Evaluating experiments (part 2)

Andy J. Wills

2016

Confounding variables

- Any variable, other than the one you are attempting to study, that varies between conditions, and which could potentially have led to the effect you observe.

Confounds discussed so far...

- Pre-existing differences (address by matching or randomization)
- Differential attrition (major issue, some partial solutions)
- Hawthorne effect / Placebo effect
- Demand charcteristics

Allocation of markers example

- 300 students, 2 markers.
- Marker 1 gets first 150 scripts handed in. Average mark B+
- Marker 2 gets the rest. Average mark B.

- What might explain the difference between markers?

Possible responses

- Use a correlation analysis to study date-of-submission vs. grade relationship.
- Randomize allocation of scripts to markers.

Therapy example

- Compare meditation-based therapy with relaxation training.
- Large, randomized groups.
- No pre-treatment differences in BDI
- No differential attrition
- BDI drops more for meditation than relaxation.
- Meditation is the more effective treatment.

Therapy example - Closer look

- Compare meditation-based therapy with relaxation training.
  - Meditation - Delivered by the people who developed the treatment
  - Relaxation - Delivered by people with no particular investment in relaxation therapy, who have been on a one-week training course in relaxation therapy.
- Large, randomized groups.
- No pre-treatment differences in BDI
- BDI drops more for meditation than relaxation.
- Meditation is the more effective treatment.

Therapy example - Closer look

- Compare meditation-based therapy with relaxation training.
  - Meditation - Delivered by the people who developed the treatment
  - Relaxation - Delivered by people with no particular investment in relaxation therapy, who have been on a one-week training course in relaxation therapy.
- Alternative explanation? - It's not the type of therapy that matters. It's some combination of therapist's belief in the treatment, their experience in delivering it, their general level of therapeutic expertise.

Experimenter Effects - Data analysis - Example

- Diary entries as a measure of happiness.
- Participants write about their feelings
- Experimenter rates for level of happiness.
- If experimenter knows which condition the participant is in, this may bias their assessment of happiness.

Experimenter Effects - Data analysis

- Objective measures immune?
- No! - Data analysis typically involves many decisions, all open to bias.
  - Should I exclude outliers?
  - If so, what's the cut-off?
  - Should I use a parametric or non-parametric test?
  - Are these tests multiple comparisons I should correct for, or separate analyses (for which I don't correct)?
- If the experimenter knows which condition the participants are in, this could bias their decisions.

Data analysis - Example

- My theory predicts people react more quickly to auditory than to visual alarm signals.
- I find this result if I exclude all reaction times above 3 seconds
- But not if I keep all RTs
- and not if I exclude all reaction times below 100ms.
- I choose the 3 second cut-off
- Am I sure that decision was unbiased?

Blind testing

- Single-blind testing - participant does not know which condition they are in.
  - e.g. Drug vs. placebo. Participants do not know which condition they are in.
- Double-blind testing - single-blind testing plus the experimenters do not know which condition is which until after they have completed their analysis.

Order effects

- Auditory versus visual alarm signals, within-subjects design
- Visual (300ms) $\rightarrow$ Auditory (250ms)
- Auditory faster?
- Or, practice effect?

Order effects

- Auditory versus visual alarm signals, within-subjects design
- Auditory (250ms) $\rightarrow$ Visual (300ms)
- Auditory faster?
- Or, fatigue effect?

Order effects

- Auditory versus visual alarm signals, within-subjects design
- Randomly allocate participants to the two orders
- Auditory (250ms) $\rightarrow$ Visual (300ms)
- Visual (300ms) $\rightarrow$ Auditory (250ms)
- Auditory faster - irrespective of order.
- No practice or fatigue effect (mean RT across conditions 275 ms at time 1 and time 2).

Difference versus no difference designs

- The preferred hypothesis is that people differ in the speed with which they react to auditory and visual alarm signals.
- The alternative theory against which this is compared is that there is no difference (nil hypothesis).
- Problem - Experimental control is never perfect.
- Thus - the nil hypothesis is almost certainly wrong, and detectably so if you test enough people.
- Thus - the result of the study is known before you run it.
- Thus - There was no point in running it.

Better alternatives 1

- One-tailed tests
  - The preferred theory is that auditory is faster.
  - The alternative theory against which this is compared is that there is no difference (nil hypothesis).
  - If you find visual faster, you have disproved your theory.
  - So, whatever the result, there was a point to running this experiment (because the theory was falsifiable).

Better alternatives 2

- Ordinally different theories
  - One well-established theory predicts that auditory is faster.
  - Another well-established theory predicts that visual is faster.
  - Whatever you find in this study, you've gained information (except in the unlikely case where the nil hypothesis was true).

Better alternatives 3

- Effect size
  - Theory 1 - RT will differ between auditory and visual by more than two standard deviations.
  - Theory 2 - RT will differ between auditory and visual by less than two standard deviations.

Statistical control myth

- The idea you can "control for" group differences by clever statistics (e.g. Analysis of Covariance).
- Example
  - Group 1 - Depressed. Selected on BDI score
  - Group 2 - Nondepressed. Selected on BDI score.
  - Group 1 are worse than Group 2 on a memory test
- Depression causes poor memory? Poor memory causes depression?
- Inspection reveals Group 1 are, on average, older than Group 2.
- Perhaps age is the causal factor?
- The Myth - you can "control for" age as a factor by entering as a co-variate into your analysis.
- This is false (Miller & Chapman, 2001).

Infallible Flowchart of Paper Evaluation

1. **Central topic** (What is it about?)
2. **Central research question** (What do they wish to test?)
3. **Central rationale** (Why bother?)
4. **Essence of Method** (How did they test their question?)
5. **Key Result**
6. **Central Valid Conclusion**

Evaluating a research paper

Research Article

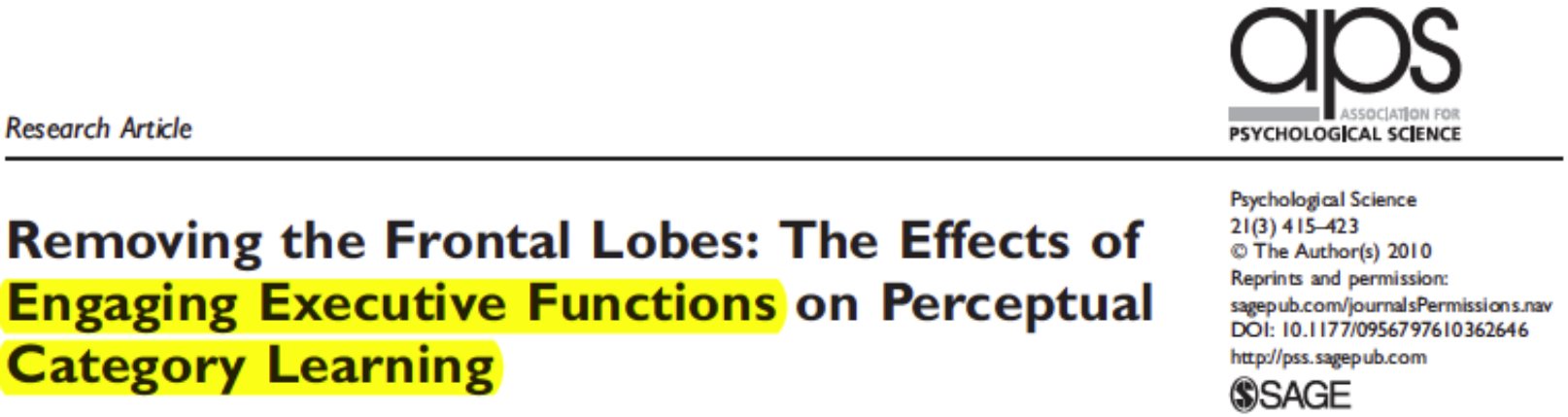
aps ASSOCIATION FOR PSYCHOLOGICAL SCIENCE

Psychological Science
21(3) 415–423
© The Author(s) 2010
Reprints and permission:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/0956797610362646
http://pss.sagepub.com
SAGE

Removing the Frontal Lobes: The Effects of Engaging Executive Functions on Perceptual Category Learning

J. Vincent Filoteo[1,2], Scott Lauritzen[3], and W. Todd Maddox[3,4]
[1]Research Service, VA San Diego Healthcare System, San Diego, California; [2]Department of Psychiatry, University of California, San Diego; [3]Department of Psychology, University of Texas at Austin; and [4]Institute for Neuroscience, University of Texas at Austin

Central topic



Research Article

aps ASSOCIATION FOR PSYCHOLOGICAL SCIENCE

Removing the Frontal Lobes: The Effects of Engaging Executive Functions on Perceptual Category Learning

Psychological Science
21(3) 415–423
© The Author(s) 2010
Reprints and permission:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/0956797610362646
http://pss.sagepub.com
SAGE

J. Vincent Filoteo[1,2], Scott Lauritzen[3], and W. Todd Maddox[3,4]
[1]Research Service, VA San Diego Healthcare System, San Diego, California; [2]Department of Psychiatry, University of California, San Diego; [3]Department of Psychology, University of Texas at Austin; and [4]Institute for Neuroscience, University of Texas at Austin

- Category learning - The act of learning how to divide the world into groups of things.
- Executive function - Management of cognitive processes.

Research question

learn the information-integration task (i.e., in the information-integration three-dimensional working memory, II-3D-WM, condition vs. the II-3D condition). This would lead to the paradoxical effect of enhanced information-integration learning under secondary-sequential-task conditions. In essence, the addition of the secondary task would behaviorally limit the contribution of the frontal lobes by overly engaging working memory processes during the processing of the corrective feedback, thus allowing the procedural-based system to operate without competition. On the other hand, a secondary working memory task that behaviorally removes the contributions of the frontal lobes should further tax the hypothesis-testing system, leading to an even greater performance decrement for conjunctive learning (i.e., in the conjunctive three-dimensional working memory, CJ-3D-WM, condition relative to the CJ-3D condition).

(c) 2010. J.V. Filoteo, S. Lauritzen, W.T. Maddox.

Research question

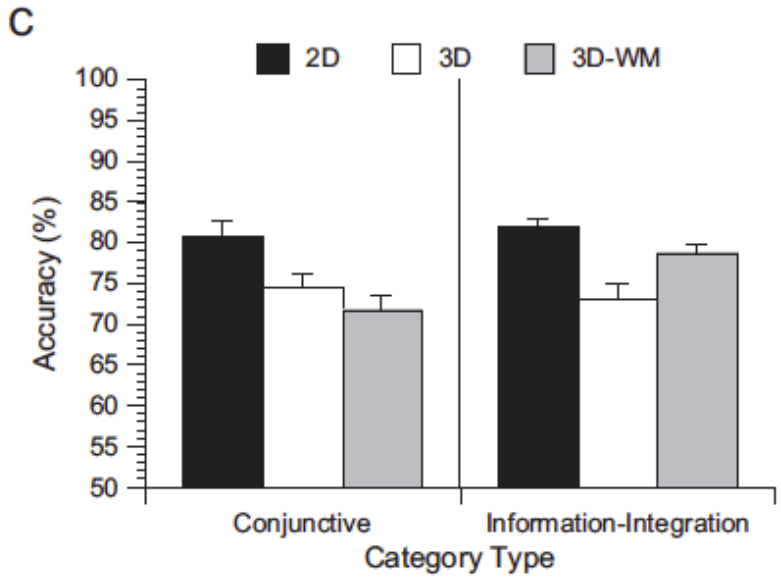
implicit systems theory (COVIS; Ashby et al., 1998) proposes two distinct systems: an explicit, hypothesis-testing system, which mediates rule-based category learning and relies on working memory and attention, and an implicit, procedural-based system, which mediates information-integration category learning and relies on reinforcement-based learning processes. These systems are thought to have different neuroanatomical bases, with the hypothesis-testing system mediated to a greater extent in frontal cortices and the procedural-based system relying more on posterior regions of the striatum (Ashby et al., 1998; Filoteo et al., 2005; Nomura et al., 2007; Seger & Cincotta, 2002).

(c) 2010. J.V. Filoteo, S. Lauritzen, W.T. Maddox.

Methodology - Key questions

- What is the dependent variable (DV)?
- Is the DV appropriate for the question?
- What are the independent variables (IV)?
- Are the IVs appropriate for the question?
- Are the IVs confounded with any other variables?

DV and IVs



- DV - Accuracy. Appropriate.
- IV1 - Category type (conjunctive vs. information integration). Appropriate.
- IV2 - Cognitive load (digit load vs. no digit load). Appropriate.

Image (c) 2010. J.V. Filoteo, S. Lauritzen, W.T. Maddox.

IV confound

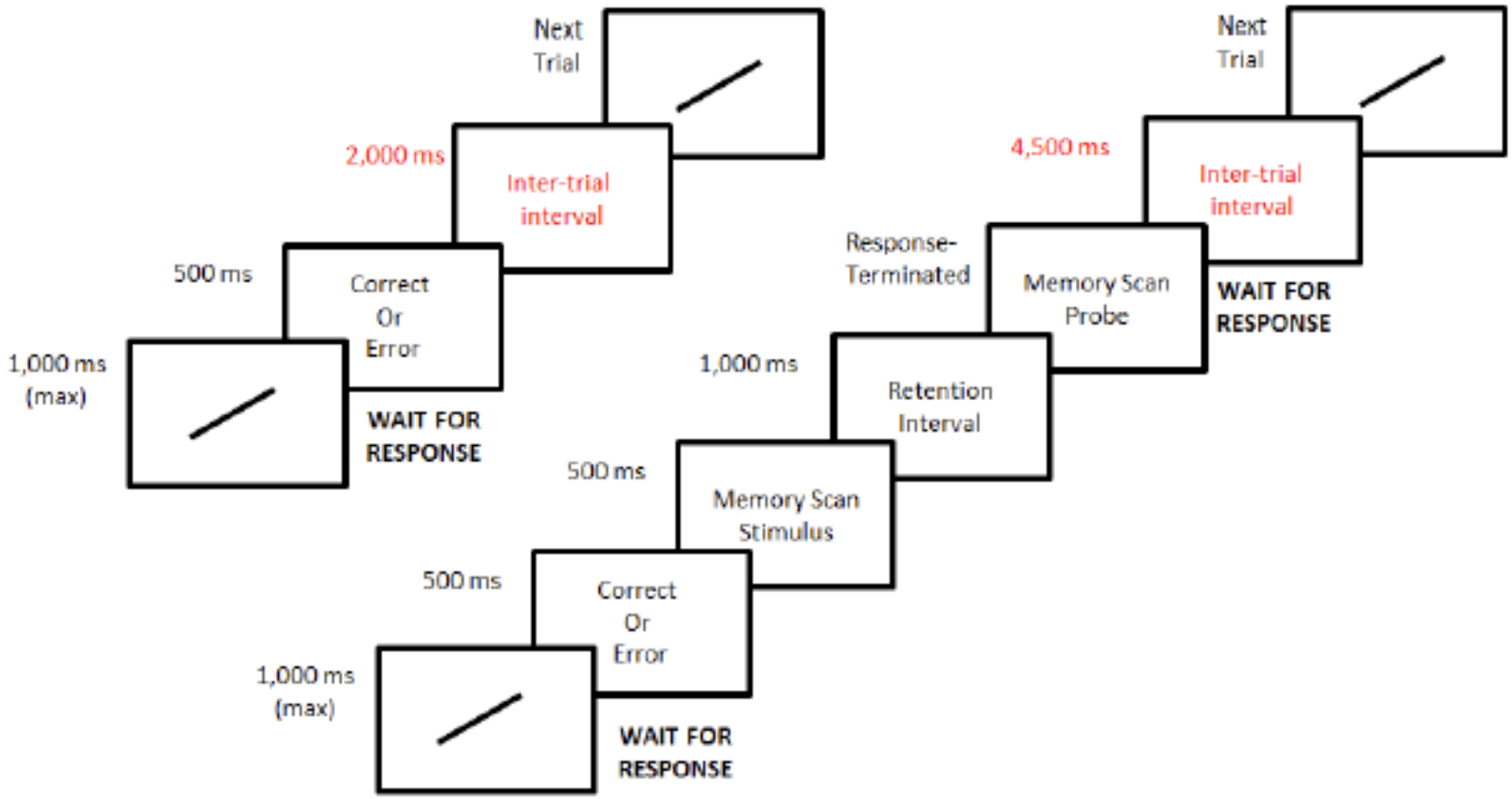
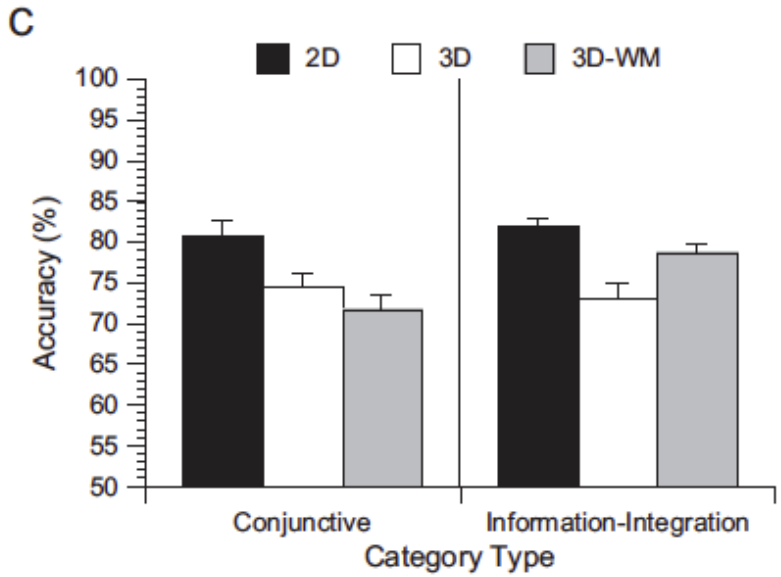


Figure S2. Trial structure for Filoteo et al.'s II3D (left) and II3D-WM (right) conditions.

(c) 2013. Newell, Moore, Milton & Wills

Key Result



- Concurrent load increases accuracy for II structure.
- Concurrent load decreases accuracy for CJ structure.

Infallible Flowchart of Paper Evaluation

- **Central topic** (What is it about?)
  Effect of "engaging" executive functions on category learning
- **Central research question** (What do they wish to test?)
  Secondary task will improve information-integration category learning but hurt conjunctive category learning.
- **Central rationale** (Why bother?)
  Prediction of COVIS theory

Infallible Flowchart of Paper Evaluation

- **Essence of Method** (How did they test their question?)
  DV - Accuracy. IV1 - category structure (CJ vs. II). IV2 - digit load (present vs. absent). IV2 confounded with ITI. The Method is not a particularly good test of the central research question.
- **Key Result**
  Digit load increases accuracy in II. Digit load has no significant effect in CJ. These two conditions differ significantly. The results are partially consistent with the hypothesis.
- **Central Valid Conclusion**
  Digit load, combined with extra time to think, improves accuracy on an information integration category structure.

Except where noted, this work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International Licence. Last update: September 15, 2016