Reading minds: using EEG pattern classification to predict behavior in negative priming

Ehren L Newman & Kenneth A Norman*

Princeton University, Princeton NJ, USA

see http://compmem.princeton.edu/publications.html for reprints

Abstract

Competition in cognitive processing has lasting consequences for the subsequent accessibility of competing representations. Negative priming (NP) demonstrates that, when representations compete, the representations that lose the competition are subsequently harder to access. To better understand the competitive dynamics that generate these effects, we developed a method of tracking the activation of the competing representations at the sub-trial time scale. Our methods rely on a pattern classification analysis of EEG data. Replicating our previous results, we show that the method is capable of 1) detecting which of four image categories a subject is viewing, 2) generalizing to EEG collected during a separate session of the experiment, and 3) detecting the classes of simultaneously presented images. Building on these previous findings, we present preliminary results that give insight how subjects ignore an image such that they are subsequently slower to name it on the following trial. Specifically, we compare the detected pattern of competitor activation on the negative priming trials for which the subjects show the slowest vs fastest reaction times. In addition to these results, we discuss the strengths and weaknesses of our EEG pattern classification technique.

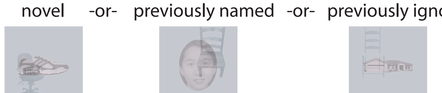
Introduction

Making a choice has consequences
- Chosen item subsequently stronger
- Non-chosen item is subsequently weaker

Example - Negative priming (Tipper, 1985)

Task:
- Two stimuli simultaneously presented
- Subject asked to name one & ignore the other
  - e.g. "Name the red tinted image in the center"
- Later: the image to be named could be:

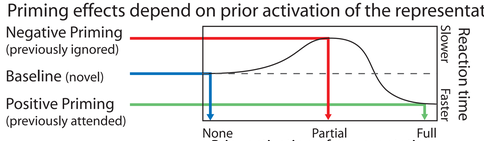
novel -or- previously named -or- previously ignored



Effect:
- Compared to the response time to a novel image
- 1) subjects are faster to name a previously attended image
- 2) subjects are slower to name a previously ignored image
- Sensitive to display layout and presentation parameters
  - e.g. smaller gap between stimuli induces more NP(Fox,1994)
  - however, others find smaller gap removes NP (Fuentes 1998)

Hypothesis:

Priming effects depend on prior activation of the representation



Our Goal:

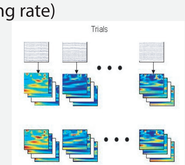
Identify when subjects fail to ignore the distractor and test our hypothesis :
- when subjects did not ignore the distractor we should find positive priming
- when subjects did ignore distractor we should find negative priming

Prediction: Amount of negative priming should be negatively correlated with ability to classify distractor image

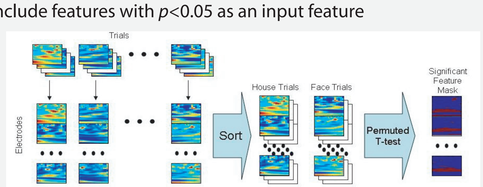
Our approaches:
1. Attempt to train classifier to identify distractor image category
   - This should not be possible if subjects successfully ignore the distractor
2. Use classifier trained to recognize the image as target to measure activity of image as the distractor
   - Detection of image as the distractor should predict faster reaction times

Decoding EEG via distributed pattern analysis

Data preparation
- Collect data with 79 electrode cap (1000Hz sampling rate)
- Remove trials with excessive noise or blinks
- Perform Frequency decomposition
  - Wavelet decomposition (6 cycle Morlet wavelet)
  - 49 frequency bands between 2 & 128Hz
  - Extract power of each frequency band
- Collapse data to form 20ms time bins (averaging)

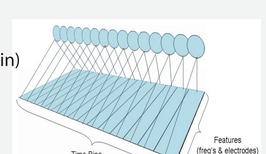


Classification preparation -
- Perform feature selection across time *bins / frequencies / electrodes*
  - Compute non-parametric *p*-value for each combination
  - Include features with $p<0.05$ as an input feature

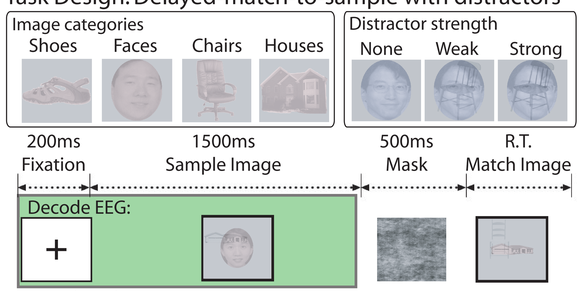


- Build a ridge regression classifier for each time bin
  - Input patterns -
    - Significantly discriminating frequency / electrode combinations
  - Output patterns -
    - Binary regressors

Classification procedure -
- Training the classifiers (for each time bin)
  - Use N-1 cross validation approach
    - train on 9/10th of the trials
    - test on remaining 1/10th
    - repeat 10x



Task Design: Delayed-match-to-sample with distractors

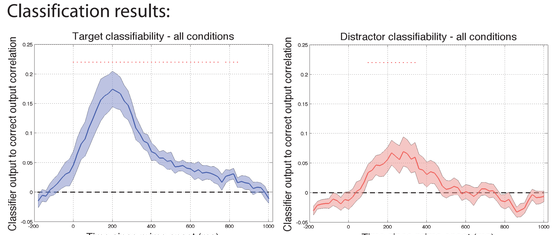


On each trial:
- subject sees two displays and asked to either:
  - say "match" if the red tinted objects are identical
  - name the second image if the objects differ
- All displays include a distractor image in the background
  - subjects are instructed to ignore this image
- 1/3rd of the trials subjects must name the object just ignored

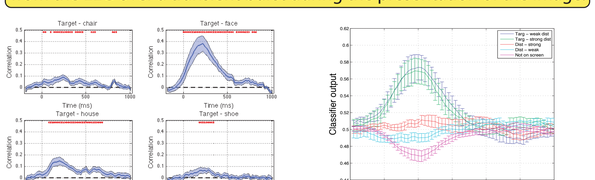
Results:

Behavioral results:
- Time to name novel image: **907ms**
- Time to name ignored image: **921ms**
- Negative priming effect: **14ms** (t(17)=-2.39, p<0.03)

Classification results:



We are able to classify the category of the attended and ignored image from the EEG over a transient burst during the presentation of the image.

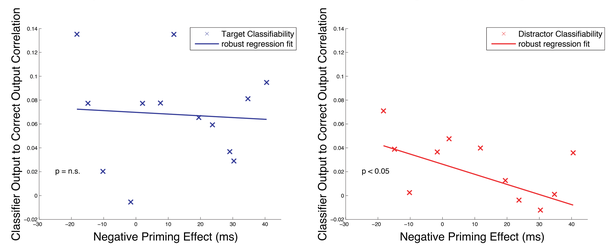


Target classification by category

Classification generalization varies with strength of the distractor

Predicting Behavior:

Ability to train classifier predicts amount of negative priming



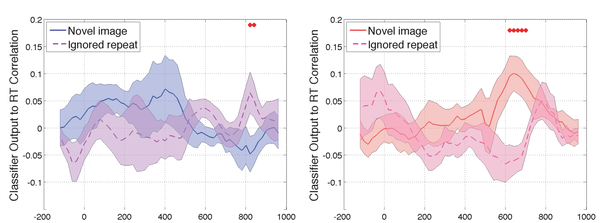
Ability to classify target does not predict negative priming effect

Ability to classify distractor does predict negative priming effect

In other words: Subjects who fail to demonstrate the negative priming effect are the same subjects whose EEG data contains enough information to identify the category of the distractor image.

- This suggests that these subjects are failing to ignore the distractor images

Activity of trained classifier predicts reaction time



Positive correlation between target classifier activity and RT (relative to baseline) during negative priming trials

Negative correlation between distractor classifier activity and RT (relative to baseline) during negative priming trials

In other words: Greater distractor activation (and lower target activation) during the cue were predictive of faster reaction times on negative priming trials.

- This provides further evidence that when subjects fail to ignore the distractor images they generate positive priming instead of negative priming.

Fox, E.(1994) Interference and negative priming from ignored distractors: The role of selection difficulty. *Perception &Psychophysics*, 56 (5), 565-574.

Fuentes LJ, Humphreys GW, Agis IF, Encarna C, Catena A (1998)Object-based perceptual grouping affects negative priming. *Journal of Experimental Psychology: Human Perception and Performance,* 24, 664-672.

Tipper, S. P. (1985). The negative priming effect: Inhibitory priming by ignored objects. The Quarterly Journal of Experimental Psychology, 37A, 571-590.

Funding for this research was provided by
NIMH grant R01 MH069456 awarded to KAN & NIMH grant F31 MH077469 awarded to ELN