Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Psychology

On priming action: conclusions from a meta-analysis of the behavioral effects of incidentally-presented words

Evan Weingarten[1], Qijia Chen[1], Maxwell McAdams[1], Jessica Yi[1], Justin Hepler[2] and Dolores Albarracin[3]

This paper presents a summary of the conclusions drawn from a meta-analysis of the behavioral impact of presenting words connected to an action or a goal representation [1]. The average and distribution of 352 effect sizes from 133 studies (84 reports) revealed a small behavioral priming effect (d_{YE} = 0.332, d_{RE} = 0.352), which was robust across methodological procedures and only minimally biased by the publication of positive (vs. negative) results. More valued behavior or goal concepts (e.g. associated with important outcomes or values) were associated with stronger priming effects than were less valued behaviors. In addition, opportunities for goal satisfaction appeared to decrease priming effects.

Addresses
[1] University of Pennsylvania, USA
[2] Facebook
[3] University of Illinois at Urbana-Champaign, USA

Corresponding authors: Weingarten, Evan (ewein@wharton.upenn.edu) and Albarracin, Dolores (dalbarra@illinois.edu)

Current Opinion in Psychology 2016, **12**:53–57

This review comes from a themed issue on **Social priming**

Edited by **Fritz Strack** and **Norbert Schwarz**

http://dx.doi.org/10.1016/j.copsyc.2016.04.015

2352-250/© 2016 Elsevier Ltd. All rights reserved.

Now, goal-priming experiments are coming under scrutiny — and in the process, revealing a problem at the heart of psychological research itself. [2]

In 1996, Bargh *et al.* asked a group of New York University undergraduates to complete a brief research task and then request a second task from another researcher in a nearby room. The first task was comprised of scrambled sentences (e.g. *they her respect see usually*, *they her bother see usually*, and *they her exercising see usually*). After unscrambling the sentences, participants sought out the experimenter who was chatting with a friend.

Naturally, the NYU students were able to correctly form 15 sentences from the grabled ones (e.g. *they usually respect her*, *they usually bother her*, *and they usually see her exercising*). Surprisingly, the content of the unscrambled sentences (containing polite, rude, or neutral themes that were varied among participants) influenced the amount of time students took to interrupt the experimenter when requesting permission to proceed with the second study. College students who had unscrambled sentences about rudeness were more likely to interrupt the experimenter's conversation than those who had unscrambled sentences about politeness or unrelated topics.

In light of this evidence, Bargh *et al.* [4] argued that the effects of priming were not limited to social perception [3] but instead reached the more substantial domain of action. Since that time and for nearly two decades, social psychologists and scholars in many other fields have attempted to understand the perceptual and motivational principles responsible for the intriguing observations presented by Bargh *et al.* [4] in their seminal study. For example, Bargh and his colleagues (2001) tasked students with solving a series of word search puzzles that either contained synonyms of achievement (e.g. *win*, *achieve*) or control words (e.g. *building*, *staple*). All students were reminded of achievement goals, leading to improved intellectual performance before the exercise. However, those who first found achievement words located more words on subsequent word search puzzles than did those who initially found neutral words.

Despite the excitement surrounding the effects of primes on performance, the Zeitgeist changed as a result of failure to directly replicate the phenomenon. The last five years have shown a dramatic shift toward the more somber intellectual climate reflected in the quotes below.

As all of you know, of course, questions have been raised about the robustness of priming results. The storm of doubts is fed by several sources, including the recent exposure of fraudulent researchers, general concerns with replicability that affect many disciplines, multiple reported failures to replicate salient results in the priming literature, and the growing belief in the existence of a pervasive file drawer problem that undermines two methodological pillars of your field: the preference for conceptual over literal replication and the use of meta-analysis. [. . .] For all these reasons, right or wrong, your field is now the poster child for doubts about the integrity of psychological research. [5]

The worst is yet to come for priming . . . over the next two or three years you're going to see an avalanche of failed replications published. [6]

Could a well-executed meta-analysis of the behavioral effects of incidentally presented concepts transform this controversy and inform the many disciplines concerned with this phenomenon? Weingarten *et al.* [1] thought so, especially with the use of sophisticated methods to detect the systematic elimination of null and negative findings (a form of publication bias often referred to as the *file drawer* problem) (see [7,8]). With the objective of gathering the most comprehensive data available on this issue, Weingarten *et al.* [1] obtained published and unpublished research on the performance effects of priming concepts compared with a control condition. They then calculated Cohen's *g* by subtracting the mean of the control group from the mean of the priming group and dividing that by the pooled *SD*, or used analogous methods for categorical dependent measures. The results from this project were reported by Weingarten *et al.* [1].

Weingarten *et al.* [1] synthesized 352 published and unpublished effect sizes, obtained from research conducted in the US and internationally. Priming methods included various forms of supraliminal and subliminal word presentation clearly linked to a concept (*win*, *affiliate*). The most commonly primed concepts were presented supraliminally (e.g. via scrambled sentences and word puzzles) and pertained to achievement, although social behaviors such as helping were also prevalent. Among many other measurements, performance measures included a score for test performance (number of solved problems), time spent on a task, and various ratings of overt behavior. Non-performance measures such as concept accessibility, or measures of attitudes, beliefs, or knowledge, were deemed ineligible in an attempt to model effects on *actual* cognitive and motor performance.

Summary of average findings

On average, Weingarten *et al.* [1] obtained a small but significant effect size comparable to many findings in psychological, sociological, and medical research (see e.g. $d. = 0.21$ [9]). Weighted mean effect sizes and associated heterogeneity statistics appear in Table 1 and indicate considerable non-random variability.

Detailed analyses of inclusion and publication bias were conducted, including funnel plot analyses [10–13], trim-and-fill methods [14], various failsafe *N* estimates [15–17], and cumulative meta-analysis [18]. A trim-and-fill analysis via the $R0^2$ estimator filled nine studies that yielded new, estimated fixed-effects $d. = 0.295$ (95% CI [0.264, 0.325]; $z = 19.08$, $p < .001$) and random-effects $d. = 0.312$ (95% CI [0.257, 0.366], $z = 11.15$, $p < .001$), suggesting a significant effect even after accounting for publication bias. Similarly, an Egger OLS modeled at the study level indicated a small study effect (fixed effects: $t(218) = 4.25$, $p < .001$; random effects: $t(218) = 5.19$, $p < .001$). Weingarten *et al.* [1] also used these analyses to identify outliers and remove them from further analyses. As a consequence, nine effect sizes from eight studies in subsequent analyses were removed ([19], Study 1; [20], Study 1; [21], Study 1; [22], Study 2; [23], Study 2; [24], Study 4; [25], Study 1; [26], Study 3). *P*-curves [27] were then fit to estimate potential bias in the selection of the statistical findings reported in the synthesized studies. (For a complete list of included studies, see [1]). Results suggested an absence of both *p*-hacking and selective reporting in the synthesized studies. Overall, these findings suggest a real performance effect that is not attributable to publication bias.

Table 1

Average effect (*d.*) and heterogeneity (*k* = 352)

Weighted mean effect	95% confidence interval	Heterogeneity indexes
FE $d. = 0.323$	0.277–0.387	$Q_{351} = 934.77^{***}$
RE $d. = 0.352$	0.294–0.409	$I^2 = 62.45$

*** $p < .001$.

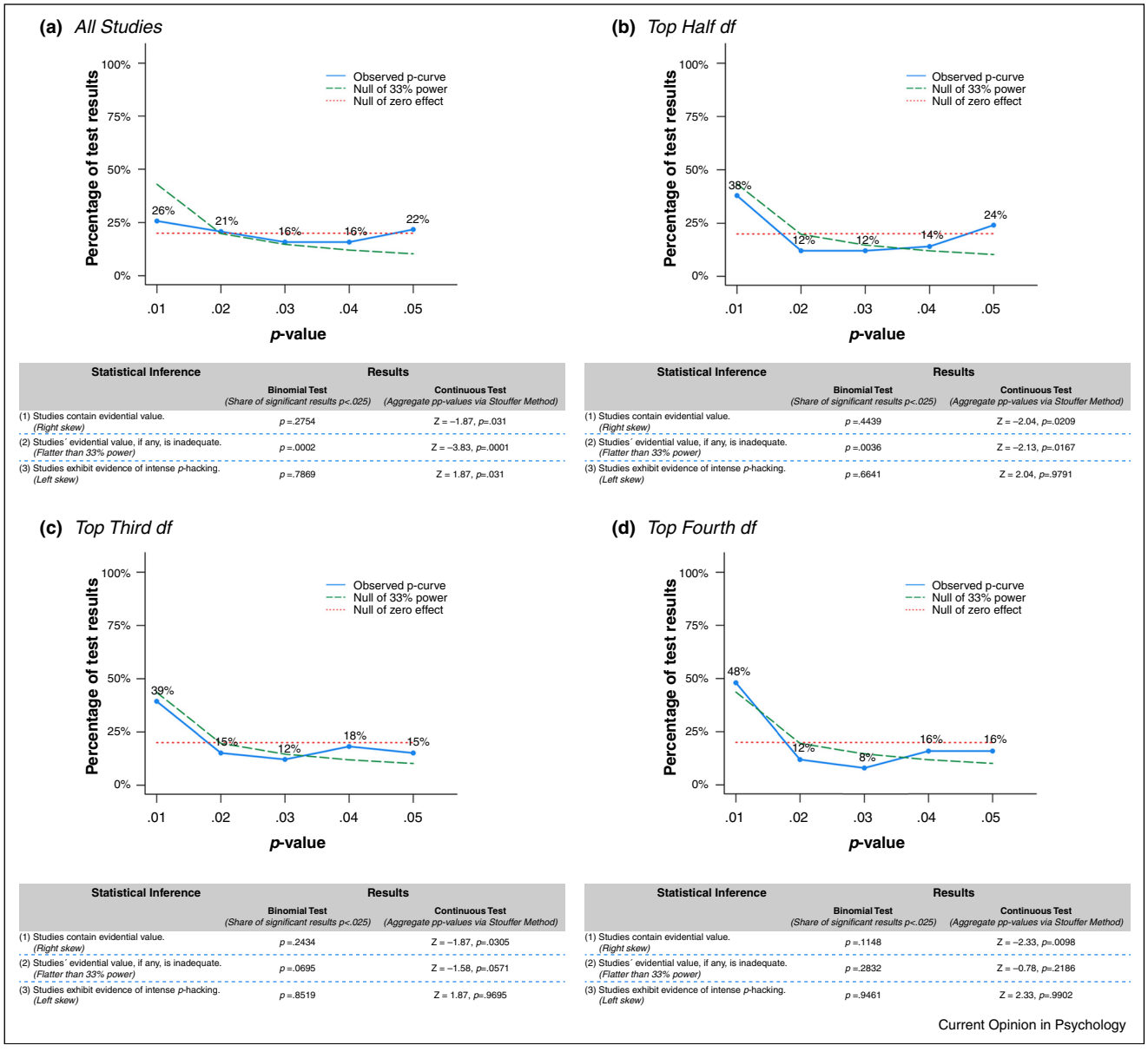
After removing effect sizes from the trim-and-fill analysis and modeling the studies at the study level, Weingarten *et al.* [1] obtained an average effect size of $d. = 0.315$ (95% CI [0.263, 0.368]; $t(132) = 11.75$, $p < .001$) from fixed-effects and $d. = 0.323$ (95% CI [0.270, 0.376]; $t(132) = 11.95$, $p < .001$) random-effects models. Both of these analyses rejected the null homogeneity hypothesis ($Q(342) = 806.43$, $p < .001$) and had a similar I^2 value of 57.59% (95% CI [54.08, 60.72]), demonstrating between moderate and large heterogeneity. This new effect size had a Rosenthal (Rosenberg) failsafe number of 46 930 (31 623), which exceeds the $5k + 10$ threshold and thus suggested that publication bias is unlikely to fully explain the meta-analytic findings [28,29].

Weingarten *et al.* [1] likewise ruled out that the effect sizes emerge from two different distributions using a normal-quantile plot of the 343 individual effect sizes, which also checks for non-normality of the data [30,31]. The normal-quantile plot examines potential publication bias by reviewing whether the shape of the curve has any discontinuities around 0 (indicative of publication bias) or has an S-shaped structure that might signal two underlying populations [30]. A Shapiro–Wilk normality test on the 343 data points yielded a marginally significant *p*-value ($W = 0.992$, $p = .073$), suggesting non-normality of the data, possibly indicating publication bias. The authors characterized this value as insufficient to explain their effect. The shape of the distribution did not immediately suggest that the studies come from two populations (curve, not S-shaped [30]).

Finally, *p*-curve analyses suggested that selective reporting could not explain the results of the set of studies from which Weingarten *et al.* [1] drew their effect sizes [27]. They presented two sets of *p*-curve analyses (based on continuous tests) of the studies in this meta-analysis: (a) a

p-curve on all studies conducted using *p*-values based on the researchers' focal hypotheses [27], and (b) *p*-curves based on studies with the largest error degrees of freedom. The curves can be found in Figure 1 and are based on the focal hypotheses of authors (often including interaction effects rather than mere differences between prime and control conditions). When Weingarten *et al.* [1] included all studies (see Panel A, published or unpublished) with clear hypotheses for behavioral measures (as outlined in the paper's *p*-curve disclosure table), The researchers found no evidence of *p*-hacking (no left-skew), but dual evidence of a right-skew and flatter than 33% power. The *p*-curve being flatter than 33% indicated that, on average, the studies considered in this meta-analysis lacked the statistical power to uncover the effect of interest in the study, although selective reporting alone cannot explain the entirety of the evidence. Weingarten *et al.* [1] again found this pattern when they restricted the *p*-curve to studies in the top half of error degrees of freedom (see panel B of Figure 1). However, when they restricted the *p*-curve to studies in the top third (see panel C of Figure 1) or top quartile

Figure 1



Statistical Inference	Binomial Test (Share of significant results p<.025)	Continuous Test (Aggregate pp-values via Stouffer Method)
(1) Studies contain evidential value. (Right skew)	p =.2754	Z = −1.87, p=.031
(2) Studies' evidential value, if any, is inadequate. (Flatter than 33% power)	p =.0002	Z = −3.83, p=.0001
(3) Studies exhibit evidence of intense p-hacking. (Left skew)	p =.7869	Z = 1.87, p=.031

Statistical Inference	Binomial Test (Share of significant results p<.025)	Continuous Test (Aggregate pp-values via Stouffer Method)
(1) Studies contain evidential value. (Right skew)	p =.4439	Z = −2.04, p=.0209
(2) Studies' evidential value, if any, is inadequate. (Flatter than 33% power)	p =.0036	Z = −2.13, p=.0167
(3) Studies exhibit evidence of intense p-hacking. (Left skew)	p =.6641	Z = 2.04, p=.9791

Statistical Inference	Binomial Test (Share of significant results p<.025)	Continuous Test (Aggregate pp-values via Stouffer Method)
(1) Studies contain evidential value. (Right skew)	p =.2434	Z = −1.87, p=.0305
(2) Studies' evidential value, if any, is inadequate. (Flatter than 33% power)	p =.0695	Z = −1.58, p=.0571
(3) Studies exhibit evidence of intense p-hacking. (Left skew)	p =.8519	Z = 1.87, p=.9695

Statistical Inference	Binomial Test (Share of significant results p<.025)	Continuous Test (Aggregate pp-values via Stouffer Method)
(1) Studies contain evidential value. (Right skew)	p =.1148	Z = −2.33, p=.0098
(2) Studies' evidential value, if any, is inadequate. (Flatter than 33% power)	p =.2832	Z = −0.78, p=.2186
(3) Studies exhibit evidence of intense p-hacking. (Left skew)	p =.9461	Z = 2.33, p=.9902

P-curve analyses from all of the effects reported in the included studies. This analysis refers to study-level reporting and power rather than the strength or significance of the priming effect.

(see Panel D of Figure 1) of degrees of freedom, there was a clear right skew, further indicating that selective reporting alone cannot explain the study results.

The role of theoretically meaningful moderators

The replicability of priming effects, however, depends not only on statistical considerations, but also on theoretically relevant features of priming studies. Is it possible that some failures to replicate are the result of researchers pursuing priming effects under moderating conditions that eliminate those effects, and thus should not be taken as evidence of the non-existence of priming per se? This possibility is consistent with the observed heterogeneity of the pooled effect size, so Weingarten *et al.*'s [1] meta-analysis also considered theoretical moderators of priming effects. As implied by the definition of goals as *desirable end states* ([32]; see also [33,34]), conditions in which performance had high value (e.g. often the goal was important to participants due to an accompanying reward) were associated with stronger priming effects than conditions of low value (for other effects of value, see [35]; see Barsalau, this issue). Second, as indicated by theories about goals ([36]; see also [37,38]), behavioral priming effects remained even in the absence of satisfaction opportunity when performance was valuable. In contrast, behavioral priming effects decayed in the absence of a satisfaction opportunity when performance was limited in value. Contrary to much speculation, there was generally an absence of methodological effects such as subliminal vs. supraliminal priming.

As quoted above, priming has become *the poster child* for concerns about the replicability and veracity of behavioral science research: These concerns attract the attention of the popular press (e.g. [2]) and even President Obama and his Council of Advisors on Science and Technology [39]. Because many of the raised concerns cannot be addressed by single studies, but only through the analysis of publication bias in the context of a meta-analysis, the Weingarten et al's [1] meta-analysis represented the first empirical investigation of these concerns. Through this approach, the meta-analysis excluded publication bias and selective reporting of analyses as an alternative explanation for the existence of priming effects. Furthermore, the *p*-curve techniques showed that the field is not plagued by widespread academic misconduct in the form of *p*-hacking, despite persistent conjectures in academic and popular debates to the contrary. In addition to clarifying these issues for the area of priming research, this approach may serve as a general model for addressing replicability concerns. Illuminating this pathway is an important, timely, and broad contribution. Psychology is not alone when it comes to replicability failures, given that reproducibility challenges trouble virtually all scientific disciplines [40], including medicine [41], behavioral genetics [42], and neuroscience [43] among others.

Conflict of interest statement

Nothing declared.

References

1. Weingarten E, Chen Q, McAdams M, Yi J, Helper J, Albarracin D: **From primed concepts to action: a meta-analysis of the behavioral effects of incidentally-presented words**. *Psychol Bull* 2016.
2. Satel SL: **Primed for controversy**. *New York Times* 2013 http://www.nytimes.com/2013/02/24/opinion/sunday/psychology-research-control.html?_r=2&. (retrieved 10.07.14).
3. Bargh JA: **The four horsemen of automaticity: awareness, efficiency, intention, and control in social cognition**. In *Handbook of Social Cognition,* edn 2. Edited by Wyer Jr RS, Srull TK.Hillsdale, NJ: Erlbaum; 1994:1-40.
4. Bargh JA, Chen M, Burrows L: **Automaticity of social behavior: direct effects of trait construct and stereotype priming on action**. *J Pers Soc Psychol* 1996, **71**:230-244.
5. Kahneman D: *Personal Communication*. 2012, September.
6. Barlett T: *Power of Suggestion*. Chronicle of Higher Education; 2013 http://chronicle.com/article/Power-of-Suggestion/136907/. (retrieved 10.7.14).
7. Cooper H: **Research synthesis and meta-analysis: a step by step approach**. edn 4. **Applied Social Research Methods Series**. Thousand Oaks, CA: Sage; 2010.
8. Cooper HM, Hedges LV (Eds): *The Handbook of Research Synthesis*. New York: Russell Sage Foundation; 1994.
9. Johnson BT, Scott-Sheldon LAJ, Carey MP: **Meta-synthesis of health behavior change meta-analyses**. *Am J Public Health* 2008, **100**:2193-2198 http://dx.doi.org/10.2105/AJPH.2008.155200.
10. Begg CB, Mazumdar M: **Operating characteristics of a rank correlation test for publication bias**. *Biometrics* 1994, **50**:1088-1101.
11. Egger M *et al.*: **Bias in meta-analysis detected by a simple, graphical test**. *Br Med J* 1997, **315**:629-634.
12. Light RJ, Pillemer DB: *Summing Up: The Science of Reviewing Research*. Cambridge, MA: Harvard University Press; 19840-674-85431-4.
13. Peters JL, Sutton AJ, Jones DR, Abrams KR, Rushton L *et al.*: **Assessing publication bias in meta-analyses in the presence of between-study heterogeneity**. *J R Stat Soc A Stat Soc* 2010, **173**:575-591 http://dx.doi.org/10.1111/j.1467-985X.2009.00629.x.
14. Duval SJ, Tweedie RL: **Practical estimates of the effect of publication bias in meta-analysis**. *Aust Epidemiol* 1998, **5**:14-17.
15. Orwin RG: **A fail-safe N for effect size in meta-analysis**. *J Educ Stat* 1983, **8**:157-159.
16. Rosenberg MS: **The file-drawer problem revisited: a general weighted method for calculating fail-safe numbers in meta-analysis**. *Evolution* 2005, **59**:464-468 http://dx.doi.org/10.1111/j.0014-3820.2005.tb01004.x.
17. Rosenthal R: **The "file drawer problem" and tolerance for null results**. *Psychol Bull* 1979, **86**:638-641.
18. Rothstein HR, Sutton AJ, Borenstein M: *Publication Bias in Meta-Analysis: Prevention, Assessment and Adjustments*. New York: John Wiley & Sons; 2005.
19. Jefferis VE, Fazio RH: **Accessibility as input: the use of construct accessibility as information to guide behavior**. *J Exp Soc Psychol* 2008, **44**:1144-1150 http://dx.doi.org/10.1016/j.jesp.2008.02.002.
20. Keatley DA, Clarke DD, Ferguson E, Hagger MS: **Effects of pretesting implicit self-determined motivation on behavioral**

engagement: evidence for the mere measurement effect at the implicit level. *Front Psychol* 2014 http://dx.doi.org/10.3389/fpsyg.2014.00125.

21. Legal J, Meyer T, Delouvee S: **Effect of compatibility between conscious goal and nonconscious priming on performance**. *Curr Res Soc Psychol* 2007, **12**:80-90.

22. Levesque CS: *Automatic activation of intrinsic and extrinsic motivation*. (Dissertation) 1999.

23. Macrae CN, Johnston L: **Help, I need somebody: automatic action and inaction**. *Soc Cogn* 1998, **16**:400-417 http://dx.doi.org/10.1521/soco.1998.16.4.400.

24. Oettingen G, Grant H, Smith PK, Skinner M, Gollwitzer PM: **Nonconscious goal pursuit: acting in an explanatory vacuum**. *J Exp Soc Psychol* 2006, **42**:668-675 http://dx.doi.org/10.1016/j.jesp.2005.10.003.

25. Roehrich L: *Priming the Pump: Activation of the Alcohol Expectancy Construct Increases Drinking Behavior*. (Dissertation) 1992.

26. Sela A, Shiv B: **Unraveling priming: when does the same prime activate a goal versus a trait?** *J Consum Res* 2009, **36** http://dx.doi.org/10.1086/598612.

27. Simonsohn U, Nelson LD, Simmons JP: ***P*-curve: a key to the file drawer**. *J Exp Psychol: Gen* 2014 http://dx.doi.org/10.1037/a0033242.

28. Rosenthal R: *Meta-Analytic Procedures for Social Research*. 1991.

29. Rosenthal R, Rosnow R: *Essentials of Behavioral Research: Methods and Data Analysis*. edn 3. McGraw-Hill; 2008.

30. Wang MC, Bushman BJ: **Using the normal quantile plot to explore meta-analytic data sets**. *Psychol Methods* 1998, **3**:46-54 http://dx.doi.org/10.1037/1082-989X.3.1.46.

31. Wang M, Bushman B: *Integrating Results through Meta-Analytic Review Using SAS Software*. SAS Publishing; 1999.

32. Lewin K: **Defining the 'field at a given time'**. *Psychol Rev* 1943, **50**:292-310.

33. Bargh JA, Gollwitzer PM, Lee-Chai AY, Barndollar K, Troetschel R: **The automated will: nonconscious activation and pursuit of behavioral goals**. *J Pers Soc Psychol* 2001, **81**:1014-1027.

34. Kruglanski AW, Shah JY, Fishbach A, Friedman R, Chun WY, Sleeth-Keppler D: **A theory of goal-systems**. In *Advances in Experimental Social Psychology*, vol. 34. Edited by Zanna MP. New York: Academic Press; 2002:331-378.

35. Johnson BT, Eagly AH: **The effects of involvement on persuasion: a meta-analysis**. *Psychol Bull* 1989, **106**:290-314 http://dx.doi.org/10.1037/0033-2909.106.2.290.

36. Zeigarnik B: **On finished and unfinished tasks**. In *A Sourcebook of Gestalt Psychology*. Edited by Ellis WD. New York, NY: Humanities Press; 1967:300-314.

37. Marsh RL, Hicks JL, Bink ML: **Activation of completed, uncompleted, and partially completed intentions**. *J Exp Psychol: Learn Mem Cogn* 1998, **24**:350-361 http://dx.doi.org/10.1037/0278-7393.24.2.350.

38. Förster J, Liberman N, Higgins ET: **Accessibility from active and fulfilled goals**. *J Exp Soc Psychol* 2005, **41**:220-239 http://dx.doi.org/10.1016/j.jesp.2004.06.009.

39. Begley S: **U.S. science officials take aim at shoddy studies**. *Reuters* 2014 http://uk.reuters.com/article/2014/01/27/science-reproducibility-idUKL2N0KX18S20140127. (retrieved 10.07.14).

40. Ioannidis JP: **Why most published research findings are false**. *PLoS Med* 2005, **2**:e124.

41. Prinz F, Schlange T, Asadullah K: **Believe it or not: how much can we rely on published data on potential drug targets?** *Nat Rev: Drug Discov* 2011, **10**:712.

42. Sullivan PF: **Spurious genetic associations**. *Biol Psychiatry* 2007, **61**:1121-1126.

43. Button KS, Ioannidis JP, Mokrysz C, Nosek BA, Flint J, Robinson ES, Munafò MR: **Power failure: why small sample size undermines the reliability of neuroscience**. *Nat Rev Neurosci* 2013, **14**:365-376.