Developmental Science

Developmental Science 12:6 (2009), pp 1070–1082 DOI: 10.1111/j.1467-7687.2009.00856.x

PAPER

How do symbols affect 3- to 4-year-olds′ executive function? Evidence from a reverse-contingency task

Ian A. Apperly[1] and Daniel J. Carroll[2]

1. School of Psychology, University of Birmingham, UK
2. Department of Psychology, University of Sheffield, UK

Abstract

In two experiments, 330 3- to 4-year-olds competed for stickers in a game in which the optimal response strategy was to point to an empty box that their opponent would receive in order to obtain a baited box for themselves. When the baited box contained stickers, children showed a strong tendency to point at the baited box and therefore lose the stickers to their opponent. In Experiment 1 children performed better when the number of stickers to be won was represented with one of five different types of symbol: numerals, number words, dots, a photograph or sweets. In Experiment 2 children transferred their improved performance in symbolic conditions to non-symbolic conditions. These findings suggest that symbols enable children to formulate an efficient response strategy, and that this effect may be qualitatively different in children from the effect of symbols in non-human primates.

Introduction

Between 3 and 4 years of age there are substantial changes in children's executive control (e.g. Zelazo, Muller, Frye & Marcovitch, 2003) and in their understanding of symbols (e.g. Bialystok, 2000; Liben, 1999). Importantly, these domains of development interact, consistent with the long-standing contention in cognitive science that symbols are essential to flexible thought and action (e.g. Werner & Kaplan, 1963). On the one hand, there are empirical relationships between executive tasks and tasks designed to test 3–4-year-olds' understanding of symbols. For example, Bialystok (1999) found that performance on the moving word test of symbolic understanding (e.g. Apperly, Williams & Williams, 2003; Bialystok, 1991) was correlated with performance on a test of executive function (the dimensional change card sort task: Frye, Zelazo & Palfai, 1995). On the other hand, symbols can affect children's performance on tests of children's executive control. For example, Carlson, Davis and Leach (2005) tested children's executive control on a task that required them to select a smaller quantity of a reward in order to receive a larger quantity, and found that children performed better when the rewards were represented symbolically than when they were not (see also Russell, 1996, for comparable findings on a related task). However, although there is recently emerging evidence to suggest that symbols can help children to perform better on some tests of executive control, it is not known which aspects of symbols help children, nor which aspects of executive function are facilitated. Regarding the first point, there are several potentially important factors that are introduced when objects are replaced with symbols, including changed levels of desirability, varying resemblance to the object and varying degrees of transparency between symbol and object. However, it is not clear whether the effect of symbols results from one of these individual aspects, from several of them, or from all of these aspects in combination. Regarding the second point, it is unclear whether symbols help children to formulate a response strategy, or whether they reduce the executive costs of executing a response on a task. The present study will directly address these two points by investigating the effect of different kinds of symbol on children's performance on a reverse-contingency task.

The reverse-contingency task that has been most widely studied in 3- to 4-year-old children is Russell *et al.*'s 'Windows task' (e.g. Carroll, Apperly & Riggs, 2007a; Hala & Russell, 2001; Russell *et al.* 1991; Russell, Jarrold & Potel, 1994; Samuels, Brooks & Frye, 1996; Simpson, Riggs & Simon, 2004). In this task there is a counterintuitive reverse-contingency between the child's action and the result of the action. The child is presented with two boxes and a simple rule: they must point to a box for the opponent to open, and will themselves open the other box and receive its contents. The child can see

Address for correspondence: Ian Apperly, School of Psychology, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK; e-mail: i.a.apperly@bham.ac.uk

© 2009 The Authors. Journal compilation © 2009 Blackwell Publishing Ltd, 9600 Garsington Road, Oxford OX4 2DQ, UK and 350 Main Street, Malden, MA 02148, USA.

an attractive reward (e.g. a sticker) in one box, and can see that the other box is empty. Thus, in order to obtain the reward, the child must point to the empty box for an opponent to open. Whereas 4-year-olds typically perform well, 3-year-old children commonly fail this task, pointing to the box containing the reward that they wish to obtain, and thereby losing it. This failure is surprisingly resistant to feedback, with many children persisting with their incorrect strategy over 15 or 20 repeated trials. This perseverative pattern is consistent with the view that children's difficulty is executive in nature (see e.g. Russell, 1996), and mirrors findings with reverse-contingency tasks in non-human species (e.g. Boysen & Berntson,1995; Boysen, Berntson, Hannan & Cacioppo, 1996; Kralik, Hauser & Zimlicki, 2002; Vlamings, Uher & Call, 2006).

There is evidence that children's performance on reverse-contingency tasks can be improved in a variety of ways. Children perform better if they indicate their chosen box by rotating an arrow or by placing a marker, rather than by pointing with their hand (e.g. Carlson, Moses & Hix, 1998; Carroll, Apperly & Riggs, under submission; Hala & Russell, 2001). Children perform better if they play with an ally than if they play alone (Hala & Russell, 2001), if they are prompted with instructions that explicitly relate their response to the outcome of winning or losing the reward (e.g. Carroll *et al.*, 2007a; Samuels *et al.*, 1996), or if the box without the reward contains a non-desirable scrap of paper rather than being empty (Carroll, Apperly & Riggs, 2007b). Of particular relevance to the current study, there is also evidence that children perform better when symbolic representations of stimuli, rather than real stimuli, are used. In a study by Carlson, Davis and Leach (2005), two boxes were baited with a smaller and a larger number of sweets, or with symbols representing the sweets (e.g. 5 versus 2 small stones; a card with many dots versus a card with fewer dots; an elephant picture versus a mouse picture to represent 5 versus 2 sweets). The children's task was to point to the box with the smaller number in order to obtain the box with the larger number and they were rewarded with the corresponding number of sweets in all conditions. There was a trend for better performance in all conditions with symbols, although only one symbol type (Mouse = small number/ Elephant = large number) had a statistically significant effect.

Convergent evidence that symbols might help performance on reverse-contingency tasks comes from research with non-human primates. A variety of non-human species have difficulty with such tasks (see e.g. Boysen & Berntson,1995; Boysen *et al.*, 1996; Kralik, *et al.*, 2002; Vlamings *et al.*, 2006). There is evidence that some individuals can learn to respond at above-chance levels after many trials (many tens or hundreds of trials), but other individuals do not seem to learn. Boysen and Berntson (1995; Boysen, Berntson, Hannan & Cacioppo, 1996) report results from six chimpanzees who failed to improve performance over several hundred trials of a reverse-contingency task. The animals also performed poorly in a condition where the sweets were represented with an analogue symbol (stones). Interestingly, however, the chimpanzees had been trained in the representation of small numerical quantities with Arabic numerals, and when sweets were represented with numerals all six animals performed significantly above chance.

These data from children and non-human primates indicate that symbols may improve performance on reverse-contingency tasks, but the precise nature of this effect is not clear. Carlson *et al.* (2005) invoke the idea of a 'gradient of symbolic representation' to explain why symbolizing sweets with larger versus smaller sets of stones was less effective than symbolizing sweets with pictures of a larger versus a smaller animal. However, although a 'gradient of symbolic representation' has some intuitive appeal, Carlson *et al.* (2005) do not offer any objective criteria by which the symbolic 'gradient' or 'distance' between a symbol and a referent can be determined a priori. Moreover, even in principle, it is unclear whether such criteria could be determined, or whether a single dimension is really the most informative way to characterize the relationship between symbol and referent. In the current work we test more specific hypotheses about which features of a symbol can help 3- to 4-year-old children.

What features of symbols might help children respond optimally on reverse-contingency tasks?

Children's difficulty on reverse-contingency tasks seems to be with exploiting their knowledge of the task rule, rather than with comprehending the rule (indeed, comprehension of the rule is usually a criterion that children must meet before entering the test-phase of a study). When faced with a box containing the desirable stimulus, therefore, unsuccessful children either cannot work out the correct response strategy of pointing to the empty box, or they cannot act on this strategy. Given these possible problems, and considerations from the broader literature on children's symbolic development (see e.g. Carlson & Zelazo, 2008; DeLoache, 1991; Liben, 1999), we suggest three factors that might enable symbols to improve children's performance.

One possibility is that 3- to 4-year-olds perform poorly on reverse-contingency tasks because the objects to be won (stickers or sweets) are highly salient stimuli that would attract children's attention even if they were not part of the task. Symbols might therefore assist children to the extent that they are less intrinsically desirable and attention-grabbing than the objects that they symbolize. This account predicts that children's performance might be improved if desirable stickers are symbolized with stimuli that are of low intrinsic desirability, such as dots,

© 2009 The Authors. Journal compilation © 2009 Blackwell Publishing Ltd.

a photograph, a numeral or a number word[1]. In contrast, this account predicts that children's performance should *not* be enhanced if the goal of stickers is represented by intrinsically attractive, attention-grabbing objects, such as sweets.

A second possibility is that what constitutes a distracting stimulus is determined by the specific goal of the task (in this instance, obtaining stickers), rather than by any intrinsic desirability of the stimulus itself. This account predicts that children's performance might be improved if the symbols used bear little resemblance to the reward items that are the child's goal on the task. Thus, if the child is trying to win stickers, their performance might be improved by the use of symbols dissimilar in appearance to stickers (e.g. dots, numerals, number words, or sweets), even if these symbols are themselves intrinsically desirable (e.g. sweets). In contrast, this account predicts that children's performance will not be enhanced if stickers are symbolized with photographs of stickers, because the photographs' close visual resemblance to the stickers means that they are likely to be potent reminders of the stickers that the child wishes to obtain.

A third possibility that we considered was that children's performance might be determined by their ability to read the symbol. Three- to four-year-old children can typically read analogue representations (e.g. dots, photographs and sweets) and written representations of small numbers when they are numerals. However, children of this age are much less likely to be able to read number words (e.g. Bialystok, 2001). Nonetheless, children are familiar with having words read to them, meaning that the symbol's meaning can be introduced to them by the experimenter (e.g. Apperly *et al.*, 2004; Bialystok, 1991, 2001). Competing predictions can be made about the effects of being able to read the symbols used on a reverse-contingency task, and these are of interest because children's reading fluency improves with age, and at different rates with different symbols. One possibility is that symbols that children can read would be more effective at enhancing performance, because children would find it easier to remember the quantity signified by the symbols. Alternatively, symbols that children can read may be less effective at enhancing performance, because if children can directly access their meaning then they may be more potent reminders of the distracting properties of the desirable goal (Carlson *et al.*, 2005). A third prediction would be that if symbols enhance children's performance simply because they are not the reward that children wish to obtain then children's ability to read the symbols' meaning for themselves might make no difference to any beneficial effect on a reverse-contingency task .

All three of the above factors were investigated in Experiment 1 of the current paper. In all of our experimental conditions children competed to win stickers. In the Stickers condition – a standard version of the Windows task – one box was empty and the other was baited with N stickers. In each of the five Symbol conditions, one box was empty and the other contained one of five types of symbol: a photograph of N stickers, a card with N printed dots, N sweets, the numeral 'N' and the number word 'N'. In these conditions, if the child pointed to the empty box then they obtained the box with the symbol and were rewarded with N stickers. These specific symbol types were chosen to allow us to assess the contribution of each of the factors identified above to children's performance on reverse-contingency tasks.

Experiment 1

Method

Participants

We tested 180 children (92 boys) with a mean age of 3;11 (range 3;2–4;6). Children were from nurseries serving an upper working class or lower middle class population in Birmingham, UK. Approximately 15% of children were of Asian or African descent. The remainder were White. All spoke English as their first language. Children were randomly assigned to each condition, yielding 6 groups of 30. A univariate ANOVA showed no difference in the mean ages of children in each group, $F(5,174) = 1.59$, $p = .16$, with the mean age of children in the Stickers condition (3;11) falling in the middle of the range of the means for the Symbol conditions (3;9–4;0). The distribution of ages and genders for each condition is detailed in Table 1.

Procedure

The 'Windows task' has been used in a number of previous studies, which show that children find the task challenging regardless of whether their opponent is a real person (Hala & Russell, 2001), or a puppet (e.g. Simpson *et al.*, 2004), or indeed if they have no opponent at all (Carroll, Apperly & Riggs, 2007a, b; Russell, Jarrold & Potel, 1994). Similarly, previous work has shown that children have problems whether the treat to be won is a sweet (e.g. Hala & Russell, 2001) or a sticker (e.g. Carroll *et al.*, 2007a, 2007b). Some reverse-contingency tasks require participants to choose between a smaller and a larger quantity (e.g. Boyson *et al.*, 1996; Carlson *et al.*, 2005). However, the Windows task has always been presented with one baited box and one empty box, and that was the procedure followed here. The current procedure was

[1] It is conceivable that the effect of symbols might be greater or different in kind if the rewards were food (which has appetitive value) rather than stickers. However, the literature clearly shows that preschoolers have difficulty on reverse-contingency tasks played with stickers or with sweets. It may be fruitful for future work to test whether the effects of symbols differ for rewards with and without appetitive value, but in the absence of current evidence to the contrary we shall assume that findings from studies using stickers as rewards would generalize to studies using sweets or other foods.

© 2009 The Authors. Journal compilation © 2009 Blackwell Publishing Ltd.

Table 1 *Sex and age distributions across the conditions of Experiments 1 and 2*

	Male: Female	Mean age (months) Younger half of sample	Older half of sample	Overall
Experiment 1 conditions				
Stickers	16:14	43.9	49.9	46.9
Numerals	12:18	44.5	50.2	47.3
Number words	17:13	45.1	50.5	47.8
Dots	16:14	43.6	49.2	46.4
Sweets	14:16	42.7	48.1	45.3
Photograph	17:13	43.0	49.7	46.3
Experiment 2 conditions				
Stickers	12:18	42.1	47.8	45.0
Number words	12:18	42.3	47.3	44.8
Number words–Stickers	14:16	42.5	47.5	45.0
Photograph	12:18	41.9	46.8	44.4
Photograph–Stickers	22:8	42.0	47.4	44.7

modelled on the version of the task used by Simpson *et al.* (2004). Children were told that they would play a game with a puppet, Kevin the crocodile, to see which of them could win more stickers. All children began with a training phase to ensure that they understood the rules of the game and the method of responding.

Training phase – Stickers condition

Children were shown two covered boxes (approximate dimensions 10 × 6 × 6 cm) and told that one box had no stickers inside whereas the other had a small number, N (but not which box was which). They were told that they could choose which box Kevin looked in, and that Kevin kept any stickers he found in that box, while the child would keep any in the other box. The child was then asked to point to the box they wanted Kevin to look in. The experimenter opened the box, said how many stickers were inside (either no stickers or N stickers), and gave any stickers to Kevin. The child looked inside the other box, the experimenter said how many stickers there were (either no stickers or N stickers), and the child received any stickers. After five such training trials we checked that the child had grasped the rule of the game. After the first box was opened the child was asked 'So who gets the stickers this time? You or Kevin?' This check trial was repeated until the child gave three consecutive correct responses. All children reached this criterion within five further training trials.

Testing phase – Stickers condition

Boxes with windows cut into them were introduced and the experimenter pointed out that it was now possible to see which box had stickers inside. This was the first point at which children could respond with a systematically correct strategy of pointing to the empty box for the opponent to open. As in the training phase, one box had N stickers and the other had none. Across trials, N was 1, 2 or 3, determined pseudo-randomly such that each N appeared the same number of times across trials. Each trial began with the experimenter telling the child how many stickers would be used on that trial. The stickers were placed in one of the two boxes and the boxes were placed on the table with the windows facing the child, enabling the child to see which box contained the stickers. The child was prompted with the words 'Point to a box for Kevin to look in'. After the child responded the experimenter said '[Kevin/You] get(s) [N] stickers this time', as appropriate. This procedure was repeated for nine trials, this being the number of trials shown by Carroll, Apperly and Riggs (2007b) to be sufficient to observe improved performance both in a standard version of a task and when the wording of the test instructions was manipulated.

The *Symbol conditions* were closely matched to the Stickers condition, with the single difference that throughout training and testing phases, one of the boxes contained a symbolic representation of number on a small card (approximate dimensions 4 × 4 cm) instead of stickers. Whoever obtained the box containing the symbol card was then rewarded with the number of stickers corresponding to the number represented on the card. The stickers were drawn from a container kept by the experimenter. Children's understanding of the relationship between the symbol and the stickers was checked in the warm-up phase, in which they were asked to say who was to receive stickers as a result of finding the symbol. Three consecutive correct responses were required before children entered the test phase of the experiment.

The level of reward was the same as in the Stickers condition. At the start of each trial, before the symbol-card was placed face-up in a box, the experimenter showed the card to the child and told the child the number depicted on the card. Thus, even if the child could not read the symbol for themselves, they were made aware of its numerical value. In the test phase of the experiment the windows in the side of the boxes meant that the symbol was clearly visible. Five distinct symbol types were presented: Dots, Photographs, Sweets, Numerals and Number words. The 'Dots' symbols were cards with 1, 2 or 3 black dots (5 mm in diameter) on a white background. 'Photograph' symbols were a printed digital photograph of 1, 2 or 3 stickers on a black background. Each sticker in the photograph was of a similar size to the stickers that children won – approximately 15 mm diameter. 'Sweets' symbols were 1, 2 or 3 wrapped candies approximately 20 mm × 10 mm × 10 mm. 'Numeral' symbols were the numerals '1', '2' or '3' printed in black on a white background in 60-point font. 'Number word' symbols were the words 'one', 'two' or 'three' printed in black on a white background in 48-point font.

Symbol reading test

To test the ability of children in our sample to read word and numeral symbols, children were presented with the words 'one', 'two' and 'three' and the numerals 1, 2 and 3

© 2009 The Authors. Journal compilation © 2009 Blackwell Publishing Ltd.

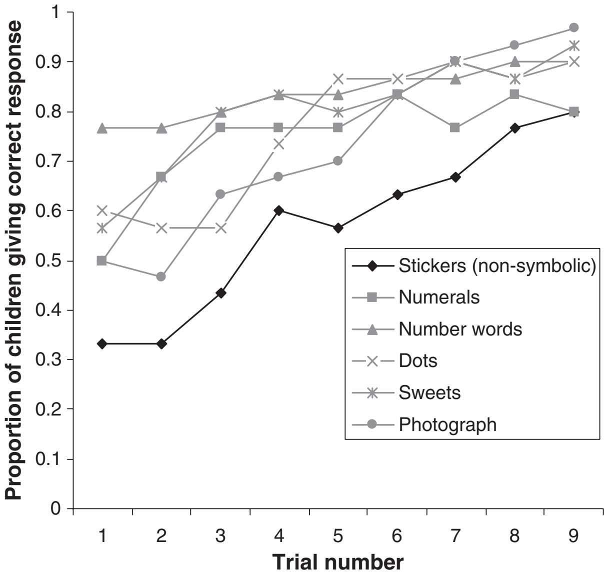

Figure 1 *Proportion of children giving a correct response on each trial in the six conditions of Experiment 1.*

on flash-cards in a varied order across children. This reading test was always presented after the main experiment. No feedback was given to children on the reading test.

Results and discussion

The proportion of children giving a correct response on each test trial of each condition is plotted in Figure 1. With only one exception (Numerals condition, trial 9), on every trial, each of the five Symbol conditions had a higher proportion of children giving a correct response than the non-symbolic Stickers condition. Although there was some variability from trial to trial, all conditions also showed a trend for improved performance over time.

Overall performance

An initial analysis compared the effects for males and females. Each child's number of correct responses (out of 9) was entered into an ANOVA with condition and sex as between-child factors. This analysis revealed no effect of sex, $F(1, 168) = .598$, $p = .44$, $n_p^2 = .004$, and no significant interaction between sex and condition, $F(5, 168) = .392$, $p = .85$, $n_p^2 = .012$. There was a significant main effect of condition, $F(5, 168) = 2.53$, $p = .031$, $n_p^2 = .07$. Data were combined for males and females in all subsequent analyses, and the main effect of condition was explored fully in a further analysis.

For an overall analysis of children's performance in each condition, taking into account the possibility of change over time, we calculated the average number of correct responses for each child for trials 1–3 (Time 1), trials 4–6 (Time 2), and trials 7–9 (Time 3) (see Table 2). This analysis method was employed successfully by Carroll, Apperly and Riggs (2007a, 2007b). We conducted an analysis of variance (ANOVA) with Time as a within-child factor, and Condition (Stickers, Dots, Pictures, Sweets, Numerals and Number words) and Age (younger versus older half of the children in each condition) as between-child factors. There was a significant main effect of condition, $F(5, 168) = 2.79$, $p = .019$, $n_p^2 = .08$. *T*-tests showed that performance in each Symbol condition was higher than performance in the Stickers condition (all *p*s < .038), but no symbol condition differed from any other (all *p*s > .19). Because our hypotheses concerned possible differences between Symbol conditions and the non-symbolic Stickers condition, we corrected for multiple comparisons using Dunnett's test for multiple comparisons against a single baseline. The one-tailed form of this test was used because the prediction was for higher scores in the Symbol conditions than in the Stickers condition. With this correction, the difference between Stickers and Number words, Dots and Sweets remained significant ($p = .002$, $p = .03$ and $p = .008$, respectively), whereas the differences for Stickers versus Numerals and for Stickers versus Photographs were marginally significant ($p = .052$ and $p = .071$, respectively).

The same ANOVA also revealed a significant main effect of Time, $F(2, 336) = 68.1$, $p < .001$, $n_p^2 = .29$,

Table 2 *Number of correct responses on trial 1, frequency with which children gave 0, 1, 2 or 3 correct responses over the first three trials, and comparison of this distribution against chance. Mean number of correct responses for the first, second and third set of three trials in each condition in Experiment 1*

		Time 1 (Trials 1–3)						Time 2 (Trials 4–6)	Time 3 (Trials 7–9)
Number of correct responses	Correct responses on Trial 1	0	1	2	3	Comparison between observed distribution and chance²	Mean/3	Mean/3	Mean/3
Expected by chance for $n = 30$	15	3.75	11.25	11.25	3.75		1.5	1.5	1.5
Stickers	10	11	12	0	7	χ^2 (1,30) = 19.6 $p < .001$	1.1	1.8	2.2
Numerals	15	7	3	5	15	χ^2 (1,30) = 37.4 $p < .001$	1.9	2.4	2.4
Number words	23	5	1	3	21	χ^2 (1,30) = 60.8 $p < .001$	2.3	2.5	2.7
Dots	18	9	5	1	15	χ^2 (1,30) = 48.4 $p < .001$	1.7	2.5	2.7
Sweets	17	5	3	8	14	χ^2 (1,30) = 23.5 $p < .001$	2.0	2.5	2.7
Photograph	15	9	6	3	12	χ^2 (1,30) = 32.4 $p < .001$	1.6	2.2	2.8

© 2009 The Authors. Journal compilation © 2009 Blackwell Publishing Ltd.

and of Age, $F(1, 168) = 16.3$, $p < .001$, $n_p^2 = .09$, and significant interactions between Condition and Time, $F(10, 336) = 2.48$, $p = .007$, $n_p^2 = .07$, and between Time and Age, $F(2, 336) = 3.94$, $p = .02$, $n_p^2 = .023$. All other effects were non-significant (all *p*s > .28).

To interpret the interaction between Condition and Time in the main analysis we conducted univariate ANOVAs to examine the effects of condition at each time point. At Time 1 there was a significant main effect of condition, $F(5, 174) = 3.53$, $p = .005$, $n_p^2 = .10$. *T*-tests showed that performance in each Symbol condition was higher than performance in the Stickers condition (all *p*s < .016), and that performance with Number words was significantly higher than performance with Dots ($p = .049$) and Photographs ($p = .016$). No other comparisons were significant (all *p*s > .19). To correct for multiple comparisons against the Stickers baseline condition we conducted one-tailed Dunnett's tests, which revealed significant effects for Numerals ($p = .02$), Number words ($p < .001$) and Sweets ($p = .008$), and non-significant effects for Dots ($p = .087$) and Photographs ($p = .189$). At Time 2 the main effect of condition was only marginally significant, $F(5, 174) = 2.25$, $p = .086$, $n_p^2 = .06$. *T*-tests showed that performance in the Numerals, Number words, Dots and Sweets conditions was higher than performance in the Stickers condition (all *p*s < .035). No other comparisons were significant (all *p*s < .14). One-tailed Dunnett's tests between Symbol conditions and the Stickers condition revealed significant effects for Number words ($p < .018$), Sweets ($p = .034$) and Dots ($p = .034$) and non-significant effects for Photographs ($p = .23$) and Numerals ($p = .077$). At Time 3 the main effect of condition was non-significant, $F(5, 174) = 1.53$, $p = .182$, $n_p^2 = .05$. In sum, all Symbol conditions apart from Photographs differed significantly from the Stickers condition at one or more points in time, but the overall difference between Stickers and Symbol conditions decreased over time as children's performance tended towards ceiling. It must be noted that, because children's performance improved over time in all conditions, there was less room for a beneficial effect of symbols to be apparent at later points in time. Nonetheless, the findings from this analysis are more consistent with symbols helping children either immediately or rapidly (i.e. within the first three test trials) than with symbols helping children to learn gradually from feedback. This point will be returned to below.

To interpret the interaction between Age and Time we examined the effects of Age at Time 1, Time 2 and Time 3. Three univariate ANOVAs with Age as a between-child factor showed significant effects of age at each point in time (all *F*s > 7.26, all *p*s < .008). The interaction was a result of the difference in performance between the younger and the older children being largest at Time 1 (Younger = 1.4/3 correct, Older = 2.2/3 correct), smaller at Time 2 (Younger = 2.1/3 correct, Older = 2.6/3 correct) and smallest at Time 3 (Younger = 2.4/3 correct, Older=2.8/3 correct).

Trial 1 performance

It is important to determine whether symbols improved children's performance from trial 1 or whether they allowed rapid learning from negative feedback in the first three trials. To investigate this, we used one-tailed Fisher's exact tests to examine whether the frequency of correct responses in each Symbol condition was greater than the frequency of correct responses in the Stickers condition (these frequencies are recorded in Table 2). All Symbol conditions showed a trend for more correct responses than the Stickers condition. This difference was significant for Number words ($p = .001$) and Sweets ($p = .035$), but not for Dots ($p = .059$), Photographs ($p = .13$) or Numerals ($p = .15$). Indeed, once a correction for multiple comparisons was made, only Number words were significantly different from Stickers.

Comparison against chance over trials 1–3

From Table 2 it is apparent that the mean number of correct responses over the first three trials did not greatly exceed that expected by chance in several of the Symbol conditions. There are two possible explanations for this. One is that children in the Symbol conditions were largely guessing during the early trials; this random responding might be statistically better than the performance of children in the Stickers condition, who were systematically wrong, but clearly would not constitute good performance. The second explanation is that some children in the Symbol conditions were responding consistently poorly, while others were responding consistently well, leading to overall condition means of near-chance levels. In this latter case, the distribution of scores over the first three trials should deviate from that expected by chance, even though the group mean might not. This pattern would indicate that the correct interpretation for the effect of symbols was that they enabled more children to perform consistently and correctly from trial 1. To distinguish between these two possibilities we compared the frequency distribution of 0, 1, 2 or 3 correct responses over the first three test trials with that expected by chance for each experimental condition[2]. The frequency distributions and the results of the chi-squared analyses

[2] If all children in a sample of 30 are guessing then the expected frequencies for 0 and 3 correct responses are both 3.75 and the expected frequencies for 1 and 2 correct responses are both 11.25. Because the expected frequencies for 0 and for 3 correct responses are each less than 5 it would be inappropriate to conduct a chi-squared analysis to compare separate observed frequencies of 0, 1, 2 and 3 correct responses against chance. To solve this problem we combined the frequencies of 0 and 3 correct responses (total expected frequency = 7.5) and the frequencies of 1 and 2 correct responses (total expected frequency = 22.5). Comparing observed and expected frequencies for these combined responses using a chi-squared test ($df = 1$) showed that the distribution of children's scores differed from that expected by chance in every condition.

© 2009 The Authors. Journal compilation © 2009 Blackwell Publishing Ltd.

are presented in Table 2. The chi-squared analyses show that the frequency distributions in all six conditions differ significantly from those expected by chance. That is to say, the group means for several conditions may have been relatively close to 50% correct, but children were not guessing in any condition. Inspection of the frequency distributions suggests that most children were either consistently correct or consistently incorrect over the first three trials, with more children giving consistently correct responses in Symbol conditions than in the Stickers condition. In a further analysis we examined this hypothesis directly.

Consistency or improvement over trials 1–3?

For a more powerful test of whether Symbol conditions were helping children to deploy the correct strategy immediately, or helping them to learn from feedback over the first test trials, we examined the pattern of correct responses over trials 1 to 3. If symbols help children to respond consistently and correctly from the very first test trial, then more children should respond correctly on all of the first three test trials in the Symbol conditions than in the Stickers condition. It is clear from Table 2 that there was indeed a trend for more children to respond consistently correctly in the Symbol conditions. Across all of the Symbol conditions, 84 children were consistently correct compared with 66 children who were not. In the Stickers condition, 7 children were consistently correct compared with 23 children who were not. A chi-squared test confirmed that there was a significant difference between these patterns of performance, $\chi^2(1,180) = 10.7$, $p < .01$.

In contrast, if symbols were enabling children to learn rapidly from initial negative feedback then Symbol conditions should have more children who began with an error on trial 1, but who switched to the correct response strategy on trial 2 or trial 3. In fact, the number of children showing this pattern of improvement was relatively low in each condition: Stickers = 6, Numerals = 8, Number words = 2, Dots = 2, Sweets = 8, Photographs = 6. We conducted a chi-squared test to compare the frequency of this pattern in the Stickers condition (6 children, compared with 24 children who showed other response patterns) with the frequency of this pattern combined over the Symbol conditions (26 children, compared with 124 who showed other response patterns). This analysis confirmed that these patterns did not differ, $\chi^2(1,180) = .12$, $p > .99$, suggesting that children were no more likely to improve over the first three trials in the Symbol conditions than in the Stickers condition. Thus, the effect of symbols seems to have been to increase the likelihood that children would perform correctly and consistently from the very first test trial.

On the symbol reading test, the overall sample of children read 2.4/3 (77%) of the numerals '1', '2' and '3' correctly, and 0.13 (4%) of the number words 'one', 'two' and 'three' correctly, consistent with Bialystok (1991).

In sum, we found clear trends for children to perform better on reverse-contingency tasks when the stickers that could be won were represented symbolically. Although older children performed better than younger children, both younger and older children performed better in the Symbol conditions. Most conditions showed improved performance over time, but the effect of symbols was to enable more children to perform correctly from the very first test trial, not to increase the rate of learning over time. This effect was apparent even when the symbols were intrinsically attractive (sweets), or unreadable (number words), and there was a trend for an effect when the symbols were direct reminders of the children's goal (photographs of stickers). This pattern suggests that none of these factors was essential for children's performance to be better in a Symbol condition. However, because photograph symbols showed only a trend towards helping children's performance, we tested the robustness of this effect on two new samples of children in Experiment 2. Therefore we reserve a more complete discussion of the effects of different kinds of symbol for the general discussion.

The main objective of Experiment 2 was to examine the nature of the effect of symbols. The findings from Experiment 1 suggest that symbols do not help children to learn from feedback. However, this finding that symbols result in improved performance from the very first test trial could be accounted for by two very different explanations. Explanation 1: symbols might be helping children to formulate the correct strategy (on the first test trial or during training). Explanation 2: symbols might be helping children to execute a strategy that they successfully formulate on all versions of the task (symbolic and non-symbolic). To distinguish between these possibilities we took our inspiration from Boysen and Berntson (1995), who showed that chimpanzees that performed *above* chance in a symbolic (numeral) version of a reverse-contingency task performed *below* chance when, within the same block of trials, they switched to a non-symbolic version of the task (presented with sweets). For Boysen and Berntson (1995) this finding was consistent with Explanation 2 (above): symbols were helping chimpanzees to execute a strategy that they had already learned over many previous trials. When subsequently faced with a non-symbolic version of the task, the animals were unable to execute this strategy because they could not inhibit the prepotent response of pointing to the larger amount of sweets.

It is possible that Explanation 2 also applies to children. However, recent findings by Carroll *et al.* (2007a, 2007b) suggest that, unlike chimpanzees, children's difficulties on the task can be alleviated by helping them to formulate the correct strategy (Explanation 1). Carroll *et al.* (2007a) found that children performed better on a reverse-contingency task when the prompt to respond on a test trial related the child's response explicitly to the outcome of winning or losing the reward

© 2009 The Authors. Journal compilation © 2009 Blackwell Publishing Ltd.

('point to a box for me [the opponent] to open so that I don't get the sticker'), compared with children who received 'standard' prompts ('point to a box for me to open'). Critically, children who had been given the experimental prompt continued to perform well even on later trials in which they were prompted with 'standard' prompts. This transfer suggests that the novel prompt helped children to formulate a generalizable strategy for responding. The same pattern was found in Carroll *et al.* (2007b), when children played a two-object variant reverse-contingency task in which one box contained a desirable sticker and the other contained a non-desirable piece of paper. Not only did children in this condition perform significantly better than children in a 'standard' condition where there was no non-desirable object, but children who began with the two-object condition continued to perform better on later trials in which they played a 'standard' version of the task.

No previous study of children has tested whether the beneficial effect of symbols will transfer to a non-symbolic condition. Our aim in Experiment 2 was to use this method to distinguish between the two alternative explanations for the effects of symbols on children's performance.

We selected two symbol types from Experiment 1, Number words and Photographs, because they have a number of contrasting symbolic properties, enabling us to test whether any of these properties were necessary for observing a generalizable effect of symbols. Words are representationally opaque and cannot be read by most 3- to 4-year-old children; they have low intrinsic desirability; and they do not resemble stickers and so do not serve as a direct reminder of the task goal of winning stickers. In contrast, photographs of stickers are representationally transparent and do serve as a direct reminder of the task goal of stickers. The use of Photograph symbols also enabled us to check the robustness of the non-significant effect of Photograph symbols in Experiment 1.

In Experiment 2 we repeated the Sticker, Photograph, and Number word conditions of Experiment 1. We also included two 'mixed' conditions, in which children completed three initial test trials with either Number words or Photographs, before completing a further six test trials with Stickers. Experiment 1 showed clear effects of Symbol conditions after just three test trials, so we gave children in the two mixed conditions three trials with Symbols before changing to Stickers.

Experiment 2

Method

Participants

We tested 150 children with a mean age of 3;10 (range 3;2–4;10). Children were randomly allocated to one of five conditions: Stickers, Number words, Number words–Stickers, Photograph, and Photograph–Stickers, with 30 children in each group. A univariate ANOVA showed no difference in the mean ages of children in each group, $F(4, 145) = .194$, $p = .94$, with the mean age of children in the Stickers condition (3;9) falling within the range of the means for the other conditions (3;8–3;9). The distribution of ages and genders for each condition is detailed in Table 1.

Procedure

The warm-up and test protocols followed the same basic procedure as in Experiment 1. The only difference from Experiment 1 was in the Number words–Stickers and Photograph–Stickers conditions. Here, the first three test trials were identical to the Number word and Photograph conditions respectively. At the start of the fourth test trial, children were told that for the rest of the game, stickers would be placed in one of the boxes instead of cards. Six of these trials were completed, and were identical to the trials in the Stickers condition.

Results and discussion

An initial analysis compared the effects for males and females. Each child's number of correct responses (out of 9) was entered into an ANOVA with condition and sex as between-child factors. This analysis revealed no effect of sex, $F(1, 140) = 1.14$, $p = .29$, $n_p^2 = .008$, and no significant interaction between sex and condition, $F(4, 140) = 1.97$, $p = .10$, $n_p^2 = .05$. There was a significant main effect of condition $F(4, 140) = 3.63$, $p = .008$, $n_p^2 = .09$. Data were combined for males and females in all subsequent analyses, and the main effect of condition was explored fully in a further analysis.

The proportion of children giving a correct response on each test trial for the five conditions is plotted in Figure 2 (for clarity, separate graphs are presented, although the same Stickers condition appears in each graph). There was a clear trend for performance in the Symbol conditions to be better than that in the Stickers condition. In both the Number words–Stickers and Photograph–Stickers conditions there was a noticeable dip in performance on the first post-switch trial (trial 4), but after this performance seemed to recover and more closely resembled performance in the corresponding Symbol condition than in the Stickers condition.

Overall analysis of trials 1–3

We began by analysing mean performance for each condition over the first three test trials. We conducted a univariate ANOVA with Condition and Age (younger versus older half of the children in each condition) as between-subject factors. This analysis showed a main effect of Condition, $F(4, 140) = 4.81$, $p = .001$, $n_p^2 = .12$. *T*-tests showed that performance in all Symbol

© 2009 The Authors. Journal compilation © 2009 Blackwell Publishing Ltd.

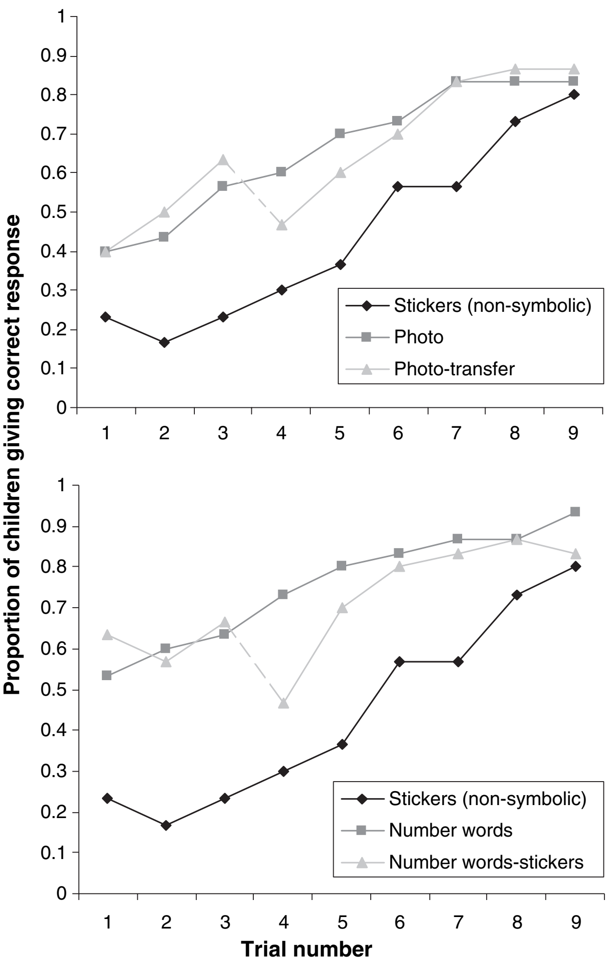

Figure 2 *Proportion of children giving a correct response on each trial in the five conditions of Experiment 2. The dashed lines from trials 3 to 4 in the Symbol–Stickers conditions indicate the switch from Symbol to Stickers trials. NB. The 'Stickers (non-symbolic)' condition corresponds to the same sample of children in each panel of the figure.*

conditions was higher than performance in the Stickers condition (all *p*s < .016), but no symbol condition differed from any other (all *p*s > .139). To correct for multiple comparisons against the standard baseline condition we conducted one-tailed Dunnett post-hoc tests, which showed that performance on all four symbol conditions (Number words, Number words–Stickers, Photograph, and Photograph–Stickers) was better than performance on the Stickers condition (all *p*s < .027). Thus, we twice replicated the finding from Experiment 1 that children perform significantly better when the task is presented with number words. Moreover, the fact that children also performed better in two Photograph conditions suggests that the non-significant effect of Photographs in Experiment 1 did reflect a genuine trend for performance to be better than in the standard condition. The analysis also showed a main effect of Age, $F(1, 140) = 12.2$, $p = .001$, $n_p^2 = .08$, with older children responding correctly more often than younger children. There was no interaction between these factors, $F(4, 140) = .34$, $p = .92$, $n_p^2 = .007$.

Trial 1 performance

We used one-tailed Fisher's exact tests to examine whether the frequency of correct responses in each Symbol condition on trial 1 was greater than the frequency of correct responses in the Stickers condition (these frequencies are recorded in Table 3). All Symbol conditions showed a larger number of correct responses than the Stickers condition. This difference was significant for the Number words ($p = .016$) and Number words–Stickers ($p = .0019$) conditions, but not for either Photograph condition (both *p*s > .13). Once a Bonferroni correction was applied for four multiple comparisons only the Number words–Stickers comparison remained significant.

Comparison against chance over trials 1–3

From Table 3 it is apparent that the mean number of correct responses over the first three trials did not greatly exceed that expected by chance in some of the Symbol conditions. As in Experiment 1, we compared the frequency distribution of 0, 1, 2 or 3 correct responses over the first three test trials with that expected by chance for each experimental condition. The frequency distributions and the results of the chi-squared analyses are presented in Table 3. The chi-squared analyses show that the frequency distributions in all five conditions differ significantly from that expected by chance. As in

Table 3 *Frequency of correct responses over the first three trials of each condition of Experiment 2 and comparison of this distribution against chance. Mean number of correct responses in each condition in Experiment 2*

		Time 1 (Trials 1–3)						Time 2 (Trials 4–6)	Time 3 (Trials 7–9)
Number of correct responses	Correct responses on Trial 1	0	1	2	3	Comparison between observed distribution and chance[2]	Mean/3	Mean/3	Mean/3
Expected by chance for $n = 30$	15	3.75	11.25	11.25	3.75		1.5	1.5	1.5
Stickers	7	19	7	0	4	χ^2 (1,30) = 42.7 $p < .001$	0.6	1.2	2.1
Number words	16	7	6	4	13	χ^2 (1,30) = 27.8 $p < .001$	1.8	2.4	2.7
Number words–Stickers	19	7	4	5	14	χ^2 (1,30) = 32.4 $p < .001$	1.9	2.0	2.5
Photograph	12	12	4	4	10	χ^2 (1,30) = 37.4 $p < .001$	1.4	2.0	2.5
Photograph–Stickers	12	11	4	3	12	χ^2 (1,30) = 42.7 $p < .001$	1.5	1.8	2.6

© 2009 The Authors. Journal compilation © 2009 Blackwell Publishing Ltd.

Experiment 1, the group means for several conditions may have been relatively close to 50% correct, but children were not guessing in any condition. Inspection of the frequency distributions suggests that most children were either consistently correct or consistently incorrect over the first three trials, with more children giving consistently correct responses in Symbol conditions than in the Stickers condition. In a further analysis we examined this hypothesis directly.

Consistency or improvement over trials 1–3?

To test whether Symbol conditions were enabling children to deploy the correct strategy immediately, or enabling more children to learn rapidly over the first test trials, we examined the frequency of correct responses over trials 1 to 3. Table 3 shows that there was a trend for more children to respond consistently correctly on trials 1–3 in the Symbol conditions than in the Stickers condition. Across all of the Symbol conditions, 49 children were consistently correct compared with 71 children who were not. In the Stickers condition, 4 children were consistently correct compared with 26 children who were not. A chi-squared test confirmed that these patterns of performance differed significantly, $\chi^2(1,180) = 7.94$, $p < .01$.

In contrast, the number of children who showed evidence of learning by beginning with an error on trial 1 but switching to the correct response strategy on trial 2 or trial 3 was relatively low in each condition: Stickers = 3, Number words = 4, Number words–Stickers = 3, Photograph = 6, Photograph–Stickers = 7. In the Stickers condition, 3 children showed this improvement compared with 27 children who did not. Across all of the Symbol conditions, 20 children showed this improvement compared with 100 children who did not. A chi-squared test analysis confirmed that these patterns did not differ, $\chi^2(1, 150)=.82$, $p > .99$, suggesting that children were no more likely to improve over the first three trials in the Symbol conditions than in the Stickers condition. Thus, as in Experiment 1, the effect of symbols seems to have been to increase the likelihood that children would perform correctly and consistently from the very first test trial.

Analysis of trials 4–9

To test whether the effect from symbols would be sustained when children switched to playing with stickers, we conducted an ANOVA with children's mean performance over trials 4–6 and trials 7–9 as a within-child factor (Time) and Condition as a between-child factor. There was a significant main effect of Condition, $F(4, 140) = 3.14$, $p = .016$, $n_p^2 = .08$. *T*-tests showed that performance in all Symbol conditions was higher than performance in the Stickers condition (all *p*s < .046) but no symbol condition differed from any other (all *p*s > .16). To correct for multiple comparisons against the standard baseline condition we conducted one-tailed Dunnett post-hoc tests, which showed that performance in the Number words, Number words–Stickers and Photograph conditions was significantly better than in the Stickers condition (all *p*s < .034), and showed a non-significant trend for better performance in the Photograph–Stickers condition ($p = .071$). The analysis also revealed main effects of Age, $F(1, 140) = 17.9$, $p < .001$, $n_p^2 = .11$, and Time, $F(1, 140) = 53.6$, $p < .001$, $n_p^2 = .28$, and no significant interaction between any factors (all *p*s > .145). Older children performed better than younger children, and performance on trials 7–9 was better than performance on trials 4–6.

On the symbol reading test the overall sample of children read 2.3/3 (76%) of the numerals '1', '2' and '3' correctly and 0.09/3 (3%) of the number words 'one', 'two' and 'three' correctly.

In sum, the results suggest that using Number word and Photograph symbols results in children performing better, with the magnitude of this effect being somewhat larger for Number words than for Photographs. This effect resulted from a larger number of children performing consistently correctly from the very first test trial in the Symbol conditions. Critically, Experiment 2 also showed that this effect can be transferred to a non-symbolic condition. For trials 4–9, children in the Stickers, Photograph–Stickers and Number words–Stickers conditions were presented with exactly the same task in which one of the two boxes was baited with stickers. Nonetheless, children in both the Number words-Stickers and Photograph-Stickers conditions continued to show an effect of having played the game with symbols on earlier trials. This effect was statistically significant for the Number words–Stickers condition and showed a substantial though non-significant trend in the Photograph–Stickers condition. Our tentative interpretation of the Photograph–Stickers condition is that this constitutes a genuine transfer effect of the same kind as for the Number words–Stickers condition, but that this fails to reach statistical significance because photograph symbols resulted in a smaller (though still significant) initial improvement in children's initial performance than did word symbols.

General discussion

Consistent with the existing literature on the Windows task (e.g. Carroll *et al.*, 2007a, 2007b; Hala & Russell, 2001; Russell *et al.*, 1991, 1994; Simpson *et al.*, 2004) and on other reverse-contingency tasks (e.g. Carlson *et al.*, 1998, 2005), Experiments 1 and 2 both showed that 3- to 4-year-old children find it difficult to point to an empty box to give away in order to keep a baited box for themselves. Consistent with the existing literature, this effect was larger in younger children than in older children, and occurred when the treat was stickers (both stickers and sweets have been used in previous studies), and when the child had a puppet as an opponent (previous studies have found similar effects with puppet

© 2009 The Authors. Journal compilation © 2009 Blackwell Publishing Ltd.

opponents, human opponents and no opponent). The current experiments showed learning over time in all conditions, which has been observed in some earlier studies (e.g. Carroll *et al.*, 2007a, 2007b; 4-year-olds in Carlson *et al.*, 2005) but not in others (e.g. 3-year-olds in Carlson *et al.*, 2005; Hala & Russell, 2001; Russell *et al.*, 1991, 1994). The overall level of performance was similar to that observed in some earlier studies (e.g. Carroll *et al.*, 2007a, 2007b) but higher than that observed in others (e.g. Hala & Russell, 2001; Russell *et al.*, 1991, 1994). A possible reason for these differences is that studies in which learning has been observed and in which children tend to perform better have also tended to have somewhat older children as participants. However, the fact that the effect of symbols was observed in both the older and the younger halves of the samples in the current studies suggests that our results are not specific to older 3- to 4-year-olds. Future work might examine how far the effect of symbols is apparent in even younger children.

At the outset we noted that the existing literature provides evidence that symbols can lead to better performance among 3- to 4-year-olds on reverse-contingency tasks (e.g. Carlson *et al.*, 2005; Russell, 1996) but that it was far from clear how symbols affected the cognitive processes involved in performing these tasks, or what properties of symbols were necessary for such effects.

In Experiment 1 we examined three factors in the relationship between symbols and the objects they represent that might be relevant for the beneficial effects of symbols on children's performance: inherent desirability, resemblance to the goal object, and children's ability to read the symbol. First, we reasoned that children's problem on non-symbolic reverse-contingency tasks might be with resisting distraction from the intrinsically desirable and attention-grabbing reward (e.g. stickers). If this were correct then symbols might assist children to the extent that they are less intrinsically desirable and attention-grabbing than the reward that they symbolized. Contrary to this suggestion, the symbol that we expected to be most intrinsically desirable – Sweets – was as effective as any other symbol at improving children's performance. (Many studies showing poor performance on the Windows task in preschool children have used sweets as the non-symbolic treat to be won, and the poor performance in these studies suggests that sweets are indeed desirable and attention-grabbing when they are children's goal, rather than a symbolic stand-in for stickers.)

Second, we reasoned that symbols might improve children's performance because they did not typically resemble the child's goal (of obtaining stickers). Contrary to this suggestion, Photograph symbols were effective in enhancing children's performance (this effect was marginal in Experiment 1 but significant in Experiment 2), despite the fact that photographs were a clear reminder of the child's goal of obtaining stickers. It seems reasonable to assume that children understood that the photograph symbolized the stickers, because much younger children can be shown to understand the symbolic nature of photographs (e.g. DeLoache, 1991; DeLoache, Pierroutsakos & Uttal, 2003).

Third, although children in all conditions were told the numerical meaning of the symbol, we reasoned that children's ability to read the symbol for themselves might affect whether their performance was improved. On the symbol reading test very few children could read number words, while many could read the corresponding numerals. However, this difference appeared to have little effect, as both types of symbol led to better performance.

Our finding that children's performance on reverse-contingency tasks is improved by several types of symbol is consistent with data from Carlson *et al.* (2005), who also found a trend for improved performance whether sweets were symbolized with rocks (least effective), dot patterns (somewhat more effective) or pictures of an elephant and a mouse (the only condition for which the effect was statistically significant). In addition, our data suggest that a symbol's intrinsic desirability, the degree to which a symbol might resemble the object it symbolizes, and children's ability to read the symbol are not critical influences on the effect that symbols have on children in reverse-contingency tasks.

In Experiment 2 we asked whether the effects of symbols would transfer to a non-symbolic version of the task. We found clear evidence that children do transfer the effect of a symbolic condition when they switch to a non-symbolic condition, with a statistically significant effect for Number word symbols and a similar trend for Photograph symbols. Together with the findings from Experiment 1, this suggests that symbols are not enabling children to execute a correct response strategy that is beyond their abilities in non-symbolic versions of the task. Rather, the findings suggest that the effect of symbols is to help children from the very first test trial to *formulate* the correct response strategy, and that, once formulated, this strategy is sufficient to sustain good performance in a non-symbolic condition. In the current studies the non-symbolic condition followed immediately after the symbolic condition, and it would be interesting for future work to determine the duration of this effect of symbols.

Why do children have difficulty formulating a response strategy and how do symbols help?

Theories of executive function in adults typically distinguish between processes involved in the online performance of each trial of a given task (such as the need to hold task rules and other information in mind and to inhibit incorrect responses), and processes involved in deciding, monitoring and updating what is attempted on each trial (such as formulating what to do on the basis of the task rules, and adjusting this as necessary on the basis of performance) (e.g. Shallice & Burgess, 1996). The current findings suggest that

© 2009 The Authors. Journal compilation © 2009 Blackwell Publishing Ltd.

children's difficulty is with the latter process of formulating a response strategy, but there are different ways in which this difficulty can be characterized.

Carlson and Zelazo (2008) suggest that success on reverse-contingency tasks requires children to hold in mind a relatively complex higher-order symbolic rule, and that symbol conditions may help children with formulating the rule. The rule suggested by Carlson and Zelazo (2008) relates children's desire for the reward to the reverse-contingency rule of the game: '……[successful] children are able to say to themselves "Yes I want the [reward], but in this game, if I want the [reward], then I have to point to the [empty box]" ' (p.9). On this account, children who would normally be unable to hold in mind such higher-order rules are helped by the presence of symbols with the problem of 'representing and reflecting upon the situation from more than one angle' (p.8). Although the authors do not use this account to explain how good performance in a symbol condition might be sustained in later, non-symbolic trials of the task, it is conceivable that once symbols have done their work of enabling children to reflect upon the situation from more than one angle then it will be possible for children to hold the high-order rule in mind for later trials in which symbols are not used. What is less clear is whether this explanation could also explain the effects of other manipulations (of the wording of the test question, the mode of responding, or the objects in the boxes) that also result in better performance on reverse-contingency tasks.

In contrast, Carroll *et al.* (2007a, b) suggest that success on reverse-contingency tasks may be achieved with much simpler reasoning. Their explanation takes inspiration from the fact that simply telling children to point to the empty box[3] leads most 3-year-olds to perform at ceiling on a reverse-contingency task (Simpson *et al.*, 2004). This clearly shows that 3-year-olds are not irretrievably bound to point to rewards, provided they are equipped with a suitable rule to guide their behaviour. Carroll *et al.* suggest that, on standard forms of reverse-contingency tasks, 3-year-olds fail to derive such a rule because their decision-making is dominated by their desire to obtain the reward. Therefore 3-year-olds are led to point repeatedly and incorrectly at the baited box, even in the face of negative feedback from repeated losses. Like Carlson and Zelazo (2008), Carroll *et al.* (2007a, b) suppose that children's difficulty is with 'thinking outside the box' of the most obvious decision to point to the desirable reward. However, Carroll *et al.*'s evidence that several, apparently diverse, manipulations yield sustained improvement in children's performance on reverse-contingency tasks (i.e. performance that generalizes to a 'standard' form of the task) leads them to suggest that successful children need not be doing anything as complicated as representing higher-order rules about the task (cf. Carlson & Zelazo, 2008).

One way in which children can be helped to perform better on reverse-contingency tasks is by directing the child to think in terms of which box to give away, either by manipulating the wording used to prompt the child to respond (e.g. 'point to a box for me [the opponent] to open so that I don't get the sticker'; Carroll *et al.*, 2007a) or by adding a salient but non-desirable object to the empty box (Carroll *et al.*, 2007b). Another way is by interrupting the child's tendency to point impulsively at the desirable sticker by having them point with an arrow, rather than with their finger (e.g. Carlson *et al.*, 1998; Carroll *et al.*, under submission; Hala & Russell, 2001). The idea in this case is that preventing the child from pointing impulsively on the basis of their desire affords them the possibility of formulating an alternative response strategy. We suggest that the current findings with symbols might fit with this latter explanation. That is to say, presenting the reverse-contingency task with symbolic stand-ins for the rewards to be won helps children avoid pointing impulsively on the basis of their desire, affording them the possibility of formulating an alternative response strategy such as 'point to the empty box'. Formulating such a response strategy may not require very sophisticated reasoning abilities. Nonetheless, it can explain how it is possible for some children to respond consistently and correctly from the very first test trial. Moreover, if symbolic conditions are helping children to formulate a general response strategy, then it is also clear why children continue to perform well even when they switch to a non-symbolic condition. Importantly though, this analysis would also suggest that symbols are not helping children to a unique new way of solving reverse-contingency tasks, but are just one among a variety of ways in which they can be helped to infer a relatively simple alternative response strategy.

In sum, the current paper is the first study to examine whether the effect of symbols on children's reverse-contingency task performance transfers to a non-symbolic condition (see also Beck & Carlson, 2007, for recent findings consistent with our own). Children do transfer the effect of symbols to subsequent non-symbolic trials whereas chimpanzees do not (e.g. Boysen & Berntson, 1995), indicating that the effect of symbols on children may be qualitatively different from the effect of symbols on chimpanzees. Our suggestion is that symbols enable children to overcome their tendency to point impulsively on the basis of their desire, and that this enables them to *think* flexibly in order to formulate a response strategy that allows them to respond successfully. The medium of such response strategies is currently unclear. An interesting possibility is that children formulate their strategies in 'inner speech', which may be involved in the solution of other

[3] Studies of non-human primates typically require the animal to choose between a larger and a smaller quantity, whereas almost all studies of children have required children to choose between a baited box and an empty box (Carlson *et al.* 2005 is the only exception). The relevance of this difference is unclear, and should be investigated.

© 2009 The Authors. Journal compilation © 2009 Blackwell Publishing Ltd.

executive problems such as the Tower of London (e.g. Fernyhough & Fradley, 2005; for a related view see Carlson & Zelazo, 2008). Whether or not this turns out to be the case, the current findings offer an insight into how children gain increasingly flexible control over their behaviour not only by increasing their capacities for memory and inhibitory control (see Carlson, Moses & Hix, 1998) or thinking with progressively more complex structures (e.g. Carlson & Zelazo, 2008; Zelazo *et al.*, 2003) but also by formulating efficient strategies for responding, given the rules of the game or constraints of a given situation.

Acknowledgements

This research was supported by a grant from the ESRC, UK, grant number ESRC RES-000-22-0796.

References

Apperly, I.A., Williams, E., & Williams, J. (2004). Domain specificity in 3–4 year olds' understanding of picture versus word symbols. *Child Development*, **75** (5), 1510–1522.

Beck, D., & Carlson, S.M. (2007) *Executive function and symbolic representation: improving children's performance on the less is more task*. Poster presented at the Biennial meeting of the Jean Piaget Society, Amsterdam, May 2007.

Bialystok, E. (1991). Letters sounds and symbols: changes in children's understanding of written language. *Applied Psycholinguistics*, **12**, 75–89.

Bialystok, E. (1999). Cognitive complexity and attentional control in the bilingual mind. *Child Development*, **70** (3), 636–644.

Bialystok, E. (2000). Symbolic representation across domains in preschool children. *Journal of Experimental Child Psychology*, **76** (3), 173–189.

Boysen, S. T., & Berntson, G. G. (1995). Responses to quantity: perceptual versus cognitive mechanisms in Chimpanzees (Pan Troglodytes). *Journal of Experimental Psychology: Animal Behavior Processes*, **21** (1), 82–86.

Boysen, S.T., Berntson, G.G., Hannan, M.B., & Cacioppo, J. T. (1996). Quantity-based inference and symbolic representations in Chimpanzees (Pan troglodytes). *Journal of Experimental Psychology: Animal Behavior Processes*, **22** (1), 76–86.

Carlson, S.M., & Zelazo, P.D. (2008) Symbolic thought. In M.M. Haith & J.B. Benson (Eds.), *Encyclopaedia of infant and early childhood development*. (pp. 288–297) London: Academic Press.

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** (3), 672–691.

Carlson, S.M., Davies, A.C., & Leach, J.G. (2005). Less is more: executive function and symbolic representation in preschool children. *Psychological Science*, **16** (8), 609–616.

Carroll, D.J., Apperly, I.A., & Riggs, K.J. (2007a). The executive demands of strategic reasoning are modified by the way in which children are prompted to think about the task: evidence from 3- to 4-year-olds. *Cognitive Development.*, **22** (1), 142–148.

Carroll, D.J., Apperly, I.A., & Riggs, K.J. (2007b). Choosing between two objects reduces 3-year-olds errors on a reverse-contingency test of executive function. *Journal of Experimental Child Psychology*, **98** (3), 184–192.

DeLoache, J.S. (1991). Symbolic functioning in very young children: understanding of pictures and models. *Child Development*, **62**, 736–752.

DeLoache, J.S., Pierroutsakos, S.L., & Uttal, D.H. (2003). The origins of pictorial competence. *Current Directions in Psychological Science*, **12**, 114–118.

Fernyhough, C., & Fradley, E. (2005). Private speech on an executive task: relations with task difficulty and task performance. *Cognitive Development*, **20** (1), 103–120.

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

Hala, S., & Russell, J. (2001). Executive control within strategic deception: a window on early cognitive development? *Journal of Experimental Child Psychology*, **80** (2), 212–241.

Kralik, J.D., Hauser, M.D., & Zimlicki, R. (2002). The relationship between problem solving and inhibitory control: cotton-top tamarind (*Saguinus oedipus*) performance on a reversed contingency task. *Journal of Comparative Psychology*, **116**, 39–50.

Liben, L. S. (1999). Developing an understanding of external spatial representations. In E. Sigel (Ed.), *Development of mental representations* (pp. 297–321). Mahwah, NJ: Lawrence Erlbaum Associates.

Russell, J. (1996). *Agency: Its role in mental development*. Hove, UK: Erlbaum.

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.

Russell, J., Jarrold, C., & Potel, D. (1994). What makes strategic deception difficult for children – the deception or the strategy? *British Journal of Developmental Psychology*, **12**, 301–314.

Samuels, M.C., Brooks, P.J., & Frye, D. (1996). Strategic game playing through the windows task. *British Journal of Developmental Psychology*, **14**, 159–172.

Shallice, T., & Burgess, P. W. (1996). he domain of supervisory processes and the temporal organisation of behaviour. *Philosophical Transactions of the Royal Society of London B*, **351**, 1405–1412.

Simpson, A., Riggs, K.J., & Simon, M. (2004). What makes the windows task difficult for young children: rule inference or rule use? *Journal of Experimental Child Psychology*, **87** (2), 155–170.

Vlamings, P.H.J.M., Uher, J., & Call, J. (2006). How the great apes (*Pan troglodytes*, *Pongo pygmaeus*, *Pan paniscus*, and *Gorilla gorilla*) perform on the reversed contingency task: the effects of food quantity and food visibility. *Journal of Experimental Psychology: Animal Behaviour Processes*, **32** (1), 60–70.

Werner, H., & Kaplan, B. (1963). *Symbol formation*. New York: John Wiley.

Zelazo, P. D., Muller, U., Frye, D., & Marcovitch, S. (2003). *The development of executive function in early childhood*. Monographs of the Society for Research in Child Development. Serial No. 274 [68 (3)].

Received: 19 June 2007
Accepted: 8 October 2008

© 2009 The Authors. Journal compilation © 2009 Blackwell Publishing Ltd.