Journal of Experimental Psychology:
Human Perception and Performance
2010, Vol. 36, No. 3, 649–661

0096-1523/10/$12.00 DOI: 10.1037/a0017178

Task-Relevant Chunking in Sequence Learning

Amotz Perlman, Emmanuel M. Pothos,
and Darren J. Edwards
Swansea University

Joseph Tzelgov
Achva Academic College and Ben-Gurion University
of the Negev

In the present study, we investigated possible influences on the unitization of responses. In Experiments 1, 2, 3, and 6, we found that when the same small fragment (i.e., a few consecutive responses in a sequence) was presented as part of two larger sequences, participants responded to it faster when it was part of the sequence that was presented more often. This indicates that chunking can be driven by task-relevant considerations, as opposed to co-occurrence. The results are discussed in the context of chunking theories and the relevant motor learning literature.

Keywords: automaticity, sequence learning, unitization of responses, chunking

In a typical sequence learning study, participants have to respond to a sequence of stimuli. In every trial, a stimulus is presented at one of a number of locations on a computer screen and participants are asked to respond by pressing a key corresponding to the position of the stimulus. Unknown to the participants, the series of stimuli follow a repeating sequence. Reaction times (RTs) tend to decline with time throughout training, but rise noticeably if the replicated pattern is distorted. Sequence learning is at the heart of many fundamental cognitive activities, such as writing, typing, language production, and complex motor skills (riding a bicycle or driving). In this paper we examine the kind of knowledge acquired in a sequence learning task. As we will shortly see, an influential hypothesis is that co-occurring elements are organized into single units/chunks (we consider the terms *unit*, *chunk*, and *fragment*, to be equivalent). An alternative hypothesis is that organization of individual elements is driven by task demands, for example the requirement to produce a set of responses together. In what follows, our purpose is to motivate these two hypotheses from the existing literature and provide a contrast within a *principled experimental paradigm*.

Unitization/chunking on the basis of co-occurrence statistics has been a core mechanism in many influential cognitive models. For example, standard backpropagation (back prop) networks involve three layers, an input layer, an output layer, and a hidden layer whose function is to transform the input in a way that the learning problem can be more easily solved. Back prop networks have been one of the dominant computational paradigms for modeling cognitive processes, because of their ability to model complex cognitive phenomena (e.g., default rules in inflectional morphology; Hare, Elman, & Daugherty, 1995; Plunkett & Marchman, 1993) and also because of their broad biological plausibility (see Rumelhart & McClelland, 1986; for a collection of early arguments). A simple recurrent network (SRN) is a modification of the standard back prop network, such that there is an additional set of units which retain a copy of the hidden unit activation at each time step (Elman, 1990). In the next time step in the operation of an SRN, the activation at the input layer is a function both of the actual input and the hidden unit activation at the previous time step. This architecture makes SRNs particularly suitable for learning temporal/sequential structure, and indeed SRNs have been widely employed in modeling linguistic processing (e.g., Christiansen & Chater, 1999, 2001; Elman, 1991) and sequential learning tasks, such as the serial reaction time task (e.g., Boucher & Dienes, 2003; Buchner, Steffens, & Rothkegel, 1998; Cleeremans & McClelland, 1991).

A key characteristic of SRNs is that their learning involves progressive sensitivity to an increasing number of elements of the preceding sequence. In other words, early in learning, an SRN may learn that after one particular element, another element is likely to follow. After extensive learning the network will be able to take into account the previous several elements in predicting the next element. This type of learning can be interpreted as showing that, with learning, an SRN becomes sensitive to sequence chunks (early in learning, pairs or triplets of sequence elements). Put simply, the creation of such chunks is driven by co-occurrence statistics (of course, there are many different kinds of co-occurrence statistics, but this issue is not presently relevant; see Perruchet & Peereman, 2004).

The chunking hypothesis refers to the intuition that the knowledge acquired through experience often corresponds to chunks. The chunking hypothesis is an extremely influential set of related intuitions, which have been employed in a variety of empirical contexts. For example, Miller (1956) formed his hypothesis regarding short-term memory not in terms of individual elements, but rather chunks and Simon and Barenfeld (1969) were able to explain the superiority of expert chess players relative to novice ones, in terms of the availability of more complex ‘chunks’ for remembering and representing chess-piece arrangements. Consider briefly, for example, the formulation of the chunking hypothesis

---

Amotz Perlman, Emmanuel M. Pothos, and Darren J. Edwards, Swansea University; and Joseph Tzelgov, Achva Academic College and Ben-Gurion University of the Negev.

This research was partly supported by ESRC grant R000222655 to E.M.P. We thank Abby Petrie and Simon Hughes for collecting the data for Experiment 5. We also thank Bernhard Hommel and Pierre Perruchet for their many useful comments on previous versions of this paper.

Address correspondence to Amotz Perlman, Department of Psychology, Ben-Gurion University of the Negev, PO Box 653, Beer Sheva 84105, Israel. E-mail: amotz@bgu.ac.il

by Servan-Schreiber and Anderson (1990; Servan-Schreiber, 1991). The familiarity of a new stimulus is assumed to be proportional to $e^{1-nchunks}$, where *nchunks* is the number of chunks activated when processing the stimulus (Servan-Schreiber, 1991). Accordingly, the fewer the chunks employed the greater the stimulus familiarity. The probability that a chunk will be activated when processing a new stimulus will depend by how frequently and recently it has been activated. This model effectively predicts that frequently occurring chunks will have a greater influence in the processing of new information. Servan-Schreiber and Anderson's (1990) model is well-suited for sequential learning tasks.

The same ideas can be easily adapted in cases in which there are distinct training and test phases. For example, in artificial grammar learning, Knowlton and Squire (1996, 1994; see also Perruchet & Pacteau, 1990; Pothos, 2007) suggested that a test item would be more likely to be endorsed (as being in the same category as the training items) if it contains chunks which have been more frequently encountered in the training phase. Specifically, the associative chunk strength of each chunk (only bigrams and trigrams were considered) was defined as the number of times it appeared in the training items. The global associative chunk strength of a test item is simply computed by averaging the associative strengths of all chunks in the item. The widely supported behavioral prediction is that higher global associative chunk strength leads to a higher likelihood of endorsement (see also Boucher & Dienes, 2003; Johnstone & Shanks, 1999; Pothos & Bailey, 2000).

So, it is clearly the case that chunks can be created as a function of co-occurrence. The literatures regarding SRNs and chunking are only two examples in which this idea has been widely advocated and supported. However, there is an alternative perspective. In the motor literature, co-occurring elements (typically responses) can be organized into a single chunk (or unit) depending on the relevant task demands. For example, the act of vocalizing or writing a word, such as 'a-p-p-l-e', involves a sequence of particular motor actions, which (appear) unitized into a single unit (but see Rosenbaum, 1987). Co-occurrence information probably plays some role but, in this case, it appears that the relevant task requirement is the driving force in unitization. In the example of 'apple' the fragment 'p-l-e' occurs in the context of many other words, such as 'please,' yet the responses required to vocalize or write 'apple' would still appear to be organized in a single unit. Regardless of the validity of this argument, our point is that, *conceptually*, unitization on the basis of co-occurrence may be different from unitization based on task demands, such as the requirement to perform a particular sequence of actions. However, it is difficult to make this intuition (and the corresponding contrast) more concrete in the case of a behavior such as language, since the relevant responses would be highly overlearned and also because we would have poor control of the amount of experience of different individuals with particular sequences of responses. Our objective in this paper is to provide a concrete, experiment-based examination of which factor is more important in unitization: co-occurrence or task demands. We do it in the context of learning simple motor sequences.

We next consider in more detail the relevant motor learning literature, so as to both better motivate the hypothesis of unitization as function of task demands and also evaluate the novelty of the proposed contrast. Rosenbaum and colleagues (e.g., Rosenbaum, Hindorff, & Munro, 1987; Rosenbaum, Kenny, & Derr, 1983) have been advocating a model for the control of rapid motor movements, according to which individual actions are organized into a hierarchy of behavioral units—behavioral units further up the hierarchy control more individual responses (cf. Dirlam, 1972, for a computational examination of hierarchies). Examples of related ideas are Rumelhart and Norman's (1982) neural network model, in which the decision to type a word is associated with a unit corresponding to that word, which in turn activates each of the letters in the word. Likewise, van Galen (1991; see also Berg, 2002; Portier and van Galen, 1992) proposed a model of handwriting such that handwriting was assumed to be the result of several, hierarchically organized, processing modules, each of which corresponds to a different aspect of the handwriting process.

By contrast, Palmer and Pfordresher (2003) suggested that in the case of music performance, the ability of humans to produce long sequences of music events cannot be explained by hierarchical models of sequence production, since it is unlikely that the cognitive system can concurrently represent all information in a long sequence. Therefore, they argued that an incremental model for sequence production is required, whose main characteristic is that planning a sequence occurs incrementally, so that planning and producing co-occur in time. Incremental planning has been carefully considered in the case of speech production as well (Dell, Burger, & Svec, 1997), but other researchers have advocated hierarchical models of speech production (Sternberg, Knoll, Monsell, & Wright, 1988). In speech production a major problem is how the cognitive system can keep track of the serial order of phonemes, in which it is typically postulated that a time-dependent signal is used to ensure that the correct phonemes are produced at the correct serial order (e.g., Burgess & Hitch, 1997; Houghton, 1990; Vousden, Brown, & Harley, 2003). Resolving this problem appears to conflict with a simpler class of speech production models, based on the chaining of individual responses.

This is an extremely brief overview of an extensive and complex literature. It is meant to help derive two conclusions. First, speech production, music production, handwriting, and related behaviors share some key characteristics, such that the fact that they concern the ability to produce novel sequences of elements (e.g., a novel sequence of words in an uttered sentence) and that they involve material which is already highly learned. By contrast, in sequence learning tasks the sequences of responses are fixed and of interest is the development of unitized knowledge, less so how it is processed after extensive training. These differences between sequence learning tasks in the laboratory and speech/music production, or handwriting, suggest caution in drawing analogies between the two literatures. Second, theoretical approaches such as hierarchical models or anticipatory ones all assume that responses can be organized into chunks or units. Regarding the present research, the key question is how such units arise. In this respect, the literature on speech/ music production, handwriting, etc. is less informative; it tends not to distinguish between chunking which is driven by co-occurrence information and by task demands (such as the requirement to vocalize a particular word). We will return to this issue again once we have described our proposed manipulation.

The work of Hommel and colleagues is a notable exception. Hommel considered a stimulus-based control mode, according to which the cognitive system intentionally delegates control to external stimuli (Hommel, 2000). A stimulus-based control mode is related to what we termed learning on the basis of co-occurrence statistics, since in both cases learning/responding is guided by the stimuli. Hommel (2003) proposed an alternative control mode, which he labeled plan-based control. Plan-based control corresponds to responses provided as a result of a particular intention (e.g., the intention to write or vocalize a word) or, in our terms, tasks demands. Tubau, Hommel, and Lopez-Moliner (2007) examined in a series of experiments the conditions under which responding would reflect a stimulus-based or a plan-based control mode. For example, they found that explicit instructions enhanced and auditory noise reduced the tendency to adopt a plan-based control mode.

In Tubau et al.'s work there was only one sequence. Therefore, the frequency of any chunk would be the same as the frequency of the entire sequence (assuming positional-dependent encoding; but such an assumption makes sense, especially in the case of sequences of many repeating elements). In such a case, there would be an expectation that, to the extent that the sequence is learned, the entire sequence would be eventually unitized, instead of any particular part of the sequence. This expectation is consistent with co-occurrence models of chunking, and is supported by Tubau et al.'s findings, since there was an overall tendency for plan-based responding for explicit learners. So, Tubau et al.'s work cannot help regarding the present research question, that is, whether co-occurrence statistics will be a greater influence in the acquired knowledge, compared to the relevant task demands, when the two are in contrast—Tubau et al.'s objective was to identify the methodological manipulations which would enhance/ suppress a plan-based control mode.

The general design of the present experiments involved two sequence orders, which can be represented as ACBD and ACDB. The fragment AC appeared in both orders. Such a design guarantees that whatever the frequency of either sequence, the frequency of the shared fragment would still be higher. Accordingly, co-occurrence models of chunking are forced to predict that unitization of the shared fragment would be stronger than for the entire sequence, so that the common fragment AC is organized as a single unit, but not the entire sequences of four responses. In such a case, sequence ACBD will be performed as the separate chunks AC, B and D and ACDB will be performed as the separate chunks AC, D and B. Importantly, in this situation the fragment AC will be identified as AC each time it is presented *regardless* of whether it is presented as part of the sequence ACBD or ACDB. Conversely, suppose that unitization is driven by task demands. In such a case, the two sequences will be learned and performed as single units. For example, ACBD will be performed as a single unit ACBD and not as the separate chunks AC, B and D.

How can we empirically determine whether chunking occurs for AC or for the entire sequences? In our experimental investigation, the sequence ACBD was presented 80% of the time and the sequence ACDB was presented 20% of the time. If AC forms a chunk independent of task demands, responding to C should be the same whether it appears in the more (ACBD) or less (ACDB) frequent sequence. But if unitization occurs as a result of task demands, so that the responses for ACBD are organized in one unit and the responses for ACDB in another, we expect that C will be responded to faster in the case of the more frequent sequence. To anticipate our conclusion, most of our results indicate unitization according to task demands, rather than co-occurrence.

At this point we can ask again whether theoretical work in motor learning can address the above empirical question. We consider the hierarchical editor model of Rosenbaum and colleagues, since this has been a carefully developed model examined under a wide range of empirical conditions (for an overview see Rosenbaum, 1987). Hierarchical organization of information is one aspect of the model, as is the fact that preparing responses for two different sequences involves creating an abstract program which includes all the common features of the sequences. In other words, in the case of sequences ACAD and ACAB, the prediction would be that participants prepare ACA and decide online only with respect to the final response, which involves the choice between D and B. Accordingly, the hierarchical editor model does not provide any reason to expect that RT might differ for the responses towards the common fragment in the two sequences. This prediction is the same as that from co-occurrence models of unitization; as mentioned, it is not supported in our results. The frequency manipulation we describe above is, we believe, the most novel aspects of the present work and one which can provide a test of unitization on the basis of co-occurrence vs. task demands.

There is a separate, less major, issue we have to address. So far we have assumed that a well-practiced sequence of responses will be organized in a single unit (our particular question being whether this unit will reflect co-occurrence or task demands). It is desirable (though not essential) that we demonstrate that unitization occurs in our experiments. This is because, for example, we can then interpret our main question in terms of what kinds of *units* are formed. A compelling empirical marker of unitization relates to a so-called preparation effect, according to which the time required to initiate list recall and the time to recall each item in a list increases with the length of the list (e.g., Luria & Meiran, 2003; Schneider & Logan, 2006; Sternberg, Monsell, Knoll, & Wright, 1978; for an early related demonstration see Henry & Rogers, 1960). To quote from Sternberg et al.'s (1988, p. 175) review, "a 'motor' program for the whole utterance, prepared in advance, controls the execution of each of its 'units'." A preparation effect has been observed in speech (e.g., Dell et al., 1997; Sternberg et al., 1988) and other motor behavior (for an overview see Rosenbaum, 1987). Perlman, Tzelgov, Pothos and Millar (under review) have demonstrated a preparation effect with learning of sequences just like the ones employed in the current study—the main difference between these two investigations is that the Perlman et al. (under review) one does not include the critical sequence overlap manipulation. Specifically, Perlman et al. (under view) observed that, as training advanced, participants delayed their response to the first stimulus, while gradually becoming faster to respond to the second, third, and fourth stimuli. Finally, an additional important empirical marker of unitization we will consider relates to how the RT for each response varies with serial position of the response. It can be argued that when all responses are produced as a single unit there should be no difference in the RT for all responses after the first one (e.g., see Perlman et al., under review).

Experiment 1

Experiment 1 is the simplest test of our hypothesis. Participants went through extensive training for two sequences of four responses. Before the start of a sequence, a cue indicated which sequence was going to be presented. The relative frequency of the two sequences was manipulated and the two sequences had a common fragment (the two sequences can be represented as ACBD and ACDB). Following the last response for a sequence, there was a response cue interval (RCI) of 1,000 ms before another cue appeared to indicate the next sequence.

Method

Participants. Twenty one students from introductory psychology courses at Swansea University participated in the experiment for course credit. All participants reported normal or corrected-to-normal vision.

Apparatus. The experiment was conducted using IBM compatible Pentium III computers with 17″ monitors. The screen was placed approximately 60 cm from the participants. The participants responded by using a response box (from Cedrus corp., which allows ms accuracy in RT measurement). The onset of a stimulus started the timer; the stimulus disappeared as soon as the participant responded. The responses were indicated by pressing the 1 (left most), 2 (second from left), 3 (second from right), and 4 (right) keys on the response box. Participants were asked to use the index and the middle fingers of both hands for key presses.

Stimuli. The stimuli were small Xs, half a centimeter wide and half a centimeter tall. The Xs could appear in each of four horizontally arranged locations: A rectangle, 12 centimeters wide and three centimeters tall was presented in the middle of the screen. This rectangle was divided into four squares, three centimeters wide and three centimeters tall each. In every trial, an X could appear in one of four squares.

The experiment was organized in 10 blocks of 160 trials each. In these trials, the locations of the Xs appeared in either of two fixed, repeating orders. With the four locations represented by A (leftmost square), B, C, and D (rightmost square) respectively, the two orders we employed can be denoted as ACBD and ACDB. The ACBD sequence was always presented with a red rectangle as a cue and the ACDB sequence was always presented with a blue rectangle as a cue. In other words, the rectangle, divided into four squares, was used as a cue. It appeared with the Xs inside, while participants responded to the Xs. The rectangle was displayed until the last stimulus of a sequence was responded to (at which point, it disappeared for a RCI of 1,000 ms and re-appeared again in red or blue).

Procedure. Each participant took part in ten blocks of training of 160 trials each. In each trial, participants had to respond, as fast as possible, to the location of an appearing X, by pressing the corresponding key. The participants did not receive any information regarding the sequence contingency and they were not told that each cue indicates a different sequence. Each block started with the appearance of the message: "press any key to continue." After the participant responded, the screen went black for 1,000 ms. Subsequently, a blue or red rectangle appeared (the cue) that stayed on until the last stimulus of the sequence was responded to. A participant response triggered the immediate appearance of the next stimulus in the sequence. The response of the last stimulus in a sequence was followed by a RCI of 1.000 ms, when the screen was black; the next cue followed. Each block of training included 32 presentations of the sequence ACBD and 8 presentations of the sequence ACDB. The 40 sequences (of four trials each) were arranged in a random order.

Results and Discussion

Both RT and error data for the experimental trials were recorded. Comparable analyses were run on both of these measures, yielding similar results, except that some effects that were significant in the RT data were not always so in the error data. There was no evidence of a speed-accuracy trade-off, thus, only the RT data are presented in detail. Average error rates were 2.5% for the sequence ACBD and 2.7% for the sequence ACDB.

Only the RTs from correct responses were included in the analyses. For each participant, the median RT was calculated separately for each block. The mean of the median RTs is presented in Figure 1 as a function of block of trials, for each condition.

The mean RTs of each block of responses were submitted to a three-way within subject analysis of variance (ANOVA) with, Sequence (ACBD and ACDB) Block and Stimulus (Stimulus 1, Stimulus 2, Stimulus 3, and Stimulus 4), as the manipulated factors. In all statistical analyses, the significance level was set to .05. The Block effect was significant, [$F(9, 180) = 35.677$, $MSE = 8897.201$, $\acute{\eta}_p^2 = 0.640$], as was the Stimulus effect [$F(3, 60) = 40.012$, $MSE = 66388.73$, $\acute{\eta}_p^2 = 0.666$], the Sequence effect [$F(1, 20) = 42.218$, $MSE = 32035.234$, $\acute{\eta}_p^2 = 0.678$] and the interactions of Block and Sequence [$F(9, 180) = 2.216$, $MSE = 2646.070$, $\acute{\eta}_p^2 = 0.099$], Block and Stimulus [$F(27, 540) = 1.756$, $MSE = 3576.231$, $\acute{\eta}_p^2 = 0.080$] and Sequence and Stimulus [$F(3, 60) = 10.217$, $MSE = 11884.535$, $\acute{\eta}_p^2 = 0.338$].

The interaction of Block and Sequence is a result of faster learning for sequences that appear 80% of the time. The interaction of Block and Stimulus is a result of different Block effect for Stimulus 1 compared with the other three stimuli. In particular, for Stimulus 1, the slope of the curve for RT as a function of block is less steep, indicating that the psychological process underlying the response for Stimulus 1 is different to that for the other stimuli. In light of previous research in motor learning (e.g., Perlman et al., under review; Sternberg et al., 1988), we suggest that this difference indicates that before responding to Stimulus 1 the entire sequence of responses is prepared; the amount of time required for this preparation effect does not diminish with time (hence the less steep slope). However, the comparison between RT for Stimulus 1 vs. the remaining stimuli and the corresponding linear trend according to block was not significant [$F(1, 20) = 2.15$, $MSE = 29094.71$, $p > .1$, $\acute{\eta}_p^2 = 0.097$]. This difference in processing the response for Stimulus 1 also explains the larger effect of sequence for the first stimulus and the interaction between sequence and stimulus. Accordingly, the interaction of comparison between the contrast Stimulus 1 vs. the remaining stimuli and the Sequence factor was significant [$F(1, 20) = 14.48$, $MSE = 24325.4$, $\acute{\eta}_p^2 = 0.419$].

Of particular interest is the difference in latency for the second stimulus. If AC forms a separate chunk, there should be no reason that responding to C would be faster in the ACBD order (that appears 80% of the time) compared to the less frequent order; in both cases, presumably, participants would be processing AC as the same chunk. However, performing the second response was

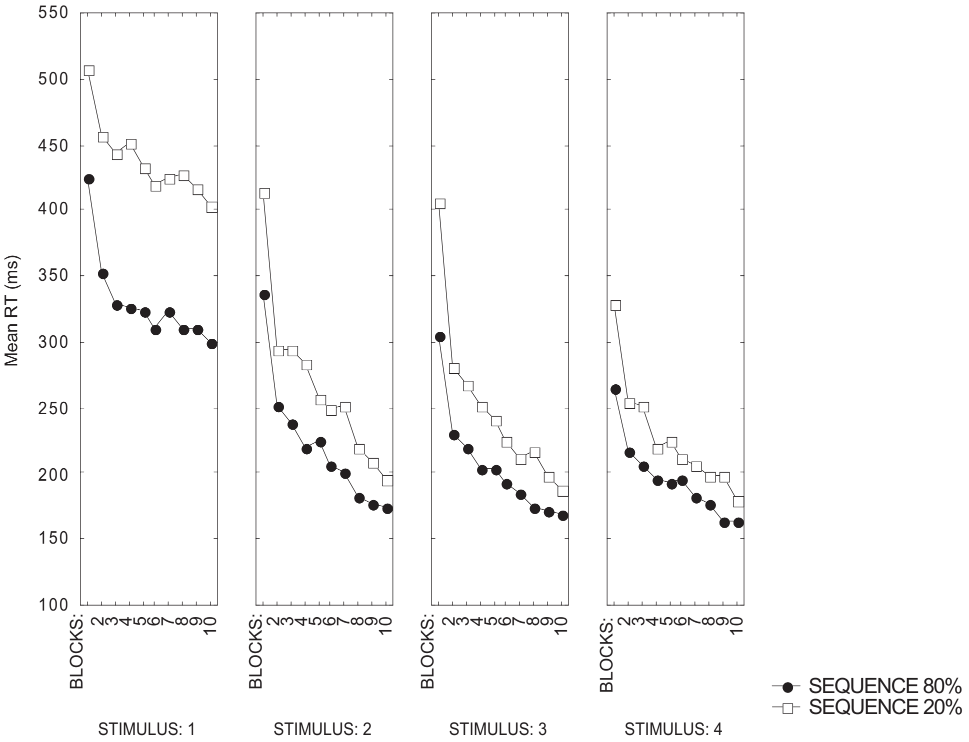

Figure 1. Mean of the median reaction times (RTs) for each training block of trials in each condition in Experiment 1.

faster when it was made within the sequence which appeared more frequently [$F(1, 20) = 27.052$, $MSE = 7944.91$, $\acute{\eta}_p^2 = 0.574$]. This result supports the hypothesis that responding is organized in terms of behavioral units which reflect task demands, but not co-occurrence information. The prediction that follows from an organization of responses into two behavioral units determined by task demands is also supported by a triple interaction between Sequence, Block, and Stimulus and also by the fact that in advanced stages of training the advantage of processing C in the more frequent sequence is larger. However, an advantage for responding to C in the more frequent sequence was evident from the outset a finding slightly surprising (since we did not expect participants to become sensitive to the difference in the relative frequency of the sequences so quickly).

Experiment 2

The results of Experiment 1 indicated that unitization of responses was determined by task demands (the requirement to respond to two different sequences of responses), rather than co-occurrence statistics. The particular mode of unitization is perhaps a result of the fact that the cognitive system needs to distinguish between two sets of responses on a continuous basis (in other words, on any given trial participants could see either of the two sequences and they had to be constantly prepared to respond on the basis of either of them). Hence, it makes sense to organize these responses into single units, to facilitate the task of constantly having to shift between responding according to the two sequences. In Experiment 2 we sought to extend these results, by having the two sequences appearing in different blocks. In such a situation there would be a reduced requirement to unitize the responses in a task-relevant way, so that unitization might be determined by co-occurrence statistics. The sequence that was intended to appear more frequently was presented in blocks having more trials and the sequence that appeared less frequently was presented in blocks having fewer trials.

Method

Twenty-six university students, who did not take part in Experiment 1, were employed in the experiment. We used the same sequences as in Experiment 1. The sequence ACBD was presented 32 times in each long block, and the sequence ACDB was presented 8 times in each short block. The first training block consisted of a long one followed by a short one, followed by long one and so on. Accordingly, there were five long blocks, and each one of them was followed by a short block. For both sequences, the rectangle was presented in blue.

Results and Discussion

Visual inspection of the mean latencies in the various conditions presented in Figure 2 show that the results of the experiment are broadly similar to those of Experiment 1. Of particular interest is the difference in the RT for the second response (C), in the case of the more frequent sequence and the less frequent one. This comparison allows us to examine whether unitization was determined by task demands or co-occurrence information.

The mean RTs of each block of responses were submitted to a three-way within subject analysis of variance (ANOVA) with

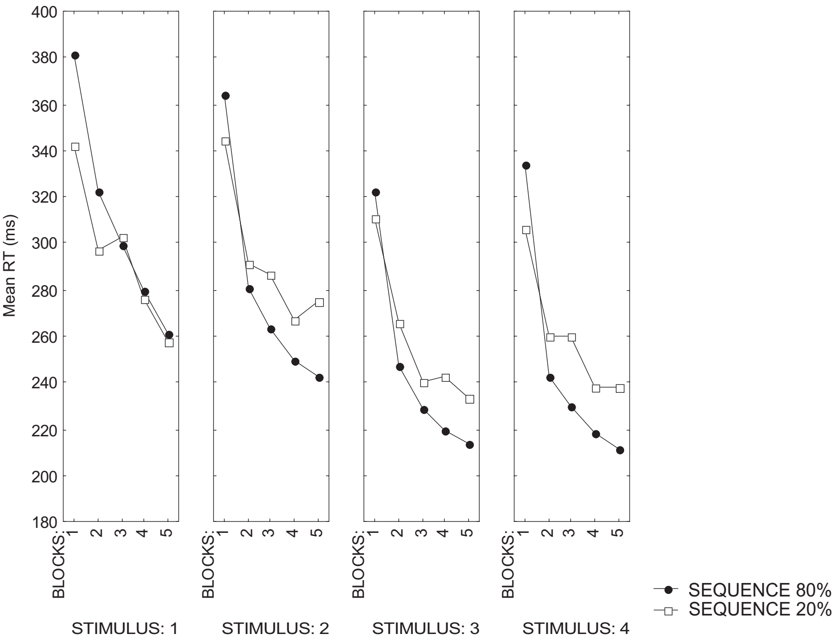

Figure 2. Mean of the median reaction times (RTs) for each training block of trials in each condition in Experiment 2.

Sequence, Block (five blocks), and Stimulus as the manipulated factors. Average error rates were 1.5% for the sequence ACBD and 1.5% for the sequence ACDB. The Block effect was significant, [$F(4, 100) = 40.95$, $MSE = 7539.495$, $\acute{\eta}_p^2 = 0.620$], as was the Stimulus effect [$F(3, 75) = 5.207$, $MSE = 30014.254$, $\acute{\eta}_p^2 = 0.172$]. Moreover, significant were the interactions between Block and Sequence [$F(4, 100) = 7.239$, $MSE = 2336.463$, $\acute{\eta}_p^2 = 0.224$] and of Sequence and Stimulus [$F(3, 75) = 5.381$, $MSE = 2108.793$, $\acute{\eta}_p^2 = 0.177$]. The interaction of Block and Sequence indicates that there was no difference between the sequences at the first blocks but there was a difference in later blocks. The interaction of Sequence and Stimulus indicates that there was a smaller difference in latency between the sequence that appeared more and the sequence that appeared less in the first stimulus and there was a larger difference in the second third and fourth stimuli [$F(1, 25) = 15.48$, $MSE = 2188.57$, $\acute{\eta}_p^2 = 0.382$].

Figure 2 shows a clear difference in mean RTs between the sequence that was presented more often and the sequence that was presented less often, for Stimulus 2, such that for the sequence that appeared more, RTs were shorter [$F(1, 25) = 4.657$, $MSE = 2259.263$, $\acute{\eta}_p^2 = 0.157$]. This result replicates and extends the results of Experiment 1, where we also observed an effect of sequence unitization determined by task demands, but not in different blocks, as we presently do.

Experiment 3

In both Experiment 1 and Experiment 2, our results suggest that the two sequences of responses were processed and responded to as single units. This particular mode of unitization indicates unitization according to task demands, not co-occurrence information. In Experiment 3, we looked for a stronger test for this effect. In particular, we used the sequences ACAD and ACAB. The shared fragment in these two sequences is larger. Participants could perform the first three responses as a unit and perform the last response separately (as indeed the hierarchical editor model of Rosenbaum, 1987, would predict). In such a case, we wouldn't expect any difference in latency between the two sequences for the second and third responses, if unitization is driven by co-occurrence and not by the task-relevant demands. By contrast, if the entire sequence of four responses is performed as one unit (as was the case in Experiments 1 and 2), we anticipate lower reactions time for the second and third responses for the sequence that is observed more often.

Method

Eighteen students were employed in the experiment. The experiment was exactly the same as Experiment 1. The only difference was that we used the sequences ACAD and ACAB. The sequence ACAD was presented 80% of the times.

Results and Discussion

Inspection of the mean latencies in the different conditions, as shown in Figure 3, readily illustrates that the results were very similar to those of Experiment 1. For all stimuli RTs were lower for the sequence that was observed more frequently compared to the sequence that was observed less frequently. Particularly interesting for theoretical reasons are the reaction times for Stimuli 2 and 3.

As in Experiments 1 and 2, these results are consistent with the idea that the two sequences of four responses were organized into two separate units. Likewise, as before, there was no evidence that chunks were formed depending the frequency of co-occurrence between individual responses.

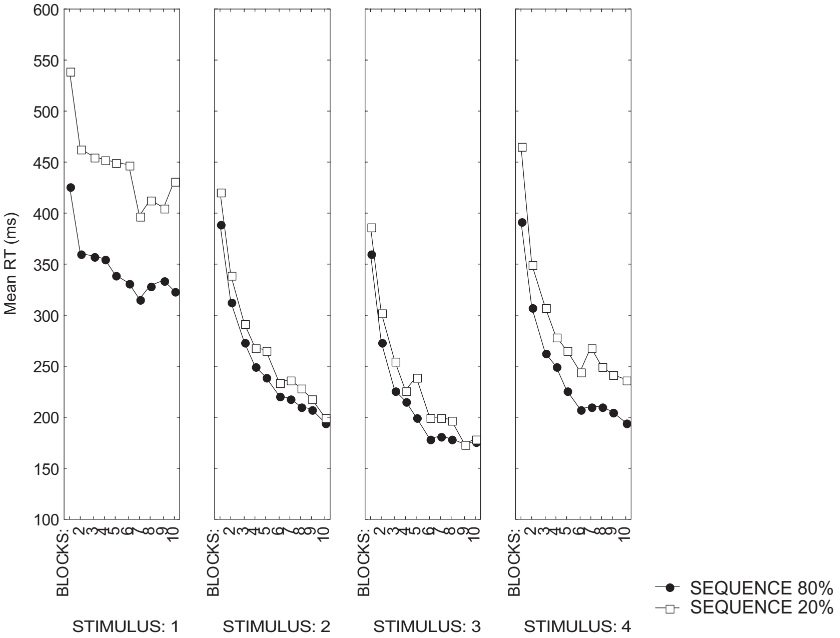

Figure 3. Mean of the median reaction times (RTs) for each training block of trials in each condition in Experiment 3.

The analyses reported below verify the above points. RTs were submitted to a three-way ANOVA with Sequence, Block, and Stimulus, as within-participant factors. Error rates were 2.5% for the sequence ACAD and 2.7% for the sequence ACAB. The Block effect was significant, [$F(9,\ 153) = 35.65$, $MSE = 12680.946$, $\eta_p^2 = 0.677$], as was the Sequence effect [$F(1,\ 17) = 83.908$, $MSE = 8764.904$, $\eta_p^2 = 0.831$], the Stimulus effect [$F(3,\ 51) = 18.890$, $MSE = 105606.181$, $\eta_p^2 = 0.526$], the interactions between Stimulus and Block [$F(27,\ 459) = 2.988$, $MSE = 3412.645$, $\eta_p^2 = 0.149$] and Sequence and Stimulus [$F(3,\ 51) = 21.600$, $MSE = 5875.261$, $\eta_p^2 = 0.559$]. The interaction between Stimulus and Block is a result of a steeper slope in the curve showing the reduction of reaction time with learning for Stimuli 2, 3, and 4, as compared with Stimulus 1 [$F(1,\ 17) = 8.22$, $MSE = 20640.7$, $\eta_p^2 = 0.325$]. As before, we suggest that this is a result of a preparation effect for the entire sequence, which is ‘included’ in the RT for Stimulus 1 (cf. Perlman et al., under review). A similar suggestion applies to the interaction between Sequence and Stimulus [$F(1,\ 17) = 31.02$, $MSE = 11004.2$, $\eta_p^2 = 0.154$].

The figure also shows a clear difference in mean RT for responding to Stimuli 2 and 3, so that the RT for these stimuli is lower for the sequence that was presented more frequently, compared to the sequence that was presented less frequently (for Stimulus 2, $F(1,\ 17) = 14.390$, $MSE = 2103.429$, $\eta_p^2 = 0.458$; for Stimulus 3, $F(1,\ 17) = 19.227$, $MSE = 1769.753$, $\eta_p^2 = 0.530$) such that for the sequence that was presented more, RTs are shorter. As noted previously, this effect would be expected if the sequences of responses were processed, and responded to, as single units. Note, however, that the effect was not significant at block 10 [$F < 1$]. This could be due to the fact that, because of the larger shared fragment, unitization according to task demands took longer to achieve.

Experiment 4

This experiment is a replication of Experiment 2 but we used the sequences ACAD and ACAB. In other words, in this experiment there are two reasons which may enhance chunking due to co-occurrence, as opposed to unitization due to task demands. First, the shared part between the two sequences was greater. Second, each sequence appeared in different blocks, so that participants would not need to constantly switch between responding on the basis of one sequence as opposed to the other. Twenty-three participants were employed in the experiment.

Results and Discussion

Visual inspection of the mean latencies in the various conditions presented in Figure 4 show that the results of the experiment are different from those of the previous experiments. In particular, it seems that there is no difference in RT for the second response, between the sequence that was presented more frequently compared to the sequence that was presented less frequently.

RTs were submitted to a three-way ANOVA with Sequence, Block, and Stimulus, as within-participant factors. Error rates were 1.5% for the sequence ACAD and 1.5% for the sequence ACAB. The Block effect was significant, [$F(4,\ 88) = 43.574$, $MSE = 16533.698$, $\eta_p^2 = 0.664$], as was the Stimulus effect [$F(3,\ 66) = 5.278$, $MSE = 15389.602$, $\eta_p^2 = 0.193$], the interactions between Stimulus and Block [$F(12,\ 264) = 1.986$, $MSE = 2983.330$, $\eta_p^2 = 0.082$] and Sequence and Stimulus [$F(3,\ 66) = 6.688$, $MSE = 3661.197$, $\eta_p^2 = 0.233$]. The interaction between Stimulus and Block is probably a result of steeper reduction of RT with training for Stimuli 2, 3, and 4, as compared with Stimulus 1 (as noted in previous experiments). The interaction of

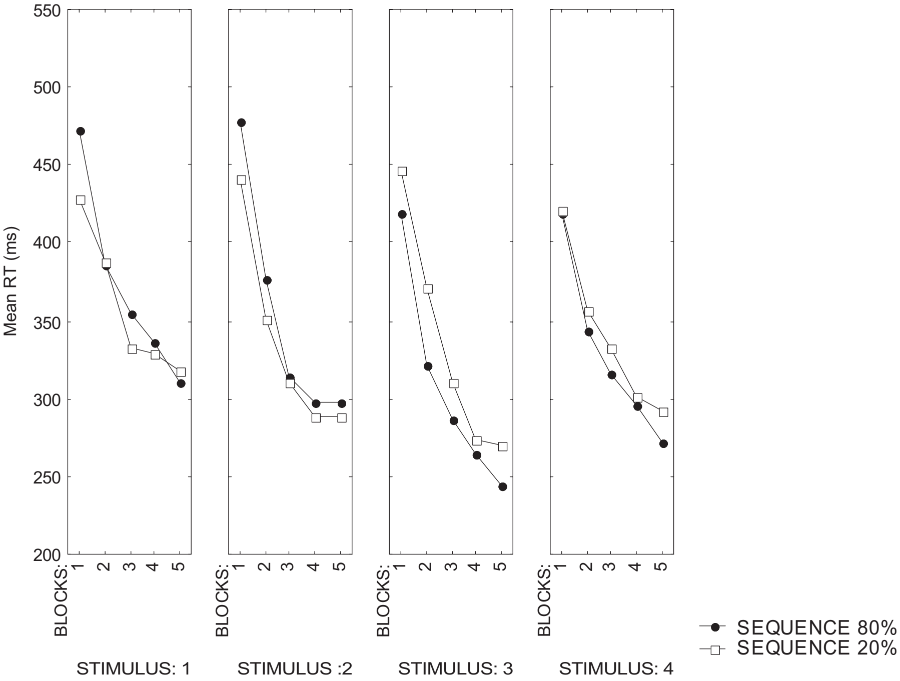

Figure 4. Mean of the median reaction times (RTs) for each training block of trials in each condition in Experiment 4.

Sequence and Stimulus indicates that there was no difference between the sequence that appeared more and the sequence that appeared less frequently, with respect to RTs for the first and last stimuli [F $<$ 1], but there was some difference regarding the second and third.

If the sequence of responses was processed and responded to as a single unit, we would expect a clear difference in mean RTs for Stimuli 2 and 3, for the sequence that was presented more frequently compared to the one that was presented less frequently. Our results showed no such effect. Although there was a difference for the third response [$F(1, 22) = 5.06$, $MSE = 8415.19$, $\acute{\eta}_{\text{p}}^2 = 0.187$], the difference for the second response was in the opposite direction [$F(1, 22) = 4.17$, $MSE = 3775.04$, $\acute{\eta}_{\text{p}}^2 = 0.159$]. This pattern of results is not sufficient to conclude unitization for the entire sequence of four responses. Rather, it indicates unitization due to co-occurrence. Accordingly, in this experiment, we obtained the first evidence of chunking that is driven by co-occurrence, as opposed to chunking of the responses in a task-relevant way.

Experiment 5

Experiment 5 provides a re-examination of a puzzling effect in Experiment 4. In Experiment 4, if the sequence of four responses was processed and responded to as a single unit, we would expect a clear difference in mean RTs for Stimuli 2 and 3, for the sequence that was presented more frequently compared to the one that was presented less frequently. Our results showed no such effect. Although there was a difference for the third response the difference for the second response was in the *opposite* direction. This effect cannot be explained in terms of unitization. Accordingly we replicated Experiment 4. Twenty participants were employed in the experiment; all aspects of Experiment 5 were identical to those of Experiment 4.

Results and Discussion

Visual inspection of the mean latencies in the various conditions presented in Figure 5 show that the results of the experiment are different to those of the previous experiments. In particular, there is no difference in RT for the second response, between the sequence that was presented more frequently compared to the sequence that was presented less frequently.

RTs were submitted to a three-way ANOVA with Sequence, Block, and Stimulus, as within-participant factors. Error rates were 2.1% for the sequence ACAD and 2.3% for the sequence ACAB. The Block effect was significant, [$F(4, 76) = 32.740$, $MSE = 9557$, $\acute{\eta}_{\text{p}}^2 = 0.632$], as was the Stimulus effect [$F(3, 57) = 28.809$, $MSE = 6746$, $\acute{\eta}_{\text{p}}^2 = 0.602$], and the interactions between Sequence and Stimulus [$F(3, 57) = 2.856$, $MSE = 3661.197$, $\acute{\eta}_{\text{p}}^2 = 0.130$]. If the sequence of responses was processed and responded to as a single unit, we would expect a clear difference in mean RTs for Stimuli 2 and 3, for the sequence that was presented more frequently compared to the one that was presented less frequently. Our results showed no such effect. Although there was a difference for the third response [$F(1, 19) = 3.758$, $MSE = 4919.69$, $\acute{\eta}_{\text{p}}^2 = 0.165$], there was no difference for the second response [F $<$ 1]. This pattern of results is not sufficient to conclude unitization for the entire sequence of four responses. Accordingly, in this experiment, we obtained more evidence of chunking that is driven entirely by co-occurrence, as opposed to chunking of the responses in a task-relevant way.

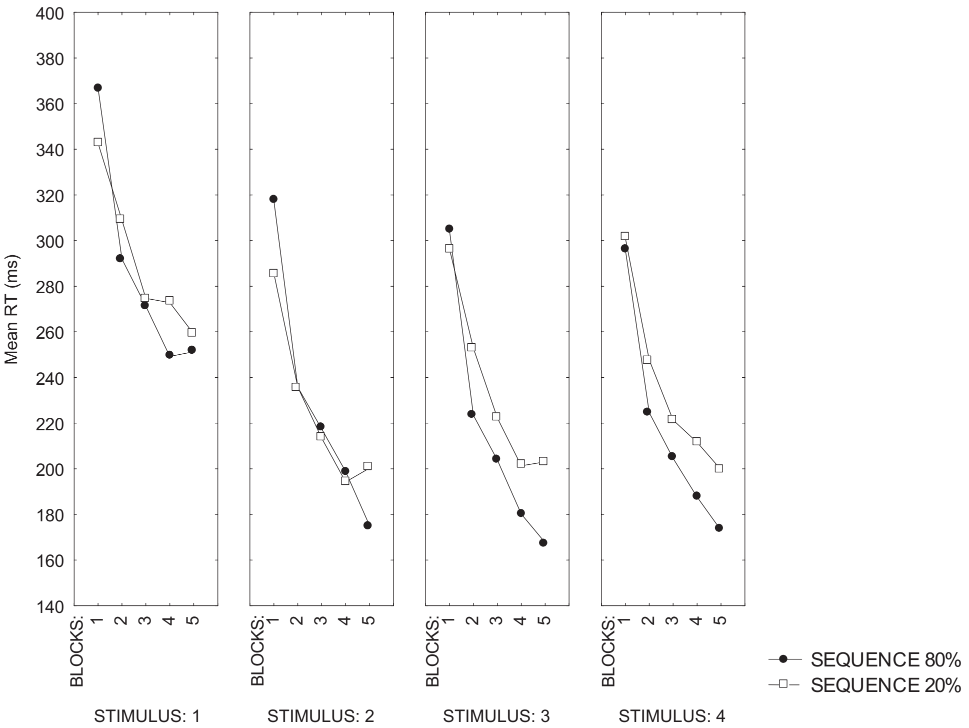

Figure 5. Mean of the median reaction times (RTs) for each training block of trials in each condition in Experiment 5.

Finally, in Experiment 4 we observed a puzzling effect, namely a trend for the RT for the second response to be faster in the less frequent sequence compared to the more frequent sequence. As this effect was not replicated in Experiment 5, we can provisionally consider it a peculiarity of Experiment 4, rather than a genuine psychological effect in need of an explanation. With the exception of this difference, Experiment 5 replicated Experiment 4.

Experiment 6

The results of Experiments 1, 2, and 3 support our hypothesis that responding is organized in terms of one of two behavioral units, each corresponding to a sequence of responses. The organization of responses into two behavioral units leads to a prediction that in advanced stages of training the RT for C in the more frequent sequence would be lower than its RT in the less frequent sequence (Experiments 1, 2; correspondingly for responding to CA in Experiment 3).

However, organization of responses into single units is supposed to be a time consuming process. Therefore, the finding of unitization early in training for Experiment 1 is slightly problematic. The aim of Experiment 6 is to investigate performance early in training in more detail. The experiment is an exact replication of Experiment 1. The only difference is that each block of training included seven presentations of the sequence ACBD and three presentations of the sequence ACDB. We used a much shorter training schedule since we were interested in the early stages of training. We expected no advantage in processing C, in the more frequent sequence, early in training; rather, we expected the advantage to emerge in later stages of training.

Results and Discussion

Inspection of the mean latencies in the different conditions, as shown in Figure 6, readily illustrates that the results were very similar to those of Experiment 1. For all stimuli, and particularly for Stimulus 2, RTs were lower for the sequence that was observed more frequently compared to the sequence that was observed less frequently. These results are consistent with the results and conclusions of Experiments 1, 2, and 3.

The analyses reported below verify the above points. RTs were submitted to a three-way ANOVA with Sequence, Block, and Stimulus, as within-participant factors. Error rates were 2.3% for the sequence ACBD and 3.5% for the sequence ACDB. The Block effect was significant, [$F(9, 108) = 22.99$, $MSE = 20655$, $\hat{\eta}_p^2 = 0.657$], as was the Sequence effect [$F(1, 12) = 55.42$, $MSE = 12848$, $\hat{\eta}_p^2 = 0.822$], the Stimulus effect [$F(3, 36) = 15.70$, $MSE = 73436$, $\hat{\eta}_p^2 = 0.566$], the interactions between Stimulus and Block [$F(27, 324) = 1.63$, $MSE = 9129$, $\hat{\eta}_p^2 = 0.119$], and the triple interaction of Stimulus, Sequence, and Block [$F(27, 324) = 1.64$, $MSE = 5089$, $\hat{\eta}_p^2 = 0.120$]. The interaction between Stimulus and Block is a result of a steeper slope in the curve showing the reduction of RT with learning for Stimuli 2, 3, and 4, compared to Stimulus 1. As before, we suggest that this is a result of a preparation effect for the entire sequence, which is "included" in the RT for Stimulus 1 (cf. Perlman et al., under review). The actual contrast, however, was not significant [$F(1, 12) = 2.43$, $MSE = 567542$, $p = .144$, $\hat{\eta}_p^2 = 0.168$, CI(Stimulus 1) = 51-698, CI(Stimuli 2, 3 & 4) = 89-525].

The prediction that follows from organization of responses into two behavioral units according to task demands is a triple inter-

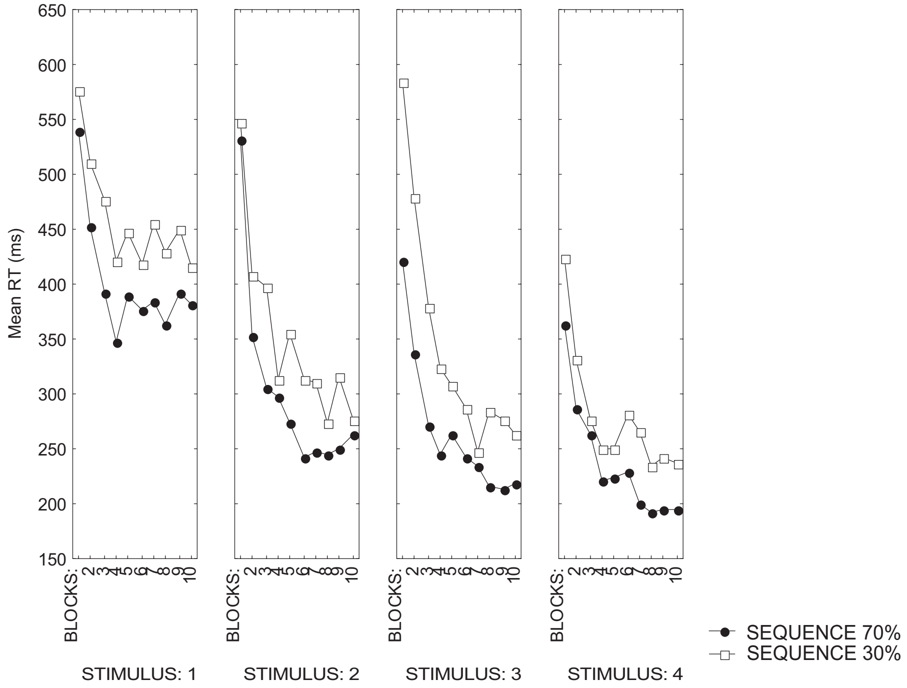

Figure 6. Mean of the median reaction times (RTs) for each training block of trials in each condition in Experiment 6.

action among Sequence, Block, and Stimulus, as observed. In particular, in the advanced stages of training the advantage of processing C in the more frequent sequence is larger. Importantly, such an advantage was *not* observed from the outset, at Block 1, [F < 1], but was observed only from Block 2 onwards [$F(1, 12) = 16.29$, $MSE = 8690.9$, $\acute{\eta}_{\text{p}}^{2} = 0.575$]. This finding is consistent with the assumption that organization of responses into a single unit does not take place immediately, but only after extensive training.

General Discussion

Unitization or chunking is one of the most important psychological processes, as it provides a means of organizing the information we encode from our environment. There are many ways in which unitization could work. Many computational models (e.g., the chunking hypothesis of Servan-Schreiber and Anderson, 1990, the measures developed by Knowlton and Squire, 1996, and the various computational hypotheses of language acquisition; e.g., Elman, 1990), suggest that unitization is a function of co-occurrence: elements co-occurring together end up being represented as a single unit. There is contrasting hypothesis, according to which elements organized together (e.g., the elements of the same response sequence) end up being represented as a single unit. This hypothesis is alluded to by some models of motor learning, in which unitization corresponds to sequences which have behavioral significance (e.g., the production of phonemes corresponding to a single word would be organized in a single unit).

The purpose of this work is to consider whether there might be tension between unitization according to task demands vs. unitization according to co-occurrence. Is it meaningful to talk about two types or modes of unitization processes? If yes, which type is dominant? Finally, are there factors which promote one type as opposed to the other? Tubau et al. (2007) examined an analogous contrast, between stimulus-based control vs. plan-based control of behavior (Hommel, 2000, 2003). However, Tubau et al.'s examination is not entirely equivalent to the one we are interested in, namely whether unitization is primarily driven by co-occurrence statistics vs. task demands. By adapting the serial reaction time task, we were able to develop an experimental paradigm in which these two types of unitization can be directly contrasted.

Participants learned two sequences of responses that can be represented as ACBD and ACDB. The first sequence was presented 80% of the time, the second 20%. If unitization is driven by task demands, we expect that responses ACBD and ACDB would be organized into two separate units. Accordingly, when C is encountered in the more frequent sequence it should be responded to faster than when it is encountered in the less frequent sequence. If unitization is determined by co-occurrence information, then the shared fragment AC should form one unit, so that C should be responded to equally fast regardless of whether it is encountered in the more or less frequent sequence. The results from Experiments 1, 2, 3, and 6 provided evidence for the first mechanism and from Experiments 4 and 5 for the second (Experiment 5 was a replication of 4). These results are consistent with other related, recent work. For example, Giroux and Rey (in press) reported that learning of an artificial language leads to learning of *words* (in the language), rather than sub-lexical units. Likewise, Fiser and Aslin (2005) showed that in perceptual learning, memory for features was biased in favor of combinations which corresponded to (previously learned) whole objects. In other words, if a combination of features corresponded to a subset of the features of an object, then this combination would be remembered more poorly, compared to features which were not such subsets.

Our results are difficult to reconcile with models of chunking based on co-occurrence information. They are also difficult to reconcile with prominent models of motor behavior, such as Rosenbaum's (Rosenbaum, 1987; Rosenbaum et al., 1983, 1987) hierarchical editor model. Rosenbaum's model predicts that the shared component(s) between two sequences should be likewise shared in the planning of the two sequences. In our experiments, AC should have been a common element in the planning of the two sequences, so that C should have been responded to equally fast regardless of whether it appeared in the less or more frequent sequence. We hasten to add that the development of the hierarchical editor model did not take into account frequency manipulations such as the one we employed. It is possible that the model could be successfully revised to take into account our results.

Evaluating the implications of our results for chunking models of learning, such as SRN's, also requires some qualifications. The simplest variety of SRN's, as proposed in the seminal work of Elman (1990, 1991), appears inconsistent with the finding that unitization can be determined by task demands *instead* of co-occurrence information. However, in general SRNs and derivative models have provided excellent fits to many aspects of human behavior and it is possible that they can be developed to account for the present results as well.

Similar considerations apply the various versions of the chunking hypothesis. In this case, there has been a recent variation which does appear consistent with our main finding, the PARSER model by Perruchet, Vinter, Pacteau, and Gallego (2002). Perruchet et al. suggested that humans are able to parse input into structurally relevant units, in cases where there are no obvious boundaries between units. PARSER works by attempting to encode input with a range of possible chunks. Perruchet et al. argued that the primitives that are perceived within the attentional focus (as a result of their experienced temporal or spatial proximity) become the constituents of a new representational unit. Relevant to our results, the chunks which turn out to be useful in encoding are prioritized for encoding future input, while the chunks which are not are gradually eliminated from the system's memory. In this way, PARSER can predict that when a bigger unit is learned (e.g., a unit corresponding to the responses in an entire sequence in our experiments), smaller sub-units *no longer* play a role in performance. For example, in learning sequences ACBD and ACDB, PARSER would originally encode input in terms of the common, frequent fragment AC. However, once the model has recognized that AC is part of larger sequence, it will start using these larger sequences in order to encode input and abandon the use of AC. Similar predictions are possible from Pothos and Wolff's (2008) Simplicity and Power, which is a chunking model derived from principles of information theory. Finally, regarding the predictions from PARSER and the Simplicity and Power model, it is worth noting that in our work the degree of unitization was affected by actual task demands (e.g., whether presentation of the two sequences was intermixed or blocked) as well. Whether computational models of learning can explain this finding is an interesting issue for future work.

In Experiments 4 and 5 our results indicated that unitization was driven by co-occurrence, rather than by task demands. It therefore looks like in some (most?) cases unitization is determined by task demands, but in other cases by co-occurrence. What is the basis for preferring one unitization strategy as opposed to another? We can consider the key differences between Experiment 4/5 and Experiments 1, 2, 3, and 6. First, in Experiment 4/5 the fragment shared by the two sequences was larger. Second, in Experiment 4/5 sequence presentation was blocked. We can adopt a functional perspective and ask under which circumstances would it make more sense for the cognitive system to organize information in terms of task demands or co-occurrence. We can minimally suggest that if there is no or little shared information between the two sequences, then it would be more efficient to organize responses according to task demands. Also, the mode of responding might make a particular type of unitization more useful. In other words, if there are frequent shifts between responding on the basis of one sequence and another sequence, then, arguably organizing the sequences into units would facilitate switching. Experiment 4 suggests that appropriate values for both of these factors together can lead to unitization according to co-occurrence.

Clearly, the above is not a *model* of when different kinds of unitization are expected to occur—rather, it is merely a descriptive statement of our results. A proper corresponding model for unitization would need to not only explain the observed pattern of results, but also reconcile the insights about unitization from influential ideas in motor learning (such as from the hierarchical editor model of Rosenbaum, 1987).

An issue which we have avoided to consider is that of whether the implementation of knowledge about a unitized sequence of responses is explicit or implicit. The recent work of Tubau et al. (2007) indicated that plan-based control of behavior is explicit. In our experiments, it is extremely likely that participants had explicit knowledge of the two sequences, since the sequences were reasonably short and they were extensively trained. However, note that we did not directly assess the explicitness of sequence knowledge. Also, given the briefness of the RTs we observed, there might be an argument that the actual implementation of the knowledge was not explicitly done, even if the knowledge per se could be explicitly accessed (for discussions see Dienes & Perner, 1999; Perlman & Tzelgov, 2006; Perruchet & Vinter, 2002; Pothos, 2007). A proper resolution of this issue should, at the very least, involve interference manipulations like the ones employed by Tubau et al. (2007).

Summing up, the notions of unitization and chunking have attracted widespread interest in psychology research. The purpose of this study was to explore some of the factors which may be affecting the type of unitization which occurs. In most cases, we found unitization in terms of task demands. As well, we were able to identify unitization driven by co-occurrence information alone, when the task did not require participants to constantly switch between responding on the basis of one sequence, as opposed to the other, and when the two sequences overlapped considerably with each other. We hope these findings will further aid the development of unitization and motor learning models.

References

Berg, T. (2002). Slips of the typewriter key. *Applied Psycholinguistics, 23,* 185–207.

Boucher, L., & Dienes, Z. (2003). Two ways of learning associations. *Cognitive Science, 27,* 807–842.

Buchner, A., Steffens, M. C., & Rothkegel, R. (1998). On the role of fragmentary knowledge in a sequence learning task. *The Quarterly Journal of Experimental Psychology, 51,* 251–281.

Burgess, N., & Hitch, G. J. (1997). Memory for serial order: A network model of the phonological loop and its timing. *Psychological Review, 106,* 551–81.

Christiansen, M. H., & Chater, N. (1999). Toward a connectionist model of recursion in human linguistic performance. *Cognitive Science, 23,* 157–205.

Christiansen, M. H., & Chater, N. (2001). Connectionist psycholinguistics: Capturing the empirical data. *Trends in Cognitive Sciences, 5,* 82–88.

Cleeremans, A., & McClelland, J. L. (1991). Learning the structure of event sequences. *Journal of Experimental Psychology: General, 120,* 235–253.

Dell, G. S., Burger, L. K., & Svec, W. R. (1997). Language production and serial order: A functional analysis and a model. *Psychological Review, 104,* 123–147.

Dienes, Z., & Perner, J. (1999). A theory of implicit and explicit knowledge. *Behavioral and Brain Sciences, 22,* 735–808.

Dirlam, D. K. (1972). Most efficient chunk sizes. *Cognitive Psychology, 3,* 355–359.

Elman, J. L. (1990). Finding structure in time. *Cognitive Science, 14,* 179–211.

Elman, J. L. (1991). Distributed representations, simple recurrent networks, and grammatical structure. *Machine Learning, 7,* 195–225.

Fiser, J., & Aslin, R. N. (2005). Encoding multielement scenes: Statistical learning of visual features hierarchies. *Journal of Experimental Psychology: General, 134,* 521–537.

Giroux, I. & Rey, A. (in press). Lexical and sub-lexical units in speech perception. *Cognitive Science.*

Hare, M., Elman, J. L., & Daugherty, K. G. (1995). Default generalization in connectionist networks. *Language and Cognitive Processes, 10,* 601–630.

Henry, F. M., & Rogers, D. E. (1960). Increased response latency for complicated movements and a memory drum theory of neuromotor reaction. *Research Quarterly, 31,* 448–458.

Hommel, B. (2000). The prepared reflex: Automaticity and control in stimulus-response translation. In S. Monsell & J. Driver (Eds.), *Control of cognitive processes: Attention and performance XVIII* (pp. 247–273). Cambridge, MA: MIT Press.

Hommel, B. (2003). Planning and representing intentional action. *The Scientific World Journal, 3,* 593–608.

Houghton, G. (1990). The problem of serial order: A neural network model of sequence learning and recall. In R. Dale, C. Mellish, & M. Zock (Eds.), *Current research in natural language generation* (pp. 287–319). London: Academic Press.

Johnstone, T., & Shanks, D. R. (1999). Two mechanisms in implicit grammar learning? Comment on Meulemans and Van der Linden (1997). *Journal of Experimental Psychology: Leaning, Memory, and Cognition, 25,* 524–531.

Knowlton, B., & Squire, L. R. (1994). Artificial grammar learning and implicit memory: Reply to Higham and Vokey. *Psychological Science, 5,* 61.

Knowlton, B. J., & Squire, L. R. (1996). Artificial grammar learning depends on implicit acquisition of both abstract and exemplar-specific information. *Journal of Experimental Psychology: Learning, Memory and Cognition, 22,* 169–181.

Luria, R., & Meiran, N. (2003). Online order control in the psychological refractory period paradigm. *Journal of Experimental Psychology: Human Perception and Performance, 29,* 556–574.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *The Psychological Review, 63,* 81–97.

Palmer, C., & Pfordresher, P. Q. (2003). Incremental planning in sequence production. *Psychological Review, 110,* 683–712.

Perlman, A., & Tzelgov, J. (2006). Interactions between encoding and retrieval in the domain of sequence-learning. *Journal of Experimental Psychology, Learning Memory and Cognition, 32,* 118–130.

Perlman, A., Tzelgov, J., Pothos, E. M., & Millar, D. (under review). The emergence of response–ready sequences in the extended Cue Target Interval procedure.

Perruchet, P., & Pacteau, C. (1990). Synthetic grammar learning: Implicit rule abstraction or explicit fragmentary knowledge? *Journal of Experimental Psychology: General, 119,* 264–275.

Perruchet, P., & Peereman, R. (2004). The exploitation of distributional information in syllable processing. *Journal of Neurolinguistics, 17,* 97–119.

Perruchet, P., & Vinter, A. (2002). The self-organizing consciousness. *Behavioral and Brain Sciences, 25,* 297–388.

Perruchet, P., Vinter, A., Pacteau, C., & Gallego, J. (2002). The formation of structurally relevant units in artificial grammar learning. *Quarterly Journal of Experimental Psychology, 55*A, 485–503.

Plunkett, K., & Marchman, V. (1993). From rote learning to system building: Acquiring verb morphology in children and connectionist nets. *Cognition, 48,* 21–69.

Portier, S. J., & van Galen, G. P. (1992). Immediate vs. postponed visual feedback in practicing a handwriting tasks. *Human Movement Science, 11,* 563–592.

Pothos, E. M. (2007). Theories of artificial grammar learning. *Psychological Bulletin, 133,* 227–244.

Pothos, E. M., & Bailey, T. M. (2000). The importance of similarity in artificial grammar learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 847–862.

Pothos, E. M., & Wolff, J. G. (2006). The Simplicity and Power model for inductive inference. *Artificial Intelligence Review, 26,* 211–225.

Rosenbaum, D. (1987). Successive approximations to a model of human motor programming. In G. H. Bower (Ed.) *The psychology and learning and motivation,* pp. 153—182. Academic Press: San Diego.

Rosenbaum, D. A., Hindorff, V., & Munro, E. M. (1987). Scheduling and programming of rapid finger sequences: Tests and elaborations of the hierarchical editor model. *Journal of Experimental Psychology: Human Perception and Performance, 13,* 193–203.

Rosenbaum, D. A., Kenny, S. B., & Derr, M. A. (1983). Hierarchical control of rapid movement sequences. *Journal of Experimental Psychology: Human Perception and Performance, 9,* 86–102.

Rumelhart, D. E., & McClelland, J. L. (1986). *Parallel distributed processing: Explorations in the microstructure of cognition.* Cambridge, MA: MIT Press.

Rumelhart, D. E., & Norman, D. A. (1982). Simulating a skilled typist: A study of skilled cognitive-motor performance. *Cognitive Science, 6,* 1–36.

Schneider, D. W., & Logan, G. D. (2006). Hierarchical control of cognitive processes: Switching tasks in sequences. *Journal of Experimental Psychology: General, 135,* 623–640.

Servan-Schreiber, E. (1991). *The competitive chunking theory: Models of perception, learning, and memory.* Doctoral Dissertation, Department of Psychology, Carnegie-Mellon University.

Servan-Schreiber, E., & Anderson, J. R. (1990). Learning artificial grammars with competitive chunking. *Journal of Experimental Psychology: Learning Memory and Cognition, 16,* 592–608.

Simon, H. A., & Barenfeld, M. (1969). Information-processing analysis of perceptual processes in problem solving. *Psychological Review, 76,* 473–483.

Sternberg, S., Knoll, R. L., Monsell, S., & Wright, C. E. (1988). Motor programs and hierarchical organization in the control of rapid speech. *Phonetica, 45,* 175–197.

Sternberg, S., Monsell, S., Knoll, R. L., & Wright, C. E. (1978). The latency and duration of rapid movement sequences: Comparisons of speech and typewriting. In G. E. Stelmach (Ed.), Information processing in motor control and learning. New York: Academic Press.

Tubau, E., Hommel, H., & Lopez-Moliner, J. (2007). Models of executive control in sequence learning: From stimulus-based to plan-based control. *Journal of Experimental Psychology: General, 136,* 43–63.

van Galen, G. P. (1991). Handwriting: Issues for a psychomotor theory. *Human Movement Science, 10,* 165–191.

Vousden, J. I., Brown, G. D. A., & Harley, T. A. (2003). Serial control of phonology in speech production: A hierarchical model. *Cognitive Psychology, 41,* 101–175.

Received July 12, 2008
Revision received May 20, 2009
Accepted June 4, 2009 ■

Members of Underrepresented Groups:
Reviewers for Journal Manuscripts Wanted

If you are interested in reviewing manuscripts for APA journals, the APA Publications and Communications Board would like to invite your participation. Manuscript reviewers are vital to the publications process. As a reviewer, you would gain valuable experience in publishing. The P&C Board is particularly interested in encouraging members of underrepresented groups to participate more in this process.

If you are interested in reviewing manuscripts, please write APA Journals at Reviewers@apa.org. Please note the following important points:

- To be selected as a reviewer, you must have published articles in peer-reviewed journals. The experience of publishing provides a reviewer with the basis for preparing a thorough, objective review.
- To be selected, it is critical to be a regular reader of the five to six empirical journals that are most central to the area or journal for which you would like to review. Current knowledge of recently published research provides a reviewer with the knowledge base to evaluate a new submission within the context of existing research.
- To select the appropriate reviewers for each manuscript, the editor needs detailed information. Please include with your letter your vita. In the letter, please identify which APA journal(s) you are interested in, and describe your area of expertise. Be as specific as possible. For example, “social psychology” is not sufficient—you would need to specify “social cognition” or “attitude change” as well.
- Reviewing a manuscript takes time (1–4 hours per manuscript reviewed). If you are selected to review a manuscript, be prepared to invest the necessary time to evaluate the manuscript thoroughly.