



Crowdsourcing the Credibility of Empirical Research

Etienne P. LeBel, PhD
University of Western Ontario / IGDORE
etienne.lebel@gmail.com
etiennelebel.com
CurateScience.org

Goals of Science

- Advance theory
- Address social & medical problems
 - Cancer, heart disease, stroke, Alzheimer's
 - Suicide, violence, extremism, discrimination, poverty



- To achieve such goals, scientific findings must be *credible*.
- ***But what makes a published finding credible?***



“...if you’re doing an experiment, you should report everything that you think might make it invalid—not only what you think is right about it: other causes that could possibly explain your results; ...

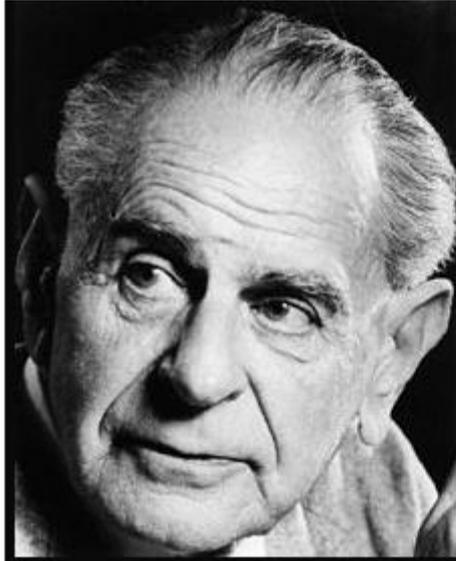
Details that could throw doubt on your interpretation must be given, if you know them. You must do the best you can—if you know anything at all wrong, or possibly wrong—to explain it. If you put out a theory, for example, then you must also put down all the facts that disagree with it, as well as those that agree with it. ... ***In summary, the idea is to try to give all of the information to help others to judge the value of your contribution; not just the information that leads to judgment in one particular direction or another.***”

(Richard Feynman, 1974, Cargo Cult Science)

- In other words, to evaluate the credibility of a published finding, it needs to be reported with *sufficient transparency*.
- *Sufficient transparency* allows independent researchers to:
 - ruthlessly scrutinize a study's methods/data to identify design flaws and data reporting errors
 - conduct replication attempts in independent samples

- If a finding “survives” such scrutiny, i.e., independent researchers fail to identify
 - fatal study design flaws,
 - data processing/statistical errors and result fragilities, and
 - replicability issues,then, and only then, can a finding be (temporarily) retained as *not-yet-proven-wrong* and hence treated as *credible*.
- In other words, *sufficient transparency maximizes falsifiability*.

Falsifiability



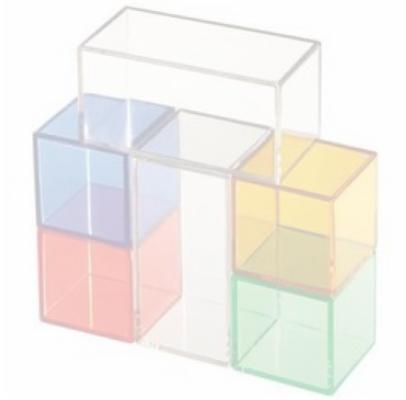
In so far as a scientific statement speaks about reality, it must be falsifiable; and in so far as it is not falsifiable, it does not speak about reality.

(Karl Popper)

- Possibility of proving a hypothesis wrong
 - i.e., to conceive of observations that could refute a hypothesis
- But 50 shades (or degrees) of falsifiability
 - High falsifiability → highly likely to prove a false hypothesis wrong
 - Low falsifiability → highly unlikely to prove a false hypothesis wrong

Violently strong falsification attacks

All hypothesis bricks start fragile



Surviving hypothesis become strong bricks leading to solid cumulative knowledge.

Recap

- Sufficient transparency required for falsifiability
- Higher levels of transparency afford higher levels of falsifiability.
 - The more scrutiny a published finding “survives”, the more *credible* it can be considered.
 - 3 fundamental dimensions:
 - Method/data transparency
 - Analytic reproducibility/robustness
 - Effect replicability

Concrete example: Bem's (2011) ESP studies

- “Retroactive recall” (Bem's Study 8 & 9)
 - **Finding:** Enhanced memory for words rehearsed *after* a test?

Feeling the Future: Experimental Evidence for Anomalous Retroactive Influences on Cognition and Affect

Daryl J. Bem
Cornell University

The term *psi* denotes anomalous processes of information or energy transfer that are currently unexplained in terms of known physical or biological mechanisms. Two variants of psi are *precognition* (conscious cognitive awareness) and *premonition* (affective apprehension) of a future event that could not otherwise be anticipated through any known inferential process. Precognition and premonition are themselves special cases of a more general phenomenon: the anomalous retroactive influence of some future event on an individual's current responses, whether those responses are conscious or nonconscious, cognitive or affective. This article reports 9 experiments involving more than 1,000 participants

1. Method/Data transparency

- **1.1 Methodological details:** Seemingly fully reported, but actually opaque according to new reporting standards.
- **1.2 Experimental materials availability:** Available upon request, however, not publicly available.
- **1.3 Pre-registration:** Studies and hypotheses not pre-registered; cannot distinguish exploratory vs. confirmatory analyses.
- **1.4 Data files availability:** Available upon request, however, not publicly available.

2. Analytic reproducibility

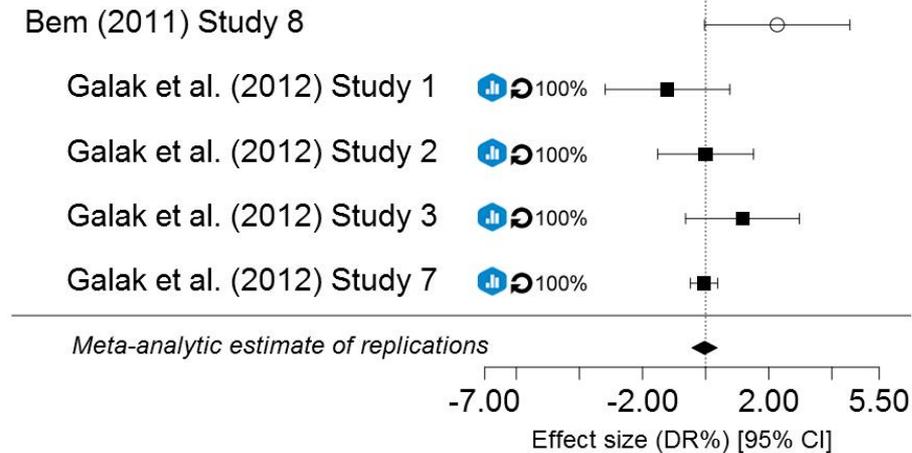
- Schimmack's (2018) email correspondence w/ Bem

Dataset	Sample	Year	N	Experiment
5	1	2002	50	#5: Retroactive Habituation I (Neg only)
5	2	2002	50	#5: Retroactive Habituation I (Neg only)
6	1	2002	91	#6: Retroactive Habituation II (Neg & Erot)
6	2	2002	19	#6: Retroactive Habituation II (Neg & Erot)
6	3	2002	40	#6: Retroactive Habituation II (Neg & Erot)
7	1	2005	200	#7: Retroactive Induction of Boredom
1	1	2006	40	#1: Precognitive Detection of Erotic Stimuli
1	2	2006	60	#1: Precognitive Detection of Erotic Stimuli
3	1	2008	100	#2: Precognitive Avoidance of Negative Stimuli
8?	1	2008	50	#2: Precognitive Avoidance of Negative Stimuli
2	1	2007	100	#3: Retroactive Priming I
2	2	2008	100	#4: Retroactive Priming II
4?	1	2008	100	#8: Retroactive Facilitation of Recall I
9	1	2009	50	#9: Retroactive Facilitation of Recall II

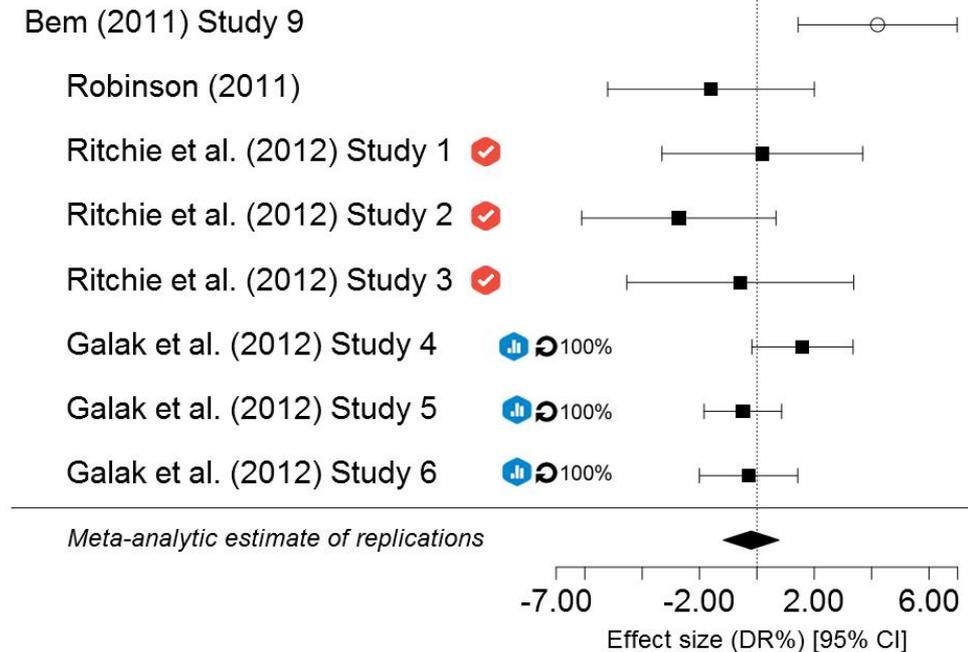
3. Effect replicability

Retroactive Recall Effect

Original instantiation of effect



Method generalization of effect



Bem (2011) summary

- Sufficient transparency would have revealed cherry picking of samples, studies, items, and analyses
- Independent analytic reproducibility verifications would have revealed data processing/analytic problems
- Hence:
 - Paper likely would never have been published
 - Replicators likely would have decided not worth time/resources to attempt replication of Bem's findings



- **Goal:** To crowdsource the credibility of empirical research by curating these 3 (fundamental) falsifiability-related study characteristics:
 1. Method/data transparency
 2. Analytic reproducibility/robustness
 3. Effect Replicability

→ The more scrutiny a published finding “survives” along these 3 dimensions, the more *credible* it is.
- **Current focus:**
 - Developing standards for a web platform (user interface) to curate the method/data transparency of studies

Demo: Latest Prototype

Search for transparently reported studies!

Search:

Include only studies with:

-  Reporting standard/Methodological disclosure statements
-  Open/public experimental materials
-  Pre-registered/Registered Reports protocol
-  Open/public data

authors/study.number	journal.name	article.title	DOI
Pittarello, Leib et al. (2015) Study 1     Study 2    	<i>Psychological Science</i>	 Justifications Shape Ethical Blind Spots	10.1177/0956797615571018
Colby, DeWitt & Chapman (2015) Study 1     Study 2     Study 3    	<i>Psychological Science</i>	 Grouping Promotes Equality: The Effect of Recipient Grouping on Allocation of Limited Medical Resources	10.1177/0956797615583978
Birmingham et al. (2015)    	<i>Psychological Science</i>	 Implicit Social Biases in People With Autism	10.1177/0956797615595607
Tworek & Cimpian (2016) Study 1     Study 2     Study 3     Study 4     Study 5      	<i>Psychological Science</i>	 Why Do People Tend to Infer 'Ought' From 'Is'? The Role of Biases in Explanation	10.1177/0956797616650875
Willén & Strömwall (2012)    	<i>Legal and Criminological Psychology</i>	 Offenders' uncoerced false confessions: A new application of statement analysis?	10.1111/j.2044-8333.2011.02018.x
Campbell et al. (2018)     	<i>Journal of Research in Personality</i>	 Self-esteem, relationship threat, and dependency regulation: Independent replication of Murray, Rose, Bellavia, Holmes, and Kusche (2002) Study 3	10.1016/j.jrp.2017.04.001
Vize, Collison et al. (2018)    	<i>European Journal of Personality</i>	 Examining the Effects of Controlling for Shared Variance among the Dark Triad Using Meta-analytic Structural Equation Modelling	10.1002/per.2137
Butler, Karpowitz et al. (2017)     	<i>Political Science Research and Methods</i>	 Who Gets the Credit? Legislative Responsiveness and Evaluations of Members, Parties, and the US Congress HTML	10.1017/psrm.2015.83

End.

Thank you for your attention.



Crowdsourcing the Credibility of Empirical Research

Etienne P. LeBel, PhD
University of Western Ontario / IGDORE
etienne.lebel@gmail.com
etiennelebel.com
CurateScience.org

*Thanks to Felix Schönbrodt for hosting CurateScience.org

Table 1. The intimate relation between analytic reproducibility, analytic robustness, effect replicability, and effect generalizability in terms of the similarity/dissimilarity of the analytic approach, methodology, and sample data used across original and follow-up studies.

		Methodology (manipulation, measurement instrument, paradigm)			
		Same		Different	
		Sample data		Sample data	
		Same	Different	Same	Different
Analytic approach	Same	Analytic Reproducibility	Effect Replicability	-	-
	Different	Analytic Robustness	-	-	Effect Generalizability