

A Chain-Retrieval Model for Voluntary Task Switching

André Vandierendonck¹, Jelle Demanet², Baptist Liefoghe³,

and

Frederick Verbruggen⁴

¹ Department of Experimental Psychology, Ghent University, Henri Dunantlaan 2, B-9000 Gent, Belgium, E-mail: Andre.Vandierendonck@UGent.be

² Department of Experimental Psychology, Ghent University, Henri Dunantlaan 2, B-9000 Gent, Belgium, E-mail: Jelle.Demanet@UGent.be

³ Department of Experimental-Clinical and Health Psychology, Ghent University, Henri Dunantlaan 2, B-9000 Gent, Belgium, E-mail: Baptist.Liefoghe@UGent.be

⁴ Psychology, University of Exeter, Washington Singer Building, Perry Road, Exeter EX4 4QG, UK, F.L.J.Verbruggen@exeter.ac.uk

Running Head: Modeling Voluntary Task Switching

Address Corresponding Author: André Vandierendonck,
Department of Experimental Psychology, Ghent University,
Henri Dunantlaan 2,
B-9000 Gent,
Belgium
Phone: +32-(0)9-264 64 37, Fax: +32-(0)9-264 64 96
E-mail: Andre.Vandierendonck@UGent.be

Abstract

To account for the findings obtained in voluntary task switching, this article describes and tests the chain-retrieval model. This model postulates that voluntary task selection involves retrieval of task information from long-term memory, which is then used to guide task selection and task execution. The model assumes that the retrieved information consists of acquired sequences (or chains) of tasks, that selection may be biased towards chains containing more task repetitions and that bottom-up triggered repetitions may overrule the intended task. To test this model, four experiments are reported. In Studies 1 and 2, sequences of task choices and the corresponding transition sequences (task repetitions or switches) were analyzed with the help of dependency statistics. The free parameters of the chain-retrieval model were estimated on the observed task sequences and these estimates were used to predict autocorrelations of tasks and transitions. In Studies 3 and 4, sequences of hand choices and their transitions were analyzed similarly. In all studies, the chain-retrieval model yielded better fits and predictions than statistical models of event choice. In applications to voluntary task switching (Studies 1 and 2), all three parameters of the model were needed to account for the data. When no task switching was required (Studies 3 and 4), the chain-retrieval could account for the data with one or two parameters clamped to a neutral value. Implications for our understanding of voluntary task selection and broader theoretical implications are discussed.

Keywords: voluntary task switching, task selection, cognitive control, random generation

A Chain-Retrieval Model for Voluntary Task Switching

1. Introduction

Goal-directed behavior relies on a determination to achieve the current goal, whilst being adaptive to changes in the environment. This balance between goal persistence, on the one hand, and goal flexibility, on the other hand, is achieved by means of executive control (e.g., Logan & Gordon, 2001; Norman & Shallice, 1986). Executive control processes are required to ensure that behavior consistent with the goal is produced, while remaining sensitive to changes in the environment. Such changes may require a shift to another goal, in which case the system must be flexible enough to release the previous goal and install a new one adapted to the environmental change. The present study focuses on an essential feature of such cognitive flexibility, namely the processes underlying the behavioral choice between different goals or tasks.

Task switching has been the preferred paradigm to study such flexible changes in the laboratory (see Kiesel et al., 2010; Monsell, 2003; Vandierendonck, Liefoghe, & Verbruggen, 2010, for reviews). Many studies have shown that task switching comes with a cost, which is indicated by longer reaction times and higher error rates on task-switch trials than on task-repeat trials. This switch cost has been attributed to task-set reconfiguration processes, interference, or both (e.g., Allport, Styles, & Hsieh, 1994; Mayr & Kliegl, 2000; Meiran, 1996, 2008; Rogers & Monsell, 1995; Waszak, Hommel, & Allport, 2003). In most task-switching studies, participants are instructed when to repeat and when to switch. Behavioral flexibility, however, also involves the possibility to *choose* voluntarily for a particular course of action or for a particular goal. Such goal-selection or task-selection processes can be examined with the voluntary task switching (VTS) procedure (Arrington & Logan, 2004, 2005). This procedure seems to tap into the same behavioral mechanisms that are studied in the traditional task-switching procedures as testified by the finding that task switch costs are driven by top-down (Arrington & Logan, 2005; Mayr & Bell, 2006) as well as bottom-up factors (Arrington, 2008; Arrington & Rhodes, 2010; Mayr & Bell, 2006; Yeung, 2010). Importantly, the VTS procedure was designed to allow voluntary choice or selection of a particular task or task goal by giving subjects the *freedom to select and execute the task of their choice on every trial*. This makes it an interesting procedure because it not only provides the usual task-

performance measures, such as the switch cost, but also enables investigation of the characteristics of the task choices made over a series of trials. The latter feature allows for the study of *processes involved in choosing or selecting a task* to be performed. Understanding the processes underlying voluntary choice of courses of action is of paramount theoretical importance. even though thus far not so much is known about these processes.

Previous research using the VTS procedure has shown that people tend to repeat the same task more often than expected on the basis of chance (Arrington & Logan, 2004). This is known as the *task-repetition bias*, and consists of a preference of task repetitions over task switches. This repetition bias becomes weaker when more time is available to select a task (i.e., when the response-stimulus interval (RSI) increases; Arrington & Logan, 2005). This observation suggests that choosing to perform a task depends—at least to some extent—on endogenous (top-down) time-consuming task selection processes (see also e.g., Arrington & Logan, 2005; Arrington & Yates, 2009; Liefoghe, Demanet, & Vandierendonck, 2009). However, bottom-up factors clearly play an important role as well, especially when the amount of time available to select a task is small. This has been shown in several ways: the task repetition bias is affected by priming due to stimulus repetitions (Mayr & Bell, 2006), stimulus availability (Arrington, 2008), previously learned stimulus-task associations (Arrington, Weaver, & Pauker, 2010; Demanet, Verbruggen, Liefoghe, & Vandierendonck, 2010), processing efficiency (Arrington & Rhodes, 2010), and differences in task difficulty (Liefoghe, Demanet, & Vandierendonck, 2010; Yeung, 2010). Furthermore, the presence of a working memory load that interferes with endogenous control processes increases the task-repetition bias in the presence of stimulus repetitions (Demanet et al., 2010). In sum, task choice seems to result from an interaction between exogenous (bottom-up) influences and endogenous (top-down) cognitive control operations.

The present study aims to contribute to our understanding of the processes involved in task choice by elaborating and testing a model of task choice in a context in which participants voluntarily and frequently switch between tasks. When participants are free to select any of a range of tasks for execution, in the absence of specific constraints, task switches are rarely selected. In a study of Kessler, Shencar and Meiran (2009), for example, proportions of

spontaneously selected switches were rather low (between 0.04 and 0.13). Because the switch cost is measured by comparing repetition and switch trials, for statistical reasons a less skewed distribution of switches and repetitions is needed. In order to approach a more balanced distribution, instructions in VTS typically stress that the tasks should be selected (and performed) about equally often, and that tasks should be selected independently from the previous trial so as to form an unpredictable series of choices. This is usually clarified by the coin-tossing metaphor (imagine that on each trial, a coin with one task name on one side and the other task name on the other side is flipped to decide which task to perform). These instructions have been frequently used in VTS and result in distributions with 20 to 50 percent switches.

Under the constrained instructions, voluntary task choice seems similar to free and independent selection of events and responses, which has traditionally been studied in the human random-generation literature. In a typical random-generation experiment, participants are requested to generate sequences of outcomes that would be obtained by repeatedly tossing a coin or throwing dice. Furthermore, the heuristics of availability (Baddeley, 1996) and representativeness (Rapoport & Budescu, 1997) seem to underlie random event-generation and voluntary task choice (see Arrington & Logan, 2005, p. 699). This suggests that independent selection of tasks and independent selection of events are based on common underlying processes.

In view of all this, our purpose was to develop a model that not only accounts for voluntary task choice behavior but also for independent event selection. Although this may seem straightforward, the model must be able to account for some differences between voluntary task selection and independent event generation. A first important difference between the selection of tasks and the selection of events is that patterns of task selection show a repetition bias (see above), whereas other patterns of event selection typically show an alternation bias which is a tendency to generate more alternations than repetitions and to produce more frequently short runs of repetitions of the same event (Lopes, 1982; Lopes & Oden, 1987; Neuringer & Allen, 1986; Rapoport & Budescu, 1992; Treisman & Faulkner, 1987; Wagenaar, 1972). A second difference concerns the context in which selection of tasks and events takes place. In VTS, the selected task

has to be applied, whereas no further consequences are associated with the selection of an event. Because in VTS the selected task has to be executed, bottom-up factors that are external to the process of task selection, such as stimulus repetition priming, may directly influence the selection process (Demanet et al., 2010; Mayr & Bell, 2006; Orr & Weissman, 2011). Furthermore, variations in the demands placed on top-down control during the execution of a task may equally affect the selection process. If the selected task is likely to result in a performance error or takes more time to execute, the probability of selecting this particular task may become smaller. The experience that a particular task is easier (i.e., fewer errors and shorter RTs) than another one, may affect the decision to repeat that task or to switch to the other task (Liefoghe et al., 2010). Similarly, task repetitions are likely to be perceived as easier than task switches (also in VTS, Arrington & Logan, 2005), which could explain the preference for task repetitions over task switches. The preference for selecting the easier tasks or choosing task repetitions is a case of going for the least mental effort (Botvinick, 2007; Hull, 1943), which can be defined as “a preference for activities or strategies that minimize cognitive demand” (Botvinick & Rosen, 2009, p. 835).

Thus, our model assumes that task selection and event selection are based on the same selection mechanism. However, these mechanisms are biased by events external to selection, namely experience-based preferences for particular tasks or events, and bottom-up triggered intrusions leading to repetition of the same task or event. Such biases are more likely to occur in VTS than in event selection, because participants have to execute the selected tasks in the VTS paradigm.

2. Modeling task selection processes

For the mechanism common to task and event selection, already existing models of event selection may provide a useful basis. Some of these models use Markov chains to account for biased event selection (Budescu, 1987; Vandierendonck, 2000a). Such models—in particular those that can predict perseverations—could be useful to describe what goes on in VTS.

However, it will become clear later in this article that these models are insufficient to account for task-choice processes. A completely different type of model was proposed by Rapoport and

Budescu (1997). It is most explicit in specifying the processes involved in selection. According to this model, people do not represent all possible combinations of events that may occur when generating sequences of independent events, but only the events that are believed to be representative of the situation at hand. This particular feature may be considered as a bias in conceptualization of the situation; it can also be interpreted as a consequence of limited cognitive capacity to overview all possibilities. This model further assumes that events are selected while a monitoring process follows retrospectively within a window of a particular width (capacity constraint) whether the part of the generated sequence within that window deviates from independence. When the monitoring process detects a deviation from this subjective idea of independence, the next event will be selected so as to restore the person's conception of statistical independence within the window. This model provides an excellently fitting description of some forms of independent event selection. However, because task selection in VTS has been shown to be susceptible to bottom-up intrusions (e.g., Arrington et al., 2010; Demanet et al., 2010; Mayr & Bell, 2006), application of the model of Rapoport and Budescu to task choice in VTS leads to some difficulties. The intruding bottom-up events, which essentially consist of repetitions of the same task executed on the previous trial, should show up in the window of generated tasks, and consequently, the monitoring mechanism should detect these occurrences as deviations from the person's conception of statistical independence (representativeness). This should then result in a corrective action, namely the selection of as many task switches as needed to restore the sequence, and an alternation tendency would still be present in the end. Clearly, the problem arises because the monitoring mechanism of the model inspects *retrospectively* the recent window of the sequence. The problem does not arise, however, if the generation mechanism would elicit prospectively the presence of repetitions by selecting series of independent events in advance. In what follows, we describe a model of task selection in VTS that replaces the retrospective monitoring mechanism in the Rapoport and Budescu (1997) model by a mechanism that selects a sequence of events in advance.

2.1. Modeling assumptions

The model proposed here builds on the idea of Rapoport and Budescu (1997) that when humans freely select independent events, they do not consider all possible combinations of events; instead they rely on the representativeness of the sequence, which means that they prefer sequences that are balanced (equiprobable events) and that the order in which events appear is representative of randomness (unsystematic orderings). In line with the considerations regarding the differences between event and task selection, we further adopt the hypotheses that (1) task performance difficulty affects the outcomes of the task-selection mechanism (Liefoghe et al., 2010; Yeung, 2010) and (2) that bottom-up events such as repetition priming compete with the top-down task-selection processes (Demanet et al., 2010; Mayr & Bell, 2006). First, we describe the event/task selection mechanism. Next, we discuss how the outcome of this mechanism can be modified by task difficulty and bottom-up events.

In line with the view of Arrington and Logan (2005), the instantiation of the task-selection mechanism calls upon the availability heuristic. This heuristic was used in studies of independent event selection and entails that the selected tasks are retrieved from long-term memory (cf. Baddeley, 1996). Because most studies of task switching use only two tasks, here we restrict the development of the model to situations in which there are two tasks. Selection of one task (say task A) on one trial may prime retrieval of the same task again on the next trial (which affects again availability; e.g., Baddeley, Emslie, Kolodny, & Duncan, 1998). In order not to end up in long and dependent series of the same task, the tendency to select task A must occasionally be suppressed in order to allow selection of the other task (B) and achieve more or less independent events (see Mayr & Bell, 2006). Soon, participants could then learn that they could work with short series of tasks, such as AABB, ABBA, and ABAB, to generate task sequences. Because of the large amount of mental effort required on each trial to select a new task independently and to avoid repetitions by priming while being engaged in executing the tasks, participants will be motivated to learn action sequences that are less effortful. Working with short sequences diminishes the burden of checking independence of the selected tasks on each trial. We propose, therefore, that during the early stages of the experiment, the participants learn to select such short sequences by

retrieving a sequence from long-term memory (LTM) on the basis of the task names (cf. Schneider & Logan, 2007). Initially, it may be necessary to produce the sequences one element at a time. As practice progresses, the sequences are stored in LTM as a chain of task names and participants would soon use such sequences to guide task-selection behavior.

These chains are continuously applied throughout the experimental session, but their selection from LTM is biased by a number of factors. The following assumptions are central to the development of the model.

(1) *Subset of chains.* In accordance with the central hypothesis of Rapoport and Budescu (1997), only a subset of all possible two-task sequences of a certain length are stored (the example above shows the subset for sequences of length 4): each of these sequences is balanced so that each of the two tasks is equally frequent. Due to these characteristics, these sequences are biased towards an alternation bias as predicted by the model of Rapoport and Budescu (1997).

(2) *Preference for 'easier' chains.* The application of the generated chain of tasks to the targets may have implications for preferences among chains such that tasks that elicit fewer errors and that are executed more swiftly (easier tasks) are preferred to tasks that lead to more errors or that require longer response times, i.e., tasks that are more difficult (Liefoghe et al., 2010). Similarly, some transitions between tasks are easier in that they take less time and evoke less errors; these 'easier' transitions are preferred to difficult transitions. As task switches are generally more difficult than task repetitions, we assume that the experience of differences in difficulty between tasks or between task transitions will bias the selection of chains from LTM towards chains that are on average easier. Because the switch cost is a general phenomenon in task switching, in the model we will only consider the preference for chains that can be executed with less effort, namely the chains with less switches (cf. the law of least mental effort; Botvinick, 2007; Botvinick & Rosen, 2009). This principle leads to a preference of chains like AAABB (only one switch) over AABAB (three switches). It is also possible to further distinguish between switches that are easier and those that are more

difficult (Kessler et al., 2009; Yeung, 2010), but in the present paper we will ignore these as well as the difficulty differences at the level of tasks.

(3) *Limited working memory capacity.* In order to monitor the produced sequence of events, the model of Rapoport and Budescu (1997) assumes that a window of the sequence is maintained in short-term memory. Our model does not call on monitoring processes, but there is still a window of generated events that has to be maintained in working memory to guide task selection over the next trials. We assume, therefore, that the size of this window is constrained by working memory capacity (e.g., Baddeley et al., 1998; Baddeley, 1966). Such constraint is also evident in the experiment on the task span reported by Logan (2004b) which revealed that the task span (the number of tasks that could be remembered and correctly executed) was essentially not different from the memory span (the number of tasks that could be remembered). This proposal is also consistent with the findings reported by Bryck and Mayr (2005) which suggest that verbalization of sequential information in task switching is critical for maintenance of sequential courses of actions or sequential plans. These and other studies (Baddeley, Chincotta, & Adlam, 2001; Emerson & Miyake, 2003; Miyake, Emerson, Padilla, & Ahn, 2004; Saeki & Saito, 2009) strongly indicate that working memory forms a basis for task switching without implying that shortage of working memory capacity would completely disrupt task-switching performance.

(4) *Coupling chains.* Thus far we propose that short chains of tasks are retrieved from LTM. After using a chain to guide task execution on the series of targets presented, a new chain must be retrieved. The issue here is how the new chain is coupled to the previous one. Given that priming may violate representativeness of the produced sequence (cf. *supra*), this can be achieved in three different ways. One possibility is that the next chain is concatenated to the previous one (e.g., AABB followed by ABBA leads to AABBABBA). In the example, the end (B) of the first chain is coupled to the head (A) of the second chain, resulting in a task switch. If the second chain started with a B, a repetition would have occurred. Ideally, each time a new chain is selected, this would occur independently taking into account the preference for the easier chains. Thus repetitions and switches would be added in a balanced

way. However, such a balance is difficult to achieve because the last task executed may prime chains that begin with the same task (cf. supra). To ensure a balance of repetitions and switches, top-down control over such repetition priming would be needed. A second possibility avoids the need for top-down control, so that the last element of the chain primes chains that begin with the same element. Combined with simple concatenation, this would result in the addition of a repetition at every chain junction. This systematic addition of a repetition between every pair of chains would result in an overall sequence that violates the representativeness principle that lays at the basis of each individual chain (Rapoport & Budescu, 1997). There is a third possibility that does not have the drawback of the previous one. It shares the priming assumption of the previous possibility, but when the last task of the chain is reached (the last task to be executed), this task is not executed but instead primes the chains starting with the same task. In other words, arriving at the last task (A), this task primes chains that start with an A, and thus chains such as ABBA, ABAB and AABB may become active. This simplifies selection of a chain as only half of the chains stored in LTM are activated. One of these is selected, and used to guide behavior in the following trials. For example, at the end of the chain BBAA, the chain ABAB may now be selected, resulting in the execution of BBA(AA)BAB, where the task between parentheses is executed only once as it was used to prime the next chain. Because only the last of the three possible chain coupling methods guarantees maintenance of the representativeness of the composite sequence without the need for extra assumptions about top-down control, this coupling is preferred in the present modeling.

2.2. Elaboration

These assumptions are now elaborated in a framework that defines a family of formal models. We will define three parameters, m , b , and r , which express quantitatively the operation of underlying processes.

The first parameter, m , expresses the chain length that the person prefers. As explained above, the chain length is partly constrained by available working memory capacity, but as will become clear later on, other factors also affect chain length. Nevertheless, based on assumption 3

above (limited working memory capacity), the parameter m specifies the length of the chains stored in and retrieved from LTM. We assume that the minimum length is 3. Chains of two elements are of course possible, but chains of this length do not allow enough variability: only AB and BA are balanced sequences, and because such sequences do not contain repetitions, a repetition bias based on experience with the easier repetitions cannot develop (see assumption 2: preference for easier chains). Therefore, we adopted 3 as the lower limit. Furthermore, because many studies show that working memory span is limited to about 5-6 elements (e.g., Atkinson & Shiffrin, 1968; Baddeley, 2007; Baddeley & Hitch, 1974), we adopted 6 as the maximum limit; Table 1 displays the sequences that are possible at each of the lengths 3-6. The sequences considered are balanced in the sense that the two tasks occur $m / 2$ times in the sequences with an even length and occur minimally $m / 2$ times and maximally $m / 2 + 1$ times in odd numbered sequences. This way, both tasks will be selected approximately equally often. At the same time, sequences show, on average, a tendency to alternate, which is shown in Table 1 in the column labeled “# Rep”. We assume that all the sequences of the person’s preferred length m are stored in LTM.

---- Table 1 about here ----

The second parameter, b , is related to the strength of a chain in accordance with assumption 2 (preference for easier chains). In order to model the variability in the retrieval of chains, each chain is supposed to have an associative strength that specifies its likelihood of being retrieved. The associative strength is a common way to represent effects of learning (e.g., Anderson & Lebiere, 1998; Bush & Estes, 1959; McClelland & Rumelhart, 1986; Rescorla, 1988; Rumelhart & McClelland, 1986). In situations where feedback is explicitly provided, the associative strength of a memory trace is changed in accordance with the feedback. If the memory retrieval is “correct” or leads to a positive outcome, its strength is increased; if the retrieval is “incorrect” or leads to a negative outcome, the strength is decreased. In task selection no explicit feedback is provided. However, implicit feedback still remains present because behavioral consequences of task execution may be used to strengthen a chain. Thus chains with more task repetitions may be

experienced as easier so that these chains are strengthened more frequently, and the degree of strengthening will be larger the more repetitions the chain contains.

To implement the idea of chain strength or chain weight, w , we define a parameter $0 < b < 1$, which is the bias of a repetition in the stored sequence ($1 - b$ is the likelihood of a switch). The value of b determines the weight (w) of the chain by the simple multiplicative rule $w = b^R(1 - b)^S$, where R and S are respectively the numbers of repetitions and switches in the chain. With a given value of b , the larger the proportion of repetitions in the chain, the larger the weight of the chain becomes. With a fixed number of repetitions in the chain, the larger b , the larger the weight of the chain will be. In other words, b represents an acquired *bias* towards more repetitions. This way, the consideration is implemented that experience with execution of the tasks (repetitions are easier than switches) will influence the retrieval of task chains. The parameter b indicates the probability of a single element within a sequence of tasks and w defines the overall strength of a sequence.

The sequence AABB, for example, contains two repetitions and one switch: thus $w = b^2(1 - b)$.

Each time a new chain has to be selected, the chains compete for being selected by producing an amount of activation that is proportional to their strength. Using Luce's choice rule (Luce, 1959), the probability that a particular sequence i with weight w_i will be selected, depends on the relative degree of activation with respect to the total amount of activation of all chains. Let W be the sum of the weights w of all the sequences in the set, the probability to select a particular sequence i with weight w_i is then w_i / W . The underlying process is thus simply based on competition of the alternatives for selection.

Whereas b expresses a bias towards chains with more repetitions, its value determines the probability (p) of *top-down selection* of a repetition. For example, for chains of length 3 as described in Table 1, if $b = 0.5$ (i.e., repetitions and switches have an equal weight), the overall likelihood of a repetition, $p = .33$, because 1/3 of the transitions within this set are repetitions. If the value of b were 0.7, the weights of the three chains would be respectively 0.21 (0.7×0.3), 0.21 (0.7×0.3), and 0.09 (0.3^2). Given the weight of the first chain (0.21) and the fact that half of the events in that chain are repetitions, the weight of a repetition in that chain is 0.21×0.5 . The

overall probability of a repetition, p , would thus be $(0.21 \times 0.5 + 0.21 \times 0.5)/(0.21 + 0.21 + 0.09) = 0.41$. The set of chains intrinsically constrain the maximum value of p ; for example, in the set of chains with size 3 (2 transitions), p can never become larger than 0.5 (when b approaches 1) because each chain always contains at least one switch. The longer the chain, the larger p can become.

The third parameter, r , is related to bottom-up priming. This parameter is introduced because several studies have shown that bottom-up events, such as stimulus repetitions (Demanet et al., 2010; Mayr & Bell, 2006), stimulus availability (Arrington, 2008), and processing efficiency (Arrington & Rhodes, 2010) affect task choice in voluntary task switching. The probability r , then, represents the likelihood of a *bottom-up event* overruling planned or intended task selections. We assume that with an unknown probability events occur that activate a tendency to repeat the same task. On trials on which a repetition is intended, this parameter has no effect, but on trials on which a switch is intended, this leads to a conflict between the tendency to switch to another task and the tendency to repeat the previous task. On some trials this conflict will be resolved in favor of the intended action (switch), on other trials, the triggered repetition will win the competition; we assume the latter occurs with a probability r . Importantly, in combination with the b parameter, this r parameter could account for the task-repetition bias typically observed in VTS. Several choices as to what happens with the already selected chain are possible; in the present model, the primed repetition overrules the current event in the selected chain, and execution of the retrieved chain continues after the intrusion. The reason for implementing an overruling mechanism in the model is based on the rationale that the intended task (i.e., the task specified in the chain) has been selected, but then loses the competition with the primed task repetition. In testing the model, other possible choices will also be considered.

Thus far this *chain-retrieval model* specifies that short balanced sequences (chains) avoiding too many repetitions are stored in memory, and have a strength that determines the probability of being selected. This model has three important distinctive characteristics. First, the model specifies a selection mechanism that is based on retrieval of short chains of events from LTM (see Table 1). Because the chains stored in LTM are in agreement with the representation bias

specified by Rapoport and Budescu (1997), the produced sequences deviate from statistical independence without the need for a retrospective monitoring mechanism to guarantee this. Second, because the model allows the selection mechanism to be sensitive to the difficulty of the transitions (the b parameter coupled to strengths of chains on the basis of the relative difficulty of the task transitions), it may be useful to predict the characteristics of the sequence of task names as well as the characteristics of the sequence of transitions¹. The advantage of this feature will become clear in the application of the model to the data. The third characteristic of the model allows that bottom-up events in the form of task repetitions could gain control over task selection (with a probability r). This is in line with the nature of the bottom-up events reported in the literature (Demant et al., 2010; Mayr & Bell, 2006; Vandamme, Szmalec, Liefoghe, & Vandierendonck, 2010).

The fundamental assumptions adopted allow for some freedom of implementation in a formalized model, which suggests that it is possible to build similar models with slightly different implementations of some assumptions. Appendix B contains an overview of variations in the implementation and how they fare when used to fit actual data. As these variations do not result in better fits and predictions, we do not consider them in the main text. The purpose of the present investigation was to test the usefulness of this framework and its formalization. Because the framework generates a series of hypotheses about processes involved in selection of independent events as well as voluntary task selection, an exhaustive test of all these hypotheses is beyond the scope of this article. For that reason, we focus on testing how well the presented model can account for voluntary task selection and independent-event selection performed by participants in experimental settings. In a first study, the model is applied to data of 17 participants in voluntary task switching based on the standard procedure (Arrington & Logan, 2004). In a second study, the model is applied to data of 80 participants in a voluntary task switching study based on the double registration procedure (Arrington & Logan, 2005). This second study is also used to test the assumption that the length of the sequences stored in LTM depends on experience over trials, and the simplifying assumption that task selection over an experimental session is quite stable. The

¹ For this reason the model could also be framed as a transition-chain retrieval model.

third study presents a fit of the model to a situation where a response hand is selected for execution of one single task, and the fourth study presents a fit of the model when participants are required to independently generate events instead of tasks. In each case, we describe the experiment and its results, and then we apply the model to the reported data.

3. Study 1

As a first test of the modeling, we focus on the task choice data of an experiment that has been published before (Liefoghe et al., 2009). This experiment used the standard method of voluntary task switching, where one single key-press informs about the task selected and the categorization of the target (Arrington & Logan, 2004, 2005). This version of the procedure has been shown to be very sensitive to bottom-up triggered repetitions (Demanet et al., 2010; Mayr & Bell, 2006). Because the study has been published, we briefly describe the methods and only report the results that are relevant for the present purpose.

3.1. Method

Eighteen first-year psychology students of Ghent University participated for course requirements and credit. They were assigned to one of two between-subjects conditions, a condition with an RSI of 100 ms or a condition with an RSI of 1000 ms. Nine other students in the original study were assigned to a within-subjects condition. These participants were not included in the present study, because the RSIs were varied randomly from trial to trial, leading to confounding of the manipulation with other factors affecting task choice. Data for one subject (in the RSI-100 condition) were not included in the present analysis, because the proportion of selections of one of the tasks amounted to .85.

Stimuli consisted of letters (A, E, B, or D) surrounded by a geometric figure (circle, ellipse, square, or rectangle) to which either a letter categorization task (consonant vs. vowel) or a form categorization task (quadrangle or ellipsoid) could be applied. A keyboard was used to register the responses. The form task was performed with the left hand (keys f or g) and the letter task was performed with the right hand (keys j or k).

Participants were tested individually on a PC with a 17-inch color monitor running Tscope (Stevens, Lammertyn, Verbruggen, & Vandierendonck, 2006). Instructions of Arrington and Logan

(2004) were used. In each condition, participants received one practice block and four blocks of 64 test trials each. A trial started with presentation of the stimulus in the center of the screen until a response was given within a response deadline of 3000 ms. Immediately after the response or when the response deadline was attained, the stimulus disappeared. After an incorrect response, the screen turned red for 200 ms before the RSI started (100 ms in one condition; 1000 ms in the other).

3.2. Data analysis

The analysis focuses on the series of task choices and task transitions. As the task choices and the transitions within the sequence of tasks are binary events, the sequence of events can be expressed as a series of binary digits (0 or 1). A sequence of tasks is a series of task names; using the letters L (letter categorization) and F (form categorization), an example of a series of selected tasks may be LLFLFFL. Similarly, the letters R (repetition) and S (switch) can be used to describe a sequence of transitions (RSSRS in the example). To convert the sequence to binary values, L can be recoded as 1 and F as 0 (or vice versa) and R can be coded as 1 and S as 0 (or vice versa).

Such sequences of random events are often summarized by using a runs statistic, which yields a proportion of the runs of the same event at a series of lengths. The runs statistic (Sternberg, 1959a; see also Vandierendonck, 2000a), can be defined as follows

$$r_k = \frac{1}{N - k + 1} \sum_{i=1}^{N-k+1} x_i x_{i+1} \cdots x_{i+k-1} \quad (1)$$

where N is the number of events in the complete sequence and r_k is the proportion of runs with length k in which $x_k = 1$. Consider a sequence like “0110111010100011”, where the target outcome is coded 1. The number of runs of length 1 equals the number of 1s in the sequence, which is 9 (code 1 occurs 9 times). Runs of length 2 consist of two consecutive 1s; there are four such groups of 1s in the sequence. Runs of length 3 consist of three consecutive 1s; there is only one such group. By dividing these counts by the number of possible runs of a particular length, a proportion is obtained for each length. Although the runs statistic captures large deviations from

independence, it is not particularly sensitive in detecting very small deviations from independence because the values become smaller as the run length increases (see example). However, the statistic is useful because it captures deviations from statistical independence in both directions: when there are more repetitions, there will be fewer short and more long runs, and when there are more alternations, the opposite pattern will occur (relatively more short and less long runs).

In addition to the runs statistic, we used the *autocorrelation statistic* (Sternberg, 1959a; see also Vandierendonck, 2000a), which is very sensitive to deviations from independence, and therefore useful to make a more fine-grained analysis of the data. This statistic expresses the tendency for pairs of events in the sequence to correlate with each other. The pairs of events that are considered can be close together or further apart (short or long lag). The autocorrelation statistic lag k (c_k) is defined as

$$c_k = \frac{1}{N - k} \sum_{i=1}^{N-k} x_i x_{i+k} \quad (2)$$

The correlation expresses the probability that both elements in the pair (separated by lag k) are the same. When lag is 1, for example, the autocorrelation expresses the probability that the current event is the same as the previous one. Considering the example we had before “0110111010100011”, the autocorrelation lag 1 looks at all occurrences of two consecutive 1s; there are 4 of these. Actually, by definition this is the same as runs of length 2. The autocorrelation lag 2 looks at two occurrences of a 1 separated by another (not relevant) outcome. The triplets to consider are 011, 110, 101, 011, 111, 110, 101, 010, 101, 010, 100, 000, 001, 011 and there are only four cases out of these 14 where the first and the third element are both 1. The autocorrelation statistic is a measure that is sensitive to statistical dependencies based on learning to repeat an event: the autocorrelation will tend to be larger when a learning process governs the production of the events in the series (Sternberg, 1959a, 1959b). Application of the statistic requires that one task is coded 1 and the other 0. By applying the statistic twice to the sequence of task choices, once with Letter coded as 1 (Form 0) and once with Form coded as 1 (Letter 0), the joint outcome specifies all correlations in the data. This joint outcome is complementary to all

tendencies to alternate instead of to repeat. Hence, the statistic applied in this way is sufficient to describe all deviations from independence.

3.3. Results

The data analysis will be performed in two steps. In the first step, the analysis focuses on the sequence of task choices; in the second step, on sequences of transitions. Although the transitions are derived from the sequence of task choices, due to the regrouping of successive events in terms of repetition or switch, the second analysis may detect different statistical properties within the sequence. Inspection of the sequence of transitions may, for example, clarify whether a bias to repeat tasks as seen in the task analysis is a local phenomenon or a more global phenomenon. If the repetition bias is local, task repetitions will occur in only one or a few subsequent transitions, soon interrupted by a switch. On the contrary, when the repetition bias is more generalized, long as well as short series of task repetitions will occur. Performing these two analyses, then, may reveal information about the sequence of tasks that would otherwise remain undetected. Within each type of analysis, the results of the runs and the autocorrelation statistic are reported separately.

In the task-based analysis, both tasks were selected equally often and the proportions of runs decreased at the same rate in both tasks. Run proportions were on average smaller (i.e., less repetition bias) in the condition with long RSI (see Figure 1, top-left). In the autocorrelation data, the size of the correlation varied with lag with lower values at lags 2-4 than at the other lags (see Figure 1, bottom-left). Even though, the autocorrelations depended only marginally on RSI, the effect on lags 2-4 was stronger in the long RSI condition. The autocorrelation findings confirm a repetition tendency at lag 1 followed by a tendency to alternate at lags 2-4; as this alternation tendency was stronger in the long RSI condition, the tendency to repeat tasks was stronger in the short RSI condition.

In the transition-based analysis, repetitions were reliably more frequent than switches and this difference was maintained at all run lengths (see Figure 1, top-right). Repetitions were also more frequent at short than at long RSI. This clearly confirms the presence of a repetition bias which is stronger at short RSI. The autocorrelations were lower at lag 1 than at the other lags and

were stable from lag 2 on (see Figure 1, bottom-right). This difference did not interact with transition type but it did interact with RSI. The fact that autocorrelations tend to be lower at lag 1 indicates that although repetitions are more frequent overall, an immediate repetition of the same transition is lower than on average at other lags.

These findings are described with more statistical detail in the following paragraphs 3.3.1 and 3.3.2. Readers who prefer to skip these details can do so and move on directly to the discussion of the results in section 3.4.

3.3.1. Focus on tasks.

For the analysis with focus on the letter task, the 4 x 64 task selections of each participant were coded 1 when the letter task was selected and 0 otherwise (form or no selection). Similarly, for the form task, the selections were coded 1 when form was selected and 0 otherwise. In only one participant, about 1.5% of the trials were non-selections; in the other participants, no non-selections occurred. Anyway, due to the coding, non-selections did not contribute to the run length or the autocorrelation data.

These binary data were used to calculate the proportion of runs of lengths 1-8 and autocorrelations at lags 1-8². Per statistic, the multivariate general linear model was applied to the data on the basis of a 2 (RSI: 100 or 1000 ms) x 2 (Task: letter vs. form) x 8 (Lengths or Lags) factorial design with repeated measures on the last two factors. For all analyses, $\alpha = .05$, unless otherwise mentioned.

3.3.1.1. Runs. On average, the selection proportions of letter and form were both .50. Run proportions depended on RSI and run length only. Proportions of runs decreased with run length, $F(7,10) = 870.46$, $\eta_p^2 = 1.0$. Run proportions were smaller when RSI was long (.10 for long versus .16 for short RSI). Neither task (.14 vs. .13), nor any of the interactions attained significance (smallest $p > .20$). This shows that there was a tendency to repeat tasks more at short than at long RSI. Figure 1 (top-left) displays the runs proportions as a function of RSI, task, and run length.

---- Figure 1 about here ----

² We calculated lengths and lags 1-10, but in order not to violate the assumptions of the analytic tool, we restricted the analysis to 8 lengths or lags.

3.3.1.2. Autocorrelations. The autocorrelations are shown in Figure 1 (bottom-left) as a function of RSI, task, and lag. Only the effect of lag was reliable: $F(7,10) = 4.10$, $\eta_p^2 = 0.74$. The effect of RSI was marginally significant, $F(1,16) = 3.10$, $\eta_p^2 = 0.16$. Figure 1 shows that the correlations start high, drop off at lags 2-4 and stabilize from lag 5 on. This pattern is more clearly present at longer RSI, though. Analysis confirmed that overall autocorrelations were higher at the first than at the other lags, $F(1,16) = 27.68$, $\eta_p^2 = 0.63$. A contrast between the autocorrelations at lags 2-4 and lags 6-8 was not reliable overall ($F < 1$), but it interacted with RSI, $F(1,16) = 4.69$, $\eta_p^2 = 0.23$, such that the difference between lags 2-4 and lags 6-8 was larger in the condition with the long RSI. These findings show that there is a rather strong repetition tendency (autocorrelation) at lag 1, but that in lags 2-4 in the long RSI condition, the autocorrelation is rather weak. This suggests that the repetition bias is stronger at short RSI.

3.3.2. Focus on Transitions.

In the sequential analysis of the transitions, repeating the same task was coded 1 and changing the task or failing to select a task was coded 0; in the calculation targeting on switches, changing tasks was coded 1 and repeating the same task or failing to select a task was coded 0. In all other respects, the same data-analytic method was used as for the analysis focusing on task selections. In order not to overload the report with an enumeration of statistical tests, the outcomes of this analysis are presented in Table 2; only the effects that are central to our main purpose are reported in the text.

---- Table 2 about here ----

3.3.2.1. Runs. Transition runs are shown in Figure 1 (top-right) as a function of RSI, transition and length. Clearly, repetitions ($M = 0.63$) were more frequent than switches ($M = 0.37$) and they also differed (.17 vs. .06) at all lengths. Run proportions were larger on short ($M = 0.14$) than on long RSI ($M = 0.10$). Interactions of transition and length and RSI and length were reliable (see Table 2). The dominant presence of repetitions confirms the repetition bias already observed at the level of tasks. Repetitions were repeated more often than switches especially at short RSI.

---- Figure 1 about here ----

3.3.2.2. Autocorrelations. Figure 1 (bottom-right) displays the transition-based autocorrelations as a function of RSI, transition, and lag. Lag correlations were larger for repetitions ($M = .41$) than for switches ($M = .15$). They also varied over lags. Transition and lag did not interact, but transition was involved in an interaction with RSI, as displayed in Figure 1. Finally, also the interaction of RSI and lag was significant (Table 2). In contrast to the task-based autocorrelations, the transition-based autocorrelations seem quite stable, except that they were lower at lag 1 ($M = .26$) than at other lags ($M = .29$), $F(1,16) = 13.00$, $\eta_p^2 = 0.45$. This contrast explains the major part of the variance among the means per lag: $r^2 = .82$. Repetitions are selected more often than switches and therefore repetitions also tend to be repeated more than switches. However, for both, repetitions and switches, this tendency is smaller at lag 1 than at longer lags, where transitions rather show a pattern of independence, but repetitions are still repeated more than switches.

3.4. Discussion

These analyses confirm the repetition bias that has been reported in the literature (e.g., Arrington & Logan, 2004; Mayr & Bell, 2006), as well as the finding that this bias is stronger at shorter RSI (Arrington & Logan, 2005). The autocorrelations of the task choices show that the repetition bias is strong at all lags except lags 2-4. Transition autocorrelations are stable from lag 2 on and indicate in fact that the repetition tendency of both repetitions and switches is a global phenomenon, except for lag 1 which suggests a local tendency to repeat the previous transition less often than average. This pattern of findings is not consistent with existing perseveration models (e.g., Vandierendonck, 2000a), as they predict a decrease of autocorrelation over lags.

3.5. Model testing

In this section, we report the results of the model tests performed on the task-selection data of this experiment. First, we will present the results of fitting the chain-retrieval model to the runs proportions obtained in the experiment. In order to show that the chain-retrieval model yields a better account than simpler models, also the results of the statistical independence (Bernoulli) model and two statistical dependency models are presented. One of these statistical dependency models is the perseveration model (Vandierendonck, 2000a). It assumes that with probability a the

previous event is repeated and if no such repetition occurs, with probability q an event is sampled independently from the previous event:

$$Pr(x_i) = a + (1 - a)q \quad (3),$$

where $Pr(x_i)$ refers to the probability that a certain event (x) occurs at time i , a represents the probability that the previous event is repeated (perseverates) and q is the probability that the present event is sampled independently from the previous trial. The second dependency model is the alternation model (Vandierendonck, 2000a). This model is also based on equation (3), but now parameter a represents the probability that an alternation occurs (i.e., the previous event is not repeated) and if no alternation occurs, with a probability q an event is sampled independently from the previous trial.

In a next step, the parameter estimations obtained in these fits of the four models will be used to predict the autocorrelations in sequences of tasks and sequences of transitions. After this phase of global model testing, more specific tests of the chain-retrieval model will be reported. To that end, each of the free parameters will in turn be clamped to a particular value and the other parameters will be estimated resulting in new fits and new predictions.

3.5.1. Model fitting and parameter estimation.

First, we used the run proportions of the task-based data analysis to estimate the parameters of the task-chain retrieval model and the three comparison models. Because the data of each individual participant are sufficient to fit the models, the runs data of each subject were used to

estimate the free parameters of the best fitting model for that subject. The fits per subject could then be entered in statistical analyses comparing the merits of the models³.

All model fits were based on maximum likelihood estimation. The parameters were estimated jointly for the letter task and the form task data on the basis of the observed runs proportions length 1-10. The likelihood function to estimate the correspondence between the model and the data is:

$$L = \prod_{i=1}^K \binom{N-i}{n_i} p_i^{n_i} (1-p_i)^{N-n_i-i} \quad (4)$$

where L is the likelihood function, K is the number of lengths considered, p_i is the estimated proportion of runs of length i , N is the total number of observations (number of trials) and n_i is the number of observed runs of length i . In order to find the maximum of such function, it is easier to take $\ell = \ln L$, and to minimize $-\ell$:

$$-\ell = \sum_{i=1}^K \ln \binom{N-i}{n_i} n_i \ln p_i - (N - n_i - i) \ln(1 - p_i) \quad (5)$$

This minimization was performed by means of a univariate search method (Brent, 1973). On each step of the iteration, the current parameter values were used to generate a sequence of 50,000 task choices. On the basis of this generated sequence, estimated runs for the letter task

³ This procedure has several advantages over the alternative procedure based on fitting the models on the between-subject average of runs. First, the processes described in the chain-retrieval model and also those of the other dependency models are constrained by the skills and capacities of each subject. This variability is given the best chance by using individual data. Second, the between-subject average of the runs statistics does not adequately represent the processes that resulted in the generated random sequences. It is easy to imagine that the average runs of a subject generating a sequence with a repetition bias and another subject generating a sequence with an alternation bias will show either a very small bias or no bias at all.

were calculated. Next the sequence was converted (0 was recoded to 1 and 1 to 0) to calculate the estimated runs for the form task. The estimated proportions were weighted as specified in formula (5) to yield $-\ell$. The search procedure would then sample new parameter values and start a new step. This continues until a minimum is obtained. By performing the estimation jointly for the two tasks, the characteristic of random succession of the two tasks is captured in the parameter fit. The search procedure finds a local minimum in a very efficient way. In order to maximize the chance of finding the global minimum, the search procedure was applied ten times with random starting values. In the application for the chain retrieval model, this estimation procedure was repeated for each of the values 3-6 of m . Per participant, the value of the three parameters of the best fitting model was then selected.

The top panel of Table 3 displays the fit and the estimated parameters of the four models. This table shows that the averaged minimization over all participants was quite good for the chain-retrieval model but was much worse for the three statistical models (Bernouilli, perseveration and alternation). In order to test whether the difference between the fit of the chain-retrieval model and the other models was statistically reliable, Akaike's Information Criterion (Akaike, 1974) with correction for the number of observations (AICc, Hurvich & Tsai, 1989) was calculated per subject and per model. The proportion of subjects in which the AICc was smaller for the chain-retrieval model than for the comparison model is displayed in Table 3 on the row labeled "Prop(AICc)". To test whether this proportion was significant, a z-score was calculated in the Gaussian approximation to the binomial distribution of proportions; this z-score is also shown in the table together with the probability that this would occur under the null hypothesis that both models are equivalent. These tests show that the chain-retrieval model yielded a better fit than each of the other three models, except the perseveration model. In view of this, further tests will have to clarify how well the chain-retrieval model captures the data.

---- Table 3 about here ----

3.5.2. Model predictions.

The estimated parameters of these four models were used to predict the autocorrelation statistic for the task data. Predictions of the models were calculated on a sequence of 50,000

generated events on which the likelihood function was calculated given the parameters estimated on the runs data. In order to measure the degree of correspondence between the predictions and the data, the root mean squared deviation was calculated:

$$rmsd = \sqrt{\frac{1}{K} \sum_{i=1}^K (o_i - e_i)^2}$$

where o_i refers to the data points and e_i refers to the predicted values. This measure obtained for each of the models is displayed in the middle panel of Table 3, and shows that the prediction of chain-retrieval model is better than that of the three other models. The chain-retrieval model's prediction accuracy was compared pairwise with that of the other models. Table 3 shows the proportion of participants performing better on the chain-retrieval model than on each of the other models. The probability of this proportion was assessed by calculating a z-score (also shown in the table). Figure 2 (top-left panel) illustrates that the task-chain retrieval model yielded the best correspondence between predictions and data. When applied to data that are more sensitive to deviations from statistical independence than the runs statistic, it seems that the chain retrieval model significantly better accounts for these deviations than the other dependence models.

---- Figure 2 about here ----

The estimated parameters of these four models were also used to predict the autocorrelation statistic for the transition data. In order to keep task and transition statistics equivalent in terms of number of events, only lags 1-9 were included for the transitions. The sequence generated for the prediction of the task autocorrelations was converted into a sequence of transitions, once with focus on repetitions and once with focus on switches. The rmsd was calculated per subject. Table 3 (bottom panel) displays the rmsd and the proportion of participants having smaller rmsd on the chain-retrieval model than on the comparison models. Again, the probability of this observation under the null hypothesis is derived by calculating z-scores. Figure 2 (bottom-left panel) illustrates that the task-chain retrieval model yielded the best correspondence between predictions and data, and this was confirmed by the significant difference with each of the other models.

3.5.3. Model validation.

As the model has three free parameters, it is also important to know whether each of the parameters is indispensable. To that end, each of the parameters in turn was clamped to a neutral value, and with one parameter fixed, new estimations of the other two parameters were obtained. A neutral value for the free parameters was considered to be a value at which the effect of the corresponding process was the lowest; so for m and r the lowest value, respectively 3 and 0, was taken, and for b , a value of 0.5 (no bias in either direction) was considered to be neutral. Table 4 shows that $-\ln L$ yielded a lower value for the full than for the restricted models (see also, Figure 2, right panels). For testing the significance of the differences between these nested models, the likelihood ratio test was used (Buse, 1982). Let ℓ_A refer to the log likelihood of the full model and ℓ_B to the log likelihood of the nested model, then the statistic W , which is defined as,

$$W = 2(\hat{\ell}_A - \hat{\ell}_B)$$

is distributed according to a χ^2 distribution with the difference in the number of parameters between the two models as the number of degrees of freedom. Per subject, W was calculated for all pairs consisting of the full (three-parameter) model and a model where one of the parameters was fixed. In order to test whether the majority of subjects was in favor of the full model, a normal approximation to the binomial test for proportions was performed. On the line labeled " $P(\text{LR})$ " the proportion of subjects with a significantly smaller likelihood of the full model is displayed, and on the next line the z-score with the probability of the observation under the null hypothesis is listed.

---- Table 4 about here ----

It appears that the fit of the full model is better than the fit of the nested models, but only significantly for the variant with parameter r fixed at 0. The other two nested models ($m = 3$ and $b = 0.5$) could not be rejected. However, when also the predictions of the full and the nested models are taken into account (see lower panels of Table 4), it appears that in the task autocorrelation

predictions only the nested model with b fixed is equivalent with the full model, and in the predictions of the transition autocorrelations, the full parameter model was significantly better than each of the nested variants. These findings suggest that the m and the r parameter may play an important role, whereas the evidence in favor of the b parameter is not convincing. Possibly the statistical power of the present study is not strong enough for supporting strong conclusions.

Table 4 also shows the estimate of the proportion of repetitions due to top-down control and the estimate of the proportion of bottom-up triggered repetitions. Together these two proportions account for the observed proportion of repetitions. Clearly, the proportion of top-down triggered repetitions is much smaller than the value of the bias parameter (b). Also the proportion of bottom-up triggered repetitions is smaller than expressed in parameter r . This is because some bottom-up repetitions coincide with top-down repetitions and do not change the top-down controlled actions. Comparison of these proportions across the different model variants shows that when r is 0, there are no bottom-up events whatsoever, with as a result that the other parameters (m and b) are adapted to account as much as possible for the observed proportion of repetitions. Similarly, when b is clamped to a neutral value, m and r become larger to achieve an acceptable estimation. The observation that m increases to compensate for the lack of repetitions, suggests that this parameter is not only constrained by the length of the chains that can be handled but also by the frequency with which top-down repetitions do occur. Finally, even when m is fixed, the estimated value of the two other parameters is adapted (increased), again suggesting that m affects the number of top-down repetitions.

3.5.4. Discussion.

The chain-retrieval model seems quite promising when fitted to the run length proportions in sequences of tasks. For a significant majority of the participants, the obtained fit was better than the fits obtained for the statistical independence model and existing statistical dependence models. Also the predictions of the autocorrelations in sequences of tasks as well as in sequences of transitions were better for a significant majority of the participants than the predictions of the comparison models.

In a final test designed to investigate whether the processes underlying the three free parameters are all involved in achieving the very good correspondence of the chain-retrieval model and the data, each of the parameters in turn was clamped to a neutral value so as to exclude or to minimize the role of the underlying process. These analyses indicated that the m and r parameters are important: even when corrected for the number of estimated parameters, the full parameter model still yielded significantly better fits to the runs proportions and significantly better predictions of the task and transition autocorrelations. The case for the b parameter was less convincing in the present study, although the findings clearly indicate that this parameter also affects the top-down repetition tendency.

4. Study 2

Although the application of the chain-retrieval model to the task choice data of Study 1 shows that the model yields a better account of task choices in VTS than statistical independence and dependence models, some concerns may be raised about the generality of these findings. First, the size of the subject sample and the number of choices per subject were both rather small. Second, the data were collected with the single-registration procedure which is known to be very sensitive to bottom-up intrusions. As the occurrence of such intrusions is a central assumption of the model, the test in Study 1 may be biased in favor of the model. In order to counter the latter criticism, in Study 2 the double registration procedure (Arrington & Logan, 2005) was used. In this procedure, two responses have to be emitted in each trial. First, a probe appears to which the subject responds by indicating the task that will be used in the current trial. Next, the target stimulus is presented and the selected task is applied to this stimulus. The concern regarding the amount of data was countered by having more subjects (80) and longer sequences of task choices (2 blocks of 256 trials).

This design has the additional advantage that it becomes possible to include the effect of practice in the study. If task-choice behavior changes with practice, then the pattern of choices might be different between the first and the second block of 256 trials. In the same vein, this design allows to investigate whether the best fitting model parameters change with practice. The rationale for this is that in developing the chain-retrieval model, it was assumed that the chains in

LTM are learned from experience within the constraints of working memory because once retrieved, the chain has to be unfolded in working memory in order to access the next element in the sequence on consecutive trials. In order to be able to fit the model parameters, it was further assumed that learning was confined to the practice trials so that chain length could be considered constant over the experimental session. A secondary aim of this study is to test whether this simplified assumption is tenable: are changes in the m parameter negligible or are there important changes with further practice?

4.1. Method

4.1.1. Participants and design

Eighty first-year psychology students at Ghent University participated for course requirements and credit. All participants had normal or corrected-to-normal vision, were right-handed, and all were naïve to the purpose of the experiment. Participants were randomly assigned to two conditions (forty participants per condition) that differed in the stimulus onset asynchrony (SOA) of the stimulus and the probe (see below).

4.1.2. Materials

Stimuli were the digits 1-9, excluding 5. Participants were required to classify the digits either on the basis of their magnitude (smaller or larger than five) or their parity (odd or even). Responses were registered by means of the numeric pad of a standard keyboard. One hand was used for pressing the task-selection keys; the other hand was used for pressing the task-execution keys.

4.1.3. Procedure

Although we only need the task-selection data, we describe the complete procedure of data collection. However, for completeness and to show that the experiment replicates typical task-switching results, the main findings regarding the task-execution data obtained in this experiment are briefly presented in Appendix A. Pentium III personal computers with a 17-inch color monitor running the Tscope C/C++ library (Stevens et al., 2006) were used. Each session lasted for approximately 45 minutes. After participants signed an informed consent, instructions were presented on screen and paraphrased if necessary. The instructions concerning unpredictability of

voluntary task switches were the same as those used by Arrington and Logan (2005), namely that each task should be performed about equally often and that the sequence should be as unpredictable as in coin tossing.

On each trial, a probe (“?”) was presented in a square 5 mm above the centre of the screen. This probe disappeared when participants pressed one of the task-selection keys. This was followed 400 ms later by the appearance of the target stimulus, 5 mm below the centre of the screen. The target remained on screen until participants responded on the basis of the previously selected task or until a maximal response time of 2500 ms elapsed. The probe of trial n appeared either 50 ms (short SOA condition) or 1500 ms (long SOA condition) after the presentation of the stimulus of trial $n-1$ ⁴.

After two practice blocks of 64 trials, participants performed two experimental blocks of 256 trials. In the first practice block, the emphasis was on familiarization of the procedure of selecting and executing the different tasks. In the second practice block, the emphasis was on unpredictable selection of the tasks. In order to increase participants’ awareness of their selection behavior, the warning “do not forget to switch tasks” was shown for 1000 ms whenever the participants selected the same task four times in a row. When the participants switched between tasks four times in a row, the warning “do not forget to repeat tasks” appeared for 1000 ms. This task-selection feedback was presented in the second practice block only.

⁴ This variation of SOA between the probe and the target is reminiscent of the PRP (psychological refractory period) paradigm. However, unlike in the PRP paradigm, in the present procedure, task selection does not have to wait for the probe to start. In fact, task selection can start at any time, even at an earlier trial. The function of the probe is to indicate that a task-indication response must be emitted. On the trials where the task choice has not yet been made, the task-selection response may result in postponement of the task execution response. Appendix A shows, however, that probe RTs are fairly short suggesting that on average task selection has been completed before presentation of the probe. For all these reasons, strictly speaking, the present design is not a PRP paradigm, although it may bear similarities to it and it may be the case that task choice processes compete with task execution.

During the entire experiment, participants received on-line feedback about their performance. A red screen appeared for 50 ms when they made an error on the target. When they were too slow to select a task ($RT > 2500$ ms), the message “no task selected” was displayed for 1500 ms. Following each block (practice and experimental), a general summary about the performance during that block was shown. This feedback included the mean reaction times on the targets, the percentage of errors, the selection percentage of each task, the percentage of failures to select a task, and the percentage of task repetitions and task switches. If necessary, participants were corrected: they were urged to switch more or to repeat more when the proportion of repetitions or switches was above .70, to make fewer errors when percentage of errors was above 15%, to respond faster when mean task-execution reaction time was above 1200 ms or when the proportion of trials without a task-indication response was above 10%, and to be more random when they selected a particular task on more than 75% of the trials.

4.2. Results

As in Study 1, the data analysis was performed in two steps. In the first step, the analysis focused on the sequence of task choices; in the second step, on sequences of transitions. The analyses of the runs and the autocorrelation statistic are reported separately. As for Study 1, we first provide an overall summary of the findings before presenting the statistical details.

In the task-based analyses, runs proportions were larger in the parity task, in the second block and in the short SOA condition (Figure 3, top-left). The drop in proportions with run length was steeper in the short SOA condition and also in the second block. The autocorrelations varied with task and lag (see Figure 3, bottom-left). At lag 1, the autocorrelations were rather larger and then showed a dip at lags 2-4. The size of the dip was larger at long SOA. The size of the dip was also larger in the second block than in the first.

The transition-based analyses revealed that runs proportions were higher for repetitions than for switches, and higher for short than for long SOA (see Figure 3, top-right), and higher in the second block as compared to the first block. At short SOA repetitions were repeated more often than switches and in the second block runs proportions tended to be larger than in the first block at all lengths. The autocorrelations were characterized by a lower value at lag 1 than at the other

lags and overall there was not much variation from lag 2 on (see Figure 3, bottom-right). The autocorrelations were larger for repetitions than for switches and were larger in block 2 than in block 1.

In the following paragraphs, these findings are substantiated with the statistical details.

Readers who prefer to skip these details can continue at section 4.3 (Discussion).

4.2.1. Focus on tasks

For the magnitude task, the task selections of each participant were coded 1 when magnitude was selected and 0 otherwise (parity or no selection). Similarly, for the parity task, the selections were coded 1 when parity was selected and 0 otherwise. About 1% of the trials were non-selections. Due to the coding, non-selections did not contribute to the run length or the autocorrelation data.

These data were used to calculate the proportion of runs of lengths 1-10 and autocorrelations at lags 1-10 separately in each block. Per statistic, the multivariate general linear model was applied to the data on the basis of a 2 (SOA: 50 or 1500 ms) x 2 (Task: magnitude vs. parity) x 2 (Blocks) x 10 (Lengths or Lags) factorial design with repeated measures on the last two factors. For all analyses, $\alpha = .05$, unless otherwise mentioned. In order not to overload the report with an enumeration of statistical tests, the outcomes of the complete analyses are presented in Table 5; only the effects that are central to our main purpose are reported in the text.

---- Table 5 about here ----

4.2.1.1. Runs. Figure 3 (top-left) displays the runs proportions as a function of SOA, task, and run length. On average, the selection proportions of magnitude and parity were 0.48 and 0.51 respectively. As can be seen in Table 5, run proportions depended on task, blocks, length, and SOA. Proportions of runs were higher for the parity task ($M = .13$) than for the magnitude task ($M = .11$), and they decreased with run length. Run proportions were larger in the second block ($M = .12$) than in the first block ($M = .11$), and they were smaller when SOA was long (.10 for long versus .13 for short SOA). Length interacted with task and with SOA. The drop in the proportions was less steep for the parity task and for short SOA. This shows that there was a tendency to

repeat the parity task more than the magnitude task and also a tendency to repeat tasks more at short than at long SOA. Blocks interacted with length, SOA and length x SOA.

---- Figure 3 about here ----

In the second block, more short run lengths were observed and the proportions of runs decreased over lengths at a slower rate than in the first block. This is confirmed in the interaction of the contrast between short (1-2) and medium lengths (4-5) by blocks, $F(1,78) = 16.58$, $\eta_p^2 = .18$ and in the interaction of this contrast, SOA and blocks, $F(1,78) = 7.13$, $\eta_p^2 = .08$.

4.2.1.2. Autocorrelations. The autocorrelations are shown in Figure 3 (bottom-left) as a function of SOA, task, and lag. Only the effects of task and lag were reliable (see Table 5). Figure 3 shows that the correlations start high and then quickly drop off and stabilize from lag 5 on. Overall, autocorrelations were lower at lags 2-4 ($M = .22$) than at lags 7-9 ($M = .24$), $F(1,78) = 23.76$, $\eta_p^2 = 0.23$. This contrast interacted with SOA, $F(1,78) = 6.47$, $\eta_p^2 = 0.08$. The contrast was smaller at short (.22 vs. .23) than at long SOA (.21 vs. .25). These findings show that there was a rather strong repetition tendency (autocorrelation) at lag 1, but that at lags 2-4, the autocorrelation was rather weak. This suggests that a tendency to immediately repeat the task is soon followed by one or more switches.

There was no main effect of blocks, and blocks interacted only with lags. Autocorrelations at short lags (1-2) tended to be larger in the second block (.269 vs. .258), whereas from lags 4-5 on, the difference tended to vanish (.228 vs. .225). This was confirmed in an interaction of the contrast between short (1-2) and medium (4-5) lags, $F(1,78) = 6.04$, $\eta_p^2 = .07$.

4.2.2. Focus on transitions

In the sequential analysis of the transitions, repeating the same task was coded 1 and changing the task or failing to select a task was coded 0; in the calculation targeting on switches, changing tasks was coded 1 and repeating the same task or failing to select a task was coded 0. In all other respects, the same data-analytic method was used as for the analysis focusing on task selections. The statistical analyses are reported in Table 6.

---- Table 6 about here ----

4.2.2.1. Runs. Transition runs are shown in Figure 3 (top-right) as a function of SOA, transition, and length. Overall, runs of repetitions ($M = 0.14$) were more frequent than runs of switches ($M = 0.07$). Run proportions were only slightly but reliably larger on short ($M = 0.103$) than on long SOA ($M = 0.102$). All interactions were reliable (see Table 6). The dominant presence of repetitions confirms the repetition bias already observed at the level of tasks. Repetitions were repeated more often than switches especially at short SOA.

In the first block, run proportions were smaller ($M = .10$) than in the second block ($M = .11$), and this effect interacted with all other factors except transition; all higher-level interactions with blocks were significant. The contrast between the first (1-5) and the second half of the lengths (6-10) revealed an average drop of .17 in the first block and a drop of .18 in the second block. Although values started higher in the second block and dropped more, at length 10 they were still higher than in the first block. This contrast accounts to a large extent for the interactions involving run length and blocks.

4.2.2.2. Autocorrelations. Figure 3 (bottom-right) displays the transition-based autocorrelations as a function of SOA, transition, and lag. In contrast to the task-based autocorrelations, the transition-based autocorrelations seem quite stable, except at lag 1. Overall, autocorrelations were larger for repetitions ($M = .38$) than for switches ($M = .17$). They were larger in the second block ($M = .28$) than in the first block ($M = .27$). They also varied over lags. In particular, correlations were lower at lag 1 ($M = .22$) than at other lags ($M = .28$), $F(1,78) = 146.54$, $\eta_p^2 = 0.65$. This contrast explains most of the variance among the means per lag: $r^2 = .98$.

Transition interacted with SOA and lag. The latter interaction basically corresponds to an interaction of transition with the contrast between lag 1 and lags 2-10, $F(1,78) = 72.16$, $\eta_p^2 = 0.48$. Blocks interacted with SOA but not with transition or lag. Finally, the triple interactions of SOA, transition and blocks and of SOA, transition and lag were significant, as well as the interaction of all four factors.

Repetitions were selected more often than switches and therefore repetitions also tended to be repeated more than switches. However, for both, repetitions and switches, this tendency was

smaller at lag 1 than at longer lags, where transitions rather showed a pattern of independence, but repetitions were still repeated more than switches.

4.3. Discussion

The pattern of findings replicates the pattern observed in Study 1: a repetition bias was observed and this bias was stronger at short than at long SOA (Arrington & Logan, 2004, 2005; Mayr & Bell, 2006). The pattern of autocorrelations for both the tasks and the transition was also completely similar to that observed in the first study. Again, it may be concluded that the repetition bias is global with the exception that immediate (lag 1) repetition of a transition is less frequent than the global average.

The presence of more alternations in lags 2-4 is consistent with the observation of Lien and Ruthruff (2008) that task repetitions at lag 2 (i.e., ABA and BAB) are avoided. These authors attribute their results to persistence of the inhibition (e.g., Mayr & Keele, 2000) of the task at trial $n-2$ on the present trial (n). However, in the present data, the alternation tendency is present in the window at lags 2-4 so that even if persistence of inhibition could account for the findings at lag 2, this explanation fails to account for the observed alternation tendency at lags 3 and 4.

Practice seems to result in a tendency to have more shorter runs of the same task, but this effect depends on combinations of the other variables in the design such as SOA, lag and task. The transition-based analysis showed evidence for an effect of practice with an increase in the number of repetitions in the transition data. Basically, practice seems to enhance the repetition bias.

4.4. Model testing

The same model-testing procedure was used as in Study 1. Additionally, the effect of practice on the obtained parameter estimates was investigated by comparing model fits based on the first versus the second block of task choices. Next, the chain-retrieval model was compared to model variations where each parameter in turn is clamped to a neutral value. In each case, first the tests are applied to the entire data sequence ignoring practice effects before presenting data that show how practice modulates the estimated parameters, the fits and the predictions.

4.4.1. Model fitting and parameter estimation.

Table 7 (top panel) displays the fit and the estimated parameters of the four models to the complete sequence of task choices per subject. This table shows that the average fit over all participants was very good for the chain-retrieval model, while the fits of the statistical independence and dependence models were rather poor. Based on the AICc criterion which balances the goodness of fit and the number of free parameters, the chain-retrieval model yielded the best fit for a large majority of the participants (73 of the 80 participants for the independence and alternation models; 62 out of 80 participants for the perseveration model). This significantly better fit to the task run proportions of the chain-retrieval model is interesting in that it suggests that the model better captures the relevant information in the data.

---- Table 7 about here ----

As an additional investigation of the goodness of fit, we tested whether the estimated parameter values captured differences between the subjects due to a short versus long SOA between the previous target and the present probe stimulus in the experiment. The estimated values of parameters m and b did not depend on the SOA, but the value of the r parameter did: with short SOA the parameter value was larger ($M = 0.31$) than with long SOA ($M = 0.18$), $F(1,78) = 10.11$, $\eta_p^2 = .11$. This is consistent with the assumption that r represents the probability of intrusions which occur more often at short SOA. Similarly, in the perseveration model, only the perseveration parameter was sensitive to SOA, with a larger perseveration tendency at short (.19) than at long (.07) SOA, $F(1,78) = 8.34$, $\eta_p^2 = .10$. This is again consistent with the findings of a larger repetition tendency at short SOA. In the alternation model, the alternation parameter was sensitive to SOA with a smaller probability at short (.07) than at long (.13) SOA, $F(1,78) = 4.92$, $\eta_p^2 = .06$. The smaller tendency to alternate at short SOA is also consistent with a larger repetition tendency at short SOA. In neither of these two models, the q parameter varied with SOA, and similarly the p parameter of the Bernoulli model did not depend on SOA. These findings show that the repetition bias is something that is not captured in a general probability of selecting a task, but that the chain-retrieval model and the statistical dependence models are able to capture its variability.

The assumption that parameter b is related to the switch cost was tested by means of the correlation between this parameter and the switch cost. Because the perception of the difference in difficulty between repetitions and switches should be driving the bias, the correlation of the error switch cost per participant with the individual's estimated value of b was calculated. The error switch cost is the difference between the average proportion of incorrect repetitions and the average proportion of incorrect switches. The product-moment correlation amounted to 0.28, $t(78) = 2.58$, $p < .05$. This indicates that b is related to the difference in difficulty of repetitions and switches.

4.4.2. Modulation of model fits due to practice

Next, the chain-retrieval model parameters were estimated for each block separately. The results of these fits are also shown in Table 7. The minimized $-\ell$ did not significantly differ between the two blocks, $F(1,78) = 2.30$, $\eta_p^2 = .03$, $p = .13$, but more participants (47 out of 80) obtained a lower minimum in the first block than in the second block. The average m and b parameter values did not differ between the two blocks, respectively $F(1,78) = 1.54$, $\eta_p^2 = .02$, $p = .22$ and $F < 1$. Parameter r , on the contrary, was smaller in the first block than in the second block 2, $F(1,78) = 5.34$, $\eta_p^2 = .06$, which would suggest that vulnerability to bottom-up intrusions increases from the first to the second block, possibly due to fatigue or a related drop in attentional focusing. That m did not change significantly over the two blocks, suggests that a short practice period suffices to achieve stable performance.

4.4.3. Model predictions.

The same procedure was followed as in Study 1 to obtain predictions of the autocorrelation statistic for both the task data and the transition data. The outcomes of the statistical tests of this correspondence are displayed in Table 7 (lower panels). The correspondence of the predictions and the data is shown in Figure 4 for the predictions of the task autocorrelations (top-left) and the transition autocorrelations (bottom-left). This figure illustrates that the task-chain retrieval model yielded the best correspondence between predictions and data. The overview in Table 7 shows that the chain-retrieval model yields better predictions for both the task autocorrelations and the transition autocorrelations than the other models in a significant majority of the participants. The

present observations confirm that when applied to data that are more sensitive to deviations from statistical independence than the runs statistic, it seems that the chain-retrieval model significantly better accounts for these deviations than the other dependence models.

---- Figure 4 about here ----

The correspondence between the observed and the predicted autocorrelations was on average very similar in the two blocks, as can be seen in the lower panels of Table 7. In the predictions of the autocorrelations of tasks and transitions, the difference in the number of participants with a smaller rmsd on the first than on the second block (respectively 43/80 and 32/80) was not significant.

4.4.4. Model validation.

As in Study 1, each of the parameters in turn was clamped to a neutral value and with one parameter fixed, new estimations of the other two parameters were obtained. Table 8 and Figure 4 (right panels) show that except for the model with $m = 3$, each of these restricted models yielded a significantly worse fit than the full model. However, because the m parameter varies stepwise in units of 1, it is the case that for 15 subjects, the full model is exactly the same as the model with m fixed. When these participants—for whom no distinction is possible between the full and the nested model—are excluded from the test, the proportion of subjects for which the full model yielded a better fit amounted to 0.64; this proportion has a z -score of 2.36 ($p < .01$). Moreover, in each of the three nested models, the predictions of the autocorrelations of tasks were dramatically worse. In the predictions of the transition autocorrelations, the results were similar except that here the full model was not significantly better than the model with $m = 3$. When again, the 15 overlapping cases were excluded from the test, the proportion of subjects with better predictions for the full model amounted to .77 ($z = 4.34$, $p < .001$). Taken together, these findings support the conclusion that all three parameters and the underlying processes are needed to account for the data.

---- Table 8 about here ----

4.4.5. Discussion.

The application of the chain-retrieval model in the present experiment confirms and strengthens the findings reported in Study 1. Compared to the statistical independence model and

existing statistical dependence models, both the parameter fit to the runs proportions and the predictions of the autocorrelation statistic of tasks and transitions yielded a better correspondence to the data. In the validation test designed to investigate whether the processes underlying the three free parameters are all involved in achieving this good correspondence, each of the parameters in turn was clamped to a neutral value so as to exclude or to minimize the role of the underlying process. These analyses indicated that all three parameters of the model and the underlying processes are important in achieving the good correspondence with the data. Other assumptions for the underlying processes of the model were also investigated; these are reported in Appendix B. The variations considered relate to the way the chains are connected and the way bottom-up events affect task choice. Instead of working with overlapping chains, the possibility to work with concatenated chains was tested and instead of bottom-up repetitions that overrule the intended action, other types of intrusions were also considered. More details are provided in the appendix that shows that changing the present assumptions does not seem to improve the chain-retrieval model fitted to task sequences.

Separate model fitting to each of the two blocks in the present study shows a modest effect of practice, only the size of the r parameter changed with practice. These findings suggest that a small amount of practice at the start of the experiment suffices to obtain stable behavior. Over time, bottom-up events seem to become more important (increase in r), which suggests that top-down control over task selection becomes less strict as practice proceeds. In view of the high degree of similarity between the two blocks in both the degree of fit to the runs data and the degree of correspondence with the autocorrelations, it seems that the model yields quite stable results and is not jeopardized by accepting the simplifying assumption that the chain length is stable over the session.

If the hypothesis is correct that the b parameter expresses the sensitivity to the difference in difficulty between task repetitions and task switches, then as soon as task switching is involved, a bias towards choosing more repetitions would be expected. The question raised here, is whether such a bias would still be expected if the switching is irrelevant to the task choice. For example, would the bias still exist when participants are required to randomly choose which hand they will

use to perform the task? Will there be a bias towards selecting more hand repetitions or may it be expected that b would take a value close to neutral (close or equal to 0.5)? This question was investigated in Studies 3 and 4.

5. Study 3

In study 3, participants were requested to perform *only one task* throughout the experimental session (e.g., magnitude judgment) so that, strictly speaking, there were only task repetitions. Thus, they did not have to select tasks but they had to select the hand with which they wanted to execute the response. Hand switching is associated with a cost (Stelzel, Basten, & Fiebach, 2011), but this cost is due to associative carry-over in the hand repetition trials. Pashler and Baylis (1991) proposed that during execution of a task, a transient association between the stimulus identity and the executed response is formed, and if this association can be reused on the next trial, a repetition advantage is observed as the normal response-selection stage is bypassed. Applied to the present context and assuming that subjects represent their hand selection by means of an internal label, this implies that a shortcut between this internal label and the selected hand can be reused on hand repetition trials, bypassing the top-down controlled hand-selection stage. No such advantage is present in the hand switching situation, where both the internal label and the selected hand would be different. This account suggests that the difference between hand repetitions and hand switches is not based on a difference in the amount of mental effort needed for preparing and executing the tasks. As the bias towards repetitions is assumed to result from such differences in mental effort (Botvinick, 2007), it is expected, therefore, that in the case of hand switching no bias towards selecting more hand repetitions will develop. Hence, it is predicted that in such a hand-switching design, a hand switch cost (or hand alternation cost) may be observed (cf. Pashler & Baylis, 1991), but this will not result in a bias toward more frequent selection of hand repetitions. In other words, it is expected that the model with b fixed to 0.5 will not yield a worse fit and worse predictions than the full model. In addition, it should be noted that the present procedure constitutes a dual-task situation in which random hand selection is required while performing another task. Therefore, bottom-up events may be expected to play an important role.

5.1. Method

5.1.1. Participants and design

Thirty-five first-year psychology students at Ghent University participated for course requirements and credit. None of them had participated in one of the previous studies. They all had normal or corrected-to-normal vision, were right-handed, and all were naïve to the purpose of the experiment. Participants were randomly assigned to two conditions that differed in the task they performed throughout the session (see below). One participant was excluded for not following the instructions and producing errors on about half of the trials; a second participant was excluded for producing a pattern of task choices that was completely deviant from that of the other participants, so that as well the proportions of runs as the proportions of autocorrelations in this participant's choice data deviated more than 2 standard deviations from the sample mean.

5.1.2. Materials and procedure

The same procedure (single registration) as in Study 1 was used in combination with the stimuli and the tasks of Study 2. Participants were instructed to perform one single task throughout the experimental session that consisted of two blocks of 65 trials, preceded by a practice block of the same length. Half of the participants performed the magnitude task and pressed the left key for smaller and the right key for larger than 5. The other half was requested to execute the parity task and pressed the left key for odd and the right key for even. They were further instructed to switch hands in an unpredictable manner with the restriction of using both hands about equally often. During the practice block, they received feedback in the same way as in Study 2. At the end of each block a feedback screen showing performance during the block was presented, also completely similar as in Study 2.

5.2. Results

For completeness, analyses of task performance are presented in Appendix C. Here we focus on the sequential analysis. This data analysis was completely similar to that of Studies 1 and 2. The analysis was performed once with a focus on tasks (hands in the present study), and once with a focus on transitions. Both runs and autocorrelations were analyzed on the basis of the

multivariate linear model in a 2 (Task: magnitude or parity) \times 2 (Hand: left or right) \times 10 (Length or Lag) with repeated measures on the last two factors.

Previewing the findings, the hand-based analyses revealed larger runs proportions for the left than for the right hand, but decrease of runs proportions over lengths did not differ across the two hands. As in Studies 1 and 2, the autocorrelations showed a dip at lags 2-4, and although the average autocorrelations were higher for the left than for the right hand, the size of the dip did not differ across hands. In the transition-based analyses, proportion of hand repetitions was larger than the proportion of hand switches. Decrease of runs proportions over lengths was faster for repetitions than for switches. Autocorrelations were larger for repetitions than for switches and as in the previous studies, the autocorrelation at lag 1 was smaller than at the other lags.

More statistical details about these findings are presented in the following paragraphs.

Readers who prefer to skip these details can move directly to session 5.3 (Discussion).

5.2.1. Focus on hands

Hand selection data were coded as binary series of events in the same way as in Studies 1 and 2. More specifically, the 2 \times 64 hand choices was first coded 1 for left and 0 for right to generate runs and autocorrelation statistics. The codes were then reversed to calculate the values of the same statistics with the focus on the right hand (right = 1, left = 0). Also the transitions were coded, namely first with hand repetition coded as 1 and hand switching as 0 and then with the coding reversed. This way completely the same analysis was performed as in Studies 1 and 2, with the understanding that in the present study, there was no task switching but only hand switching.

The pattern of runs was completely similar to the one obtained in Studies 1 and 2. Selection proportions of the left hand and the right hand were .52 and .48 respectively. Overall, runs proportions were larger on the left hand ($M = .11$) than on the right hand ($M = .10$). This difference was significant, $F(1,31) = 7.59$, $\eta_p^2 = 0.20$. Furthermore, only the main effect of run length was reliable, $F(9,23) = 91434$, $\eta_p^2 = 1.00$. Importantly, hand did not interact with run length, $F(9,23) = 1.97$, $p = .09$.

Similarly, average autocorrelations were larger when focus was on the left hand ($M = .26$) than when focus was on the right hand ($M = .23$), $F(1,31) = 13.59$, $\eta_p^2 = 0.30$. Besides, autocorrelations varied over lags, $F(9,23) = 4.11$, $\eta_p^2 = 0.62$. As in Studies 1 and 2, autocorrelations were lower in the window at lags 2-4 than at longer lags (7-9), $F(1,31) = 15.17$, $\eta_p^2 = 0.33$. No other effects or interactions were reliable.

5.2.2. Focus on transitions

Proportions of hand repetitions ($M = .53$) were larger than proportions of hand switches ($M = .47$), but these values were not significantly different from 0.5, $t(32) = 1.37$, $p = 0.18$. On average runs proportions were larger for repetitions ($M = .11$) than for switches ($M = .09$), but this difference was not reliable ($F < 1$). Run length was significant, $F(9,23) = 23894$, $\eta_p^2 = 1.00$, as well as its interaction with transition type, $F(9,23) = 10.76$, $\eta_p^2 = 0.81$.

On average autocorrelations were larger for repetitions ($M = .30$) than for switches ($M = .23$), but this difference was not reliable. Autocorrelations only varied over lags, $F(9,23) = 10.70$, $\eta_p^2 = 0.81$. As in Studies 1 and 2, transition autocorrelations were smaller at lag 1 ($M = .23$) than at the other lags ($M = .27$), $F(1,31) = 72.22$, $\eta_p^2 = 0.70$. This contrast did not interact with the other effects.

5.3. Discussion

The present study differs from Studies 1 and 2 by not showing a significant repetition bias for hand selection even though the hand switch cost was reliable (see Appendix C), and replicates Studies 1 and 2 in revealing a similar pattern of findings both in the analysis of the hand choice and in the analysis of the transitions between hands. These similarities are obtained notwithstanding an important difference between the previous studies and the present one, namely that in the present study no task switching is needed as there is only one task, so that participants only have to select the hand they will use to perform the task.

5.4. Model fits

The present study was designed to specifically test the validity of the chain-retrieval model. In the experimental procedure of the present study, task switching was replaced by hand switching while performing one single task. As explained in the introduction to this study, on the basis of the

considerations that a hand switch cost may be observed but that this switch cost is not driven by task-set preparations, it was predicted that no bias would develop towards chains with more hand repetitions. Because a hand switch cost was observed, as well in error data as in RT data (see Appendix C for details), in order to verify the prediction, we must consider two possibilities. First, due to the presence of a hand-switch cost, participants may learn from this performance difference to prefer chains with more hand repetitions to chains with less hand repetitions. If that is the case, the present model application should simply replicate the results of Studies 1 and 2. The second possibility, is that notwithstanding the presence of a hand-switch cost, participants did not learn from the performance difference between hand repetitions and hand switches and did not acquire a bias towards more repetitions. Such failure to acquire a bias, given that a performance difference is present, is probably due to the fact that switching hands does not involve mental effort to prepare for using another hand, whereas in task switching, changing from one task to another involves a set of preparatory processes that can be modulated by performance monitoring (Botvinick, Nystrom, Fissell, Carter, & Cohen, 1999; Botvinick, 2007; Botvinick, Braver, Barch, Carter, & Cohen, 2001; Botvinick, Cohen, & Carter, 2004). If participants did not acquire a bias towards chains with more repetitions, then the present model application should not simply replicate Studies 1 and 2. More particularly, it would be expected that the b parameter would take a value close to neutral (0.5).

Because of this focused research question, the present section therefore reports only the tests of the chain-retrieval model. The full parameter models, and the three variants with one of the parameters clamped to a neutral value were tested. The questions addressed were whether the fit of the model with b fixed would be close to the fit obtained for the full parameter version and whether any of the other two-parameter models would be sufficient to explain the data.

Table 9 displays the fit and estimated parameters of the four model variants studied. The fit of the full-parameter chain-retrieval model was quite good. The same goes for the variants with m fixed and b fixed. In fact, the fit of the full model was not statistically better than the fit with either parameter fixed. The fit of the model with r fixed was substantially worse, although the difference fell short from statistical significance. When considering the predictions of the task

autocorrelations, the full parameter model was significantly better than the fixed- r model. The difference between the full and the m -fixed model was also significant when the overlapping cases were excluded ($z = 3.67, p < .001$). Only the difference with the b -fixed model was not statistically significant. The predictions of the transition autocorrelations yielded all significant differences in favor of the full-parameter model.

---- Table 9 about here ----

These findings show that the bias parameter (b) does not play a role when task difficulty differences are not at issue: neither in the parameter estimation nor in the prediction of autocorrelations, the full model was better than the model with b fixed. This observation is further substantiated by the finding that the correlation between the value of the b in the full model and the RT switch cost ($r = -0.43$, opposite direction than expected) and the correlation of b with the error switch cost ($r = 0.27, p = 0.13$) are not reliable. Even though hand switching was more difficult than hand repetition, this does not seem to intrude into the selection of a hand for executing a task.

That there is no bias towards chain with more repetitions does not mean that the repetition bias is completely abolished. The difference between the r -fixed and the full model shows that hand selection is still affected by bottom-up events. As well in the model fits as in the prediction of autocorrelations, the full model was reliably better than the model with r fixed to 0. The estimated proportion of bottom-up triggered repetitions seems to be smaller than in Studies 1 and 2. A test of the difference showed a marginally significant difference, $F(1,127) = 3.30, p = .07, \eta_p^2 = 0.03$.

Finally, in comparing the full model to the model with m fixed, the fits showed that the full model was not better than the restricted model, but in the prediction of the autocorrelations, the difference between both sets of predictions was significant. This observation indicates that the preferred chain length is an important modulator of selection and that it plays an important role in achieving the tendency to repeat present in this data set, the more so that bias does not seem to be effective to augment proportions of repetitions and that fewer bottom-up repetitions do occur.

6. Study 4

Study 3 tested the chain-retrieval model in a dual-task context: it required hand switching while performing another task. Although a hand-switch cost was observed, no bias was acquired

on the basis of this difference in difficulty between hand repetitions and hand switches. In that context, the m and r parameters could not be discarded. The question addressed in Study 4 is whether the processes underlying these parameters would still play an important role, when the context is further simplified, i.e., when the dual-task characteristic is removed. In Study 4, the situation was completely similar as in Study 3, except that no target-based task processing was required. On every trial, a fixed probe was presented to which participants responded by pressing a key either with the left or with the right hand. Hence, the task situation only required voluntary hand selection. The research questions are (a) whether the chain-retrieval model would still be better than the statistical independence and dependence models, and (b) whether the three parameters would still be instrumental in obtaining good fits and predictions.

6.1. Method

6.1.1. Subjects.

Twenty-four participants were selected from the same pool as Study 3. They all served in a single condition in which they randomly switched between hands. One subject failed to emit a response on about one third of the trials and was not included in the data analysis.

6.1.2. Materials and procedure

Participants were instructed to switch hands in an unpredictable manner with the restriction of using both hands about equally often. Each trial consisted of a probe (exclamation mark) that signaled that a response was required, namely to either press a left key with the left hand or to press a right key with the right hand. Time allowed for responding was 3000 ms. During the practice block, they received feedback. At the end of each block a feedback screen showing performance during the block was presented completely similar as in Study 2. The practice block contained 65 trials; this was followed by two experimental blocks of 65 trials each.

6.2. Results

RTs were not faster for repetitions than for switches (Ms respectively 332 and 359 ms), $F(1,22) = 2.04$, $p = 0.17$, $\eta_p^2 = 0.08$. The sequence data were analyzed once with a focus on hands and once with a focus on transitions between hands. Both runs and autocorrelations were

analyzed on the basis of the multivariate linear model in a 2 (Hand: left or right) \times 10 (Length or Lag) repeated measures design .

6.2.1. Focus on hands

Hand selection data and transition data were coded in the same way as in Study 3. This way we could conduct completely the same analyses as in the previous studies, with the understanding that in the present study, hand switching was the only task.

The pattern of runs was completely similar to the one obtained in the previous studies. Selection proportions of the left hand and the right hand were .51 and .49 respectively. Overall, runs proportions were not larger on the left hand ($M = .10$) than on the right hand ($M = .10$), $F < 1$. The main effect of run length was reliable, $F(9,14) = 6.16$, $\eta_p^2 = 0.80$. Importantly, hand did not interact with run length, $F(9,14) = 1.13$, $p = .40$.

Average autocorrelations were slightly larger when focus was on the left hand ($M = .25$) than when focus was on the right hand ($M = .24$), but this difference was not reliable, $F < 1$. Autocorrelations varied over lags, but again not reliably, $F(9,23) = 1.40$, $p = .28$, $\eta_p^2 = 0.47$. In contrast to the previous studies, autocorrelations were not lower in the window at lags 2-4 than at longer lags (7-9), $F < 1$, $\eta_p^2 = 0.04$. No other effects or interactions were reliable.

6.2.2. Focus on transitions

Proportions of hand repetitions ($M = .53$) were larger than proportions of hand switches ($M = .47$), but these values were not significantly different from 0.5, $t(22) = 1.08$, $p = 0.29$. On average runs proportions were not larger for repetitions ($M = .10$) than for switches ($M = .10$), $F < 1$. Run length was significant, $F(9,14) = 4.81$, $\eta_p^2 = 0.76$, but its interaction with transition type was not, $F(9,14) = 2.02$, $p = .12$, $\eta_p^2 = 0.57$.

On average, autocorrelations were larger for repetitions ($M = .27$) than for switches ($M = .25$), but this difference was not reliable, $F < 1$. Autocorrelations did not vary reliably over lags, $F(9,14) = 1.23$, $p = .35$, $\eta_p^2 = 0.44$.

6.3. Discussion

Like Study 3, the present study did not show a repetition bias for hand selection. The pattern in the autocorrelations was completely different from the pattern found in the three previous

studies, both in the hand choice and in the transitions. The most striking feature is that neither a hand preference, nor a repetition bias is observed. On the basis of the random generation literature, an alternation bias could have been expected. In comparison to typical random generation experiments, the time allowed per trial was rather large in the present study (3 s), so that there was enough time to overcome such bias. Note, however, that hand choice has so far not been used in random generation tasks, and possibly hand selection is sensitive to bottom-up intrusions. The model application will show that this may indeed have been the case.

6.4. Model fits

Like Study 3, the present study was designed to specifically test the validity of the chain-retrieval model, presently in a situation where only random choice of hands is required. Because this is the kind of situation that corresponds most closely to typical random generation tasks, first a comparison is made between the chain-retrieval model and the three statistical models tested in Studies 1 and 2, namely the independence model, and the perseveration and alternation models. Next, an analysis like the one presented in Study 3 is performed to check whether the assumptions of the chain-retrieval model all hold in this more simple kind of task setting.

Table 10 presents an overview of the findings comparing the chain-retrieval model to the statistical models. Although the goodness-of-fit of the statistical models now seems to be slightly better than in Studies 1 and 2, the chain-retrieval model still yields a better fit, even when the number of free parameters is taken into account; this model yields a statistical significant better fit than each of the other models. In terms of the predictions, all models seem to be doing equally well. This is not very informative on the relative quality of the models, as the data analysis had already shown that the autocorrelation data did not vary over the lags.

---- Table 10 about here ----

The comparison of the full and nested variants of the chain-retrieval model is summarized in Table 11. The table shows that each of the nested models provided an equally good fit as the full model. When considering also the predictions, only the model with $r = 0$, produced significantly worse predictions than the full-parameter model. Similar to Study 3, the proportion of bottom-up triggered repetitions tended to be smaller. Indeed, when contrasting the number of bottom-up

triggered repetitions in Studies 1 and 2 with the number in Studies 3 and 4, less such repetitions did occur in the latter two studies, $F(1,149) = 4.92$, $p < .05$, $\eta_p^2 = 0.03$. As in Study 3, the full model was not better than the model with b fixed; the correlation of b with the hand-switch cost was close to zero ($r = 0.09$, $p = 0.70$).

---- Table 11 about here ----

By and large, this study shows that the chain model accounts well for random generation data, while parameters m and b are fixed. A good fit was obtained with r fixed to 0, but predictions of autocorrelations were better when r varied freely. Apparently, bottom-up repetitions (e.g., due to repetition priming) did occur in hand switching. In a task context with less opportunities for bottom-up repetitions, fixing r to 0 may be expected to yield good correspondence. That the chain-retrieval model achieves a good account of random generation data with so many of the free parameters fixed further supports the assumptions that go back to the model of Rapoport and Budescu (1997).

7. General Discussion

7.1. Overview of the findings

The objective of the present study was to develop a framework that specifies the processes involved in voluntary task selection, to define a formal model on the basis of the framework, test this model in situations involving task selection, and test the limits of the model by applying it also to event selection. On the basis of functional similarities between VTS and the selection of independent events as studied in the random generation experiments, we formulated the central hypothesis of the framework, namely that in order to produce sequences of task choices in which both tasks occur independently and about equally often, people quickly learn to retrieve short balanced sequences (in line with the model of Rapoport & Budescu, 1997) of task names from LTM to guide the trial-by-trial choice of tasks. We assumed that the length of the chains is constrained by working memory capacity, although we noticed that chain length is also related to the proportion of task repetitions in the chains. We further assumed that there is a bias towards retrieval of chains with more repetitions, and that bottom-up triggered repetitions may overrule the currently intended event. This view was specified in a model with three free parameters

corresponding to these hypothesized processes. This model's performance was compared with that of a statistical independence model (Bernoulli) and with performance of two statistical dependence models, one with a repetition bias (perseveration model) and one with an alternation bias (alternation model).

To compare these models, we analyzed task-choice data from two perspectives, namely as a sequence of tasks and as a sequence of transitions between tasks. The task-based analysis of the VTS data in Studies 1 and 2 confirmed the task repetition bias in task selection (Arrington & Logan, 2004). The autocorrelation statistic showed that this bias was strongly present at lag 1 and dipped on subsequent lags, which is consistent with the avoidance of lag-2-repetitions in VTS (Lien & Ruthruff, 2008) but then stabilized at an intermediate level from lag 5 on. In other words, when participants executed task A, there is a tendency to immediately repeat task A (lag-1 repetition/ autocorrelation), and to switch to the other task one or two times in the lag 2-5 window. At lags longer than 5, there is no strong evidence for any statistical dependency. This pattern is different from what is typically found in the random-generation literature: when people select random events, they often tend to alternate at Lag 1. The transition-based analysis of the choice data confirmed the presence of a global repetition bias supported by a dominance in the proportion of repetitions and a stable level of relatively high autocorrelations on lags greater than 1. At lag 1, the tendency to repeat the same transition was lower, which suggests that actual lag-2 repetitions were not avoided. The same pattern of findings was also present in Study 3 which did require hand switching instead of task switching, although the repetition bias was not so strong in this study.

In Studies 1 and 2, the parameters of the chain-retrieval model and of the three statistical models were estimated separately for each participant by means of a maximum likelihood method taking into account the number of free parameters. This analysis showed that the chain-retrieval model yielded a better account of the sequential choices than any of these other models. Also in a comparison of the full-parameter model to restricted models in which one of the parameters was clamped to a neutral value, the full model produced better fits and better predictions than the restricted models.

In order to further test the processes underlying the chain-retrieval model, in Study 3 the model was applied to a dual-task situation in which no task switching, but instead hand switching was required. The fit of the full-parameter model was not statistically better than that of the nested two-parameter models. Taking into account also the predictions of the autocorrelation data, it appeared that only the variant with $b = 0.5$ was completely equivalent to the full model. Even though a hand switch cost was present in RT and error data, this did not result in the acquisition of a bias towards the easier chains with more hand repetitions. It was argued that this is due to a qualitative difference between task switching and hand switching. Task switching requires task-set configuration, a process that involves mental effort and is modulated by performance monitoring (Botvinick, 2007; Botvinick et al., 2001). As explained in the introduction to Study 3, the reason for an RT difference in hand repetitions and hand switches fits with the pattern of data reported by Pashler and Baylis (1991), and can be accounted for by re-usage of a response-selection shortcut in the case of hand repetitions. The difference in RTs is not based on a difference on mental effort and hence could not be used to develop a preference for the more easy hand repetitions.

Study 4 compared the chain-retrieval model to the statistical models as well as to the variants with one of the parameters clamped to a neutral value. In this study, participants only switched between response hands. Even though the statistical dependence models should provide a good fit to such data, the chain-retrieval model's fit (corrected for the number of free parameters) was still better than that of the three statistical models. The comparison of the full model to the models with one parameter fixed revealed no longer an advantage for the full model. Based on the goodness-of-correspondence of data and predictions, the model without bottom-up events produced poorer correspondence to the data, so that this study provides evidence that the bottom-up events do also occur in simple event selection. Interestingly, however, the likelihood of these bottom-up triggered repetitions was smaller than in the task switching studies (Studies 1 and 2). This observation supports the hypothesis that VTS is more vulnerable to bottom-up repetitions than is the case in independent event selection as in Studies 3 and 4. A possible reason why bottom-up events may affect VTS more relates to the larger number of requirements in VTS (select

tasks, randomly and correct execution of the selected tasks) than in random event selection (random selection without any behavioral consequences).

By way of summary, the chain-retrieval model yields a better account than the statistical independence and the statistical dependence models, not only in contexts of VTS with single (Study 1) and double registration (Study 2), but also in a context of hand switching (Study 4). The analyses support the conclusion that in the context of task switching, the three postulated processes of the model (namely chain selection, chain-retrieval biasing and bottom-up triggered repetitions), all have a role to play. Even though, the extent to which either of these processes contribute to the generation of task choices can be modulated for the absence of the others, this modulation does not suffice to completely account for the data when one of these processes is blocked in the model. In contexts without task switching, the model still accounts for the observed selections, but this can be achieved by means of only two parameters (r and m), and in a context of simple hand choice, one parameter (r) seems to suffice.

7.2. Theoretical implications

In this section, we discuss the theoretical basis and implications of the assumptions we made in developing the chain-retrieval model. This discussion considers the assumptions related to the three free parameters.

The three parameters of the model were introduced to grasp particular constraints of the task selection process in VTS. Parameter m was introduced to specify the variability in sequences of choices produced by different persons. In developing the sets of balanced chains of different lengths, two observations may be of interest. First, all these sets of chains are characterized by an alternation bias, which is less strong in chains of lengths 5-6 than in chains of lengths 3-4. Second, by increasing the bias (b) towards more repetitions, there is more room for creating a repetition bias in the longer than in the shorter chains. In other words, the preferred chain length does not only depend on the limited capacity available for processing chains of a particular length, the preference for a particular length may also depend on the maximum proportion of repetitions that can be produced by chains of a particular length. Taken together, these observations suggest that

preferred chain length may be based on a balance between capacity limits and limits in achievable repetition bias.

A second parameter, b , biases the selection of sequences in such a way that sequences are more likely to be retrieved when they contain more repetitions compared to when they contain more alternations. The inclusion of the b parameter in the modeling shows that the probability of retrieving particular chains depends on the ease of execution, as was confirmed by the significant correlation between b and the accuracy difference between repetitions and switches. In Studies 3 and 4 which did not require task switching, the bias parameter was not needed to account for the data. As already extensively argued, the task-switch cost, but not the hand-switch cost, is intricately related to task-set reconfiguration and performance monitoring to achieve the best performance with the least mental effort.

The good correspondence of the model's predictions and the data is consistent with the idea that in voluntary task choice, there is a bias towards a selection of sequences with more repetitions. How can this be explained? One obvious possible explanation relates to the assumption we adopted to motivate the inclusion of the b parameter, namely that participants develop such a bias during task execution on the basis of the experience that repetitions are easier to execute than switches. Whether participants indeed develop such a bias, is one of the avenues for further research that follows from the present modeling attempt. Making switching easier or more difficult to execute should affect the value of this parameter. This is in line with Botvinick's integrative account of the functioning of the anterior cingulate cortex (Botvinick, 2007).

The third parameter, r , specifies the probability that a repetition overrules the intended action. The value of the parameter overestimates the real proportion of bottom-up intrusions, because on some trials the bottom-up triggered event and the intended event are the same. Correction for this overlap shows that the estimated proportion of bottom-up triggered repetitions hovers around .11 in task switching and around .07 when no task switching is involved. This parameter was included to account for previous findings regarding bottom-up effects in VTS. The parameter value also suggests that these kinds of events are happening quite often. This confirms the difficulty of remaining in control of task selections while trying to avoid to be distracted by exogenous events

(Mayr & Bell, 2006), although in the context of Study 2, the effect cannot be attributed to the presence of stimulus repetitions, as the task choice occurs in response to a prompt, before stimulus presentation.

The particular choice for bottom-up intrusions that overrule the intended action has some advantages over other possible rules that could be used to implement the model. Other possibilities are that the bottom-up intrusions intervene between the intended events but do not overrule the intention or that an intrusion intervenes and blocks further execution of the chain. These alternatives are conceptually less attractive because in these alternatives the presently intended action is never overruled, but only postponed. Also empirically, these alternative rules do not seem to account so well for the data, as can be seen in Appendix B that reports tests of model variations using these alternative rules.

Another interpretation of the r parameter is that it may reflect execution errors on the part of the participants. Erroneously executing a repetition instead of a switch would have the same effect. Given that error rates in task-switching research are usually around 5%, errors cannot completely account for the value of r . The interpretation that r represents bottom-up intrusions that could not be blocked by top-down control mechanisms is also supported by the observation that the size of r depended on the SOA between the previous stimulus and the present task probe. The value of r was namely larger with a short SOA. This is exactly what would be expected if this parameter is related to intrusions. On the same count, no such effect is expected for the other parameters and this was also confirmed. Further tests of the model could include other well chosen design variations, such as the number of target repetitions (Mayr & Bell, 2006) or invoked response repetitions. Data based on such procedural variations should lead to larger values for r in conditions that result in more repetitions.

While the present study was completely framed within the VTS procedure, the modeling reported here, does have implications that go beyond VTS. First, the chain-retrieval model was designed to provide an explanation of task choice in VTS, but it can also account for the different pattern of findings in the production of sequences of independent events. The model is, for example, also applicable to binary randomization tasks, such as coin tossing. Study 4 reports

such an application based on random hand switching. The findings showed that, as expected, bias towards repetitions did not play a role. However, the task was still sensitive to bottom-up intrusions, be it that the amount of intrusions was significantly smaller than in a task-switching context. Considering that the chain-retrieval model was developed starting with assumptions from the Rapoport and Budescu (1997) model, it can be considered as an extension of the latter model complementing it with other processes such as biased selection and bottom-up intrusions. The results of Study 4 and these considerations raise the question, whether and to what extent typical random choice behavior is also affected by bottom-up processes.

7.3. Avenues for further research

One obvious question for further research concerns the role of working memory in relation to the present framework. As was already mentioned, the preferred chain length is, among other factors, probably constrained by working memory capacity. Also the frequency with which bottom-up intrusions occur may depend on the availability of attentional resources, which are central in most working memory models.

There is little doubt that further empirical work to substantiate these relationships between the hypothesized mechanisms of task selection and working memory would be most valuable. One major reason is that, thus far, many attempts to clarify the relationship between task switching and working memory have stranded in null results (see Vandierendonck et al., 2010, for a brief discussion), and only a few studies have reported positive results. As the potential role of working memory is explicit in at least two processes of the present framework, further tests of these relations do seem worthwhile.

Further testing of the relationship between preferred chain length and working memory capacity does seem straightforward. Important evidence has already been reported as to the role of working memory in sequence generation. Baddeley et al. (1998), for example, found that the deviation of randomness in generated key presses increased when an irrelevant memory load was larger, which shows that maintaining a memory load may interfere with cognitive control processes needed for the generation task. Several studies have also reported poorer recall on a memory task when an unrelated random generation task was performed concurrently (e.g., Fisk & Sharp,

2003; Macizo, Bajo, & Soriano, 2006; Towse & Cheshire, 2007; Vandierendonck, 2000a, 2000b; Vandierendonck, De Vooght, & Van der Goten, 1998a, 1998b; Vandierendonck, Kemps, Fastame, & Szmalec, 2004), suggesting that random generation interferes with maintenance of unrelated memory contents. Based on these findings, the expectation may be formulated that performing VTS under a memory load could result in the usage of shorter chains. This may be modulated by two factors. First, if the tendency to avoid the more difficult switches is strong, longer chains may be preferred and efforts may be deployed to work with longer chains. Second, the impact of a memory load depends on the amount of working memory capacity available. Persons with a larger working memory capacity are more able to maintain a load while performing another demanding cognitive task (e.g., Engle, Kane, & Tuholski, 1999). Therefore, the expectation that shorter chains would be used in VTS under a memory load would probably mainly affect persons with a low working memory capacity. Therefore, these modulating factors must be taken into account in a test of the relationship between working memory and actually used chain length. It is worth noting that one study of the role of working memory in VTS did not find a relationship between task choice and working memory capacity (Butler, Arrington, & Weywadt, 2011). As the study was performed outside the present framework, unfortunately, it did not provide a direct test of the hypothesis proposed here.

The second issue of interest is whether working memory mediates the frequency of bottom-up intrusions. Starting from the finding that stimulus repetitions enhance task repetitions (Mayr & Bell, 2006), one study already showed that this effect is strengthened under a working memory load (Demantet et al., 2010). It is quite likely that a memory load sets the stage for more bottom-up intrusions (see e.g., Lavie, Hirst, de Fockert, & Viding, 2004); consequently, the r parameter is expected to increase when load increases, also because the control processes that block upcoming intrusions are expected to be less efficient when there is a working memory load. In this sense, voluntary task selection involves blocking of “unwilled” actions. In terms of the task-chain model, this means that the more choices are voluntary, the lower the value of r should be. As the average value was lower at long than at short SOA, it seems that at long SOA more intrusions were blocked and that consequently more voluntary task choices were made. Within the

application of the model, it is possible to estimate how frequently an intended switch was overruled by a bottom-up triggered repetition. This is, however, merely an estimate. Under the hypothesis that under working memory load more bottom-up intrusions overrule an intended action, it follows that the estimated proportion of overruled switches should be larger under load than without load, a prediction worthwhile to follow up.

If it is the case that under load, task switching would become more difficult, this would also imply that a larger b -value would be expected under load than without load. This could also imply that low working memory subjects would also be more susceptible to the effects of load and develop a stronger bias towards repetitions, which would require the selection of longer chains. Taken together, in an individual differences approach, participants with larger working memory capacity would be observed to prefer longer chains, and they would be expected to be less sensitive to bottom-up intrusions. In terms of the modeling, high working memory participants would show a larger value for m and a smaller value for r and b than low working memory participants. Clearly such a test will not be easy because of the possible compensatory effects between the processes involved.

Apart from the role of working memory, the present framework also provides a basis for further exploring the conditions under which differences in difficulty between transitions or between tasks may affect the propensity to avoid the more difficult tasks or transitions in task choice. This assumption of the framework seems to work well, but it was not explicitly tested in the present project. The work of Botvinick (Botvinick, 2007; Botvinick & Rosen, 2009) regarding the role of mental effort offers one entry for further investigation. This also includes a link to work on performance monitoring. In fact, a bias towards the more easy repetitions as proposed in the chain-retrieval model must be based on such processes. This entails, checking for errors, for response slowing, for conflicts between response alternatives, and using this information as feedback in the acquisition of the bias towards more easy tasks, which in the end helps to maintain an acceptable level of performance. In the context of VTS, the effect of asymmetric switch costs on task choice (e.g., Liefoghe et al., 2010) and the role of lag 2 repetition costs (e.g., Lien &

Ruthruff, 2008) may be further pursued by linking these effects to performance and conflict monitoring

As a final remark, the idea of retrieving stored chains of tasks to guide further action is quite similar to but also different from retrieving a plan of actions. When people make a plan, in order to achieve the goal, appropriate subgoals are retrieved and chained into a sequence of steps. These chains are stored in LTM and can be retrieved for later usage. The retrieved chains guide task execution which subsequently results in achieving the goal. Apart from this similarity, there are also differences. The chains considered in the chain-retrieval model clearly are no action plans. The chains are rather a means to achieve the goal of performing well in a particular task setting with particular constraints. Nevertheless, further research on how chain retrieval could develop into plan retrieval would, no doubt, be worthwhile. The step from chain-retrieval to sequence planning may be rather small and if so, it is no coincidence that there are some similarities between the present modeling and the task-span procedure introduced by Logan (2004a, 2006).

7.4. Conclusion

The present study showed that the repetition bias in VTS is due to a globally-based statistical dependency in the sequence of selected tasks or transitions. The model based on chain retrieval accounts very well for these data and suggests (a) that short sequences of tasks are retrieved from LTM to guide task selection, (b) that the length of these sequences varies over individuals and (c) is affected by the probability that the sequences contain repetitions linked to ease of execution and may also be constrained by working memory capacity, and finally (d) that bottom-up intrusions of repetitions play an important part in the repetition bias. This model has a number of implications that can be tested in future research. The memory processes postulated in the model propose a hypothesis on how sequences of independent events in general and more particularly, how task selection in VTS may occur. The assumption that task selection is based on chains of events retrieved from memory provides a possible answer regarding the cognitive control of task selection, namely that in VTS, not single tasks but chains of tasks or possibly even action plans are selected. Finally and not least important, the modeling also indicates that the adjective

“voluntary” in voluntary task switching may refer to the process of selecting tasks but also to the process of blocking intrusions.

Acknowledgement

The research reported in this article was supported by grant 01J07707 of the Research Council of Ghent University to the first author; Baptist Liefoghe is supported by grant BOF09/01M0029 of the Research Council of Ghent University.

We are indebted to Marcel Brass, Wouter De Baene, and Arnaud Szmalec for their helpful comments on an earlier version the present article. We also want to express our gratitude to Erik Altmann, Catherine Arrington, an anonymous reviewer, and the action editor, Tom Carr, for their constructive and useful comments on earlier versions of this article.

Correspondence about the paper can be directed to: André Vandierendonck, Department of Experimental Psychology, Ghent University, Henri Dunantlaan 2, B-9000 Gent, Belgium. E-mail: Andre.Vandierendonck@UGent.be.

References

- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*, 716–723. doi: 10.1109/TAC.1974.1100705.
- Allport, D. A., Styles, E. A., & Hsieh, S. (1994). Shifting attentional set: Exploring the dynamic control of tasks. In C. Umiltá & M. Moscovitch (Eds.), *Attention and Performance XV: Conscious and nonconscious information processing* (pp. 421-452). Cambridge, MA: MIT Press
- Anderson, J. R., & Lebiere, C. (1998). *The Atomic Components of Thought*. New York: Lawrence Erlbaum Associates.
- Arrington, C. M. (2008). The effect of stimulus availability on task choice in voluntary task switching. *Memory & Cognition*, *36*, 991-997. doi: 10.3758/mc.36.5.991.
- Arrington, C. M., & Logan, G. D. (2004). The cost of a voluntary task switch. *Psychological Science*, *15*, 610-615. doi: 10.1111/j.0956-7976.2004.00728.x.
- Arrington, C. M., & Logan, G. D. (2005). Voluntary task switching: Chasing the elusive homunculus. *Journal of Experimental Psychology: Learning Memory and Cognition*, *31*, 683-702. doi: 10.1037/0278-7393.31.4.683.
- Arrington, C. M., & Rhodes, K. M. (2010). Perceptual asymmetries influence task choice: The effect of lateralised presentation of hierarchical stimuli. *Laterality*, *15*, 501-513. doi: 10.1080/13576500902984695.
- Arrington, C. M., Weaver, S. M., & Pauker, R. L. (2010). Stimulus-Based Priming of Task Choice During Voluntary Task Switching. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *36*, 1060-1067. doi: 10.1037/a0019646.
- Arrington, C. M., & Yates, M. M. (2009). The role of attentional networks in voluntary task switching. *Psychonomic Bulletin & Review*, *16*, 660-665. doi: 10.3758/pbr.16.4.660.
- Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation* (Vol. 2, pp. 89-195). New York: Academic Press

- Baddeley, A. (1996). Exploring the central executive. *Quarterly Journal of Experimental Psychology, 49A*, 5-28.
- Baddeley, A. (2007). *Working memory, thought, and action*. Oxford: Oxford University Press.
- Baddeley, A., Chincotta, D., & Adlam, A. (2001). Working memory and the control of action: Evidence from task switching. *Journal of Experimental Psychology: General, 130*, 641-657. doi: 10.1037//0096-3445.130.4.641.
- Baddeley, A., Emslie, H., Kolodny, J., & Duncan, J. (1998). Random generation and the executive control of working memory. *Quarterly Journal of Experimental Psychology, 51A*, 819-852.
- Baddeley, A. D. (1966). The capacity for generating information by randomization. *Quarterly Journal of Experimental Psychology, 18*, 119-129.
- Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 8, pp. 47-89). New York: Academic Press
- Botvinick, M., Nystrom, L. E., Fissell, K., Carter, C. S., & Cohen, J. D. (1999). Conflict monitoring versus selection-for-action in anterior cingulate cortex. *Nature, 402*, 179-181.
- Botvinick, M. M. (2007). Conflict monitoring and decision making: Reconciling two perspectives on anterior cingulate function. *Cognitive Affective & Behavioral Neuroscience, 7*, 356-366. doi: 10.3758/CABN.7.4.356.
- Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review, 108*, 624-652. doi: 10.1037//0033-295x.108.3.624.
- Botvinick, M. M., Cohen, J. D., & Carter, C. S. (2004). Conflict monitoring and anterior cingulate cortex: an update. *Trends in Cognitive Science, 8*, 539-546.
- Botvinick, M. M., & Rosen, Z. B. (2009). Anticipation of cognitive demand during decision-making. *Psychological Research, 73*, 835-842. doi: 10.1007/s00426-008-0197-8.
- Brent, R. P. (1973). *Algorithms for minimization without derivatives*. Englewood Cliffs, NJ: Prentice-Hall.
- Bryck, R. L., & Mayr, U. (2005). On the role of verbalization during task set selection: Switching or serial order control? *Memory & Cognition, 33*, 611-623.

- Budescu, D. V. (1987). A Markov model for generation of random binary sequences. *Journal of Experimental Psychology: Human Perception and Performance*, *13*, 25-39.
- Buse, A. (1982). The likelihood ratio, Wald, and Lagrange Multiplier Tests: An expository note. *The American Statistician*, *36*, 153-157.
- Bush, R. R., & Estes, W. K. (Eds.). (1959). *Studies in mathematical learning theory*. Stanford: Stanford University Press.
- Butler, K. M., Arrington, C. M., & Weywadt, C. (2011). Working memory capacity modulates task performance but has little influence on task choice. *Memory & Cognition*, *39*, 708-724. doi: 10.3758/s13421-010-0055-y.
- Demant, J., Verbruggen, F., Liefvooghe, B., & Vandierendonck, A. (2010). Voluntary task switching under load: Contribution of top-down and bottom-up factors in goal-directed behavior. *Psychonomic Bulletin & Review*, *17*, 387-393. doi: 10.3758/PBR.17.3.387.
- Emerson, M. J., & Miyake, A. (2003). The role of inner speech in task switching: A dual-task investigation. *Journal of Memory and Language*, *48*, 148-168. doi: 10.1016/S0749-596X(02)00511-9.
- Engle, R. W., Kane, M. J., & Tuholski, S. W. (1999). Individual differences in working memory capacity and what they tell us about controlled attention, general fluid intelligence, and functions of the prefrontal cortex. In A. Miyake & P. Shah (Eds.), *Models of working memory. Mechanisms of active maintenance and executive control* (pp. 102-134). Cambridge: Cambridge University Press
- Fisk, J. E., & Sharp, C. A. (2003). The role of the executive system in visuo-spatial memory functioning. *Brain and Cognition*, *52*, 364-381.
- Hull, C. L. (1943). *Principles of behavior. An introduction to behavior theory*. New York: Appleton, Century & Crofts.
- Hurvich, C. M., & Tsai, C. L. (1989). Regression and time-series model selection in small samples. *Biometrika*, *76*, 297-307.

- Kessler, Y., Shencar, Y., & Meiran, N. (2009). Choosing to switch: Spontaneous task switching despite associated behavioral costs. *Acta Psychologica, 131*, 120-128. doi: 10.1016/j.actpsy.2009.03.005.
- Kiesel, A., Steinhauser, M., Wendt, M., Falkenstein, M., Jost, K., Philipp, A. M., et al. (2010). Control and interference in task switching—A review. *Psychological Bulletin, 136*, 849-874. doi: 10.1037/a0019842.
- Lavie, N., Hirst, A., de Fockert, J. W., & Viding, E. (2004). Load theory of selective attention and cognitive control. *Journal of Experimental Psychology: General, 133*, 339-354.
- Liefooghe, B., Demanet, J., & Vandierendonck, A. (2009). Is advance reconfiguration in voluntary task switching affected by the design employed? *Quarterly Journal of Experimental Psychology, 62*, 850-857. doi: 10.1080/17470210802570994.
- Liefooghe, B., Demanet, J., & Vandierendonck, A. (2010). Persisting activation in voluntary task switching: It all depends on the instructions. *Psychonomic Bulletin & Review, 17*, 381-386. doi: doi:10.3758/PBR.17.3.381.
- Lien, M.-C., & Ruthruff, E. (2008). Inhibition of task set: Converging evidence from task choice in the voluntary task-switching paradigm. *Psychonomic Bulletin & Review, 15*, 1111-1116.
- Logan, G. D. (2004a). Working memory, task switching, and executive control in the task span procedure. *Journal of Experimental Psychology: General, 133*, 218-236.
- Logan, G. D. (2004b). Working memory, task switching, and executive control in the task span procedure. *Journal of Experimental Psychology: General, 133*, 218-236.
- Logan, G. D. (2006). Out with the old, in with the new: More valid measures of switch cost and retrieval time in the task span procedure. *Psychonomic Bulletin & Review, 13*, 139-144.
- Logan, G. D., & Gordon, R. D. (2001). Executive control of attention in dual-task situations. *Psychological Review, 108*, 393-434.
- Lopes, L. L. (1982). Doing the impossible: A note on induction and the experience of randomness. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 8*, 626-636.
- Lopes, L. L., & Oden, G. C. (1987). Distinguishing between random and nonrandom events. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13*, 392-400.

- Luce, R. D. (1959). *Individual choice behavior*. New York: Wiley.
- Macizo, P., Bajo, T., & Soriano, M. F. (2006). Working memory and executive control: Inhibitory processes in updating and random generation tasks. *Psicothema, 18*, 112-116.
- Mayr, U., & Bell, T. (2006). On how to be unpredictable: Evidence from the voluntary task-switching paradigm. *Psychological Science, 17*, 774-780. doi: 10.1111/j.1467-9280.2006.01781.x.
- Mayr, U., & Keele, S. W. (2000). Changing internal constraints on action: The role of backward inhibition. *Journal of Experimental Psychology: General, 129*, 4-26.
- Mayr, U., & Kliegl, R. (2000). Task-set switching and long-term memory retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 1124-1140.
- McClelland, J. L., & Rumelhart, D. E. (1986). *Parallel distributed processing. Explorations in the microstructure of cognition. Vol. 2. Psychological and Biological models*. Cambridge, MA: MIT Press.
- Meiran, N. (1996). Reconfiguration of processing mode prior to task performance. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*, 1423-1442.
- Meiran, N. (2008). The dual implication of dual affordance: Stimulus-task binding and attentional focus changing during task preparation. *Experimental Psychology, 55*, 251-259.
- Miyake, A., Emerson, M. J., Padilla, F., & Ahn, J. C. (2004). Inner speech as a retrieval aid for task goals: the effects of cue type and articulatory suppression in the random task cuing paradigm. *Acta Psychologica, 115*, 123-142. doi: 10.1016/j.actpsy.2003.12.004.
- Monsell, S. (2003). Task switching. *Trends in Cognitive Sciences, 7*, 134-140.
- Neuringer, & Allen. (1986). Can people behave "randomly": The role of feedback. *Journal of Experimental Psychology: General, 115*, 62-75.
- Norman, D. A., & Shallice, T. (1986). Attention to action: Willed and automatic control of behavior. In R. J. Davidson, G. E. Schwartz & D. Shapiro (Eds.), *Consciousness and self-regulation* (Vol. 4, pp. 1-18). New York: Plenum Press
- Orr, J. M., & Weissman, D. H. (2011). Succumbing to bottom-up biases on task choice predicts increased switch costs in the voluntary task switching paradigm. *Frontiers in Psychology, 2*: 31 doi: 10.3389/fpsyg.2011.00031.

- Pashler, H., & Baylis, G. (1991). Procedural learning: 2. Intertrial repetition effects in speeded choice tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*, 33-48. doi: 10.1037/0278-7393.17.1.33.
- Rapoport, A., & Budescu, D. V. (1992). Generation of random series in two-person strictly competitive games. *Journal of Experimental Psychology: General*, *121*, 352-363.
- Rapoport, A., & Budescu, D. V. (1997). Randomization in individual choice behavior. *Psychological Review*, *104*, 603-617.
- Rescorla, R. A. (1988). Pavlovian conditioning. It's not what you think it is. *American Psychologist*, *43*, 151-160.
- Rogers, R. D., & Monsell, S. (1995). Costs of a predictable switch between simple cognitive tasks. *Journal of Experimental Psychology: General*, *124*, 207-231.
- Rumelhart, D. E., & McClelland, J. L. (1986). *Parallel distributed processing. Explorations in the microstructure of cognition. Vol. 1. Foundations*. Cambridge, MA: MIT Press.
- Saeki, E., & Saito, S. (2009). Verbal representation in task order control: An examination with transition and task cues in random task switching. *Memory & Cognition*, *37*, 1040-1050. doi: 10.3758/mc.37.7.1040.
- Schneider, D. W., & Logan, G. D. (2007). Task switching versus cue switching: Using transition cuing to disentangle sequential effects in task-switching performance. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *33*, 370-378.
- Stelzel, C., Basten, U., & Fiebach, C. J. (2011). Functional Connectivity Separates Switching Operations in the Posterior Lateral Frontal Cortex. *Journal of Cognitive Neuroscience*, *23*, 3529-3539.
- Sternberg, S. H. (1959a). A path-dependent linear model. In R. R. Bush & W. K. Estes (Eds.), *Studies in mathematical learning theory* (pp. 308-339). Stanford: Stanford University Press
- Sternberg, S. H. (1959b). Application of four models. In R. R. Bush & W. K. Estes (Eds.), *Studies in Mathematical Learning Theory* (pp. 340-381). Stanford: Stanford University Press

- Stevens, M., Lammertyn, J., Verbruggen, F., & Vandierendonck, A. (2006). Tscope: A C library for programming cognitive experiments on the MS Windows platform. *Behavior Research Methods, 38*, 280-286.
- Towse, J. N., & Cheshire, A. (2007). Random number generation and working memory. *European Journal of Cognitive Psychology, 19*, 374-394.
- Treisman, M., & Faulkner, A. (1987). Generation of random sequences by human subjects: Cognitive operations or psychophysical process? *Journal of Experimental Psychology: General, 116*, 337-355.
- Vandamme, K., Szmalec, A., Liefoghe, B., & Vandierendonck, A. (2010). Are voluntary switches corrected repetitions? *Psychophysiology, 47*, 1176-1181. doi: 10.1111/j.1469-8986.2010.01032.x.
- Vandierendonck, A. (2000a). Analyzing human random time generation behavior: A methodology and computer programs. *Behavior Research Methods, Instruments, & Computers, 32*, 555-565.
- Vandierendonck, A. (2000b). Is judgment of random time intervals biased and capacity limited? *Psychological Research, 63*, 199-209.
- Vandierendonck, A., De Vooght, G., & Van der Goten, K. (1998a). Does random time interval generation interfere with working memory executive functions? *European Journal of Cognitive Psychology, 10*, 413-442.
- Vandierendonck, A., De Vooght, G., & Van der Goten, K. (1998b). Interfering with the Central Executive by means of a Random Interval Repetition Task. *Quarterly Journal of Experimental Psychology, 51A*, 197-218.
- Vandierendonck, A., Kemps, E., Fastame, C., & Szmalec, A. (2004). Working memory components in the Corsi blocks task. *British Journal of Psychology, 95*, 57-79.
- Vandierendonck, A., Liefoghe, B., & Verbruggen, G. (2010). Task switching: Interplay of reconfiguration and interference control. *Psychological Bulletin, 136*, 601-626. doi: 10.1037/a0019791.

- Wagenaar, W. A. (1972). Generation of random sequences by human subjects: A critical review of the literature. *Psychological Bulletin*, *77*, 65-72.
- 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.
- Yeung, N. (2010). Bottom-up influences on voluntary task switching: The elusive homunculus escapes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*, 348-362. doi: 10.1037/a0017894.

Table 1. Overview of the balanced sequences of tasks A and B that are possible at lengths 3-6. The corresponding sequences of transitions (R = repetition; S = switch) are also displayed as well as the number of repetitions in the sequence.

Length 3			Length 4		
Tasks	Transitions	# Rep	Tasks	Transitions	# Rep
AAB	RS	1	AABB	RSR	2
ABB	SR	1	ABBA	SRS	1
ABA	SS	0	ABAB	SSS	0
Length 5			Length 6		
Tasks	Transitions	# Rep	Tasks	Transitions	# Rep
AAABB	RRSR	3	AAABBB	RRSRR	4
AABBB	RSRR	3	AABBBA	RSRRS	3
AABBA	RSRS	2	AABBAB	RSRSS	2
AABAB	RSSS	1	AABABB	RSSSR	2
ABBBA	SRRS	2	ABBBA	ARRSR	3
ABBAA	SRSR	2	ABBAAB	SRSRS	2
ABBAB	SRSS	1	ABBABA	SRSSS	1
ABABB	SSSR	1	ABAABB	SSRSR	2
ABAAB	SSRS	1	ABABBA	SSSRS	1
ABABA	SSSS	0	ABABAB	SSSSS	0

Table 2. Results of the analyses of variance of the proportions of runs and autocorrelations in the transition-based analysis of Study 1 in a 2 (RSI: 100 vs 1000 ms) \times 2 (Task: letter vs. form) \times 8 (Lengths/Lags 1-8) factorial design with repeated measures on the last two factors.

Effect	df	Runs		Autocorrelations	
		<i>F</i>	η_p^2	<i>F</i>	η_p^2
RSI (I)	1,16	3.5†	.18	3.5†	.18
Task (T)	1,16	13.2**	.45	25.3***	.61
Lag (L)	7,10	2644.5***	1.00	3.6*	.72
T \times L	7,10	8.6**	.86	1.0	.41
I \times T	1,16	7.3*	.31	7.3*	.31
I \times L	7,10	1.7	.55	3.3*	.70
I \times T \times L	7,10	1.3	.48	0.4	.21

Note. * $p < .05$, ** $p < .01$, *** $p < .001$; † $p = .10$

Table 3. Comparison of the chain-retrieval model in Study 1 to the statistical independence model, and the statistical dependence models. For each model, the table shows $-\ell$ (Fit), the proportion of cases in which the chain-retrieval model is better than the other models on the basis of the AICc measure, and the estimated parameter values. The table also shows the degree of correspondence (rmsd) between predictions of task and transition autocorrelations and data, as well as the proportion of the participants for which the prediction of the chain-retrieval model is better than that of the other models, and the probability the z-score based on this proportion.

	Chain-retrieval model	Bernouilli model	Perseveration model	Alternation model
Fit				
$-\ell$	42.3	249.2	51.6	244.9
Prop(AICc)		0.88	0.65	0.88
z		3.15 ($p < .001$)	1.70 ($p < .05$)	3.15 ($p < .001$)
Parameter 1	$m = 5.35$	$p = 0.48$	$q = 0.49$	$q = 0.48$
Parameter 2	$b = 0.62$		$a = 0.16$	$a = 0.05$
Parameter 3	$r = 0.27$			
Prediction task autocorrelation				
rmsd	0.020	0.041	0.035	0.044
Prop(rmsd)		0.88	0.82	0.88
z		3.15 ($p < .001$)	2.67 ($p < .01$)	3.15 ($p < .001$)
Prediction transition autocorrelation				
rmsd	0.018	0.043	0.026	0.045
Prop(rmsd)		0.82	0.71	0.82
z		2.67 ($p < .01$)	1.69 ($p < .05$)	2.67 ($p < .01$)

Table 4. Comparison of the full-parameter chain-retrieval model to variations with one of the parameter clamped to a neutral value in Study 1. For each model, the table shows $-\ell$ (Fit), the proportion of cases in which the chain-retrieval model is better than the other models on the basis of the likelihood ratio (LR), and the estimated parameter values. The table also shows degree of correspondence (rmsd) between predictions of task and transition autocorrelations and data, as well as the proportion of the participants for which the prediction of the chain-retrieval model is better than that of the other models, and the probability the z-score based on this proportion.

	Full model	$m = 3$	$b = 0.5$	$r = 0$
Fit				
$-\ell$	42.3	46.2	46.2	904.3
$P(\text{LR})$		0.47	0.59	0.82
z		-0.25 ($p = 0.60$)	0.73 ($p = 0.23$)	2.67 ($p < 0.001$)
Parameter m	5.35	3.00	5.65	5.82
Parameter b	0.62	0.79	0.50	0.79
Parameter r	0.27	0.37	0.33	0.00
Top-down repetitions				
	0.48	0.41	0.39	0.65
Bottom-up repetitions				
	0.12	0.15	0.17	0.00
Prediction task autocorrelation				
rmsd	0.020	0.044	0.028	0.050
Prop(rmsd)		0.82	0.65	0.94
z		2.67 ($p < 0.001$)	1.21 ($p = 0.12$)	3.64 ($p < 0.001$)
Prediction transition autocorrelation				
rmsd	0.018	0.035	0.027	0.036
Prop(rmsd)		0.88	0.94	0.94
z		3.15 ($p < 0.001$)	3.64 ($p < 0.001$)	3.64 ($p < 0.001$)

Table 5. Results of the analyses of variance of the proportions of runs and autocorrelations in the task-based analysis of Study 2 on the basis of a 2 (SOA: 50 vs 1500 ms) \times 2 (Task: magnitude vs parity) \times 2 (Blocks) \times 10 (Lengths/Lags 1-10) factorial design with repeated measures on the last three factors.

Effect	df	Runs		Autocorrelations	
		<i>F</i>	η_p^2	<i>F</i>	η_p^2
SOA (S)	1,78	8.2**	.09	1.0	.01
Task (T)	1,78	34.1***	.30	40.2***	.34
Blocks (B)	1,78	9.0**	.10	0.0	.00
Lag (L)	9,70	6552.3***	1.00	27.0***	.78
T \times B	1,78	0.0	.00	0.5	.07
T \times L	9,70	4.6***	.37	0.6	.07
B \times L	9,70	3.8**	.33	2.6*	.25
S \times T	1,78	3.1	.04	.7	.01
S \times B	1,78	11.1**	.12	.7	.01
S \times L	9,70	2.0†	.20	1.5	.16
S \times T \times B	1,78	0.1	.00	0.1	.00
S \times T \times L	9,70	1.7	.18	1.0	.12
S \times B \times L	9,70	2.8**	.26	1.9	.19
S \times T \times B \times L	9,70	1.1	.12	1.1	.12

Note. * $p < .05$, ** $p < .01$, *** $p < .001$; † $p = .051$

Table 6. Results of the analyses of variance of the proportions of runs and autocorrelations in the transition-based analysis of Study 2 on the basis of a 2 (SOA: 50 vs. 1500 ms) \times 2 (Tasks: magnitude vs. parity) \times 2 (Blocks) \times 10 (Lengths/Lags 1-10) factorial design with repeated measures on the last two factors.

Effect	df	Runs		Autocorrelations	
		<i>F</i>	η_p^2	<i>F</i>	η_p^2
SOA (S)	1,78	4.4*	.05	2.8	.03
Trans(T)	1,78	26.8***	.26	55.1***	.41
Blocks (B)	1,78	12.9***	.14	35.8***	.31
Lag (L)	9,70	9424.5***	1.00	20.5***	.72
T \times B	1,78	0.1	.00	0.8	.01
T \times L	9,70	37.9***	.83	9.2***	.54
B \times L	9,70	3.9***	0.33	1.3	.14
S \times T	1,78	8.5**	.10	6.1*	.07
S \times B	1,78	4.1*	.05	21.1***	.21
S \times L	9,70	2.2*	.22	1.3	.14
S \times T \times B	1,78	2.5	.03	4.4*	.05
S \times T \times L	9,70	6.3***	.45	5.4***	.41
S \times B \times L	9,70	2.7*	.26	1.8	.19
S \times T \times B \times L	9,70	2.1*	.21	2.1*	.22

Note. * $p < .05$, ** $p < .01$, *** $p < .001$

Table 7. Comparison of the chain-retrieval model in Study 2 to the statistical independence model, and the statistical dependence models. For each model, the table shows $-\ell$ (Fit), the proportion of cases in which the chain-retrieval model is better than the other models on the basis of the AICc measure, and the estimated parameter values. The table also shows the degree of correspondence (rmsd) between predictions of task and transition autocorrelations and data, as well as the proportion of the participants for which the prediction of the chain-retrieval model is better than that of the other models, and the probability the z-score based on this proportion.

	Chain-retrieval model			Bernouilli model	Perseverati on model	Alternation model
	All	Block 1	Block 2			
Fit						
$-\ell$	49.2	38.2	41.0	302.5	83.0	282.9
Prop(AICc)				0.91	0.78	0.91
z				7.38 ($p < .001$)	4.92 ($p < .001$)	7.38 ($p < .001$)
Parameter 1	$m = 4.80$	$m = 4.85$	$m = 5.01$	$p = 0.48$	$q = 0.49$	$q = 0.48$
Parameter 2	$b = 0.67$	$b = 0.67$	$b = 0.67$		$a = 0.13$	$a = 0.10$
Parameter 3	$r = 0.24$	$r = 0.21$	$r = 0.24$			
Prediction task autocorrelation						
rmsd	0.028	0.042	0.049	0.048	0.043	0.052
Prop(rmsd)				0.90	0.85	0.91
z				7.15 ($p < .001$)	6.26 ($p < .001$)	7.38 ($p < .001$)
Prediction transition autocorrelation						
rmsd	0.022	0.131	0.119	0.044	0.032	0.046
Prop(rmsd)				0.86	0.80	0.91
z				6.48 ($p < .001$)	5.37 ($p < .001$)	7.38 ($p < .001$)

Table 8. Comparison of the full-parameter chain-retrieval model to variations with one of the parameter clamped to a neutral value in Study 2. For each model, the table shows $-\ell$ (Fit), the proportion of cases in which the chain-retrieval model is better than the other models on the basis of the likelihood ratio (LR), and the estimated parameter values. The table also shows degree of correspondence (rmsd) between predictions of task and transition autocorrelations and data, as well as the proportion of the participants for which the prediction of the chain-retrieval model is better than that of the other models, and the probability the z-score based on this proportion.

	Full model	$m = 3$	$b = 0.5$	$r = 0$
Fit				
$-\ell$	27.6	28.6	28.4	121.8
$P(\text{LR})$		0.53	0.64	0.79
z		0.45 ($p = 0.67$)	2.45 ($p < 0.01$)	5.14 ($p < 0.001$)
Parameter m	4.73	3.00	4.97	5.39
Parameter b	0.59	0.69	0.50	0.69
Parameter r	0.18	0.25	0.21	0.00
Top-down repetitions				
	0.47	0.42	0.38	0.61
Bottom-up repetitions				
	0.11	0.12	0.14	0.00
Prediction task autocorrelation				
rmsd	0.028	0.047	0.036	0.050
Prop(rmsd)		0.78	0.76	0.74
z		4.92 ($p < 0.001$)	4.70 ($p < 0.001$)	4.25 ($p < 0.001$)
Prediction transition autocorrelation				
rmsd	0.022	0.033	0.032	0.035
Prop(rmsd)		0.63	0.80	0.76
z		2.24 ($p < 0.05$)	5.37 ($p < 0.001$)	4.70 ($p < 0.001$)

Table 9. Comparison of the full-parameter chain-retrieval model to variations with one of the parameter clamped to a neutral value in Study 3. For each model, the table shows $-\ell$ (Fit), the proportion of cases in which the chain-retrieval model is better than the other models on the basis of the likelihood ratio (LR), and the estimated parameter values. The table also shows degree of correspondence (rmsd) between predictions of task and transition autocorrelations and data, as well as the proportion of the participants for which the prediction of the chain-retrieval model is better than that of the other models, and the probability the z-score based on this proportion.

	Full model	$m = 3$	$b = 0.5$	$r = 0$
Fit				
$-\ell$	27.6	28.6	28.8	122.3
$P(\text{LR})$		0.21	0.21	0.61
z		-3.31 ($p = 0.99$)	-3.31 ($p = 0.99$)	1.22 ($p = 0.11$)
Parameter m	4.73	3.00	4.97	5.39
Parameter b	0.59	0.69	0.50	0.69
Parameter r	0.18	0.25	0.21	0.00
Top-down repetitions	0.43	0.38	0.38	0.56
Bottom-up repetitions	0.08	0.09	0.10	0.00
Prediction task autocorrelation				
rmsd	0.028	0.043	0.032	0.048
Prop(rmsd)		0.64	0.58	0.82
z		1.57 ($p = 0.06$)	0.87 ($p = 0.19$)	3.66 ($p < 0.001$)
Prediction transition autocorrelation				

rmsd	0.024	0.034	0.030	0.032
Prop(rmsd)		0.67	0.70	0.76
z		1.91 ($p < 0.05$)	2.26 ($p < 0.05$)	2.96 ($p < 0.01$)

Table 10. Comparison of the chain-retrieval model in Study 4 to the statistical independence model, and the statistical dependence models. For each model, the table shows $-\ell$ (Fit), the proportion of cases in which the chain-retrieval model is better than the other models on the basis of the AICc measure, and the estimated parameter values. The table also shows the degree of correspondence (rmsd) between predictions of task and transition autocorrelations and data, as well as the proportion of the participants for which the prediction of the chain-retrieval model is better than that of the other models, and the probability the z-score based on this proportion.

	Chain-retrieval model	Bernouilli model	Perseveration model	Alternation model
Fit				
$-\ell$	24.4	60.6	39.2	50.3
Prop(AICc)		0.96	0.96	0.87
z		4.38 ($p < .001$)	4.38 ($p < .001$)	3.54 ($p < .001$)
Parameter 1	$m = 4.87$	$p = 0.50$	$q = 0.50$	$q = 0.50$
Parameter 2	$b = 0.60$		$a = 0.07$	$a = 0.16$
Parameter 3	$r = 0.15$			
Prediction task autocorrelation				
rmsd	0.038	0.036	0.036	0.036
Prop(rmsd)		0.43	0.43	0.43
z		-0.63 ($p = .73$)	-0.63 ($p = .73$)	-0.63 ($p = .73$)
Prediction transition autocorrelation				
rmsd	0.039	0.035	0.035	0.035
Prop(rmsd)		0.30	0.30	0.30
z		-1.88 ($p = 0.97$)	-1.88 ($p = 0.97$)	-1.88 ($p = 0.97$)

Table 11. Comparison of the full-parameter chain-retrieval model to variations with one of the parameter clamped to a neutral value in Study 4. For each model, the table shows $-\ell$ (Fit), the proportion of cases in which the chain-retrieval model is better than the other models on the basis of the likelihood ratio (LR), and the estimated parameter values. The table also shows degree of correspondence (rmsd) between predictions of task and transition autocorrelations and data, as well as the proportion of the participants for which the prediction of the chain-retrieval model is better than that of the other models, and the probability the z-score based on this proportion.

	Full model	$m = 3$	$b = 0.5$	$r = 0$
Fit				
$-\ell$	24.4	25.6	25.9	134.2
$P(\text{LR})$		0.26	0.35	0.43
z		-2.29 ($p = 0.99$)	-1.46 ($p = 0.93$)	-0.63 ($p = 0.73$)
Parameter m	4.87	3.00	5.30	5.43
Parameter b	0.60	0.75	0.50	0.69
Parameter r	0.15	0.23	0.19	0.00
Top-down repetitions				
	0.45	0.41	0.38	0.56
Bottom-up repetitions				
	0.07	0.08	0.10	0.00
Prediction task autocorrelation				
rmsd	0.038	0.045	0.032	0.050
Prop(rmsd)		0.61	0.30	0.74

<i>z</i>		1.04 (<i>p</i> = 0.15)	-1.88 (<i>p</i> = 0.97)	2.29 (<i>p</i> < 0.01)
----------	--	----------------------------	-----------------------------	----------------------------

Prediction transition autocorrelation

rmsd	0.039	0.037	0.033	0.042
------	-------	-------	-------	-------

Prop(rmsd)		0.48	0.35	0.78
------------	--	------	------	------

<i>z</i>		-0.21 (<i>p</i> = .58)	-1.46 (<i>p</i> = 0.93)	2.71 (<i>p</i> < 0.01)
----------	--	----------------------------	-----------------------------	----------------------------

Figure Captions

Figure 1. Task-based analysis (left panels) and transition-based analysis (right panels) of task choices in Study 1. Top-left panel: Proportion of runs length 1-10 as a function of RSI (100 vs. 1000 ms) and task (letter vs. form). The runs proportions for the short RSI in both tasks show more frequent and on average longer runs of the task. Bottom-left panel: Proportions of autocorrelations lag 1-10 as a function of RSI and task. At lag 1 the proportion of correlations is quite high; it is lower (alternation tendency) at lags 2-4 especially at long RSI and gradually increases again from lag 5 on. Top-right panel: Proportion of runs length 1-10 as a function of RSI (10 vs. 1000 ms) and transition (repetition vs. switch). More frequent and longer runs of repetitions occur in the short RSI condition. Bottom-right panel: Proportions of autocorrelations lag 1-10 as a function of RSI and transition. At lag 1, the proportion of correlations is lower than at the later lags, where it is quite stable. Correlations are higher for repetitions especially at short RSI.

Figure 2. Observed and predicted proportions of autocorrelations lag 1-10 in the model estimations in Study 1. Top-left panel: The observed task choice autocorrelations, the predictions of the chain-retrieval, the Bernoulli, the perseveration and the alternation models are shown. The predictions of the chain-retrieval model seem to correspond best with the data. Bottom-left panel: Observed autocorrelations and predictions by the same models for the transition data. The figure shows lags 1-10, although for the predictions only 9 lags were used. The predictions of the chain-retrieval model seem to correspond best with the data. Top-right panel: The predictions of the task choice autocorrelations by the full chain-retrieval model (3 parameters) and the three versions with the value of one parameter clamped to a neutral value. The predictions of the complete chain-retrieval model correspond best with the data. Bottom-right panel: Same for the transition autocorrelations lag 1-10. The figure shows lags 1-10, although for the predictions only 9 lags were used. Except for the model with r clamped to 0, the predictions of all models seem to correspond well with the data.

Figure 3. Task-based analysis (left panels) and transition-based analysis (right panels) of task-indication responses in Study 2. Top-left panel: Proportion of runs length 1-10 as a function of SOA (5 vs. 1500 ms) and task (magnitude vs. parity). The runs proportions for the short SOA in both tasks show more frequent and on average longer runs of the task. Bottom-left panel: Proportions of autocorrelations lag 1-10 as a function of SOA and task. At lag 1 the proportion of correlations is quite high; it is lower (alternation tendency) at lags 2-4 especially at long SOA and gradually increases again from lag 5 on. Top-right panel: Proportion of runs length 1-10 as a function of SOA and transition (repetition vs. switch). More frequent and longer runs of repetitions occur in the short SOA condition. Bottom-right panel: Proportions of autocorrelations lag 1-10 as a function of SOA and transition. At lag 1, the proportion of correlations is lower than at the later lags, where it is quite stable. Correlations are higher for repetitions especially at short SOA.

Figure 4. Observed and predicted proportions of autocorrelations lag 1-10 in the model estimations in Study 2. Top-left panel: Besides the observed task autocorrelations, the predictions of the chain-retrieval, the Bernoulli, the perseveration and the alternation models are shown. The predictions of the chain-retrieval model seem to correspond best with the data. Bottom-left panel: Observed transition autocorrelations and predictions by the same models. The figure shows lags 1-10, although for the predictions only 9 lags were used. The predictions of the chain-retrieval model seem to correspond best with the data. Top-right panel: The predictions of the task autocorrelations by the full chain-retrieval model (3 parameters) and of the three versions with the value of one parameter clamped to a neutral value. The predictions of the complete chain-retrieval model correspond best with the data. Bottom-right panel: Same for the transition autocorrelations lag 1-10. The figure shows lags 1-10, although for the predictions only 9 lags were used. Except for the model with parameter r clamped, the predictions of all the models seem to correspond well with the data.

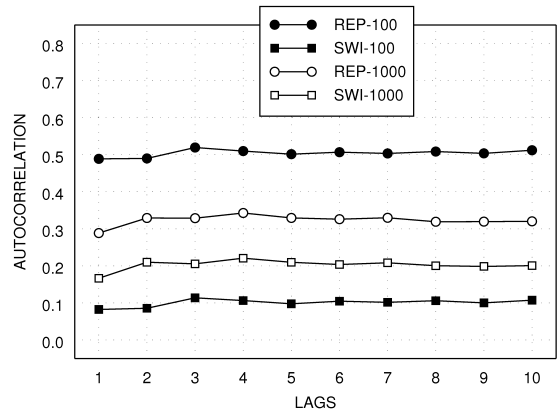
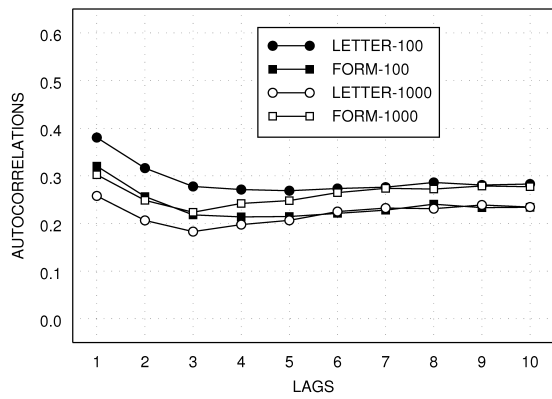
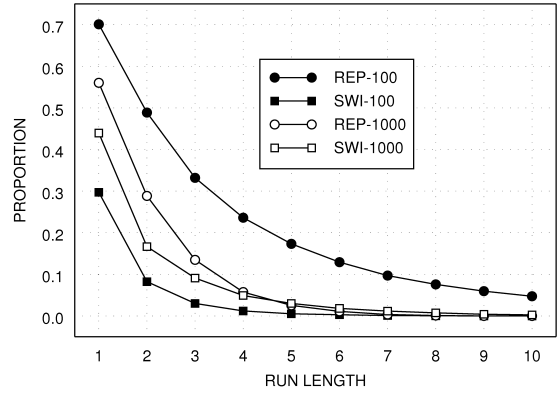
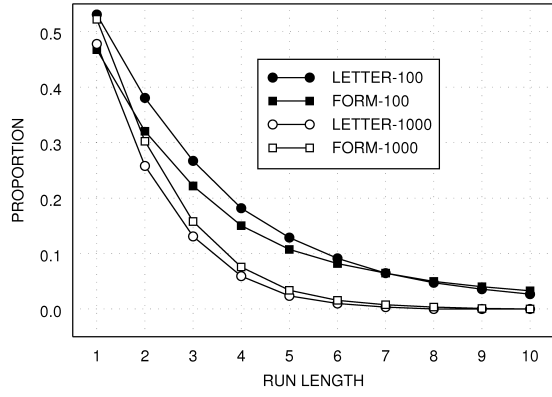


Figure 1

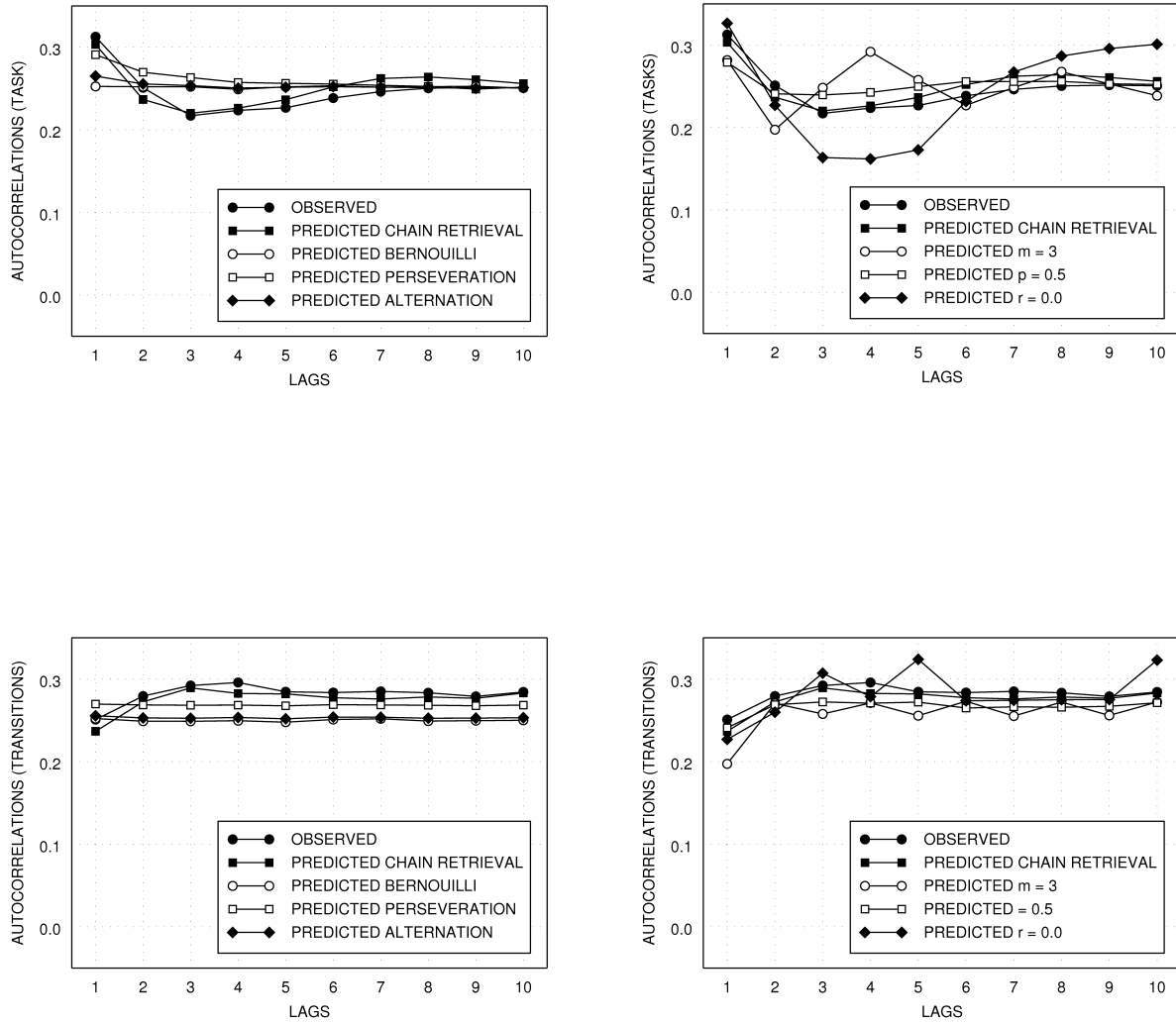


Figure 2

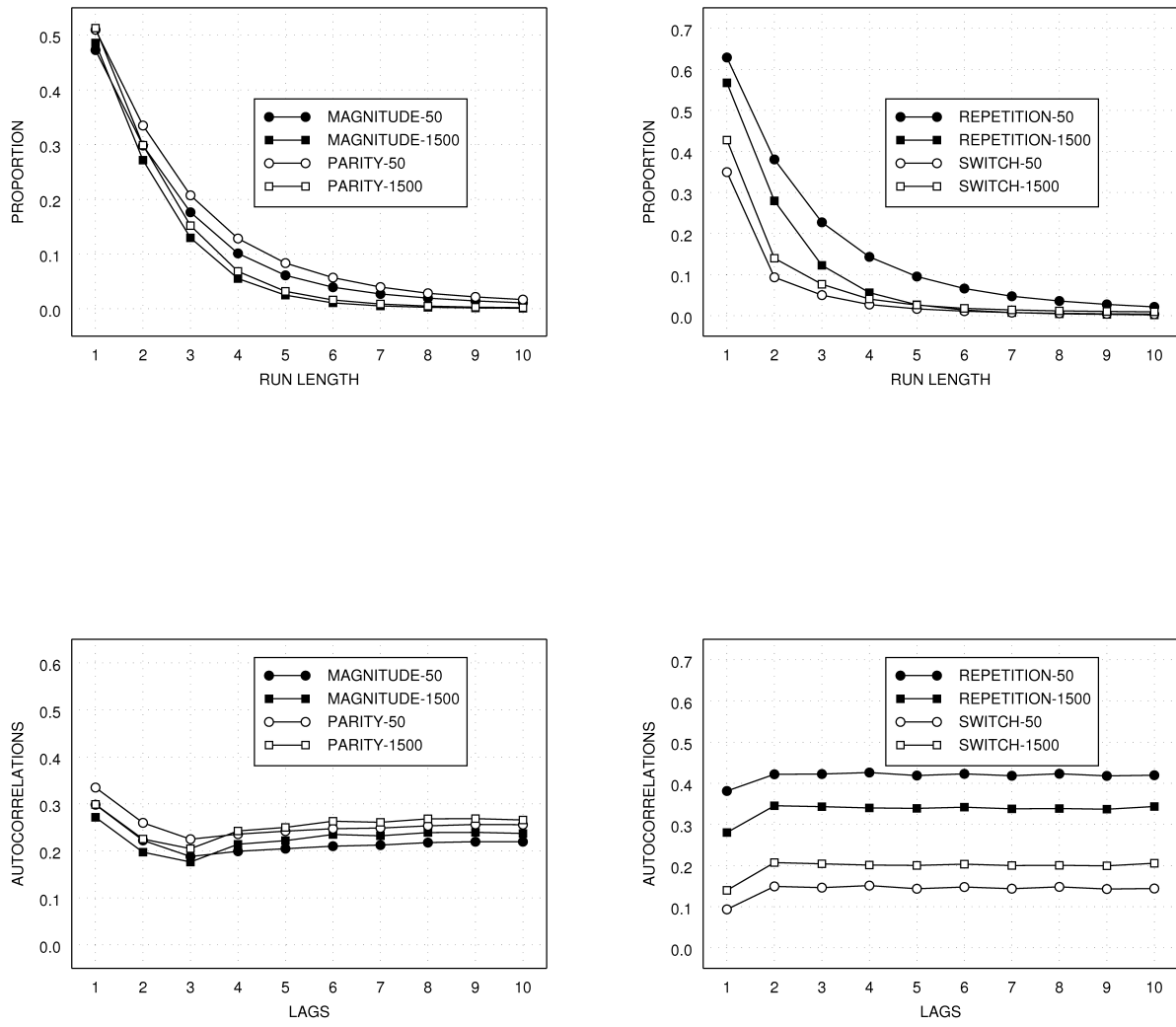


Figure 3

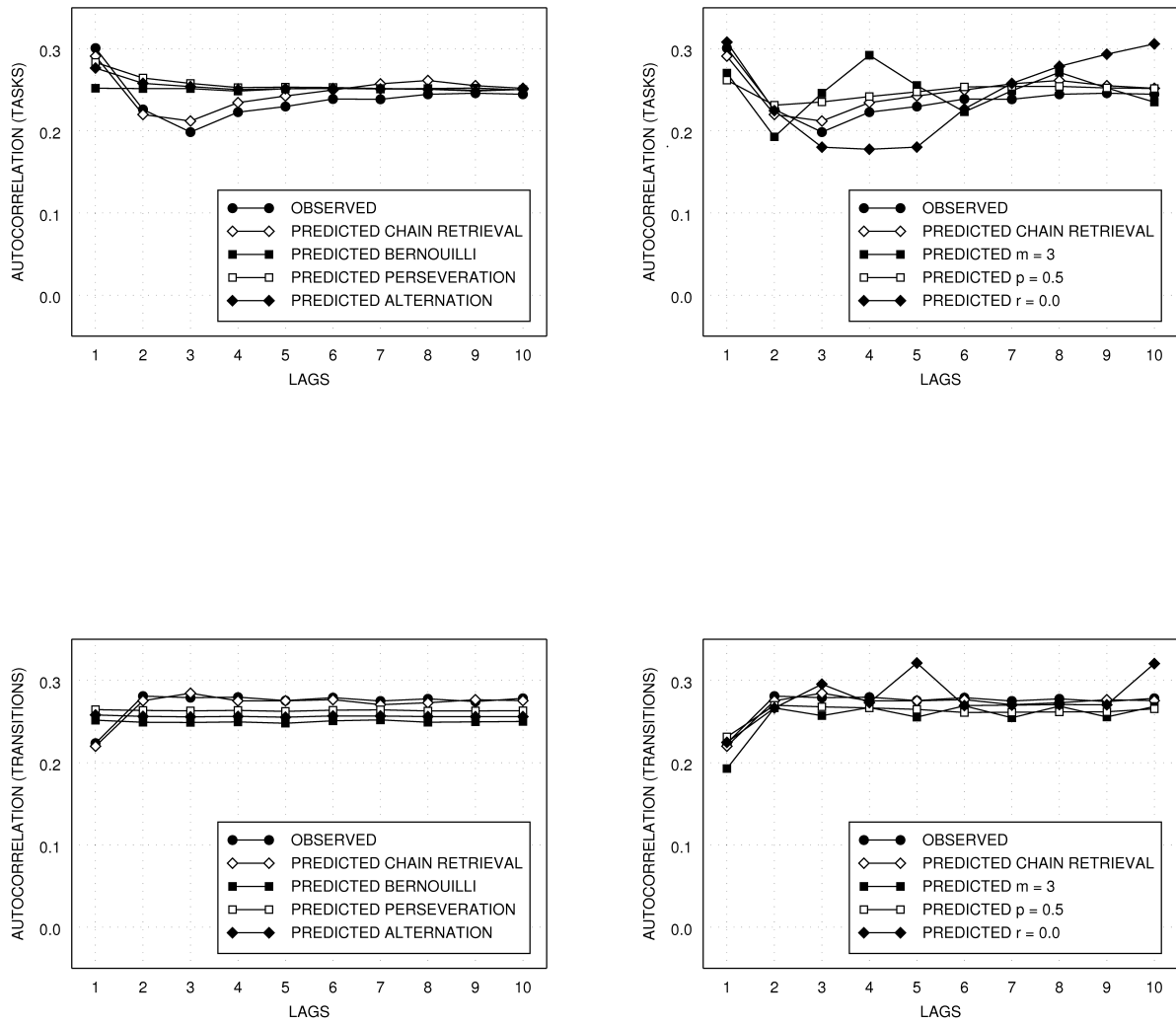


Figure 4

Appendix A: Results task execution Study 2

Task execution RTs of correct trials and error rates were calculated. Task-execution RTs and error rates were analyzed by means of a 2 (SOA: 50ms vs. 1500ms) by 2 (task transition: task repetition vs. task switch) mixed ANOVA with repeated measures on the last factor. The main effect of SOA was significant, $F(1,78) = 13.49$, $p < .01$, $\eta_p^2 = .15$, indicating slower responses in the SOA-50ms ($M = 706\text{ms}$, $SE = 14$) than in the SOA-1500ms ($M = 632\text{ms}$, $SE = 14$) condition. The RTs confirmed the significant main effect of task transition, $F(1,78) = 135.01$, $p < .01$, $\eta_p^2 = .63$. Task switches ($M = 715\text{ms}$, $SE = 13$) were slower than task repetitions ($M = 621\text{ms}$, $SE = 9$). The interaction between task transition and SOA was not significant, $F(1,78) = 1.97$, $p = .16$, $\eta_p^2 = .02$.

For the error rates, the analysis showed that the main effect of SOA was not significant, $F(1,78) = 2.16$, $p < .15$, $\eta_p^2 = .03$. The main effect of task transition was significant, $F(1,78) = 13.49$, $p < .01$, $\eta_p^2 = .15$, indicating that more errors were made on task switches ($M = .10$, $SE = .007$) than on task repetitions ($M = .07$, $SE = .005$). The interaction between SOA and task transition was not reliable, $F(1,78) = 2.31$, $p = .13$, $\eta_p^2 = .03$.

In sum, with this analysis we confirmed the observation of Arrington and Logan (2004, 2005) that also with a double registration procedure for voluntary task switching, a switch cost is present on task execution performance.

Appendix B: Additional tests of chain-retrieval model

Using the same formal basis defined by the three parameters of the chain-retrieval model, other assumptions can be made to implement model variations. Basically, the critical issues are the way chains are combined to form a sequence and what happens when a bottom-up intrusion occurs. The chain-retrieval model assumes (a) that the last element of the current chain is overlaid with the first element of the next chain and (b) that a bottom-up repetition overrules the intended action. The number of possible variations and combinations of variations that can be considered are numerous. In order to explore the usefulness of the two structural assumptions made, we consider alternative models in which each time one of the assumptions is replaced with another one.

Instead of the first assumption, namely that the chains are overlaid, it is possible to assume that chains are simply concatenated. Two possible ways of concatenation are considered, namely random concatenation and primed concatenation. In random concatenation the chains are connected together without any restrictions. In primed concatenation, the last element of the current chain primes the start of the next chain, so that when a chain ends with one task, the next starts with the same task (e.g., ABA followed by AAB). This mechanism produces a not intended repetition at every junction of two consecutive chains.

The second assumption regarding the effect of a bottom-up event can also be implemented in several ways. In the model the event replaces the intended action. An alternative is that the event intrudes into the intended chain, so that the presently intended element is postponed to the next position in the sequence. Another alternative is that after an intrusion, the intended chain is lost from working memory and a new chain has to be retrieved.

The results of the tests of these model variations are presented in Table B1. The tests are based on the data set of Study 2. This table is constructed in the same way as the tables in the main text. It compares the goodness-of-fit and the correspondence of the predictions of the task-chain retrieval model with other possible implementations of the model. The table shows that the model yields a better fit to the observed runs proportions than the overlay-include, the overlay-

restart and the concatenate-random variants and that the fit is equivalent for the concatenate-primed variant. The latter variant tends to yield better predictions of the task autocorrelations, but is very bad in prediction the transition autocorrelation data. It seems, then, that the structural assumptions made are probably the best one possible.

Table B1. Comparison of model variants of the chain-retrieval model. The table compares the basic version of the model to variations of the model in which one assumption was changed. For each model variant the following results are shown: $-\ell$ (Fit), the proportion of cases in which the chain-retrieval model is better than the other models on the basis of the AICc measure, and the estimated parameter values. The table also shows the degree of correspondence (rmsd) between predictions of task and transition autocorrelations and data, as well as the proportion of the participants for which the prediction of the chain-retrieval model is better than that of the other models, and the probability the z-score based on this proportion.

Junction	Overlay			Concatenation	
	overrule	intrude	restart	random	primed
Fit					
$-\ell$	49.2	49.9	52.2	56.4	48.9
Prop(AICc)		0.61	0.91	0.89	0.46
z		2.01 ($p < .05$)	7.38 ($p < .001$)	6.93 ($p < .001$)	-0.67 ($p = .75$)
m	4.80	4.15	4.51	5.70	4.55
b	0.67	0.64	0.65	0.62	0.50
r	0.24	0.27	0.21	0.25	0.17
Prediction task autocorrelation					
rmsd	0.028	0.027	0.029	0.031	0.027
Prop(rmsd)		0.45	0.51	0.53	0.40
z		-0.89 ($p = .81$)	0.22 ($p = .41$)	0.45 ($p = .33$)	-1.79 ($p = .96$)
Prediction transition autocorrelation					
rmsd	0.022	0.020	0.022	0.024	0.041
Prop(rmsd)		0.39	0.51	0.54	0.86
z		-2.01 ($p = .98$)	0.22 ($p = .41$)	0.67 ($p = .25$)	6.48 ($p < .001$)

Appendix C: Task execution results in Study 3

Error RTs and RT on trials following an error were excluded from data analysis, leading to 6.7% of data loss. These RTs were averaged per condition and subjected to a 2 (Task: magnitude or parity) \times 2 (Hand: left or right) \times 2 (Transition: repetition or switch) ANOVA with repeated measures on the last two factors. Hand switches were slower ($M = 813$ ms) than hand repetitions ($M = 692$ ms), $F(1,31) = 17.75$, $\eta_p^2 = 0.36$, and this was the only significant main effect. Task interacted with Hand, $F(1,31) = 10.44$, $\eta_p^2 = 0.25$ so that in the magnitude task, left responses were slower ($M = 737$ ms) than right responses ($M = 680$ ms), $F(1,31) = 10.85$, $\eta_p^2 = 0.26$, whereas the difference was not significant in the parity task, $F(1,31) = 1.55$, $p = 0.22$, $\eta_p^2 = 0.05$. The triple interaction of Task by Hand by Transition was also significant, $F(1,31) = 5.94$, $\eta_p^2 = 0.16$, indicating that the hand switch cost size depended on both task and hand.

More errors were committed on hand switches ($M = 0.09$) than on hand repetitions ($M = 0.05$), $F(1,31) = 15.28$, $\eta_p^2 = 0.33$. This effect interacted with Task, $F(1,31) = 5.52$, $\eta_p^2 = 0.15$. On the magnitude task, the switch cost was not reliable ($M = 0.04$ for hand repetitions and $M = 0.06$ for hand switches), $F(1,31) = 1.18$, $p = 0.29$, $\eta_p^2 = 0.04$, but on the parity task the contrast was large ($M = 0.05$ for hand repetitions and $M = 0.12$ for hand switches), $F(1,31) = 20.20$, $\eta_p^2 = 0.39$. Apart from these effects, no other significant effects were observed, except that the interaction of Task and Hand was close to significance, $F(1,31) = 3.90$, $p = 0.06$, $\eta_p^2 = 0.11$. On the magnitude task, less errors were committed on the right hand ($M = 0.04$) than on the left hand ($M = 0.07$), $F(1,31) = 3.48$, $p = 0.07$, $\eta_p^2 = 0.10$, while no difference was observed on the parity task (both $M = 0.09$).

The observation of a hand switch cost in RT and error data is in line with the findings reported by Stelzel et al. (2011). Interestingly, in the present case, the size of the switch cost seems to depend on the main task performed and on the hand selected to execute the main task.