



Published in final edited form as:

Anim Cogn. 2014 May ; 17(3): 619–631. doi:10.1007/s10071-013-0693-0.

Rapid cognitive flexibility of rhesus macaques performing psychophysical task-switching

Ema Avdagic,

Department of Psychology, Columbia University, New York, NY, USA

Greg Jensen,

Department of Psychology, Columbia University, New York, NY, USA

Drew Altschul, and

Department of Psychology, University of Edinburgh, Edinburgh, Scotland, UK

Herbert S. Terrace

Department of Psychology, Columbia University, New York, NY, USA

Herbert S. Terrace: terrace@columbia.edu

Abstract

Three rhesus monkeys (*Macaca mulatta*) performed a simultaneous chaining task in which stimuli had to be sorted according to their visual properties. Each stimulus could vary independently along two dimensions (luminosity and radius), and a cue indicating which dimension to sort by was random trial to trial. These rapid and unpredictable changes constitute a task-switching paradigm, in which subjects must encode task demands and shift to whichever task-set is presently activated. In contrast to the widely reported task-switching delay observed in human studies, our subjects show no appreciable reduction in reaction times following a switch in the task requirements. Also, in contrast to the results of studies on human subjects, monkeys experienced enduring interference from trial-irrelevant stimulus features, even after exhaustive training. These results are consistent with a small but growing body of evidence that task-switching in rhesus macaques differs in basic ways from the pattern of behavior reported in studies of human cognition. Given the importance of task-switching paradigms in cognitive and clinical assessment, and the frequency with which corresponding animal models rely on non-human primates, understanding these differences in behavior is essential to the comparative study of cognitive impairment.

Keywords

Comparative cognition; Cognitive assessment; Mental set; Task-switching; Rhesus macaques

Introduction

The complex coordination of sensory feedback, motor action, information processing, and memory that underlies the execution of a skill has been described as a “mental set” (Jersild

1927), which is a representation of the relevant rules and action frameworks associated with a given task. Embedding task requirements into a mental representation follows naturally from the use of hierarchical control structures (Kleinsorge and Heuer 1999) and is an important aspect of organizing motor actions into functional behaviors (Rogers and Monsell 1995). Although task representations facilitate the efficient deployment of response strategies, “task-switching” (whereby subjects must change which representation is in active use) results in a corresponding drop in efficiency (Mayr and Kliegl 2000; Monsell 2003).

Healthy human adults consistently slow down and make less accurate responses during a brief transitional period following a switch between tasks (Koch 2001). These switches occur so rapidly as to give the subjective feeling of performing more than one task at once, but the costs of switching remain evident in impaired task performance (Pashler 1994; Ophir et al. 2009). The intransigence of these task-switching costs suggests that they are a necessary by-product of the manner in which executive control systems are organized (Pashler 2000). Furthermore, various forms of cognitive impairment magnify these costs, whether they arise from aging (e.g., Clapp et al. 2011), brain injury (Mecklinger et al. 1999), or other clinical conditions (Gu et al. 2008). It is therefore important to understand the mechanisms by which the evident task-switching bottleneck arises.

When stimuli unambiguously signal the task in which a subject must engage, task-switching engenders only a small delay, particularly if past responses need not be recalled (Spector and Biederman 1976; Allport et al. 1994). In such cases, subjects form a “task-set,” a superordinate representation made up of several well-learned tasks. When the tasks required are known and are signaled by cues, task-sets permit subjects to shift between tasks rapidly (Monsell et al. 2000). However, when task cues are subject to interference, selective attention is required to switch tasks effectively, resulting in an increased cognitive cost (Rogers and Monsell 1995; Logan and Gordon 2001). This is particularly important when subjects must distinguish between ‘cues,’ which are necessary to succeed at a task, and ‘primes,’ which may be informative but are not required. When the cues for one task are mixed with contradictory primes, selective attention is needed to focus on relevant cues and disregard mismatched primes (Sudevan and Taylor 1987).

Because of their clinical implications, most studies of the relationship between selective attention and task-switching have focused on human cognition. Recently, however, attempts have been made to establish an animal model of selective attention and task-switching, using rhesus monkeys. Such a model can inform cognitive development (Weed et al. 2008) and age-related cognitive decline (Moore et al. 2003; Zeamer et al. 2011), both domains in which cognitive assessment is crucial.

Experiments on non-human primates have two important independent benefits. They not only provide opportunities for unit recording, but they also avoid the confound of verbal processing. The ways in which the human capacity for language impacts cognitive performance are often subtle, manifesting even in tasks that can be replicated by non-human subjects (Prado et al. 2013). This risk is particularly salient given the verbal component in many human assessment procedures, such as the Stroop task (Washburn 1994). Although

verbal mechanisms are important, it is also important to dissociate them from non-verbal mechanisms, as can be accomplished with animal models.

There are, however, several aspects of the comparative literature on task shifting that make direct comparison difficult. For example, the importance of studying frequent shifts between tasks is recognized in the human literature, both for its theoretical implications (Pashler 2000) and its applicability to scenarios of deficient cognitive control (Clapp et al. 2011). Despite this, most studies of macaque task-switching initiate changes in the task on the basis of performance criteria, allowing subjects to persist until they ‘get it right’ (e.g., Moore et al. 2003; Weed et al. 2008). Typically, these studies do not analyze the reaction-time delay arising from task-switching. In one of the few studies to do so (Stoet and Snyder 2003a, b), rhesus monkeys performed a task that randomly switched from one requirement to another with a 50 % probability on each trial. These monkeys displayed minimal delay when switching between tasks (unlike human participants). However, they also showed considerable interference from distracting primes, a pattern of error that persisted despite extended training (also unlike humans).

As discussed earlier, retrieval of task-set representations and the suppression of irrelevant stimuli are important consequences in task-switching scenarios with human subjects. Thus, Stoet and Snyder’s results raise the possibility of a qualitative species difference in how memory and attention interact in human and non-human subjects. For example, it may be that Stoet and Snyder’s unusual experimental procedure included unaccounted-for, laboratory-specific, or task-specific confounds. In order to examine these issues with a fresh perspective, we performed an experiment with a different task that retained three key features of Stoet and Snyder’s experiment: extensive training (to ensure robust, well-encoded task representations), frequent and unpredictable task-switching, and substantial interference between stimulus properties that acted as (necessary) cues and those that acted as (potentially distracting) primes.

In our experiment, rhesus macaques were presented with groups of stimuli that varied along two psychophysical dimensions. Subjects were required to sort these stimuli according to one of the dimensions using a simultaneous chaining paradigm (Terrace 1984, 2005). During each trial, between 3 and 6 stimuli were simultaneously presented, each of which had a different radius and a different luminosity. On the basis of a context cue (the background color), subjects were required to respond to the stimuli in a particular order according to a particular psychophysical dimension, while disregarding the other dimension. Subjects were trained incrementally, but were eventually required to switch between the two orderings solely on the basis of background cues that varied randomly from trial to trial.

Methods

Subjects

Subjects were 3 male rhesus monkeys (*Macaca mulatta*), Lashley, MacDuff, and Oberon. At the start of the experiment, the subjects were 13, 15, and 15 years old, respectively. Throughout the duration of the experiment, they were housed at the New York State Psychiatric Institute. All three subjects had extensive experience with the apparatus (Terrace

et al. 2003; Kornell et al. 2007). Subjects were fed a mixed diet of primate chow and mixed fruit on a daily basis, immediately following testing. Water was available ad libitum.

Apparatus

The experiments were conducted in sound-attenuated booths (127 cm high \times 97 cm wide \times 97 cm deep) that each contained an operant chamber made of Plexiglas and stainless steel (53 cm \times 48 cm \times 53 cm). Each booth contained a pellet dispenser (Med Associates), used for reinforcing the subjects with a single 190-mg banana pellet (Bioserv) after every correct trial, and a video camera for recording the performance on the task. Within the operant chamber, subjects had access to a touch-sensitive 15-inch (38 cm) computer monitor, which served to both present stimuli and record responses. All experimental tasks were programmed using Real Studio (formerly RealBASIC), controlled by an iMac computer (model: MA710xx/A). Unless otherwise noted, the apparatus was identical to that reported in earlier experiments on monkey cognition (Subiaul et al. 2004).

Procedure

Subjects responded under a variation of the simultaneous chaining paradigm (Terrace 1984, 2005), hereafter identified as the “SimChain” task. Subjects were presented simultaneously with a set of stimuli, each consisting of a single filled circle on a rectangular white backdrop. Subjects were required to respond to each of the stimuli in the correct order, with a food pellet delivered immediately upon correctly selecting the last item. If the items were selected in an incorrect order, the screen went black and subjects experienced a timeout. In both cases (reward vs. blackout), the inter-trial interval was 4 s. Thus, if four stimuli were presented, a subject would need to select each item once. Any out-of-order touches prematurely ended the trial. Each experimental session consisted of 50 trials.

The stimuli differed according to two dimensions: *luminosity*, ranging from black (RGB 0, 0, 0) to light gray (RGB 220, 220, 220), and *radius*, ranging from 10 pixels (0.15 cm) to 70 pixels (1.05 cm). Subjects were required to touch items in a sequence that was determined by the item’s ordinal position along these dimensions. During early training, one dimension was held constant while the other was permitted to vary. In later stages, both dimensions were allowed to vary independently, and subjects were required to order the stimuli according to one dimension or the other, as indicated by the background color of the screen. In all phases of the experiment, a red background was a contextual cue for ordering stimuli on the basis of the stimulus radius. A blue background signaled that stimuli should be ordered on the basis of stimulus luminosity (ranging from light gray to black). Lashley and Oberon were required to select stimuli in an ascending order, and MacDuff was required to select stimuli in a descending order.

Subjects were first trained to order stimuli on the basis of radius, with luminosity held constant (Training Phase 1). Training began with 3-item lists. List length was increased to 4-item, 5-item, and ultimately 6-item lists, advancing when an 80 % overall performance criterion was met. In the final stage of radius training, subjects were received a mix of SimChain lists, ranging in length from 3 to 6 items. Radius training was considered complete when subjects were able to successfully complete 50 % of 6-item lists in a session.

Subjects were then trained to order stimuli according to their luminosity, while radius was held constant (Training Phase 2). As in radius training, subjects began by responding to 3-item lists, with 4- and 5-item lists introduced as performance reached the 80 % overall performance criterion.

Once subjects displayed high levels of accuracy at ordering stimuli according to each dimension in isolation, they were trained on both list types (luminosity and radius) during the same session (Training Phase 3). During this phase, the “target” dimension was varied but the “dis-tractor” dimension was not. Accordingly, subjects were required to order list items on only one dimension at a time. As noted previously, the background color provided a context cue indicating the relevant dimension (red for radius, blue for luminosity). Subjects responded to a mix of 3-, 4-, 5-, and 6-item lists in this phase of training, again working until an 80 % performance criterion was met.

In Experimental Phase 1 (biased ordering), subjects were presented with a mix of 3- and 4-item lists that varied along both dimensions. Thus, every stimulus had a different radius and a different luminosity. This phase was “biased” because the distractor dimension varied by smaller increments than the target dimension, thereby making the target dimension more salient. Figure 1a shows a biased luminosity trial, and Fig. 1b, a biased radius trial. The bias was expected to improve performance, because subjects had two sources of information about the task: the background color and which dimension contained more extreme (and therefore more salient) differences between list items. This phase lasted 36 sessions (1,800 trials).

In Experimental Phase 2 (unbiased ordering), subjects were presented with a mixture of 3- and 4-item lists that varied along both dimensions. During unbiased ordering, the background color was the only contextual indicator of the correct dimension by which to order list items. As such, the stimuli themselves provided no differential emphasis to either dimension. Figure 1c, d shows trials with an exemplar set of stimuli, showing how a different response order was cued by the background color. Unlike Fig. 1c, d, however, the positions and stimulus particulars were randomized from trial to trial. Experimental Phase 2 was expected to be the more difficult than Experimental Phase 1 because of interference from the distractor dimension. This phase lasted 24 sessions (1,200 trials) (Table 1).

Results

Subjects' accuracy on 3-item lists routinely exceeded 75 % correct. Accordingly, we will only focus on their performance on 4-item lists, to avoid ceiling effects.

Experimental Phase 1: biased ordering

Across the 36 sessions of Experimental Phase 1, subjects were trained on 900 4-item lists, split randomly between the luminosity and radius conditions. As shown in Fig. 2, all subjects exceeded 60 % accuracy to individual list items. Further, they completed at least 30 % of the full 4-item lists correctly (compared to chance performance of 1.7 %).

In order to assess the pattern of errors being made by participants, we focused on the accuracy of the first response, as a function of its rank order according to each of the two stimulus dimensions. Figure 3 presents the distribution of first responses made to 4-item lists in Experimental Phase 1, with the diameter of each circle corresponding to the frequency of each choice. The zone between the dashed lines represents the correct responses, while the zone between the dotted lines would be the correct responses according to the distractor dimension. Thus, a response made *outside* the dashed lines, but *inside* the dotted lines, can be interpreted as arising from confusion about the appropriate dimension.

As can be seen in Fig. 3, subjects' error patterns showed that they had no difficulty ordering items according to the target dimension during Experimental Phase 1. Relatively few erroneous responses to the distractor dimension were made.

To obtain a nonparametric measure of response uncertainty, we used Shannon information (Jensen et al. 2013b), measured in bits and commonly identified as H . The equation for H and for "information explained" I is as follows:

$$H = - \sum_{i=1}^4 p_i \cdot \log_2(p_i) \quad (1)$$

$$I = H_{\max} - H$$

Here, p_i is the probability of a response to category i . For example, Lashley's probability of selecting the lowest ranked luminosity item on a luminosity trial would be $(106 + 98 + 90 + 81)/475 = 375/475 = 0.79$, while on a radius trial, it would only be $(85 + 36 + 11 + 4)/425 = 136/425 = 0.32$. The minimum value H_{\min} is 0.0, which occurs when only a single alternative is selected; the maximum value H_{\max} is 2.0, which occurs when all four alternatives are selected with equal probability. In our data, lower H and higher I correspond to more accurate responding.

Table 2 shows that, for each subject, there was more information explained (I) for the target dimension than there was for the distractor dimension, a result consistent with sufficiently effective distractor suppression to succeed on most trials. Nevertheless, there were still signs of interaction between the two dimensions, which suggests interference.

In the event that luminosity and radius were independent from one another, the rank of a stimulus on one dimension should be independent of its rank on the other dimension. This null hypothesis was tested by calculating Spearman's rank correlation coefficient (Myers and Well 2003), denoted by r_s , for the degree of interrelation between the two dimensions for a subject's first response (thus, each response had two ranks associated with it, both ranging from 1 to 4). All subjects displayed mild to moderate rank correlations on both luminosity trials (Lashley $r_s = 0.151$; MacDuff $r_s = 0.350$; Oberon $r_s = 0.302$) and radius trials (Lashley $r_s = 0.148$; Mac-Duff $r_s = 0.238$; Oberon $r_s = 0.257$), with higher correlations in cases in which the difference ($I_{\text{target}} - I_{\text{distractor}}$) was closer to zero.

In order to test whether these rank correlations were significantly different from 0.0, we used the Mantel-Haenszel statistic (Mantel and Haenszel 1959), denoted by ψ_{MH} . The

Mantel–Haenszel statistic is a repeated-measures test for categorical independence and was used to minimize assumptions about the (potentially nonlinear) effects of stimulus rank; it follows a chi-squared distribution with one degree of freedom. We found a significant level of interaction was obtained from all subjects on luminosity trials (Lashley $\psi_{MH} = 10.81$, $p < .002$; MacDuff $\psi_{MH} = 59.02$, $p < .001$; Oberon $\psi_{MH} = 49.56$, $p < .001$) and radius trials (Lashley $\psi_{MH} = 9.29$, $p < .003$; MacDuff $\psi_{MH} = 23.68$, $p < .001$; Oberon $\psi_{MH} = 23.45$, $p < .001$).

An analysis of reaction times was performed for each subject to test for a task-switching effect. Figure 4 plots log(reaction time), split according to that response's list position, and whether the corresponding list came after a switch in task demands (black) or not (white). The primary effect, which was consistent across subjects, was that early list items elicited longer reaction times. This is consistent with a process-of-elimination visual search. There were, however, no consistent effects of task-switching.

An ANOVA was performed for each subject independently to quantify the effect of task-switching and list position on reaction times. A significant main effect of list position was unambiguously observed in all subjects (Lashley: $F(3,2482) > 152.3$, $p < .0001$; MacDuff: $F(3,2066) > 104.4$, $p < .0001$; Oberon: $F(3,2305) > 130.2$, $p < .0001$). A significant main effect for task-switching was only observed in MacDuff [$F(1,2066) > 5.99$, $p < .02$] and Oberon [$F(1,2305) > 5.85$, $p < .02$], but these effects were quite small, a slowing down of 7 and 6 % for each subject, respectively. The interaction between task-switching and list position was not significant.

The ANOVA above was supplemented with an analysis of effect size (Hentschke and Stüttgen 2011). The *omega squared* (ω^2) statistic for list position was considerable (Lashley $\omega^2 = 0.283$; MacDuff $\omega^2 = 0.226$; Oberon $\omega^2 = 0.188$), whereas the omega squared for task-switching was infinitesimal, even in cases where it was statistically significant (Lashley $\omega^2 = 0.0002$; MacDuff $\omega^2 = 0.0019$; Oberon $\omega^2 = 0.0017$).

Experimental Phase 2: unbiased ordering

During the 24 sessions of Experimental Phase 2, subjects responded on approximately 1,100 4-item lists, split randomly between the luminosity and radius conditions.

As shown in Fig. 5, subjects performed consistently above chance, but were considerably more likely to make errors than in Experimental Phase 1. Additionally, subjects had considerably more difficulty with radius trials than with luminosity trials, a reversal of the pattern observed in the biased ordering phase.

The pattern of errors displayed in Fig. 6 suggests that, on radius trials, subjects nevertheless made a substantial number of errors favoring the luminosity ordering. Mac-Duff's performance was particularly weak in this regard, with most responding clustering in the lower left-hand corner, rather than being evenly spread along the target dimension.

Table 3 shows the information explained (I) for target and distractor dimensions, in which subjects had considerably greater difficulty making the appropriate discrimination. Both

Lashley and Oberon favored the luminosity dimension even on radius trials, suggesting a biased response strategy. While MacDuff was unbiased and favored target over distractor in both conditions, his margin of error was considerable. These results suggest that the background cue exerted no more than moderate stimulus control in Phase 2.

This interference was only somewhat evident in comparisons with rank correlation on luminosity trials (Lashley $r_s = 0.108$; MacDuff $r_s = 0.346$; Oberon $r_s = 0.271$), likely because two subjects displayed a luminosity bias. The evidence for interference was clear in radius trials (Lashley $r_s = 0.267$; MacDuff $r_s = 0.336$; Oberon $r_s = 0.314$). In all cases, these interactions were significant according to the Mantel–Haenszel statistic in luminosity trials (Lashley $\psi_{MH} = 6.43$, $p < .02$; MacDuff $\psi_{MH} = 67.45$, $p < .001$; Oberon $\psi_{MH} = 40.73$, $p < .001$) and radius trials (Lashley $\psi_{MH} = 39.46$, $p < .001$; Mac-Duff $\psi_{MH} = 60.08$, $p < .001$; Oberon $\psi_{MH} = 55.45$, $p < .001$).

As in Phase 1, an analysis of reaction times was performed for each subject. Figure 7 plots $\log(\text{reaction time})$, split according to that response's list position, but split according to whether the corresponding list came after a switch in task demands (black) or not (white). As in Phase 1, there was a consistent effect of list position, but not of task-switching. Notably, Oberon appeared to respond marginally *faster* following a switch, unlike Phase 1.

An ANOVA testing the effect of task-switching and list position on reaction times was performed for each subject independently. The significant main effect of list position was even stronger than in Phase 1 (Lashley: $F(3,3293) > 460.9$, $p < .0001$; MacDuff: $F(3,2778) > 175.0$, $p < .0001$; Oberon: $F(3,3116) > 147.0$, $p < .0001$). However, only Oberon showed a significant main effect for task-switching [$F(1,3116) > 5.16$, $p < .03$], and the direction of the effect was for responding 6 % *faster* when switching tasks, rather than slower. No other effects or interactions were significant.

Effect sizes in Phase 2 remained substantive with respect to list position in all subjects (Lashley $\omega^2 = 0.447$; MacDuff $\omega^2 = 0.264$; Oberon $\omega^2 = 0.225$) and were still very small with respect to task-switching (Lashley $\omega^2 = 0.0002$; MacDuff $\omega^2 = 0.0003$; Oberon $\omega^2 = 0.0010$).

Discussion

During both phases of this experiment, monkeys were able to use an item's psychophysical properties to order ambiguously defined stimuli with substantially greater accuracy than chance. In each instance, subjects' performance was significantly influenced by an item's ordinal position on the relevant psychophysical dimension. Task-switching from trial to trial did not yield a consistent delay in response onset. Such delays had very small effect sizes when detected, and most did not differ significantly from zero. Subjects nevertheless made consistent errors when the distractor dimension conflicted with the target dimension. This interference persisted despite prolonged training.

The effects of task-switching manipulations are frequently identified as “congruency effects,” particularly in studies of human cognition (e.g., Mayr and Kliegl 2000). Because the principal finding of human task-switching studies has been an increase in reaction time

rather than an increase in errors driven by distracting primes, the term “congruency effect” is often assumed to be measured in milliseconds. Although the patterns of error we observed could be labeled as congruency effects, we have avoided doing so for two reasons. The first is to minimize confusion with respect to our reaction-time results. The second is that we interpret the errors in Figs. 3 and 6 as being a function of the salience of stimulus properties as well as the congruency between a response alternative and the task demands. Thus, we favor the term “interference” when discussing these effects.

According to a Spearman’s rank correlation, all subjects displayed at least some interference between the psychophysical dimensions on which they were trained. The interference was slightly higher in the “unbiased ordering” phase, during which no secondary cues were available to assist in identifying the dimension by which to order the relevant stimuli. This interference was not equal. Our analysis of Shannon Information (Tables 2, 3) showed that Lashley and Oberon favored the luminosity dimension over the radius dimension in all conditions. Macduff, however, did not display this bias. Given that this overshadowing preference was not consistent across subjects, it is unclear whether the order of training was a factor. Importantly, all subjects, regardless of which dimension they preferred, modulated their response on the basis of the context cue. Despite their strong biases, they could successfully switch tasks when required to do so.

Most animal studies of task-switching have used simple tasks that were based on single responses. Our SimChain paradigm required multiple responses to cope with large amounts of simultaneously presented information. Because floor and ceiling effects can make it difficult to evaluate comparative claims, we were encouraged by the intermediate performance of our subjects (responding above chance but consistently showing signs of interference). Because natural scenarios are rarely interference free, we consider the relative difficulty of the task to be an experimental strength. Furthermore, despite the difficulty of the tasks, subjects did not reliably show a delay when switching their attention to the relevant stimulus property. Although subjects displayed interference effects from the distractor dimension, these were never sufficient to overwhelm their ability to switch when required to by the background cue.

Our results agree with those of Stoet and Snyder (2003a, b), who were the first to identify a low-delay, high-interference pattern of performance in rhesus macaques. In contrast to their human subjects, their primate subjects did not display any switching delay. Even when human subjects were given extensive training on the experimental task consisting of tens of thousands of trials, human task-switching costs did not abate (Stoet and Snyder 2007). However, whereas humans could suppress distracting primes and focus only on task-relevant cues, Stoet and Snyder’s primate subjects never overcame interference effects. On the basis of these findings, they argued that although macaques provide a suitable animal model for some forms of cognitive processing, such as early cognitive development (Gómez 2005), species-specific effects must qualify the translational application of comparative studies.

Caselli and Chelazzi (2011) presented a dissenting view, contesting Stoet and Snyder’s general result and arguing that although macaques consistently display task interference

effects, they also display the task-switching delay characteristic of human performance. However, their analysis of reaction times is suspect for several reasons. The first relates to differential training: Stoet and Snyder's subjects received approximately 100,000 trials of training before formal data collection began (2003a), and the authors performed their statistical tests on approximately 1,000 "critical trials" in each experiment (2003b), as can be inferred from the degrees of freedom in their F tests. Contrastingly, Caselli and Chelazzi (who do not report the number of trials performed) only provided subjects with "several training sessions" of undetermined duration and based their ANOVAs not on individual reaction times, but on mean reaction times per session. Because reaction times rarely follow a normal distribution, ANOVAs of sample means are an inappropriate statistical test that can be powerfully influenced by a handful of long response latencies (Whelan 2008). Although both studies performed imperfect analyses, Stoet and Snyder's results are far more likely to be trustworthy because they did not perform an intermediary averaging operation. Because Stoet and Snyder based their inference on the full data, and not on derived descriptive statistics, the measurement error associated with their inferential tests had greater opportunity to converge on Gaussian as a result of the central limit theorem. Our own analysis used log-transformed reaction times (whose distributions were approximately Gaussian), and our results were consistent with Stoet and Snyder's analysis. Whether these differences are due to insufficient training or to inappropriate analysis, the reaction-time results reported by Caselli and Chelazzi should be taken with a grain of salt.

An important contrast between the present study and that of Caselli and Chelazzi is that the present study permitted subjects to respond at will, without a tight time constraint. One hypothesis is that this permitted subjects to "prepare" for each trial, and this preparation explains the lack of a consistent switching penalty on reaction times. We consider this unlikely, because our procedure did not include a "trial-initiating" response (such as the fixation point used in eye-tracking studies), and thus, the reaction times for first responses depicted in Figs. 4 and 7 are representative of the inter-trial intervals (plus the 4-s interval following all trials). Even in the cases where the differences reached statistical significance, the effect size for switch versus non-switch trials was insubstantial. So far as we are able to tell, subjects responded at a fairly steady rate throughout both experimental phases, whether or not a trial included a switch in task demands.

Dreisbach et al. (2007) suggest that discrepancies in task-switching results may hinge on the degree to which experimental tasks can be solved by stimulus-response association, rather than by cognitive representations. To support their claim, they cited an elegant series of experiments in which a simple task-switching paradigm was learned by human participants with explicit verbal instruction regarding task rules introduced at different times (or not at all). Participants who received verbal instruction from the outset showed a robust and enduring task-switching cost, while those who received no instruction (learning only by performing the task) did not. Crucially, a third group of participants received verbal instructions halfway through the experiment, and this group only showed a task-switching cost *after* the introduction of the instructions and *despite* later reporting that they did not consciously make use of the instructions provided. This suggests that verbal instructions

were implicitly encoded and that human stimulus–response learning (a) bypasses task-switching costs and (b) is superseded by verbally conveyed information.

It is important to distinguish stimulus–response learning *in principle* from the associative framework of reinforcement learning, because the latter is not able to account for many kinds of response sequencing (Lashley 1951). One of the advantages of the SimChain paradigm is that it cannot be simply mapped onto S–R associations (Jensen et al. 2013a). Even if it could, the set of mappings would be massive and switching costs on a SimChain task should be similarly large (Dixon 1981). The absence of a switching cost in rhesus macaques points to a procedural form of information processing, and these results are consistent with human performance (with respect to task-switching delays) only when verbal instruction is entirely omitted. Thus, although humans and macaques appear to share many basic processes, this pattern of results suggests additional processing that is specific to verbally instructed humans.

Postponement of central processing, or queuing, has been proposed as being the most important source of inference in task-switching (Pashler 2000). Other evidence collected using the SimChain paradigm suggests that macaques rarely plan their response more than one response in advance (Scarf et al. 2011), consistent with an earlier finding suggesting limited planning in both macaques and chimpanzees (Beran et al. 2004). These results support the view that non-human primates engage in cognitive tasks with a minimal reliance on queuing processes.

Comparative cognitive flexibility in a clinical context

Task-switching has long been a hallmark of cognitive assessment in human participants. A prominent example is the Stroop test, in which the relevant cue (the color in which a word is printed) suffers from powerful interferences from a highly trained but mismatched prime (the semantic meaning of the word) (for review, see MacLeod 1991). Another popular procedure is the Wisconsin Card Sorting Test (WCST), which requires participants to consider stimuli that differ along several dimensions (shape, color, etc.) and infer sorting rules based on a changing schedule of reinforcement (Berg 1948). Both tasks have been used extensively in applied contexts, whether studying cognition in healthy and aging populations (Bryan and Luszcz 2000) or in acute psychiatric illness (Rossi et al. 1997).

The validity of direct animal/human comparisons on such tasks is complicated by the difficulty of translating established tasks into an animal paradigm. For example, because the WCST cannot be administered in its canonical form to a non-human, Moore et al. (2002, 2003, 2005) developed a modified version of the WCST called the “Conceptual Set-Shifting Task” (CSST). The Stroop test is even more difficult to adapt given its linguistic requirements, and comparable animal studies focus on psychophysical competition between cues (Behar 1974; Washburn 1994).

To date, the only direct comparison of set-shifting in human and macaque fMRI made use of an adapted Wisconsin Card Sorting Test (Nakahara et al. 2002), with very similar patterns of activation observed in both. This result is consistent with subsequent studies emphasizing the comparative similarity of macaques to humans (e.g. Moore et al. 2002, 2003, 2005;

Weed et al. 2008), particularly with respect to their lack of reaction-time analysis. However, comparing the neurological systems responsible for these cognitive processes is complicated by the methodological disconnect between human and non-human neuroscience (the former largely employing fMRI and the latter largely employing invasive electrophysiology). However, functional imaging of awake macaques supports the hypothesis that they possess a frontoparietal network associated with complex attentional processes (Stoewer et al. 2010). The observed pattern of activation appears similar to a well-documented human network associated with cognitive control and differentially active in cases of task-switching (Dreher and Grafman 2003).

Several candidates have been proposed as executive bottlenecks in human imaging studies of task-switching, including posterior lateral prefrontal-cortex (Dux et al. 2006) and left superior parietal cortex (Braver et al. 2003). Although prefrontal regions have been implicated in behavioral inhibition in rhesus macaques (Sakagami et al. 2006) and parietal neurons have similarly been implicated in set shifting (Kamigaki et al. 2009), it is not yet known whether these regions function in an analogous fashion. Differential function in one or more of these regions, however, could account for both (1) the consistency of human/macaque behavior in the absence of verbal instruction and (2) the human discrepancy induced by instructions. Given our results, and those reported by Stoet and Snyder (2003a, b, 2007), we suggest that the circumstances in which these species can reliably be compared may be limited to specific cognitive scenarios in which verbal processing plays a minimally confounding role.

Our results raise several considerations for future comparative work. First, it is essential that human and non-human performance be compared under conditions that are as similar as possible, particularly with respect to task instructions. As Dreisbach et al. (2007) point out, the discrepancy between humans and non-humans may hinge solely on the implicit processing of verbal instructions regarding task objectives. Additionally, comparisons that depend primarily on task performance or “trials to criterion” (as in most WCST analyses) may provide an incomplete picture. A careful analysis of individual trials (e.g., with respect to both stimulus properties and reaction times), like the one we present here, may uncover important details about the underlying cognitive processing. In particular, rapid task-switching (of the sort that may be of particular interest to clinical measures of impaired cognitive performance) consistently engenders considerable interference between task representations in rhesus macaques, even with extended training. Further comparative research is needed to determine whether this arises as a merely quantitative difference, or instead as a qualitative difference, such as a species-wide bias with respect to speed–accuracy tradeoffs (Caselli and Chelazzi 2011).

Acknowledgments

This research was supported by National Institute of Mental Health Grant R01 MH051153 awarded to H. S. Terrace.

References

- Allport, A.; Styles, EA.; Hsieh, S. Shifting intentional set: exploring the dynamic control of tasks. In: Ultimá, C.; Moscovitch, M., editors. *Attention and performance XV: conscious and non-conscious information processing*. Bradford; Cambridge, MA: 1994. p. 421-452.
- Behar I. Performance of rhesus monkeys on two-cue problems with cues separated. *Anim Learn Behav*. 1974; 2:49–53.
- Beran MJ, Pate JL, Washburn DA, Rumbaugh DM. Sequential responding and planning in chimpanzees (*Pan troglodytes*) and rhesus macaques (*Macaca mulatta*). *J Exp Psychol Anim Behav Process*. 2004; 30:203–212. [PubMed: 15279511]
- Berg EA. A simple objective technique for measuring flexibility in thinking. *J Gen Psychol*. 1948; 39:15–22. [PubMed: 18889466]
- Braver TS, Reynolds JR, Donaldson DI. Neural mechanisms of transient and sustained cognitive control during task switching. *Neuron*. 2003; 39:713–726. [PubMed: 12925284]
- Bryan J, Luszcz MA. Measurement of executive function: considerations for detecting adult age differences. *J Clin Exp Neuropsychol*. 2000; 22:40–55.
- Caselli L, Chelazzi L. Does the macaque monkey provide a good model for studying human executive control? A comparative behavioral study of task switching. *PLoS ONE*. 2011; 6:e21489. [PubMed: 21720549]
- Clapp WC, Rubens MT, Sabharwal J, Gazzaley A. Deficit in switching between functional brain networks underlies the impact of multitasking on working memory in older adults. *Proc Nat Acad Sci USA*. 2011; 108:7212–7217. [PubMed: 21482762]
- Dixon P. Algorithms and selective attention. *Mem Cogn*. 1981; 9:177–184.
- Dreher JC, Grafman J. Dissociating the roles of the rostral anterior cingulate and the lateral prefrontal cortices in performing two tasks simultaneously or successively. *Cereb Cortex*. 2003; 13:329–339. [PubMed: 12631562]
- Dreisbach G, Goschke T, Haider H. The role of task rules and stimulus-response mappings in the task switching paradigm. *Psychol Res*. 2007; 71:383–392. [PubMed: 16397812]
- Dux PE, Ivanoff J, Asplund CL, Marois R. Isolation of a central bottleneck of information processing with time-resolved fMRI. *Neuron*. 2006; 52:1109–1120. [PubMed: 17178412]
- Gómez JC. Species comparative studies and cognitive development. *Trends Cogn Sci*. 2005; 9:118–125. [PubMed: 15737820]
- Gu BM, Park JY, Kang DH, Lee SJ, Yoo SY, Jo HJ, Choi CH, Lee JM, Kwon JS. Neural correlates of cognitive flexibility during task-switching in obsessive-compulsive disorder. *Brain*. 2008; 131:155–164. [PubMed: 18065438]
- Hentschke H, Stüttgen MC. Computation of measures of effect size for neuroscience data sets. *Eur J Neurosci*. 2011; 34:1887–1894. [PubMed: 22082031]
- Jensen G, Altschul D, Danly E, Terrace HS. Transfer of a serial representation between two distinct tasks by Rhesus macaques. *PLoS ONE*. 2013a; 8:e70285. [PubMed: 23936179]
- Jensen G, Ward RD, Balsam PD. Information: theory, brain, and behavior. *J Exp Anal Behav*. 2013b; 100 TBD.
- Jersild AT. Mental set and shift. *Arch Psychol*. 1927; 89:1–82.
- Kamigaki T, Fukushima T, Miyashita Y. Cognitive set reconfiguration signals by macaque posterior parietal neurons. *Neuron*. 2009; 61:941–951. [PubMed: 19324002]
- Kleinsorge T, Heuer H. Hierarchical switching in a multidimensional task space. *Psychol Res*. 1999; 62:300–312.
- Koch I. Automatic and intentional activation of task sets. *J Exp Psychol Learn Mem Cogn*. 2001; 27:1474–1486. [PubMed: 11713881]
- Kornell N, Son LK, Terrace HS. Transfer of metacognitive skills and hint seeking in monkeys. *Psychol Sci*. 2007; 18:64–71. [PubMed: 17362380]
- Lashley, KS. The problem of serial order in behavior. In: Jeffress, LA., editor. *Cerebral mechanisms in behavior*. Wiley; New York: 1951. p. 112-131.

- Logan GD, Gordon RD. Executive control of visual attention in dual-task situations. *Psychol Rev.* 2001; 108:393–434. [PubMed: 11381835]
- MacLeod CM. Half a century of research on the Stroop effect: an integrative review. *Psychol Bull.* 1991; 109:163–203. [PubMed: 2034749]
- Mantel N, Haenszel W. Statistical aspects of the analysis of data from retrospective studies of disease. *J Nat Cancer Inst.* 1959; 22:719–748. [PubMed: 13655060]
- Mayr U, Kliegl R. Task-set switching and long-term memory retrieval. *J Exp Psychol Learn Mem Cogn.* 2000; 26:1124–1140. [PubMed: 11009248]
- Mecklinger A, von Cramon DY, Springer A, Matthesvon Cramon G. Executive control functions in task switching: evidence from brain injured patients. *J Clin Exp Neuropsychol.* 1999; 21:606–619. [PubMed: 10572281]
- Monsell S. Task switching. *Trends Cogn Neurosci.* 2003; 7:134–140.
- Monsell S, Yeung N, Azuma R. Reconfiguration of task-set: is it easier to switch to the weaker task? *Psychol Res.* 2000; 63:250–264. [PubMed: 11004879]
- Moore TL, Killiany RJ, Rosene DL, Prusty S, Hollander W, Moss MB. Impairment of executive function induced by hypertension in the rhesus monkey (*Macaca mulatta*). *Behav Neurosci.* 2002; 116:387–396. [PubMed: 12049319]
- Moore TL, Killiany RJ, Herndon JG, Rosene DL, Moss MB. Impairment in abstraction and set shifting in aged rhesus monkeys. *Neurobiol Aging.* 2003; 24:125–134. [PubMed: 12493558]
- Moore TL, Killiany RJ, Herndon JG, Rosene DL, Moss MB. A non-human primate test of abstraction and set shifting: an automated adaptation of the Wisconsin Card Sorting Test. *J Neurosci Method.* 2005; 146:165–173.
- Myers, J.L.; Well, AD. *Research design and statistical analysis.* 2. Erlbaum; Hillsdale, NJ: 2003.
- Nakahara K, Hayashi T, Konishi S, Miyashita Y. Functional MRI of macaque monkeys performing a cognitive set-shifting task. *Science.* 2002; 295:1532–1536. [PubMed: 11859197]
- Ophir E, Nass C, Wagner AD. Cognitive control in media multitaskers. *Proc Nat Acad Sci USA.* 2009; 106:15583–15587. [PubMed: 19706386]
- Pashler H. Dual-task interference in simple tasks: data and theory. *Psychol Bull.* 1994; 116:220–244. [PubMed: 7972591]
- Pashler, H. Task switching and multitask performance. In: Monsell, S.; Driver, J., editors. *Attention and performance XVIII: control of cognitive processes.* Bradford; Cambridge, MA: 2000. p. 277-307.
- Prado J, Mutreja R, Booth JR. Fractionating the neural substrates of transitive reasoning: task-dependent contributions of spatial and verbal representations. *Cereb Cortex.* 2013; 23:499–507. [PubMed: 22275478]
- Rogers RD, Monsell S. Costs of a predictable switch between simple cognitive tasks. *J Exp Psychol Gen.* 1995; 124:207–231.
- Rossi A, Daneluzzo E, Mattei P, Bustini M, Casacchia M, Stratta P. Wisconsin card sorting test and Stroop test performances in Schizophrenia: a shared construct. *Neurosci Lett.* 1997; 226:87–90. [PubMed: 9159496]
- Sakagami M, Pan X, Uttl B. Behavioral inhibition and prefrontal cortex in decision-making. *Neural Netw.* 2006; 19:1255–1265. [PubMed: 16952442]
- Scarf D, Danly E, Morgan G, Colombo M, Terrace HS. Sequential planning in rhesus monkeys. *Anim Cogn.* 2011; 14:317–324. [PubMed: 21184125]
- Spector A, Biederman I. Mental set and mental shift revisited. *Am J Psychol.* 1976; 89:669–679.
- Stoet G, Snyder LH. Executive control and task-switching in monkeys. *Neuropsychologia.* 2003a; 41:1357–1364. [PubMed: 12757908]
- Stoet G, Snyder LH. Task preparation in macaque monkeys. *Anim Cogn.* 2003b; 6:121–130. [PubMed: 12721788]
- Stoet G, Snyder LH. Extensive practice does not eliminate human switch costs. *Cogn Affect Behav Neurosci.* 2007; 7:192–197. [PubMed: 17993205]

- Stoewer S, Ku SP, Goense J, Stdeudel T, Logothetis NK, Duncan J, Sigala N. Frontoparietal activity with minimal decision and control in awake macaque at 7 T. *Magn Reson Imaging*. 2010; 28:1120–1128. [PubMed: 20117895]
- Subiaul F, Cantlon JF, Holloway RL, Terrace HS. Cognitive imitation in rhesus macaques. *Science*. 2004; 305:407–410. [PubMed: 15256673]
- Sudevan P, Taylor DA. The cuing and priming of cognitive operations. *J Exp Psychol Hum Percept Perform*. 1987; 13:89–103. [PubMed: 2951490]
- Terrace, HS. Simultaneous chaining: the problem it poses for traditional chaining theory. In: Commons, ML.; Herrnstein, RJ.; Wagner, AR., editors. *Quantitative analyses of behavior: discrimination processes*. Ballinger; Cambridge, MA: 1984. p. 115-138.
- Terrace HS. The simultaneous chain: a new approach to serial learning. *Trends Cogn Sci*. 2005; 9:202–210. [PubMed: 15808503]
- Terrace HS, Son LK, Brannon EM. Serial expertise of rhesus macaques. *Psychol Sci*. 2003; 14:66–73. [PubMed: 12564756]
- Washburn DA. Stroop-like effects for monkeys and humans: processing speed or strength of association? *Psychol Sci*. 1994; 5:375–379. [PubMed: 11539143]
- Weed MR, Bryant R, Perry S. Cognitive development in macaques: attentional set-shifting in juvenile and adult rhesus monkeys. *Neuroscience*. 2008; 157:22–28. [PubMed: 18805462]
- Whelan D. Effective analysis of reaction time data. *Psychol Rec*. 2008; 58:475–482.
- Zeamer A, Decamp E, Clark K, Schneider JS. Attention, executive functioning and memory in normal aged rhesus monkeys. *Behav Brain Res*. 2011; 219:23–30. [PubMed: 21168445]

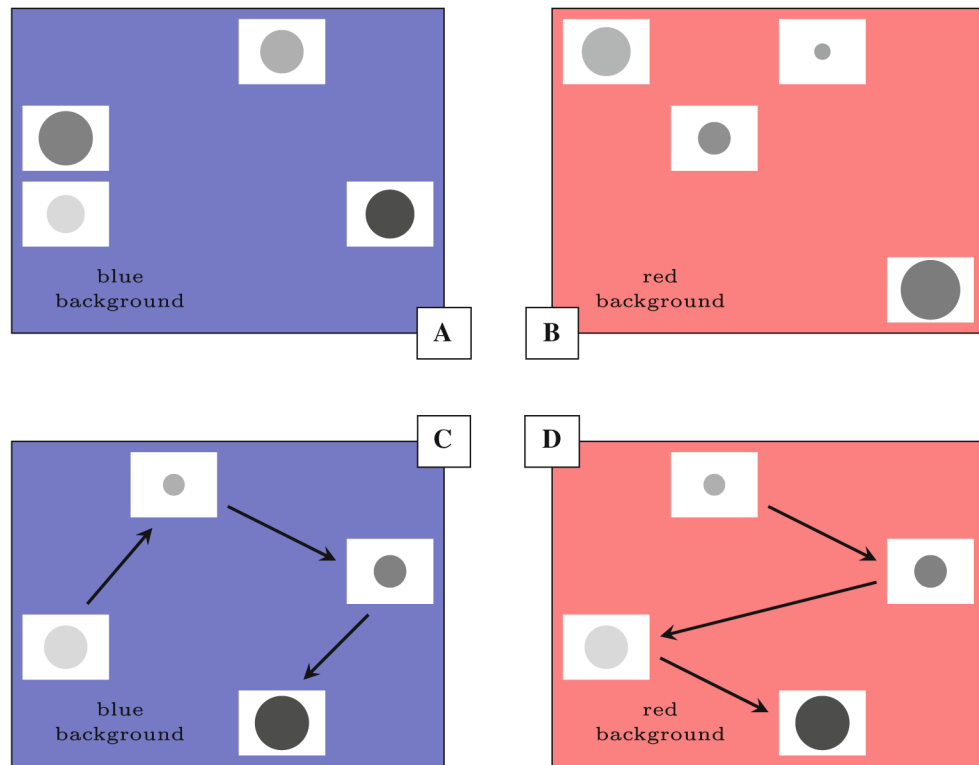


Fig. 1. Stimulus presentation in the psychophysical SimChain procedure. **a** An example of a luminosity trial during Experimental Phase 1, with stimuli biased toward luminosity salience. A *blue background* also signaled that subjects should sort according to luminosity. The exact stimulus properties and positions were randomly assigned on each trial. **b** An example of a radius trial during Experimental Phase 1, with stimuli biased toward radius salience. A *red background* signaled radius responding. **c** An example of a trial in Experimental Phase 2, in which stimuli varied across the full range possible radius and luminosity values. The *arrows* show a correct response, given the *blue background* cue. **d** Another trial in Experimental Phase 2, in which the *red background* cue signals radius responding (color figure online)

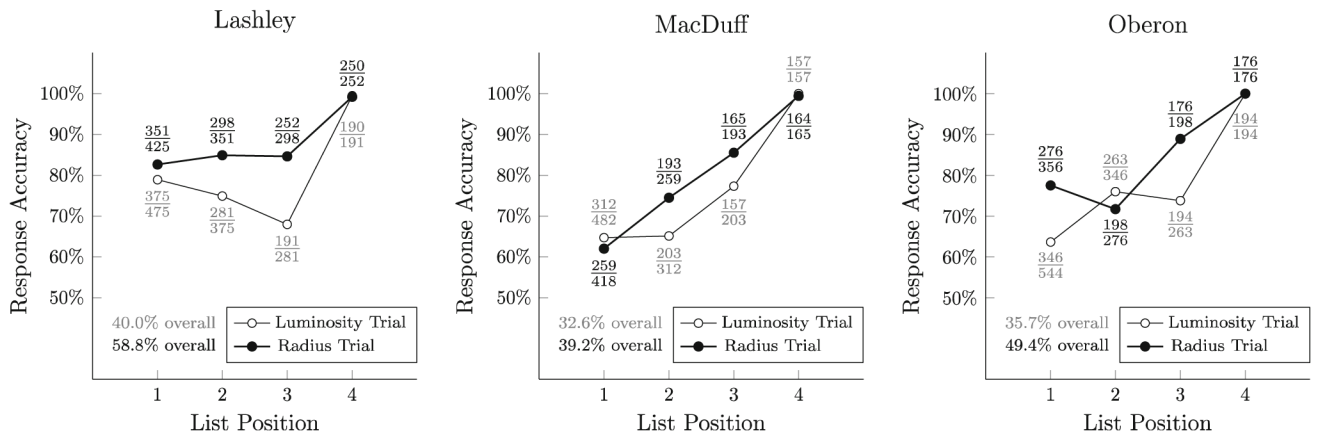


Fig. 2. Accuracy of individual responses in the Experimental Phase 1 (with biased stimulus properties), plotted as percentages for luminosity trials (*white dots*) and radius trials (*black points*). Additionally, the precise conditional ratios are presented, with luminosity trials in *gray text* and radius trials in *black text*. The overall percentages presented at the *bottom* indicate the percentage of lists in which all 4 items are correctly selected

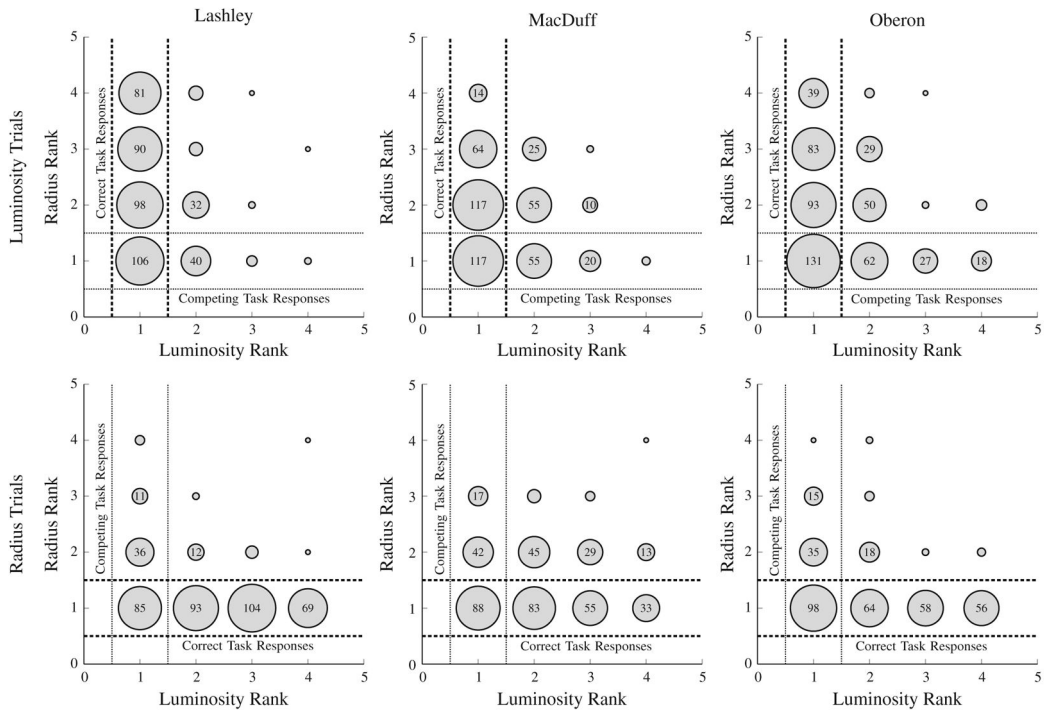


Fig. 3. Frequency plot for the first response in a 4-item list, as a function of the stimulus rank in each of the psychophysical dimensions during Experimental Phase 1. In each plot, responses falling between the *dashed lines* are correct responses, and all others are incorrect. Responses falling between the *dotted lines* are those that are ranked first according to the distractor dimension. In Experimental Phase 1, subjects consistent restricted most of their responses to within the *black dashed lines*, but nevertheless displayed some interference from the distractor dimension

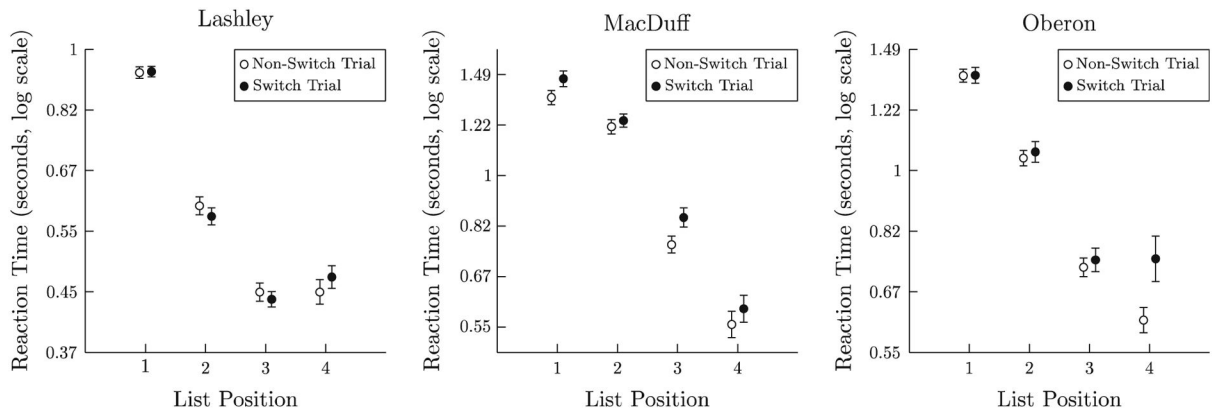


Fig. 4. Mean log_e(reaction time) to respond to each consecutive item in 4-item list, during Experimental Phase 1. “Switch trials” consist of all trials in which the currently cued psychophysical dimension differs from that cued on the preceding trial, requiring a task-shift; non-switch trials consist of all other trials. *Error bars 1 SE*

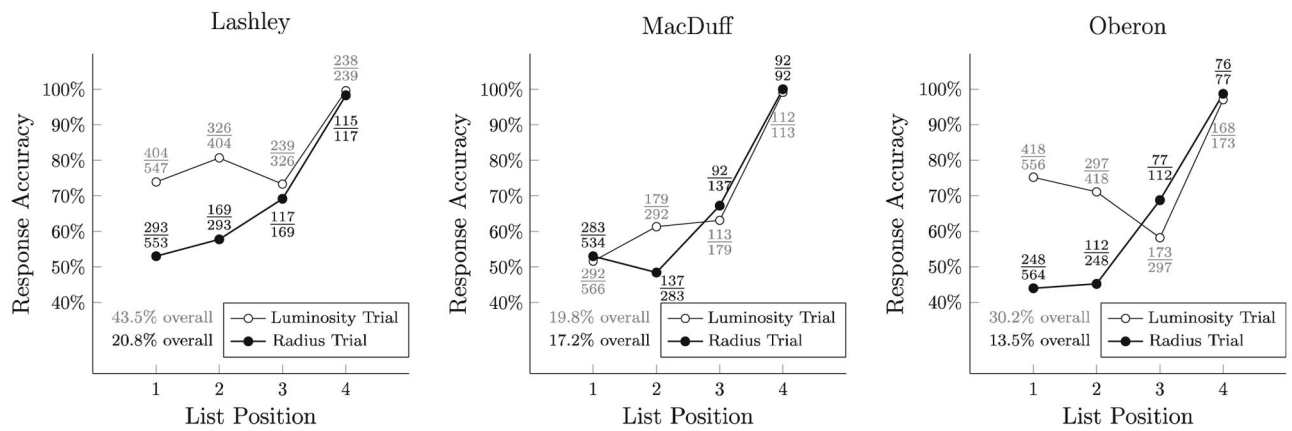


Fig. 5. Accuracy of individual responses in the Experimental Phase 2 (with unbiased stimulus properties), plotted as percentages for luminosity trials (*white dots*) and radius trials (*black points*). Additionally, the precise conditional ratios are presented, with luminosity trials in *gray text* and radius trials in *black text*. The overall percentages presented at the *bottom* indicate the percentage of lists in which all 4 items are correctly selected

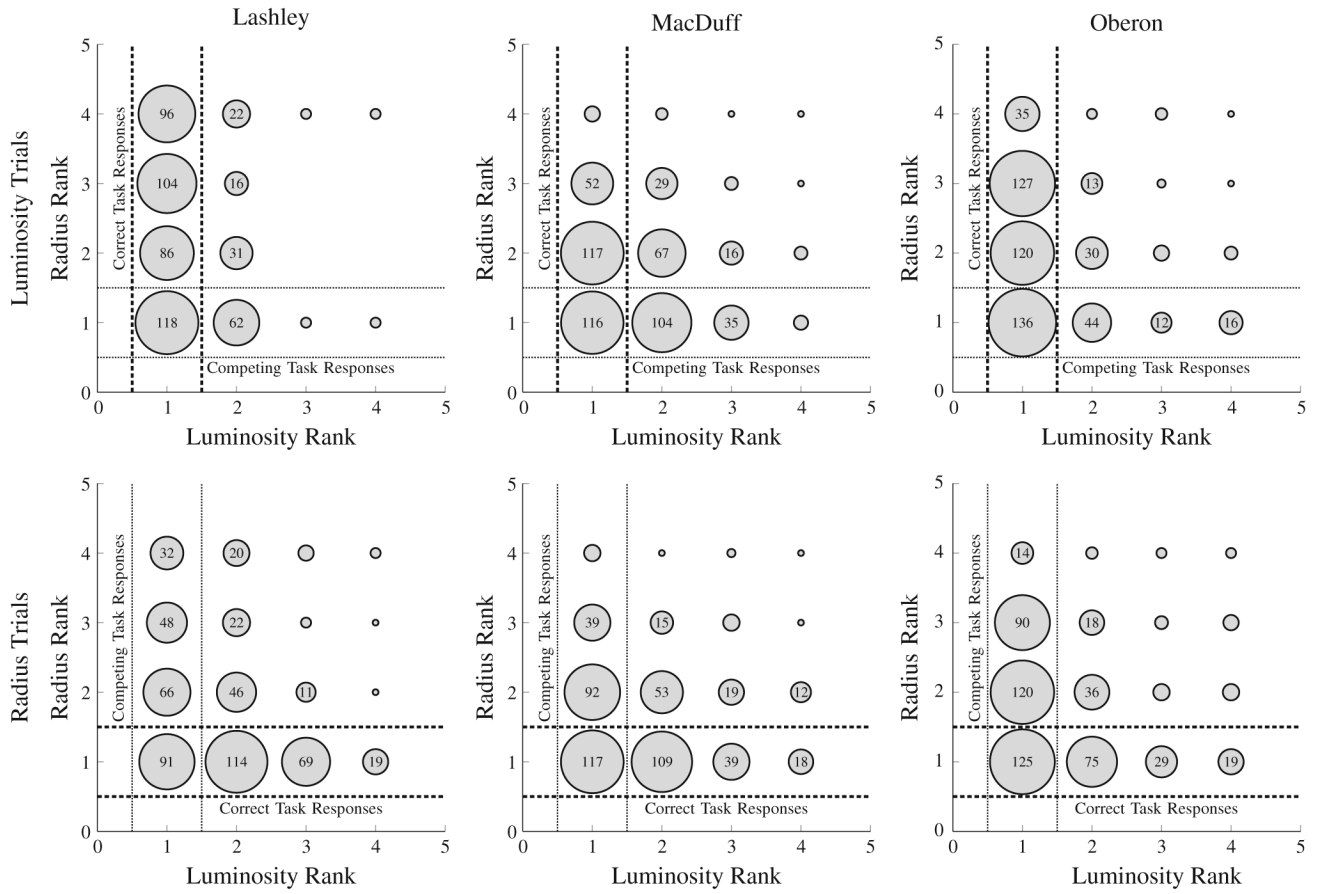


Fig. 6. Frequency plot for the first response in a 4-item list, as a function of the stimulus rank in each of the psychophysical dimensions during Experimental Phase 2. As in Fig. 3, responses falling between the *dashed lines* are correct, and those falling between the *dotted lines* are those that are ranked first according to the distractor dimension. In Experimental Phase 2, subjects showed greater interference than in Phase 1

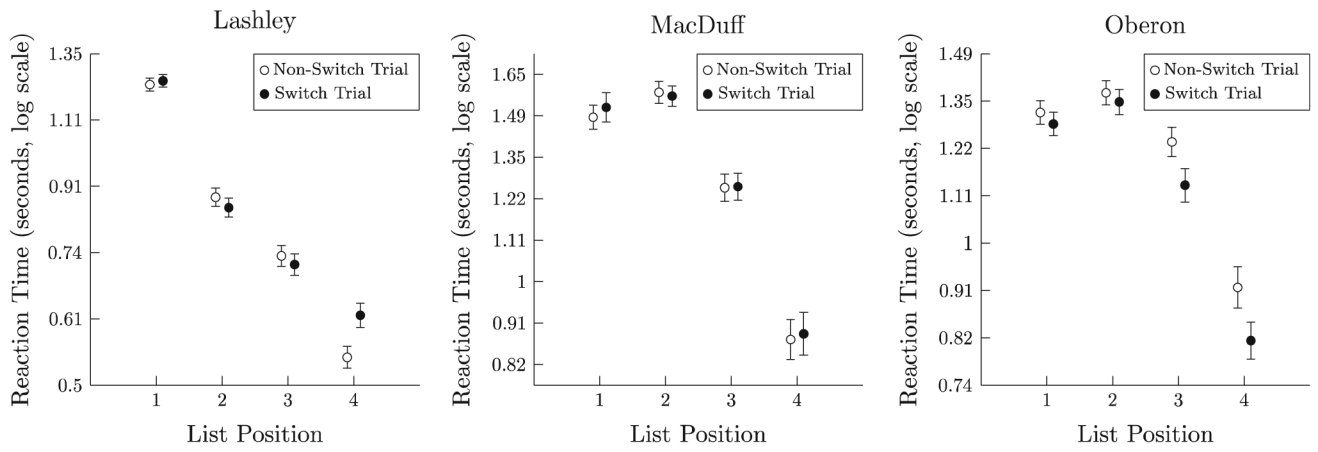


Fig. 7. Mean $\log_e(\text{reaction time})$ to respond to each consecutive item in 4-item list, during Experimental Phase 1. “Switch trials” consist of all trials in which the currently cued psychophysical dimension differs from that cued on the preceding trial, requiring a task-shift; non-switch trials consist of all other trials. *Error bars 1 SE*

Table 1

Stimulus settings used in each phase of training

	Luminosity trials			Radius trials				
	Lum. min	Lum. max	Rad. min.	Rad. max	Lum. min	Lum. max	Rad. min.	Rad. max
Training Phase 1	–	–	–	–	0	0	5	70
Training Phase 2	0	220	36	36	–	–	–	–
Training Phase 3	0	220	36	36	0	0	5	70
Experimental Phase 1 (biased)	0	220	27	57	50	170	5	70
Experimental Phase 2 (unbiased)	0	220	5	70	0	220	5	70

Table 2

Shannon information explained by dimension in Experimental Phase 1

	<u>Luminosity trials</u>		<u>Radius trials</u>	
	I_L	I_R	I_L	I_R
Lashley	1.133	0.032	0.036	1.157
MacDuff	0.774	0.339	0.115	0.761
Oberon	0.653	0.203	0.113	1.005

Table 3

Shannon information explained by dimension in Experimental Phase 2

	<u>Luminosity trials</u>		<u>Radius trials</u>	
	I_L	I_R	I_L	I_R
Lashley	1.04	0.029	0.323	0.289
MacDuff	0.518	0.414	0.341	0.500
Oberon	0.874	0.162	0.531	0.287