

Still suspicious: The suspicious coincidence effect revisited

Molly L. Lewis^{1,2} & Michael C. Frank³

¹ Computation Institute, University of Chicago

² Department of Psychology, University of Wisconsin, Madison

³ Department of Psychology, Stanford University

*To whom correspondence should be addressed. E-mail: mollylewis@uchicago.edu

Author Note

Correspondence concerning this article should be addressed to Molly L. Lewis. E-mail: mollylewis@uchicago.edu

Abstract

Enter abstract here. Each new line herein must be indented, like this line.

Keywords: keywords

Word count: X

Still suspicious: The suspicious coincidence effect revisited

Intro

What is the suspicious coincidence effect?

(Spencer, Perone, Smith, & Samuelson, 2011; F. Xu & Tenenbaum, 2007; Fei Xu & Tenenbaum, 2007)

Why is it important?

Spencer et al. paper

Methodological differences:

- simultaneous vs. sequential
- 3-1 vs. 1-3
- blocking
- same label vs. different label

other evidence relevant on this replication

Our current paper reports 10 pre-registered experiments. We recover the suspicious coincidence effect with a large effect size in both sequential and simultaneous presentation conditions. The effect only occurs, however, in experiments where the trial with one exemplar is presented *before* the key trial with three subordinate-consistent exemplars (the “suspicious coincidence”). We attribute this difference to participants’ awareness of the possibility of subordinate generalizations following the three-exemplar trial; in these conditions, we see a high level of subordinate generalizations even for the one-exemplar trial (leading to the absence of a difference between conditions). In sum, and contra SPSS, the “suspicious coincidence” effect is robust to sequential presentation. The effect is sensitive to some features of the general experimental context, however, suggesting a potential interpretation in terms of the pragmatics of the task.

Methods

We report how we determined our sample size, all data exclusions (if any), all manipulations, and all measures in the study.

Participants

50 participants were recruited on Amazon Mechanical Turk for each of our 12 experiments ($N = 600$). Participants were paid 40 cents for their participation

Our sample size was determined based on a pre-registered power calculation using a meta-analytic estimate of the effect size from studies conducted by XT and SPSS (see SI for more details). The chosen sample size was approximately twice the estimated sample size necessary to obtain a power of 1.

- Exclusions?

run: 3-same-sequential-random/1-dif-sequential-blocked

9: same-blocked-3-seq [grey triangle -> circle] 10: dif-random-1-seq [black circle]

Stimuli

Our stimuli closely resembled that of XT and SPSS. The pictures consisted of three sets of 15 photos from different basic level categories (vegetables, vehicles and animals). Within each category, five were subordinate exemplars (e.g. green pepper), four were basic level exemplars (e.g. peppers), and six were superordinate exemplars (e.g. vegetables). The exemplars were divided into a learning and test set.

Novel labels

Procedure

Manipulations:.

- (1) Timing

- (2) Order
- (3) Blocking
- (4) Labels

Data analysis

We used R (3.4.1, R Core Team, 2017) and the R-packages *bindrcpp* (0.2, Müller, 2017), *broom* (0.4.2, Robinson, 2017), *compute.es* (0.2.4, Re, 2013), *dplyr* (0.7.2, Wickham, Francois, Henry, & Müller, 2017), *forcats* (0.2.0, Wickham, 2017a), *ggplot2* (2.2.1, Wickham, 2009), *jsonlite* (1.5, Ooms, 2014), *kableExtra* (0.4.0, Zhu, 2017), *knitr* (1.17, Xie, 2015), *langcog* (0.1.9001, Braginsky, Yurovsky, & Frank, n.d.), *Matrix* (1.2.10, Bates & Maechler, 2017), *metafor* (2.0.0, Viechtbauer, 2010), *papaja* (0.1.0.9492, Aust & Barth, 2017), *png* (0.1.7, Urbanek, 2013), *purrr* (0.2.3, Henry & Wickham, 2017), *readr* (1.1.1, Wickham, Hester, & Francois, 2017), *rmarkdown* (1.6, Allaire et al., 2017), *stringr* (1.2.0, Wickham, 2017b), *tibble* (1.3.3, Müller & Wickham, 2017), *tidyr* (0.6.3, Wickham, 2017c), and *tidyverse* (1.1.1, Wickham, 2017d) for all our analyses.

Results

Discussion

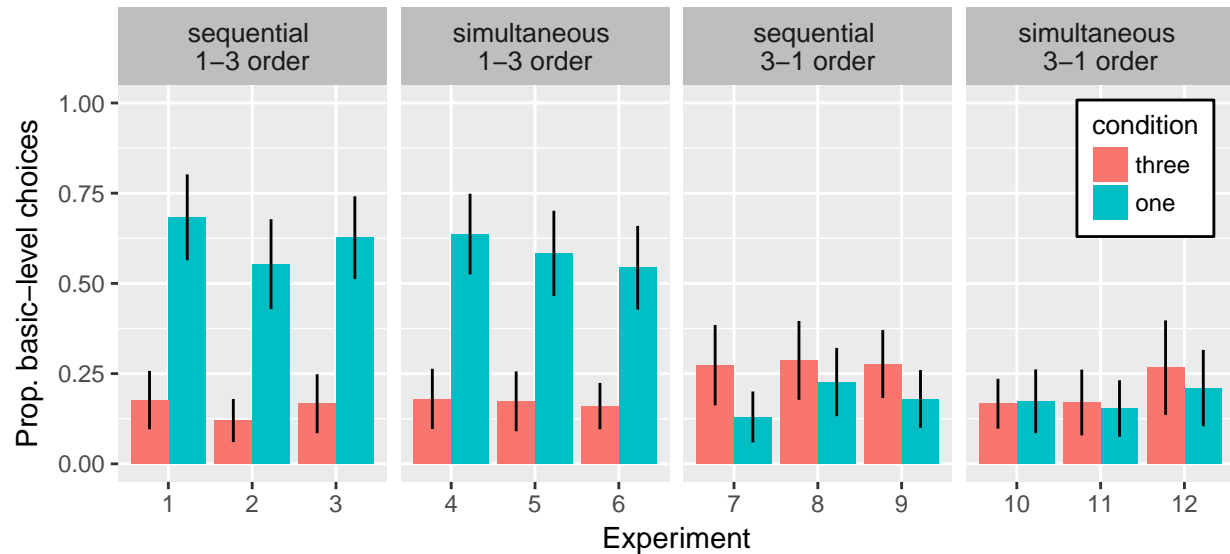


Figure 1. Mean proportion generalization to basic level exemplars in the one (green) and three (pink) subordinate exemplar conditions for all 12 of our experiments. Each facet corresponds to a pairing of presentation timing (sequential vs. simultaneous) and trial order (1-3 vs. 3-1). Error bars are bootstrapped 95% confidence intervals.

References

- Allaire, J., Cheng, J., Xie, Y., McPherson, J., Chang, W., Allen, J., ... Arslan, R. (2017). *Rmarkdown: Dynamic documents for r*. Retrieved from <https://CRAN.R-project.org/package=rmarkdown>
- Aust, F., & Barth, M. (2017). *papaja: Create APA manuscripts with R Markdown*. Retrieved from <https://github.com/crsh/papaja>
- Bates, D., & Maechler, M. (2017). *Matrix: Sparse and dense matrix classes and methods*. Retrieved from <https://CRAN.R-project.org/package=Matrix>
- Braginsky, M., Yurovsky, D., & Frank, M. (n.d.). *Langcog: Language and cognition lab things*. Retrieved from <http://github.com/langcog/langcog>
- Henry, L., & Wickham, H. (2017). *Purrr: Functional programming tools*. Retrieved from <https://CRAN.R-project.org/package=purrr>
- Müller, K. (2017). *Bindrcpp: An 'rcpp' interface to active bindings*. Retrieved from

- <https://CRAN.R-project.org/package=bindrcpp>
- Müller, K., & Wickham, H. (2017). *Tibble: Simple data frames*. Retrieved from <https://CRAN.R-project.org/package=tibble>
- Ooms, J. (2014). The jsonlite package: A practical and consistent mapping between json data and r objects. *arXiv:1403.2805 [Stat.CO]*. Retrieved from <https://arxiv.org/abs/1403.2805>
- R Core Team. (2017). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <https://www.R-project.org/>
- Re, A. C. D. (2013). *Compute.es: Compute effect sizes. R Package*. Retrieved from <http://cran.r-project.org/web/packages/compute.es>
- Robinson, D. (2017). *Broom: Convert statistical analysis objects into tidy data frames*. Retrieved from <https://CRAN.R-project.org/package=broom>
- Spencer, J. P., Perone, S., Smith, L. B., & Samuelson, L. K. (2011). Learning words in space and time: Probing the mechanisms behind the suspicious-coincidence effect. *Psychological Science*, 22(8), 1049–1057.
- Urbanek, S. (2013). *Png: Read and write png images*. Retrieved from <https://CRAN.R-project.org/package=png>
- Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, 36(3), 1–48. Retrieved from <http://www.jstatsoft.org/v36/i03/>
- Wickham, H. (2009). *Ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York. Retrieved from <http://ggplot2.org>
- Wickham, H. (2017a). *Forcats: Tools for working with categorical variables (factors)*. Retrieved from <https://CRAN.R-project.org/package=forcats>
- Wickham, H. (2017b). *Stringr: Simple, consistent wrappers for common string operations*. Retrieved from <https://CRAN.R-project.org/package=stringr>
- Wickham, H. (2017c). *Tidyr: Easily tidy data with 'spread()' and 'gather()' functions*.

- Retrieved from <https://CRAN.R-project.org/package=tidyr>
- Wickham, H. (2017d). *Tidyverse: Easily install and load 'tidyverse' packages*. Retrieved from <https://CRAN.R-project.org/package=tidyverse>
- Wickham, H., Francois, R., Henry, L., & Müller, K. (2017). *Dplyr: A grammar of data manipulation*. Retrieved from <https://CRAN.R-project.org/package=dplyr>
- Wickham, H., Hester, J., & Francois, R. (2017). *Readr: Read rectangular text data*. Retrieved from <https://CRAN.R-project.org/package=readr>
- Xie, Y. (2015). *Dynamic documents with R and knitr* (2nd ed.). Boca Raton, Florida: Chapman; Hall/CRC. Retrieved from <https://yihui.name/knitr/>
- Xu, F., & Tenenbaum, J. (2007). Word learning as Bayesian inference. *Psychological Review*, 114(2), 245.
- Xu, F., & Tenenbaum, J. B. (2007). Sensitivity to sampling in Bayesian word learning. *Developmental Science*, 10(3), 288–297.
- Zhu, H. (2017). *KableExtra: Construct complex table with 'kable' and pipe syntax*. Retrieved from <https://CRAN.R-project.org/package=kableExtra>

Table 1

Summary of our 10 experiments.

Exp.	N	Manipulations				Effect Size	Original Exp.
		Timing	Order	Blocking	Label		
1	50	seq.	1-3	random	same	1.42 [1.32, 1.52]	
2	50	seq.	1-3	random	diff.	1.26 [1.18, 1.34]	
3	50	seq.	1-3	blocked	diff.	1.31 [1.23, 1.39]	
4	50	simult.	1-3	random	same	1.32 [1.24, 1.4]	XT E1/E2
5	50	simult.	1-3	random	same	1.14 [1.06, 1.22]	XT E1/E2
6	50	simult.	1-3	random	diff.	1.16 [1.08, 1.24]	
7	50	seq.	3-1	blocked	diff.	-0.44 [-0.52, -0.36]	SPSS E2/E3
8	50	seq.	3-1	blocked	same	-0.17 [-0.25, -0.09]	
9	50	seq.	3-1	random	same	-0.31 [-0.39, -0.23]	
10	50	simult.	3-1	blocked	diff.	0.02 [-0.06, 0.1]	SPSS ES1/ES2
11	50	simult.	3-1	blocked	diff.	-0.06 [-0.14, 0.02]	
12	50	simult.	3-1	blocked	same	-0.14 [-0.22, -0.06]	

¹ N = sample size; Timing = presentation timing (sequential or simultaneous); Order = relative ordering of 1 and 3 subordinate trials; Blocking = trials blocked by category or pseudorandom; Label = same or different label in 1 and 3 trials; Effect size = Cohen's d [95% CI]; Original Exp. = corresponding experiments from prior literature.

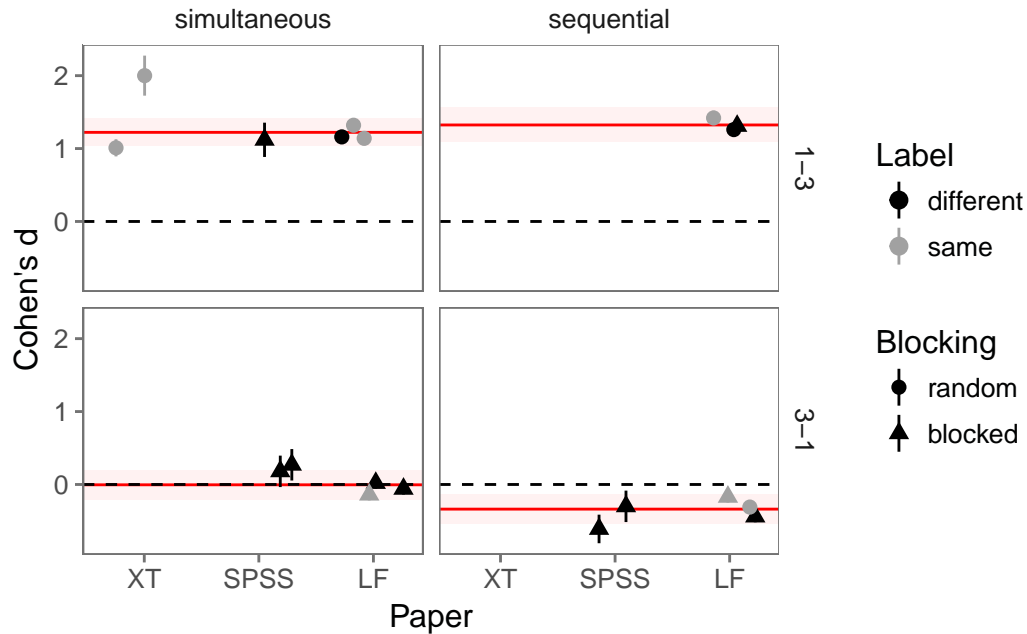


Figure 2. Cumulative plot of effect sizes for all 19 studies conducted on the suspicious coincidence effect by XT (Xu & Tenenbaum, 2007a), SPSS (Spencer, et al, 2011), and the current authors. Facets along the vertical indicate whether the single exemplar trial occurred first (1-3) or second (3-1). Facets along the horizontal indicate whether the exemplars were presented simultaneously as in XT or sequentially as in SPSS. Point color indicates whether the single exemplar and three subordinate exemplars received the same (grey) or different (black) label. Point shape indicates whether trials were blocked by category (circle) or pseudo-random (triangle). Points are jittered along the x-axis for visibility. The red line reflects the meta-analytic estimate of the effect size. All error bars are 95% confidence intervals.

Table 2

Meta-analytic model with manipulations as fixed effects.

Fixed effect	beta	z-value	p-value
Intercept	1.37 [1.09, 1.65]	9.48	<.0001
Simultaneous vs. sequential timing	-0.13 [-0.33, 0.08]	-1.18	0.24
1-3 vs. 3-1 condition order	-1.48 [-1.77, -1.18]	-9.90	<.0001
Different vs. same label	0.03 [-0.21, 0.26]	0.20	0.84
Blocked vs. random trial order	-0.09 [-0.41, 0.24]	-0.52	0.6