TITLE OF RESEARCH PAPER

ai-Sang Iao, Jens Roeser, Mark Torrance, Karl Nicholls (All Nottingham Trent University)|N0965126@ntu.ac.uk

Introduction

Visual statistical learning is a form of implicit learning of the statistical regularities among adjacent and nonadjacent elements across time and/or space.

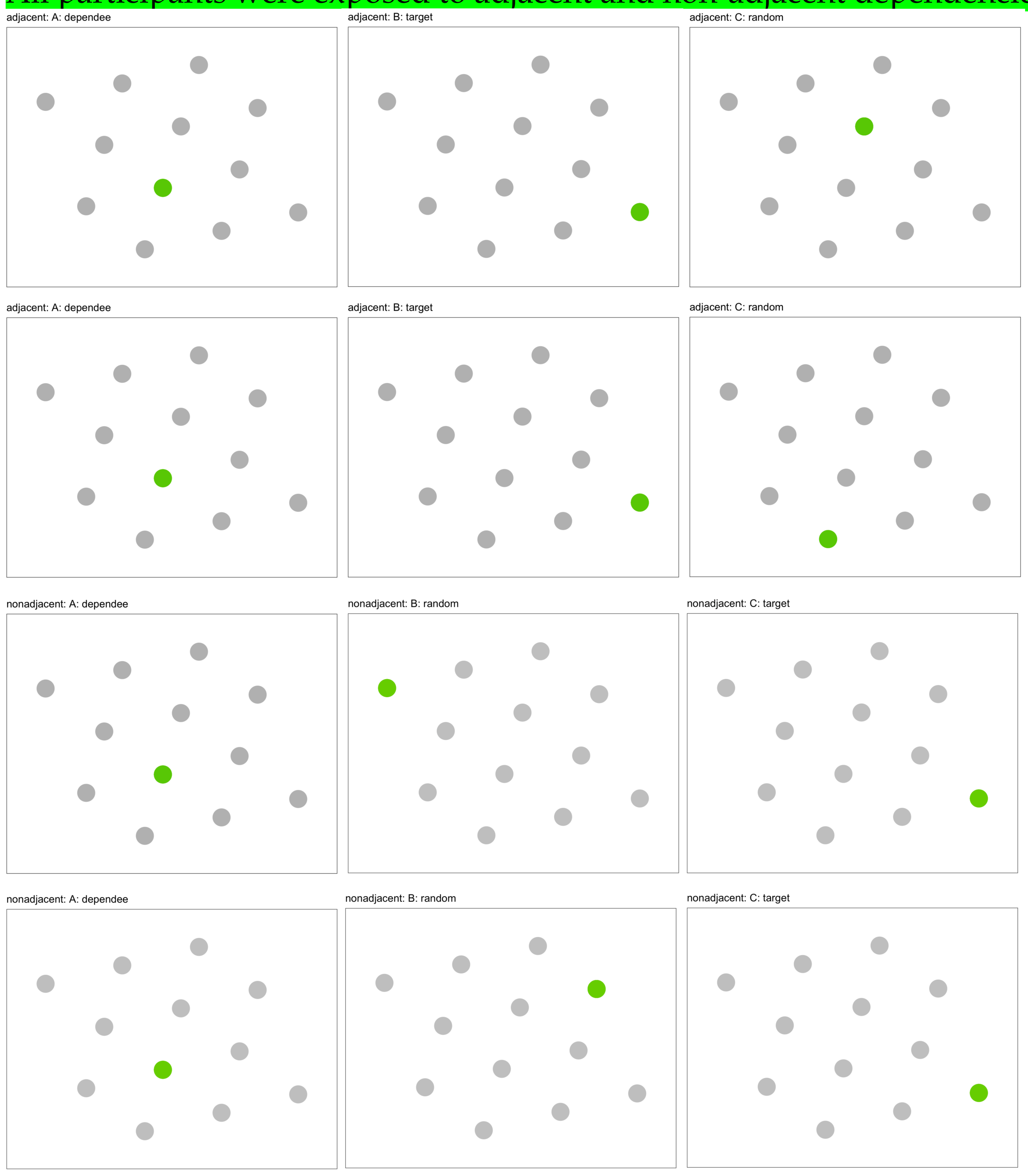
- Adjacent dependencies
  - E.g., syllable /pre/ is more likely to be followed by /ty/ than /on/
- Nonadjacent dependencies
  - E.g., morphosyntactic rules such as ***"is** X-**ing**"* (X is a verb)
- Both coexist in visual scenery, music and natural languages
  - E.g., She ***is*** go***ing*** to a party.

We use these regularities to make predictions which guide our behaviours. E.g., when we hear "she" we would predict something similar to "is", not "are", coming next.

This is what research has been focusing on – prediction based on adjacent dependency learning. Our study looks at prediction based on nonadjacent dependency learning as well as prediction of both adjacent and nonadjacent dependencies at the same time which is more ecological valid as stimuli in real life contain both dependencies.

Design

- All participants were exposed to adjacent and non-adjacent dependencies.



- Two blocks with four sequences of three elements, repeated 40 times per block
- Participants were either encouraged to predict the next target (directive instruction) or not (non-directive instructions).
- Adjacent and nonadjacent dependencies presented in each block concurrently or in separate blocks.
  - Experiment 1: non-directive instruction; dependencies not mixed (4 sequences of one dependency type per block / location set)
  - Experiment 2: directive instruction; dependencies not mixed (4 sequences of one dependency type per block / location set)
  - Experiment 3: directive instruction; dependencies mixed (2 sequences per dependency type in each block / location set)

Participants

- Experiment 1: We tested 32 participants (24 females, 7 males, 1 prefer not to say). The median age of the sample was 20 years (*SD* = 2.51) with an age range from 18 to 29 years.
- Experiment 2: We tested 32 participants (24 females, 8 males). The median age of the sample was 24 years (*SD* = 14.43) with an age range from 18 to 62 years.
- Experiment 3: We tested 32 participants (25 females, 7 males). The median age of the sample was 27 years (*SD* = 11.82) with an age range from 19 to 57 years.

Analysis method

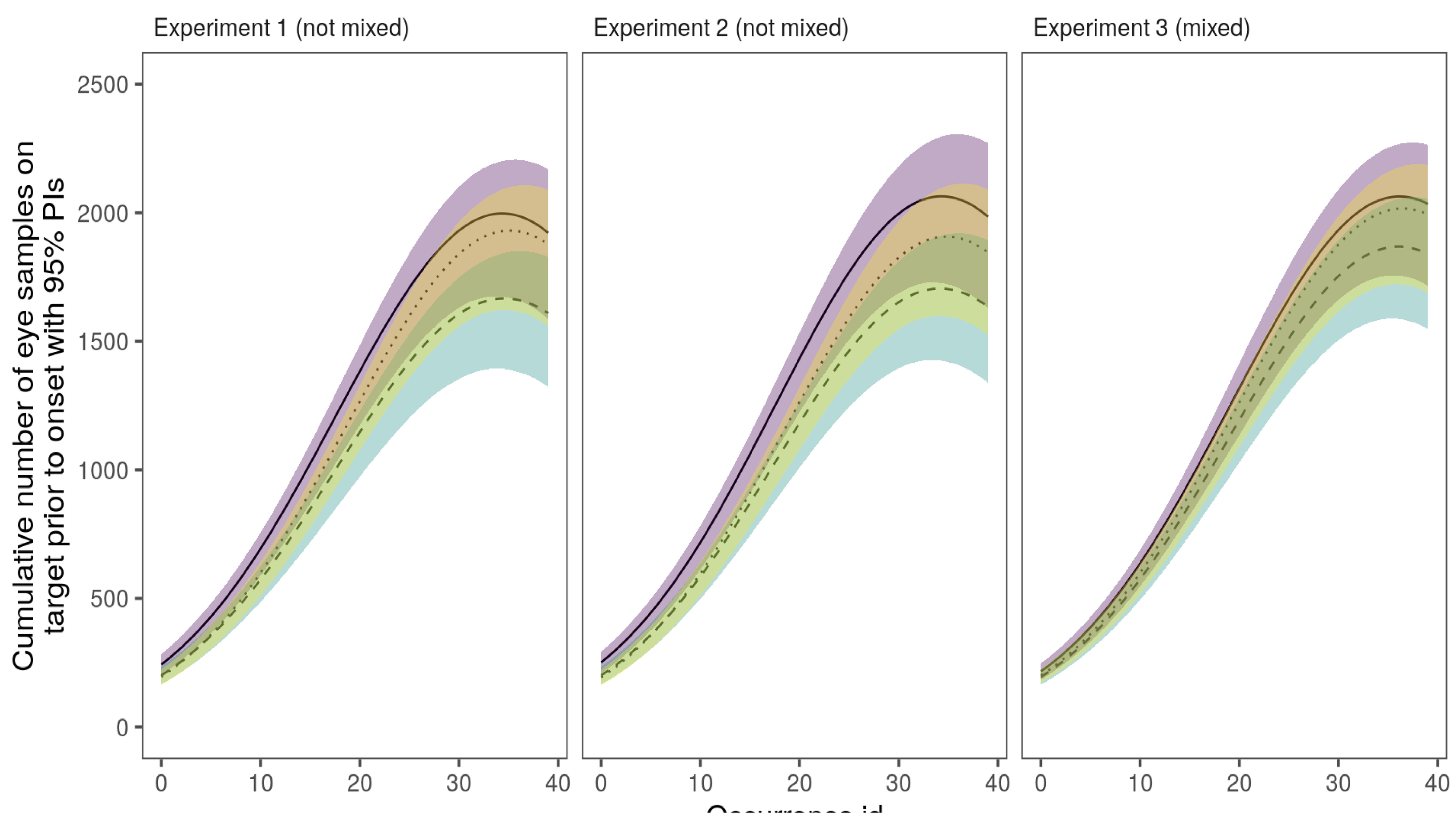
- The dependent variable was the number of eye samples on the target dot (C in the sequence A-B-C) before it illuminated (and 250 msecs after the previous target illuminated) accumulating across occurrences.
- Data was analysed in Bayesian mixed effects models following a zero-inflated negative binomial distribution (Gelman et al. 2014; McElreath 2016). The R package brms (Bürkner 2017, 2018) was used to model the data using the probabilistic programming language Stan (Carpenter et al. 2016; Hoffman and Gelman 2014). Fixed effects were main effects and interactions of occurrence id (1 to 40) and dependency type (levels: adjacent, nonadjacent, baseline). Occurrence id was treatment coded and dependency type was sum coded (for advantages of prespecifying model contrasts see Schad et al. 2020) comparing each dependency type to baseline; the baseline consists of transitions to every dot that was essentially random and not part of a dependency (neither dependee, nor dependent). Further, to model the learning curves, occurrence was modelled as quadratic function (modelled as second order orthogonal polynomial).
- Models were fitted with maximal random effects structure (Barr et al. 2013; Bates et al. 2015) including random participant intercepts with by-participant slope adjustments for the second order function of occurrence id as well as random slopes for the second-order polynomial of occurrence id[1] and their interaction with the combination of location set (2 sets), sequence (4 sequences per dependency type and location set), dependency type, and transition (levels: to A for adjacent dependencies, to B for nonadjacent dependencies, and to C) rending 16 levels for the latter.[2]
- We calculated the statistical support for the alternative hypothesis over the null hypothesis. This evidence was obtained using Bayes Factors (henceforth, BF) calculated using the Savage-Dickey method (see, e.g., Dickey, Lientz, et al. 1970; Wagenmakers et al. 2010). We calculated both the evidence for the alternative hypothesis H1
- over the null hypothesis given the data. A BF larger than 5 indicate moderate and larger than 10 strong evidence for a statistically meaningful effect compared to the null hypothesis (see, e.g., Baguley 2012; Jeffreys 1961; Lee and Wagenmakers 2014). For example BF of 2 reflect that the alternative hypothesis is two times more likely than the null hypothesis given the evidence. In contrast to traditional statistical methods (null-hypothesis significance testing), the Bayesian framework allows us to infer the evidence against the alternative hypothesis typically corresponding to BFs smaller than 0.33 (for discussion see Dienes 2014, 2016; Dienes and Mclatchie 2018; Schönbrodt et al. 2017; Wagenmakers et al. 2018).
- Models were fitted with weakly informative priors (see McElreath 2016) and run with 20,000 iterations on 3 chains with a warm-up of 10,000 iterations and no thinning. Model convergence was confirmed by the Rubin-Gelman statistic (Gelman and Rubin 1992) and inspection of the Markov chain Monte Carlo chains.

Results

Table 2.2: Fixed effects summaries for main effects and interactions of adjacent and nonadjacent dependency (compared to baseline) and occurrence id with linear and quadratic term.

	Experiment 1		Experiment 2		Experiment 3	
Predictor	**Est. with 95% PI**	**H_1**	**Est. with 95% PI**	**H_1**	**Est. with 95% PI**	**H_1**
Occurrence id (linear)	91.93 [82.42 – 98.71]	>100	91.81 [81.83 – 100.3]	>100	95.75 [88.14 – 101.06]	>100
Occurrence id (quadratic)	-30.68 [-33.48 – -27.23]	>100	-31.1 [-33.63 – -27.99]	>100	-29.41 [-32.12 – -26.36]	>100
Adjacent	**0.17 [0.09 – 0.26]**	**>100**	**0.21 [0.13 – 0.29]**	**>100**	**0.08 [0 – 0.17]**	**2.36**
Nonadjacent	**-0.16 [-0.24 – -0.08]**	**>100**	**-0.15 [-0.23 – -0.07]**	**>100**	**-0.08 [-0.17 – 0]**	**2.68**
Occurrence id (linear) : Adjacent	**-8.59 [-15.23 – -0.26]**	**15.1**	**-9.77 [-15.41 – -2.49]**	**72.31**	**-1.38 [-5.77 – 1.1]**	**1.37**
Occurrence id (quadratic) : Adjacent	0.59 [-1.53 – 3.4]	0.98	1.16 [-1.08 – 4.6]	1.3	0.1 [-2.15 – 2.46]	0.88
Occurrence id (linear) : Nonadjacent	-1.47 [-9.6 – 1.79]	1.26	-0.37 [-4.73 – 2.48]	1.04	-0.98 [-4.79 – 1.37]	1.2
Occurrence id (quadratic) : Nonadjacent	0.74 [-1.36 – 3.68]	1.03	0.24 [-1.85 – 2.56]	0.84	-0.1 [-2.45 – 2.14]	0.89

Note:
H_1 = evidence in favour of the alternative hypothesis over the null hypothesis (Bayes Factor); PI = probability interval; ':' = interaction



References

- References

NOTTINGHAM TRENT UNIVERSITY

Acknowledgements
This poster was possible because of the work of Karl Nicholls