A practical introduction to distributed collaboration for formal modelling

Andy J. Wills
Plymouth University

Schedule

10:30 Introduction and demo

11:00 Software install

11:15 Example data set

11:30 Adding a data set to **catlearn**

11:45 – short break –

Schedule

12:00 Documenting a data set

12:30 Writing an OAT function

13:00 Next steps

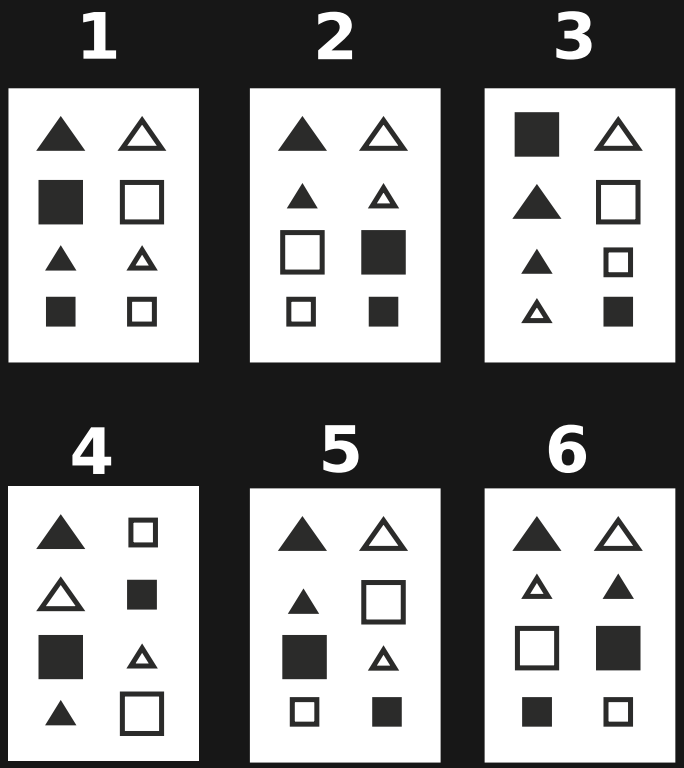
Progress through distributed collaboration in formal modelling

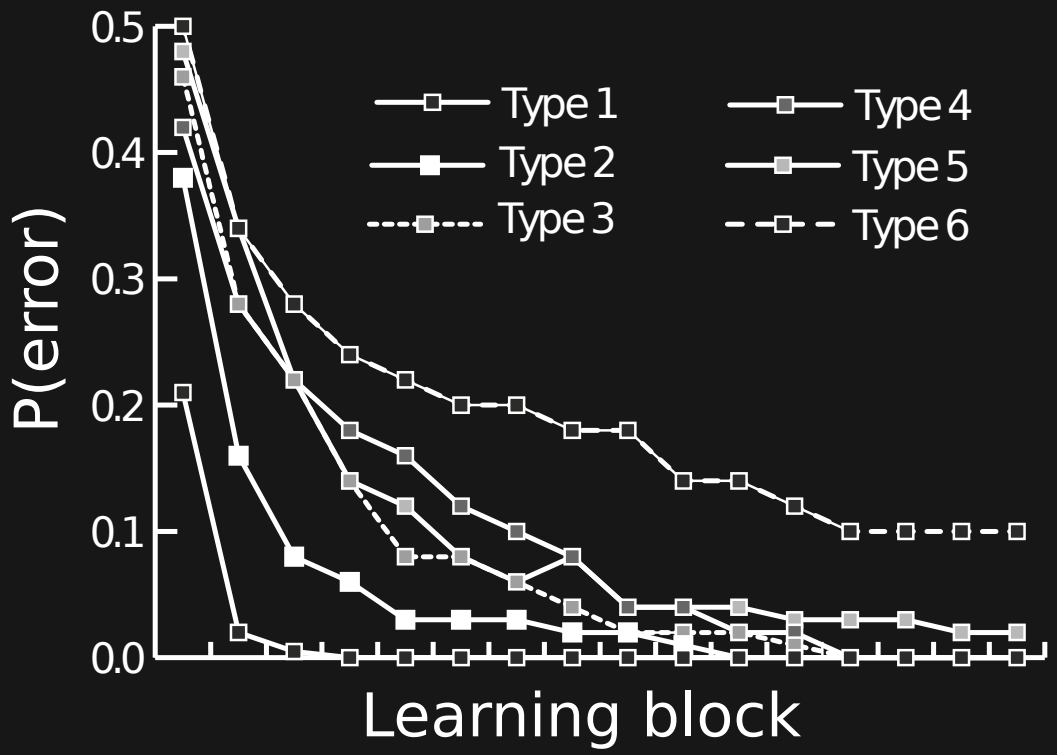
Formal modelling in psychology is failing to live up to its potential due to a lack of effective collaboration.

catlearn: Free and open-source tools for distributed collaboration.

A formal model unambiguously specifies transformations from independent variables to dependent variables.

Wills & Pothos (2012)


1
2
3
4
5
6


0.5
0.4
0.3
0.2
0.1
0.0
P(error)
Learning block
Type 1
Type 2
Type 3
Type 4
Type 5
Type 6

Formal models are specified mathematically.

- Rescorla-Wagner (1972)
- ALCOVE (Kruschke, 1992)
- DRC (Coltheart et al., 2001)

Formal models allow unambiguous comparison of the relative adequacy of theories.

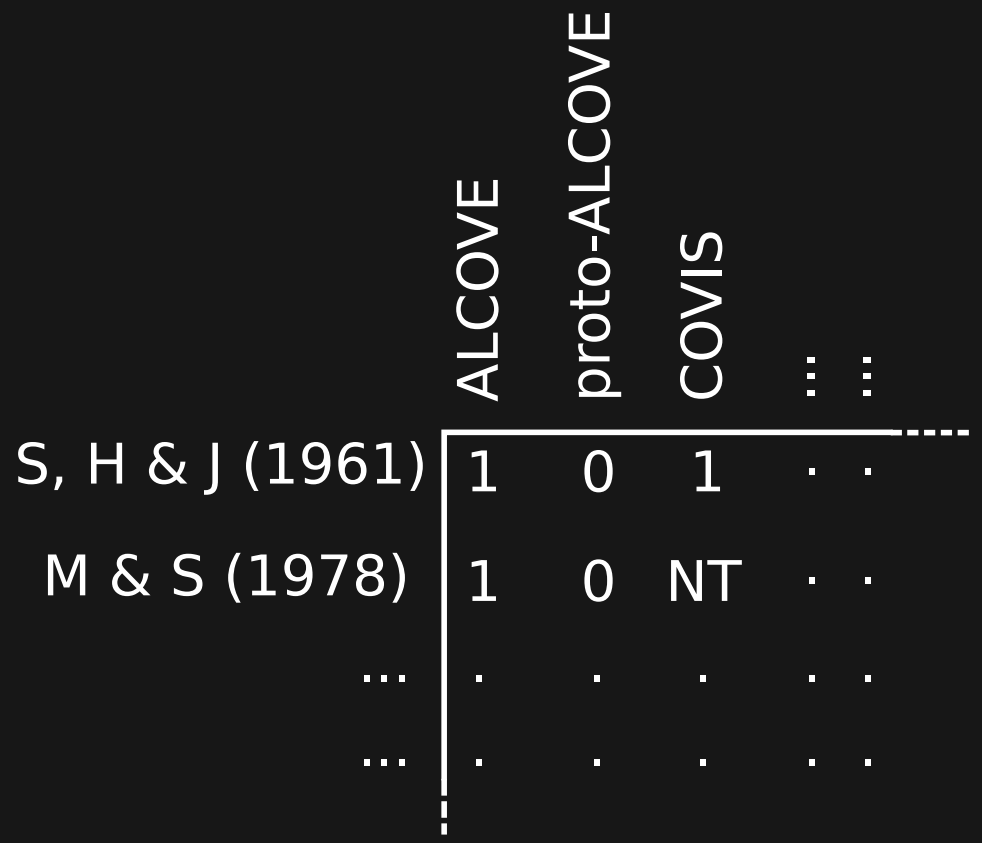
Little progress so far … because the comparisons have been too narrow.

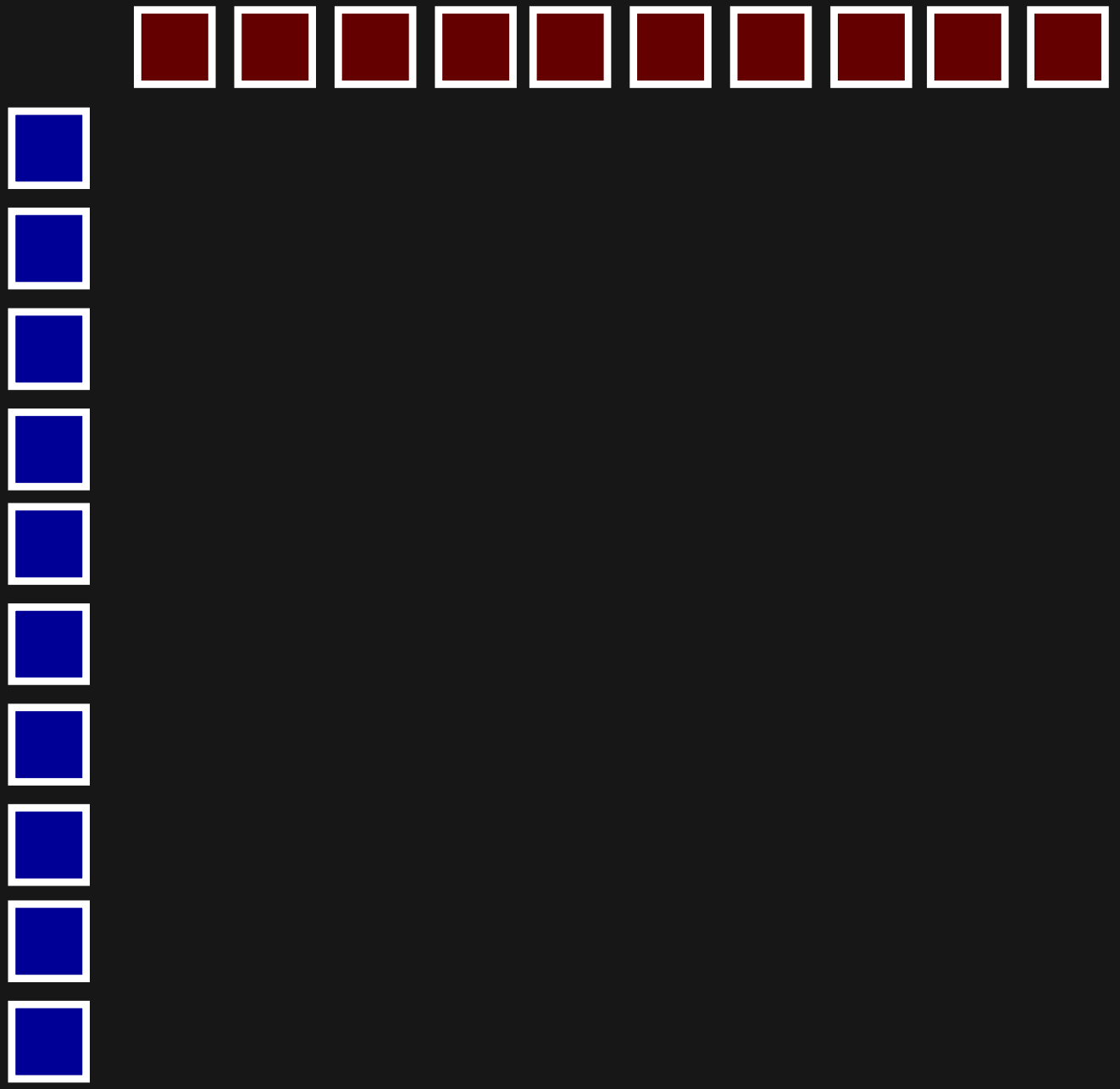
Wills & Pothos (2012)

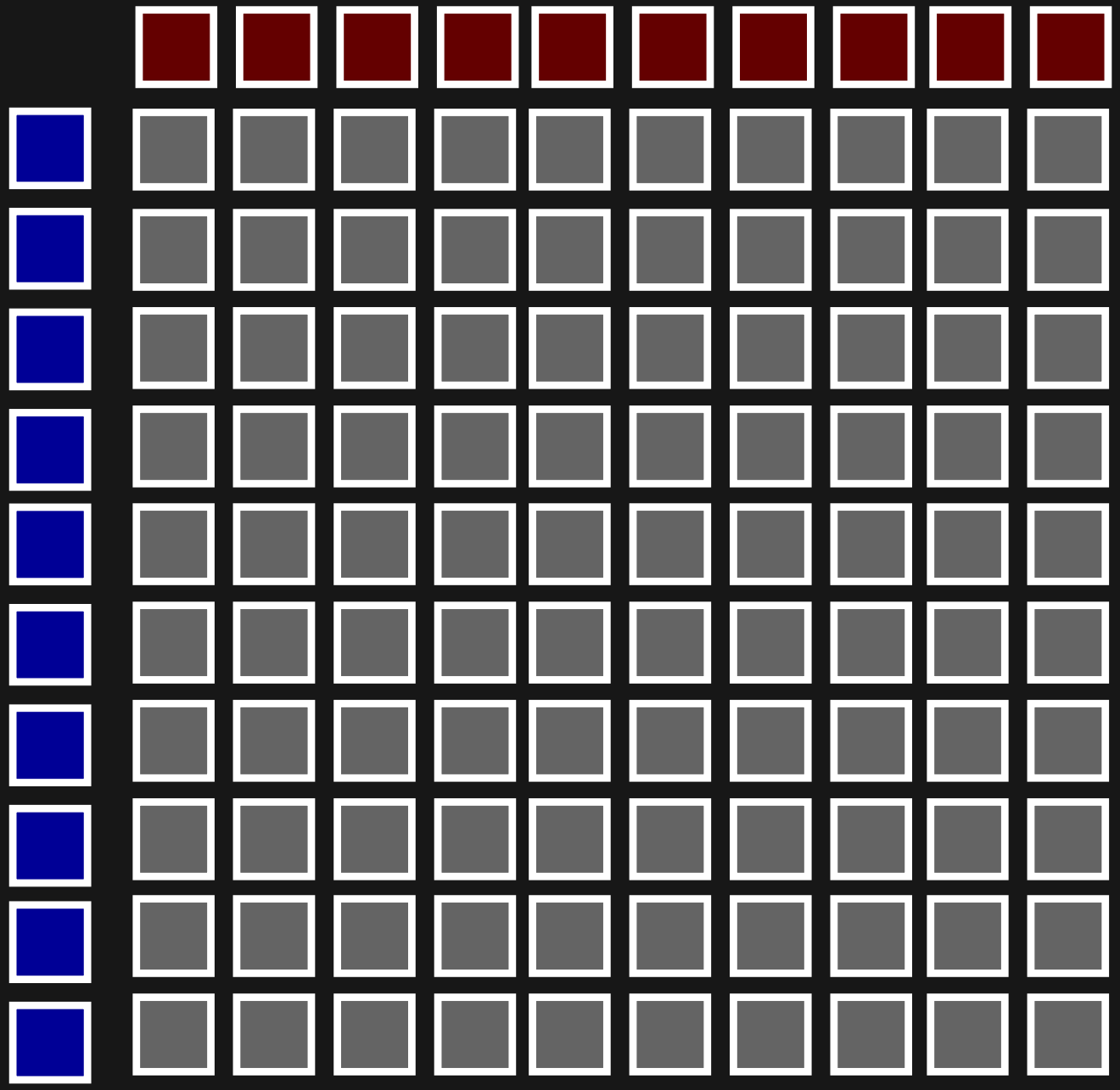
Compare models across a much broader set of phenomena (Wills & Pothos, 2012).

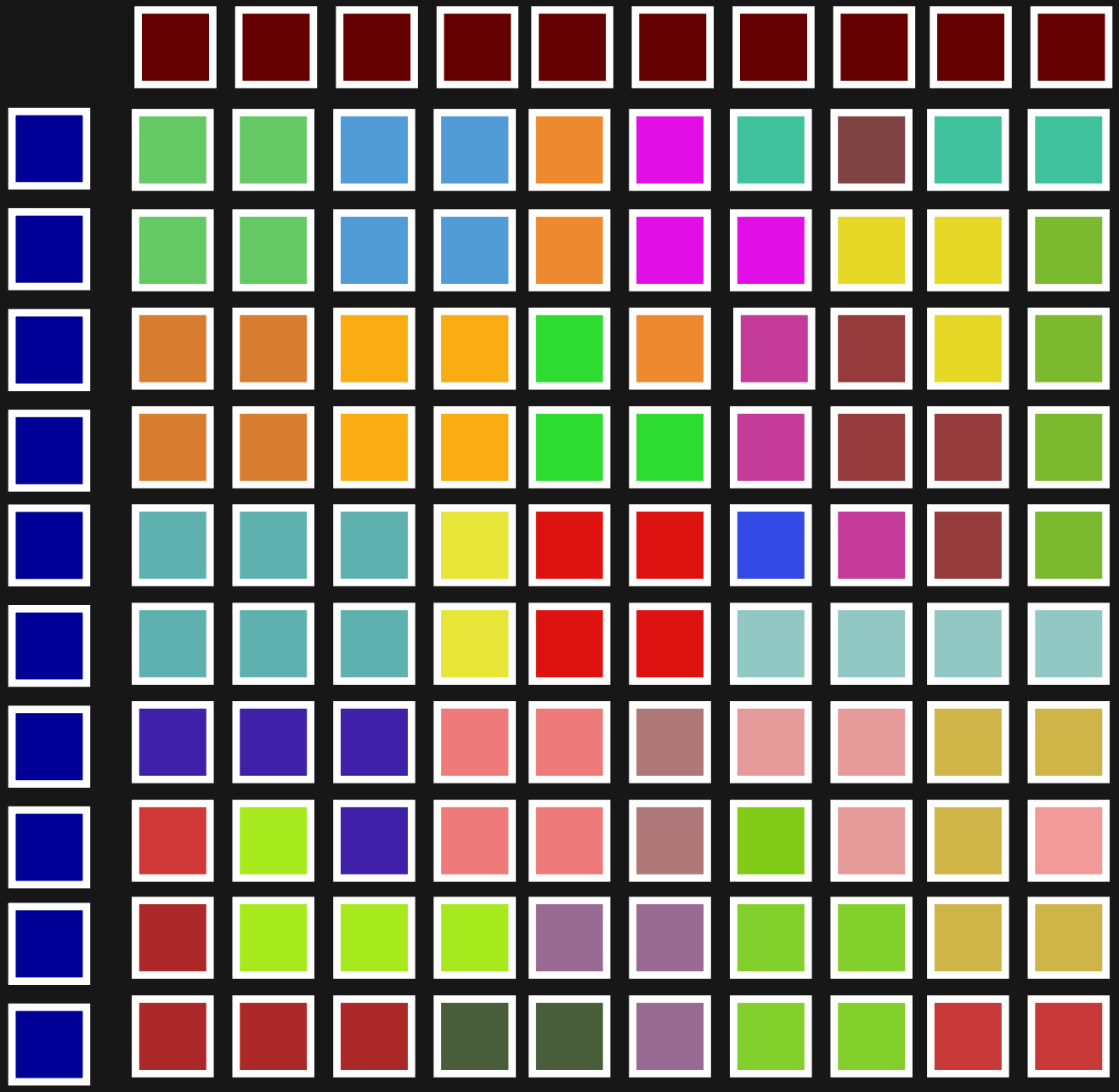
Unachievable by any individual or small group.

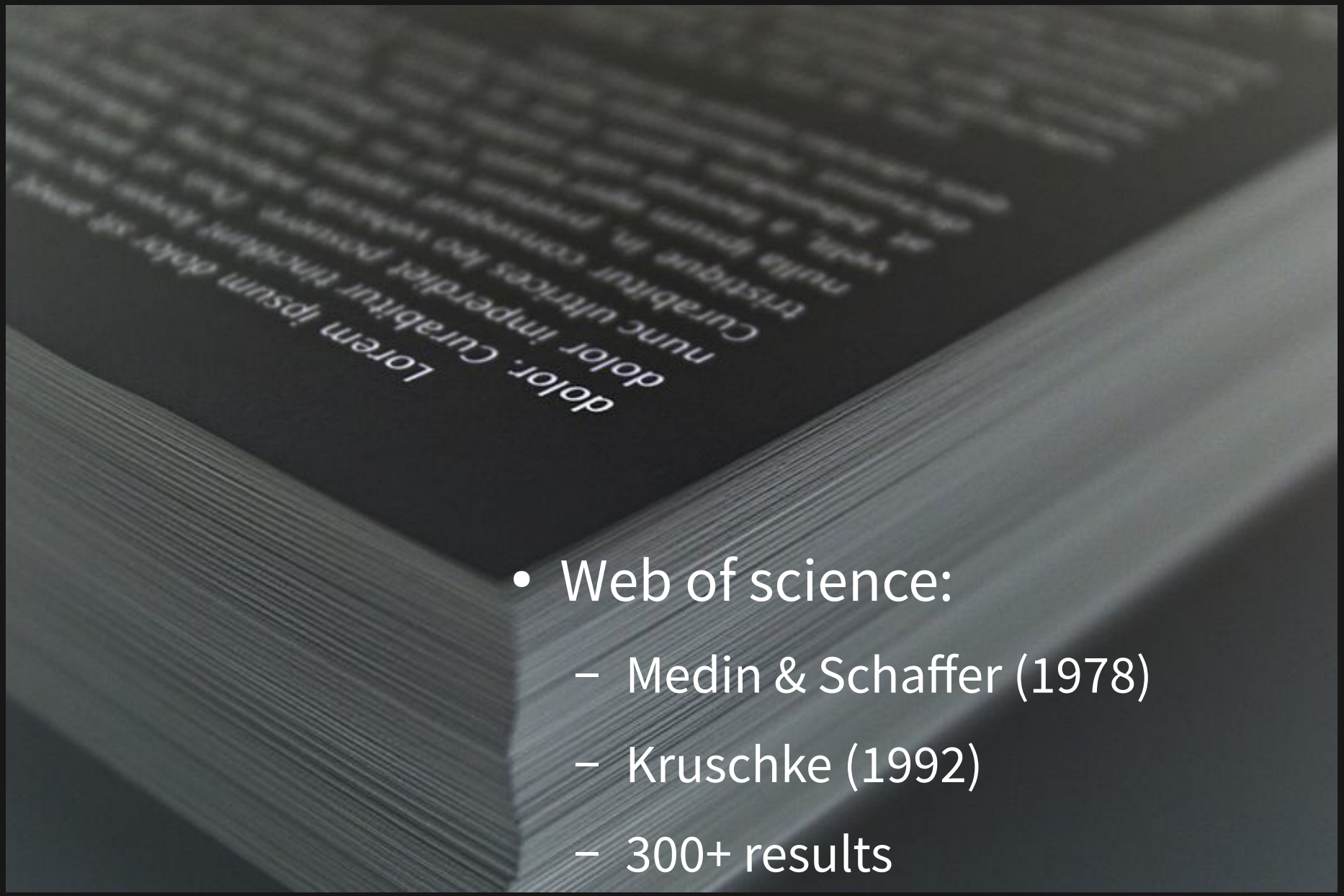
Framework for efficient distributed collaboration has been lacking.


ALCOVE
proto-ALCOVE
COVIS
S, H & J (1961)
1 0 1
M & S (1978)
1 0 NT









- Web of science:
  - Medin & Schaffer (1978)
  - Kruschke (1992)
  - 300+ results

Canonical **I**ndependently **R**eplicated **P**henomenon

Modelling non-real results:

- wastes time.
- distorts relative adequacy assessments.

Restrict relative adequacy comparisons to independently replicated phenomena.

Attempts to replicate are rare, and often unsuccessful.

What counts as a replication?

- As close to direct as the literature will allow.
- Ordinal success.
- Boundary conditions.

CIRP separate empirical phenomena from interpretations.

Model implementations are often not publicly available.

This leads to a substantial waste of effort … but **catlearn** is beginning to change this.

ALCOVE, COVIS, DIVA, Rescorla-Wagner

http://catlearn.r-forge.r-project.org/

Framework for distributed collaboration in the formal modelling of psychological processes.

Why R?

- non-linear optimization
- familiar environment
- easy to learn
- integrated documentation

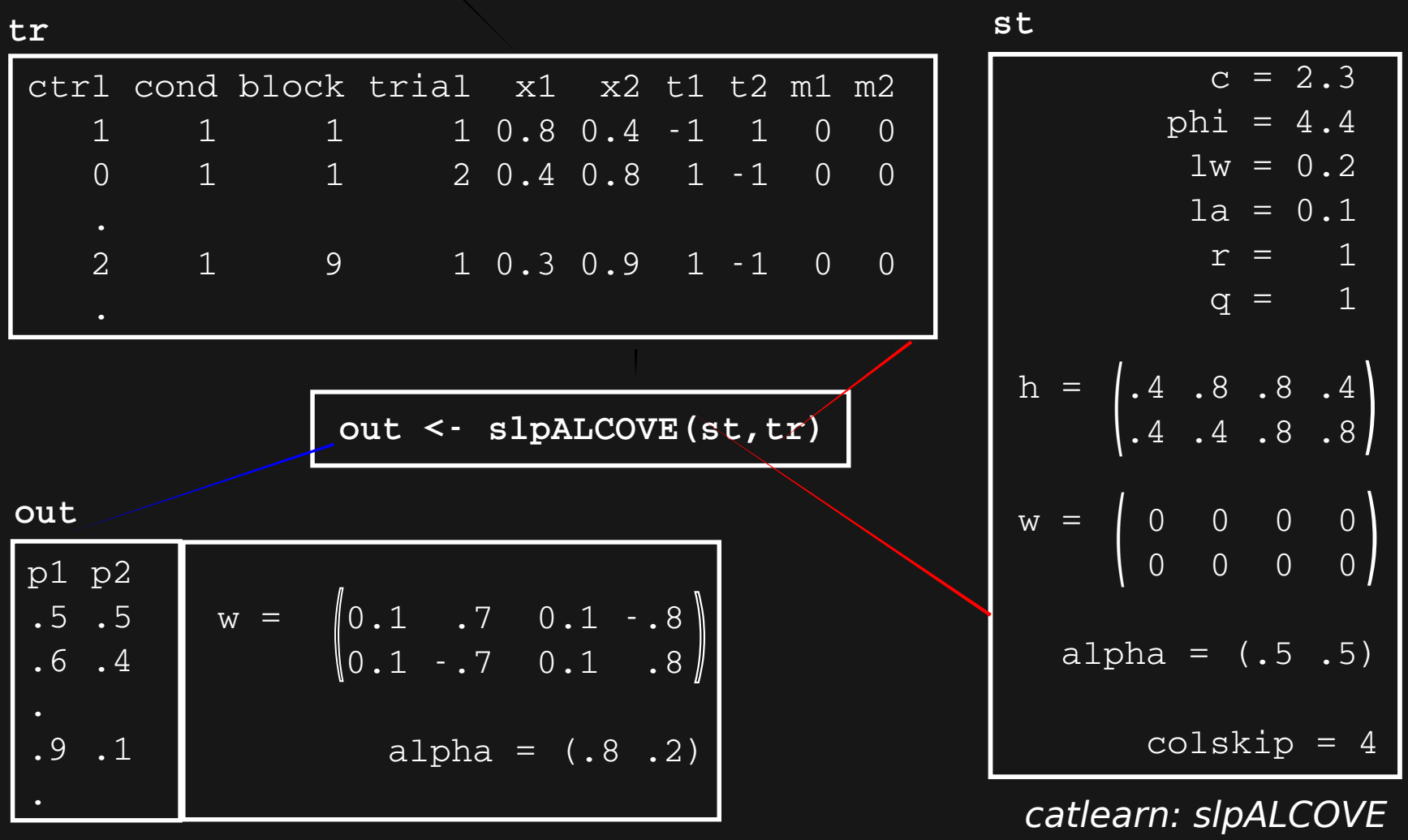
Slow?

- often instantaneous
- Model implementations can be compiled for speed
  (e.g. C++, FORTRAN)

http://catlearn.r-forge.r-project.org/

catlearn is a documented open archive of

- Model implementations (as stateful list processors)
- CIRP
- Simulations



catlearn: slpALCOVE

Live demo...

Efficient progress in the formal modelling of psychological processes requires a technical framework for distributed collaboration.

catlearn provides a free open-source framework.

Collaboration and cultural change.

Acknowledgements

Charlotte Edmunds, Garrett Honke, Angus Inkster, Garret O'Connell, Stuart Spicer.

Further reading

Wills, A.J., O'Connell, G., Edmunds, C.E.R., & Inkster, A.B. (2017). Progress in modeling through distributed collaboration: Concepts, tools, and category-learning examples. *Psychology of Learning and Motivation.*

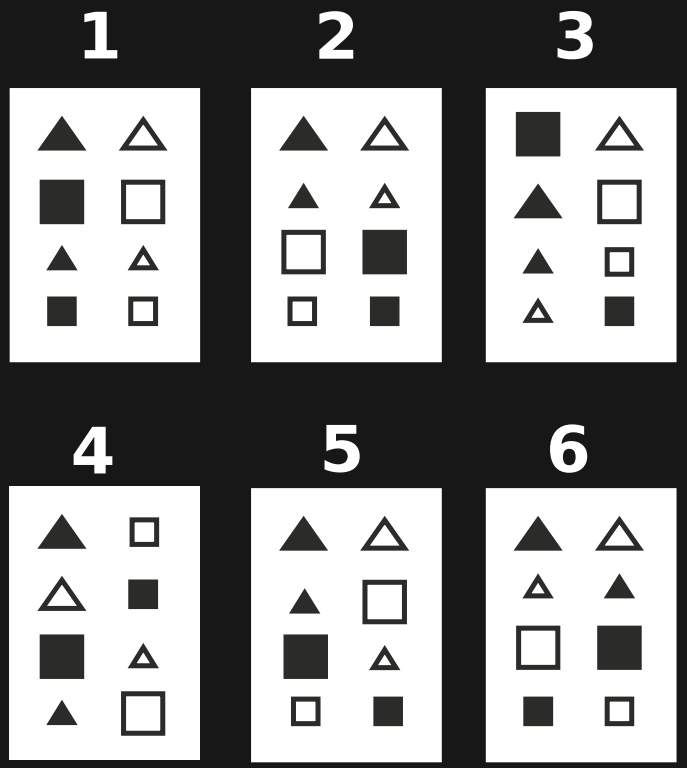
Coming in Oct 2017: **catlearn** v. 0.5. ATRIUM, Bush-Mosteller, MB/MF, and much more!

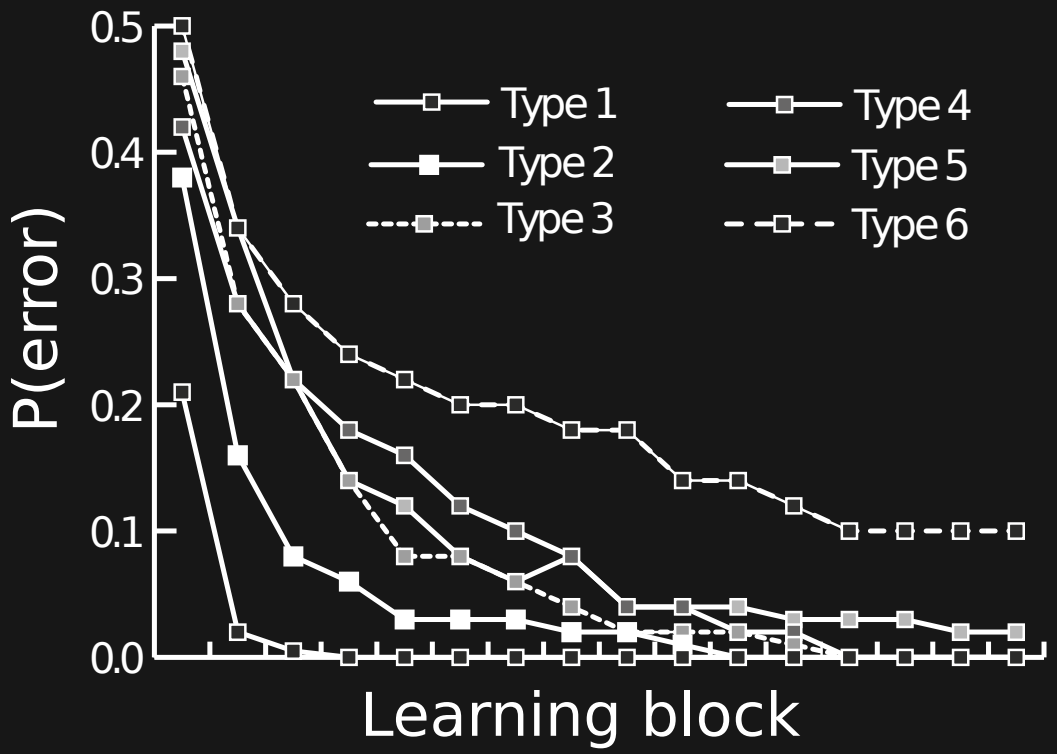
Installation instructions: http://catlearn.r-forge.r-project.org/

Software install

- Install R
- Install Rstudio
- Install **catlearn**
- Make a copy of the *resources* directory.

CIRP


1
2
3
4
5
6


0.5
0.4
0.3
0.2
0.1
0.0
P(error)
Learning block
Type 1
Type 2
Type 3
Type 4
Type 5
Type 6

1 < 2 < [3,4,5] < 6

CIRP

- Independent replications:
  - Shepard et al. (1961)
  - Nosofsky et al. (1994)
  - Smith et al. (2004)
- II < IV may depend on verbalizability of dimensions and instructions to look for rules (Kurtz et al., 2013)
- Nosofsky et al. (1994) chosen as CIRP because it has high N, and published simulations with various models.

Adding a CIRP

- Scrape data (e.g. from PDF) into spreadsheet; use long format:

type	block	error
1	1	0.211
1	2	0.025
1	3	0.003

- Save as CSV, load CSV into R:

```
nosof94 = read.csv('temp.csv', stringsAsFactors = FALSE)
```

- Save out as Rdata, into data directory:

```
save(nosof94, file='nosof94.RData')
```

Documenting a CIRP

```
\name{  name of data frame  }
\docType{data}
\alias{   name of data frame }
\title{  One-Line Title  }
\description{  One paragraph summary   }
\usage{ data(     ) }
\format{ A data frame with the following columns:
    \describe{
        \item{    }{    }
        \item{    }{    }
    }
}
\details{  see next slide }
\source{ Reference for data source }
\references{ Other references }
\author{ Name of author of Rd document \email{ email addr } }
```

What to put in \details

- Citation of review article that contains full derivation of CIRP.
- Citations establishing effect is independently replicated.
- Reason for selecting this data set as CIRP amongst the replications.
- Any further brief details you deem relevant about the study.

Writing an OAT

- OAT function tests whether a simulation captures the ordinal properties of a dataset, as defined in the CIRP.
  - Takes in data in same format, order, and column headings as the CIRP
  - Applies test(s)
  - Returns a single TRUE or FALSE
- For example, this by definition returns TRUE:

  ```
  nosof94oat(nosof94)
  ```

Coding clues

Mean p(error) by condition	`aggregate`
Test for inequality	`if(x < y)` *do_something*
Set a Boolean (TRUE, FALSE)	`Bool = TRUE, Bool = FALSE`
Might need a NOT operator?	`if(!(x < y))` *do_something*

Define a function that takes data and returns a value	`dummyoat = function(dta) {`
	`  `*do_something*
	`  return(`*something*`)`
	`}`

Loading a function saved into ‘dummyoat.R’	`source(‘dummyoat.R’)`

Demo of next steps…

Closing thoughts

- Which CIRP would you like to see in **catlearn**?
- Which model implementations?
- Keep in touch :-)

andy@willslab.co.uk