MODELLING JUDGEMENT OF SEQUENTIALLY PRESENTED CATEGORIES USING WEIGHTING AND SAMPLING WITHOUT REPLACEMENT

Petko Kusev - Kingston University London and City University London

Krasimira Tsaneva-Atanasova - University of Bristol

Paul van Schaik - Teesside University

Nick Chater - Warwick Business School

Correspondence concerning this article should be addressed to Petko Kusev,

Department of Psychology,

City University London

EC1V 0HB, United Kingdom

E-mail: p.kusev@city.ac.uk

Phone: +44(0)20 7040 8590

Fax: +44(0)20 7040 8580

Abstract

A series of experiments by Kusev, Ayton, van Schaik, Tsaneva-Atanasova, Stewart and Chater (2011) studied relative-frequency judgment of items drawn from two distinct categories. The experiments showed that judged frequencies of categories of sequentially encountered stimuli are affected by properties of the experienced sequences. Specifically, a first-run effect was observed, whereby people overestimated the frequency of a given category when that category was the first repeated category to occur in the sequence. Here we (1) interpret these findings as reflecting the operation of a judgment heuristic sensitive to sequential patterns, (2) present mathematical definitions of sequences used in Kusev et al. (2011) and (3) present a mathematical formalization of the first-run effect - the Judgments Relative to Patterns model (JRP) - to account for the judged frequencies of sequentially encountered stimuli. The model parameter (*w*) accounts for the effect of length of the first run on frequency estimates, given the total sequence length. We fitted data from Kusev et al. (2011) to the model parameter, where with increasing value of *w* subsequent items in the first run have less influence on judgment. We see the role of the model as essential for advancing knowledge in the psychology of judgments and other disciplines such as computing sciences, cognitive neuroscience, artificial intelligence and human-computer interaction.

Keywords: temporal-sequence patterns, frequency judgments, first-run effect

Many events in life occur in temporal sequence - for example sunny and rainy days. A long history of research has investigated memory for - and judgment of - the *frequency* of events encountered in temporal sequence (Brown, 1997; Hasher & Zacks, 1979, 1984). Several prominent cognitive theories have also analyzed how people reason about the processes underlying sequences and how they anticipate individual events in a sequence (Ayton & Fischer, 2004; Oskarsson, Van Boven, McClelland & Hastie, 2009; Kusev, van Schaik, Ayton, Dent & Chater, 2009; Kusev et al., 2011; Sedlmeier & Betsch, 2002). Accordingly, in this paper we attempt to offer a theoretical formalization to account for judgments of stimuli experienced in temporal sequences.

One possible source of information about the relative frequency of the elements in a sequence is the presence of *runs* - repetitions of types of stimulus in a sequence[1] (Kusev et al., 2011). For example, across a wide range of sequences varying in the relative frequency of their elements, one would be entitled to assume that, when one encountered a run of one type of stimulus, that stimulus type was likely to be more preponderant in the sequence[2]. Consistent with this notion, Kusev et al. (2011) have identified a *first-run effect* whereby, after experiencing a sequence of stimuli, people give higher estimates to the frequency of a given category of event when that category is the first *repeated* category to occur in the sequence. For example in sequence (1) below, **'O'** is the first repeated *type* of stimulus and, although the 'O's and 'X's are equally frequent, 'O' is judged to be more frequent.

(1) **XOXOOOOXXXOOOOOXXXXOOXXXOXOX**

It is important to note that, although the first run had a biasing effect on frequency judgment, the last run did not. Furthermore, the categories investigated by Kusev et al. (2011) were chosen to be abstract (checkerboard patterns, geometrical shapes or sine-wave tones) in Experiments 1-5 or very common (in Experiment 6) so that previous experience with the categories (e.g. in terms of familiarity or intuitiveness) would not affect people's frequency judgment within each experiment. Accordingly, the finding of a first-run effect by Kusev et al. (2011) is in agreement with the results of other research on the perception of randomness of stimuli (e.g., the 'hot-hand fallacy', Ayton & Fischer, 2004; Gilovich, Vallone & Tversky, 1985), where a bias in judgment occurred, in favor of the category that was repeated in a sequence.

The study of memory flashes out that human judgments are informed by consulting memory for the individual items or by their temporal order. For example, extensive research indicates that, with a list of objects, participants are likely to remember items at the beginning (Anderson, 1965; Asch, 1946), and end (Miller & Campbell, 1959) - *primacy* and *recency* effects. In contrast, the results in Kusev et al. (2011) demonstrate that judgment of the frequency of types of items in a sequence and a respondent's recall of the number of individual items of each type in the sequence are dissociated. This finding is not anticipated by theories which predict that frequency is assessed according to the ease with which individual instances can be brought to mind (e.g., Tversky & Kahneman, 1973).

We attribute this phenomenon to a simple memory heuristic. For example, across a wide range of sequences varying in the likelihood of events, respondents assume that a sequence with a first run of a particular event is more likely to be preponderantly comprised of those items. The rationale for this finding is based on the assumption that, in making frequency judgments, people are constrained by information-processing limitations and hence have a propensity to avoid cognitive load. According to Simon (1956) one way for people to achieve this is to simplify and, use satisficing strategies - rather than attempt to use optimal or normative strategies. We identified the first-run effect as one such possible strategy.

Stimuli and Method

The experiments in Kusev et al. (2011) were designed to investigate whether judgments of the frequencies of stimulus categories in a presented sequence depend on the pattern of events within the sequence. The first experiment used sequences comprising two categories of stimuli occurring in equal proportion, but randomized in their order. The method allowed us to investigate how the varying characteristics of the sequences affect people's estimation of the relative frequency of the two categories. Specifically, the experiments examined a variety of sequence characteristics (such as the number of runs, the length of the first run, the length of the last run and the length of the sequence) to explore whether they influenced the estimated frequency of occurrence of the category of stimuli. For each individual participant a sequence was generated by randomly sampling stimuli without replacement from the set of stimuli comprising each category; therefore, each individual stimulus appeared just once in the sequence. Stimuli (geometrical figures and tones varied in pitch) were presented to each participant. A computer program for generating and presenting the stimuli was employed. The respective probabilities of a category to be the first repeated category to occur in a sequence of a given length, in other words a 'first run' with length 2, 3, and i are given by Equations (2), (3), and (4) (examples are presented in Table 1).

Each respondent was presented with one random sequence of stimuli from two categories (in a pre-defined proportion). Participants were instructed that they should try to remember as much as possible about the stimuli and informed that they would be viewing checkerboard patterns, geometrical figures or listening to tones. Respondents were presented with one sequence of stimuli and then, immediately after the sequence had been presented, they were explicitly asked, via a visual message on the computer screen, to make one judgment of the frequency (as a percentage) of one of the stimulus categories experienced in the sequence.

During the presentation of the experimental trials, respondents were not required to make any explicit judgments of the stimuli; presentation of each stimulus was self-paced via the computer keyboard and there was no limit on how much time participants could spend observing each stimulus. The next stimulus appeared without delay after a participant's key press.

Judgment Relative to Patterns

In the interests of clarity we offer mathematical definitions of the sequences used in this article (see Table 1) and a formalization of the first-run effect in the Judgment Relative to Patterns (JRP) model. Specifically, we found that judgments of frequency are informed by the apprehension of patterns. In particular, after experiencing a sequence of stimuli, people give higher estimates to the frequency of a particular category of event when that category is the first repeated category to occur in the sequence (the first-run effect). The model accommodates the first-run effect by assuming that the judgment of the frequency of a given category of items appearing within the first run is directly influenced by the sequence pattern. Accordingly, the JRP model estimates the probability of the first run occurring by assuming that the elements in sequences are sampled without replacement from a finite sample equal to the sequence length; thus, the probability of a run of one stimulus category continuing diminishes the more items from that category appear. Empirically Kusev et al. (2011) found no effect of the length of the first run on frequency judgments, so the model also weights - in a decreasing manner - each probability associated with consecutive items in the first run such that the weight of each item decreases as the number of category repetitions increases. Defined in this way, the judged frequency of the first-run category produced by the JRP model is always greater than, or at least equal to, the actual pre-defined proportion.

Derivation of the model

Frequency coding is considered one of the most common coding strategies employed by neural systems (Gerstner, Kreiter, Markram & Herz, 1997). In our model what is essential to the process of the judgment of frequencies is pattern. Below we present a formal derivation of the model equations. Let the total number of items be denoted by N; l = the length of the first run, or in other words, the number of items within the first run; n = the number of items in the first-run category (the number of items in the non first-run category is then given by $N - n$); FJ_l denotes frequency-judgment estimation in the JRP model. Furthermore, let $A_n^{(1)}$ denote the event of first appearance of an item from category A in the first run, and $A_n^{(2)}$ = the second appearance of an item from the same category A in the first run. Similarly, we define the event $A_n^{(i)}$ as the *i-th* appearance of an item from category A in the first run. Thus, the probability of a category-A item appearing for the first time within the first run in position x is

$$P_{n_x}^{(1)} = \Pr\{A_{n_x}^{(1)}\} = \frac{n_x}{N_x},$$

where n_x is the number of remaining items of the first-run category before the first run and N_x is the total remaining items (of the first-run category and the non-first-run category) before the first run.

In order to illustrate our idea we consider an example where in the sequence no item of the first-run category appears before the first run. In this case the above expression simplifies to

$$P_{n_1}^{(1)} = \Pr\{A_{n_1}^{(1)}\} = \frac{n_1}{N_1}, \tag{1}$$

where $n_1 = n$ and $N_1 = N$. The probability of a first run of length 2 is the probability that the second item in the first run is from the same category as the first item.

$$P_{n_1}^{(2)} = \Pr\{A_{n_1}^{(2)} \mid A_{n_1}^{(1)}\} = \frac{n_1(n_1-1)}{N_1(N_1-1)}. \tag{2}$$

By analogy, the probability of a category A item appearing in Position 3 in the first run is then:

$$P_{n_1}^{(3)} = \Pr\{A_{n_1}^{(3)} \mid A_{n_1}^{(1)}, A_{n_1}^{(2)}\} = \frac{n_1(n_1-1)(n_1-2)}{N_1(N_1-1)(N_1-2)}. \tag{3}$$

As a generalization, the probability of a category A item appearing in Position i is:

$$P_{n_1}^{(i)} = \Pr\{A_{n_1}^{(i)} \mid A_{n_1}^{(1)}, A_{n_1}^{(2)}, \ldots, A_{n_1}^{(i-1)}\} = \frac{n_1(n_1-1)(n_1-2)\ldots(n_1-i+1)}{N_1(N_1-1)(N_1-2)\ldots(N_1-i+1)}. \tag{4}$$

Note that the above equations represent the probabilities of having a repeated sequence of items from category A starting at Position 1, that is in other words to have a first run of this category. If $x > 1$ then the numerator $n_x = n - (x-1)/2$ if x is uneven and $n_x = n - (x-2)/2$ if x is even, and the denominator $N_x = N - (x-1)$. In addition, the length of the first-run, l, has to be less or equal to the total number of items in the series, N, minus the starting position, x, (i.e., $l \leq N - 1$). This constraint is necessary in order to guarantee applicability of $P_{n_x}^{(i)}$ to the special case when the first-run appears at the very end of the test sequence. What is important is that this is the first appearance of a repeated sequence (pattern) in the experiment. The above can be generalized using the gamma function $\Gamma(\alpha) = (\alpha - 1)!$ as follows. By rewriting Equation (4) and using the n_x instead of n (which is a special case of n_x), we arrive at the probability of having *i* repeated consecutive same-category items (out of a total of *n* such items), $P_{n_x}^{(i)}$, for the first time in a total sequence of items with length *N*:

$$P_{n_x}^{(i)} = \frac{\Gamma(n_x+1)\Gamma(N_x+1-i)}{\Gamma(n_x+1-i)\Gamma(N_x+1)}. \tag{5}$$

where $1 \le i \le l$. Given the independence of the events ($A_{n_x}^{(i)} \mid A_{n_x}^{(1)}, A_{n_x}^{(2)}, \ldots, A_{n_x}^{(i-1)}$), the probability that either of these events occurred is the sum of each probability,

$$\sum_{i=1}^{p} P_{n_x}^{(i)}.$$

Results and discussion

The JRP model formally estimates the relative frequency judgment of one of the categories as a function of the probability of the response category to appear in a repeated sequence of arbitrary length. We see the role of the model as essential for advancing knowledge in the psychology of frequency estimation - it provides transferability of psychological knowledge to related disciplines such as computing sciences, artificial intelligence and human-computer interaction. Our work, formalized in the JRP model, contributes to the psychology of frequency estimation by highlighting and modeling the role of sequential patterns in stimuli in this estimation. Thus, these patterns need to be accounted for, in addition to the role of stimulus characteristics, in research on frequency estimation. In effect, our work exemplifies the ubiquity of sequence effects that have been exposed in other areas of research, such as psychophysical judgment. There, it is claimed that, due to sequence effects, none of the psychophysical laws, such as Weber's, are general (Lockhead, 2004).

The model accommodates the first-run effect by assuming that the judgment of the frequency of a given category of items appearing within the first run is directly determined by the sequence pattern. Accordingly, FJ_l assumes that in addition to summing $P_{n_x}^{(i)}$ we also weigh in a decreasing manner each probability associated with consecutive items from the first-run. In other words, we assign a weight ($\frac{1}{i^w}$) to each probability $P_{n_x}^{(i)}$, for $1 \leq i \leq l$ and $w \geq 1$. Note that chosen in this way the weight of each probability $P_{n_x}^{(i)}$ decreases as i increases. This assumption accounts for the discriminability properties of JRP, that is as the position of an item within the first-run increases, its influence on the judgment decreases, because the weight of any consecutive item in the sum (6) below decreases as fast as $\frac{1}{i^w} \to 0$. In other words, as far as their importance for (or contribution to) JRP is concerned, the subsequent members/items within the first-run have an increasingly smaller weighting.

Thus, FJ_l takes the following form:

$$FJ_l = 100 \times \sum_{i=1}^{l} \frac{1}{i^w} P_{n_x}^{(i)}. \tag{6}$$

where $\frac{1}{i^w}$ is the weight of the i^{th} item in the first-run pattern. Note that defined in this way, FJ_l is always greater than the actual percentage of the judged category, which is given by the ratio $100 \times n_x/N_x$ as shown in Figure 1. For example *the value of* FJ_l predicted by the model (6) will be always greater or equal to 50 (corresponding with a frequency estimate of 50%) for an experimental sequence of items with two equally represented categories (Figure 1).

In Figure 1 we depict the dependence of FJ (as given in Equation (6)) on the weights (w) and on the length of the first run (l) represented as a two-dimensional surface in the space spanned by (w, l, FJ). Figure 1 clearly demonstrates the frequency judgment approaches the ratio n_x/N_x with increase in w and levels off to a constant with increase in l.

In Kusev et al.'s (2011) Experiment 1, where two stimulus categories occurred with equal frequency, 47 of 78 participants judged the frequency of the first run category > 50%, consistent with the first-run effect. Another 23 correctly produced a frequency judgment/estimate of 50% and a further 8 produced a frequency judgment < 50%. In our model, the frequency of the initial repeated items in the first run make a larger contribution than later items to the value of FJ_l as described by Equation (6). The stimulus number within the first run is weighted by the parameter w. With increasing the value of w, subsequent items in the first-run length have less influence on judgment. Figure 1 depicts the relationship between the weight, w, the length of the first run, l, and frequency judgment in the model.

In model-fitting, the data were analyzed of participants (N = 47) whose response was a frequency estimate > 50%. For each participant's frequency estimate, the model parameter was estimated, using the Wolfram Mathematica 8 platform**.** For each of 47 jack-knife samples, taking into account the starting position of the first run x, the average model parameter of the jack-knife sample was then used to calculate the error in predicting the remaining participant's frequency estimate. The mean value (*SD*) of model parameter was 1.73 (.45), CI(*M*).95 = [1.60; 1.86], indicating high precision of the estimate. Therefore, the weight of subsequent Positions i in the first-run category in the modeled frequency estimate FJ_p is, on average, reduced by a factor of $i^{1.73}$. Within the jack-knife samples, model fit was excellent with maximum error < 10^{-11}. Regarding the remaining cases in the jack-knife samples, actual and predicted frequency judgments were substantially and significantly aligned, intraclass correlation coefficient = .53, F (46, 46) = 3.21, $p < .001$. Relative prediction error was relatively small, in the order of 5%: for mean absolute error relative to actual frequency of the remaining case, mean value (*SD*) was .04 (.05) with *CI*(*M*) .95 = [.03; .05], and for mean absolute error relative to predicted frequency for the remaining case, mean value (*SD*) was .04 (.05) with *CI*(*M*) .95 = [.03; .06].

Conclusion

Our findings are consistent with the idea that people's frequency judgments are achieved in a similar fashion insofar as they are made without recollecting individual items in the sequence, but - instead - they are influenced by specific properties of the sequence configuration. In particular, we propose a simple strategy that draws minimal effort from our limited-capacity attentional mechanism, whereby respondents use the first run as a cue to frequency.

The basic finding of the first-run effect and its formalization in the current model could have implications for real-world phenomena. Therefore, applied research should investigate how the effect can account for judgment in different domains (e.g. the weather, the outcome of sport matches, and visual search in human-computer interaction).

Another important consideration is the difference between judgments from memory and those made immediately after stimulus presentation (e.g., Dickert & Slovic, 2009). This distinction could also apply to the judgment of sequences of stimuli, where judgment from memory for sequentially presented stimuli (as in Kusev et al., 2011) and judgment based on simultaneous presentation might involve different types of processing. It could be expected that with simultaneous presentation holistic processing (Hsiao & Cottrell, 2009) is more likely, thereby reducing the biasing influence of the first run on judgment.

Our work shows that the frequency estimate for a run of symbols will be strongly affected by order effects. A potentially important implication concerns fragment-based approaches to learning, motivated by associative learning theory. For example, Artificial Grammar Learning (AGL), is a widely employed paradigm for studying learning processes. In an AGL task, participants are typically exposed to training sequences in a first phase and subsequently asked to identify new sequences compatible with the old ones. Several theorists have proposed that participants can perform such tasks on the basis of knowledge derived from the statistical information about symbol co-occurrence in the training items (see for review Pothos, 2007, 2010). However, if such co-occurrence information is distorted by order effects, corresponding models would be in need of revision (cf., Pothos, 2007, 2010).

While other authors have proposed that frequency information is automatically encoded with minimal demand on attentional resources (Zacks & Hasher, 2002), our proposal does not address the issue as to whether the process underlying this strategy is automatic or controlled (‘System One’ or ‘System Two’), though - plainly - this is open to investigation. In sum, however, our research provides a specification and formalization of a process by which judgments of the frequency of types of event might be made, with implications for descriptive theories of identification, categorization and decision-making as well as their practical application.

References

Anderson, N. H. (1965). Primacy Effects in Personality Impression Formation Using a Generalized Order Effect Paradigm. *Journal of Personality and Social Psychology, 2*, 1-9.

Asch, S. E. (1946). Forming impressions of personality. *Journal of Abnormal Psychology, 41*, 258-290.

Ayton, P., & Fischer, I. (2004). The hot hand fallacy and the gambler's fallacy: two faces of subjective randomness? *Memory and Cognition, 32*, 1369-1378.

Brown, N. R. (1997). Context memory and the selection of frequency estimation strategies. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23*, 898-914.

Dickert, S. & Slovic, P. (2009). Attentional mechanisms in the generation of sympathy. *Judgment and Decision Making*, *4*, 297-306.

Gerstner, W., Kreiter, A. K., Markram, H., & Herz, A. V. (1997). Neural codes: firing rates and beyond. *Proceedings of the National Academy of Science, 94*, 12740-12741.

Gilovich, T., Vallone, R., & Tversky, A. (1985). The hot hand in basketball: On the misperception of random sequences. *Cognitive Psychology, 17*, 295–314.

Hasher, L., & Zacks, R. (1979). Automatic and effortful processes in memory. *Journal of Experimental Psychology: General, 108,* 356-388.

Hasher, L., & Zacks, R. T. (1984). Automatic processing of fundamental information: the case of frequency of occurrence. *American Psychologist, 39*, 1372-1388.

Hsiao, J. H. & Cottrell, G. W. (2009). Not all visual expertise is holistic, but it may be leftist: the case of Chinese character recognition. *Psychological Science*, *20*, 455-463.

Kusev, P., Ayton, P., van Schaik, P., Tsaneva-Atanasova, K., Stewart, N., & Chater, N. (2011). Judgments relative to patterns: How temporal sequence patterns affect judgments and memory. *Journal of Experimental Psychology: Human Perception and Performance, 37*, 1874-1886.

Kusev, P., van Schaik, P., Ayton, P., Dent, J., & Chater, N. (2009). Exaggerated risk: prospect theory and probability weighting in risky choice. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*, 1487-1505.

Lockhead, G. R. (2004). Absolute judgments are relative: a reinterpretation of some psychophysical ideas. Review of General Psychology, 8, 265–272. doi:10.1037/1089-2680.8.4.265

Miller, N., & Campbell, D. T. (1959). Recency and primacy in persuasion as a function of the timing of speeches and measurements. *Journal of Abnormal Psychology, 59*, 1-9.

Oskarsson, A. T., Van Boven, L., McClelland, G. H., & Hastie, R. (2009). What's next? Judging sequences of binary events. *Psychological Bulletin, 135*, 262-285.

Pothos, E. M. (2010). An entropy model for Artificial Grammar Learning. *Frontiers in Cognitive Science*, *1*, 1-13.

Pothos, E. M. (2007). Theories of Artificial Grammar Learning. *Psychological Bulletin*, *133*, 227-244.

Sedlmeier, P., & Betsch, T. (Eds.). (2002). *Frequency processing and cognition.* Oxford: Oxford University Press.

Simon, H. A. (1956). Rational choice and the structure of the environment.

Psychological Review, 63, 129-138.

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology, 5,* 207-232.

Zacks, R. T., & Hasher, L. (2002). Frequency processing: A twenty-five year perspective. In P. Sedlmeier & T. Betsch (Eds.), Frequency processing and cognition (pp. 21-36). Oxford, England: Oxford University Press.

Author Note

Petko Kusev, Paul van Schaik and Nick Chater are supported by Economic and Social Research Council Grant RES-000-22-1768. We also thank the Nuffield Foundation (50045SP) and The British Academy (SG 47881) for supporting Petko Kusev in his research. Correspondence concerning this article should be addressed to Petko Kusev, Department of Psychology, City University, London, EC1V 0HB, United Kingdom, e-mail: p.kusev@city.ac.uk.

Footnotes

[1] In this paper we use the term ‘run’ when at least *two* consecutive occurrences of the same category appear in a sequence; accordingly, a single occurrence is not considered a run (Kusev et al., 2011).

[2] Of course, for those sequences where the category with greater relative frequency is *not* signaled by the presence of a run, there will be bias.

Table 1

Probabilities of first runs of different lengths (calculated with the JRP model)

			First run of length 2 stimuli		First run of length 3 stimuli		First run of length 6 stimuli	
Number of stimuli in sequence	p(majority category): p(minority category)	Odds ratio	Majority category	Minority category	Majority category	Minority category	Majority category	Minority category
28	50:50	1.00	24	24	11	11	1	1
28	60:40	2.25	32	15	17	5	2	0
28	70:30	5.44	45	7	30	2	7	0
28	80:20	16.00	61	3	47	0	20	0
28	90:10	81.00	79	0	70	0	47	0
30	50:50	1.00	24	24	11	11	1	1
30	60:40	2.25	35	15	20	5	3	0
30	70:30	5.44	48	8	33	2	9	0
30	80:20	16.00	63	3	50	0	23	0
30	90:10	81.00	81	1	72	0	50	0
42	50:50	1.00	24	24	12	12	1	1
42	60:40	2.25	35	14	20	5	3	0
42	70:30	5.44	47	8	32	2	9	0
42	80:20	16.00	61	3	48	0	21	0
42	90:10	81.00	77	1	68	0	44	0

Figure 1

Dependence of the relative frequency judgment (FJ) on the weights (w) and on the length of the first run (l), where N=30, x=1 see Equations (4) and (6). (a) l(majority category): l(minority category)=50:50, n=15; (b) l(majority category): l(minority category)=60:40, n=18; l(majority category): l(minority category)=40:60, n=12.