Humans use different statistics for sequence analysis depending on the task

Dinis Gökaydin (dinis.gokaydin@adelaide.edu.au)
School of Psychology, University of Adelaide

Anna Ma-Wyatt (anna.mawyatt@adelaide.edu.au)
School of Psychology, University of Adelaide

Daniel Navarro (daniel.navarro@adelaide.edu.au)
School of Psychology, University of Adelaide

Amy Perfors (amy.perfors@adelaide.edu.au)
School of Psychology, University of Adelaide

Abstract

Despite its long history (Luce, 1986) the study of sequential effects has mostly been confined to simple binary tasks such as two-alternative forced choice tasks (2AFC). Here we present experimental results from a choice task with three rather than two alternatives (3AFC) as well as a novel model that can explain them. We find that humans change the statistics they use to analyse a sequence depending on the task constraints, relying on first-order transition probabilities in a 2AFC but event relative frequencies (i.e., zeroth-order transition probabilities) in a 3AFC.

Keywords: Sequential effects; reaction time; perception; decision making

Introduction

Sequential effects exist when the response people make to a current stimulus is influenced not just by that stimulus but also on the sequence of previous stimuli. Such effects occur even if the sequence of stimuli is random, and reflect a natural tendency of humans to find patterns in randomness. One well-documented sequential effect is found in reaction times in two-alternative forced choice (2AFC) tasks where subjects are instructed to respond as quickly and as accurately as they can. Speed of response is influenced by the recent sequence history: violations of a local pattern (e.g., B after seeing AAAA or A after seeing BABA) result in longer response times and higher errors, while stimuli that fit a pattern (say B after seeing BBBB or A after ABAB) have faster responses and lower errors (Cho et al., 2002).

Despite a long history of research into sequential effects and reaction times (RTs) in general (Luce, 1986), only recently have models have been proposed that successfully explain these effects (Cho et al., 2002; Yu & Cohen, 2008; Ma-Wyatt & Navarro, 2009; Wilder, Jones, & Mozer, 2010). While these models differ in their details, there seems to be a general agreement that the discounting of past events takes the shape of an exponential function: the impact that a sequence element has on the reaction time to the current stimulus decreases exponentially with distance in the past. In fact, there appears to be an equivalence between the Markov models used to explain sequential effects and some form of exponential filtering (Yu & Cohen, 2008). Ma-Wyatt and Navarro (2009) also demonstrated that sequential structure influences motor control in a pointing task, and that a model including exponential discounting can help explain this effect.

Sequential effects in complex tasks

Although for reasons of simplicity most previous modelling work has focused on explaining performance in tasks with two choices only (2AFC), in the real world people are more commonly confronted with behavioural sequences made up of several different elements, rather than just two. Two obvious questions then arise. First, do sequential effects happen in tasks with more than two choices, like a 3AFC task? Second, can exponential discounting of past events explain these sequential effects, as it does in the simpler task?

We address these questions by first measuring human responses in a 3AFC and then presenting a novel model that can explain the overall pattern of reaction times. We demonstrate that not only do subjects show systematic sequential effects in a 3AFC but that these are well captured by a simple model that incorporates exponential discounting of past events. We also investigate what kind of statistics people are tracking in a sequential task (see below). Previous models can either handle different statistics but not multiple options (Yu & Cohen, 2008; Wilder et al., 2010), or can handle multiple options but not different order statistics (Ma-Wyatt & Navarro, 2009), while our model can handle both. This is necessary if we want to investigate the different statistics that people use across tasks.

What statistics are people tracking?

There are many different kinds of sequential effects, depending on the statistics that people might use. For instance, people might simply track the overall frequencies of the sequential stimuli (e.g., $P(A)$, which is zeroth-order). Alternatively, they might track higher-order transition probabilities, including the probability of one element given the last element seen (e.g. $P(A|B)$, which is first-order), or the probability given the last two elements (e.g., $P(A|B,C)$, which is second-order), and so on. A closely related question is whether humans always rely on the same type of statistics, or whether this varies within a single task, across different tasks, or across individuals. Here we will focus on investigating whether people use different statistics when the task changes from 2AFC to 3AFC.

One major difference between binary (2AFC) and ternary (3AFC) tasks is that the relationship between the number of variables an individual needs to keep track of and the order of the transition probabilities is different. In a binary task, if using zeroth-order transition probabilities, there are two probabilities to track ($P(A)$ and $P(B)$). This number increases to four ($P(A|A)$,$P(A|B)$,$P(B|A)$ and $P(B|B)$) if using first-order transition probabilities. In a ternary task with three different elements, the number of probabilities to track increases: there are now three zeroth-order and nine first-order probabilities. As a result, while a task with more elements is presumably always harder, higher-order statistics grow relatively more difficult to track than lower-order statistics: the number of zeroth-order probabilities varies linearly with the number of different sequence elements, while the number of first-order probabilities varies quadratically.

Objectives

This work has several goals. Our first objective is to introduce and validate an experimental paradigm with a sequential task with more than two alternatives. Our second aim is to develop a simple model that can straightforwardly predict reaction time performance in such a task as a function of tracking different types of statistics (here we focus on zeroth and first-order probability tracking). We demonstrate that our model obtains a good fit to the RT data of a 2AFC as well as a 3AFC. Interestingly, our results indicate that humans track first-order transition probabilities in a 2AFC task but only zeroth-order frequencies in the case of a 3AFC task.

Model

Our model is a simple extension of an n-gram model, common in computational linguistics (e.g., Manning & Schütze, 1999), to which we add exponential discounting of past events. An n-gram model simply keeps track of the $(n-1)$-th order transition probabilities within a sequence of elements. This framework allows for a clear distinction between tracking zeroth-order and first-order transition probabilities, corresponding to two sub-models that we shall henceforth refer to as 0th and 1st-order respectively. In principle, n-gram models can also easily be expanded to accommodate as many different different sequence elements as necessary.

The 0th order model is effectively just an exponential moving average. Given a sequence of events $S_n = \{s_1, ..., s_n\}$, the probability of the next event is given by:

$$P(s_{n+1}) = \frac{\sum_{i=1}^{n} e^{-\lambda(n-i)} x_i}{\sum_{i=1}^{n} e^{-\lambda(n-i)}} \tag{1}$$

where $x_i = 1$ if $s_i = s_{n+1}$ and $x_i = 0$ otherwise. This formula can also be written recursively as

$$P(s_{n+1}) = \alpha s_n + (1-\alpha) P(s_n) \tag{2}$$

where α is a constant depending on λ. This form highlights the 0th-order model's equivalence to an exponential filter. In fact, it is simply an exponentially weighted moving average.

The first-order model is effectively a first-order Markov Chain. The probability of a pair of sequence elements factorizes into

$$P(s_n, s_{n+1}) = P(s_{n+1}|s_n) P(s_n) \tag{3}$$

and so the predictive probability for the next element in the sequence is simply

$$P(s_{n+1}|s_n) = \frac{P(s_n, s_{n+1})}{P(s_n)} \tag{4}$$

where the probability of a given n-gram depends on an exponential weighing of past data, and is given by

$$P(s_{n-j}, ..., s_n) = \frac{\sum_{i=1}^{n} e^{-\lambda(n-i)} x_i}{\sum_{i=1}^{n} e^{-\lambda(n-i)}} \tag{5}$$

where $x_i = 1$ if $\{s_{i-j}, ..., s_i\} = \{s_{n-j}, ..., s_n\}$ and $x_i = 0$ otherwise.

We now seek to fit this model to human performance on standard 2AFC tasks as well as a novel 3AFC task.

Method

We performed three experiments. First, in order to validate the procedure, we used a similar 2AFC task to Cho et al. (2002). Second, because neither the two-fingered response nor the two elements used in Cho et al. (2002) generalize well to a three-alternative task, we conducted a different 2AFC task with one finger and a different set of elements. The third experiment was a three-alternative task with the same sequence elements as the second 2AFC task plus a third element. This sequence of experiments was necessary to establish that any differences between a standard 2AFC task and our novel 3AFC task were due to the presence of an additional choice, rather than superficial differences in the task.

Participants

Five subjects (four female, one male) participated in Experiment 1, five more (four female, one male) in Experiment 2, and an additional seven (six female, one male) in Experiment 3. All participants were volunteers recruited from the University of Adelaide and surrounding community. Four subjects were not naive to the purpose of the experiment and two were left-handed. One subject participated in the first and second experiments and another in the second an third experiments, so the total number of subjects used was 16 (sixteen). All participants gave their informed consent to participating in the experiment and had normal or corrected-to-normal eyesight.

Stimuli

Experiment 1 used a similar task to Cho et al. (2002), which used a tall and short O as the two elements. This presents a problem when seeking to extend this paradigm to more than two elements. Since a third element would have to be different in size, this might introduce an preference for an increasing or decreasing sequence. For this reason, the stimuli used in Experiments 2 and 3 were abstract figures: a square, a triangle, and a circle. The first two were used in Experiment 2 and all three were used in Experiment 3.

Procedure

In all three experiments, subjects sat approximately 60cm away from the computer screen, inside a darkened room. The stimuli were white and displayed against a grey background, approximately 3 cm high, and displayed sequentially in the same position in middle of the screen, using Psychophysics Toolbox 3 and Matlab r2008a on a 15" Macintosh MacBook Pro running MacOSX 10.6. Responses were effected on a Cedrus RT-530 response time box, which has one central round button surrounded by four rectangular buttons. The RT box was placed to the right of the computer where stimuli were displayed if the subject was right-handed, and to the left if left-handed.

In Experiment 1, following Cho et al. (2002), experiments used two fingers, one placed on the left button and one on the right button of the response box. They were instructed to respond as quickly and accurately as possible to the stimulus by pressing the button corresponding to the stimulus shown (left - small O, right - large O). After pressing the button, the stimulus disappeared and after a Response Stimulus Interval (RSI) of 800ms, another appeared. The only feedback was a beep whenever a response button was pressed. This paradigm precisely replicates Cho et al. (2002), the only differences being that our stimuli were presented on a dark gray (rather than black) background and that our measurements were taken with a response box, allowing for near-millisecond precision.

In Experiments 2 and 3, participants used only one finger for all responses. This was necessary because, in a 3AFC task, the use of more than two fingers can lead to a preference of left to right and left to right sequences, as any person who has tried to tap a table will recognise. In general, spatial mapping of stimuli to response should be avoided as much as possible, as this has been shown to have an effect in RT patterns Soetens, Boer, and Hueting (1985). Although Experiment 2 was a 2AFC task, it was made as methodologically similar to Experiment 3 as possible in order to ensure that any differences in performance were due to the addition of another element rather than extraneous characteristics of the task. In these experiments, subjects kept the middle button of the RT box depressed at stimulus onset, at which point one of the elements was displayed on the screen. They responded by moving the finger from the middle button to the rectangular button corresponding to the element shown (left - triangle, right-square, and in Experiment 3, top-circle). The stimulus disappeared after the button was pressed and participants returned their finger to the middle button. 800ms after pressing the rectangular button, the next stimulus appeared. The time between stimulus onset and middle button release and between middle button release and side button press were recorded. Feedback consisted of a high pitch beep if everything was alright and one low in pitch as a warning in case the subject forgot to return his/her finger to the middle position.

All experiments consisted of 13 blocks of 120 trials each, with a small break between each block and a longer break (around 10 min) after the seventh block. Subjects in Experiment 1 and 2 were given one block of training before beginning, while those in Experiment 3 received two. The data from these training blocks was not used in the analysis. Sequences were generated randomly for each subject, with each element sampled from a uniform distribution over the elements. The absolute frequencies of all the elements were equal within each block and so for the whole experiment.

Results

It is assumed throughout the model analysis and fits that RTs are inversely proportional to the predictive probability of the next element in the sequence, a fairly common assumption in the literature (Cho et al., 2002; Yu & Cohen, 2008; Wilder et al., 2010). This means that the higher the probability of the next sequence element, the lower the reaction time should be. Model fits were obtained by minimizing the sum of squared deviations between the models and the datasets by varying three parameters: a, b and λ. a and b are the parameters of a linear transformation of the form $a+bx$ and λ is the exponential rate of the model.

All the model results are described in terms of $(1-P(s_{n+1}))$, where $P(s_{n+1})$ is the probability of the last sequence element in each of the possible five long sequences. A 0th-order model, whose predictions are shown in Figure 1, can effectively only detect repetitions, so in a sequence such as ABAB_ it will assign a higher probability to seeing a B in the last position ($P(B)$ is higher than $P(A)$), due to exponential discounting weighing recent events more. A first-order order model can detect alternations as well as repetitions, correctly predicting an A in the previous sequence ($P(A|B)$ is higher than $P(B|B)$).

In the absence of exponential discounting (i.e., with perfect memory) a zeroth-order model would eventually predict that a sequence like ABABA would be as probable as ABABB (since it would learn that $P(A) = P(B) = 0.5$). However, because of such discounting, the predictions are different for sequences with the same number of each of the elements, meaning that even a zeroth-order model is sensitive to order in the sequence. Figure 1 shows the predictions for a zeroth-order model with $\lambda = 0.22$.

Experiment 1

Our results are shown in Figure 2. Participant performance in our experiment, as in Cho et al. (2002), is best explained by a 1st-order model. There are a few points of divergence, most notably at sequence AAABB, but overall the model fits the data well. It is evident that subjects were tracking first-order transition probabilities, paying attention to alternations and repetitions as opposed to just relative frequencies of events.

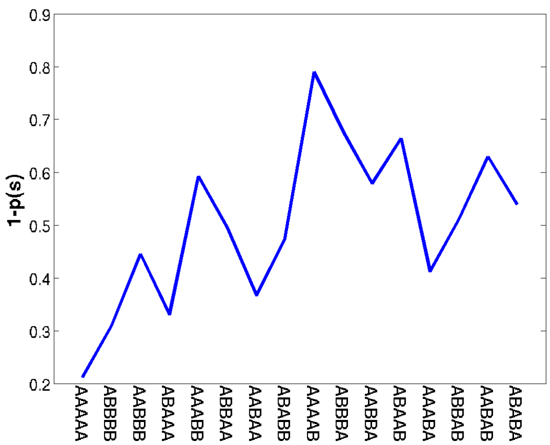

Figure 1: Expected RT pattern ($1 - P(s_n)$) for a 0th-order model. In this case $\lambda = 0.22$. The sequences in the x-axis should be read from left to right. The probabilities indicated correspond to the predictive probability of seeing the last element in the sequence give the last four.

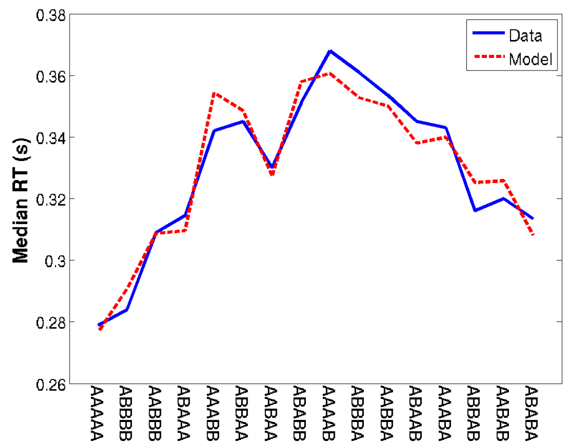

Figure 2: RT data and best model fit for Experiment 1 (the Cho et et replication experiment). $a = 0.262$, $b = 0.105$, $\lambda = 0.35$.

This can be shown quantitatively by calculating the log likelihood values for the model fits. As Table 1 demonstrates, a first-order model had a higher log likelihood, and so was the preferred model.

Experiment 2

As in Experiment 1, the reaction time pattern for Experiment 2 was well-captured by a 1st-order model, as is shown in Figure 3 and Table 1. This suggests that subjects were mostly using transition probabilities in order to predict the next element in the sequence. However, the fit is not perfect. In particular, the peak of people's RT pattern is not at AAAAB, corresponding to a violation of a perfectly repetitive pattern (as predicted by the model); rather, it is at ABABB, corresponding to the violation of a perfectly alternating pattern. The reason for this is clear upon examination of the individual-level data: several subjects displayed faster reaction times to alternations (right side of the RT pattern) as compared to repetitions, something that never happened in the simple 2AFC using two fingers. It appears that the new task involving one finger returning to a central position induces a preference for alternations in at least some subjects. For the sake of simplicity we did not include free parameters to capture this bias (as per Yu & Cohen 2008 and Wilder at al 2010), though this would not be difficult to do.

A key point is that the sequence ABABA is one of the most defining in distinguishing between a first and second order pattern: a first order model will give a much high probability of seeing an A after ABAB, whereas a zeroth order model will in fact give a slightly higher probability of seeing a B, given that it is more recent in the sequence. Clearly, participants are sensitive to first order regularities in this task.

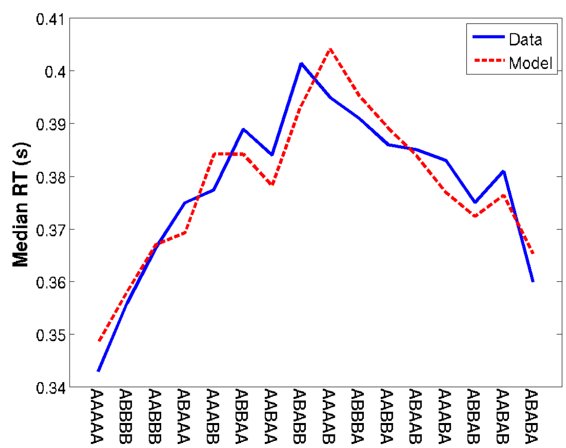

Figure 3: RT data and best model fit for Experiment 2 (the 2AFC using one finger). $a = 0.336$, $b = 0.095$, $\lambda = 0.2639$

Experiment 3

The sequence data for the 2AFC tasks contained $2^5 = 32$ possible sequences of length five, which were grouped into 16 categories, each holding two equivalent sequences, e.g AABBB = $\{00111, 11000\}$. The data from the 3AFC tasks however contains $3^5 = 243$ sequences of length five. We grouped them into 41 *equivalence classes*, each containing 6 equivalent sequences corresponding to all possible combinations of 3 elements. For instance, ABBCC = $\{01122, 02211, 10022, 12200, 20011, 21100\}$. The single exception to this is AAAAA = $\{00000, 11111, 22222\}$. This way of organising the data has the extra advantage of eliminating the need to randomise the mapping of symbols to response keys used or indeed to worry about systematic differences in RTs for different symbols and/or buttons: each data point is a median of all six possible permutations of symbols,

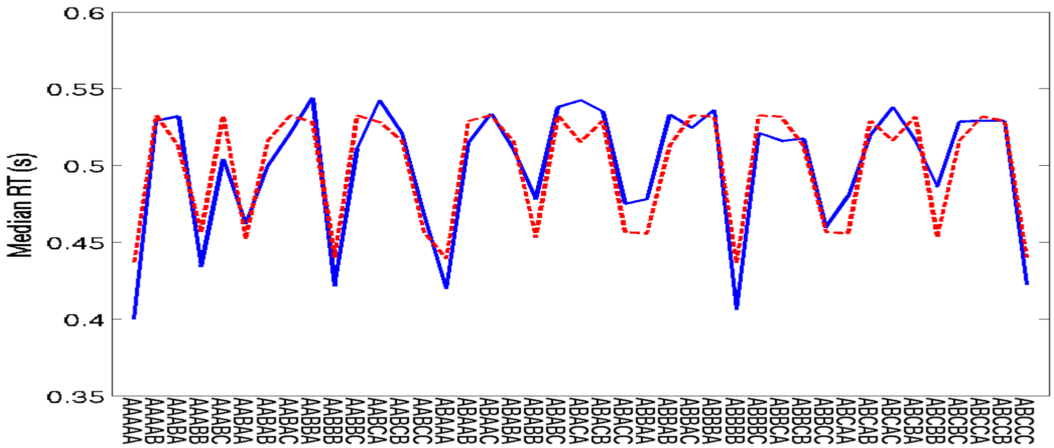

Figure 4: RT data and best model fit for Experiment 3 (the 3AFC using one finger). $a = 0.435$, $b = 0.096$, $\lambda = 1.53$

Experiment	0th order model	1st order model
Experiment 1	-0.275	**-0.043**
Experiment 2	-0.180	**-0.038**
Experiment 3	**-0.569**	-1.24

Table 1: log likelihood values for 0th-order and 1st-order models (higher values are in bold) on all three datasets estimated by assuming that the data is normally distributed. Experiment 1: classical 2AFC; Experiments 2 and 3: new paradigm 2AFC and 3AFC

which also implies a permutation of response buttons.

It is clear from Figure 4 and Table 1 that most of the variation in people's responses in our 3AFC task is explained by a 0th-order model but that no appreciable fit could be obtained from a 1st-order model. This suggests that the subjects tracked element frequency (0th-order transition probabilities) only. It seems then that, in this case, subjects were using absolute frequencies, rather than transition probabilities, in order to predict the next element in the sequence.

Discussion

In this work we introduced and validated experimental paradigm with a sequential task with more than two alternatives, and developed a model capable of predicting reaction times in such a task (as well as the simpler 2AFC tasks that already exist). Our model obtained a good fit to the median RT from both types of tasks. This work suggest that humans track and use first-order transition probabilities in a 2AFC task but only zeroth-order frequencies in the case of a 3AFC task.

Arguably not all the trends in the ternary task data are captured by the model. This could be for a number of reasons. On the experimental side it is worth noting that in a ternary task the probability or occurrence of each individual sequence is lower that in a binary task, and even when grouped into equivalence classes, each median RT reported will be a median of fewer data points, approximately one third on average. To obtain the same number of data points for each equivalence class, one would effectively have to multiply the number of subjects by 3. Since only 8 subjects were used, as opposed to 15, it is possible that there is some added noise in the RT data. Assuming that the experimental RT data is sound, further research is necessary in order to explain the observed differences. Nonetheless, this is to our knowledge the first time that a model has shown such a good fit to a ternary task. These results also clearly indicate that sequential effects persist beyond binary tasks, further supporting the findings of Ma-Wyatt and Navarro (2009) which suggest that sequential structure can influence more complex tasks than just 2AFCs.

In need of explanation is the apparent change from tracking 1st order transition probabilities to just absolute frequencies as the task changed from binary to ternary. It is possible that the increased difficulty in tracking the necessary nine 1st order transition probabilities in the ternary task put too much load on working memory, and this forces a regression towards just tracking absolute frequencies. Interestingly, λ was also much higher in the ternary task than in the binary, indicating that not only were the subjects tracking the relative frequencies of events but they were integrating over a much narrower window – in other words, they were not looking as far out into the past.

It is noteworthy that there were significant individual differences in individual RT patterns. As an example, one of our subjects in Experiment 1 (not shown) showed a near perfect fit to a zero-th order reaction time pattern, indicating that although the overall pattern of the task was clearly first order, it was made up of a mix of significantly different patterns. While beyond the scope of the present article, these individual differences are also in need of explanation.

While not in contradiction with previous models, our model has the virtue of highlighting that exponential filtering is the key to understanding sequential effects. Additionally, it handles multiple options and different order statistics at the same time, which none of the previous models did. It does fall short of allowing probabilistic inference as a complete statistical model would; in particular, it does not allow for model comparison. It also fails to capture the capture the full RT distributions as a more complete model would. Ongoing work in our lab is dedicated to addressing these issues.

Future directions

From an empirical point of view, it would be interesting to investigate what kind of binary task would produce an overall 0th order RT pattern. Conversely, it would be interesting to see if there is any kind of ternary task that would produce a first order pattern. While not able to confirm it quantitatively at present, we cannot help but note the striking similarity of the RT pattern predicted by the 0th order model in Figure 1 to Jones, Cho, Nystrom, and Cohen (2002), which was a one-to-many multi-alternative forced choice task (which would out of necessity imply a need to track many transition probabilities).

It would also be interesting to investigate why our experimental paradigm, which uses the same finger moving from a central position to different locations around it, seems to create a preference for alternations in some subjects. More generally, a rigorous investigation of the underlying causes of changes in preference for alternations and repetitions is necessary. In the past, they have been explained in terms of subjective expectancy and automatic facilitation (Soetens et al., 1985), and shown to depend on the RSI. However, this did not take into account what we now know, that a difference between RTs to repetitions and alternations can arise naturally from a simple statistical model (something Wilder et al. (2010) have also noted). It is tempting to think that a more complete statistical model can perhaps also account for the observed differences.

Conclusion

There are two key findings in this work:

1. It was demonstrated that a simple model including exponential filtering of past events can account for most of the RT variability in a 3AFC.

2. First order tracking seems to predominate in a 2AFC but this situation is inverted in the case of a 3AFC. This is the first demonstration that humans can change the statistics used to track a sequence depending on the task constraints.

Sequential effects demonstrate how humans can use the statistical regularities in a sequence and use them over short time scales. This work contributes to the literature attempting to understand how humans can achieve this, specifically in the case of tasks more complex than a simple 2AFC. Sequential effects are also intimately related to human pattern recognition, in the sense that they occur due to an attempt at optimally capturing the statistical nature of sequences, and in this sense are one of the simplest ways to study the way humans detect patterns over time. This work extends previous results analysing binary sequences to more complex ones such as ternary sequences. In addition, we show that humans can use different statistics in order to perform this sequence analysis depending on the task.

Acknowledgments

DJN was supported by an Australian Research Fellowship (ARC grant DP0773794)

References

Cho, R. Y., Nystrom, L. E., Brown, E. T., Jones, A. D., Braver, T. S., Holmes, P. J., et al. (2002). Mechanisms underlying dependencies of performance on stimulus history in a two-alternative forced-choice task. *Cognitive, Affective, & Behavioral Neuroscience*, *4*(2), 283-299.

Jones, A. D., Cho, R. Y., Nystrom, L. E., & Cohen, J. D. (2002). A computational model of anterior cingulate function in speeded response tasks: Effects of frequency, sequence, and conflict. *Cognitive, Affective, & Behavioral Neuroscience*, *4*(2), 300-317.

Luce, R. D. (1986). *Response times: their role in inferring elementary mental organization*. New York: Oxford University Press.

Manning, C. D., & Schütze, H. (1999). *Foundations of statistical natural language processing*. Cambridge, Mass.: MIT Press.

Ma-Wyatt, A., & Navarro, D. J. (2009). Exploiting sequential structure to improve visuomotor control. *Proceedings of the 31st Annual Conference of the Cognitive Science Society*, 1424-1429.

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.

Soetens, E., Boer, L., & Hueting, J. (1985). Expectancy or automatic facilitation? separating sequential effects in two-choice reaction time. *Journal of Experimental Psychology: Human Perception and Performance*, *11*(5), 598-616.

Wilder, M. J., Jones, M., & Mozer, M. C. (2010). Sequential effects reflect parallel learning of multiple environmental regularities. *Advances in Neural Information Processing Systems*.

Yu, A., & Cohen, J. (2008). Sequential effects: Superstition or rational behavior? *Advances in Neural Information Processing Systems*.