficult (i.e. reaction times were equivalent). One simple explanation is that feedback reduced the working memory demands of the *different-response-set/semantically-related* condition in Experiment 2 thus allowing equivalent performance to the *different-response-set/semantically-unrelated* condition.

When comparing the 4-year-olds' reaction time data from Experiment 2 to the accuracy data of Diamond *et al.* (2002) the findings are broadly compatible. Both sets of researchers found that a *same-response-set/semantically-related* condition (the Diamond *et al.* day–night condition; our red–blue condition) was harder than a *different-response-set/semantically-unrelated* condition (the dog–pig condition in both studies). It is notable that accuracy performance in the Diamond *et al.* (2002) *different-response-set/semantically-unrelated* condition was extremely good (and probably no different from a control condition). In contrast, reaction time performance in ours was moderately poor: intermediate between the conditions with the same response set and the control condition. Our data suggest that for 4-year-olds the *different-response-set/semantically-unrelated* condition does require inhibition while the Diamond *et al.* (2002) data suggest that it does not. A possible explanation for this discrepancy is that our reaction time measure was sensitive enough to detect an inhibitory cost not detected by an accuracy measure in the Diamond *et al.* (2002) study.

General discussion

What makes responses prepotent in the day–night task? Our data provide strong support for a response set effect at all ages, and some evidence for a semantic effect in 4-year-olds. In children of all ages between $3\frac{1}{2}$ and 11 years performance was worse in the conditions with the same response set (red–blue/blue–red and car–book/book–car) than the control. In the conditions with the same response set the names of the two pictures (e.g. 'car' and 'book') and the two conflicting responses (e.g. 'book' and 'car') were the same. Children generally performed better in the conditions with different response sets (black–yellow/white–green and sun–dog/moon–pig) in which the names of the two pictures (e.g. 'sun' and 'moon') and the two conflicting responses (e.g. 'dog' and 'pig') were different. This differential performance suggests that inhibitory demands were higher in the conditions with the same response set than the conditions with different response sets. The only exception to this pattern was 4-year-olds' low accuracy performance in the *different-response-set/semantically-related* condition (black–yellow/white–green) in Experiment 1. This finding may reflect the presence of a semantic effect in younger children that may affect the inhibitory or working memory demands of this condition.

Our data also indicate that the response set effect increases with age and so effectively stops some responses being prepotent in older children. The data from Experiment 2 suggest that between the ages of $3\frac{1}{2}$ to 5 years the inhibitory demands were reduced in the conditions with different response sets (black–yellow/white–green and sun–pig/moon–dog), while for 7- to 11-year-olds they were completely eliminated. These data suggest that performance in certain conditions of the day–night task may improve partly because the range of responses that are prepotent reduces during development. For example, it may be that the responses 'yellow' and 'green' had some prepotency in the *different-response-set/semantically-related* condition (black–yellow/white–green) for $3\frac{1}{2}$- and 4-year-olds but none at all for 7- and 11-year-olds. Perhaps then, performance on some inhibitory tasks improves during development because these tasks stop requiring inhibition. More generally it may be that some responses which are prepotent in young children are not prepotent in adults, independent of any improvements in inhibitory control.

The day–night methodology

A possible limitation of our design was that we used different stimuli in the four experimental conditions – some stimuli were colours and some were objects. We explained in the discussion of Experiment 1 why it was impossible to use identical stimuli in all four conditions. We also noted that day–night task performance is unaffected by changes in the stimuli (Simpson & Riggs, in press). Furthermore, in the present study we never found that performance mirrored the colour–object distinction in the stimuli. Performance was never the same on both the colour conditions (red/blue and black/white) yet different from both the object conditions (car/book and sun/moon). In fact in Experiment 2, one colour condition was always paired with one object condition: performance on the red/blue and car/book conditions was nearly identical, yet different from the black/white and sun/moon conditions.

In this paper we varied two aspects of the day–night methodology: main performance measure and feedback. First, with regard to main performance measure it seems likely that both reaction time and accuracy reflect the inhibitory demands of the day–night task. The reaction time measure is clearly preferable in older children because of ceiling effects in accuracy (Simpson & Riggs, in press). In young children the position is less clear and reaction time may not always be the most sensitive. Second, there is also the question of how feedback affects the inhibitory and working memory demands of the day–night task. It is likely that reminding children of the task rules helps them to remember those rules. There may also be other, more subtle, effects and these may be expressed differently in the reaction time and accuracy measures. For example, giving feedback may encourage children to be more careful, thus improving accuracy, but slowing their responses and increasing reaction times. More research is required into these methodological aspects of the day–night task.

The inhibitory mechanism in the day–night task

We begin by considering the performance of the 7- and 11-year-olds in which the response set effect was most dramatic. Our results suggest that for these older children only the two response options defined in the task rules become prepotent. So, for example, in the *same-response-set/semantically-unrelated* condition (car–book/book–car) the response options are 'book' and 'car': thus 'book' and 'car' become prepotent. In this condition these prepotent responses are triggered by the stimuli because the stimuli and responses are from the same response set. When the book picture is displayed the child has to respond 'car' but the prepotent response 'book' is triggered by the picture of a book and so must be inhibited. We suggest that the conditions with the same response set (red–blue/blue–red and car–book/book–car) require inhibition because the child is planning to make one of two responses and the wrong one is triggered by the stimulus. This contrasts with the situation in the conditions with different response sets. For example in the *different-response-set/semantically-unrelated* condition (sun–dog/moon–pig) the response options are 'pig' and 'dog': thus 'pig' and 'dog' become prepotent. However, when the sun picture is displayed and the child has to respond 'dog', neither of the two prepotent responses is triggered by the picture and so inhibition is not required. We suggest that for 7- and 11-year-olds the conditions with different response sets do not require inhibition because the child is planning to make two responses and neither of them is triggered by the stimulus.

How then do we interpret the results for the younger children in the conditions with different response sets? The reaction time data from Experiment 2 suggest that prepotency is *reduced* rather than eliminated in these conditions. For these younger children the responses they are planning to make become prepotent, but also, the responses they are *not* planning to make retain some prepotency. As previously suggested, the accuracy data from Experiment 1 are compatible with the response set effect being reduced either through an interaction with semantics or the effect of high working memory demands. Exactly what happens to the response set effect in younger children is an important question for future research.

The inhibitory mechanism in other tasks

Our interpretation suggests that prepotency can be restricted (or partially restricted) to the responses the child is planning to make in the day–night task. This method of conferring prepotency may be applicable to any inhibitory task in which the child is told how to respond. We have previously referred to these tasks as 'response-given' tasks (Simpson & Riggs, 2004). Response-given tasks include all those tasks in which the child must stop a prepotent response while choosing between a pair of conflicting responses (e.g. the grass–snow (Carlson & Moses, 2001), hand game (Hughes, 1998), card sort (Fyre *et al.*, 1995), tapping (Diamond & Taylor, 1996) and Simon (Diamond, 2003) tasks). For example, in the hand game the child is told to point a finger when the experimenter makes a fist and make a fist when the experimenter points a finger: making a fist and pointing a finger are the response options and so, we would suggest, become prepotent. When the experimenter makes a fist the child must point a finger but the prepotent response make a fist is triggered by the experimenter's gesture and so must be inhibited. As was noted in the introduction, in all these tasks the stimuli and responses are from the same response set and so, all things being equal, prepotency may be conferred in the same way.

Response-given tasks also include all those tasks in which the child must simply delay the prepotent response (e.g. the gift delay, pinball, snack delay, tower building (Kochanska, Murray, Jacques, Koenig & Vandegeest, 1996), box search (Livesey & Morgan, 1991) and Luria light (Flynn, O'Malley & Wood, 2004) tasks). For example, in the tower building task the child has to take turns with an experimenter in adding blocks to a tower. It is assumed that adding a block is the prepotent response that the child must inhibit when the experimenter takes his turn to add a block. Our interpretation, that planned responses become prepotent responses, can be applied to delay tasks. In the tower building task the child is planning to add blocks to the tower; planning this response renders it prepotent and therefore inhibition is required when the experimenter takes his turn at adding a block.

There are, however, many putative inhibitory tasks in which the child is not told how to respond and so must infer the conflicting responses. For example, in the windows task a child is shown two boxes with windows which reveal that one is empty while the other contains a treat (Russell *et al.*, 1991). The child is asked to point to a box for an opponent to look in. The child then 'wins' the contents of the other box (the treat). In order to pass the task the child must use a rule for producing the conflicting response such as 'point to the empty box'. However, because the child is not told this rule by the experimenter he or she must first infer it. We have previously referred to tasks in which the child is *not* told how to respond as 'open' tasks. Other open tasks include theory of mind (Carlson *et al.*, 1998), counterfactual conditional reasoning (Riggs, Peterson, Robinson & Mitchell, 1998), counterfactual deductive reasoning (Dias & Harris, 1988), Iowa gambling (Zelazo, 2003) and conversation (Russell, 1999) tasks.

In these open tasks it seems unlikely that prepotency is straightforwardly restricted to responses that the child is planning. The child is not told to make specific responses and thus has no reason to plan to make any. How then might prepotency be conferred in these open tasks? We suspect that it is likely to be conferred differently in different tasks. Nevertheless, we briefly set out how it may be conferred in one open task: the false belief task.

In a standard false belief story, Maxi and Mummy are in the kitchen where they place a bar of chocolate in the fridge. Maxi then leaves the house and while he is away Mummy makes a cake with some of the chocolate, and places the rest of it in the cupboard. The false belief question asks, 'Where does Maxi think the chocolate is?' To generate this conflicting response the child must use theory of mind knowledge to reason that Maxi believes that the chocolate is in the cupboard because he did not see it moved to the fridge. It is assumed by many in the literature that pointing to the chocolate in the cupboard is prepotent. However, a question that needs answering is *why* this response is prepotent. One possibility is that children assume by default that beliefs are true and this assumption may be sufficient to make pointing to the cupboard prepotent. Poor inhibitory control would then lead to the execution of this prepotent (and incorrect) response. This example illustrates how, without planned responses, unsophisticated theory of mind knowledge (i.e. assuming that beliefs are true) may give rise to a prepotent response which combines with weak inhibitory control to produce poor performance in the false belief task. Development either in theory of mind knowledge (eliminating the prepotent response) or inhibitory control (making the prepotent response easier to inhibit) would then eliminate inhibitory errors. The question of how (or if) prepotency is conferred in other open tasks is clearly a question for future research.

Conclusion

In this study we obtained evidence to suggest that response set is a major factor in creating prepotency in the day–night task for children aged between $3\frac{1}{2}$ and 11 years. Our findings, and our interpretation of them, may be applicable to other response-given tasks and perhaps provide some basis for answering how inhibitory demands are created in open tasks. Understanding what makes responses prepotent is essential if we are to understand the development of inhibitory control. This information will give us a rational basis for determining when inhibitory control is needed both in tasks used in developmental research and more importantly in children's everyday lives. We may also find that as children get older there is a reduction in the number of things that they find prepotent.

Acknowledgements

This research was carried out in partial fulfilment for the degree of PhD by the first author at the School of Psychology, University of Birmingham, UK. We would like to thank the staff, parents and pupils of Elizabeth Lansbury Nursery School and Susan Lawrence Infants School, Poplar, London.

References

Carlson, S.M., & Moses, L.J. (2001). Individual differences in inhibitory control and children's theory of mind. *Child Development*, **72**, 1032–1053.

Carlson, S.M., Moses, L.J., & Hix, H.R. (1998). The role of inhibitory processes in young children's difficulties with deception and false belief. *Child Development*, **69**, 672–691.

Dalrymple-Alford, E.C. (1972). Associative facilitation and interference in the Stroop colour–word task. *Perception and Psychophysics*, **11**, 274–276.

Diamond, A. (2003). Longitudinal and cross-sectional data on interrelations among executive function tests in children 4–12 years old. Paper presented at the meeting of the Society for Research in Child Development, Florida, 2003.

Diamond, A., Kirkham, N., & Amso, D. (2002). Conditions under which young children can hold two rules in mind and inhibit a prepotent response. *Developmental Psychology*, **38**, 352–362.

Diamond, A., & Taylor, C. (1996). Development of an aspect of executive control: development of the ability to remember what I said and to 'do as I say, not as I do'. *Developmental Psychobiology*, **29**, 315–334.

Dias, M.G., & Harris, P.L. (1988). The effect of make-believe play on deductive reasoning. *British Journal of Developmental Psychology*, **6**, 207–221.

Flynn, E., O'Malley, C., & Wood, D. (2004). A longitudinal, microgenetic study of the emergence of false belief understanding and inhibition skills. *Developmental Science*, **7**, 103–115.

Fyre, D., Zelazo, P.D., & Palfai, T. (1995). Theory of mind and rule-based reasoning. *Cognitive Development*, **10**, 483–527.

Gerstadt, C.L., Hong, Y.J., & Diamond, A. (1994). The relationship between cognition and action: performance of 3.5 to 7 year olds on Stroop-like day–night test. *Cognition*, **53**, 129–153.

Hughes, C. (1998). Finding your marbles: does preschoolers' strategic behaviour predict later understanding of mind? *Developmental Psychology*, **34**, 1326–1339.

Jones, L.B., Rothbart, M.K., & Posner, M.I. (2003). Development of executive attention in preschool children. *Developmental Science*, **6**, 498–504.

Klein, G.S. (1964). Semantic power measured through the interference of words with colour-naming. *American Journal of Psychology*, **13**, 125–135.

Kloo, D., & Perner, J. (2003). Training transfer between card sorting and false belief understanding: helping children apply conflicting descriptions. *Child Development*, **74**, 1823–1829.

Kochanska, G., Murray, K., Jacques, T.Y., Koenig, L., & Vandegeest, K.A. (1996). Inhibitory control in young children and its role in emerging internalization. *Child Development*, **67**, 490–507.

Kornulum, S., Hasbroucq, T., & Osman, A. (1990). Dimensional overlap: cognitive basis for stimulus–response compatibility – a model and taxonomy. *Psychological Bulletin*, **97**, 253–270.

La Heij, W. (1988). Components of Stroop-like interference in picture naming. *Memory and Cognition*, **16**, 400–410.

Livesey, D.J., & Morgan, G.A. (1991). The development of response inhibition in 4- and 5-year-old children. *Australian Journal of Psychology*, **43**, 133–137.

MacLeod, C.M. (1991). Half a century of research on the Stroop effect: an integrative review. *Psychological Bulletin*, **109**, 163–203.

Munakata, Y., Morton, J.B., & Yerys, B.E. (2003). Children's perseveration: attentional inertia and alternative accounts. *Developmental Science*, **6**, 449–476.

Nigg, J.T. (2000). On inhibition/disinhibition in developmental psychopathology: views from cognitive and personality psychology and a working inhibition taxonomy. *Psychological Bulletin*, **126**, 220–246.

Proctor, R.W. (1978). Sources of colour–word interference in the Stroop colour-naming task. *Perception and Psychophysics*, **23**, 413–419.

Reed, M.A., Pien, D.L., & Rothbart, M.K. (1984). Inhibitory self-control in pre-school children. *Merrill-Palmer Quarterly*, **30**, 131–147.

Riggs, K.J., Peterson, D.M., Robinson, E.J., & Mitchell, P. (1998). Are errors in false tasks symptomatic of a broader difficulty with counterfactuality? *Cognitive Development*, **13**, 73–90.

Russell, J. (1999). Cognitive development as an executive process – in part: a homeopathic dose of Piaget. *Developmental Science*, **2**, 247–295.

Russell, J., Mauthner, N., Sharpe, S., & Tidswell, T. (1991). The 'windows task' as a measure of strategic deception in preschoolers and autistic subjects. *British Journal of Developmental Psychology*, **9**, 331–349.

Simpson, A., & Riggs, K.J. (2004). What makes the windows task difficult for young children: rule inference or rule use? *Journal of Experimental Child Psychology*, **87**, 155–170.

Simpson, A., & Riggs, K.J. (in press). Investigating the inhibitory and working memory demands of the day–night task in children aged 3-and-a-half to eleven years. *British Journal of Developmental Psychology*.

Stirling, N. (1979). Stroop interference: an input and an output phenomenon. *Quarterly Journal of Experimental Psychology*, **31**, 121–132.

Zelazo, P. (2003). Development of 'hot' executive function: the children's gambling task. Paper presented at the meeting of the Society for Research in Child Development, Florida, 2003.

Received: 22 March 2004
Accepted: 12 July 2004