Baboons, Homo Sapiens, and Implicit Learning in Reading: Deep Learning or Wide Learning?

Harald Baayen & Maja Linke

Seminar für Sprachwissenschaft
Universität Tübingen

November 24, 2016

the transposed letter effect

If You Can Raed Tihs, You Msut Be Raelly Smrat

Aoccdrnig to a rscheearch at Cmabrigde Uinervtisy, it deosn't mttaer in waht oredr the ltteers in a wrod are, the olny iprmoatnt tihng is taht the frist and lsat ltteers be at the rghit pclae. The rset can be a toatl mses and you can sitll raed it wouthit porbelm. Tihs is bcuseae the huamn mnid deos not raed ervey lteter by istlef, but the wrod as a wlohe

workhorse: the visual lexical decision task

+

WORD

workhorse: the visual lexical decision task

+

FLAS

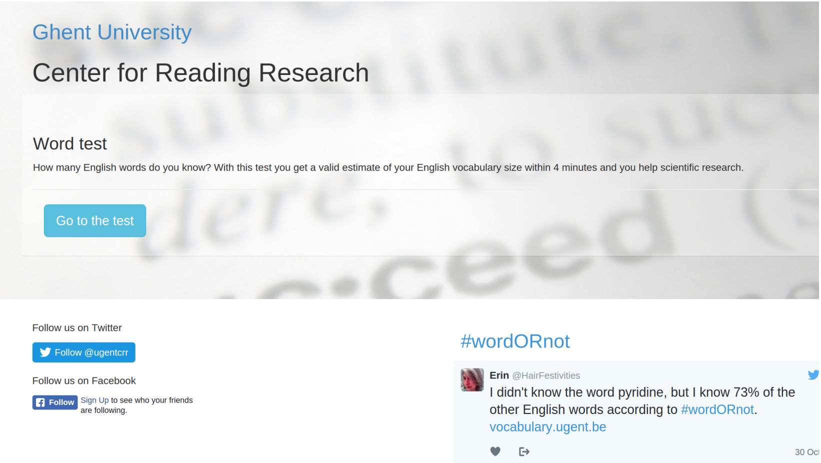
workhorse: the visual lexical decision task

+

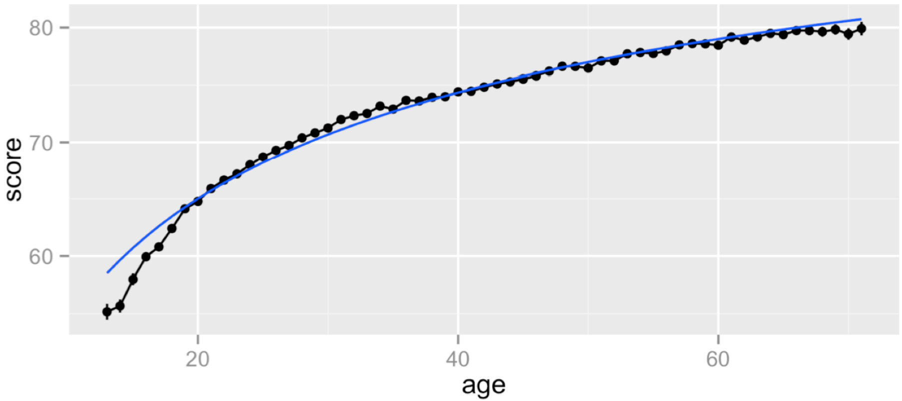
workhorse: the visual lexical decision task

TARE

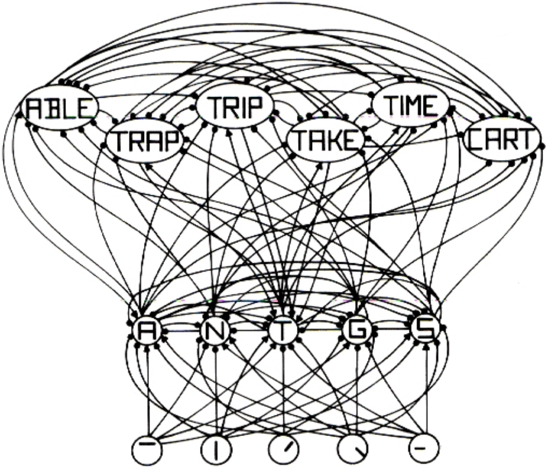
lexical decision masked as vocabulary test



lexical decision masked as vocabulary test



1981: interactive activation model



Guinea baboon (papio papio)

training baboons in visual lexical decision

- materials
  - words: TORE WEND BANG BOOR LEIS CHIC TACK KITE ...
  - nonwords: EFTD ULKH ULNX KRBA KRBO KRBU IMMF PSMI ...
- training regime
  - words: presented intensively until performance 80% correct, then repeated intermittently
  - nonwords: never repeated

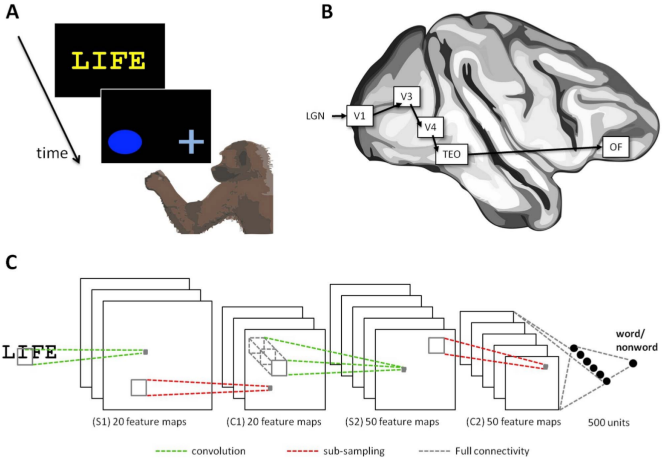
1. above chance selection of word response
   for real words at first encounter
   generalization

2. the more dissimilar a nonword is from real words,
   the more accurate the (nonword) response
   OLD20: average Levenshtein distance between a word and its 20 nearest neighbors in a lexicon
   neighborhood effect

3. the letter transposition effect: letter transpositions leave stimuli more word-like than letter substitutions
   letter transposition effect

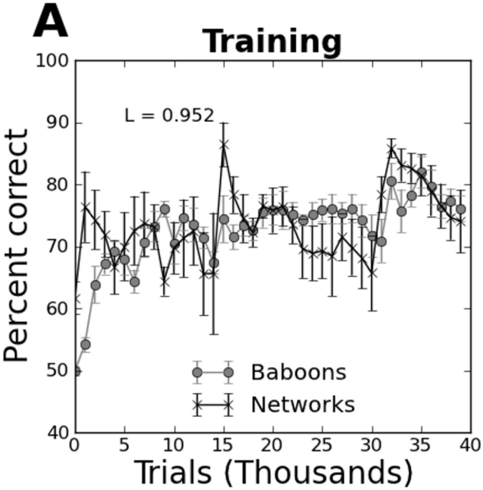
- transposed letters: LESS → LSES
- letter substitution: LESS → LUSS/LEPS
  - replace vowel by another vowel
  - replace consonant by another consonant
- nonwords derived from words by letter transpositions elicit more word responses than words with letter substitutions

2014: deep convolution network

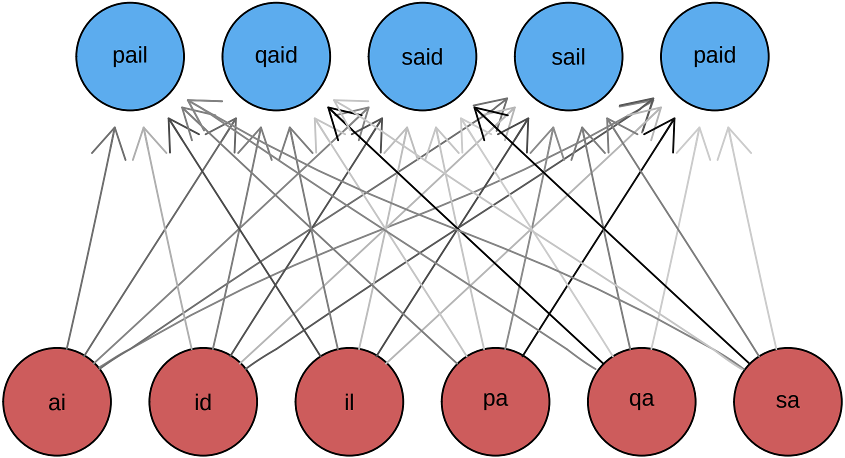


EBERHARD KARLS UNIVERSITÄT TÜBINGEN

deep convolution network performance

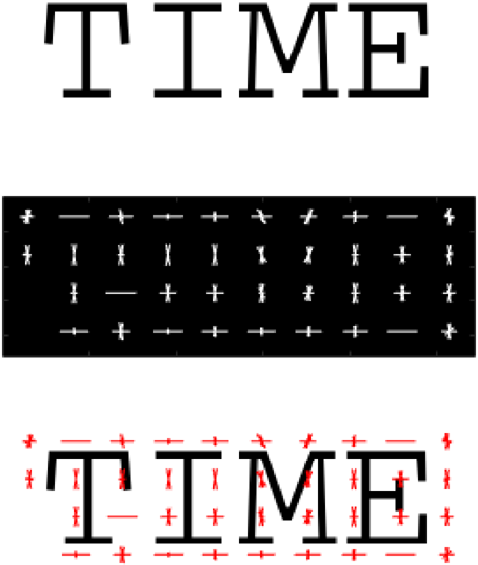


wide learning



input features

- simplest set-up of histogram orientation features (HOG) (Dalal & Triggs, 2005)
- 4 x 10 grid of cells
- for each cell gradient magnitude values for each of 9 orientations
- total number of distinct cues: 15149


TIME

output features

- a node for a YES response
- a node for a NO response

- for each baboon, the network is trained on exactly the sequence of words and nonwords encountered
- connection weights estimated
  using the Rescorla-Wagner learning equations

learning theory: the Rescorla-Wagner equations

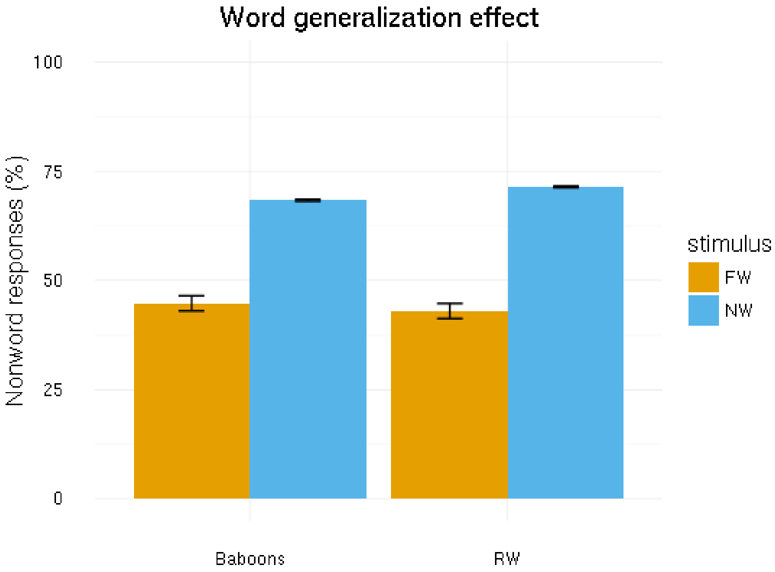
$$w_i^{t+1} = w_i^t + \Delta w_i^t$$

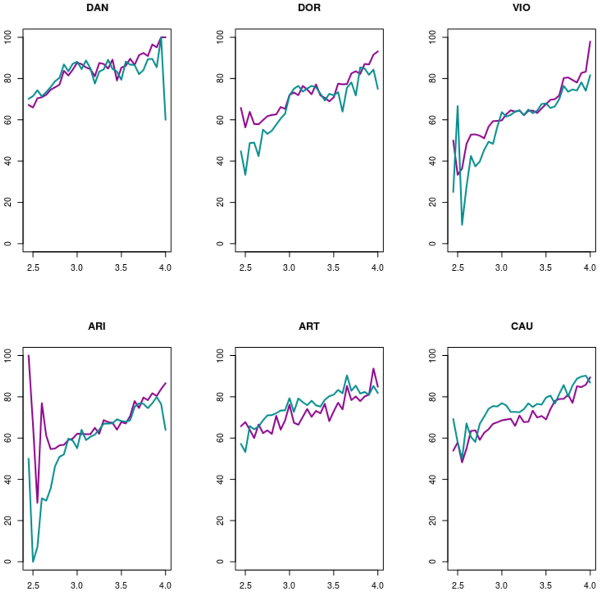
with

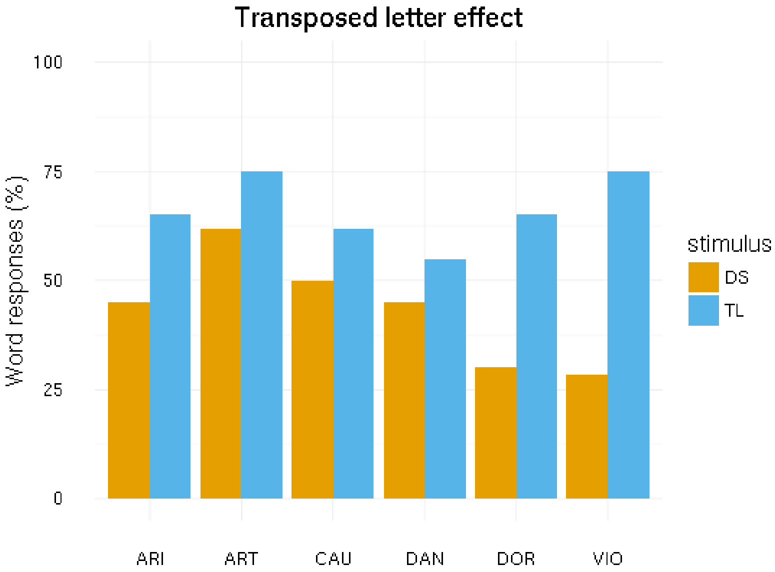
$$\Delta w_i^t = \begin{cases} 0 & \text{if } \textsc{absent}(C_i, t) \\ \alpha_i \beta_1 \left(\lambda - \sum_{\textsc{present}(C_j,\, t)} w_j \right) & \text{if } \textsc{present}(C_j, t)\ \&\ \textsc{present}(O, t) \\ \alpha_i \beta_2 \left(0 - \sum_{\textsc{present}(C_j,\, t)} w_j \right) & \text{if } \textsc{present}(C_j, t)\ \&\ \textsc{absent}(O, t) \end{cases}$$

wide learning decision criteria

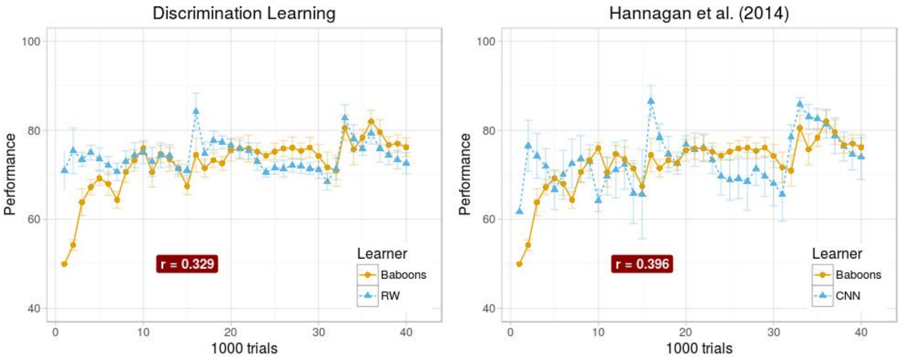
- response selection (not part of the network):
  - yes decision when $a(w) - a(nw) > \theta$
  - no decision when $a(nw) - a(w) > \theta$
  - random choice when $|a(w) - a(nw)| < \theta$
- comparison with baboon performance at 1000 trial intervals

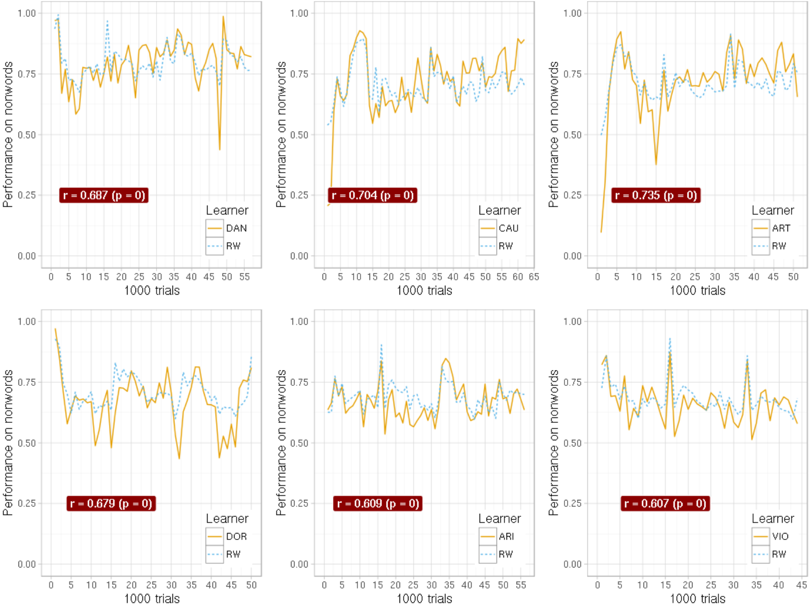

Word generalization effect
Nonword responses (%)
100
75
50
25
0
Baboons
RW
stimulus
FW
NW

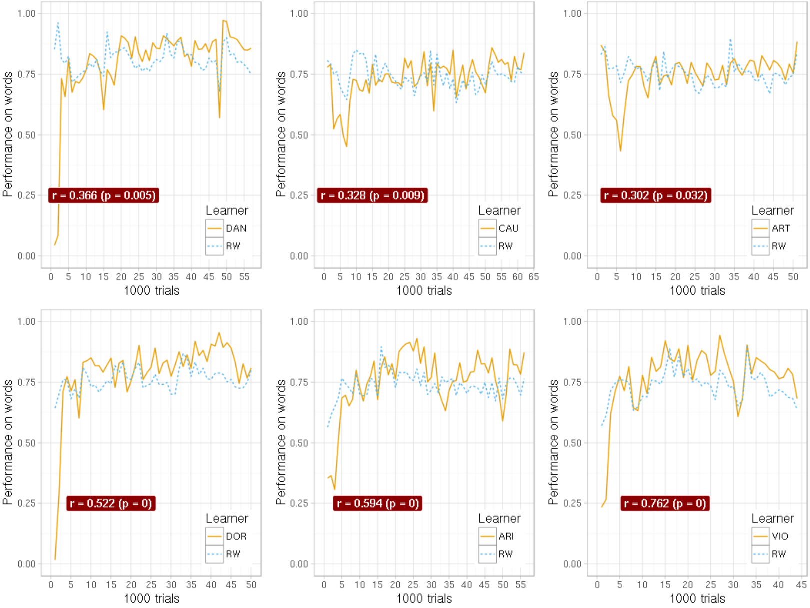

DAN
DOR
VIO
ARI
ART
CAU
0
20
40
60
80
100
2.5
3.0
3.5
4.0


Transposed letter effect
Word responses (%)
100
75
50
25
0
ARI
ART
CAU
DAN
DOR
VIO
stimulus
DS
TL

model performance




r = 0.687 (p = 0)
Learner
DAN
RW
r = 0.704 (p = 0)
CAU
r = 0.735 (p = 0)
ART
r = 0.679 (p = 0)
DOR
r = 0.609 (p = 0)
ARI
r = 0.607 (p = 0)
VIO
Performance on nonwords
1000 trials


Performance on words
1000 trials
r = 0.366 (p = 0.005)
Learner
DAN
RW
r = 0.328 (p = 0.009)
Learner
CAU
RW
r = 0.302 (p = 0.032)
Learner
ART
RW
r = 0.522 (p = 0)
Learner
DOR
RW
r = 0.594 (p = 0)
Learner
ARI
RW
r = 0.762 (p = 0)
Learner
VIO
RW

lexical units?

baboon data

- strong support for words in mixed models
- weaker support for letter bigrams

wide learning predictions

- strong support for words in mixed models
- weaker support for letter bigrams

lexical units?

- potential sources of evidence for lexical units:
  - representations in the brain
  - patterns in the input that are reflected in the brain's goal-directed response to the environment

- There is one key computational argument against a major role for holistic word-shape information in reading: it is more efficient to solve shape invariance at the level of individual letters (N=26) than at the level of whole words (N30,000).

  Grainger, J., & Hannagan, T. (2014). What is special about orthographic processing?

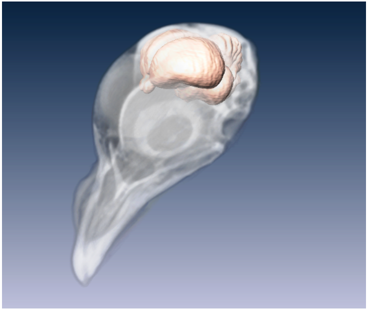
- this is not about being able to discriminate probabilistic conjunctions of three to four individual shapes. Words and nonwords were both made up of the same 26 letters. Thus, there were potentially thousands of conjunctions that had to be learned

  Ziegler et al. (2013). What Can We Learn From Humans About Orthographic Processing in Monkeys? A Reply to Frost and Keuleers (2013).

- ...monkeys use a truly orthographic rather than a visual code, that is, they decompose words into their component letters rather than process them as a whole visual shape.
- ...prior linguistic knowledge is therefore not a necessary prerequisite in order to achieve humanlike orthographic processing.
- ...orthographic processing is, at least partly, constrained by general principles of visual object processing shared by monkeys and humans, and, finally, the front end of reading is supported by neural mechanisms that are much older than the behavior itself and are not purely linguistic in nature.
- ...the rat example is nice, but Frost and Keuleers ignore that the strength of our comparison lies in the fact that, contrary to rats, humans and baboons have highly similar genetic makeups and comparable visual systems.

rock pigeon (columba livia)

pigeon brain morphology



- some pigeons, with a lot of training, can learn 26 up to 58 words
- quite remarkable, as they are neocortically rather challenged
- hypothesis: a simple food network would suffice for pigeons

Pigeons not only correctly identified novel words but also display the hallmarks of orthographic processing, in that they are sensitive to the bigram frequencies of words, the orthographic similarity between words and nonwords, and the transposition of letters.

Scarf, D., Boy, K., Reinert, A. U., Devine, J., Gunturkun, O., & Colombo, M. (2016).