

Krisis Replikasi dan Praktik Sains Terbuka

Rizqy Amelia Zein

INSAN Jurnal Psikologi dan Kesehatan Mental



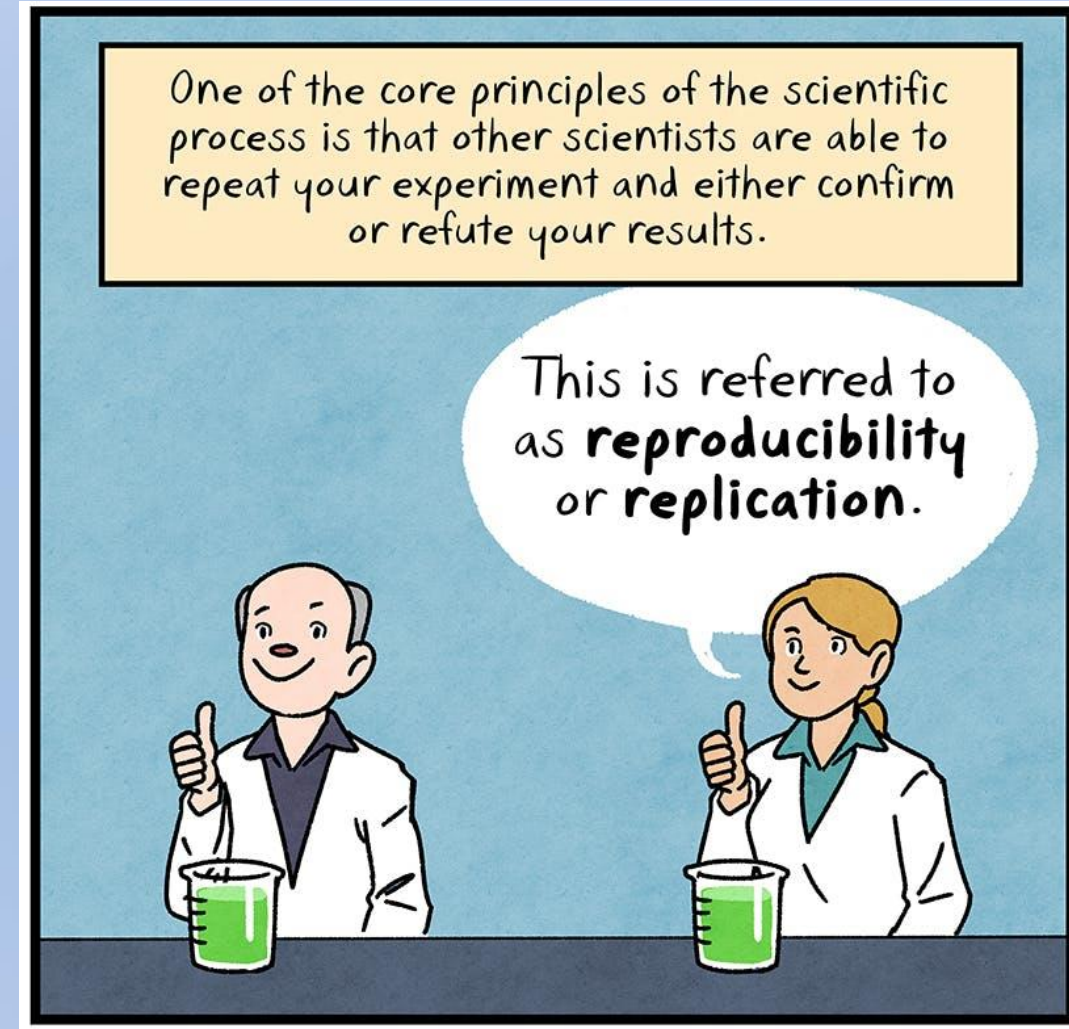
FAKULTAS
PSIKOLOGI
UNIVERSITAS AIRLANGGA

*Imagining
Learning
& Creating
for life*

psikologi.unair.ac.id

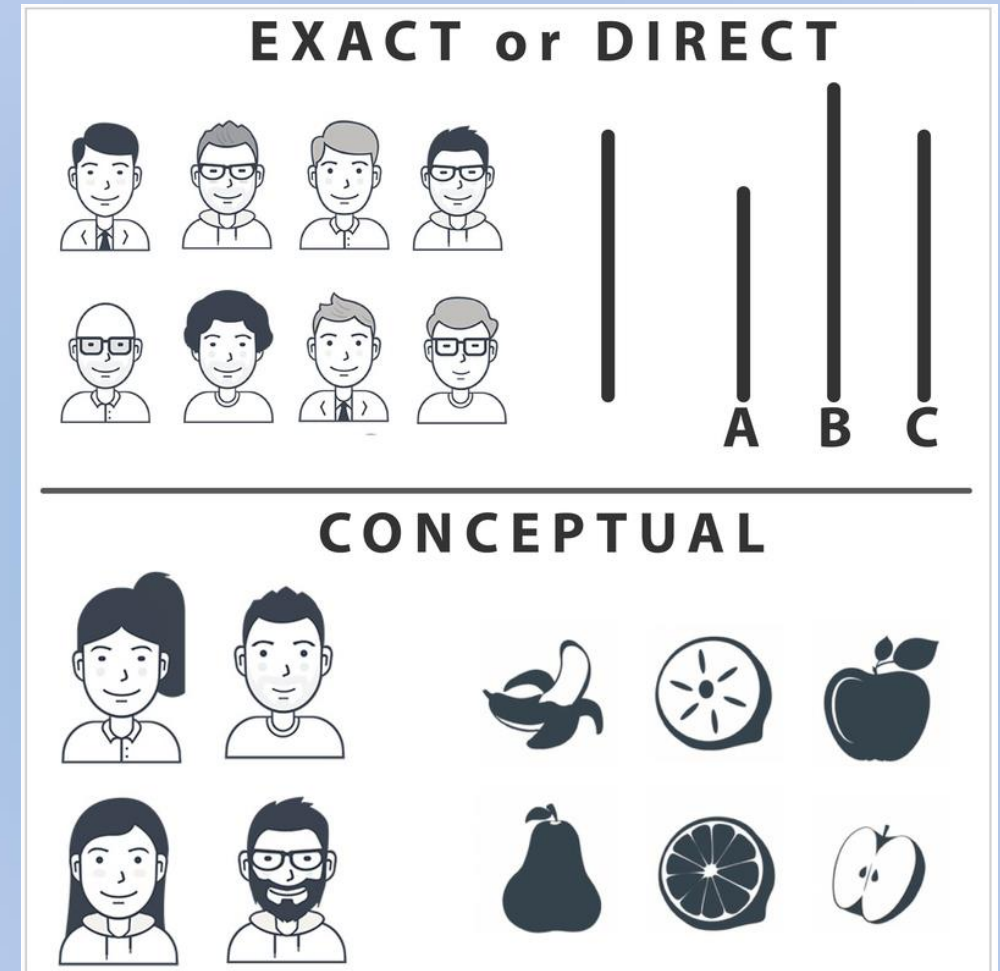
Replikasi (1)

- Seandainya anda sedang berjalan di Jalan Dharmawangsa dan melihat seseorang yang mirip mantan anda, pada pengamatan pertama mungkin anda tidak terlalu yakin bahwa orang tsb adalah mantan anda.
- Namun setelah observasi berkali-kali (dengan memutar jalan dan melihatnya untuk kedua, ketiga kali), anda akan yakin bahwa orang tsb adalah mantan/bukan mantan anda.
- Temuan penelitian baru akan kredibel apabila peneliti lain menghasilkan temuan yang kurang lebih sama, ketika penelitian direplikasi.
- Menariknya, sebagian studi-studi dalam Psikologi (Sosial) yang sangat populer, gagal direplikasi.
- <https://nobaproject.com/modules/the-replication-crisis-in-psychology>



Replikasi (2)

- Replikasi dapat dilakukan dengan 2 cara:
 - *Exact replication*, dan
 - *Conceptual replication*



Replication Continuum

	Highly Similar ←			→ Highly Dissimilar	
	Direct Replication			Conceptual Replication	
Design Facet	Exact Replication (All facets under researcher control are the same)	Very Close Replication (Procedure or physical setting is different)	Close Replication (IV or DV stimuli are different)	Far Replication (IV or DV operationalization or population is different)	Very Far Replication (IV or DV constructs are different)
Effect, Hypothesis	Same	Same	Same	Same	Same
IV Construct	Same	Same	Same	Same	Different
DV Construct	Same	Same	Same	Same	Different
IV Operationalization	Same	Same	Same	Different	
DV Operationalization	Same	Same	Same	Different	
Population (e.g., age)	Same	Same	Same	Different	
IV Stimuli	Same	Same	Different		
DV Stimuli	Same	Same	Different		
Procedural Details	Same	Different			
Physical Setting	Same	Different			
Contextual Variables	Different				
⋮	⋮				

Fig. 1. Taxonomy for classifying a replication study's methodological similarity to an original study. "Same" indicates that the design facet in question is the same as in the original study, and "different" indicates that it is different. IV = independent variable; DV = dependent variable. "Population" refers to major population characteristics, such as age and whether the sample is drawn from the community or a special clinical population. Procedural details are minor experimental particulars (e.g., task instructions, font, font size). Contextual variables are design facets beyond a researcher's control (e.g., history, culture, language).

Learning
& Creating
for Life



Many Labs 1 (1)

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, 349(6251), aac4716. (<http://osf.io/ezcu1>)

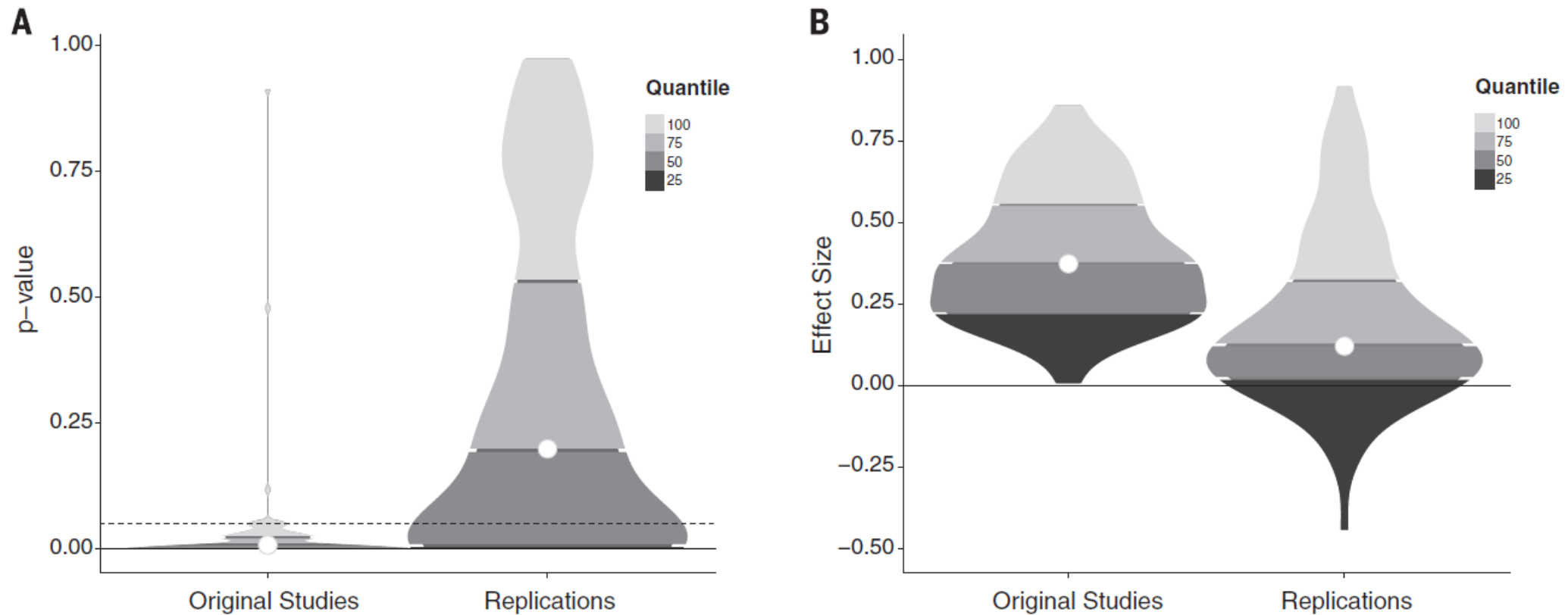


Fig. 1. Density plots of original and replication *P* values and effect sizes. (A) *P* values. (B) Effect sizes (correlation coefficients). Lowest quantiles for *P* values are not visible because they are clustered near zero.

Many Labs 1 (2)

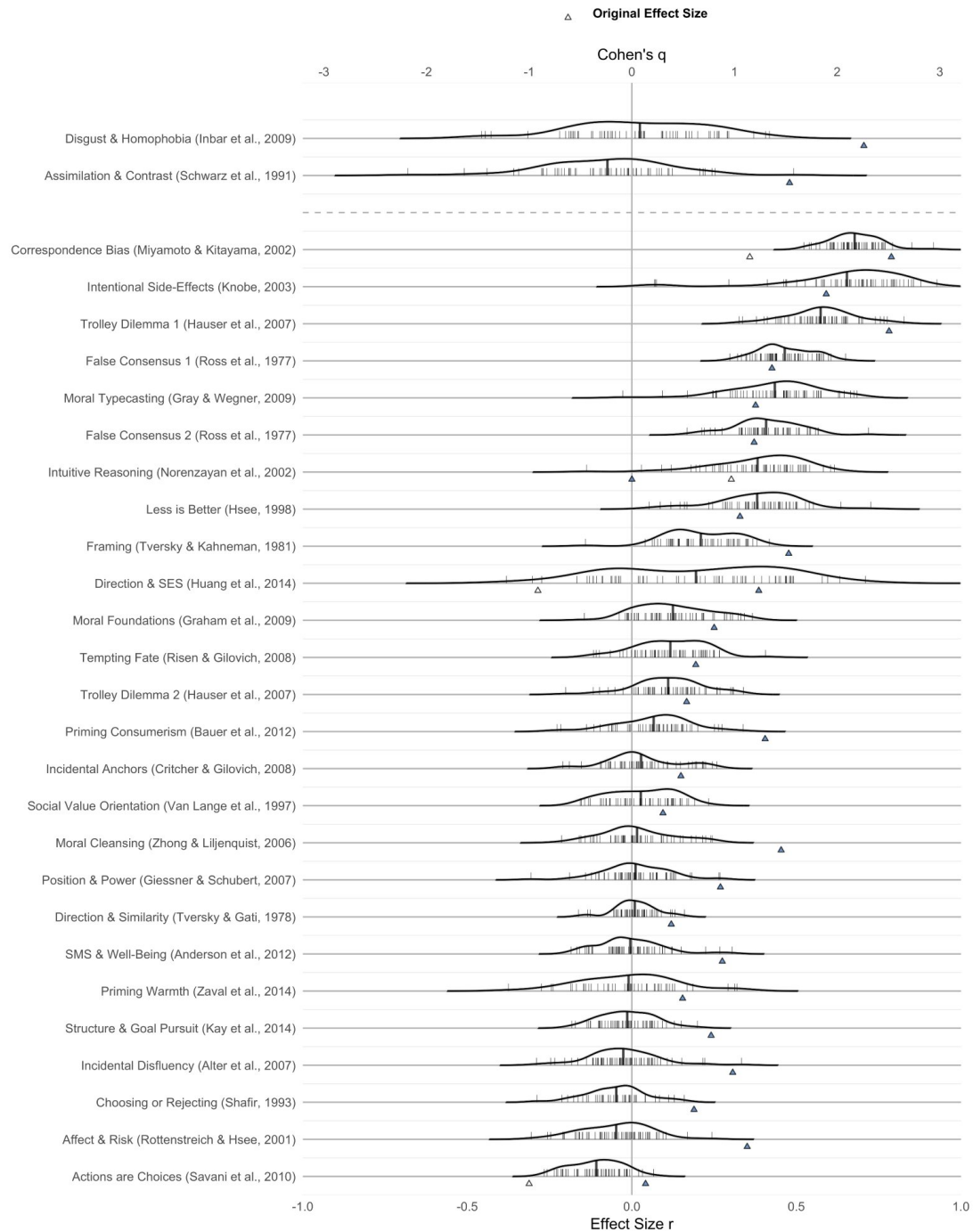
Journal	% Findings Replicated
Journal of Personality and Social Psychology: Social	23
Journal of Experimental Psychology: Learning, Memory, and Cognition	48
Psychological Science, social articles	29
Psychological Science, cognitive articles	53
Overall	36

Imagining
Learning
& Creating
for life



Many Labs 2

- <https://twitter.com/BrianNosek/status/1064548910872580101>
- *Pre-print:* <https://psyarxiv.com/9654g>
- *Data:* <https://t.co/EFhB9DR9IF>



Kasus Fluoxetine (Prozac)

- Fluoxetine (Prozac) adalah antidepresan yang sangat populer dan menghasilkan penjualan sampai miliaran dolar di Amerika Serikat saja
 - Namun studi meta-analisis yang dilakukan oleh Kirsch et al (2008) menghasilkan temuan bahwa efek yang diberikan Prozac sangat kecil, bahkan cenderung *placebo* (<https://doi.org/10.1371/journal.pmed.0050045>)
 - Kesimpulannya, Prozac sama sekali tidak bermanfaat untuk perawatan pasien depresi (<http://www.sciencemag.org/news/2008/02/problem-Prozac>)
- Ada problem metodologi yang kurang cermat dalam uji klinik Prozac → peneliti hanya mempublikasi hasil positif

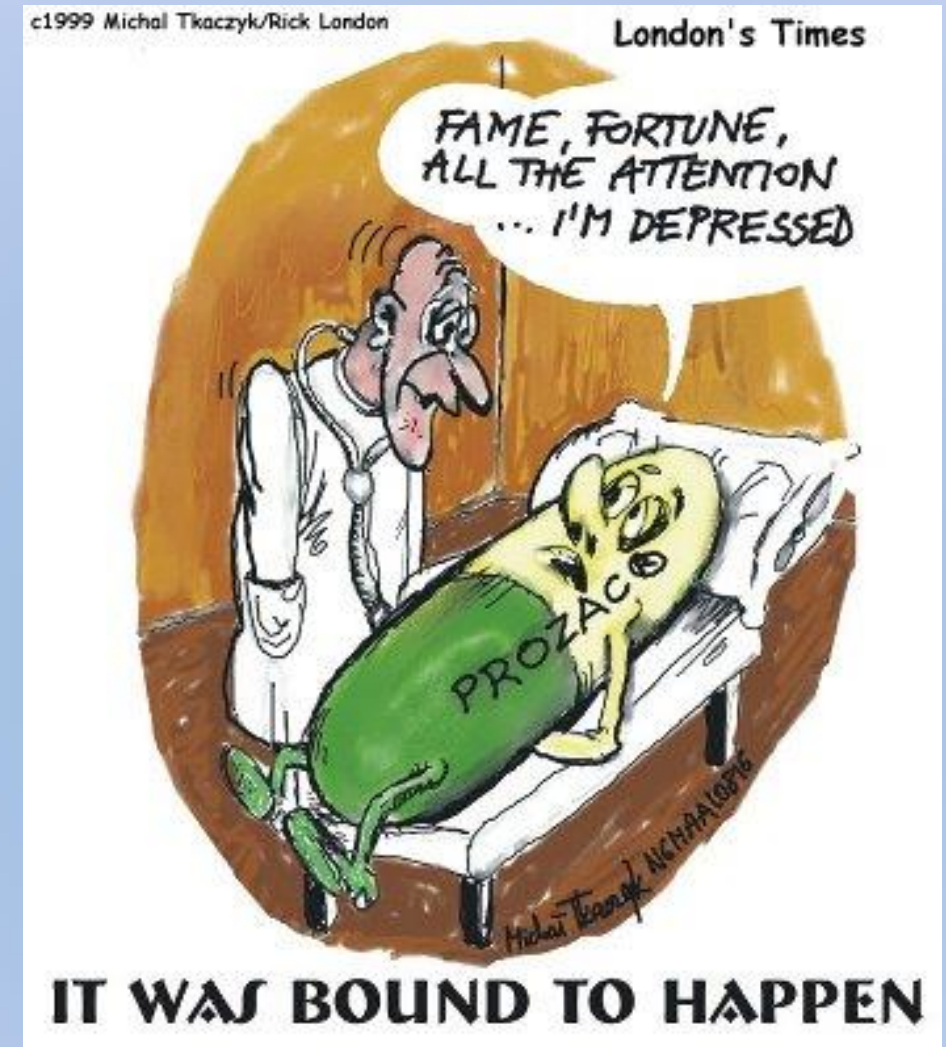


Imagining
Learning
& Creating
for life



...cont'd

- Ioannidis (2008) menyimpulkan bahwa mitos efektivitas antidepressan disebabkan karena:
 - Interpretasi yang keliru soal 'signifikansi statistik (*p-value*)'
 - Desain penelitian yang dimanipulasi
 - Pemilihan sampel yang bias
 - *Follow up study* yang cenderung berjangka waktu pendek dari pemberian perlakuan
 - ***Selective*** dan ***distortive report*** → peneliti hanya melaporkan hasil yang positif dan mengabaikan informasi lainnya yang bertentangan dengan hipotesisnya (*confirmation bias*)



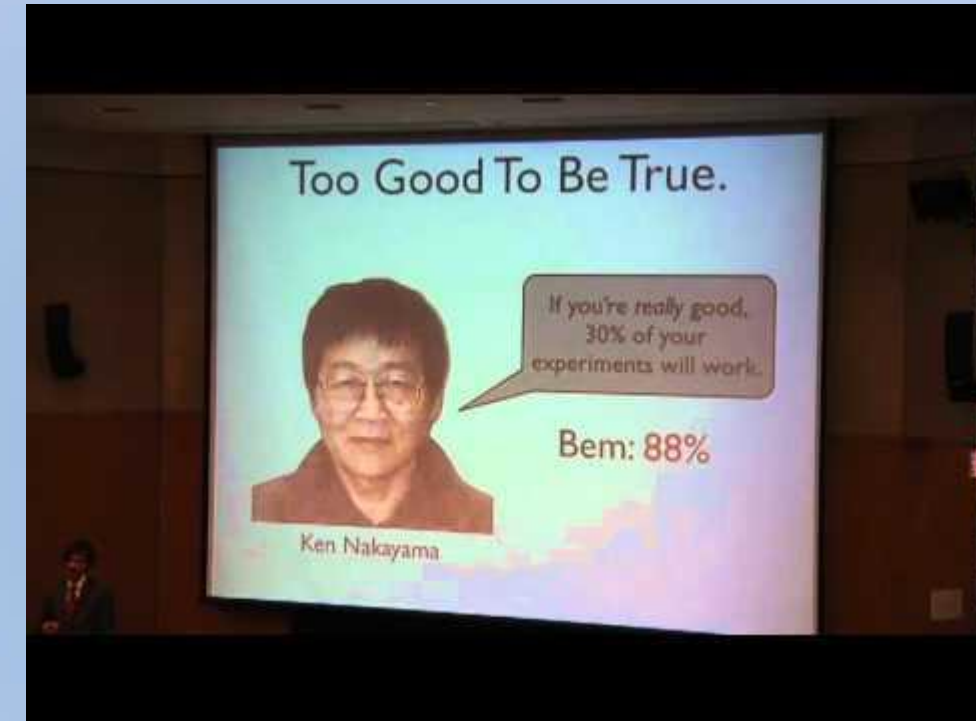
Kasus *psi* (precognition) (Bem 2011)

- Pada tahun 2011, Daryl Bem, Profesor di Cornell University menerbitkan karyanya yang berjudul *“Feeling the future: Experimental evidence for anomalous retroactive influences on cognition”* di APA Journal of Personality Psychology
 - Penelitian Bem memberikan ‘bukti’ adanya **precognition**, suatu bentuk *extrasensory perception*, dimana asumsinya manusia punya kemampuan ‘memprediksikan’ masa depan
 - 9 dari 10 eksperimennya menunjukkan *p-value* yang *significant* (<0.05) dengan *mean effect size* (d) = 0.22
 - **Direplikasi** (Ritchie et al 2012; Galek et al 2012) bahkan oleh Bem sendiri bersama timnya, tapi ***gagal*** mendapatkan hasil yang sama
 - Manuskrip Bem ini menjadi **kontroversi** global dan melahirkan gerakan **#OpenScience**



Apa yang dilakukan Bem?

- Temuan Bem yang amat unik (*time reversed causality* – tapi gagal direplikasi) mengindikasikan bahwa kemungkinan Bem melakukan:
 - **Selective reporting**: Bem hanya melaporkan hasil positif saja
 - Bem **tidak melaporkan semua kondisi** eksperimennya
 - Bem **mengukur beberapa variabel dependen** dan **hanya melaporkan yang signifikan** saja
 - Bem **membulatkan *p-value*** dengan berlebihan
 - HARKing
 - Bem **membuang data outliers**
 - Bem **berhenti mengambil data** setelah *p-valuenya* sudah mencapai $<.05$ (*optional stopping*). Atau mungkin....
 - ***Bem hanya sedang beruntung*** 😊



Mengapa ini menjadi masalah

- Lalu teori-teori yang diajarkan di buku teks, apakah bisa dipercaya?
- Apakah teori-teori tersebut saintifik?
- Kalau ternyata hasilnya *false finding*, apa bedanya dengan *hoax*?
- Penelitian yang gagal direplikasi amat mencemaskan, karena mengindikasikan:
 - Penelitian tidak dilakukan dengan cermat
 - Ada kemungkinan peneliti melakukan tindakan tak terpuji dengan melakukan manipulasi pada desain studi/cara analisis/data penelitian
 - Ada kemungkinan peneliti melakukan *data dredging*, yaitu sengaja memperbesar jumlah sampel agar mendapatkan nilai $p < .05$ (*p-hacking*)

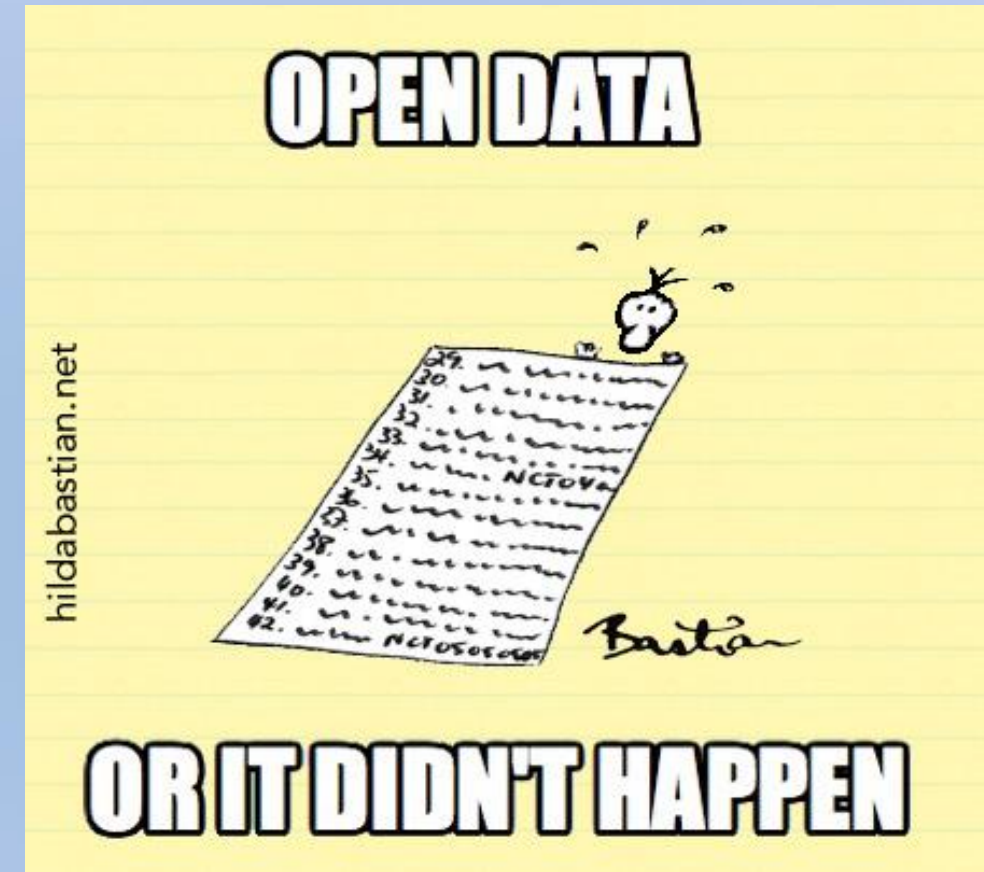


Imagining
Learning
& Creating
for life



Solusi atas problem ini

- Sains terbuka!
(https://en.wikipedia.org/wiki/Open_science)
 - Open data
 - Open access
 - Open source
 - Open methodology (& reporting)
 - Open educational resources
- Pra-registrasi
(<https://www.sciencemag.org/news/2018/09/more-and-more-scientists-are-preregistering-their-studies-should-you>)
 - <https://osf.io/k5wns/>
- Perencanaan jumlah sampel
- Optimisasi *statistical power*
- Lakukan penelitian *metascience*, yaitu *systematic review* & meta-analisis



Ioannidis' corrolaries (2008)

- 1: the smaller the studies (sample size) conducted in specific field, the less likely the research findings are to be true
- 2: the smaller the effect sizes in a scientific field, the less likely the research findings are to be true
- 3: the greater the number and the lesser selection of tested relationships in a scientific field, the less likely the research findings are to be true
 - Desain yang *confirmatory* lebih besar PPVnya
- 4: the greater the flexibility of the design, definitions, outcomes, and analytical modes of a scientific field, the less likely the research findings are to be true
- 5: the greater the financial and other interests and prejudices in a scientific field, the less likely the research findings are to be true
- 6: the hotter a scientific field (with more scientific teams involved), the less likely the research findings are to be true

Pre-registration

- Saat ini, tren *pre-registration* desain penelitian dan *analysis plan* menjadi populer utk mengatasi *replication crisis*
- Apa itu pre-registration?
 - Merupakan praktik *open research* dimana peneliti mendaftarkan semua detail penelitiannya pada suatu portal (<https://osf.io/> atau <https://aspredicted.org/>). Semua *pre-registration* akan di *time stamped*
 - Memastikan studi yang desainnya *confirmatory*, benar-benar *confirmatory*
 - Mencegah temuan penelitian yang *false positive*
- Bisakah diubah? Bolehkah menggunakan *post-hoc analysis*?
 - Dimungkinkan, tapi harus dijelaskan secara transparan dalam manuskrip penelitian



Towards a better science... (Willis & Moya 2017)

- Indicating in the method section: that the *number of participants* or *the rule for ending data collection* was determined a priori; that all the variables that were measured and all the manipulations performed are presented (and if they are not, why). Likewise, if covariables or filters were used, we also suggest that the results be presented without these covariables or filters.
- We recommend presenting an analysis and discussion of the statistical power (and if it is impossible to do this, explain why).
- We recommend clearly distinguishing in the article between exploratory and confirmatory analyses.



Perkembangan terakhir

- Sudah mulai ada desakan untuk mengadopsi prinsip sains terbuka
 - Ada *society* baru yang punya tujuan untuk mempromosikan sains terbuka dalam penelitian Psikologi yaitu Society for Improving Psychological Science (SIPS)
 - Sudah ada Tim Sains Terbuka Indonesia (TST) yang aktif mengadvokasi melalui berbagai kanal termasuk media sosial (#terbukaatautertinggal)
 - Sudah mulai ada indikasi *corrupted science* di Indonesia, sehingga sudah mulai ada kegelisahan
- Namun ide sains terbuka masih banyak mendapatkan resistensi





"Thank you. You've been a great audience."