Memory & Cognition
2005, 33 (8), 1484-1494

Mixing costs in task shifting reflect sequential processing stages in a multicomponent task

MARCO STEINHAUSER and RONALD HÜBNER
Universität Konstanz, Konstanz, Germany

We investigated the hypothesis that mixing costs in task shifting reflect the sequential selection of task components (e.g., stimulus categories) during task execution. This proposition was supported by Hübner, Futterer, and Steinhauser (2001), who showed that the amount of mixing costs depends on the number of mixed task components (e.g., stimulus level and judgment). However, their results could also be explained by a task set selection account, because task components and task sets were confounded. In Experiments 1 and 2, we compared conditions in which either the number of task sets varied and the number of mixed task components was constant or vice versa. Only the number of mixed task components was predictive for the mixing costs. In Experiment 3, we replicated the additivity of mixing costs from level and judgment mixing. Our results suggest that the mixing costs reflect a selection strategy in which interference is reduced in a stepwise manner.

If a stimulus is associated with different actions, it is possible that its appearance leads to a conflict in behavioral control (Norman & Shallice, 1986). Under such circumstances, higher order control processes are required for resolving these conflicts and maintaining goal-directed behavior. Experimentally, this situation can be examined by means of the task-shifting paradigm (e.g., Allport, Styles, & Hsieh, 1994; Rogers & Monsell, 1995), in which subjects alternate between different tasks that are required for the same set of stimuli. Usually, two dissociable types of costs in response times (RTs) and error rates can be observed in corresponding experiments (De Jong, 2001; Hübner, Futterer, & Steinhauser, 2001; Keele & Rafal, 2000; Kray & Lindenberger, 2000; Mayr, 2001; Rogers et al., 1998; Salthouse, Fristoe, McGuthry, & Hambrick, 1998). One type, the so-called *shift costs*, can be measured by comparing trials following a task shift with trials following a task repetition. The other type, which is called *mixing costs*, refers to the impaired performance throughout a whole experimental block of alternating tasks, as compared with blocks with a single task. Mixing costs can even be measured when task repetition trials from both types of blocks are exclusively compared (Mayr, 2001; Salthouse et al., 1998).

Shift costs have frequently been supposed to reflect, at least partially, the increased strength of task conflicts in shift trials (e.g., Allport et al., 1994; Ruthruff, Remington, & Johnston, 2001; Schuch & Koch, 2003; Waszak, Hommel, & Allport, 2003; Wylie & Allport, 2000). Mixing costs, however, have been related to global control strategies that counteract these conflicts. A result that supports this notion comes from Mayr (2001), who showed that mixing costs occur only under conditions in which overlapping stimulus and response sets promote task conflicts. With respect to the nature of these control strategies, some authors (Hübner et al., 2001; Keele & Rafal, 2000; Mayr, 2001) have discussed the idea that stimulus-induced conflicts are reduced by the implementation of additional selection processes that activate representations of an intended task and inhibit those of nonintended tasks. These processes can cause the mixing costs, given that they are implemented poststimulus and affect all the trials of a mixed task block.

However, details of such control processes are largely unknown. One question concerns the nature and role of the involved task-related representations. Hübner et al. (2001) have proposed that control proceeds by selecting task components, such as stimulus categories. In contrast, it is also conceivable that conflicts are resolved by selecting whole task sets that comprise all aspects of a certain task. Distinguishing between these two accounts could be crucial for solving the more general problem of whether control directly affects lower level processes or is restricted to the manipulation of and the selection between high-order representations. As the following considerations will show, existing evidence is not sufficient to answer this question. After discussing the two accounts in more detail, we will report three experiments that were conducted to distinguish between them.

Task Components Versus Whole Task Sets

Usually, task-shifting studies apply highly overlapping tasks. Most often, each task comprises the same stimuli and responses, and only the translation rule differs. For

---

This research was supported by Grant Hu 432/8-1 to the second author from the Deutsche Forschungsgemeinschaft. We thank Michel D. Druey, Ulrich Mayr, Catherine M. Arrington, and Nachshon Meiran for helpful comments. Correspondence concerning this article should be addressed to M. Steinhauser, Universität Konstanz, Fachbereich Psychologie, Fach D29, D-78457 Konstanz, Germany (e-mail: marco.steinhauser@uni-konstanz.de).

any theory on task shifting, it would be crucial to explain how such overlapping tasks are represented and on which representational levels conflicts can emerge. In this respect, two broad classes of theories can be distinguished. According to the first class, tasks are represented holistically as task sets (e.g., Allport et al., 1994; Mayr & Keele, 2000; Rogers & Monsell, 1995) that "specify the configuration of perceptual, attentional, mnemonic, and motor processes critical for a particular task goal" (Mayr & Keele, 2000, p. 5). Theories of this type frequently assume that task conflicts emerge when stimuli activate competing task sets (e.g., Allport et al., 1994; Rogers & Monsell, 1995). As a consequence, task-shifting effects are explained independently of the structure and processing demands of the single tasks.

In a second class of theories, however, it is assumed that lower level task representations are involved in conflicts. For instance, Schuch and Koch (2003) used a paradigm in which subjects alternated between magnitude and parity judgments of single digits. They proposed that task conflicts emerge on the level of associations between stimulus categories (such as *odd* or *even*) and responses. In this way, they were able to explain conflict-related phenomena, such as shift costs and response repetition effects. Corresponding explanations have also been proposed for other tasks (e.g., Meiran, 2000; Wylie & Allport, 2000). These theories predict that shifting effects are determined mainly by the structure of the tasks that are applied.

The question of whether whole task sets or lower level task representations are involved in conflicts and their solution is also relevant for explaining the mixing costs. If one assumes that the mixing costs reflect a conflict resolution strategy in which relevant representations are selected against irrelevant ones, one could further ask whether this selection takes place on the level of whole task sets or on the level of lower order task representations.

For instance, Hübner et al. (2001) proposed that mixing costs are caused by the selection of lower level task representations, which they called *task components*. Conflicts arising from the activation of multiple tasks are reduced by selecting relevant task components against irrelevant ones. For instance, when subjects have to shift between parity and magnitude judgments, Hübner et al. would assume that response selection is preceded by a category selection process. Within this process, the required set of stimulus categories (e.g., *odd*/*even*, depending on the goal) is selected. In this way, the impact of irrelevant stimulus categories (e.g., *less*/*greater*) on response selection (see Schuch & Koch, 2003) is reduced.

This account, however, implies that the selection processes that support the resolution of conflicts between tasks have access to single components of a task. In contrast, it is also conceivable that these control processes have access only to higher representations, such as task sets. In this case, the only way to reduce conflicts between tasks would be to select between these task sets. Accordingly, the mixing costs could reflect a process of *task set selection*.

One way to examine this question empirically would be to test whether mixing costs are sensitive to the structure of tasks. Experiments designed to solve this question have already been conducted by Hübner et al. (2001). They derived a simple prediction: If mixing costs reflect the time required for the selection of a task component differing between the tasks (e.g., the stimulus categories), varying more than one task component should require more than one selection process. At first glance, the results that were obtained seem to support these predictions. Unfortunately, the paradigm does not allow for an unequivocal interpretation, as will be shown in the following.

Hübner et al. (2001) used a paradigm in which different task components varied independently across trials. Their subjects had to apply one of two judgments (magnitude or parity judgment) to one of two stimulus dimensions (global or local level of a hierarchical numeral; see Navon, 1977). Therefore, two stimulus categories had to be taken into account: the level (global/local) by which a numeral had to be selected and the judgment categories (*odd*/*even*, *less*/*greater*) by which a numeral had to be translated into a response. Both task components were announced by a cue, and the cue–stimulus interval was self-paced—that is, on each trial, the subjects could prepare for the upcoming task as long as they wanted (Dixon, 1981; Dixon & Just, 1986). Since the levels and judgments were either mixed or held constant, four mixing conditions could be realized. This made it possible to separately compute the mixing costs for the individual component selection processes—that is, for level selection and judgment selection.[1]

The variation of each of these task components resulted in mixing costs. Moreover, the level mixing costs and the judgment mixing costs were additive. This was interpreted as evidence that the costs of each task component were caused in different selection steps (see Sternberg, 1969). Thus, the mixing costs seem to reflect a stepwise selection process in which the relevant stimulus level is selected first and then the relevant judgment categories are chosen.

However, are these results indeed sufficient to justify such an interpretation? What predictions would be made if the idea of task set selection were to be applied to the paradigm of Hübner et al. (2001)? An answer requires a definition of what a task set actually means in this paradigm. As was mentioned above, task sets are assumed to represent all the relevant aspects of task execution. With regard to the two-component paradigm of Hübner et al., this implies that a task set consists of a certain stimulus level, as well as of a certain judgment. In other words, all the specific level and judgment combinations are represented as task sets. Consequently, a task set selection account should predict that the mixing costs will depend on the number of level/judgment combinations that can occur in a block of trials. This is reasonable, because the duration of selecting a certain task set should covary with the number of potential task sets.

This, however, casts doubt on the interpretation of Hübner et al. (2001). According to their hypothesis, each mixed task component produces mixing costs in its corresponding selection step. Consequently, the overall costs should result from the number of mixed task components. A problem arises from the fact that with an increasing number of mixed task components, the number of possible level/judgment combinations increases as well. Assume, for instance, that a block of trials consists of two mixed judgments but a constant stimulus level. This results in one mixed component but in two mixed task sets (e.g., local–parity and local–magnitude). If we additionally mix the levels in this block, we obtain two mixed components. At the same time, however, we double the number of mixed task sets from two to four (local–parity, local–magnitude, global–parity, and global–magnitude). This implies that the number of task sets is also predictive for the amount of mixing costs.

Experimental Approach

Fortunately, it is possible to resolve the confounding of the number of task sets and the number of mixed components. Basically, the confounding arises because, with each additional mixed task component, more task sets are also possible. However, the hypothesis of Hübner et al. (2001) does not imply that each possible task set actually has to occur in a block of trials. Assume that there is no integrated representation of level and judgment and that each of these task components is selected in a separate step. In this case, the mixing costs should not be affected by the specific level/judgment combinations that occur in a block. Rather, the costs should depend exclusively on whether more than one level and more than one judgment is relevant.

This offers the possibility of deconfounding the number of task sets and the number of task components. Consider the two conditions from the example above and assume that three judgments are used, instead of two. In this case, it is possible to realize a condition in which only the three judgments are mixed but the level remains constant. We will call this condition *1C3T*, because it consists of one mixed task component but three task sets. In the second condition, both task components, levels and judgments, are mixed. If each possible level/judgment combination can occur, this condition comprises six task sets. Accordingly, we will call this condition *2C6T*. The two conditions differ in both the number of mixed components and the number of task sets. However, the increase of task sets in the 2C6T condition is due merely to the fact that all possible level/judgment combinations are realized. It is also possible to construct a condition in which both task components are mixed, but without applying each combination from the two levels and the three judgments. For instance, one judgment could occur only in combination with one level, whereas the remaining two judgments are combined exclusively with the other level. This 2C3T condition comprises two mixed components but only three task sets.

These three conditions are sufficient for testing whether the number of mixed components or the number of task sets determines the mixing costs. Since the 2C3T and the 2C6T conditions differ only in the number of task sets, any differences in performance have to be attributed to this variable. On the other hand, the 1C3T and the 2C3T conditions differ only in the number of mixed components. Consequently, a performance difference between these two conditions implies that the number of mixed task components is responsible for this effect, rather than the number of task sets involved. Thus, the predictions of the two accounts for mixing costs are straightforward: If the mixing costs reflect task set selection, the former comparison (2C3T vs. 2C6T) should result in a significant effect, whereas the latter comparison (1C3T vs. 2C3T) should be significant if the number of mixed components causes the costs. Since the two hypotheses are not mutually exclusive, it is even possible that both effects will be observed. Experiment 1 was designed to test these predictions.

EXPERIMENT 1

The question addressed by this experiment was whether the number of task sets or the number of mixed task components is relevant with respect to mixing costs. As was described above, we constructed the 1C3T, 2C6T, and 2C3T conditions by adding a third judgment to the paradigm of Hübner et al. (2001). Although the general method was similar to that in Hübner et al., here we used an experimenter-paced preparation interval, instead of a self-paced one. Because it was unclear how the results would be affected by a self-paced preparation, we decided to use a fixed but rather long cue–stimulus interval in the present experiments. Furthermore, since Hübner et al. compared the overall means from the mixing conditions, the mixing costs were contaminated with shift costs. Therefore, in the present experiments, we compared only trials on which the target level, as well as the judgment, was repeated.

Furthermore, although we have no specific hypotheses concerning shift costs, for completeness we will, nevertheless, report them. According to the results of Hübner et al. (2001), we would expect that shifting the judgment and shifting the level should produce underadditive effects. We will discuss the relation between these effects and the mixing costs briefly at the end of the present article.

Method

Subjects. Twelve subjects (5 of them male and 7 female), who ranged from 17 to 30 years in age, participated in the experiment.

Apparatus. The stimuli were presented on a 21-in. color monitor connected to a personal computer, which served for controlling stimulus presentation and response registration.

Stimuli and Tasks. The stimuli were hierarchical structured forms (Navon, 1977), whose global shape was constructed from identical local elements in a 5 × 5 grid. At a viewing distance of 127 cm, the extent of the global numerals was 1.71° of visual angle horizontally and 2.34° vertically, and the extent of the local numerals was 0.23° × 0.34°. The stimuli were white on a black background. The forms and elements were numerals, ranging from 1 to 9, excluding 5.

The tasks required responding to the numeral at one of the two stimulus levels by applying one of three judgments. We used a parity judgment (odd or even), a magnitude judgment (less or greater than five), and an inward/outward judgment. In the latter, an inward response was required for the numbers 3, 4, 6, and 7, whereas an outward response was correct for the numbers 1, 2, 8, and 9. The subjects had to respond by pressing one of two response buttons with the index (even, less than five, and inward) or the middle (odd, greater than five, and outward) finger of the same hand.

To reduce the number of possible stimuli for each combination of level and judgment, we presented only stimuli for which the levels required different responses with respect to the indicated judgment (level-inconsistent stimuli; see Hübner et al., 2001).

Procedure. Each trial started with the appearance of a cue, which was centered on the screen and could have one of three forms and one of two sizes and colors. The parity judgment was indicated by an ellipse, the magnitude judgment by a square, and the inward/outward judgment by a triangle. The target level was indicated by the size and color of the cue. The size corresponded to that of the global stimulus shape or to that of one local element. In addition, the local level was indicated by a blue cue and the global level by a red cue. For instance, a small, blue ellipse indicated that an *odd/even* judgment was required for the value of the local numerals. We used two dimensions for indicating the target level because previous experiments had revealed that some subjects occasionally failed to prepare for the correct level when only the size of the cue served as an indicator. It seems that the size of the cue alone is not salient enough for ensuring reliable preparation. The cue was displayed for 1,000 msec. After that time, a blank screen appeared for 400 msec, followed by the stimulus, which was exposed at the center of the screen for 133 msec. The cue for the next trial appeared 1,000 msec after the response. Errors were signaled by a tone.

There were three conditions, which were defined in the following way.

1C3T. This condition comprised six blocks (each of 32 trials), in which the level was constant but all three judgments were randomized. In half of the blocks, the judgments had to be performed exclusively on the local level, whereas in the remaining blocks, only the global level was used.

2C6T. This condition comprised six blocks (each of 32 trials), in which the three judgments, as well as the target levels, were randomized.

2C3T. This condition comprised six blocks (each of 36 trials), in which two of the three judgments occurred at one level and the remaining judgment at the other level. Thus, one judgment was uniquely linked to a certain level. For instance, parity and inward/outward judgments occurred at the local level, and magnitude judgments occurred only at the global level. Each possible combination of a judgment and a level served as the unique level/judgment combination in one block.

Altogether, there were 18 experimental blocks, which were distributed over a 1-h main session. The order of blocks within this session was randomized. Before each block, the subjects were instructed which level/judgment combinations could occur in the block. They were given a sheet of paper on which the relevant combinations, together with the corresponding cues, were displayed. There was also a 1-h preliminary training session, in which each level/judgment combination was practiced in a block of 32 trials, followed by two randomly chosen blocks of each mixing condition. Altogether, there were 600 valid experimental trials for each subject. The expected number of trials with task repetitions differed in the three conditions, since a task repetition was two times more frequent in the 1C3T (64 trials) and 2C3T (72 trials) conditions than in the 2C6T condition (32 trials).

Data analysis. Only data from task repetition trials—that is, trials in which the level, as well as the judgment, was repeated—were used in the analyses of the mixing effects. Outliers were controlled by excluding trials with the 10% shortest and 10% longest RTs from each condition.

Results

Mixing costs. Mean RTs for repetition trials with correct responses from the 1C3T, 2C6T, and 2C3T mixing conditions were entered into a two-way ANOVA with repeated measurement on the factors of mixing condition and judgment. The factor level was not further analyzed, because this had already been done in detail in Hübner et al. (2001). The main effects of mixing condition [$F(2,22) = 9.52$, $p < .01$; see Figure 1] and judgment [$F(2,22) = 7.67$, $p < .01$] reached significance. RTs were shortest in the 1C3T condition (527 msec), followed by the 2C3T (586 msec) and the 2C6T (592 msec) conditions. Furthermore, inward/outward judgments led to longer RTs (623 msec) than did parity (545 msec) and magnitude (538 msec) judgments. However, the interaction between these two factors was not significant [$F(4,44) = 0.34$, $p = .85$].

To test our hypothesis directly, we made two pairwise comparisons of the mixing conditions by means of one-tailed *t* tests. First, we compared the 1C3T condition with the 2C3T condition in order to isolate the contribution of the number of mixed components to the mixing costs. A significant difference was detected [$t(11) = 5.58$, $p < .001$]. Second, we compared the 2C3T condition with the 2C6T condition in order to isolate the contribution of the number of task sets to the mixing costs. However, no significant effect was obtained in this case [$t(11) = 0.36$, $p = .36$].

The same analyses for the error rates revealed no significant effects. Most important, the error rates in our three main conditions were rather similar (1C3T, 8.4%; 2C3T, 6.7%; 2C6T, 7.4%).

Shift costs. In addition, we computed shift costs within each of our conditions. The mixing conditions differed with respect to the types of shifts that could occur. In the 1C3T condition, only the level could shift. In the 2C3T condition, the judgment alone or the judgment together with the level could shift. The latter was termed *double shift*. In the 2C6T condition, level shifts, judgment shifts, and double shifts could occur. The shift costs for the different types are summarized in Appendix A.

Discussion

The aim of the present experiment was to examine whether mixing costs depend on the number of mixed task

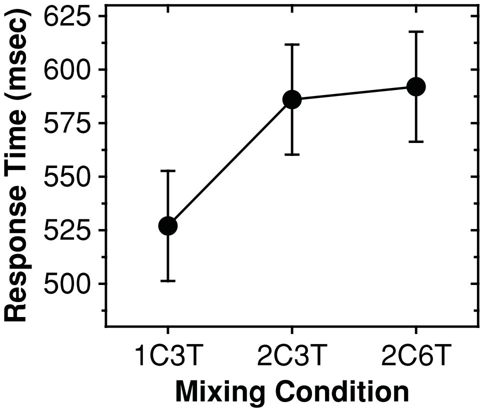

Figure 1. Experiment 1: Mean response times for repetition trials from the three mixing conditions. Error bars indicate 95% within-subjects confidence intervals for the main effect of the mixing condition factor (Loftus & Masson, 1994).

components or on the number of task sets, where the latter was defined by the number of relevant level/judgment combinations within a block of trials. Essential within our experiment was a condition that consisted of two variable components (two mixed levels and three mixed judgments) but only three task sets (the 2C3T condition). This condition was then compared with a condition with the same number of mixed task components but comprising six task sets (the 2C6T condition). There was no significant difference in performance between these two conditions. This clearly contradicts the hypothesis that the number of task sets in a block of trials affects performance. On the other hand, the comparison of the 2C3T condition with the 1C3T condition, which consisted of the same number of task sets but only one mixed component, revealed a significant difference. Thus, our data clearly show that the mixing costs were determined by the number of mixed task components. Two mixed task components produced higher costs than did one, irrespective of the number of task sets involved. This strongly supports the hypothesis that mixing costs reflect multiple selection steps, rather than a single task set selection stage.

Despite the clear results, the method we applied also has limitations. First of all, it relies on the assumption that our subjects adopted an optimal selection strategy. *Optimal* means that they focused exclusively on those level/judgment combinations that were emphasized by the instructions for the respective block of trials. However, this may not necessarily have been the case, because our conditions differed with respect to how easily the subjects could encode and hold the level/judgment conditions in working memory. For instance, it seems to be rather easy to remember that only the global level is relevant (1C3T) or that all level/judgment combinations could occur in a block (2C6T). On the other hand, it is presumably more difficult to memorize three specific level/judgment conditions that have been drawn arbitrarily from the whole set, as in the 2C3T condition. Thus, it cannot be excluded that the subjects applied the same strategy in this condition as in the 2C6T condition. Unfortunately, the assumption of such a general strategy can also explain why we observed the same performance for the 2C3T and the 2C6T conditions.

A second but related assumption was that our subjects used the relatively long cue–stimulus interval of 1,400 msec for optimal preparation. This seems to be justified, given the results of other studies (e.g., Rogers & Monsell, 1995). However, as De Jong (2000) suggested, even with such a long interval, preparation might occasionally fail. This idea could explain the performance difference between the 1C3T and the 2C3T conditions, if we assume that such preparation failures can also occur on task repetition trials. More specifically, our results could reflect the fact that an additional variable component increases the probability of a preparation failure. When only the judgment is variable, as in the 1C3T condition, only judgment preparation should be susceptible to failure, since the target level remains constant during the whole block. However, when both components are randomized, as in the 2C3T and 2C6T conditions, the subjects could fail to prepare for the judgment, as well as for the level.

Thus, our interpretation of the data is valid only if our subjects made optimal use of the strategies, both on the level of the block instruction and on the level of each single trial. Since we cannot be sure that these assumptions were met, a further experiment was designed.

EXPERIMENT 2

Our second experiment should replicate the results of Experiment 1, but with a between-subjects design. Each subject performed blocks of only one of the 1C3T, 2C3T, or 2C6T conditions. Under these circumstances, it should be less likely that the subjects would use a 2C6T strategy in the 2C3T blocks, when they had never practiced this strategy. Furthermore, the larger amount of data we could gather with this design offered the possibility of examining RT distributions for each condition. As we will argue, this should reveal whether our mixing effects were due to preparation failures.

Basically, there are two ways in which such a between-subjects design could be realized. One way would be that each subject performs only a single block type—for instance, blocks in which each judgment is applied only to the global numeral. This would require a very large sample, since there are nine different block types (two from the 1C3T condition, six from the 2C3T condition, and one from the 2C6T condition). Moreover, the frequency of the single level/judgment combinations would be different for each subject. Some subjects, for instance, would never perform a judgment on the global level. This would imply that our conditions would differ not only with respect to the selection strategy, but also with respect to the amount of interference that was caused, for instance, by the irrelevant level.

Because of these problems, we chose a design in which each subject was assigned to one of our conditions (1C3T, 2C3T, or 2C6T), and performed all corresponding block types. In this case, each level/judgment combination had to be performed equally often by each subject. Only the respective selection strategies differed. However, now the conditions differed with respect to the number of strategy changes. Since there were six block types in the 2C3T condition, the strategy had to be changed more frequently than in the 1C3T condition, which comprised only two block types. Moreover, the strategy never had to be changed in the 2C6T condition, since it included only a single block type. This could cause carryover effects from the preceding block to be more pronounced in those conditions in which the block types change more frequently. For instance, if the block type changed and a new strategy had to be applied, it could take some time to fully optimize this strategy.

To control for such carryover effects, performance was analyzed separately for the first, second, and third parts of the blocks, whose length was now increased from 32 trials to 96 trials. We expected that any carryover effects from the preceding block should be strongest in the first third of a block. Afterward, RTs should reach a steady state that reflects the current selection strategy. Accordingly, we expected that our predictions, which were similar to those in Experiment 1, should hold for the second and third parts of the blocks.

As has already been mentioned, a further aim of the experiment was to test whether the performance difference between the 1C3T and the 2C3T conditions was caused by an increased frequency of preparation failures in the latter condition. According to De Jong (2000), such an explanation predicts that our mixing effect should be present only on trials with slower responses, whereas it should be absent on trials with faster responses. This would be the result because trials in which preparation failed should produce longer RTs. As a consequence, these trials should be located at the end of the RT distribution at which responses were slower. Trials with successful preparation should produce shorter RTs and, accordingly, should be located at the end of the distribution at which responses were faster. Otherwise, if the mixing effect can be found in each region of the RT distribution, this would contradict the hypothesis that it is due to occasional failures to prepare.

Method

Three groups of 16 subjects participated in the experiment. The groups were balanced with respect to gender (7 of them male and 9 female in each group) and age (range, 19–33 years). The stimuli and procedure were similar to those in the previous experiment.

In contrast to Experiment 1, however, each subject worked exclusively through six blocks in one of our conditions. Furthermore, the length of the blocks was increased to 96 trials. Twelve subjects were assigned to the 1C3T condition, in which the level was constant but the judgments varied. Three blocks in which the local level was relevant and three blocks with a global target level alternated. Twelve subjects received blocks of the 2C3T condition. In each of these blocks, one of the level/judgment combinations was uniquely linked to a specific level. Finally, 12 subjects performed blocks of the 2C6T condition. At the beginning of the experiment, the subjects practiced each level/judgment combination in blocks of 32 trials in random order. In a second phase of practice, there were six blocks with 32 trials each of the respective condition. The total number of valid trials per subject was 576. The expected number of task repetition trials was 192 for the 1C3T and 2C3T conditions and 96 for the 2C6T condition.

Results

Time course. In a first analysis, we examined the time course of performance within the blocks. A separate analysis was computed for each of the three conditions. Mean RTs for repetition trials with correct responses were entered into a one-way ANOVA with repeated measurement on the factor of block section (Section 1, Section 2, or Section 3). In each of our three mixing conditions, at least a trend for this factor was observable (see Figure 2). Whereas performance decreased slightly in the 1C3T condition (Section 1, 635 msec; Section 2, 619 msec; Section 3, 624 msec) and the 2C3T condition (Section 1, 741 msec; Section 2, 717 msec; Section 3, 708 msec), performance increased in the 2C6T condition (Section 1, 684 msec; Section 2, 700 msec; Section 3, 707 msec). However, the effect was significant only for the 2C3T condition [$F(2,30) = 4.40$, $p < .05$]. Moreover, further testing showed that a significant decrease was present only from Section 1 to Section 2 [$F(1,15) = 4.94$, $p < .05$], but not from Section 2 to Section 3 [$F(1,15) = 0.49$, $p = .50$].

Mixing costs. The preceding analysis suggests that performance changed in the course of the blocks, at least in the 2C3T condition. However, even in this condition, asymptotic performance was reached in Section 2. As a consequence, we took the combined data from Sections 2 and 3 in order to analyze the mixing effects. Again, one-tailed *t* tests were used. First, we compared the RTs from the 1C3T and 2C3T conditions to obtain the effect of the number of components. Mean RTs were significantly shorter (621 msec) in the 1C3T condition than in the 2C3T condition (713 msec) [$t(30) = 1.76$, $p < .05$]. Second, the 2C3T and 2C6T conditions were compared in order to test the effect of the number of task sets. The analysis revealed that RTs did not significantly increase with the number of task sets [$t(30) = 0.15$, $p = .56$]. There was even a trend in the other direction (704 msec for 2C6T, 713 msec for 2C3T).

The same analyses were conducted for the error rates but revealed no significant effects. The error rates in the three mixing conditions were 6.9% (1C3T condition), 5.0% (2C3T condition), and 7.5% (2C6T condition).

RT distributions. RTs distributions were analyzed by comparing the estimated cumulative distribution functions (CDFs) for the three mixing conditions. Again, only task repetition trials in the second and third sections were used. In contrast to the preceding analyses, the data were not trimmed. In a first step, the mean RTs for 10 quantiles were computed separately for each combination of level, judgment, mixing condition, and subject. In a second step, the data were averaged across level, judgment, and subject. An inspection of Figure 3 shows three facts.

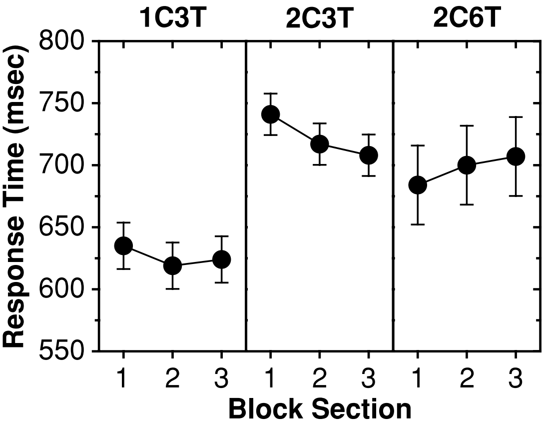

Figure 2. Experiment 2: Mean response times for repetition trials for the block sections of each mixing condition. Error bars indicate 95% within-subjects confidence intervals for the effect of block section (Loftus & Masson, 1994).

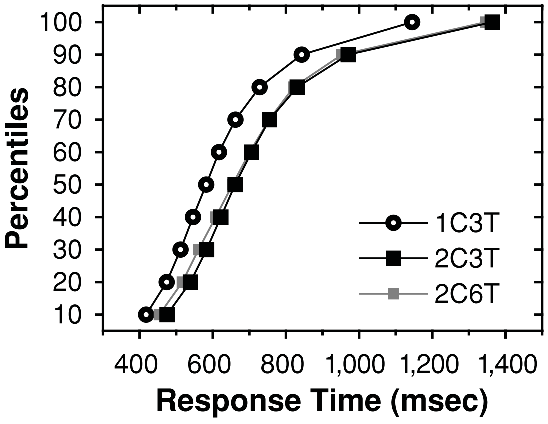

Figure 3. Experiment 2: Cumulative frequency distributions of response times for the mixing conditions.

First, the CDFs for the 2C3T and 2C6T conditions were approximately identical. Second, the CDFs for the 1C3T and 2C3T conditions differed even for the shortest RTs. This was confirmed when we compared the mean RTs for the fastest quantile for both mixing conditions [$t(30) = 2.43$, $p < .05$]. Third, the distance between the CDFs for the 1C3T and 2C3T conditions seemed to increase across the quantiles (57 msec in Quantile 1 vs. 219 msec in Quantile 10). However, the corresponding interaction between mixing factor and quantile did not reach significance [$F(9,270) = 1.08, p = .38$].

Shift costs. The same types of shift costs were analyzed as in Experiment 1. They are summarized in Appendix B.

Discussion

The aim of our second experiment was to replicate the results of Experiment 1 in a between-subjects design. This was done because we had to exclude the possibility that our results were caused by the subjects' failure to implement the instructions. It was argued that the results of Experiment 1 could reflect the fact that the subjects applied the 2C6T strategy also in the 2C3T blocks. Furthermore, we wanted to examine whether our results could be explained by the subjects' occasional failure to prepare on the single trials.

Each subject performed blocks of only one of the 1C3T, 2C3T, and 2C6T conditions. Unfortunately, a consequence of this design was that the three groups differed with respect to how frequently the strategy had to be changed between blocks. To exclude any sequential strategy effects at the beginning of each block, we first examined whether performance changed in the course of the blocks. The result was that performance was impaired at the beginning of the 1C3T and 2C3T blocks, although this was significant only in the 2C3T condition. Therefore, we excluded the first third of each block from further analysis.

However, by analyzing only the remaining data, we found the same results as in Experiment 1. Comparing the two conditions that differed only with respect to the number of relevant task sets (2C3T vs. 2C6T) revealed no significant effect. However, the two conditions that differed exclusively with respect to the number of mixed components (1C3T vs. 2C3T) were significantly different. It is unlikely that these results were due to the subjects' failure to implement the wrong selection strategy in the 2C3T condition. The subjects in this condition never practiced blocks in which all six level/judgment combinations were relevant. Thus, it would be massively counterproductive to adopt such a strategy.

To examine whether our effects were caused by the subjects' occasional failure to prepare for the indicated task, we analyzed the RT distributions for our three conditions. Such a hypothesis would predict that our mixing effect would occur only on trials with long RTs, whereas it should be absent on trials with fast responses. As our analysis revealed, the mixing effect was present in each region of the RT distribution. This suggests that this effect reflects a component that was present even in the fastest responses.

Taken together, the present results confirm our interpretation of the results of the first experiment: It is the number of mixed components that predicts the mixing costs. In other words, the costs reflect the mixing of single-task components, and not the mixing of whole task sets. However, one might still argue that our conclusions are too strong. What we actually showed is that performance decreases when the stimulus level is mixed. This does not imply that these costs are independent of the costs from judgment mixing. Rather, this assumption was taken from Hübner et al. (2001), who showed that the mixing costs of different task components are additive. However, as has already been mentioned, their experiments were slightly different from ours. They used two judgments, instead of three, and applied a self-paced procedure. Furthermore, mixing costs were computed by comparing the mean RTs for all trials from the various mixing conditions. Thus, it is not clear whether the additivity of mixing costs also holds for the present approach. Since it is important for our interpretation to show that additivity holds, a third experiment was conducted to clarify this question.

Since this experiment required a large number of mixing conditions, we again applied a within-subjects design. This seemed to be justified because, in Experiment 1, the results with a within-subjects design were no different from those in Experiment 2, in which a between-subjects design was used. Although Experiment 2 provided evidence for carryover effects from the preceding block, these effects seem not to affect the data pattern systematically when all block types are randomized in a within-subjects design.

EXPERIMENT 3

This experiment used a design similar to that in Hübner et al. (2001). However, as in Experiments 1 and 2, we applied three judgments and an experimenter-paced procedure. In each mixing condition, one, two, or three judgments were combined with one or two target levels.

As a consequence, mixing costs for level mixing and judgment mixing could be computed separately. Again, only task repetition trials were entered into the analysis. The crucial question was whether the mixing costs from level and judgment mixing are additive or not.

Method

Sixteen subjects (2 of them male, 14 female), who ranged from 20 and 28 years in age, participated in the experiment. The stimuli and the procedure were similar to those in the previous experiment.

Six different mixing conditions were examined. Each comprised six blocks with 32 trials, as follows.

1L1J. The target level, as well as the judgment, remained constant. In one of the blocks, parity judgments were required for the global target level, whereas in another block, the same judgment type had to be performed for the local level. There were corresponding blocks for the magnitude judgment and for the inward/outward judgment.

1L2J. The target level was constant, but two judgments were mixed randomly. In half of the blocks, the targets always occurred at the local level, and in the other half, they always occurred at the global level.

1L3J. The target level was constant, but all three judgments were mixed. In half of the blocks, the judgments had to be performed exclusively on the local level, whereas in the remaining blocks, only the global level was relevant.

2L1J. The target level varied across trials, whereas the judgment was held constant. There were two blocks for each judgment.

2L2J. The target level, as well as two judgments, were mixed. For each of the three possible pairs of judgments, there were two blocks.

2L3J. The target levels, as well as the three judgments, were mixed.

Altogether, there were 36 experimental blocks, which were distributed over two 1-h sessions. Mixing conditions and level/judgment combinations were counterbalanced across both sessions. The order of blocks within the sessions was randomized. There was also a 1-h preliminary training session, in which the 1L1J conditions were run first, followed by two blocks of each of the remaining conditions in a randomized order. Altogether, there were 1,152 valid experimental trials for each subject. The expected number of task repetition trials ranged from 32 in the 2L3J condition to 192 in the 1L1J condition.

Results

Mixing costs. Again, only the mean RTs for correct responses from repetition trials were entered into the data analysis. A two-way ANOVA with repeated measurements on the factors of level mode (constant or randomized) and judgment mode (constant, two randomized judgments, or three randomized judgments) was computed.

We obtained a significant main effect of judgment mode [$F(2,30) = 7.31, p < .01$]. The fastest responses occurred in conditions with a constant judgment (541 msec). Performance was reduced when two or three judgments were randomized (561 and 572 msec, respectively). Also, the level mode factor was significant [$F(1,15) = 18.4$, $p < .001$]. Responses under a constant target level were faster (541 msec) than those under randomized levels (575 msec). However, there was no significant interaction between level mode and judgment mode [$F(2,30) = 0.04$, $p = .97$; see Figure 4].

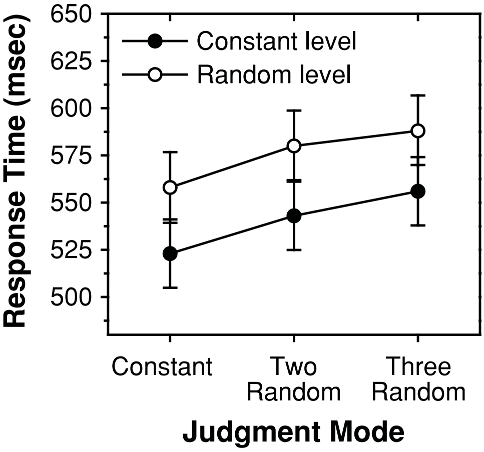

Figure 4. Experiment 3: Mean response times for repetition trials from the mixing conditions. Error bars indicate 95% within-subjects confidence intervals for the interaction effect (Loftus & Masson, 1994).

A planned contrast analysis of the main effect of judgment mode showed that the difference between conditions with constant and two randomized judgments was significant [$F(1,15) = 6.98, p < .05$], whereas the effect between conditions with two and three randomized judgments was not reliable [$F(1,15) = 1.64, p = .22$].

The same analyses were conducted with the error rates, but no significant effects were obtained. Blocks with constant level showed an error rate (5.2%) similar to that for blocks with randomized levels (5.3%). A small trend was observable for the main effect of judgment (constant judgment, 4.3%; two randomized judgments, 5.3%; three randomized judgments, 6.1%).

Shift costs. Shift costs were observable in five of our mixing conditions. Whereas only judgment shifts could occur in the 1L2J and 1L3J conditions, only level shifts were observable in the 2L1J condition. In 2L2J and 2L3J conditions, level shifts, judgment shifts, and double shifts were possible. A summary of all shift costs is given in Appendix C.

Discussion

This experiment was conducted to test whether additivity for mixing stimulus levels and judgments would also hold under the present experimental conditions. Since the additivity observed in Hübner et al. (2001) was obtained by using only two judgments and a slightly different procedure, it was not yet clear whether the effects of the individual task components would still be additive in the present situation. However, this was the case.

The observed additivity strongly supports our hypothesis that mixing costs are caused by two independent selection steps. No evidence was found, as in the preceding experiments, that mixing costs are related to the number of task sets in a block. Although there was a trend toward higher costs when three judgments were involved, as compared with the case in which two judgments were involved, this does not contradict our hypothesis. Such an effect does not reflect an effect of the number of task sets but, rather, a relation between the number of values (e.g., judgments) of each component and performance.

GENERAL DISCUSSION

The aim of the present experiments was to investigate two mechanisms that can account for the mixing costs in task shifting. Hübner et al. (2001) hypothesized that task execution under task shifting proceeds by a number of sequential selection steps. Selection steps are thought to be necessary to reduce the task conflicts that are caused by the usually employed bivalent stimuli. This conclusion was drawn from the observation that the mixing costs for individual task components were additive. Unfortunately, the mixing costs also varied with the number of relevant task sets in a block. Therefore, it is possible to explain the costs alternatively by a single selection step, in which a whole task set is selected. The present experiments, however, clearly show that this alternative does not hold.

Our hypothesis was tested in three experiments. Experiment 1 involved two conditions that differed in both the number of relevant task sets and the number of mixed components. These conditions were then compared with a third condition that included the same number of task sets as the first condition but the same number of mixed components as the second one. It turned out that only the conditions differing in the number of mixed components produced different mixing costs. On the other hand, the number of relevant task sets did not affect performance. This result implies that the number of mixed components, rather than the number of task sets, determines the mixing costs.

In Experiment 2, the same rationale was applied in a between-subjects design; that is, each subject was assigned to one of the three conditions. This should exclude the possibility that subjects also apply the 2C6T strategy to the 2C3T blocks. Since the subjects performing the 2C3T condition in Experiment 2 were never confronted with the 2C6T strategy, such a transfer should be less likely. Nevertheless, the results were the same as those in Experiment 1. Moreover, an analysis of the RT distributions showed that the observed mixing effect was present in all regions of the distribution. This contradicts a further alternative explanation that states that the observed mixing costs are due to occasional failures to prepare for the indicated task (De Jong, 2000).

Experiment 3 was conducted to determine whether the mixing costs of level and judgment mixing were still additive, even though three judgments were used, instead of two, as in Hübner et al. (2001). The results showed that this was indeed the case.

What Is the Nature of the Selection Processes?

Our results suggest that mixing costs do not reflect the selection of task sets or unitary task representations. Rather, they reflect the selection of representations that are closely linked to the mixed task components. In the following, we will describe a tentative model of the processes involved in our tasks. Basically, our main assumption is that selection processes establish constraints on response selection that ensure goal-directed responding, which has also been proposed by others (e.g., Mayr, 2001). However, most importantly, we believe that the selection of lower level task representations such as stimulus categories is sufficient to achieve this goal.

Simple models of forced choice tasks usually assume a single selection stage in which the response is selected. One could interpret response selection as a process in which a response category is selected that is activated by the stimulus. In our case, we additionally assume not only that the stimuli are associated directly with response categories, but also that these associations are mediated by other representations. More specifically, we assume that our hierarchical numerals are linked to representations of their global and local digit values (e.g., "3" and "7"). These representations are, in turn, associated with stimulus categories that correspond to the relevant judgments (e.g., *odd* or *even* from the parity judgment). Finally, these categories are associated with response categories (*left* or *right*). Note that each of these representations is directly or indirectly linked to one or more response categories. As a consequence, each of these representations can activate response categories, a mechanism that, potentially, induces a conflict at the response selection stage. However, selecting the task-relevant representations might be a strategy for avoiding strong activation by irrelevant representations of their respective response categories. We assume that in our experiments, this was done by a series of selection steps. At a first stage, the digit at the target level is selected. In a next step, the relevant stimulus category is selected. At a final stage, response selection takes place.

We believe that this model is reasonable and economical, because it does not require the additional assumption of competing higher control representations, such as task sets, for explaining the mixing costs. One merely has to assume that the representations of categories, such as numeral values or parity, are involved in task processing (see also Schuch & Koch, 2003). Reliable task performance is achieved by choosing an optimal sequence of selection processes that prevents irrelevant representations from affecting behavior. As a consequence, what actually controls behavior on a given trial is a sequence of lower level selection processes, rather than a top-down process that intervenes on a trial-by-trial basis in case of a conflict. This can be viewed as a principle of subsidiarity on the level of cognitive processes. Control is passed on to lower level processes whenever this is sufficient to enable goal-directed behavior. High-level processes are required mainly to plan and implement the structure or sequence of lower level processes.

An important question is how we can integrate the results concerning the shift costs into this model. In contrast to the mixing costs, the shift costs of the single-task components are not independent (see, e.g., Kleinsorge & Heuer, 1999). Hübner et al. (2001), for instance, obtained underadditive costs of level and judgment shifts. The present data partly support this observation. For instance, in the 2C6T condition in Experiment 1, a shift of both task components produced the same costs as did a judgment shift. This, however, is not surprising if we consider current theories on the origin of shift costs. Shift costs observed with a long preparation interval are commonly viewed as reflecting conflicts between task-related representations. For instance, shifting the level increases the conflict between the target and the distractor stimulus (e.g., Hübner, 2000; Ward, 1982). In contrast, shifting the judgment produces increased conflicts between stimulus–judgment associations (Waszak et al., 2003) and judgment–response associations (Meiran, 2000; Schuch & Koch, 2003). However, in contrast to the respective selection stages, these conflicts are not independent. For instance, an increased activation of the distractor stimulus also increases the activation of the corresponding stimulus–judgment associations (Steinhauser & Hübner, in press). As a consequence, a level shift not only causes level shift costs, but also affects the judgment conflict. Among other things, this could cause the underadditive interaction between level and judgment shift costs. However, more research is necessary to reveal the specific mechanisms that underlie these effects.

What Is Reflected by the Mixing Costs?

Independently of the question of whether task components or task sets are selected, one can ask which processes actually differ between our mixing conditions and, as a consequence, cause the mixing costs. Two hypotheses are conceivable. On the one hand, mixing costs could reflect an additional selection stage in the mixed task condition that is not present in the pure task condition (e.g., Mayr, 2001). On the other hand, Hübner et al. (2001) proposed that the mixing costs are due to the different durations of the same selection stages in pure and mixed task blocks.

Basically, the assumption of an additional stage receives support from our observation that rather similar costs are measured whenever more than one judgment occurs in a block, irrespective of the number of mixed judgments. However, the alternative view is also possible, as the following reasoning suggests. Both accounts would assume that the postulated selection stages are necessary mainly because the stimulus automatically activates representations of both tasks. In addition to the shift costs, this exogenous activation is also reflected by the so-called *congruency effects*. This refers to the reduced performance for stimuli that require different responses (incongruent) for the tasks, as compared with stimuli for which these responses are the same (congruent). In contrast to shift costs, congruency effects can be measured in both pure and mixed task conditions. Accordingly, it can serve as an indicator of task conflicts in both conditions.

Usually, it is observed that the congruency effects are also present in pure-task blocks, although they are often reduced, as compared with mixed task blocks (Hübner et al., 2001; Keele & Rafal, 2000). However, if task conflicts are also present under pure task conditions, it is conceivable that the additional selection stages are also necessary under these conditions. Rather than reflecting the number of additional selection stages, the mixing costs could be due to the different durations of these stages in both conditions, which might vary in adaptation to the actual or expected amount of task conflicts.

But how do the selection stages adapt to the requirements in the mixing conditions? In Hübner et al. (2001), it was speculated that it is the degree of attentional control that differs between the selection of a mixed or a pure task component. However, other mechanisms are also conceivable. For instance, a more conservative and error-resistant selection could be achieved by setting more or less conservative selection criteria. Subjects might accumulate as much evidence as they need to ensure that the correct component is selected. If the probability of task conflicts is high because the stimulus can activate irrelevant representations, a higher criterion is necessary to guarantee a reliable selection, at the cost of speed.

If this assumption is valid, the additivity of mixing costs from different task components could be viewed as an indicator of the mental structure controlling task performance. The costs themselves reflect strategic differences with respect to this structure. However, more research is necessary to reveal the specific type of strategies. Moreover, it is unclear how other demands, such as working memory load, are involved. Although Mayr (2001) provided evidence that working memory load cannot account for mixing costs, this was not shown for the present paradigm. For instance, if each variable component imposed additional load, this could have contributed to our results.

In conclusion, our empirical data provide evidence that mixing costs are related to sequential processing stages. We interpreted these stages as a selection strategy that reduces task conflicts in a stepwise manner. This enables reliable performance under conditions in which the same stimuli are associated with different responses, as in the task-shifting paradigm.

REFERENCES

Allport, A., Styles, E. A., & Hsieh, S. (1994). Shifting intentional set: Exploring the dynamic control of tasks. In C. Umiltà & M. Moscovitch (Eds.), *Attention and performance XV: Conscious and nonconscious information processing* (pp. 421-452). Cambridge, MA: MIT Press, Bradford Books.

De Jong, R. (2000). An intention-activation account of residual switch costs. In S. Monsell & J. Driver (Eds.), *Control of cognitive processes: Attention and performance XVIII* (pp. 357-376). Cambridge, MA: MIT Press.

De Jong, R. (2001). Adult age differences in goal activation and goal maintenance. *European Journal of Cognitive Psychology*, **13**, 71-89.

Dixon, P. (1981). Algorithms and selective attention. *Memory & Cognition*, **9**, 177-184.

Dixon, P., & Just, M. A. (1986). A chronometric analysis of strategy preparation in choice reactions. *Memory & Cognition*, **14**, 488-500.

Hübner, R. (2000). Attention shifting between global and local target levels: The persistence of level-repetition effects. *Visual Cognition*, **7**, 465-484.

Hübner, R., Futterer, T., & Steinhauser, M. (2001). On attentional control as a source of residual shift costs: Evidence from two-component task shifts. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **27**, 640-653.

Keele, S. W., & Rafal, R. (2000). Deficits of task set in patients with left prefrontal cortex lesions. In S. Monsell & J. Driver (Eds.), *Control of cognitive processes: Attention and performance XVIII* (pp. 627-651). Cambridge, MA: MIT Press.

Kleinsorge, T., & Heuer, H. (1999). Hierarchical switching in a multi-dimensional task space. *Psychological Research*, **62**, 300-312.

Kray, J., & Lindenberger, U. (2000). Adult age differences in task switching. *Psychology & Aging*, **15**, 126-147.

Loftus, G. R., & Masson, M. E. J. (1994). Using confidence intervals in within-subject designs. *Psychonomic Bulletin & Review*, **1**, 476-490.

Mayr, U. (2001). Age differences in the selection of mental sets: The role of inhibition, stimulus ambiguity, and response-set overlap. *Psychology & Aging*, **16**, 96-109.

Mayr, U., & Keele, S. W. (2000). Changing internal constraints on action: The role of backward inhibition. *Journal of Experimental Psychology: General*, **129**, 4-26.

Meiran, N. (2000). Modeling cognitive control in task-switching. *Psychological Research*, **63**, 234-249.

Navon, D. (1977). Forest before trees: The precedence of global features in visual perception. *Cognitive Psychology*, **9**, 353-383.

Norman, D. A., & Shallice, T. (1986). Attention to action: Willed and automatic control of behaviour. In R. J. Davidson, G. E. Schwartz, & D. Shapiro (Eds.), *Consciousness and self-regulation* (Vol. 4, pp. 1-18). New York: Plenum.

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

Rogers, R. D., Sahakian, B. J., Hodges, J. R., Polkey, C. E., Kennard, C., & Robbins, T. W. (1998). Dissociating executive mechanisms of task control following frontal lobe damage and Parkinson's disease. *Brain*, **121**, 815-842.

Ruthruff, E., Remington, R. W., & Johnston, J. C. (2001). Switching between simple cognitive tasks: The interaction of top-down and bottom-up factors. *Journal of Experimental Psychology: Human Perception & Performance*, **27**, 1404-1419.

Salthouse, T. A., Fristoe, N., McGuthry, K. E., & Hambrick, D. Z. (1998). Relation of task switching to speed, age, and fluid intelligence. *Psychology & Aging*, **13**, 445-461.

Schuch, S., & Koch, I. (2003). The role of response selection for inhibition of task sets in task shifting. *Journal of Experimental Psychology: Human Perception & Performance*, **29**, 92-105.

Steinhauser, M., & Hübner, R. (in press). Automatic activation of task-specific representations in task-shifting. *Memory & Cognition*.

Sternberg, S. (1969). The discovery of processing stages: Extensions of Donders' method. *Acta Psychologica*, **30**, 276-315.

Ward, L. M. (1982). Determinants of attention to local and global features of visual forms. *Journal of Experimental Psychology: Human Perception & Performance*, **8**, 562-581.

Waszak, F., Hommel, B., & Allport, A. (2003). Task-switching and long-term priming: Role of episodic stimulus-task bindings in task-shift costs. *Cognitive Psychology*, **46**, 361-413.

Wylie, G., & Allport, A. (2000). Task switching and the measurement of "switch costs." *Psychological Research*, **63**, 212-233.

NOTE

1. Hübner et al. (2001) termed these costs *residual shift costs*. However, applying the terminology we described above, we use the term *mixing costs*.

APPENDIX A
Experiment 1: Shift Types and Costs (in Milliseconds)

Condition	Shift Type	Shift Cost	t Value
1C3T	level	20	$t(11) = 2.44^{*}$
2C3T	judgment	14	$t(11) = 1.02$
	double	53	$t(11) = 2.31^{*}$
2C6T	level	36	$t(11) = 2.16^{*}$
	judgment	53	$t(11) = 2.17^{*}$
	double	53	$t(11) = 1.98^{*}$

$^{*}p < .05$ (one-tailed t test).

APPENDIX B
Experiment 2: Shift Types and Costs (in Milliseconds)

Condition	Shift Type	Shift Cost	t Value
1C3T	level	53	$t(15) = 4.15^{***}$
2C3T	judgment	79	$t(15) = 3.79^{**}$
	double	141	$t(15) = 4.75^{***}$
2C6T	level	66	$t(15) = 5.88^{***}$
	judgment	94	$t(15) = 4.59^{***}$
	double	141	$t(15) = 5.77^{***}$

$^{**}p < .01.$ $^{***}p < .001$ (one-tailed t test).

APPENDIX C
Experiment 3: Shift Types and Costs (in Milliseconds)

Condition	Shift Type	Shift Cost	t Value
1L2J	judgment	21	$t(15) = 4.02^{**}$
1L3J	judgment	14	$t(15) = 2.78^{*}$
2L1J	level	29	$t(15) = 3.77^{**}$
2L2J	level	44	$t(15) = 3.22^{**}$
	judgment	52	$t(15) = 3.24^{**}$
	double	71	$t(15) = 3.16^{**}$
2L3J	level	27	$t(15) = 2.57^{*}$
	judgment	44	$t(15) = 1.74$
	double	60	$t(15) = 2.76^{*}$

$^{*}p < .05.$ $^{**}p < .01$ (one-tailed t test).

(Manuscript received August 18, 2003;
revision accepted for publication January 2, 2005.)