Evidence for chunking vs. statistical learning in motor sequence production

Nicola J. Popp (npopp@uwo.ca)

Neda Kordjaz (nkordjaz@uwo.ca)

Paul Gribble (paul@gribblelab.org)

Jörn Diedrichsen (jdiedric@uwo.ca)

Shared affiliation: The Brain and Mind Institute, Western Interdisciplinary Research Building, Western University, London, ON N6A 5B7

Abstract:

Many complex behaviors consist of sequentially ordered actions. When acquiring a novel sequential skill, the transition between actions can be performed with increasing speed. This observation has led to the idea that the elementary actions are bound together during the learning process. Two ideas for this process have been proposed: First, statistical probabilities between different elementary actions could be acquired. Secondly, discrete groupings of elementary actions – so-called chunks - could emerge with learning. We discuss the differences between these two ideas and compare the ability of the two models to predict inter-press time intervals (IPIs) measured by a discrete sequence production task. We find a greater ability of the chunk model to predict participants' IPIs throughout learning.

Keywords: chunking; statistical learning; motor learning

Introduction

Serially ordered actions are an important component of human behaviour from typing emails to more complex actions like cooking. Most of the behaviours that we perform are divided into smaller parts that combined make up a complete action sequence. When performing a novel behavior for the first time each of the serially ordered actions is performed slowly and with care, however, with repeated executions, the transitions become increasingly faster. What kind of process could predict these changes in transition speed and are some actions more likely to be bound together?

Two main theories have been used to explain this idea of action binding. Proponents of the statistical learning theory propose that the motor system learns the relative frequency of co-occurrences of actions (transition probabilities; for a review see Perruchet & Pacton, 2006), making more frequent transitions faster. When writing an email, statistical learning would predict that letters that co-occur more frequently in our language, and hence have been performed more often in combination, would likely be executed faster.

Another process that has been suggested to play a role in action binding is "chunking" (Boucher & Dienes, 2003; Perruchet & Pacton, 2006; Perruchet & Vinter, 1998). With our limited working memory capacity executing a long sequence of actions might not be possible at the beginning of learning (Miller, 1956). To overcome this limitation, long sequences are subdivided into groups of actions - so-called chunks. Chunking is said to aid skill acquisition by decreasing cognitive load (Acuna et al., 2014; Ramkumar et al., 2016; Verwey & Dronkert, 1996). The interplay between the emergence of chunks and the learning of transitional probabilities has been under debate without a clear consensus (Boucher & Dienes, 2003; Du & Clark, 2017; Perruchet & Pacton, 2006; Perruchet & Vinter, 1998). Different hypotheses have been proposed: First, statistical learning influences transition speed in the early stages of learning and can lead to the formation of chunks as learning consolidates (Beukema & Verstynen, 2018; Nissen & Bullemer, 1987). Alternatively, chunking could also occur at the earliest stages of learning and influences learning to a greater extent than statistical probabilities (Acuna et al., 2014; Ramkumar et al., 2016). Third, statistical learning and chunking could be mutually exclusive processes, with the system adopting one or the other depending on context (Meulemans & Van Der Linden, 2003).

To distinguish the amount of influence on performance between chunks and transition probabilities in a motor sequence task, we first need to establish the differences between the two concepts (see Fig. 1). Let's take the sequence 51312315134 to exemplify the differences in predicted performance between transitions probabilities and chunks. Because 51, 13 and 31 occur 2 times in the sequence, statistical learning would predict that these 1st order transitions are executed faster than a 1st order transitions that only occurs once such as 12. The same goes for the 2nd order transition 513 which is repeated twice and hence should be performed faster. On the contrary, chunking is believed to be a winner-take-all process (Servan-Schreiber & Anderson, 1990). In our example, the sequence is arbitrarily chunked as 513, 123, 151 and 34. There are 1023 different ways to chunk the sequence; however, with our paradigm we were able to instruct the chunk structure that participants followed early in learning. In this example, 513 is chunked and hence the transitions between these presses should be fast. However, while this sequence of presses reoccurs later in the sequence, this later part might be chunked differently – e.g., as 151 and 34. This leads to a slow transition between 1 and 3 at the end of the sequence. In sum, chunks are supposed to be discrete and a winner-take-all process which does not benefit or generalize to the frequency of presentation and hence could lead to slow transitions even if they have been performed frequently. In contrast, transition probabilities generalize across sequences predicting that the frequency of the transition will benefit speed whenever the transition occurs.

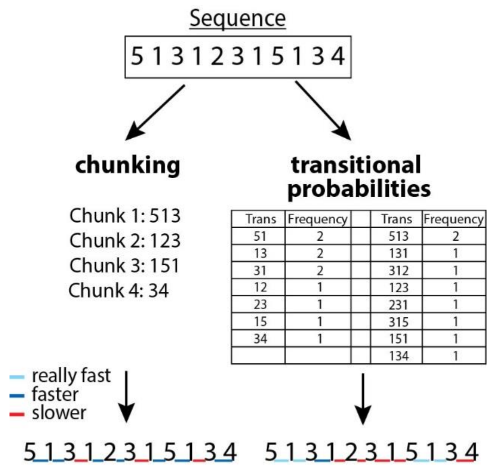

Figure 1. IPI performance predictions based on the chunking and transition model on an example sequence.

Methods

To determine the contribution of the two processes to motor learning, we used the discrete sequence production task (DSP). We were able to estimate the influence of first and second order transitions and chunking on participants' performance throughout 3 weeks of training. Participants practiced sequences of numbers on a keyboard-like device over a 3-week period with the goal to continuously improve execution speed. The first group of participants (N=32) first practiced small 2-3 digit chunks and then learned seven 11-press sequences, each of which consisted of four of the pre-trained chunks. This led to a clear chunking structure across participants at the beginning of training which enabled us to examine chunking without using the resulting data to predict it. We measured the inter-press time intervals (IPIs) between the keypresses. The frequency of 1st and 2nd order transitions were calculated for each day. We then estimated how well a combination model of 1st and 2nd order transitions, a chunk model and the combination of both predicted participants' IPIs. The data was divided into the three weeks of training to examine any changes that could occur with training. We used a cross-validated ordinary linear regression approach to estimate the R^2 between even and odd blocks of trials for each participant and each day separately (across sequences). A noise ceiling was determined through the correlation of the inter-press intervals between even and odd blocks of trials. The cut-off value for significance was corrected for the number of comparisons made within each training phase (0.05/3 = 0.0167).

Results

Comparing the two models and their combination we found that in the first week of training the chunking model predicted significantly more variability than the 1st+2nd order transition model ($t_{(31)} = 8.653$, $p = 9.002e\text{-}10$; Fig. 2). The combination of chunk and transition model did not predict significantly more variability than the chunk model ($t_{(31)} = -2.380$, $p = 0.024$), suggesting that the transition model does not predict any variability beyond what the chunk model already accounts for. While chunking still explained more variance than the transition model ($t_{(31)} = 3.135$, $p = 0.004$) in the last week of training, combining the models led to significantly higher R^2 value ($t_{(31)} = 4.834$, $p = 3.460e\text{-}05$). Overall, these findings suggest that the formation of chunks primarily influences transition speed throughout training and transition probabilities add to performance after the initial learning phase. Hence, chunking seems to play a greater role in the discrete sequence production task than the statistical learning of transition probabilities.

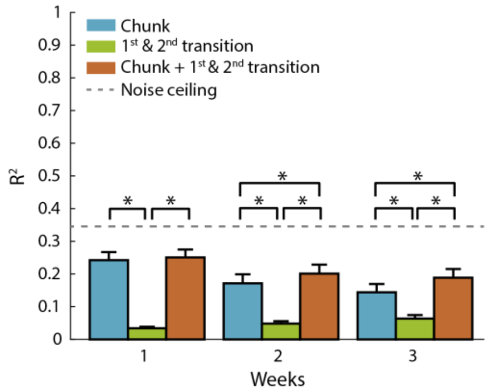

Figure 2. Chunk and transition model comparison. *Cross-validated R^2 values for the transition and the chunk model and their combination. Asterisk indicates significant difference at a significance level of 0.0167. Error-bars represent the standard-error between subjects.*

Acknowledgements

This work is supported by a James S. McDonnell Foundation Scholar award, a NSERC Discovery Grant (RGPIN-2016-04890), a Platform Support Grant from Brain Canada, and the Canada First Research Excellence Fund (BrainsCAN) to J.D., a Natural Sciences and Engineering Council of Canada Grant (RGPIN 238338), a Canadian Institutes of Health Research Grant (PJT-153447), and a National Institute of Child

Health and Human Development Grant (R01 HD075740) to P.G

References

Acuna, D. E., Wymbs, N. F., Reynolds, C. A., Picard, N., Turner, R. S., Strick, P. L., … Kording, K. P. (2014). Multifaceted aspects of chunking enable robust algorithms. *Journal of Neurophysiology*, *112*(8), 1849–1856. https://doi.org/10.1152/jn.00028.2014

Beukema, P., & Verstynen, T. (2018). Predicting and binding: interacting algorithms supporting the consolidation of sequential motor skills. *Current Opinion in Behavioral Sciences*, *20*, 98–103. https://doi.org/10.1016/j.cobeha.2017.11.014

Boucher, L., & Dienes, Z. (2003). Two ways of learning associations. *Cognitive Science*, *27*(6), 807–842. https://doi.org/10.1016/j.cogsci.2003.03.001

Du, Y., & Clark, J. E. (2017). New insights into statistical learning and chunk learning in implicit sequence acquisition. *Psychonomic Bulletin & Review*, *24*(4), 1225–1233. https://doi.org/10.3758/s13423-016-1193-4

Meulemans, T., & Van Der Linden, M. (2003). Implicit learning of complex information in amnesia. *Brain and Cognition*, *52*(2), 250–257. https://doi.org/10.1016/S0278-2626(03)00081-2

Miller, G. A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological Review*, *63*(2), 81–97. https://doi.org/10.1037/h0043158

Nissen, M. J., & Bullemer, P. (1987). Attentional requirements of learning: Evidence from performance measures. *Cognitive Psychology*, *19*(1), 1–32. https://doi.org/10.1016/0010-0285(87)90002-8

Perruchet, P., & Pacton, S. (2006). Implicit learning and statistical learning: one phenomenon, two approaches. *Trends in Cognitive Sciences*, *10*(5), 233–238. https://doi.org/10.1016/j.tics.2006.03.006

Perruchet, P., & Vinter, A. (1998). PARSER: A Model for Word Segmentation. *Journal of Memory and Language*, *39*(2), 246–263. https://doi.org/10.1006/jmla.1998.2576

Ramkumar, P., Acuna, D. E., Berniker, M., Grafton, S. T., Turner, R. S., & Kording, K. P. (2016). Chunking as the result of an efficiency computation trade-off. *Nature Communications*, *7*, 12176. https://doi.org/10.1038/ncomms12176

Servan-Schreiber, E., & Anderson, J. R. (1990). Learning artificial grammars with competitive chunking. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*(4), 592–608. https://doi.org/10.1037/0278-7393.16.4.592

Verwey, W. B., & Dronkert, Y. (1996). Practicing a Structured Continuous Key-Pressing Task: Motor Chunking or Rhythm Consolidation? *Journal of Motor Behavior*, *28*(1), 71–79. https://doi.org/10.1080/00222895.1996.9941735