Kačmár, P. (2019). Priming in the Context of Goal-Setting Theory (Review and P-Curve Analysis): The Good News, the Bad News and Some Recommendations. In P. Kačmár, J. Bavoľár, L. Lovaš (Ed.), *Work and Organizational Psychology 2018 - Past, Present, and Challenges to the Future* (pp. 208-230). Košice: ŠafárikPress. ISBN 978-80-8152-713-5

PRIMING IN THE CONTEXT OF GOAL-SETTING THEORY (REVIEW AND P-CURVE ANALYSIS): THE GOOD NEWS, THE BAD NEWS AND SOME RECOMMENDATIONS

PRIMING V KONTEXTE TEÓRIE STANOVOVANIA CIEĽA (PREHĽADOVÁ ŠTÚDIA A ANALÝZA P-KRIVKY): DOBRÉ SPRÁVY, ZLÉ SPRÁVY A ODPORÚČANIA

Pavol KAČMÁR

Abstract: *This study narratively reviewed and p-curved an emerging line of research which connects priming with the Goal-Setting Theory. The narrative review indicated that the Goal-setting theory seems to be a promising theoretical basis for priming research as a body of evidence suggests that performance could be bolstered by non/sub-conscious goal priming in a similar fashion as by conscious goals. However, present P-Curve analysis (six studies with nine inferential results; Chen & Latham, 2014; Latham et al., 2017; Latham & Piccolo, 2012; Shantz & Latham, 2011, 2009; Stajkovic et al., 2006) indicated that it cannot be concluded that the literature contains evidential value. Yet, it is also true that based on our analysis, we cannot conclude that evidential value is missing. For now, the results are inconclusive, and more research is necessary. We, therefore, further traced the merits and shortcoming of the present line of research and formulated some recommendations for future research. In a nutshell, Goal-Setting Theory still seems to be a promising theoretical basis for future priming research in the field of Industrial and Organizational Psychology and beyond, and, moreover, some merits of analyzed research can be identified. Nevertheless, in order to establish a more conclusive knowledge basis that can be trusted, some sound research practices (such as adequate statistical power; pre-registration of planned studies; and high-powered pre-registered replications) should be systematically implemented in future research.*

Keywords: *priming; Goal-Setting Theory; P-Curve; review; behavioral priming; social priming; performance priming*

1. AN INTRODUCTION

Is it possible that our behaviour is determined by subtle stimuli in our environment without even intending to use them in our decision making and/or having awareness of them or of their influence? This and similar questions are corroborated by line of research known as priming or social priming.

According to Fujita and Trope (2014; p. 68), the priming effect could be characterized as the "cognitive, motivational, affective and behavioral consequences of subtly enhancing the accessibility of a given construct independent of either available cognitive resources, awareness of this influence, or control over this influence".

Priming research in social psychology can be traced back to a study by Higgins, Rholes and Jones (1977). In this study, the participants underwent two ostensibly unrelated tasks. In the first task, the participants were unobtrusively exposed to personality traits with a positive or negative connotation. In the second task, deceptively described as a reading comprehension task, participants read about a person called Donald. His behaviour was described equivocally enough to be characterized in either a positive and negative manner. The results showed that stimuli in the first task influenced the results. If the information was applicable, the characterizations and evaluations of a person were "primed". In a similar vein, Srull and Wyer (1979) unobtrusively exposed participants to stimuli related to either hostility or not. Subsequently, participants read a vignette about a man, Donald, who behaved ambiguously enough to be considered either hostile or not. They consequently rated him on a set of personality traits. The results indicated that participants primed with the hostility evaluated Donald as more hostile than the participants in the control condition.

These studies led to new research paradigm and basis for prolific but somehow controversial research, shifting the focus away from the influence of primes in the realm of social perception towards other areas as such as motivation (e.g. Bargh, Lee-Chai, Barndollar, Gollwitzer, & Trötschel, 2001) and behaviour (e.g. Bargh, Chen, & Burrows, 1996; Dijksterhuis & van Knippenberg, 1998).

For instance, Bargh, Chen, and Burrows (1996) (Study two) exposed participants to neutral words or words related to a stereotype of the elderly (e.g. bingo) in a Scramble sentence task. Subsequently, the authors measured the speed of walking down the hall after leaving the experimental room. The results indicated that participants primed with words related to the elderly stereotype walked more slowly than the participants in the control condition. Similarly, Dijksterhuis and van Knippenberg (1998) (Study one) asked participants in the experimental condition to write an essay about a professor. In the control condition, the participants were asked to write an essay about a secretary. Following this, all participants did a general knowledge test in a subsequent ostensibly unrelated task. The results showed that the participants primed with the stereotype related to the professor performed better in the general knowledge test.

These studies have indicated that mental representation can be activated and thus influence thoughts and behaviour. Nevertheless, more recent attempts to replicate the results of previous priming studies have failed. For instance, the results of Bargh et al., (1996) were not replicated by Doyen, Klein, Pichon and Cleeremans (2012), despite the larger sample size and more advanced technology. In Study one, the results failed to show the expected priming effect. In

Study two, an effect was observed but only when the experimenter expected that participants would walk more slowly after the primes. Moreover, the results found by Dijksterhuis and van Knippenberg (1998) was not replicated by Shanks et al. (2013) over 9 experiments with 475 participants. Furthermore, a Bayesian analysis indicated support for the null hypothesis. A more recent multi-laboratory registered replication report conducted by O'Donnell et al. (2018) indicated a similar pattern of results. A Meta-analytic analysis of more than 4000 participants across various laboratories did not find an overall difference between conditions.

In fact, similar studies have recently emerged which have led some to question the existence of the priming effect. This has had the effect of polarizing the scientific community into the, so-called, "sceptics" and "believers". In general, two main explanations related to the observed discrepancy can be traced.

First, some authors have proposed that results indicating the existence of priming effects have been solely caused by questionable research practices and/or type one errors (false positives) combined with publication bias. For instance, Shanks et al. (2013) have argued that the analyzed professor priming effect is, most likely, a false positive. This argument could be in line with Simmons, Nelson, and Simonsohn (2011) who have suggested that it is possible to get a significant result even when the underlying hypothesis is false. This can occur due to researcher degrees of freedom such as flexibility in data collection, analysis and reporting. This phenomenon has been called p-hacking where a researcher tries to "hack" results in order to obtain a significant p-value, often less than 0.05. In combination with publication bias and other problematic research practices such as Harking (Hypothesizing After the Results are Known), it is possible to easily claim the existence of an effect even when the effect is not real. Given the relatively high prevalence of questionable research practices among researchers (John, Loewenstein, & Prelec, 2012), it is possible that whole area of research, such as social priming, could be tainted.

Secondly, proponents of the priming effect have argued that the current unsuccessful attempts to replicate previous studies have been caused by a lack of adequate theory capable of driving future research. There is, potentially, a variety of important mediators and moderators which could complicate replications. In his article "Welcome back theory", Dijksterhuis (2014) argues that (besides eliminating researcher degrees of freedom) theory development is crucial for future research. In fact, there are arguments suggesting crucial role of theory (see e.g. Bargh, 2006; Dijksterhuis, 2014; Higgins & Eitam, 2014) and, thus, many new promising priming theories have recently emerged (e.g. Cesario & Jonas, 2014; Higgins & Eitam, 2014; Loersch & Payne, 2014; Wheeler, DeMarree, & Petty, 2014; for a review of recent theories in the broader context of non-conscious goal activation see e.g. Kačmár & Lovaš, 2018).

One of the most recent and promising theories that can serve as a fruitful theoretical background for priming research, especially in the field of Industrial (Work) and Organizational Psychology, is the Goal-Setting Theory (Latham, 2016; Latham et al., 2017; Latham, Stajkovic, & Locke, 2010; Locke & Latham, 2006).

2. PRIMING IN THE CONTEXT OF THE GOAL-SETTING THEORY - REVIEW OF THE LITERATURE

According to the Goal-setting Theory, a behaviour is goal-directed (the goal can be characterized as the object or aim of the action). Therefore, performance in a task can be bolstered by stating a goal that has some concrete properties, e.g. a goal that is specific and difficult enough. This is done via mediators such as exertion of effort and is further moderated by factors, such as ability (Locke & Latham, 2015).

The Goal-Setting Theory was originally focused on conscious goals whilst neglecting the role of non/sub-conscious goals. More recently, Lock and Latham (Latham, 2016; Latham et al., 2017; Latham, Stajkovic, & Locke, 2010; Locke & Latham, 2006) have proposed a new layer to their theory which has incorporated automaticity/priming research from social psychology. Accordingly, the main tenet is that goals, as mental representations, can be activated by environmental stimuli which influence work performance and hypothetically operate in a similar way to a conscious goal. Therefore, further corroboration of the basic mechanisms and the role of proposed mediators and moderators seems to be promising[1]. Some of these basic tenets have been recently corroborated by an emerging line of research. A brief overview will be given below[2].

Some of the first research of the priming effect in the context of the Goal-Setting Theory was done by Stajkovic, Locke, & Blair (2006) who conducted two studies. In the first - pilot - study, participants in the experimental condition were exposed to achievement-related words in a word matrix. Participants in the control condition were exposed to neutral words. In the next ostensibly unrelated task, participants were asked to list various uses for an object (a coat hanger) as well as undergoing a debriefing procedure in the form of an awareness questionnaire. As expected, the analyses showed that participants in the experimental condition performed better and listed more uses than the participants in the control condition, validating task for future use.

In the main study, Scramble sentence task was used as the priming procedure. In the priming condition, words related to achievement were used. In the control condition, neutral words were used. Furthermore, participants were further assigned to one of three conscious goal conditions - do your best vs. easy goal vs. difficult goal. In the next part, the task used in the previous study was used. Participants were asked to list the various uses of a wire coat hanger as the measure of performance. Following this, an awareness questionnaire was administrated. A second measurement was conducted the next day. The results showed the effect of the conscious as well as the subconscious goal. Participants primed with the achievement-related words performed better. Moreover, the interaction between the conscious and sub-conscious goals was

[1] Additionally, the proposed theory building process is inspired by Goal-Setting Theory (see e.g. Locke, 2015). However, except briefly mentioning it in the discussion, this issue is beyond the scope of the present article.

[2] To identify the set of studies to be reviewed (and consequently P-curved), we analyzed the literature of the most recent articles by Lock and Latham; and articles that were cited by them. Furthermore, we found the oldest articles by the authors related to the topic (Shantz & Latham, 2009; Stajkovic et al., 2006) and analyzed articles that cited these studies using Google Scholar. The criteria for inclusion were (A) Empirical studies; (B) Combination of (1) Goal-Setting theory; and (2) Priming; (C) Connection to (work) performance; (D) Lock or Latham as co-author (as they are pioneering in this research area). The search of literature was conducted in February-March 2018.

significant as well. This provided the first indication of a priming effect in the context of the Goal-setting Theory.

In another study, Shantz and Latham (2009) extended previous results by conducting field experiments and substituting a word-related priming procedure with a picture related procedure. In their pilot study, participants were asked to participate when walking to work. They were asked to list the various uses for a coat hanger. The task was printed on paper with either a picture of (A) a woman winning a race; (B) a collage of various pictures; or (C) no picture was on the backdrop of the page. Participants performed best when the picture of the woman winning the race was used (Shantz & Latham, 2009; the picture of the women winning the race, as well as other pictures used as stimuli material in consequent studies, are shown in Appendix 1).

In a consequent study, Shantz and Latham (2009) conducted a laboratory experiment examining the influence of primes on implicit motivation. In the first part of the study, participants were exposed to the picture of the woman winning the race or not. The picture was printed on a backdrop of a paper with instructions. In the next part of the study, participants were asked to write an imaginative story about three objects after which their essays were analyzed. The results showed that the primed participants exhibited a greater need for achievement motivation as they used more achievement-related words.

Given the validity of the picture stimuli material in these two studies, a field experiment was carried out which examined the influence of priming on work-related performance (Shantz & Latham; 2009; Study three). The experiment took place in a call centre. Participants were assigned to four conditions. Two were related to non-conscious goal activation (priming group vs. control condition) while two were related to conscious goal instruction (specific high conscious goal vs. do your best). After the shift, a debriefing questionnaire was administered. The results showed that both factors - a conscious goal and non-consciously activated goal - were significant and participants exhibited better performance under these conditions. During a shift lasting three hours, the participants in the priming condition made more money than the participants in the control condition. As supported by the Goal-setting Theory, participants with a specific high goal made more money than participants asked to do their best.

In a following study, Shantz and Latham (2011) replicated the results of the previous field experiment in an organization with different organizational setting. Participants obtained the paper where ideas for soliciting fund were printed. Participants in an experimental condition were exposed to the picture of a woman winning a race. After the shift, participants were debriefed with the awareness questionnaire. The results showed that the primed participants had performed better. In a second study, Shantz and Latham (2011) chose another organizational setting. This time the participants spoke French and not English as they had previously. In addition, a longer period of work time was analyzed. Four-day shift each lasting four hours was chosen instead of one shift. The procedure was the same as it had been in the previous study and the results too. The participants primed with the picture of the woman winning the race raised more money than the participants in the control condition. Moreover, Shantz and Latham (2011) conducted a small meta-analysis of their three studies with averaged Cohen´s d = 0.56 CI (0.35-0.78).

In another study, Latham and Piccolo (2012) changed organizational setting one more time. This study was conducted in a call centre in the United States (previous studies had been carried

out in Canada). Furthermore, they analyzed the number of donors as well as the number of donations. As the Goal-Setting theory predicts that specific primes are more effective than general primes, a more specific prime was added in addition to the general prime used before The participants were assigned to three conditions. In the first condition, the general prime from the previous studies was used (woman winning the race). In the second condition, a more specific prime was used - a picture depicting a worker in a call centre (shown in Appendix 1). In the control condition, no pictures were presented. Data were collected over four workdays. Both the number of donors as well as the monetary gain was assessed. Furthermore, a TAT (Thematic Apperception Test) was administered. The results showed that participants primed with the specific prime obtained more pledges than the control group. Similarly, participants in the specific prime condition obtained more money than the participants in the control condition. However, the difference between the control condition and the general prime condition was not significant. An analysis of implicit motives showed that both participants primed with the specific goal and participants primed with the general goal wrote more words related to achievement in comparison to the participants in control condition.

Goal-setting theory differentiates between two types of goals - learning and performance goals. As a result, Chen and Latham (2014) carried out an experiment where participants were primed with either a learning or performance goal. In the first pilot study, participants assessed how much the picture made them think. According to this procedure, a picture of Rodin, The Thinker (in a specific format), was chosen (illustrated in Appendix 1). In the second pilot study, participants were primed with a new picture. This was either The Thinker, the woman winning the race, neither picture or both pictures. Following this, they had to write a story based on the presented pictures. The participants in The Thinker group produced more words related to insight than the other groups; while the race-winning woman group produced more achievement related words.

In the main experiment, Chen and Latham (2014) examined the influence of a performance/learning goal in a scheduling task, requiring the acquisition of knowledge. Firstly, the participants learned about a task. Next, the priming stimuli were presented on a computer screen. The task was to write a story about the pictures in one of the four conditions (thinker; winner; both; or neither). During the three trials, the participants had to complete as many class schedules as possible. At the end, an awareness check was administered. The results indicated that performance in a scheduling task was better when the participants were primed with a learning goal in comparison to the control condition and performance goal. However, the difference between the performance goal and priming was not significant.

In another series of studies, Latham, Brcic, and Steinhauer (2017) examined additional proposal derived from the Goal-Setting Theory. By this, that performance is determined by the difficulty of a goal. In a pilot study, participants were assigned to an experimental condition (picture of a man lifting 400 pounds; illustrated in Appendix 1); or a photograph of a rock. Following this, they had to write an imaginative story. The results showed that participants in the lifter condition provided more words related to effort than the participants in the control condition.

In the first experiment, participants were primed with a photograph of a man lifting 20 pounds (easy goal); 200 pounds (moderately difficult); or 400 pounds (difficult goal). After this, they were asked to press a digital food scale with a finger. At the end, an awareness check was conducted. The results showed a correlation between weight and task performance. Participants

who had been primed with the man lifting 400 pounds pressed harder than the participants primed with 200 and 20 pounds.

In the second experiment, Latham, Brcic, and Steinhauer (2017) were interested if priming difficult goals led to choosing a difficult conscious goal. The priming procedure was similar to the previous study. Participants were primed with the picture of a man lifting 20 or 400 pounds, after which they were asked to set a goal. The goal was to generate arguments in support of a statement. In addition, personality was assessed. The results showed that participants primed with the picture of the man lifting 400 pounds set a more difficult goal as well as producing more arguments. Moreover, the mediation analysis showed that self-set conscious goals mediated the relationship between prime and performance, and that conscientiousness, as a personality variable, moderated the relationship.

In a nutshell, body of evidence exists that indicates the effect of primes on the work performance and related issues. This is in line with the argument that the Goal-Setting theory could be a potentially relevant theoretical background for conducting priming research, especially in comparison to more classic priming theories. Nevertheless, a set of statistically significant results does not guarantee that the research encompasses evidential value. It is still possible that the studies were p-hacked and/or due to publication bias, only statistically significant results occurred in the published literature and, as such, the literature lacks evidential value. This issue, however, cannot be addressed by a narrative review. In order to overcome such shortcomings, we attempted to evaluate the evidential value of the reviewed articles with a P-curve analysis (Simonsohn et al., 2014a, 2014b, 2015).

3. P-CURVE ANALYSIS

Simmons and Simonsohn (2017) have illustrated that even a plethora of statistically significant studies can lack evidential value. They P-curved 33 statistically significant power posing studies from a previously published narrative review that affirmatively concluded that the effect exists. Crucially, the results of the P-Curve analysis indicated that the power posing literature lacked empirical support despite the numerous statistically significant results. Similarly, Lakens (2017) P-curved the two mainstream fields of priming research - elderly priming and professor priming. He concluded that while the professor priming contained some evidential value, the elderly priming did not. Furthermore, he obtained evidence of p-hacking in the elderly priming literature. This indicated that the statistically significant results could have occurred as a result of problematic research practices. Based on these observations, this study is interested in whether a similar or different pattern of results will occur in a present case. Therefore, a P-curve analysis (Simonsohn et al., 2014a, 2014b, 2015) was carried out to estimate the evidential value of the previously reviewed emerging line of research (connecting priming and the Goal-

Setting Theory)[3]. In total, six empirical articles were included [4] (Chen & Latham, 2014; Latham et al., 2017; Latham & Piccolo, 2012; Shantz & Latham, 2011, 2009; Stajkovic et al., 2006).

As some articles included more than one study, nine inferential results were included (two were excluded as they did not reach .05 value) in the total for analysis in (1) Variant A (main experiments only).

As only statistically unrelated results could be analyzed and one study could, hypothetically, report more related p-values (e.g. more dependent variables), in order to assess if the results were not contaminated by the specific results that were selected, (2) a robustness analysis using alternative measures reported as dependent variables was additionally conducted (Variant B).

Furthermore, as some articles reported up to 2 pilot studies, and we were interested if the pattern of results will change, (3) we additionally analyzed the results by adding the statistics from the pilot studies (13 values in total - Variant C).

The reason for conducting three p-curve analyses - (A) main analysis; (B) robustness analysis; and (C) additional analysis with pilot studies included - is to provide more nuanced and replicable results that are less influenced by selection of results.

The P-curve analysis was conducted according to the authors of the P-curve procedure (Simonsohn et al., 2014b, 2014b, 2015) and a disclosure table and further analysis details can be found online at DOI 10.17605/OSF.IO/DQZNT

Before going into the analysis, it is important to briefly mention that the p-curve analysis is based on the assumption that, under the null hypothesis (the underlying effect does not exist), all p-values are uniformly distributed. This is not only expected for non-significant results but for significant results as well. It also should be noted that, assuming publication bias is ubiquitous and hard to estimate, only significant results are analyzed in a p-curve analysis. If the effect exists, the p-curve should be skewed to the right (more p values between .01 to .02 than .04 to .05). Alternatively, if the results were p-hacked (e.g. various analyses were conducted, only the significant analysis were reported; covariates were added; cases were dropped; etc.), more p-values between .04 to 0.5 will be present. If there is no effect at all, the p-values should be uniformly distributed - e.g. there will be as many p values between .04 and 0.5 as between 0.1 and 0.2 (Simonsohn et al., 2014b, 2014b, 2015).

Firstly, we analyzed if the p-curved studies contained evidential value. According to the Simonsohn et al., 2014a, 2014b, 2015; evidential value is present when the p-curve is skewed to the right (lower p-values are more frequent - e.g. there are more .01 than .04 p values). This represents the alternative hypothesis and this condition is met when: A, there is evidence of

[3] Note that we are not primarily interested in whether the specific priming effect is real or not, but rather in estimating the evidential value of the research dedicated to connecting priming (non-conscious goal activation) and the Goal-Setting Theory; analyzing the indicators of p-hacking and statistical power of the analyzed studies. However, to illustratively compare the present state of art of this line of research to other, more established and mainstream research related to priming where the replication crisis has been explicitly stated by many authors, the obtained pattern of results are, only for illustration purposes, narratively compared to the P-curve analysis of professor and elderly priming studies conducted by Lakens (2017).

[4] Note that the criteria and procedure for the selection of studies for P-curve analysis were the same as for narative review and are listed above.

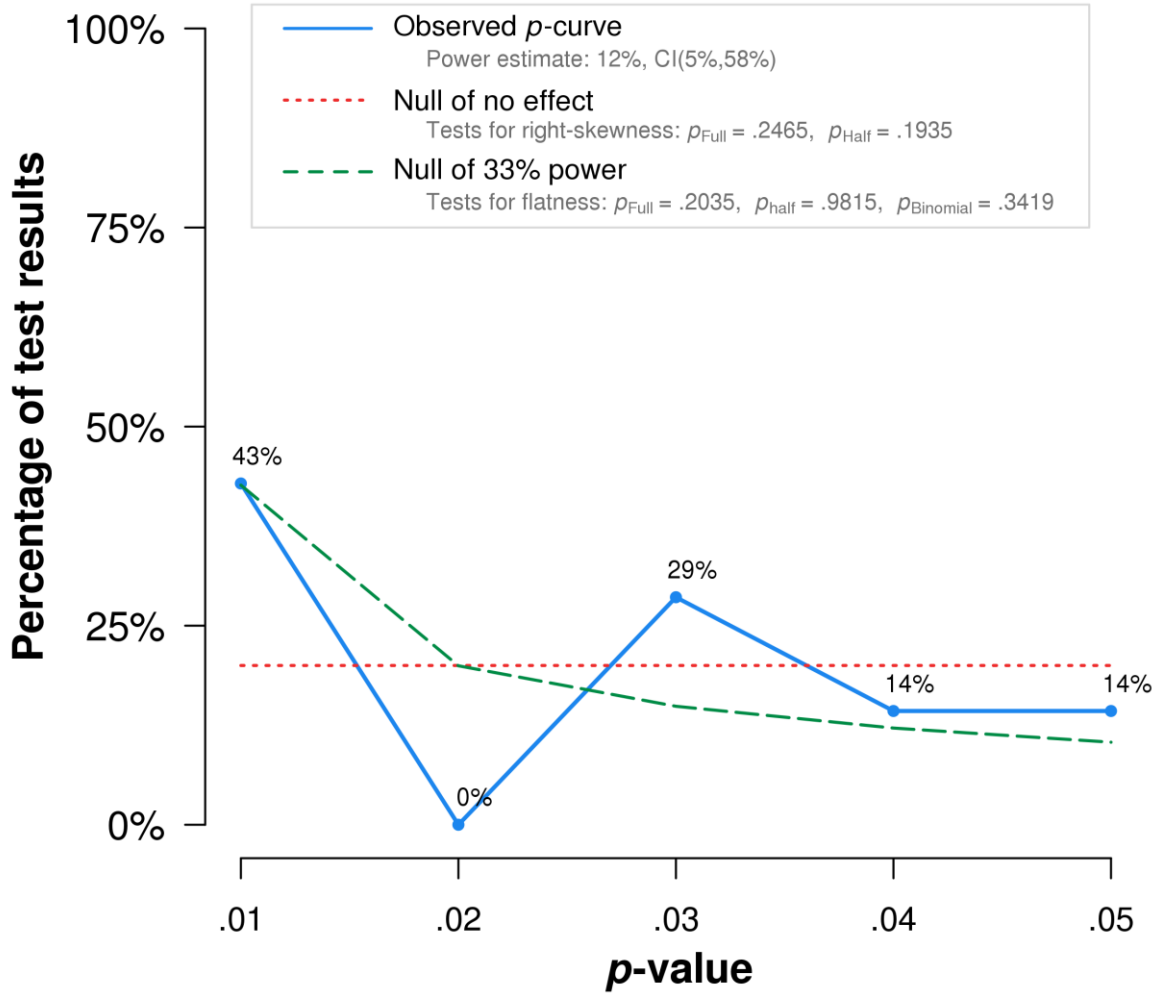
right skewing for half a p-curve (with $p < .05$); or B, for both half and full p-curves (both p values < .10). On the contrary, the null hypothesis states that there is no evidential value in the literature that was analyzed.

While a visual inspection seems to be more in line with a right-skewed p-curve, especially with all pilot studies included, the null hypothesis was not rejected as the p-curve was not significantly right skewed. According to the data analysis, neither criteria were met for 1, the main curve (full p-curve, Z=-0.69, p=.2465; half p-curve Z=-0.86, p=.1935); 2, nor robustness check (full p-curve, Z=-0.25, p=.3996; half p-curve Z=0.7, p=.7593); 3, nor when the pilot studies were included (full p-curve, Z=-1.31, p=.0957; half p-curve Z=-0.28, p=.3894) - In summary, the null hypothesis cannot be rejected. The test does not indicate that the studies contain evidential value. This is in line with the p-curve of the elderly priming (Lakens, 2017).

However, two important things should be highlighted. Firstly, based on a visual inspection of the p-curve, it should be noted that obvious evidence is lacking for obvious p-hacking. Indeed, the studies do not seem to be skewed to the left (however, see further discussion below). Secondly, it is important to note that a situation when the null hypothesis is not rejected can be caused by both A, the set of studies lacks evidential value and B, there is not enough information available - p-curve is too noisy. Therefore, in the following analysis, we were interested in if the p-curve was flatter than what would be expected if the studies were powered at 33%.

We, therefore, analyzed if the observed p-curve was significantly flatter in comparison to the existing p-curve from studies with an average of 33% power. The alternative hypothesis states that the evidential value is inadequate. The null hypothesis states that the evidential value is not inadequate or absent. The binomial test, as well as continuous test computed with Stouffer's method, indicated that the null hypothesis could not be rejected in any of the three conducted p-curves. In summary, the analysis does not indicate that the evidential value is inadequate or absent. Thus, we are unable to reject the null hypothesis that the body of analyzed studies examines a detectable effect. This mean that there is no evidence for the presence or absence of evidential value. The results are inconclusive and thus, more p-values are needed to make a definitive conclusion.

Additionally, we analyzed the statistical power of the studies that were included. The power with 90% confidence intervals was 12% (90% CI, 5-58%) for the main results; 7% (90% CI, 5-43%) for the robustness analysis; and 19% (90% CI, 5-56%) for all the reported studies (pilot studies included). This value represents an estimate of how many studies would have significant results if replicated. From the main studies, approximately 12% would be expected to be replicable. From 9 experiments, it is expected that one experiment (maximum 4-5) would be replicable. This is, hypothetically, more than 5% the value expected if the results were solely due to random variation, nevertheless, this is far from an optimal scenario. The results are graphically depicted in Figure1, Figure 2 and further detailed in Supplement 2.



Note: The observed *p*-curve includes 7 statistically significant ($p < .05$) results, of which 4 are $p < .025$
There were 2 additional results entered but excluded from *p*-curve because they were $p > .05$.

Figure 1 Main analysis *Note. The red dotted line illustrates the distribution of p-values that is expected if there is no effect, while the green dashed line illustrates the distribution of p-values if the effect exists and studies were powered at 33%. The blue line indicates the distribution of the analyzed p-values. The figure was generated by P-curve app v.4.06*

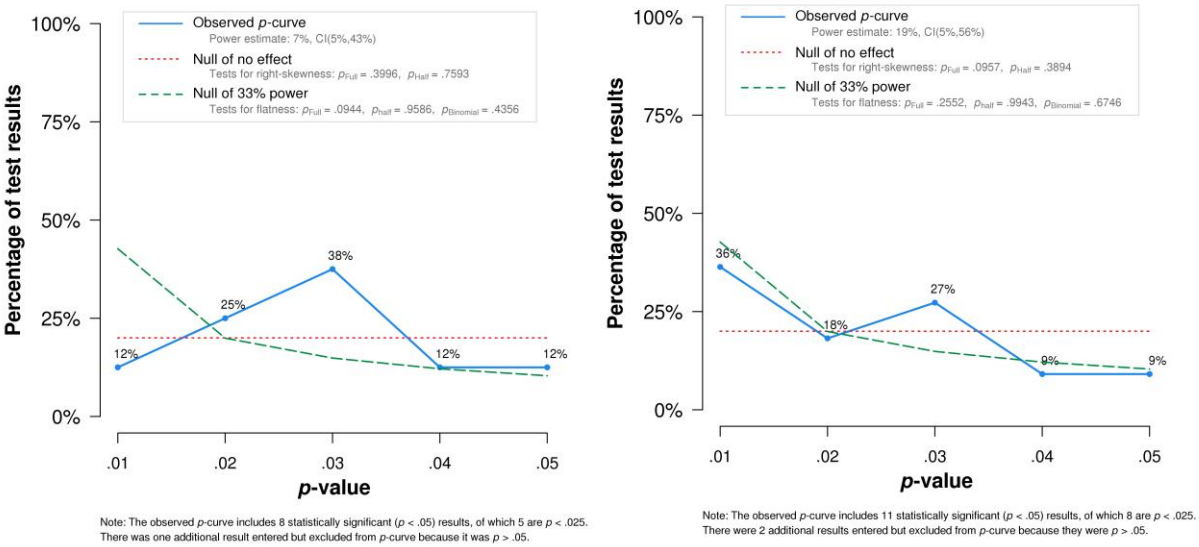


Figure 2 (A) Robustness results and (B) results from all studies (pilot studies included). *Note. The red dotted line illustrates the distribution of p-values that is expected if there is no effect; while the green dashed line illustrates the distribution of p-values if the effect exists and studies were powered at 33%. The blue line indicates the distribution of analyzed p-values. The figure was generated by P-curve app v.4.06*

As the estimated power was rather small - 12% - 19% (depending on the analyzed variant) we further analyzed the sample size of all studies. The results are depicted in Figure 3. The figure indicates low sample sizes. For all studies, the mean N per cell was 28 participants; the median and mode were 22 participants.

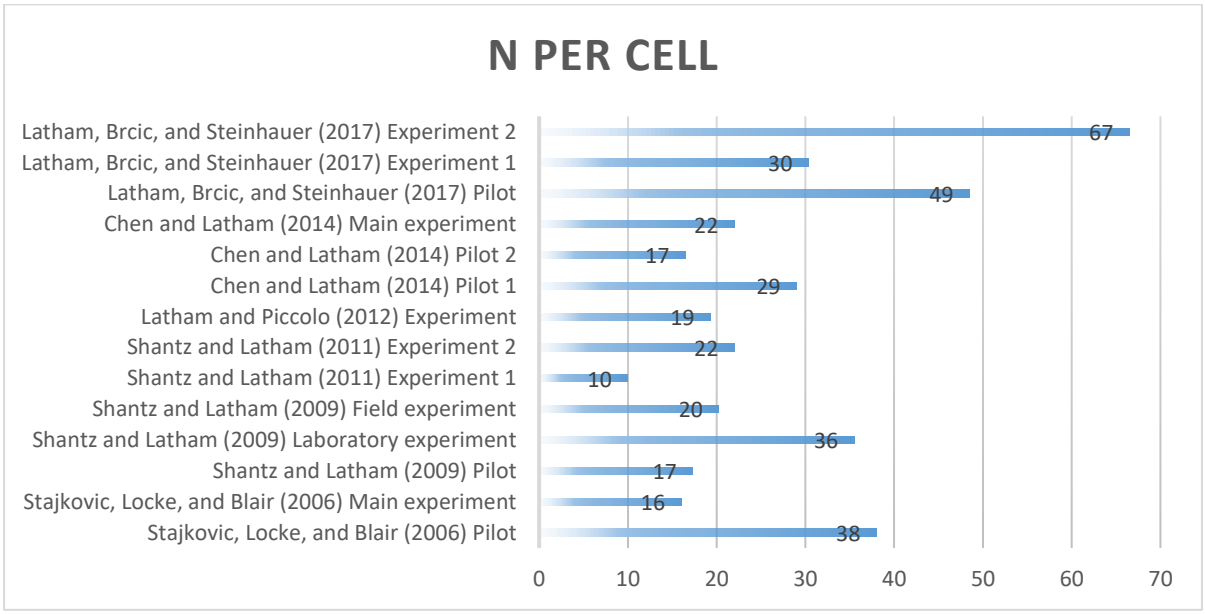


Figure 3. Sample size (per experimental cell) of reviewed studies.

In order to have power of at least 80%, far more participants than 28 (in one extreme case 10 participants per cell) seem to be needed, especially considering the effect size that could be

expected. In fact, according to a meta-analysis of the behavioral priming effect caused by words (Weingarten et al., 2016), it is necessary to have more than 145 participants per cell for 80% power to obtain a small to medium effect size of d=.3. Even when medium effect size of d=.56 from small meta-analysis conducted by Shantz and Latham (2011) will be used, more than 41 (one tailed test) or 52 (two tailed test) participants per cell will be necessary. However, it is reasonable to expect that much more participants would be necessary as meta-analytically examining the effect size could be considerably overestimated due to publication bias. The observed N (in many cases lower than approximately 25 per condition) is only adequate if the effect size is more than d=80. This is, however, highly unlikely. Therefore, more power to detect an effect of interest seems to be crucial for future studies.

4. SOME RECOMMENDATIONS FOR FUTURE RESEARCH AND LIMITATIONS OF THE PRESENT STUDY

Based on the present narrative review and consequent P-curve analysis, some recommendations can be drawn regarding future research.

Firstly, considering need for stronger theoretical basis, the Goal-Setting Theory seems to be a promising theoretical platform. Dijksterhuis (2014) has argued that productive theory-driven research is crucial but is somehow absent. Similarly, Bargh (2006) has argued in his article "What have we been priming all these years" that it is necessary to move to a new stage of research where moderators and mediators are systematically examined. Similar points have been made by other authors (e.g. Higgins & Eitam, 2014; Latham, 2016; Loersch & Payne, 2014), leading to the proposal of various theoretical models (for a review see e.g. Kačmár & Lovaš, 2018). Among them, Goal-setting theory could be found (see e.g. Latham, 2016; Latham et al., 2010; Locke & Latham, 2006).

However, problematic research practices have been identified in recent years as potentially relevant topic (e.g. John, Loewenstein, & Prelec, 2012; Simmons et al., 2011); and many previous findings have been questioned, priming studies including. Therefore, is was proposed that it is necessary to ensure that sound research practices should be implemented for creating body of literature that can be trusted (De Boeck & Jeon, 2018; Munafò et al., 2017; Nelson et al., 2018). Indeed, our p-curve analysis did not find evidential value in the analyzed research literature. Although there many statistically significant results, this does not guarantee that the research has any evidential value and that could be trusted.

On the other hand, it is also true that we cannot say that the analyzed research is lacking evidential value. For now, the results have been rather inconclusive. Therefore, we have additionally tried to more qualitatively analyze the merits and limitations of the present line of research in the following part, proposing some recommendation for future research. The analysis has indicated that while some aspects have been established in a very good manner, others need improved.

Firstly, in order to establish a knowledge base that can be trusted, it seems to be crucial to conduct a replication of existing studies (Brandt et al., 2014). Some argue that direct replication is preferable (Simons, 2014); while other prefer conceptual replication (Crandall & Sherman, 2016). In our view, both are crucial considering the purpose of the study and various phases of research (see e.g. Hüffmeier, Mazei, & Schultze, 2016). In the reviewed literature, multiple

types of replications can be found, even field experiments considered as rare and valuable (Eden, 2017; Maner, 2016). For instance, Shantz and Latham (2011) replicated the previous results by Shantz and Latham (2009) in a call-centre. This practice should be continued. However, it is necessary to extend it further to implement various kinds of replication in diverse phases of the research whilst reflecting on its strengths, limitations and purpose (Hüffmeier et al., 2016). Lock (2015) has proposed replication with variations as the basic building block of theory. However, without direct replications that can establish the effect of interest in a first place, this could be rather problematic.

Additionally, other sound research practices should be implemented simultaneously. Some authors have argued that that using problematic research practices (e.g. P-hacking) seems to be both (A) problematic (Simmons et al., 2011); and (B) widespread (John et al., 2012). In the present analysis, we did not find strong (or more blatant) evidence for p-hacking (although the present method has some limitations as discussed below). These positive trends should be bolstered by explicitly avoiding problematic research practices such as P-hacking and Harking, decreasing research degrees of freedom by pre-registering planned analysis (Munafò et al., 2017; van ’t Veer & Giner-Sorolla, 2016).

Moreover, another important factor, publication bias, could be highly problematic. However, it is hard to estimate publication bias in the present context. Only six studies were analyzed, and it is hard to find all the non-significant studies that have been conducted. It is, however, possible that they exist in a file-drawer form. Therefore, we tried to overcome this limitation by conducting a p-curve analysis which only analyzed the statistically significant results. The file-drawer problem was, therefore, not so prominent in the present analysis, nevertheless, publication bias could and should be minimalized in the long run. This could be done by implementing a registered-report (RR) format. In RR, peer-review is conducted before data acquisition and accepted in advance if the reviewed procedure was implemented regardless of results. Accordingly, the article is not published based on "sexy" significant results but based on strong theory and sound method irrespective of results (Nosek & Lakens, 2014). Moreover, using a more classical publishing schema, non-significant findings should be published if the method and theory is sound. Public repository of data should be used as well and data from studies that were not published should be systematically added. Moreover, Open-science practices could be bolstered (Munafò et al., 2017; Spellman, Gilbert, & Corker, 2017). Data could be available for future re-analysis and meta-analysis. Furthermore, additional statistical analysis could be implemented (e.g. Multiverse analysis; Bayesian analysis; and so on), bolstering interpretations.

In future research, it is critical to address the issue of low statistical power. This is not only a problem in the reviewed research. Some authors have pointed out that psychological research is highly underpowered in general (Fraley & Vazire, 2014) and this problem can be even beyond psychological research (see e.g. Button et al. 2013 for a similar problem in the field of neuroscience). In the present analysis, only 7-19% (90% CI, 5-56%) of power was estimated. Therefore, only approximately 7-19% (56% at best) could be expected to be replicated.

Furthermore, some studies found an unlikely big effect size while they were unable to detect lower effect size that would be more realistic to expect in the present scenario. In order to address this issue in future research, a power analysis should be performed when planning a study. The expected number of participants could be estimated, based on the smallest effect

size of interest (SOI) (Albers & Lakens, 2018). There are many tools and easily implementable manuals (see e.g. Perugini, Gallucci, & Costantini, 2018), even for Bayes factor analysis (Schönbrodt & Wagenmakers, 2017). It is true that beyond computational issues, implementation issues could occur as it is necessary to sample more participants, nevertheless, as LeBel, Campbell, & Loving (2017) have pointed out, the benefits outweigh the costs in a long run. Furthermore, various procedures such as Sequential analysis (Daniël Lakens, 2014); or big multi-lab collaborations (e.g. "Psychological Science Accelerator"; Moshontz et al., 2018) could be implemented to handle the high-powered studies.

Additionally, as illustrated by the present study, narrative reviews could be replaced or at least complemented by quantitative methods. In the analyzed line of research, a meta-analysis conducted by Shantz and Latham (2011) could be found which is a good sign, nevertheless, it is important to address publication bias in the future. For instance, in the present study, we conducted a p-curve analysis to partially overcome this limitation. However, it is important to note that even p-curve have some limitations. For instance, when analyzing potential p-hacking, the p-curve analysis implicitly expects the researcher to reach p-value of 0,05. While this expectation is reasonable, more ambitious p-hacking could occur and thus complicate the interpretation of the p-curve (although see e.g. Simonsohn et al., 2015 for a discussion on this topic and provided solution).

Therefore, it is important to note that an anomaly worth mentioning was found in the present analysis. As pointed out by reviewer, some studies had relatively low sample sizes. However, relatively big effect size was observed (see e.g. Schanz and Latham, 2009, 2011). In combination with the lower number of p-curved studies, it is possible that such studies led to more right skewed p-curves, hiding more blatant p-hacking. Based on this observation, we strongly recommend conducting more types of analysis in the Meta-analytical realm (such as p-uniform) when bigger body of literature will accumulate; and, crucially, to directly replicate studies with a surprisingly big effect size with much more power in combination with pre-registration.

There are also some limitations related to other aspects of the present article. In the present analysis, only six studies with 9 to13 results were analyzed. Even though we tried our best to find adequate studies and to provide results from not one, but three p-curves, more studies would certainly provide more conclusive results. We, thus, encourage other researchers to conduct a p-curve analysis with more research literature in the future.

One can ask, why we conducted p-curve analysis with "such small" amount of studies, however, it is important to remind that even less studies can provide evidence for evidential value (see e.g. Simonsohn et al., 2014b), and the goal was to assess the attributes of literature existing so far. When empirical evidence accumulates, more studies should be included, providing more representative results.

In addition, the studies selected for this review and p-curve analysis belonged to one specific line of research, being heterogeneous in nature. They were not focused on one specific priming effect, but on priming (non-conscious goal activation) in the context of the Goal-Setting Theory. While one could argue that this factor could play some role, it is important to note that it is not necessary for studies included in the P-curve analysis to be homogenous - to have uniform effect sizes (Simonsohn et al., 2014b).

In a similar fashion, one could ask what effect we are trying to analyze. It is, however, not the question we tried to answer. We were primarily interested in reviewing and p-curving one specific line of research rather than one specific kind of priming effect. As pointed out by Simonsohn et al., (2014a), it is possible to assess the evidential value of a specific effect as well as the "evidential value of findings aggregated by article, author, journal, or method of analysis" (p. 543). Accordingly, our question seems to be legitimate as well as it will be legitimate to analyze a specific priming effect in a future. In fact, we encourage analyzing a specific priming effect when more literature related to the topic accumulates. An analysis of moderation factors could also be beneficial in a future, although, another method of analysis should be implemented for this purpose.

Despite trying to do our analysis well and as transparently as possible, a mistake could occur. Therefore, we encourage additional analyses by other researchers. Simultaneously, we believe that even if the results of the present analysis had been more conclusive to provide or dismiss the evidential value, recommendations such as adequate power; pre-registration; RR replications; and open science practices could be beneficial for this (and any) research avenue irrespective of potential errors in the present study.

In a nutshell, this article reviewed and p-curved one specific line of research - priming (non-conscious goal activation) in the context of the Goal-Setting Theory. The good news is that this line of research has some merits such as potentially strong theoretical background; various existing replications of previous studies; field experiments; conducted meta-analysis and so on. Nevertheless, according to our analysis, some aspects should be improved - insufficient statistical power; conducting registered reports/pre-registrations, limiting researchers' degrees of freedom; eliminating publication bias and conducting highly powered direct pre-registered replications.

Even though many significant results exist in the reviewed line of research, the p-curve analysis could not conclude that they have evidential value. On the other hand, it was not possible to conclude that evidential value is lacking. This pattern of results indicates that more high-quality studies are necessary to be conducted and, meanwhile, some research recommendations should be implemented for creating a knowledge base with some degree of verisimilitude that can be trusted.

References

Albers, C., & Lakens, D. (2018). When power analyses based on pilot data are biased: Inaccurate effect size estimators and follow-up bias. Journal of Experimental Social Psychology, 74(Supplement C), 187–195. https://doi.org/10.1016/j.jesp.2017.09.004

Bargh, J. A. (2006). What have we been priming all these years? On the development, mechanisms, and ecology of nonconscious social behavior. European Journal of Social Psychology, 36(2), 147–168. https://doi.org/10.1002/ejsp.336

Bargh, J. A., Chen, M., & Burrows, L. (1996). Automaticity of social behavior: Direct effects of trait construct and stereotype activation on action. Journal of Personality and Social Psychology, 71(2), 230–244. https://doi.org/10.1037/0022-3514.71.2.230

Bargh, J. A., Lee-Chai, A., Barndollar, K., Gollwitzer, P. M., & Trötschel, R. (2001). The Automated Will: Nonconscious Activation and Pursuit of Behavioral Goals. Journal of personality and social psychology, 81(6), 1014–1027.

Brandt, M. J., Ijzerman, H., Dijksterhuis, A., Farach, F. J., Geller, J., Giner-Sorolla, R., … Van't Veer, A. (2014). The replication recipe: What makes for a convincing replication? Journal of Experimental Social Psychology, 50, 217–224.

Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., & Munafò, M. R. (2013). Power failure: why small sample size undermines the reliability of neuroscience. Nature Reviews Neuroscience, 14(5), 365–376. https://doi.org/10.1038/nrn3475

Cesario, J. (2014). Priming, replication, and the hardest science. Perspectives on Psychological Science, 9(1), 40–48.

Cesario, J., & Jonas, K. J. (2014). Replicability and Models of Priming: What a Resource Computation Framework can Tell us About Expectations of Replicability. Social Cognition; New York, 32(Supplement), 124–136. http://dx.doi.org/101521soco201432supp124

Crandall, C. S., & Sherman, J. W. (2016). On the scientific superiority of conceptual replications for scientific progress. Journal of Experimental Social Psychology, 66, 93–99. https://doi.org/10.1016/j.jesp.2015.10.002

De Boeck, P., & Jeon, M. (2018). Perceived crisis and reforms: Issues, explanations, and remedies. Psychological Bulletin, 144(7), 757–777. https://doi.org/10.1037/bul0000154

Dijksterhuis, A. (2014). Welcome Back Theory! Perspectives on Psychological Science, 9(1), 72–75.

Dijksterhuis, A., & van Knippenberg, A. (1998). The relation between perception and behavior, or how to win a game of Trivial Pursuit. Journal of Personality and Social Psychology, 74(4), 865–877. https://doi.org/10.1037//0022-3514.74.4.865

Doyen, S., Klein, O., Pichon, C.-L., & Cleeremans, A. (2012). Behavioral priming: it's all in the mind, but whose mind? PloS one, 7(1), e29081.

Eden, D. (2017). Field Experiments in Organizations. Annual Review of Organizational Psychology and Organizational Behavior, 4(1), 91–122. https://doi.org/10.1146/annurev-orgpsych-041015-062400

Fraley, R. C., & Vazire, S. (2014). The N-Pact Factor: Evaluating the Quality of Empirical Journals with Respect to Sample Size and Statistical Power. PLoS ONE, 9(10), e109019. https://doi.org/10.1371/journal.pone.0109019

Fujita, K., & Trope, Y. (2014). Structured Versus Unstructured Regulation: On Procedural Mindsets and the Mechanisms of Priming Effects. Social Cognition; New York, 32(Supplement), 68–87. http://dx.doi.org/101521soco201432supp68

Higgins, E. T., & Eitam, B. (2014). Priming… Shmiming: it's about knowing when and why stimulated memory representations become active. Social Cognition, 32(Supplement), 225–242.

Higgins, E. T., Rholes, W. S., & Jones, C. R. (1977). Category accessibility and impression formation. Journal of experimental social psychology, 13(2), 141–154.

Hüffmeier, J., Mazei, J., & Schultze, T. (2016). Reconceptualizing replication as a sequence of different studies: A replication typology. Journal of Experimental Social Psychology, 66, 81–92. https://doi.org/10.1016/j.jesp.2015.09.009

Chen, X., & Latham, G. P. (2014). The effect of priming learning vs. performance goals on a complex task. Organizational Behavior and Human Decision Processes, 125(2), 88–97. https://doi.org/10.1016/j.obhdp.2014.06.004

John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the Prevalence of Questionable Research Practices With Incentives for Truth Telling. Psychological Science, 23(5), 524–532. https://doi.org/10.1177/0956797611430953

Kačmár, P., & Lovaš, L. (2018). Nonconscious goal activation in the broader perspective of social priming. Ceskoslovenska Psychologie, 62(1), 75–91.

Kerr, N. L. (1998). HARKing: Hypothesizing After the Results are Known. Personality and Social Psychology Review, 2(3), 196–217. https://doi.org/10.1207/s15327957pspr0203_4

Lakens D. (2014). Performing high-powered studies efficiently with sequential analyses, Eur. J. Soc. Psychol., 44, pages 701–710, doi: 10.1002/ejsp.2023.

Lakens, D. (2017). Professors Are Not Elderly: Evaluating the Evidential Value of Two Social Priming Effects through P-Curve Analyses. PsyArXiv. https://doi.org/10.17605/OSF.IO/3M5Y9

Latham, G. P. (2016). Goal setting: a possible theoretical framework for examining the effect of priming goals on organizational behavior. Current Opinion in Psychology, 12, 85–88. https://doi.org/10.1016/j.copsyc.2016.07.005

Latham, G. P., Brcic, J., & Steinhauer, A. (2017). Toward an Integration of Goal Setting Theory and the Automaticity Model: GOAL DIFFICULTY, EFFORT, AND CHOICE. Applied Psychology, 66(1), 25–48. https://doi.org/10.1111/apps.12087

Latham, G. P., & Piccolo, R. F. (2012). The effect of context-specific versus nonspecific subconscious goals on employee performance. Human Resource Management, 51(4), 511–523. https://doi.org/10.1002/hrm.21486

Latham, G. P., Stajkovic, A. D., & Locke, E. A. (2010). The Relevance and Viability of Subconscious Goals in the Workplace. Journal of Management, 36(1), 234–255. https://doi.org/10.1177/0149206309350777

LeBel, E. P., Campbell, L., & Loving, T. J. (2017). Benefits of open and high-powered research outweigh costs. Journal of Personality and Social Psychology, 113(2), 230–243. https://doi.org/10.1037/pspi0000049

Locke, E. A., & Latham, G. P. (2006). New directions in goal-setting theory. Current directions in psychological science, 15(5), 265–268.

Locke, E. A., & Latham, G. P. (2015). Chapter Four-Breaking the Rules: A Historical Overview of Goal-Setting Theory. Advances in Motivation Science, 2, 99–126.

Loersch, C., & Payne, B. K. (2014). Situated Inferences and the What, Who, and Where of Priming. Social Cognition; New York, 32(Supplement), 137–151. http://dx.doi.org/101521soco201432supp137

Maner, J. K. (2016). Into the wild: Field research can increase both replicability and real-world impact. Journal of Experimental Social Psychology, 66, 100–106. https://doi.org/10.1016/j.jesp.2015.09.018

Moshontz, H., Campbell, L., Ebersole, C. R., IJzerman, H., Urry, H. L., Forscher, P. S., … Chartier, C. R. (2018). The Psychological Science Accelerator: Advancing Psychology Through a Distributed Collaborative Network. Advances in Methods and Practices in Psychological Science, 1(4), 501–515. https://doi.org/10.1177/2515245918797607

Munafò, M. R., Nosek, B. A., Bishop, D. V. M., Button, K. S., Chambers, C. D., Sert, N. P. du, … Ioannidis, J. P. A. (2017). A manifesto for reproducible science. Nature Human Behaviour, 1(1), s41562-016-0021–016. https://doi.org/10.1038/s41562-016-0021

Nelson, L. D., Simmons, J., & Simonsohn, U. (2018). Psychology's Renaissance. Annual Review of Psychology, 69(1), 511–534. https://doi.org/10.1146/annurev-psych-122216-011836

Nosek, B. A., & Lakens, D. (2014). Registered Reports: A Method to Increase the Credibility of Published Results. Social Psychology, 45(3), 137–141. https://doi.org/10.1027/1864-9335/a000192

O'Donnell, M., Nelson, L. D., Ackermann, E., Aczel, B., Akhtar, A., Aldrovandi, S., … Zrubka, M. (2018). Registered Replication Report: Dijksterhuis and van Knippenberg (1998). Perspectives on Psychological Science, 1745691618755704. https://doi.org/10.1177/1745691618755704

Perugini, M., Gallucci, M., & Costantini, G. (2018). A Practical Primer To Power Analysis for Simple Experimental Designs. International Review of Social Psychology, 31(1). https://doi.org/10.5334/irsp.181

Shanks, D. R., Newell, B. R., Lee, E. H., Balakrishnan, D., Ekelund, L., Cenac, Z., … Moore, C. (2013). Priming intelligent behavior: An elusive phenomenon. PloS one, 8(4), e56515.

Shantz, A., & Latham, G. (2011). The effect of primed goals on employee performance: Implications for human resource management. Human Resource Management, 50(2), 289–299. https://doi.org/10.1002/hrm.20418

Shantz, A., & Latham, G. P. (2009). An exploratory field experiment of the effect of subconscious and conscious goals on employee performance. Organizational Behavior and Human Decision Processes, 109(1), 9–17. https://doi.org/10.1016/j.obhdp.2009.01.001

Schönbrodt, F. D., & Wagenmakers, E.-J. (2017). Bayes factor design analysis: Planning for compelling evidence. Psychonomic Bulletin & Review, 1–15. https://doi.org/10.3758/s13423-017-1230-y

Schweinsberg, M., Madan, N., Vianello, M., Sommer, S. A., Jordan, J., Tierney, W., … Uhlmann, E. L. (2016). The pipeline project: Pre-publication independent replications of a single laboratory's research pipeline. Journal of Experimental Social Psychology, 66, 55–67. https://doi.org/10.1016/j.jesp.2015.10.001

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant. Psychological Science, 22(11), 1359–1366. https://doi.org/10.1177/0956797611417632

Simmons, J. P., & Simonsohn, U. (2017). Power Posing: P- Curving the Evidence. Psychological Science, 28(5), 687–693. https://doi.org/10.1177/0956797616658563

Simons, D. J. (2014). The value of direct replication. Perspectives on Psychological Science, 9(1), 76–80.

Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014a). P-curve: A key to the file-drawer. Journal of Experimental Psychology: General, 143(2), 534–547. https://doi.org/10.1037/a0033242

Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014b). P-Curve and Effect Size: Correcting for Publication Bias Using Only Significant Results. Perspectives on Psychological Science, 9(6), 666–681. https://doi.org/10.1177/1745691614553988

Simonsohn, U., Simmons, J. P., & Nelson, L. D. (2015). Better P-curves: Making P-curve analysis more robust to errors, fraud, and ambitious P-hacking, a Reply to Ulrich and Miller (2015). Journal of Experimental Psychology: General, 144(6), 1146–1152. https://doi.org/10.1037/xge0000104

Spellman, B., Gilbert, E., & Corker, K. S. (2017). Open Science: What, Why, and How. PsyArXiv. https://doi.org/10.17605/OSF.IO/AK6JR

Stajkovic, A. D., Locke, E. A., & Blair, E. S. (2006). A first examination of the relationships between primed subconscious goals, assigned conscious goals, and task performance. The Journal of Applied Psychology, 91(5), 1172–1180. https://doi.org/10.1037/0021-9010.91.5.1172

van ’t Veer, A. E., & Giner-Sorolla, R. (2016). Pre-registration in social psychology—A discussion and suggested template. Journal of Experimental Social Psychology, 67, 2–12. https://doi.org/10.1016/j.jesp.2016.03.004

Weingarten, E., Chen, Q., Mcadams, M., Yi, J., Hepler, J., & Albarracín, D. (2016). From primed concepts to action: A meta-analysis of the behavioral effects of incidentally presented words. Psychological bulletin, 142(5), 472–497.

Wheeler, S. C., DeMarree, K. G., & Petty, R. E. (2014). Understanding Prime-to-Behavior Effects: Insights from the Active-Self Account. Social Cognition; New York, 32(Supplement), 109–123. http://dx.doi.org/101521soco201432supp109

SUPPLEMENT 1:

Supplement figure 1 Priming stimuli used in reviewed research (Adapted from Chen & Latham, 2014; Latham, Brcic, & Steinhauer, 2017; Latham & Piccolo, 2012; Shantz & Latham, 2011, 2009; Stajkovic, Locke, & Blair, 2006).

SUPPLEMENT 2:

Supplement table 1

More extensive results of the P-Curve analysis

	(1) Main analysis			(2) Robustness analysis			(3) Pilot studies included		
	Binomial Test	**Continuous Test**		**Binomial Test**	**Continuous Test**		**Binomial Test**	**Continuous Test**	
	(Share of results p<.025)	*(Aggregate with Stouffer Method)*		*(Share of results p<.025)*	*(Aggregate with Stouffer Method)*		*(Share of results p<.025)*	*(Aggregate with Stouffer Method)*	
		Full p-curve (p's<.05)	Half p-curve (p's<.025)		Full p-curve (p's<.05)	Half p-curve (p's<.025)		Full p-curve (p's<.05)	Half p-curve (p's<.025)
1) Studies contain evidential value. *(Right skew)*	p =.5	Z= -0.69, p =.2465	Z= -0.86, p =.1935	p =.3633	Z= -0.25, p =.3996	Z= 0.7, p =.7593	p =.1133	Z= -1.31, p =.0957	Z= -0.28, p =.3894
2) Studies' evidential value, if any, is inadequate. *(Flatter than 33% power)*	p =.3419	Z= -0.83, p =.2035	Z= 2.09, p =.9815	p =.4356	Z= -1.31, p =.0944	Z= 1.73, p =.9586	p =.6746	Z= -0.66, p =.2552	Z= 2.53, p =.9943
	Statistical Power			**Statistical Power**			**Statistical Power**		
Power of tests included in *p*-curve *(correcting for selective reporting)*	Estimate: 12%			Estimate: 7%			Estimate: 19%		
	90% Confidence interval: (5% , 58%)			90% Confidence interval: (5% , 43%)			90% Confidence interval: (5% , 56%)		

Contact

Pavol Kačmár
Department of Psychology, Faculty of Arts, Pavol Jozef Šafárik University in Košice
Moyzesova 9, 040 01 Košice, Slovakia
E-mail: pavol.kacmar@upjs.sk