15 An Intention-Activation Account of Residual Switch Costs

Ritske De Jong

ABSTRACT *Residual switch costs* are performance costs associated with a shift of task that persist even when there is ample time to prepare in advance for the new task. I present a mixture-model approach for evaluating the contributions of two possible causes of residual switch costs: (1) failures to take advantage of opportunities for advance preparation, and (2) limitations to the completeness of task-set reconfiguration attainable by fully endogenous means. The proposed intention-activation hypothesis of failures to engage in advance preparation is shown to provide a coherent account of the influences of a variety of factors on residual switch costs. Two new experiments tested predictions of the hypothesis regarding the effects of task duration and of time on task on the incidence of preparatory failures.

Although people can perform an almost endless variety of tasks, they are limited in the number of tasks they can perform concurrently, and they generally devote themselves to just one task at any moment. As pointed out by Simon (1994), serial organization of activities should perhaps be viewed not as the result of resource scarcity prohibiting a presumably more efficient parallel organisation, but as an efficient solution to the problem of getting a powerful parallel processing device, the human brain, to support coherent behavior in environments that provide multiple affordances for action. The division of labor by time segments, with processing resources devoted, in turn, to satisfying successive goals, requires signaling and attention control mechanisms to establish priorities, to protect task performance in progress from interference, to update priorities, and to switch from one task to another.

The *task-switching* paradigm provides a simple experimental framework for systematic study of the control processes underlying our ability to switch from one task goal to another and to reconfigure the processing system for engaging in another task. This chapter presents the approach we have developed for detailed analysis and modeling of task-switching performance (De Jong et al. forthcoming) and outlines the novel perspective our approach provides on the causes of performance limitations in task switching. It reports the findings of two new experiments investigating the effects of time on task and of expected task duration on task-switching performance.

Figure 15.1 Mean correct reaction time and error rate as a function of trial type and response-stimulus interval.

15.1 THE TASK-SWITCHING PARADIGM

In the task-switching paradigm, the task to be performed on each trial is selected from a set of alternative tasks, usually choice reaction time (RT) tasks. In the "explicit cue" version, tasks are presented in an unpredictable order. At the start of each trial a cue or instruction signal signals the task to be performed (e.g., Meiran 1996), followed by the presentation of the imperative stimulus after a fixed or random delay, called the "preparation interval." In the "implicit cue" version, the tasks are presented in a predictable order, either in a simple alternating order (e.g., Allport, Styles, and Hsieh 1994) or in a more complex pattern (e.g., Rogers and Monsell 1995), with the response-stimulus interval (RSI) serving as the preparation interval.

There are two basic types of trials. On *nonswitch trials*, the task to be performed is the same as that on the previous trial, and the task set remains in place. On *switch trials*, the task changes, and the task set must be reconfigured. Longer preparation intervals provide more time for advance preparation, that is, for the selection and configuration of the relevant task set before the imperative stimulus is presented. Thus *switch costs*, defined as the difference in performance between switch and nonswitch trials, are expected to diminish gradually as the preparation interval is prolonged.

Residual Switch Costs

In this chapter, I will focus on *residual switch costs*, defined as switch costs at long preparation intervals, that should provide ample time for advance preparation to be completed. Figure 15.1 presents a representative example. The stimuli in the experiment (De Jong et al. forthcoming, exp. 1) were red or blue letters. Both tasks involved the same two keypress response alternatives but in one task, the response was to the color of the letter, and in the other, to its category (consonant or vowel). The experiment used the implicit cue version of the task-switching paradigm, with tasks alternating across trials according to a fixed AABB scheme. Following Rogers and Monsell (1995), clockwise cycling of the position of successive stimuli in a 2×2 grid was used to help subjects keep track of the tasks. Subjects were required to perform one task when the stimulus was displayed in one of the two top positions, and the other, when it was displayed in one of the two bottom positions. Mean initial switch costs at the shortest preparation interval (150 msec) were 240 msec; switch costs declined to 143 msec at the longest interval (1,200 msec). Virtually all task-switching studies have yielded residual switch costs, although the magnitude of such costs relative to initial switch costs varies widely across studies, ranging from very large (e.g., Allport, Styles, and Hsieh 1994; Rogers and Monsell 1995, exp. 2) to very small (e.g., De Jong et al., forthcoming, exp. 3; Meiran 1996, chap. 16, this volume).

Two basic accounts of residual switch costs have been proposed (a third account, by Allport and Wylie, chap. 2, this volume, will be considered later). The first one, which I refer to as the "additional process" (AP) hypothesis, is best exemplified by Rogers and Monsell (1995), who argued that, while the endogenous component of task set reconfiguration could be carried out during the preparation interval, completion of the reconfiguration process had to await triggering by a task-relevant stimulus. The duration of this exogenous component results in residual switch costs. The second account, which I refer to as the "failure-to-engage" (FTE) hypothesis, starts from the notion that advance preparation is optional. Advance preparation is useful because it promotes fast responding to the imperative stimulus, but postponing task set reconfiguration until the arrival of the imperative stimulus still suffices to ensure an accurate, albeit slow, response. According to this perspective, residual switch costs are due to intermittent failures to engage in advance preparation, rather than to a fundamental inability to attain a complete reconfiguration of task set during the preparation interval (i.e., by fully endogenous means). Some broader implications of the FTE hypothesis will be discussed later (see also De Jong, Berendsen, and Cools 1999). First, I will present the modeling approach that we have developed to evaluate the relative merits of these alternative hypotheses (De Jong et al., submitted).

15.2 A MIXTURE MODEL OF RESIDUAL SWITCH COSTS

According to the AP hypothesis, residual switch costs should be manifest on all switch trials. In contrast, the FTE hypothesis holds that such costs should be concentrated within that subset of switch trials on which, for

Figure 15.2 Cumulative distribution functions as a function of trial type and response-stimulus interval (RSI). The fit was produced by the restricted mixture model with $\alpha = 0.51$.

reasons to be discussed later, subjects failed to prepare in advance for the change of task. This suggests that, to distinguish between these hypotheses, entire RT distributions, rather than only their means, should be considered. We therefore computed cumulative distribution functions (CDFs) by dividing the rank-ordered RTs for each subject, for each condition into deciles (10% bins) and then computing the mean RT for each decile. Figure 15.2 shows these functions averaged across subjects, collapsed across the different preparation intervals for nonswitch trials, and at the shortest and longest intervals for switch trials.

The most striking feature of the figure concerns the shape of the CDF for switch trials at the longest preparation interval. In the fast-response range this function approaches that for nonswitch trials, whereas at the slow-response range it approaches the function for switch trials at the shortest preparation interval. This feature is consistent with the FTE hypothesis that responses on switch trials at the longest preparation interval consist of a mixture of two basic types. When advance preparation is carried out, the long preparation interval should provide ample time to attain a suitably reconfigured task set. Responses in this *prepared* state should be relatively fast and have the same RT distribution as those on nonswitch trials, where, by definition, a properly configured set is assumed to be in place. When advance preparation fails to be triggered, the system will remain unprepared throughout the preparation interval. Responses in this *unprepared* state should be relatively slow and have about the same distribution as responses on switch trials at the shortest preparation interval, where preparation may be assumed to have hardly gotten under way when the imperative stimulus arrives.

The FTE hypothesis holds that switch trials at a long preparation interval should yield a mixture of outcomes, with task set reconfiguration either being completed or not having been attempted by the end of the interval. This hypothesis can be formalized in terms of CDFs by the following equation:

$$F_{\text{switch,long PI}}(t) = \alpha\, F_{\text{prepared}}(t) + (1-\alpha)\, F_{\text{unprepared}}(t), \tag{15.1}$$

where $_{\text{switch,long PI}}$ is the CDF for switch trials and a long preparation interval, F_{prepared} and $F_{\text{unprepared}}$ the theoretical CDFs for the prepared and unprepared state, and α the probability that preparation is carried out and completed during the long preparation interval, which I refer to as the "mixing probability." From the definition of residual switch costs (in RT) as the RT difference between switch and nonswitch trials at the longest preparation interval, it follows that $F_{\text{nonswitch,long PI}}$ provides the proper empirical estimate of F_{prepared}. The best available estimate of $F_{\text{unprepared}}$ is provided by $F_{\text{switch,short PI}}$. However, this latter estimate may well be somewhat biased. For one thing, we cannot exclude the possibility that a significant amount of preparation might be carried out within the shortest preparation interval (see De Jong et al. forthcoming for a discussion of other potential problems regarding this estimate). As shown in the chapter appendix, this difference in a priori appropriateness of the two estimates can be effectively dealt with in the model-testing procedure. Substitution of these estimates into equation 15.1 gives a testable version of the mixture model (i.e., the FTE hypothesis) in terms of a relation between the CDFs for three experimental conditions:

$$F_{\text{switch,long PI}}(t) = \alpha\, F_{\text{nonswitch,long PI}}(t) + (1-\alpha)\, F_{\text{switch,short PI}}(t). \tag{15.2}$$

This mixture model can be generalized to allow also for a possible contribution of any additional exogenous component of task set reconfiguration to residual switch costs. Let δ represent the average duration of this hypothetical exogenous component. Even when advance preparation is carried out during the long preparation interval, with probability α, a response on switch trials should then yet incur an average time cost of δ, as compared to a response on nonswitch trials. Incorporating this hypothetical time cost in equation 15.2, we arrive at the following expression for the generalized mixture model for residual switch costs:

$$F_{\text{switch,long PI}}(t) = \alpha\, F_{\text{nonswitch,long PI}}(t-\delta) + (1-\alpha)\, F_{\text{switch,short PI}}(t). \tag{15.3}$$

which assumes that the duration of the exogenous component is invariant. Although this simplifying assumption imposes some restrictions on the generality of our approach, I suggest that it is unlikely to compromise our main objective, for two reasons. First, the assumption yields a first-order approximation that should give the generalized mixture model a substantial advantage over the restricted model ($\delta = 0$) whenever an exogenous component of appreciable mean duration actually contributes to residual switch costs. Second, the approximation may in fact be quite close. A consistent finding in our experiments has been that the relation between the two basis distributions of the mixture model, $F_{\text{nonswitch,long PI}}$ and $F_{\text{switch,short PI}}$, is captured quite accurately by a simple shift on the time axis (e.g., figures 15.2, 15.4, and 15.6). If the RT distributions for the two most extreme preparatory states are related through such a shape-conserving shift, it would seem reasonable to assume that the distributions for completely and partially prepared states are similarly related.

With $\delta = 0$, the generalized mixture model reduces to the pure FTE hypothesis; with $\alpha = 1$, it reduces to the pure AP hypothesis. The intermediate cases, $0 < \alpha < 1$ and $\delta > 0$, comprise a range of models in which various proportions of residual switch costs are attributed to failures to engage in advance preparation and to an exogenous component of task set reconfiguration. We used the "multinomial maximum likelihood method" (MMLM), developed by Yantis, Meyer, and Smith (1991), to determine maximum likelihood estimates of α for the restricted mixture model ($\delta = 0$), and of α and δ for the generalized or full model, and to compute goodness-of-fit statistics for the two models. (Details of the model-testing procedure are given in the chapter appendix.)

Application of this procedure to the present data set (De Jong et al. forthcoming, exp. 1) gave the following results. The average estimate of δ was a nonsignificant 12($\pm$8) msec: $t(15) = 1.48$, $p > 0.15$. The fit of the restricted mixture model was fairly good: $G^2(48) = 64.8$, $p > 0.05$, $\alpha = 0.51$ (average estimate). Thus the residual switch costs can be adequately explained by the hypothesis that subjects engaged in advance preparation on only about 50% of the switch trials. Figure 15.2 depicts the corresponding fit of $F_{\text{switch,long PI}}$, which can be seen to be close in the fast-response range, but less so in the slow-response range. This progressive worsening of the fit over the slower RT range can plausibly be attributed to the likelihood, discussed above, that $F_{\text{switch,short PI}}$ provides a biased estimate of $F_{\text{unprepared}}$ (De Jong et al. forthcoming).

These results lend credence to the hypothesis that failures to engage in advance preparation were the predominant cause of residual switch costs in our first experiment. Although they do not completely rule out a possible contribution of an exogenous component of task set reconfiguration, they indicate that this contribution must have been at best a minor one (see De Jong et al. forthcoming for discussion of the power of the MMLM analyses).

Corroborating evidence for this conclusion was obtained in two other experiments. In our second experiment (De Jong et al. forthcoming), we contrasted the implicit cue version of the task-switching paradigm with the explicit cue version. In both versions, the vertical position of the imperative stimulus determined the relevant task. In the implicit cue version, the position of the next stimulus could be easily predicted from the clockwise cycling of stimulus position in a 2×2 grid. In the explicit cue version, the cue, consisting of a square above or below a horizontal midline, was presented, followed by the display of the stimulus within that square after the preparation interval had elapsed. Although initial switch costs were very similar for the two versions, residual switch costs were about twice as large for implicit as for explicit cues. Mixture model analyses attributed this latter difference to the finding that failures to engage in advance preparation were twice as likely with implicit as with explicit cues. The difference can be easily understood in terms of the prompting effect of the explicit cue on triggering advance preparation.

In our third experiment (De Jong et al. forthcoming), we succeeded in eliminating residual switch costs altogether by using a combination of explicit cues and short trial blocks. That residual switch costs can be eliminated under suitable conditions is consistent with the notion that such costs have a strategic origin and do not arise from the fundamentally limited effectiveness of endogenously initiated preparation.

To summarize, mixture model analyses of residual switch costs have provided consistent support for the hypothesis that such costs are primarily, if not exclusively, due to failures to engage in advance preparation. I would like to stress, however, that these results were obtained for experiments that used young college students as subjects and pairs of relatively simple, speeded tasks. Thus preparatory limitations of the sort assumed by the AP hypothesis may yet prove to make a substantial contribution to residual switch costs in different populations or with pairs of more complex tasks associated with more intricately structured task sets. Indeed, elderly subjects have already been found to exhibit marked preparatory limitations, at least in initial stages of practice (De Jong et al. forthcoming).

15.3 THE ORIGIN OF TRIGGER FAILURES IN TASK SWITCHING

What might cause intermittent failures to engage in anticipatory preparation in the task-switching paradigm? As pointed out, advance preparation is optional, serving primarily to optimize performance on switch trials. Effective use of the option requires (1) that an explicit goal or intention to engage in advance preparation be added to the basic goal structure that governs performance in the task-switching paradigm; and (2) that this intention be retrieved and carried out at the proper time, namely, at the start of the preparation interval. We can thus see a marked correspondence in this aspect of performance between the task-switching paradigm and prospective memory tasks requiring subjects to carry out their intentions at a future time. This suggests that an answer to the question above may be informed by current ideas regarding the factors and mechanisms that determine success and failure in prospective memory tasks.

In prospective memory tasks, people may be assumed to form an associative encoding of a target cue-action pairing and to hold this representation in a state of extra activation (Goschke and Kuhl 1993; Yaniv and Meyer 1987). Success in subsequently retrieving the intention or action depends on the joint influence of two factors: (1) the activation level of the associative encoding; and (2) the characteristics of the target cue (Mantyla 1996). The application of these ideas to the case of advance preparation in task switching is straightforward.

One possible reason for failures to trigger advance preparation might be that the associative encoding of a cue-action pair, with advance preparation as the action, was never formed in the first place. This might happen if subjects failed to understand or appreciate the benefits to be gained by advance preparation, and it should be associated with complete and consistent failures (i.e., $\alpha = 0$). In our experiments, we have encountered such cases only rarely, presumably because our instructions explicitly pointed out such benefits and generally emphasized speed of responding. On the other hand, this factor may have played a role in studies that have found little or no reduction of switch costs with preparation interval (Allport, Styles, and Hsieh 1994; Rogers and Monsell 1995, exp. 2).

Another reason for trigger failures might be that the activation level or strength of the cue-action representation was too low for the cue to reliably trigger its associated action. Several factors may influence the activation level of the cue-action representation. A prominent factor would be the subjective utility of the expected benefits of the action, a low utility being associated with reduced activation of the representation. Because enhanced response speed is the primary benefit to be gained by advance preparation, we can predict that trigger failures should be especially prevalent when response speed is assigned low priority. Two pieces of evidence bear out this prediction. First, manipulation of speed-accuracy instructions in the task-switching paradigm strongly affects both the magnitude of residual switch costs and the estimated mixing probability α, which is much smaller when instructions emphasize accuracy over speed than when speed and accuracy were equally emphasized (De Jong, Schellekens, and Meyman in preparation). Second, correlational analysis of individual differences in task switching within a group of college students has yielded a strong negative correlation (-0.72) between estimated α and mean RT on nonswitch trials (De Jong et al. forthcoming). On the assumption that differences in mean RT reflect, at least in part, differences in priority assigned to response speed, this result nicely corroborates the evidence from explicit manipulation of this priority.

An important factor that may influence the activation level of the cue-action representation is the ability or capacity to generate and maintain goals or intentions in working memory. This ability has been held by some to be a primary determinant of success and failure in prospective memory tasks (Duncan et al. 1996) and in other tasks requiring organization and management of a hierarchy of goals (Anderson, Reder, and Lebiere 1996; Carpenter, Just, and Shell 1990). Following reports of a relation between this ability and "general intelligence" or Spearman's *g*

(Carpenter, Just, and Shell 1990; Duncan et al. 1996), we can predict a similar relation between general intelligence and α. The results of a recently completed study are generally consistent with this prediction (Cools 1998). Like high-*g* normals, low-*g* normals performed with high accuracy in the task-switching paradigm, indicating they were able to switch between tasks according to instructions. Moreover, estimates of δ did not significantly differ from zero for either group, although estimated α was much lower for low-*g* than for high-*g* normals, especially in the implicit cue version of the paradigm.

The intention-activation account also provides a ready explanation for the higher incidence of trigger failures in the implicit cue than in the explicit cue version of the paradigm. The implicit cue version requires subjects to anticipate or predict a change of task on the basis of the regular ordering of tasks; failure to do so would obviously prevent the triggering of advance preparation. This potential cause of trigger failures does not apply to the explicit cue version. Moreover, the commandlike nature of an explicit cue may be assumed to make it a particularly powerful trigger of preparatory activities.

Finally, holding the cue-action representation at a high level of activation may require substantial effort—effort that can be maintained for only brief periods of time. This suggests that failures to engage in advance preparation may become more prevalent as a function of task duration or time on task, a possibility investigated in the following two experiments.

15.4 EXPERIMENT 1: EFFECTS OF TASK DURATION AND TIME ON TASK ON TASK SWITCHING

It may take considerable effort to hold the intention to engage in advance preparation at a sufficiently high level of activation to ensure that advance preparation will be successfully triggered. If people are able to sustain this effort for only brief periods of time, trigger failures should be expected to be more prevalent during long than during short blocks of trials. There is some evidence to support this conjecture. The only experiment finding residual switch costs to be virtually eliminated among individual college students used short blocks of trials (De Jong et al. forthcoming, exp. 3), although procedural details other than block length may have been responsible for this exceptional finding. The two experiments reported here were designed to provide more definitive evidence on this issue.

The experiments addressed two related questions. First, does block length exert reliable effects on the incidence of trigger failures in the task-switching paradigm? Second, if it does, are such effects present right from the start of the block or do they gradually emerge during the course of long blocks? The former possibility would suggest that people pace themselves, setting intention-activation at a level that they expect to be able to sustain for the duration of the block. The latter possibility would suggest that, irrespective of known block length, people initially set the activation at a high level that they cannot sustain for prolonged periods of time. The first experiment required subjects to alternate between blocks of 12 or 48 trials, with subjects being informed about the block length at the beginning of a block. The second experiment used a between-subjects design, with blocks of 12 trials being used for one group of subjects and blocks of 96 trials for the other.

Subjects

Eight students from the University of Groningen, 3 women and 5 men between 19 and 24 years of age, were paid to participate in the experiment.

Apparatus and Procedure

Subjects sat approximately 70 cm in front of a VGA color monitor of an IBM-compatible PC. A white 2×2 grid, consisting of a 6 cm square subdivided into four 3 cm squares, was displayed continuously at the center of the display against a black background. On each trial, the stimulus was a red- or blue-colored letter displayed at the center of one of the small squares; on the next trial, the stimulus was presented in the next square clockwise. Half of the subjects were instructed to perform the letter-classification task when the stimulus appeared in either of the two top squares and the color-classification task when the stimulus appeared in either of the two bottom squares; for the other half, the assignment of tasks to positions was reversed. Because stimulus position cycled in a clockwise fashion, the task changed predictably on every second trial, according to an AABB scheme.

Letters were displayed in an uppercase sansserif font, 1.0 cm wide and 1.4 cm tall. On each trial, the letter was sampled randomly from the set *A*, *E*, *Y*, *U*, *G*, *K*, *M*, and *R*, and its color was sampled randomly from the set red and blue. The stimulus remained on the screen until a response was registered or until 5,000 msec had elapsed. After a response was registered and the stimulus extinguished, the next stimulus appeared after a response-stimulus interval (RSI) with a randomly determined duration of 150, 600, or 1,500 msec.

Subjects received written instructions, which also told them to minimize RT while avoiding errors, and that, to do so, they should make effective use of the RSI to prepare for the upcoming task. An abbreviated version of the instructions appeared for 5,000 msec on the screen at the beginning of a new trial block, after which the first stimulus appeared in the top left square.

Figure 15.3 Experiment 1: Mean correct reaction time and error rate as a function of trial type, block length, and response-stimulus interval.

Design

The experiment consisted of a single session lasting about two hours. The first three trial blocks consisted of 60 trials each and were used for training. Subjects practiced the individual letter and color tasks in the first two blocks and then practiced the task switch condition in the third block. They subsequently completed 124 experimental blocks, with blocks of 12 trials and blocks of 48 trials randomly intermixed in a 4:1 ratio. At the start of a new block, a message on the screen informed them about the length of the block. Because subjects had to start a new block by pushing the space bar, they had ample opportunity to take short breaks in between blocks and were encouraged to do so.

There were two responses: a *left* response, made by pressing the "v" key of the computer keyboard with the left index finger; and a *right* response, made by pushing the "n" key with the right index finger. For the letter task, vowels required one response and consonants the other. For the color task, red letters required one response and blue letters the other. The four possible stimulus-response mapping combinations (two possible mappings for each task) were counterbalanced across subjects.

Results

Reaction Time and Errors Figure 15.3 shows mean correct RT and error rate for switch and nonswitch trials as a function of RSI and block length. Although switch costs decreased with RSI, sizable residual switch costs were obtained for both short and long trial blocks. Responses were somewhat faster in short than in long blocks, especially on switch trials.

Figure 15.4 Experiment 1: Cumulative distribution functions for short and long trial blocks, as a function of trial type and response-stimulus interval (RSI). The fits were produced by the restricted mixture model with $\alpha = 0.64$ (short blocks) and $\alpha = 0.58$ (long blocks).

An ANOVA with block length, RSI, and trial type (switch/nonswitch) as within-subjects factors yielded, for RT, main effects of trial type: $F(1, 7) = 28.0$, $p < 0.001$; of RSI: $F(2, 14) = 14.0$, $p < 0.001$; and of block length: $F(1, 7) = 20.9$, $p < 0.003$. These effects were qualified by interactions of trial type and RSI: $F(2, 14) = 120.9$, $p < 0.001$; and of trial type and block length: $F(1, 7) = 12.5$, $p < 0.01$. No other effects on RT approached significance. Analysis of error rate yielded a significant effect only of trial type: $F(1, 7) = 6.7$, $p < 0.05$. Mean error rate was 4.1% for nonswitch trials and 6.1% for switch trials.

Reaction Time Distributions and Modeling Results Figure 15.4 shows averaged CDFs of RT for the relevant conditions (nonswitch trials at the longest RSI and switch trials at the shortest and longest RSIs) separately for short and long trial blocks. Average estimates of δ were 10 ($\pm$11) msec and 13 ($\pm$9) msec for short and long blocks, respectively; neither value differed significantly from zero (ps > 0.15). The restricted mixture model ($\delta = 0$) gave very good fits for short blocks, $G^2(24) = 25.0$, $p > 0.40$, as well as for long blocks, $G^2(24) = 26.1$, $p > 0.35$. The average estimate of α was 0.64 for short and 0.58 for long blocks; although this difference was in the predicted direction, it was not significant: $F(1, 7) = 2.6$, $p > 0.20$. The corresponding fits of $F_{\text{switch,long RSI}}$ produced by the restricted model are depicted in figure 15.4.

Time on Task To assess possible time-on-task effects, the data for the long trial blocks were reanalyzed with the factor block half (the first versus the last 24 trials) included. Though this test is admittedly crude, the limited number of trials did not permit a more precise decomposition. For RT, this analysis yielded as new results a main effect of block half: $F(1, 7) = 14.0$, $p < 0.01$; and an interaction of block half and trial type:

$F(1, 7) = 9.1$, $p < 0.02$. Mean RT in the second half of the block was longer than that in the first half by 25 msec on nonswitch trials and by 51 msec on switch trials. Error rates did not differ between the first and second half. The average estimates of α were 0.61 and 0.57 for the first and second half, respectively; this difference did not approach significance.

Discussion

Substantial residual switch costs were obtained for both short and long trial blocks. Replicating previous findings (De Jong et al. forthcoming), the modeling results indicated that these residual costs could be attributed almost exclusively to failures to engage in advance preparation, rather than to an additional poststimulus component of task set reconfiguration. Contrary to predictions, however, there was only a nonsignificant tendency for such failures to be more prevalent in long trial blocks. Instead, responses tended to be somewhat slower in long blocks, especially on switch trials, and this effect appears to have been largely due to a decline in response speed from the first to the second half of long blocks. Because these effects of block length did not interact with RSI, we suggest that they may reflect a gradual slowing of both preparation and poststimulus task execution during the course of a long block.

Although these results seem to refute our assumption that people may have trouble sustaining for prolonged periods the effort needed to keep the intention to prepare in advance highly activated, an alternative interpretation is possible. Faced with a mixture of short and long blocks, subjects may have adopted a conservative, worst-case strategy, and have set intention-activation at a level that they could sustain for the duration of the long blocks. Though admittedly ad hoc, this interpretation receives some support from several recent studies that suggest a marked lack of flexibility in adjusting control settings in response to different instructions or task requirements (Los 1996; Strayer and Kramer 1994). Experiment 2 was designed to address these remaining uncertainties.

15.5 EXPERIMENT 2

Using a between-subjects design, one group of subjects was exposed only to short trial blocks and the other group only to long blocks. If the suggested interpretation of the absence of clear effects of block length on α in experiment 1 is correct, then two predictions follow for experiment 2. First, a clear effect of block length on α should now be present. Second, if the intermixing of short and long blocks indeed caused subjects to adopt an overly conservative level of intention-activation for the short blocks of experiment 1, then α for the short blocks should be larger than that in experiment 1.

Figure 15.5 Experiment 2: Mean correct reaction time and error rate as a function of trial type, block length, and response-stimulus interval.

Method

There were 20 new paid participants, 10 male and 10 female, all students at the University of Groningen between 19 and 26 years of age. The apparatus and procedure were the same as in experiment 1, as was the design, with two important exceptions. After three practice blocks, half the subjects completed 100 blocks of 12 trials whereas the other half completed 12 blocks of 96 trials. Every fourth block was a pure-task block exactly similar to the mixed-task blocks except that only one of the two alternative tasks was relevant throughout the block.

Results

RT and Errors Figure 15.5 shows mean correct RT and error rate for pure task, nonswitch, and switch trials as a function of RSI and block length. As in experiment 1, switch costs declined with RSI but sizable residual switch costs were obtained, especially in long blocks. Responses were substantially faster in short than in long blocks, especially on switch trials.

An ANOVA with block length as a between-subjects factor and with RSI and trial type (switch/nonswitch) as within-subject factors yielded, for RT, main effects of trial type: $F(1, 18) = 93.6$, $p < 0.001$; of RSI: $F(2, 36) = 59.7$, $p < 0.001$; and of block length: $F(1, 18) = 9.8$, $p < 0.01$. These effects were qualified by interactions of trial type and RSI: $F(2, 36) = 109.4$, $p < 0.001$; and of trial type and block length: $F(1, 18) = 6.8$,

Figure 15.6 Experiment 2: Cumulative distribution functions for short and long trial blocks, as a function of trial type and response-stimulus interval (RSI). The fits were produced by the restricted mixture model with $\alpha = 0.80$ (short blocks) and $\alpha = 0.57$ (long blocks).

$p < 0.02$. No other effects on RT approached significance. Analysis of error rates yielded a significant effect only of trial type: $F(1, 18) = 12.1$, $p < 0.01$. Mean error rate was 3.0% for nonswitch trials and 4.6% for switch trials.

We conducted a separate analysis of the difference between pure task and nonswitch trials. For RT, this analysis yielded main effects of trial type: $F(1, 18) = 58.2$, $p < 0.001$; and of block length: $F(1, 18) = 4.9$, $p < 0.05$; and a significant interaction of trial type and block length, reflecting a larger pure task/nonswitch RT difference in long than in short blocks: $F(1, 18) = 8.9$, $p < 0.01$. No other effects on RT approached significance. Mean error rate in pure task blocks was 2.6%.

Reaction Time Distributions and Modeling Results Figure 15.6 shows averaged CDFs of RT for the relevant conditions separately for short and long trial blocks. Average estimates of δ were -1 (± 14) msec and 4 (± 11) msec for short and long blocks, respectively; neither value differed significantly from zero (ps > 0.25). The restricted mixture model gave excellent fits for both short and long blocks: $G^2(30) = 24.6$, $p > 0.70$ and $G^2(30) = 29.2$, $p > 0.50$, respectively. The average estimate of α was 0.80 for short blocks and 0.57 for long blocks; this difference was highly significant: $F(1, 18) = 10.9$, $p < 0.01$. The corresponding fits of $F_{\text{switch, long RSI}}$ are depicted in figure 15.6.

Time on Task In order to assess time-on-task effects, the data for the long blocks were reanalyzed with the factor block half included. For RT, this analysis yielded as the only new result a significant main effect of block half, reflecting an increase in RT of 26 msec from the first to the second half: $F(1, 9) = 7.1$, $p < 0.05$, Error rates did not differ between the two halves. The average estimate of α was 0.55 and 0.61 for the first and second half, respectively; this difference did not approach significance.

Comparison with Experiment 1 The average estimate of α for short blocks in experiment 1 was 0.64, as compared to 0.80 in experiment 2; this difference was significant: $F(1, 16) = 5.7$, $p < 0.05$.

Discussion

The two key predictions for this experiment were confirmed. First, differences in estimated α indicate that failures to engage in advance preparation were about twice as prevalent in long as in short trial blocks. Moreover, the absence of negative time-on-task effects on α in long blocks suggests that subjects paced themselves, setting intention-activation at an initial level they could sustain for the duration of the block. Second, trigger failures were more prevalent in short blocks when such blocks were intermixed with long blocks (experiment 1) than when only short blocks were administered (experiment 2). This provides further evidence for pacing because such a difference could have occurred only if subjects did take block length into account in setting the initial level of intention-activation. It also lends credence to the idea that the intermixing of short and long trial blocks in experiment 1 led subjects to adopt a compromise setting, rather than to adjust the setting for each of the two block lengths.

Consistent with the intention-activation account of residual switch costs, the combined results from the two experiments indicate that holding the intention to engage in advance preparation at a high level of activation requires considerable effort, and that, as in distance running, people can adaptively manage these requirements to maintain a steady level of performance in a prospective, rather than only a reactive manner. On the other hand, the results also suggest clear limits to the flexibility with which people adjust the level of intention-activation on the basis of expected task duration in the task-switching paradigm.

Finally, responses in pure task blocks were found to be considerably faster than those on nonswitch trials, especially in long trial blocks. As has been emphasized by Allport and Wylie (chap. 2, this volume), this finding suggests that task set reconfiguration could not have been optimal or complete even on nonswitch trials. The implications of this for the present theoretical approach will be discussed in section 15.6.

15.6 CONCLUSIONS

The mixture model approach has yielded consistent support for the FTE hypothesis that residual switch costs stem from intermittent failures to take advantage of opportunities for advance preparation. The all-or-none conception of advance preparation implicit in this hypothesis should be taken quite literally. For instance, consider the alternative hypothesis that the degree of advance preparation has, on a trial-to-trial basis, a continuous and smooth distribution with 0% and 100% as extremes and with α

representing its central tendency. While such a continuous conception of advance preparation would seem perfectly plausible, it can be shown to be incompatible with the small and nonsignificant estimates of δ that have consistently been obtained (De Jong et al. forthcoming). Clearly, although we cannot exclude the possibility that some other, yet-unspecified, model may offer an equally precise account of residual switch costs, at this point, the FTE hypothesis seems to come close to identifying the actual primary cause of this intriguing empirical phenomenon.

This conclusion, it must be stressed, is based exclusively on evidence from experiments that used young college students as subjects and pairs of simple tasks, and should therefore not be generalized beyond such cases at this point. Rather, it should provide a clear point of reference for the evaluation and interpretation of residual switch costs in other populations or for pairs of more complex tasks, where limitations to the completeness of task set reconfiguration attainable by fully endogenous means might well be present. Indeed, the mixture model approach should be most useful, perhaps even indispensable, when residual switch costs are jointly due to such preparatory limitations and to intermittent failures to engage in advance preparation, and it becomes important to assess the relative contributions of these different causes (De Jong et al. forthcoming).

An intention-activation account of intermittent failures to engage in advance preparation was proposed, based on a marked correspondence between this aspect of task-switching performance and prospective memory performance. The account was argued to provide a coherent explanation of the influence of a variety of factors on the incidence of such failures, including the effects of task duration and time on task in the two experiments reported here. Admittedly, pertinent empirical evidence is still scant and potentially important factors, such as the predictive validity of the task cue, task complexity, and training, remain to be explored. Nevertheless, these initial results are encouraging and suggest that the intention-activation account may provide a versatile theoretical framework for future studies of strategic control in the task-switching paradigm.

Residual Switch Costs and Nonswitch/Pure Task Differences

The intention-activation account may also shed light on another intriguing finding in the recent task-switching literature. Responses on nonswitch trials are usually considerably slower than those in pure task blocks, and Stroop-like interference by the competitor task is usually observed on nonswitch trials (see Allport and Wylie, chap. 2, this volume). These findings indicate that previously relevant task sets are generally not fully disengaged on nonswitch trials. It is important to note that this does not undermine the all-or-none conception of advance preparation embodied by the FTE hypothesis. The FTE hypothesis holds that the task set in place on nonswitch trials can also be attained on switch trials by fully endogenous means, and will be if advance preparation is carried out and completed. It does not assume or require this set to be fully reconfigured, with competing task sets fully disengaged, or to be the same as the task set in pure task blocks. Thus there is no logically necessary relation between residual switch costs and performance differences between pure task and nonswitch trials. Yet the question of whether and how these two phenomena might be related is an important one.

Allport and Wylie (chap. 2, this volume) outlined an interesting theoretical perspective on this issue. They suggest that incomplete disengagement of prior task sets is caused by involuntary residual priming. Residual priming of prior task sets can retard the system's settling to a unique response. Such proactive interference can be long-lasting, which explains differences between pure task and nonswitch trials, but is typically largest for the first trial of a run (the "restart" effect), which explains residual switch costs. On the other hand, the notion that residual switch costs are due largely to involuntary residual priming of prior task sets is clearly incompatible with the present evidence that such costs reflect, not fundamental preparatory limitations, but inconsistent use of preparatory capabilities.

I would like to argue that an integrative account should probably be based on the notion that, like residual switch costs, pure task/nonswitch differences depend on the control strategies that subjects adopt. Even though capable of completely disengaging prior task sets, subjects might opt not to fully exercise this capability when, for instance, these sets may need to be reinstalled shortly. The effort requirements for executive control might be significantly reduced by such a conservative control strategy, but at the expense of suboptimal task performance and potential interference effects.

The hypothesis that pure task/nonswitch performance differences may reflect a strategic compromise between minimizing control effort and maximizing task performance closely resembles the intention-activation account of residual switch costs. Combining the two accounts, we can predict that greater effort invested in executive control should have the dual effect of enhancing α and reducing pure task/nonswitch performance differences. The previously mentioned strong negative correlation between α and nonswitch RT is consistent with this prediction. Two more specific predictions can be made. First, experimental factors that affect α by influencing the level of intention-activation should also affect pure task/nonswitch differences. This prediction is borne out by the finding in experiment 2 that α was substantially larger and the pure task/nonswitch RT difference substantially smaller in short blocks than in long trial blocks. Note also that the very presence of an effect of block length on the pure task/nonswitch RT difference would seem to argue against the notion that this difference is due to involuntary persistence of the competing task set. Second, factors that affect α by influencing trigger strength should leave the pure task/nonswitch difference unaffected. This is the pattern obtained when implicit and explicit cues were contrasted: estimated α was substantially larger for explicit cues whereas the pure task/nonswitch RT difference was the same for the two types of cue (De Jong et al. forthcoming, exp. 2). These considerations would seem to provide reason to take seriously the possibility that pure task/nonswitch performance differences are not an inevitable result of involuntary persistence of competing task sets but, like residual switch costs, have a largely strategic origin.

APPENDIX

The multinomial maximum likelihood method (MMLM) for testing mixture models requires grouping of rank-ordered RTs into a finite number of bins (Yantis, Meyer, and Smith 1991). In our analyses, we used five bins, with bins 1 to 4 comprising the consecutive first four 8% portions of RTs in the mixture condition and bin 5 comprising the remaining 68% slowest RTs. This choice of bins served to reduce unwanted effects of a possibly biased estimate of $F_{\text{unprepared}}$ on goodness-of-fit statistics (see De Jong et al. forthcoming for details).

The log likelihood ratio statistic G^2 served as the goodness-of-fit statistic. For a valid restricted mixture hypothesis ($\delta = 0$) and with five bins, this statistic has an asymptotic X^2 distribution with 3 degrees of freedom (Yantis, Meyer, and Smith 1991). Because the generalized (full) mixture model is not linear in its parameters α and δ, the asymptotic distribution of the G^2 statistic for a valid model is not known a priori and, from Monte Carlo simulations, depends on such factors as the true value of α and the degree of overlap between the two basis distributions. This complicates the application of the common likelihood ratio procedure to test for a significant improvement of fit by the generalized model. Though this technical problem is not insurmountable, we (De Jong et al. forthcoming) used an alternative and less complex procedure that sufficed for the analysis of the experimental data presented in this chapter. In the first step of the procedure, maximum likelihood estimates of the models' parameters were computed for each subject. In the second step, we tested whether the average estimate of δ differed significantly from zero across subjects. This test is based on the notion that a significantly improved fit by the generalized model would be meaningful only if accompanied by consistently positively valued estimates of δ. Because the null hypothesis of $\delta = 0$ could not be rejected for any of the experimental conditions presented in this chapter, precise assessment of the relative adequacy of the generalized model was unnecessary. In the third and final step, the overall adequacy of the restricted model was therefore assessed by summing the individual G^2 values and their associated degrees of freedom for subjects as a group.

REFERENCES

Allport, D. A., Styles, E. A., and Hsieh, S. (1994). Shifting attentional set: Exploring the dynamic control of tasks. In C. Umiltà and M. Moscovitch (Eds.), *Attention and Performance XV,* pp. 421–452. Cambridge, MA: MIT Press.

Anderson, J. R., Reder, L. M., and Lebiere, C. (1996). Working memory: Activation limitations on retrieval. *Cognitive Psychology, 30,* 221–256.

Carpenter, P. A., Just, M. A., and Shell, P. (1990). What one intelligence test measures: A theoretical account of the processing in the Raven Progressive Matrices Test. *Psychological Review, 97,* 404–431.

Cools, R. (1998). Frontal deficits and symptoms in schizophrenia. Master's thesis, University of Groningen.

De Jong, R., Berendsen, E., and Cools, R. (1999). Goal neglect and inhibitory limitations: Dissociable causes of interference effects in conflict situations. *Acta Psychologica, 101,* 379–394.

De Jong, R., Emans, B., Eenshuistra, R., and Wagenmakers, E-J. (Forthcoming). Structural and strategical determinants of intentional task control.

De Jong, R., Schellekens, J., and Meyman, T. F. (In preparation). Ansatz, Aufgabe, and Einstellung in task switching.

Duncan, J., Emslie, H., Williams, P., Johnson, R., and Freer, C. (1996). Intelligence and the frontal lobe: The organization of goal-directed behavior. *Cognitive Psychology, 30,* 257–303.

Goschke, T., and Kuhl, J. (1993). Representation of intentions: Persisting activation in memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 1211–1226.

Los, S. A. (1996). On the origin of mixing costs: Exploring information processing in pure and mixed blocks of trials. *Acta Psychologica, 94,* 145–188.

Mantyla, T. (1996). Activating actions and interrupting intentions: Mechanisms of retrieval sensitization in prospective memory. In M. A. Brandimonte, G. O. Einstein, and M. A. McDaniel (Eds.), *Prospective memory: Theory and applications,* pp. 241–275. Hillsdale, NJ: Erlbaum.

Meiran, N. (1996). Reconfiguration of processing mode prior to task performance. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 1423–1442.

Rogers, R. D., and Monsell, S. (1995). Costs of a predictable switch between simple cognitive tasks. *Journal of Experimental Psychology: General, 124,* 207–231.

Simon, H. A. (1994). The bottleneck of attention: Connecting thought with motivation. In D. Spaulding (Ed.), *Integrative views on motivation, cognition, and emotion: The Nebraska Symposium on Motivation,* vol. 41, pp. 1–21. Lincoln, NE: University of Nebraska Press.

Strayer, D. L., and Kramer, A. F. (1994). Strategies and automaticity: 2. Dynamic aspects of strategy adjustment. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 342–365.

Yaniv, I., and Meyer, D. E. (1987). Activation and metacognition of inaccessible stored information: Potential bases for incubation effects in problem solving. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13,* 187–205.

Yantis, S., Meyer, D. E., and Smith, J. E. K. (1991). Analyses of multinomial mixture distributions: New tests for stochastic models of cognition and action. *Psychological Bulletin, 110,* 350–374.

16 Reconfiguration of Stimulus Task Sets and Response Task Sets during Task Switching

Nachshon Meiran

ABSTRACT A tentative model of task switching was tested in two experiments. The model accounts for the switching costs observed in previous experiments by attributing them to multivalent task elements, in the present paradigm bivalent stimuli (relevant for both tasks) and bivalent responses (used in both tasks). It assumes that stimulus task sets enable nearly univalent mental representations of bivalent stimuli, and that response task sets enable nearly univalent mental representations of bivalent responses. Results support two novel predictions of the model: (1) the residual switching cost is substantial with bivalent responses, but negligible with univalent responses; and (2) the preparatory cost is substantial when bivalent target stimuli follow bivalent stimuli, but negligible when either the current target stimulus or the previous one is univalent. Hence there is an approximate one-to-one mapping between preparatory cost and reconfiguration of stimulus task set, on the one hand, and between residual switching cost and reconfiguration of response task set, on the other.

Despite its obvious importance to the study of cognitive control, task switching was barely studied until recently. Furthermore, what used to be the dominant experimental paradigm (i.e., Jersild 1927) suffers from serious shortcomings (see Pashler, chap. 12, this volume), limiting the usefulness of most previous results. Although two better-controlled paradigms were developed, the alternating-runs paradigm (Fagot 1994; Rogers 1993; Rogers and Monsell 1995; Stablum et al. 1994) and the cuing paradigm (e.g., De Jong 1995; Meiran 1996; Shaffer 1965; see also Sudevan and Taylor 1987), extensive work with these paradigms is so recent that our understanding of the phenomena remains rudimentary, and models based on them should be regarded as first approximations. The present chapter introduces such a model, which accounts successfully for previous results and two of whose novel predictions were tested in two experiments.

16.1 THE EXPERIMENTAL PARADIGM

Two and sometimes more different tasks were performed over a long series of trials; in most of the experiments, the tasks required locating a target stimulus within a 2×2 grid (figure 16.1). Subjects were instructed to indicate either the vertical position (the *up-down* task) or the horizon-