

Gender differences in the Stroop Colour-Word task: a meta-analysis

Espen Sjoberg

Geoff Cole

BLACK

RED

YELLOW

WHITE

BLUE



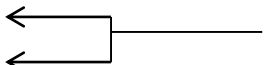
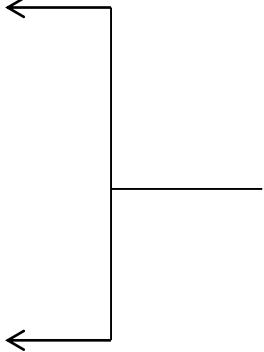
University of Essex

The Stroop task

- First paper published in 1935 (though similar experiments existed before this).
- Consists of several subtasks, but of interest to this meta-analysis is the Colour-Word subtask.

YELLOW BLUE ORANGE
BLACK RED GREEN
PURPLE YELLOW RED

Gender on the Stroop task

- Gender differences have never been systematically investigated, although papers do exist that have explicitly tested the hypothesis.
- First mention of gender differences in the original Stroop paper:
 - No significant gender effects were observed.
 - However, 9 effect sizes were possible to calculate:
 - **0.18** ←  *Used in meta-analysis*
 - **0.51** ←
 - 0.39 ←  *Practice trials*
 - 0.41
 - 0.58
 - 0.58
 - 0.27
 - 0.36
 - 0.34 ←
 - These all point to a female advantage of a moderate size.

Reasons for analysis

Theoretical:

- Female advantage exist due to superior verbal abilities (*verbal ability hypothesis*).
- Female advantage exist due to superior inhibition abilities (*evolved inhibition hypothesis*).

Practical:

- Some clinical studies use the Stroop task to assess inhibition abilities or cognitive function (e.g. eating disorders, ADHD, schizophrenia), under the assumption that no gender effects exist in the task.

Previous reviews

<u>Study</u>	<u>No. papers reviewed:</u>
• Jensen & Rohwer (1966) <ul style="list-style-type: none">– “Significant sex differences on [...] CW have not been found in any study (p.65)”	7
• MacLeod (1991) <ul style="list-style-type: none">– “There are no significant sex differences in Stroop interference at any age (p.184)”	19
• Bjorklund & Kipp (1996) <ul style="list-style-type: none">– “Researchers from the majority of studies investigating the Stroop phenomenon reported no gender differences (p. 178)”	11
• Rovainen (2001) <ul style="list-style-type: none">– “The effects of gender are inconsistent (p.147)”	2
• Mitrushina et al. (2005) <ul style="list-style-type: none">– “...effects of [...] gender on Stroop performance are equivocal (p.133)”	30

Previous studies

These experiments had gender differences as one of their main hypothesis.

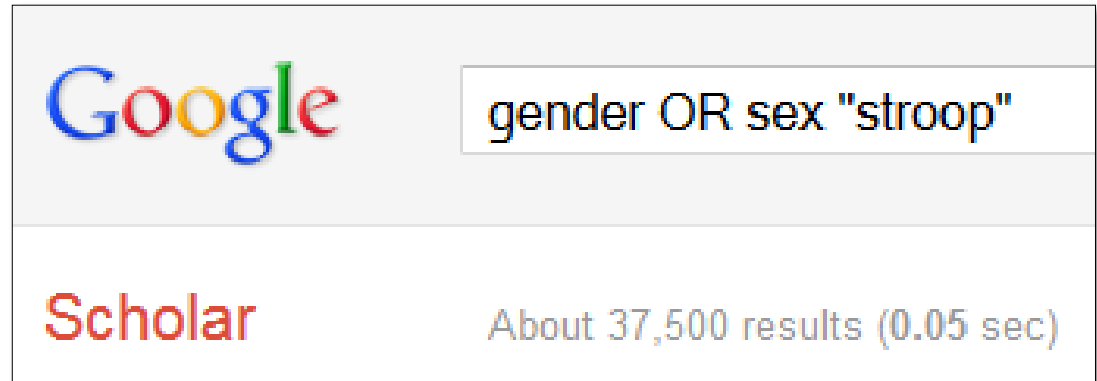
Study	Finding (sig)	N	Avg. ES
• Peretti (1969)	None	300	0.24
• Peretti (1971)	Female	100	1.26
• Golden (1974)	Female	438	0.18
• Sarmany (1977)	Female	57	0.89
• Cohen & Fischer (1980)	None	144	0.34
• Sladekova & Daniels (1981)	Inconsistent	268	0.42
• Wolff et al. (1993)	Inconsistent	130*	0.60 / NA
• Wolf & Gow (1986)	Inconsistent	174	0.30
• Pati & Dash (1990)	Inconsistent	144	0.41
• Singh (1991)	Inconsistent	100	0.61
• Von Kluge (1992)	Male	128	-1.02
• Mekarski et al. (1996)	Female	16	NA
• Owens & Broida (1998)	Male	63	-1.23
• Daniel et al. (2000)	Inconsistent	120	0.54
• Alansari & Baroun (2004)	None	210	0.16
• Baroun & Alansari (2006)	Female	504	0.27
• Sjoberg (in prep)	Female	63	0.51

Data collection

Huge number of Stroop studies available!

Data collected via:

- Google Scholar
- SAGE
- PsychArticles
- References in other papers
- Contacting Stroop authors (response rate 68%).



- The vast majority of studies had gender effects as a secondary hypothesis.

Getting a representative sample

Selection criteria:

1. Only use data from healthy participants (*control groups in diagnosed patients acceptable*).
2. Exclude experimental manipulations that is likely to affect results (*e.g. alcohol, competitiveness, anxiety*).
3. Exclude practice trials (*first trial is acceptable*).
4. Only use the classic CW version of the Stroop task (*e.g. exclude picture-word or emotional Stroop versions*).

Stroop versions

Version	Items used	Measurement	Trials	Comment
Single trial	1	Reaction time	CW, WC	Reaction time is the time from stimulus onset to verbal response. Number of trials varies from study to study.
Bohnen	100	Reaction time	W, CN, CW, SCW	CN trial uses blocks
Comalli	100	Reaction time	W, CN, CW	SCW is identical to CW trial except some trials have coloured rectangles around the word
Daniel	100	Reaction time	W, CN, CW	CN trial uses rectangles
Dodrill	176	Reaction time	WC, CW	Participants wear headphones and listen to colour names, 75% of which are incongruent to the trial.
Golden	100	No. of items	W, CN, CW	CN trial uses "XXXX"
Graf	27	Reaction time	W, CN, CG, CW	CN trial uses "XXXX"
Kaplan	100	Reaction time	W, CN, CW	CN trial uses blocks
Malayalam	40	Reaction time	W, CN, CW	CN trial uses squares
Stroop	100	Reaction time	W, CN, WC, CW	CN trial uses strips
Trenerry	112	No. of items	WC, CW	CN trial uses dots
Victoria	24	Reaction time	CN, CNW, CW	CNW trial involves naming the ink colour of neutral words

Summary of literature search

- Approx. 1200 abstracts screened
- 311 articles screened
- 134 articles relevant to the topic
- 60 articles relevant to meta-analysis
 - 14 of these articles did not report sufficient data, but the missing data was generously provided by the respective authors.

Preliminary observation

Looking at significance only, the following pattern emerges:

- No difference 74%
- Female advantage 22%
- Male advantage 4%

Based on 236 trials

Considering how the majority of studies find no significant gender difference, it is easy to see how previous reviews have reached a similar conclusion.

However, if we look at the studies which reported no difference, but where effect sizes was possible to calculate, 77.45% of them had an effect size in favour of women.

Summary of final effect size count

60 articles for the quantitative synthesis

- 56 published articles
- 2 unpublished articles
- 1 Master's thesis
- 1 PhD thesis

Total of effect sizes: **126**

- Of these 58 were only possible with the help of authors
- 21,314 participants (53% women)
- 21 countries, from 4 continents
- Ages 7-92

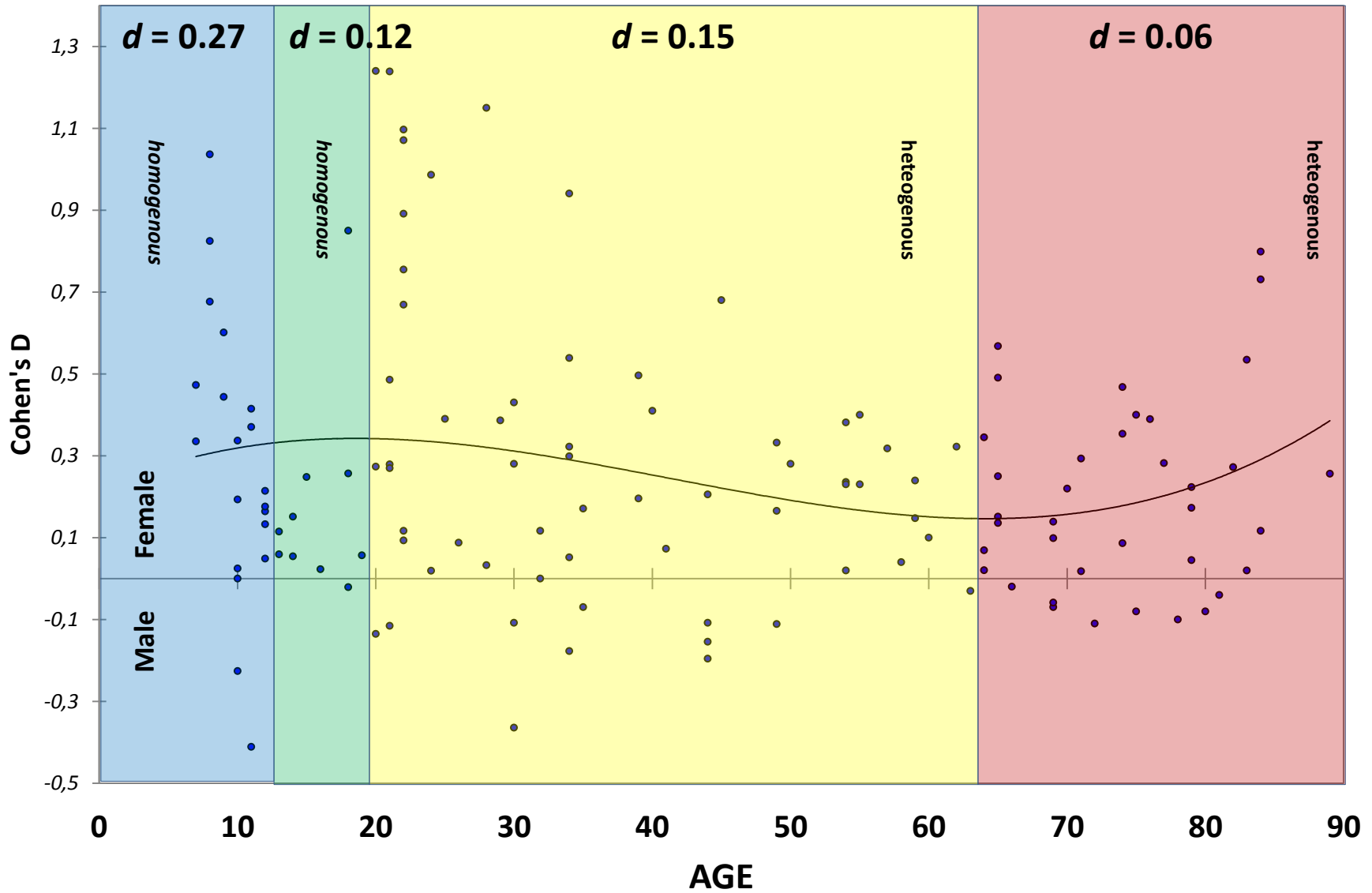
Results

Overall weighted effect size:

$$**d = 0.12, p < .0001**$$

- Resistant to the file drawer problem
- Was not homogenous, though partitioning into age and Stroop version made all subcategories homogenous.

Age



All $p < .01$ or lower

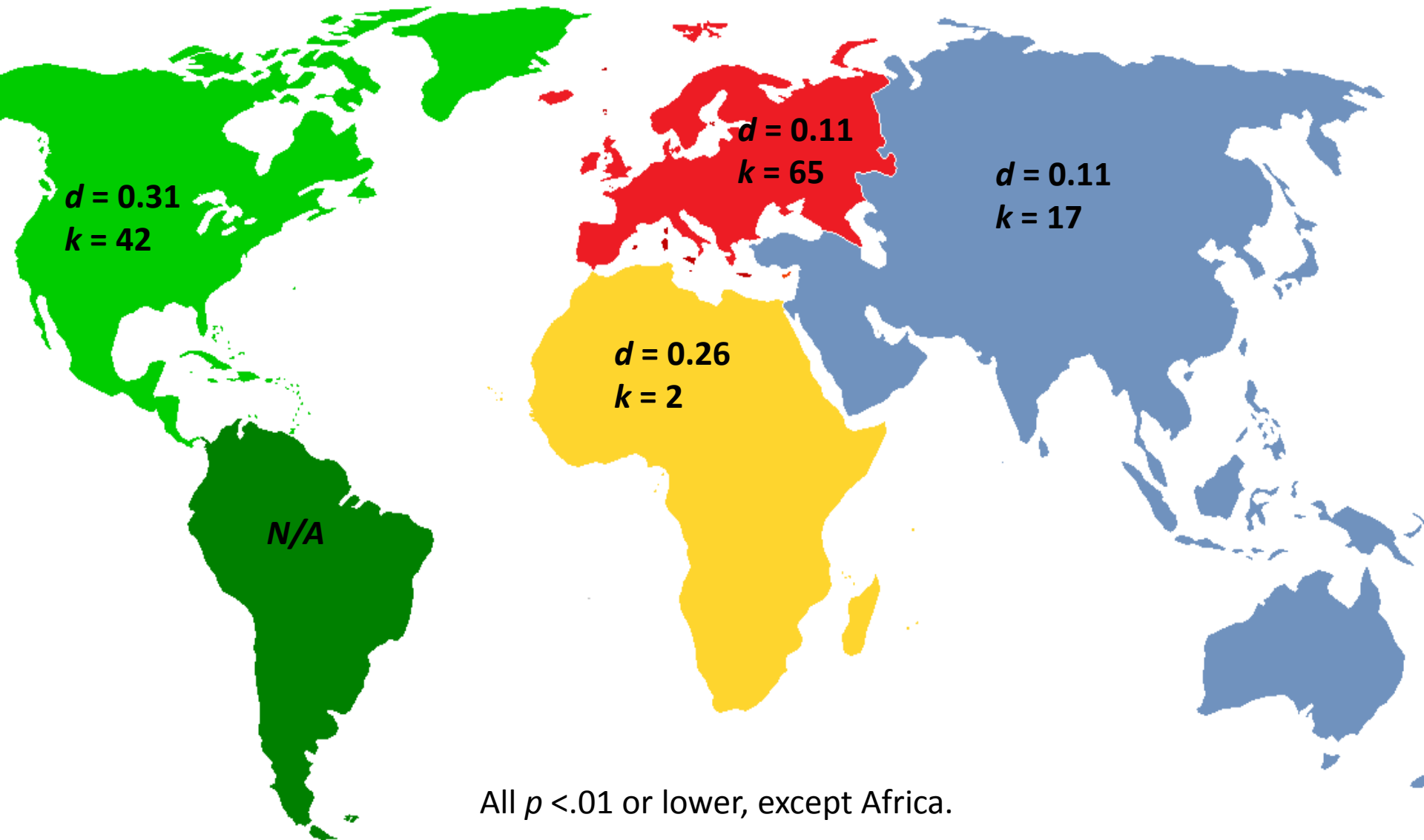
Culture

Partitioning into ethnicity was not possible:

Only 8 trials reported the ethnicity of participants.

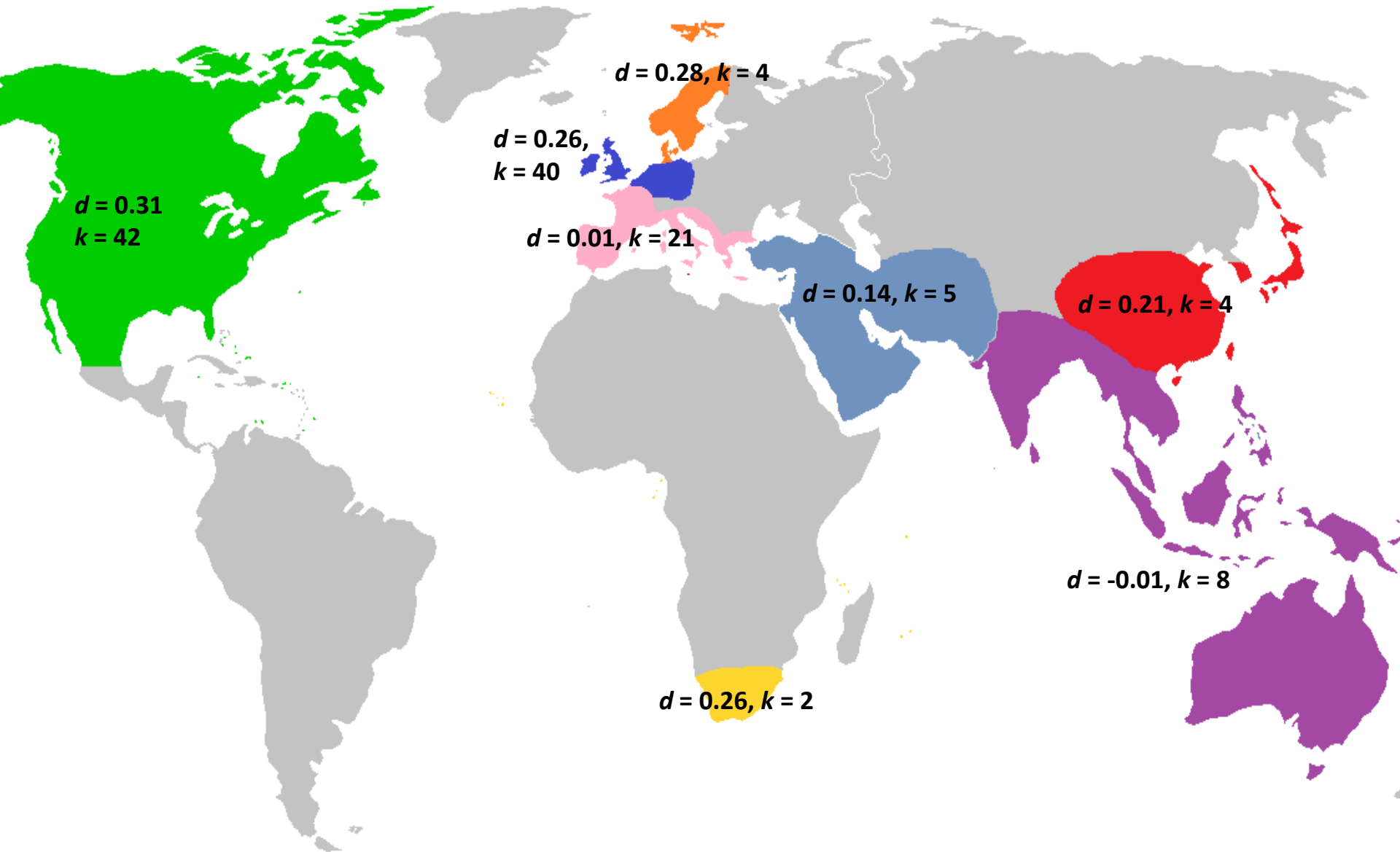
- Instead, partitioning was made into continents.

Culture



All $p < .01$ or lower, except Africa.

Culture



Stroop versions

	Stroop version		<i>k</i>	Weighted effect size (<i>d</i>)	Significance (<i>Z</i>)	Homogeneity statistic (χ^2)
1-item	Single trial	<i>d</i> = 0.48	7	0.476	5.433*	11.038†
	Comalli		10	0.231	4.197*	10.515†
RT	Daniel	<i>d</i> = 0.27	3	0.478	3.034*	6.839‡
	Victoria		10	0.161	1.909	6.119†
	Kaplan		3	0.222	1.333	0.325†
	Malayalam		3	-0.455	-3.999*	15.323*
	Stroop		43	0.234	8.488*	63.294‡
Counting	Trenerry	<i>d</i> = 0.04	5	0.030	0.436	8.653†
	Golden		29	0.043	2.190*	49.319‡
	Other/Custom		13	0.160	2.193*	49.967*

* $p < .05$

† homogeneity achieved

‡ near-homogeneity achieved

Exploratory: Negative Priming

Study	Stroop version	Country	Mean age	M	F	<i>d</i>
<i>Christiansen and Oades (2010)</i>	Single trial	Germany	11	15	22	0.48
<i>Steel, Hemsley, and Jones (1996)</i>	Single trial	United Kingdom	28	19	17	0.51
<i>Visser, Das-Smaal, and Kwakman (1996)</i>	Stroop	Netherlands	10	98	112	0.35*
<i>Sjoberg & Cole (in prep)</i>	Stroop	United Kingdom	30	32	32	0.51*
<i>* significant gender difference</i>	<i>k = 4</i>			N: 164	183	

***d* = 0.42, *p* < .05, homogenous**

Summary

- First meta-analysis of gender effects on healthy adults using the Stroop Colour-Word task.
 - 60 studies, 126 effect sizes, weighted $d = 0.12$
 - Persistent across continents and ages
 - No difference in counting Stroop versions
 - Small difference in reacting time versions
 - Moderate difference in single-item versions
 - Female advantage most likely due to verbal abilities

Huge thanks to Geoff Cole, Rachael Wilner, Rick O’Gorman, Eva Sjoberg, and David Bjorklund for their support and contributions.