

**{rant}**

**The methods don't matter**

# The woman who wasn't there: Converging evidence that subliminal social comparison affects self-evaluation

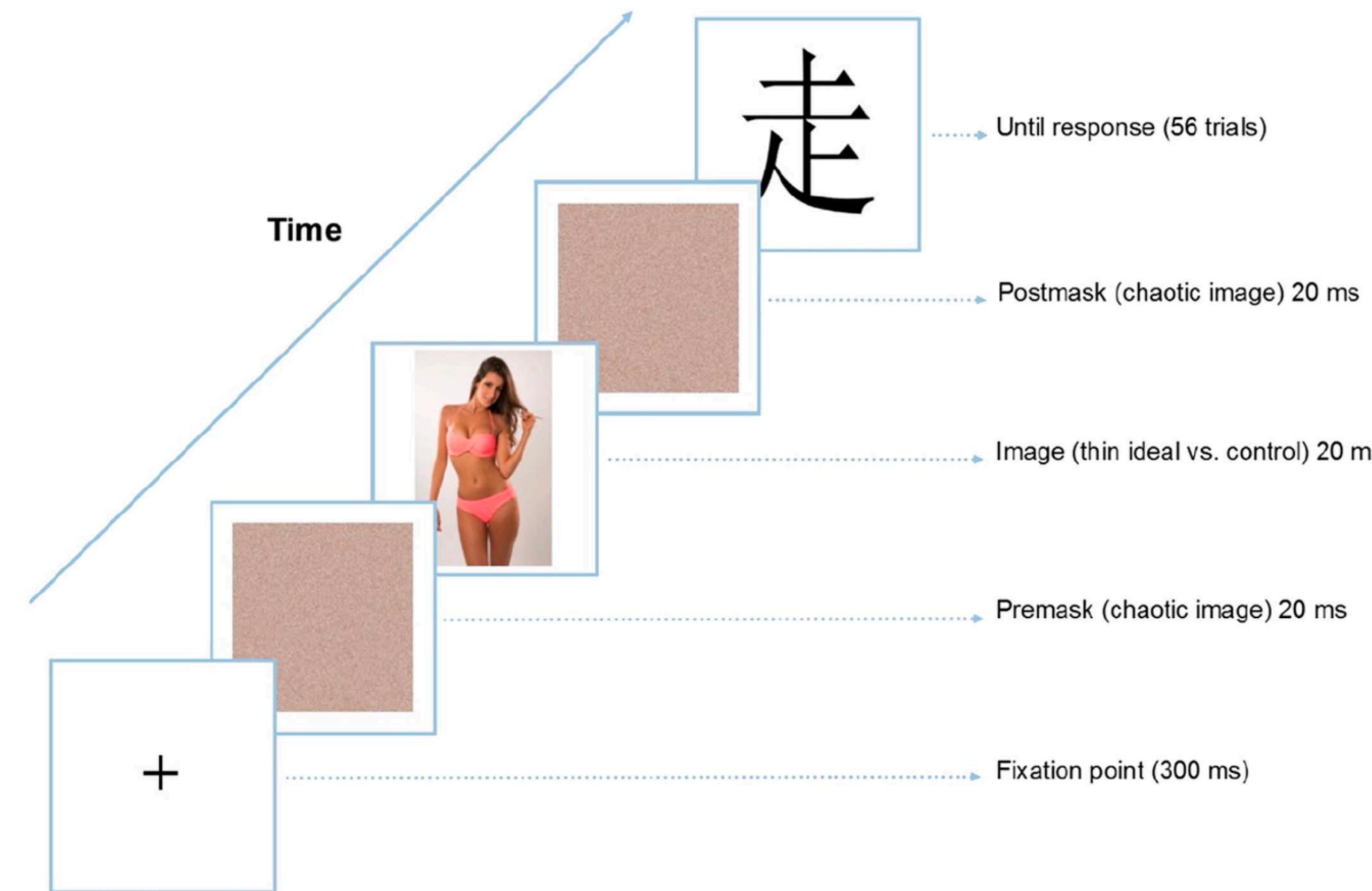
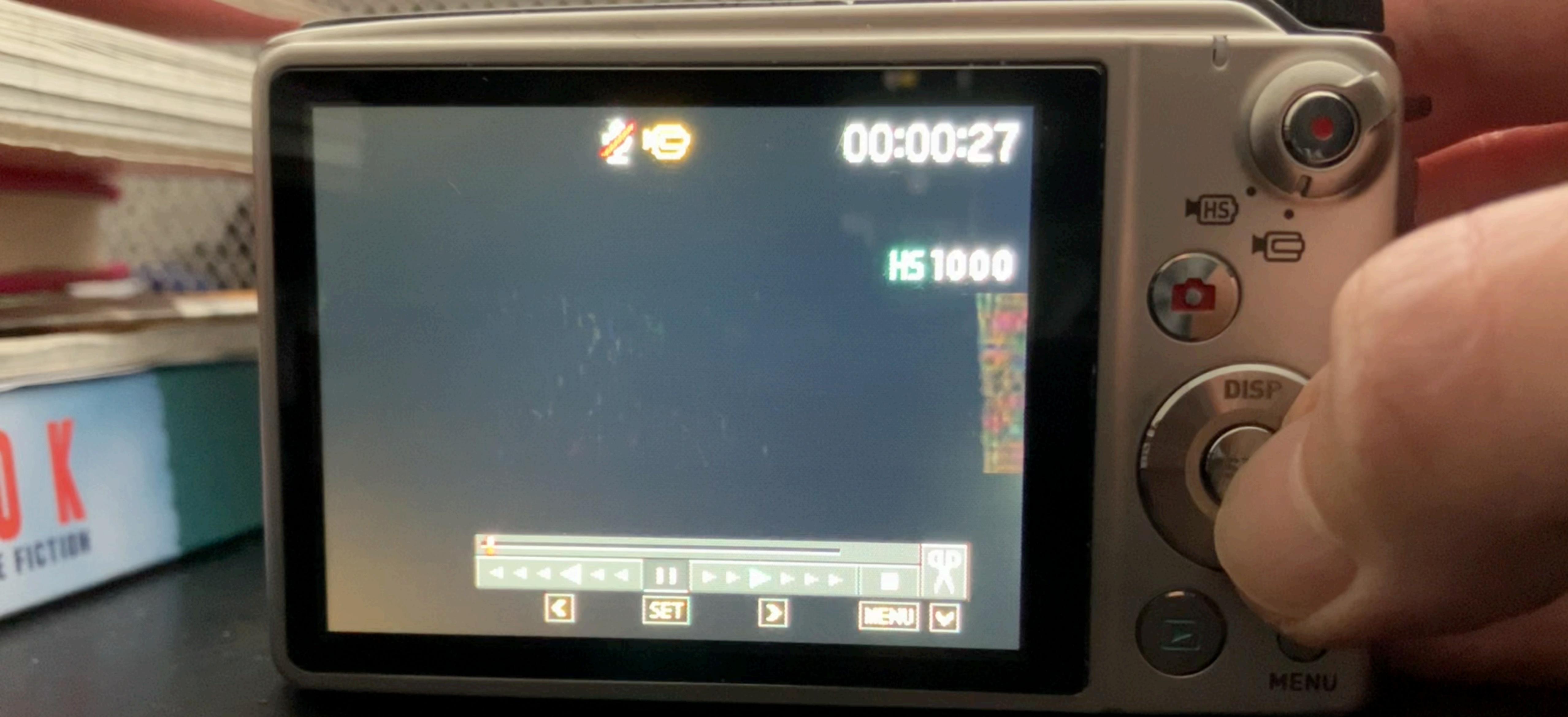


Fig. 1. Illustration of the subliminal procedure used to prime body images in Experiments 1 and 2.



# **The woman who wasn't there**

# The woman who wasn't there



“It has come to our attention that some stimuli may not have been presented correctly due to a coding error.”

**Assignment**

# **Statistical red flags**

Ian Hussey

Digitalisation of Psychology

# **What can you look for?**

## **A checklist of statistical red flags**

- 🚩 Reliance on  $p$  values between between .02 and .05, especially if one-tailed
- 🚩 Reliance on  $p$  values only with no effect sizes (standardized or unstandardized).
- 🚩 Subgroup analyses
- 🚩 Use of mediation analysis (or variants, eg mediated moderation, moderated mediation, etc: use of the PROCESS macro generally)
- 🚩 Use of MANOVA
- 🚩 3+ way interactions in ANOVA/ANCOVA/MANOVA

# **What can you look for?**

## **A checklist of statistical red flags**

- 🚩 Use of Stepwise Regression
- 🚩 Covariate selection on the basis of  $p$  values or effect sizes rather than causal justification
- 🚩 Post hoc power analysis
- 🚩 A priori power analysis but for a different test than the one conducted (e.g., power analysis for ANOVA but ran mixed models)
- 🚩 Conditioning on a post-treatment variable in a randomized experiment, eg excluding on the basis of a manipulation check

	category	percent_error_in determinable	percent_error_no	percent_error_yes
	barely_sig_p_values	19.3	43.9	36.8
	no_effect_sizes	5.3	61.4	33.3
	subgroup_analyses	NA	83.9	16.1
	mediation	NA	91.2	8.8
	manova	1.8	98.2	NA
	multi_way_interactions	NA	89.5	10.5
	stepwise_regression	1.9	98.1	NA
	poor_covariate_selection	13.0	81.5	5.6
	post_hoc_power	10.5	89.5	NA
	wrong_power	11.1	88.9	NA
	conditioning_on_post_treatment_variable	3.2	96.8	NA

**GRIM & GRIMMER**

# **GRIM**

Explain it to me

What did Brown & Heathers (2017) find?

# GRIM

I have 2 participants.

I calculate their mean age.

I round their mean age to 1 decimal place to report in a manuscript.

Without knowing either of their ages,  
we know the rounded mean age of 2 participants must end in .0 or .5

Possible: Mean age = 23.5

Impossible: Mean age = 23.1

# GRIM

GRIM test is consistent when:

`reported_mean * n_participants * n_items = total_sum`

`round(total_sum, 0) / (N_participants * N_items) = recalculated_mean`

`round(recalculated_mean, n_digits) == reported_mean`

“The 44 participants had a mean age of 22.14 (SD = 5.34)”

# Exercise 1

Fifty-two university students were asked “How likely are you to use ChatGPT in your coursework?”, with responses given via a 1–7 Likert-type scale (1=very unlikely; 7=very likely). Results showed a high likelihood ( $M = 5.22$ ,  $SD = 1.33$ ).

- Is the mean GRIM Consistent?

# Exercise 1

Fifty-two university students were asked “How likely are you to use ChatGPT in your coursework?”, with responses given via a 1–7 Likert-type scale (1=very unlikely; 7=very likely). Results showed a high likelihood ( $M = 5.22$ ,  $SD = 1.33$ ).

- Is the mean GRIM Consistent?
- What would happen if the scale was 1 to 10?

# Exercise 2

Festinger & Carlsmith (1959) Cognitive Dissonance Theory

*Cited roughly 6000 times*

- Are the means GRIM Consistent?
- <http://nickbrown.fr/GRIM>

AVERAGE RATINGS ON INTERVIEW QUESTIONS FOR EACH CONDITION

Question on Interview	Experimental Condition		
	Control (N = 20)	One Dollar (N = 20)	Twenty Dollars (N = 20)
How enjoyable tasks were (rated from -5 to +5)	-.45	+1.35	-.05
How much they learned (rated from 0 to 10)	3.08	2.80	3.15
Scientific importance (rated from 0 to 10)	5.60	6.45	5.18
Participate in similar exp. (rated from -5 to +5)	-.62	+1.20	-.25

# Exercise 2

Festinger & Carlsmith (1959) Cognitive Dissonance Theory  
*Cited roughly 6000 times*

- Are the means GRIM Consistent?
- <http://nickbrown.fr/GRIM>

AVERAGE RATINGS ON INTERVIEW QUESTIONS FOR EACH CONDITION

Question on Interview	Experimental Condition		
	Control (N = 20)	One Dollar (N = 20)	Twenty Dollars (N = 20)
How enjoyable tasks were (rated from -5 to +5)	-.45	+1.35	-.05
How much they learned (rated from 0 to 10)	3.08	2.80	3.15
Scientific importance (rated from 0 to 10)	5.60	6.45	5.18
Participate in similar exp. (rated from -5 to +5)	-.62	+1.20	-.25

# **GRIMMER**

Just like GRIM, but for Standard Deviations (SDs)

Math is more complex but principle is the same: only some values are possible

# Interpretation

What do GRIM/MER inconsistencies mean?

- Maybe you made an error
  - Extracted the wrong number, made a typo etc
  - Double check
- Maybe the authors made an error
  - Reported the wrong number, made a typo, rounded it inappropriately
- Maybe the authors changed some of the real values inappropriately
  - Started with the real results, changed some values to make them more favourable (fabrication type 1)
- Maybe there was no real data at all
  - Maybe the numbers are completely invented (fabrication type 2)

# Remember

- Understand how to use “N items”
- Understand when the GRIM test will always consistent, even when the reported values are errors or fabrications:
  - When the mean is reported to 1 decimal places (eg  $M = 4.4$ ), GRIM stops working at  $N_{participants} * N_{items} = 10$
  - When the mean is reported to two decimal places (eg  $M = 4.37$ ), GRIM stops working at  $N_{participants} * N_{items} = 100$
  - Etc.
  - **This makes GRIM useless above these values!**

# Exercise 3

- <https://errors.shinyapps.io/scrutiny/>
- We'll use my example data set

## *Readings*

- Foody, M., Barnes-Holmes, Y., Barnes-Holmes, D., & Luciano, C. (2013). An empirical investigation of hierarchical versus distinction relations in a self-based ACT exercise. *International Journal of Psychology and Psychological Therapy*, 13(3), 373-388.
  - Find everything weird, wrong, or questionable that you can in this article.
  - PDF on Ilias.

## *Assignment*

Use the collaborative Google Sheet to code GRIM & GRIMMER tests for your articles

- The articles assigned to you for StatCheck
- Use <https://errors.shinyapps.io/scrutiny/>
- Use the template “summary\_statistics.csv” file
- Full instructions will be on Ilias

# Questions?