

## Sequential planning in rhesus monkeys (*Macaca mulatta*)

Damian Scarf · Erin Danly · Gin Morgan ·  
Michael Colombo · Herbert S. Terrace

Received: 24 July 2010 / Revised: 25 November 2010 / Accepted: 30 November 2010 / Published online: 24 December 2010  
© Springer-Verlag 2010

**Abstract** In the current study, we examined the planning abilities of rhesus monkeys (*Macaca mulatta*) by training them on a five-item list composed of coloured photographs and then testing them on switch and mask trials. In contrast to previous studies where monkeys made responses using a joystick, in the current study, monkeys made responses directly to a touch screen. On switch trials, after a response to the first list item, the on-screen positions of two list items were exchanged. Performance on trials in which the second and third list items were exchanged was poorer compared to normal (non-switch) trials for all subjects. When the third and fourth items were exchanged, however, only one subject continued to show performance deficits. On mask trials, following a response to the first item, the remaining items were covered by opaque white squares. When two items were masked, all four subjects responded to each masked item at a level significantly above chance. When three items were masked, however, only one subject was able to respond to all three masked items at a level significantly above chance. The results of the present study indicate that three of our four monkeys planned one response ahead while a single monkey planned two responses ahead.

The significance of these findings is discussed in relation to previous studies on planning in chimpanzees and monkeys.

**Keywords** Planning · Serial order · Simultaneous chain

### Introduction

Recently, comparative psychologists have developed paradigms to investigate whether nonhuman animals share our ability to plan (e.g. Beran et al. 2004; Biro and Matsuzawa 1999; Correia et al. 2007; Inoue and Matsuzawa 2007, 2009; Mulcahy and Call 2006; Ohshima 1997; Osvath 2009; Osvath and Osvath 2008; Paxton and Hampton 2009; Raby et al. 2007; Scarf and Colombo 2009, 2010). One of the most frequently used paradigms is the *simultaneous chaining paradigm* (Terrace 1984), which has been used to test the planning abilities of a range of species (Humans: Cook and Wilson 2010; Silberberg and Kearns 2009, Chimpanzees: Biro and Matsuzawa 1999; Inoue and Matsuzawa 2007, 2009; Kawai 2001; Kawai and Matsuzawa 2000, Monkeys: Beran et al. 2004; Ohshima 1997; Scarf and Colombo 2009, and Pigeons: Scarf and Colombo 2010). The basic paradigm presents subjects with  $n$  number of simultaneously displayed stimuli. The subject's task is to respond to these stimuli in a prescribed order learnt during training. For ease of exposition, we refer to the stimuli with letters of the alphabet, and the order in which they should be pressed with an arrow. Thus, the order of responses to a list of five items is shown as  $A \rightarrow B \rightarrow C \rightarrow D \rightarrow E$ .

Matsuzawa and colleagues have conducted a series of experiments on the planning abilities of chimpanzees using variations of the simultaneous chaining paradigm. In the first of these, Biro and Matsuzawa (1999) trained their female chimpanzee Ai to respond to three numerals, ranging

D. Scarf (✉) · M. Colombo  
Department of Psychology, University of Otago,  
PO Box 56, Dunedin, New Zealand  
e-mail: scada298@student.otago.ac.nz

G. Morgan  
Department of Psychology, New Mexico State University,  
P.O. Box 30001, Las Cruces, NM 88003-8001, USA

E. Danly · H. S. Terrace  
Department of Psychology, Columbia University,  
1190 Amsterdam Ave, New York, NY 10027, USA

from 0 to 9, in ascending order. Once Ai had acquired this task, ‘switch trials’ were introduced. On switch trials, following a response to the first item (A), the positions of the second (B) and third (C) items were exchanged. Ai’s accuracy on switch trials was significantly lower than on normal (non-switch) trials (45% vs. 90%). Also, Ai’s response time to the second numeral on successfully completed switch trials was significantly longer than the response time to the second numeral on normal trials ( $\approx 900$  ms vs.  $\approx 500$  ms). Biro and Matsuzawa (1999) suggested that Ai’s lower accuracy on switch trials was the result of her executing a response plan formed at the start of a trial. Similarly, the higher response latency to the second item on correct switch trials was thought to reflect Ai updating the response plan. By “response plan”, we mean that at the start of each trial, the subject encodes the spatial position of each list item and then prepares an appropriate sequence of motor responses.

Matsuzawa and colleagues have also trained six chimpanzees on a ‘mask task’ (Inoue and Matsuzawa 2007, 2009; Kawai and Matsuzawa 2000). On this task, when a subject responds to the first item, the remaining items are covered by opaque white squares. In order to pass this task, the subject must use the response plan formed at the outset of a trial to guide responding. On each trial, when a masked item is responded to correctly, the item disappears from the display leaving only the to-be-responded-to items onscreen. Because the masked items disappear, when the subjects reach the last item of a list, it is the only item left onscreen and is thus always responded to correctly. Accordingly, when we report the number of masked items a subject can respond to the final item is excluded. Regarding Matsuzawa and colleague’s chimpanzees, their performance ranged widely across subjects. Two of the chimpanzees (Chloe and Pan) could only respond to a single masked item significantly above chance, two (Pal and Ai) were able to respond to four, one (Cleo) was able to respond to five and one (Ayumu) subject was able to respond to seven.

Beran and colleagues conducted a similar set of experiments with their chimpanzees and monkeys. With respect to switch trials, Beran et al. (2004) switched the positions of the second and third items within a five-item list. That is, following a response to the first item (A), the positions of the second (B) and third (C) items were exchanged, while the positions of the fourth (D) and fifth items (E) remained the same. Similar to Ai, the monkeys and chimpanzees performed markedly worse on switch trials compared to normal trials (Chimpanzees: 75% vs. 89%; Monkeys: 37.5% vs. 55%). On mask trials, the chimpanzees (with the exception of one chimpanzee that could respond to two masked items) and monkeys only responded significantly above chance to the first masked item, accuracy to the second and third masked items was no different from chance.

Given that many of Matsuzawa and colleagues chimpanzees were able to respond to between four and seven masked items, it is interesting that, with the exception of one subject, Beran et al.’s chimpanzees and monkeys were only able to respond to one masked item. A potential reason for this marked difference is that Beran et al. (2004) had their subjects respond using a joystick-controlled cursor, while Matsuzawa’s chimpanzees responded using a touch screen. Although this may appear to be a trivial issue, it is possible that this subtle difference represents what Bitterman (1965) referred to as a contextual variable, that is, a non-cognitive factor that is responsible for the observed differences in performance on a task. For example, as noted by Beran et al. (2004), it is possible that, because it takes longer to make a response with a joystick than it does with a touch screen, Beran et al.’s subjects had less information available in working memory than Matsuzawa and colleagues chimpanzees. Another possibility is that the spatial discontinuity created by using a joystick-controlled cursor impaired the performance of Beran et al.’s subjects. Indeed, studies with both chimpanzees and monkeys have shown that performance on cognitive tasks is adversely affected when there is a large degree of stimulus–response spatial discontinuity (Iwai et al. 1986; Meyer et al. 1965; Yaginuma and Iwai 1986).

It is possible that if Beran et al. (2004) had trained their subjects on a touch screen, some of their subjects would have shown performance levels comparable to those of Matsuzawa and colleagues chimpanzees (Inoue and Matsuzawa 2007, 2009; Kawai and Matsuzawa 2000). To address this possibility, in the present experiment, we investigated the planning abilities of four monkeys when the stimuli were presented on a touch screen.

## Method

### Subjects and apparatus

The subjects were four 15-year-old male rhesus monkeys (Benedict, Horatio, Oberon, and Prospero). All of the monkeys had extensive experience at ordering arbitrary and numerical stimuli (e.g. Brannon et al. 2006; Brannon and Terrace 2000; Kornell and Terrace 2007; Subiaul et al. 2004; Terrace et al. 2003). The monkeys were housed in a colony room at the New York State Psychiatric Institute in accordance with the National Institutes of Health guidelines. The monkeys were neither water nor food deprived during the experiment. In addition to the food they obtained during each experimental session, subjects were fed a mixed diet of high-protein monkey chow (Ralston Purina, Richmond, IN) and various fruits.

Training and testing were conducted in four operant chambers. The chambers measured 33 in. high, 28.5 in.

long and 27 in. wide. The back and two side walls of the chamber were stainless steel. The front wall of the chamber housed a 15-in. 3 M MicroTouch™ touch screen, and two speakers located on the top right and top left of the touch screen. Positioned in front of the touch screen was a transparent Perspex panel that contained 16 (4 by 4 matrix) equally spaced rectangular cut outs (1.5 in. by 2 in.) that corresponded to the 16 positions at which stimuli could appear. A Gerbrands pellet dispenser was fitted to the outside of each chamber. Each touch screen and pellet dispenser was controlled by a Macintosh G5 desktop computer. Reinforcers were 190-mg Noyes (Noyes, Lancaster, NH) banana flavoured pellets.

## Procedure

Subjects learnt a single five-item list composed of five coloured digital images (1.5 in. by 2 in.). The list was composed as follows: Man → Tree → Parrot → Slippers → Glass Cube. All five items were displayed from the very start of training with monkeys learning the correct item order through the process of trial and error. In the first phase of training, the paradigm followed that used by Terrace et al. (2003), with the stimuli remaining on the screen throughout each trial. Subjects were trained until they reached the criterion of 60% or more correct responses in a single session. In this phase, the chance of responding to the first item correctly is one in five (20%), and the chance of responding to each of the remaining items correctly is one in four (25%). The reason the chance value for the first item is different to that for all subsequent items is because repeat responses are not counted as errors. The chance probability of responding correctly to all five stimuli is 0.08%. In the second phase of training, the paradigm followed that used by Matsuzawa and colleagues (e.g. Inoue and Matsuzawa 2009), with the stimuli disappearing following a correct response. In this phase, the probability of responding correctly by chance to the first item is 0.20, to the second 0.25, to the third 0.33, to the fourth 0.50 and to the fifth item 1.0. Thus, the chance probability of responding correctly to all five stimuli is 0.8%. Subjects were trained on this phase until they reached a criterion of 70% or more correct responses. In both phases, the spatial position of items was varied from trial to trial to ensure that subjects did not learn a rote motor sequence. With five stimuli and 16 possible positions, there are a total of 524,160 possible configurations.

## Switch condition

Once the monkeys had reached criterion on the second phase of training, switch trials were introduced. Now, each 60-trial session consisted of 54 normal (non-switch) trials

and 6 switch trials. Three switch types were used, the second (B) and third (C) items exchanging places (BC switch), the third (C) and fourth (D) items exchanging places (CD switch) and finally the fourth (D) and fifth (E) items exchanging places (DE switch). On switch trials, when a monkey made a response to the first item (A), the respective switch items exchanged places. For example, on DE switch trials, after a response to item A, the positions of items D and E were exchanged. Importantly, the brief stimulus blink created on switch trials was replicated on normal trials by, after a response to A, having the respective switch items disappear and then reappear in the same positions. Each switch type was run for 10 consecutive sessions starting with the BC switch, followed by the CD switch, and finishing with the DE switch. Across the ten sessions on each switch type, there were a total of 540 normal trials and 60 switch trials.

## Mask condition

After completing the switch condition, the monkeys were moved immediately to the mask condition. Each session was composed of 60 mask trials. On mask trials, following a response to item A, all of the remaining list items were replaced by white squares. In the first phase of training, only items A and B were presented. When the subject responded to item A, item B was masked. In the second, third and fourth phases of training, subjects were trained with three ( $A \rightarrow B \rightarrow C$ ), four ( $A \rightarrow B \rightarrow C \rightarrow D$ ) and finally five items ( $A \rightarrow B \rightarrow C \rightarrow D \rightarrow E$ ). Transition between phases was dependent on subjects responding at or above 70% to each masked item. For example, in order to reach criterion with four items, subjects had to respond correctly to items B, C and D on at least 70% of the trials within a single session.

## Results

### Training

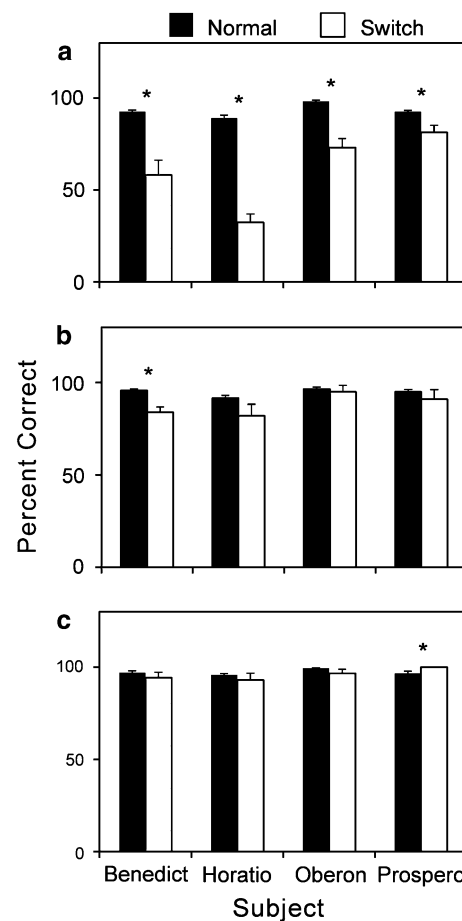
All four monkeys quickly reached the 60% criterion in the first phase of training in which the items remained on the screen for the duration of the trial. Trials to criterion for Benedict, Horatio, Oberon and Prospero were 240, 1,200, 120 and 1,020, respectively. Subjects also quickly transferred to the second phase of training, when an item disappeared following a correct response, requiring on average just 135 trials (range 60–180). The performance of our subjects on this stage is particularly impressive when compared to the time taken by Beran et al. (2004) subjects to acquire a five-item list (Experiment 2). Their three chimpanzees, Lana, Sherman and Mercury, required 1,710,

2,000 and 1,480 trials, respectively, to reach a 70% criterion. Their two monkeys, Baker and Willie, required 7,035 and 5,593 trials, respectively. It is important to remember, however, that the subjects in the present study were much more experienced in ordering lists of arbitrary items than the subjects in the Beran et al. (2004) study (e.g. Brannon et al. 2006; Terrace et al. 2003).

#### Switch condition

All subjects performed significantly better on normal trials than on BC switch trials, in which the second and third items exchanged places, (Fig. 1a, Benedict:  $t(9) = 4.23$ ,  $P = 0.002$ ; Horatio:  $t_9 = 12.01$ ,  $P < 0.001$ ; Oberon:  $t_9 = 5.10$ ,  $P = 0.001$ ; Prospero:  $t_9 = 2.85$ ,  $P = 0.019$ ). On BC switch trials, if subjects were planning a sequence of responses at the outset of a trial, we would expect them to make a larger number of ‘planning errors’. On BC switch trials, a planning error was defined as a correct response to A followed by a response to C. Indeed, all subjects made significantly more planning errors on BC switch trials than normal trials (Fig. 2a, Benedict:  $t_9 = 2.51$ ,  $P = 0.033$ ; Horatio:  $t_9 = 7.09$ ,  $P < 0.001$ ; Oberon:  $t_9 = 3.94$ ,  $P = 0.003$ ; Prospero:  $t_9 = 3.35$ ,  $P = 0.009$ ). With respect to response latencies on correct BC switch trials, all subjects took significantly longer to respond to item B on switch trials compared to normal trials (Fig. 3a, Benedict:  $t_8 = 3.54$ ,  $P = 0.008$ ; Horatio:  $t_7 = 5.56$ ,  $P = 0.001$ ; Oberon:  $t_9 = 5.28$ ,  $P = 0.001$ ; Prospero:  $t_9 = 3.10$ ,  $P = 0.013$ ). Together, the poorer performance, higher number of planning and longer response latencies to item B on BC switch trials suggests that all subjects were planning at least a response to B at the start of a trial.

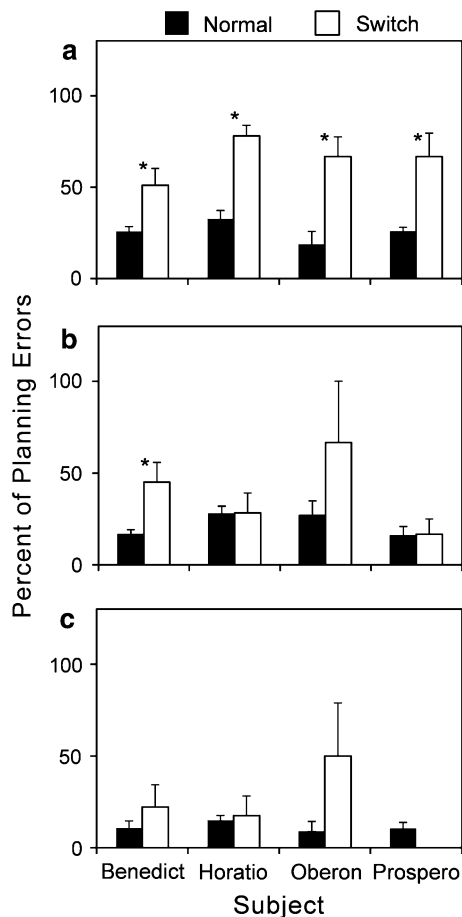
In contrast to the poor performance by all subjects on BC switch trials, only one subject (Benedict) made significantly more errors on CD switch trials compared to normal trials (Fig. 1b, Benedict:  $t_9 = 3.96$ ,  $P = 0.003$ ; Horatio:  $t_9 = 1.49$ ,  $P = 0.17$ ; Oberon:  $t_9 = 0.38$ ,  $P = 0.71$ ; Prospero:  $t_9 = 0.65$ ,  $P = 0.53$ ). On CD switch trials, a planning error was defined as correct responses to A and B followed by a response to D. Similar to the trial accuracy data, only Benedict made significantly more planning errors on CD switch trials than on normal trials (Fig. 2b, Benedict:  $t_9 = 2.73$ ,  $P = 0.023$ ; Horatio:  $t_9 = 0.56$ ,  $P = 0.96$ ; Oberon:  $t_2 = 1.64$ ,  $P = 0.24$ ; Prospero:  $t_9 = 0.18$ ,  $P = 0.86$ ). With respect to response latencies on correct CD switch trials, two subjects (Benedict and Oberon) took significantly longer to respond to item C on switch trials when compared to normal trials (Fig. 3b, Benedict:  $t_9 = 2.96$ ,  $P = 0.016$ ; Horatio:  $t_9 = 0.87$ ,  $P = 0.41$ ; Oberon:  $t_9 = 3.57$ ,  $P = 0.006$ ; Prospero:  $t_9 = 1.81$ ,  $P = 0.10$ ). It is interesting that for Oberon, the CD switch did not affect trial accuracy or the frequency of planning errors, but did have a significant effect on response latency.



**Fig. 1** Percent correct on normal and switch trials for the BC switch (a), CD switch (b) and DE switch (c). Significant differences in performance are indicated with an asterisk

This suggests that response latency may be a more sensitive measure than either trial accuracy or the percentage of planning errors. Taken together, however, Benedict was the only subject to show a consistent effect across all three measures.

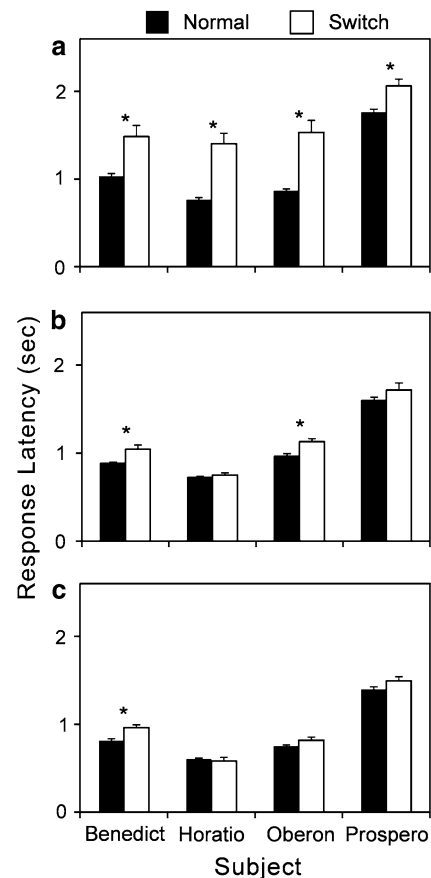
Before we conclude that Benedict was the only subject to plan a response to C at the start of a trial, we must address one issue that may explain our failure to find consistent evidence that any of our other three subjects were planning a response to C. It is possible that our other three subjects failed to display evidence of planning a response to C because they had habituated to the BC switch condition that preceded the CD switch condition. According to this view, our subjects failed to show evidence of planning a response to C in the CD switch condition because during the BC switch condition, they learnt how to respond correctly on switch trials. In order to rule out this possibility, we compared the first five sessions to the last five sessions for each of our three performance measures separately for each subject. If our subjects did habituate to the BC switch condition, then we would expect to find a difference



**Fig. 2** Percent of planning errors on normal and switch trials for the BC switch (a), CD switch (b) and DE switch (c). Significant differences in performance are indicated with an *asterisk*

between the first five sessions and the last five sessions. For no subject was there a significant difference between the first five sessions and the last five sessions for accuracy (Benedict:  $t_4 = 0.22$ ,  $P = 0.84$ ; Horatio:  $t_4 = 0.54$ ,  $P = 0.62$ ; Oberon:  $t_4 = 0.05$ ,  $P = 0.96$ ; Prospero:  $t_4 = 1.34$ ,  $P = 0.25$ ), the percentage of AC planning errors (Benedict:  $t_4 = 0.17$ ,  $P = 0.88$ ; Horatio:  $t_4 = 0.83$ ,  $P = 0.46$ ; Oberon:  $t_4 = 0.44$ ,  $P = 0.68$ ; Prospero:  $t_4 = 1.27$ ,  $P = 0.28$ ), or response latency to B on correct trials (Benedict:  $t_3 = 1.24$ ,  $P = 0.3$ ; Horatio:  $t_2 = 0.36$ ,  $P = 0.75$ ; Oberon:  $t_4 = 1.16$ ,  $P = 0.31$ ; Prospero:  $t_4 = 0.52$ ,  $P = 0.63$ ). The absence of a significant difference between the first five sessions and the last five sessions for any of the three performance measures suggests that our subjects did not habituate to the BC switch and suggests that the absence of any consistent impairment in performance for three of our subjects on the BC switch condition was due to them not planning a response to C at the start of a trial.

On DE switch trials, only one subject's performance (Prospero) was significantly different on switch DE trials than on normal trials (Fig. 1c, Benedict:  $t_9 = 1.00$ ,  $P = 0.34$ ;



**Fig. 3** Latency to the first switch item on normal and switch trials for the BC switch (a), CD switch (b) and DE switch (c). Significant differences in performance are indicated with an *asterisk*

Horatio:  $t_9 = 0.71$ ,  $P = 0.5$ ; Oberon:  $t_9 = 1.17$ ,  $P = 0.27$ ; Prospero:  $t_9 = 2.77$ ,  $P = 0.022$ ). The direction of the difference, however, was in the opposite direction of what we would predict if the difference was due to planning. Prospero actually performed slightly better on switch DE trials than on normal trials (97% correct vs. 100%). In fact, Prospero performed without error on switch DE trials. On DE switch trials, a planning error was defined as correct responses to A, B and C followed by a response to E. Because Prospero made no errors on switch DE trials, we could not make a comparison between switch DE trials and normal trials. For the remaining three subjects, there was no difference in the number of planning errors made on switch DE trials compared to normal trials (Fig. 2c, Benedict:  $t_8 = 0.98$ ,  $P = 0.36$ ; Horatio:  $t_9 = 0.24$ ,  $P = 0.82$ ; Oberon:  $t_3 = 1.41$ ,  $P = 0.26$ ). With respect to response latency, only one subject's (Benedict) response time to D on switch trials was significantly longer than that on normal trials (Fig. 3c, Benedict:  $t_9 = 5.24$ ,  $P = 0.001$ ; Horatio:  $t_9 = 0.57$ ,  $P = 0.58$ ; Oberon:  $t_9 = 1.80$ ,  $P = 0.11$ ; Prospero:  $t_9 = 1.99$ ,  $P = 0.078$ ). Taken together, the absence of a consistent effect across the three measures for



**Table 1** Performance on the mask task for the three-item, four-item and five-item lists

Subject	List length	Percent correct						Total
		Trials	A	B	C	D	E	
Benedict	3	180	100*	70*	100*	–	–	70
	4	660	97*	97*	77*	100*	–	72
	5	2,700	95*	95*	57*	52	100*	27
Horatio	3	240	98*	73*	100*	–	–	72
	4	2,400	97*	88*	53	100*	–	45
	5	–	–	–	–	–	–	–
Oberon	3	1,800	100*	72*	100*	–	–	72
	4	2,400	100*	95*	46	100*	–	43
	5	–	–	–	–	–	–	–
Prospero	3	1,320	98*	71*	100*	–	–	70
	4	2,400	90*	69*	59	100*	–	37
	5	–	–	–	–	–	–	–

Performance that is significantly above chance ( $P < 0.05$ ) is indicated with an asterisk

any subject suggests that subjects were not planning a response to item D at the start of a trial.

### Mask condition

The performance of subjects on the mask condition is shown in Table 1. On the three-item mask, in which items B and C are masked, all subjects reached the criterion of 70% or more correct responses to item B ( $P < 0.05$ , binomial test). On the four-item mask, however, only one subject (Benedict) reached the criterion within the 40 sessions (2,400 trials) on which subjects were trained. As shown in Table 1, Horatio, Oberon and Prospero responded above chance to the first masked item (B) but their performance fell to chance on the third. In contrast, Benedict performed at 97% and 77% to the two critical masked items (B and C) on his criterion session. Due to the failure of Horatio, Oberon and Prospero on the four-item mask condition, only Benedict was run on the final five-item mask phase. Benedict was unable to reach criterion on the five-item mask within the 45 sessions (2,700 trials) he was run for. As shown in Table 1, Benedict continued to respond above chance to the second and third masked item but fell to chance on the fourth.

### Discussion

The results of the present study are very similar to those of Beran et al. (2004). On BC switch trials, all four subjects were less accurate, made a larger number of planning errors and showed longer response latencies to B. This perfor-

mance mirrors that of Beran et al.'s (2004) three chimpanzees and two monkeys. On CD switch trials, only one of our four subjects (Benedict) was less accurate, made a larger number of planning errors and showed longer response latencies to C. Again, when Beran et al. (2004, Experiment 5) ran their subjects on a CD switch condition, they found that only one of their two monkeys made significantly more planning errors on CD switch trials than on normal trials (overall performance and response latencies to C were not reported). On the DE switch trials, there were no consistent effect across the three performance measures for any of our four subjects. Although our switch conditions were modelled after those used by Beran et al. (2004), it is important to note that we would have expected the same results had we switched the positions of nonadjacent items. For example, if we ran a BD switch condition, we would have expected results comparable to those we found with the BC switch condition, and if we ran a CE switch condition, we would have expected results comparable to the CD switch condition.

With respect to the mask condition, the success of our subjects followed exactly what one would predict if performance on the BC, CD and DE switch conditions reflected the degree to which a subject was planning. On the three-item mask, which corresponds to the degree of planning required to show performance deficits on the BC switch, all four subjects were able to reach the criterion of 70% or more correct responses to item B. On the four-item mask, which corresponds to the degree of planning required to show performance deficits on the CD switch, only Benedict was successful. When trained on the five-item mask, Benedict was able to accurately respond to the first two masked items (B and C) but fell to chance on the fourth (D). Benedict's performance is identical to that of a single subject in Beran et al.'s (2004) study. The comparable performance of our subjects and Beran et al.'s (2004) subjects shows clearly that their results cannot be attributed to the use of a joystick.

In the present experiment, we trained subjects on the switch conditions before transferring them to the mask condition. An interesting question is: Had we trained subjects on the mask condition first would our monkeys have displayed greater evidence of planning on the switch conditions? For example, had we trained subjects on the mask condition first would a greater number of subjects have shown consistent effects across the three measures on the CD switch? We think this is unlikely due to the fact that the BC, CD and DE switch conditions perfectly predicted the performance of subjects on the mask condition. Indeed, there was no evidence that on the mask condition, any subject ascended beyond the level of planning they display on the switch condition.

The subjects in the present study, two of Inoue and Matsuzawa's (2007, 2009) subjects, and Beran et al.'s

(2004) subjects, were all only able to respond to one or two masked items. In contrast, four of Inoue and Matsuzawa's (2009) chimpanzees were able to respond to between four and seven masked items. An important question is what separates the majority of subjects that are only able to respond to a few masked item and the subjects able to respond to many masked items? One possibility is that these two groups are using different strategies. Two strategies have been proposed to account for the performance of nonhuman primates on the mask task, a planning strategy and an eidetic strategy. As we noted in the introduction, a planning strategy is one in which a subject scans the array of list items at the start of each trial, encodes each item's spatial position and then plans a sequence of motor responses (Kawai 2001; Ohshima 1997). In contrast, an eidetic strategy is where a subject encodes an image of the item display before the items are masked and then uses that image to guide responding (Haber 1969, 1979; Inoue and Matsuzawa 2007; Jaensch 1930). Because the planning strategy requires a subject to encode information about each item separately, the number of masked items a subject can respond to is constrained by the capacity of working memory. A subject using an eidetic strategy does not face this same limitation because the subject encodes a single image that contains information about all of the list items.

The fact that the subjects in the present study, two of Inoue and Matsuzawa's (2007, 2009) subjects, and Beran et al.'s (2004) subjects, were only able to respond to a one or two masked items suggests to us they were using a planning strategy. In contrast, the performance of Inoue and Matsuzawa's (2009) four best chimpanzees suggests that they were using an eidetic strategy. Of course, we cannot rule out the possibility that the marked differences in performance on the mask task merely reflect differences in working memory capacity. However, we believe that the different strategies interpretation provides a more parsimonious explanation of the data for three reasons. First, the fact that majority of subjects were only able to respond to one or two masked items accords well with studies on the limits of serial recall in nonhuman primates (Barone and Joseph 1989; Botvinick et al. 2009; Devine and Jones 1975; Devine et al. 1979; Funahashi et al. 1997; Ninokura et al. 2003). Second, an analysis of the response time data from our monkeys revealed that as the number of masked items in the list increased, response time to the first item also increased. This relationship is consistent with a planning interpretation as the greater the number of items in the list the more time a subject would need to formulate their response plan. A look at the first-item response latencies for Inoue and Matsuzawa's (2007, online supplemental methods) best performing chimpanzees revealed that, if anything, response time to the first item actually decreased as the number of masked items increased. Obviously, this

finding is not consistent with a planning interpretation. It is important to note that although we attribute the increase in our monkeys' first-item response latency to planning, it is possible that it is due to other factors. For example, as list length increases the number of items, a subject must scan to locate the first item also increases, and this would produce the same increase in first-item response latency that we have attributed to planning.

Our third and final reason why we believe our different strategies interpretation is correct is based on the age of Inoue and Matsuzawa's (2007) chimpanzees. The three best performing chimpanzees, who were able to respond to between four and seven masked items, were all young (7 years of age) while the two chimpanzees only able to respond to a single masked item were adults. This finding corresponds well with studies of eidetic imagery in humans that suggest eidetic imagery is present in some human children but is essentially absent in adults (Haber 1969, 1979; Jaensch 1930).

Recently, studies with adult human subjects have cast doubt on the hypothesis that Inoue and Matsuzawa's (2007) chimpanzees were using an eidetic strategy. These studies have demonstrated that, on variations of the mask task, the performance of adult humans can either match or surpass that of Inoue and Matsuzawa's (2007) chimpanzees (Cook and Wilson 2010; Silberberg and Kearns 2009). Accordingly, Cook and Wilson (2010) have suggested that an appeal to eidetic imagery is unnecessary and that performance on the mask task is easily explained by processes, such as iconic memory or the use of a visual after image. An additional issue raised by these studies is the role of experience. Inoue and Matsuzawa's (2007) compared the performance of their best chimpanzees to a group of human participants and found the chimpanzees consistently outperformed them. Inoue and Matsuzawa's (2007) chimpanzees, however, were markedly more experienced than their human participants. Cook and Wilson (2010) have now shown that with comparable levels of training, human participants actually outperform Inoue and Matsuzawa's (2007) chimpanzees.

## Concluding remarks

The present study represents just one step in addressing the many questions that are raised by the marked differences in performance observed when nonhuman primates are trained on the mask task. The critical issues that must still be answered are—(1) whether eidetic imagery, iconic memory, or a visual after image, underlies high levels of performance on the mask task and (2) whether age and/or experience play a critical role in performance on the mask task.

**Acknowledgments** This study was supported in part by a Fulbright Ministry of Research, Science and Technology award to D. Scarf and a University of Otago Research Committee Postgraduate Publishing Bursary to D. Scarf.

## References

- Barone P, Joseph JP (1989) Prefrontal cortex and spatial sequencing in macaque monkey. *Exp Brain Res* 78:447–464
- Beran MJ, Pate JL, Washburn DA, Rumbaugh DM (2004) Sequential responding and planning in chimpanzees (*Pan troglodytes*) and rhesus macaques (*Macaca mulatta*). *J Exp Psychol Anim B* 30:203–212
- Biro D, Matsuzawa T (1999) Numerical ordering in a chimpanzee (*Pan troglodytes*): planning, executing, monitoring. *J Comp Psychol* 113:178–185
- Bitterman ME (1965) The evolution of intelligence. *Sci Am* 212: 92–100
- Botvinick MM, Wang J, Cowan E, Roy S, Bastianen C, Mayo JP, Houk JC (2009) An analysis of immediate serial recall performance in a macaque. *Anim Cogn* 12:671–678
- Brannon EM, Terrace HS (2000) Representation of the numerosities 1–9 by rhesus macaques (*Macaca mulatta*). *J Exp Anal Behav* 26:31–49
- Brannon EM, Cantlon JF, Terrace HS (2006) The role of reference points in ordinal numerical comparisons by rhesus macaques (*Macaca mulatta*). *J Exp Anal Behav* 32:120–134
- Cook P, Wilson M (2010) Do young chimpanzees have extraordinary working memory? *Psychon B Rev* 17:599–600
- Correia SPC, Dickinson A, Clayton NS (2007) Western scrub-jays anticipate future needs independently of their current motivational state. *Curr Biol* 17:856–861
- Devine JV, Jones LC (1975) Matching-to-successive samples: a multiple unit memory task with rhesus monkeys. *Behav Res Meth Instr* 7:438–440
- Devine JV, Burke MW, Rohack JJ (1979) Stimulus similarity and order as factors in visual short-term memory in nonhuman primates. *J Exp Anal Behav* 5:335–354
- Funahashi S, Inoue M, Kubota K (1997) Delay-period activity in the primate prefrontal cortex encoding multiple spatial positions and their order of presentation. *Behav Brain Res* 84:203–223
- Haber RN (1969) Eidetic images. *Sci Am* 220:36–44
- Haber RN (1979) Twenty years of haunting eidetic imagery: where's the ghost? *Behav Brain Sci* 2:583–629
- Inoue S, Matsuzawa T (2007) Working memory of numerals in chimpanzees. *Curr Biol* 17:1004–1005
- Inoue S, Matsuzawa T (2009) Acquisition and memory of sequence order in young and adult chimpanzees (*Pan troglodytes*). *Anim Cogn* 12:S59–S69
- Iwai E, Yaginuma S, Mishkin M (1986) Acquisition of discrimination learning of patterns identical in configuration in macaques (*Macaca mulatta* and *M. fuscata*). *J Comp Psychol* 100:30–36
- Jaensch ER (1930) Eidetic imagery and typological methods of investigation. Kegan Paul, Trench, Trubner & Co., London
- Kawai N (2001) Ordering and planning in sequential responding to Arabic numerals by a chimpanzee. *Psychologia* 44:60–69
- Kawai N, Matsuzawa T (2000) Numeric memory span in a chimpanzee. *Nature* 403:39–40
- Kornell N, Terrace HS (2007) The generation effect in monkeys. *Psychol Sci* 18:682–685
- Meyer DR, Treichler FR, Meyer PM (1965) Discrete-trial training techniques and stimulus variables. In: Schrier AM, Harlow HF, Stollnitz F (eds) *Behavior of nonhuman primates*, vol 1. Academic Press, New York
- Mulcahy NJ, Call J (2006) Apes save tools for future use. *Science* 312:1038–1040
- Ninokura Y, Mushiaki H, Tanji J (2003) Representation of the temporal order of visual objects in the primate lateral prefrontal cortex. *J Neurophysiol* 89:2868–2873
- Ohshiba N (1997) Memorization of serial items by Japanese monkeys, a chimpanzee, and humans. *Jpn Psychol Res* 39:236–252
- Osvath M (2009) Spontaneous planning for future stone throwing by a male chimpanzee. *Curr Biol* 19:R190–R191
- Osvath M, Osvath H (2008) Chimpanzee (*Pan troglodytes*) and orangutan (*Pongo abelii*) forethought: self-control and pre-experience in the face of future tool use. *Anim Cogn* 11:661–674
- Paxton R, Hampton RR (2009) Tests of planning and the Bischof-Köhler hypothesis in rhesus monkeys (*Macaca mulatta*). *Behav Process* 80:238–246
- Raby CR, Alexis DM, Dickinson A, Clayton NS (2007) Planning for the future by western scrub-jays. *Nature* 445:919–921
- Scarf D, Colombo M (2009) Eye movements during list execution reveal no planning in monkeys (*Macaca fascicularis*). *J Exp Anal Behav* 35:587–592
- Scarf D, Colombo M (2010) The formation and execution of sequential plans in pigeons (*Columba livia*). *Behav Process* 83:179–182
- Silberberg A, Kearns D (2009) Memory for the order of briefly presented numerals in humans as a function of practice. *Anim Cogn* 12:405–407
- Subiaul F, Cantlon JF, Holloway RL, Terrace HS (2004) Cognitive imitation in rhesus macaques. *Science* 305:407–410
- Terrace HS (1984) Simultaneous chaining: the problem it poses for traditional chaining theory. In: Herrnstein RJ, Wagner A (eds) *Quantitative analyses of behavior*. Ballinger Publishing Co., Cambridge
- Terrace HS, Son LK, Brannon EM (2003) Serial expertise of rhesus macaques. *Psychol Sci* 14:66–73
- Yaginuma S, Iwai E (1986) Effect of small cue-response separation on pattern discrimination in macaques (*Macaca fuscata* and *M. mulatta*). *J Comp Psychol* 100:137–142