

Research Reports

Negative Results in European Psychology Journals

Martin Rachev Vasilev*^a

[a] Department of Psychology, Sofia University "St. Kliment Ohridski", Sofia, Bulgaria.

Abstract

Psychologists have long speculated that the research literature is largely dominated by positive findings, but yet there is little data to justify these speculations. The present study investigates the extent to which negative findings exist in the literature by reviewing articles published in five European psychology journals. While no temporal change was observed, the results indicate that almost all (95.4%) articles published in 2001, 2006 and 2011 found support for at least one tested hypothesis. Moreover, a sizable number (73%) of papers found support for all tested hypotheses. It is argued that the lack of negative findings can have a detrimental effect on the ability to systemize scientific knowledge, the way science is practiced, and the rate of replications in psychology. Publishing positive findings may be very important for making progress in our field, but negative findings are also crucial for maintaining its scientific integrity. When we base our conclusions on results that support our predictions and ignore data to the contrary, we run the risk of creating a biased view of reality that gives us little confidence in the validity and applicability of our findings.

Keywords: negative results, publication bias, file-drawer problem, European journals, psychological research

Europe's Journal of Psychology, 2013, Vol. 9(4), 717–730, doi:10.5964/ejop.v9i4.590

Received: 2013-02-13. Accepted: 2013-08-26. Published (VoR): 2013-11-29.

Handling Editor: Vlad Glăveanu, Aalborg University, Aalborg, Denmark.

*Corresponding author at: Sofia University "St. Kliment Ohridski", Department of Psychology, 15 Tsar Osvoboditel Blvd., 1504 Sofia, Bulgaria. E-mail: martin.r.vasilev@gmail.com



This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

There has been a growing concern that publication decisions in scientific journals largely depend on the study outcomes. Positive results that report significant outcomes and confirm the researcher's expectations are more likely to be published and have higher odds of being fully reported (Dwan et al., 2008). Negative results, on the other hand, are often difficult to publish as they fail to reach the conventional significance levels and do not support the tested hypotheses. As a result of this, journals are filled with positive findings, while negative results are largely inaccessible to the scientific community.

Concerns with this practice are not new (e.g., McNemar, 1960; Smith, 1956; Tullock, 1959) and have sporadically occupied psychologists for decades. Yet, besides the fact that negative results are important for designing better systematic reviews and meta-analyses, few other arguments are given as to why negative results are important in psychology.

The present article revisits the problem in two ways. First, it investigates empirically the extent to which negative results exist in the literature by examining articles published in five European psychology journals. Second, it

discusses how the general lack of negative results contributes to other problems that empirical psychology faces today (e.g., see [Nosek, Spies, & Motyl, 2012](#); [Pashler & Wagenmakers, 2012](#)).

Meta-Analysis: Problems With the Lack of Negative Results

Publishing positive findings that support the tested hypotheses is important for the advancement of scientific knowledge. However, different problems can arise when the literature is dominated by a disproportionate amount of positive results. First of all, this can be a potentially major threat to systematic reviews and meta-analyses ([Bradley & Gupta, 1997](#); [Egger & Smith, 1998](#); [Rothstein, Sutton, & Borenstein, 2005b](#); [Thornton & Lee, 2000](#); [Torgerson, 2006](#); although see [Dalton, Aguinis, Dalton, Bosco, & Pierce, 2012](#)).

Meta-analytic studies can be used to shed light on important practical and empirical questions, but the confidence that we have in their results depends on the extent to which they are free from bias ([Banks, Kepes, & Banks, 2012](#)). If studies reporting positive results are more likely to be published, they will also be easier to obtain compared to negative and unpublished ones ([Torgerson, 2006](#)). This could potentially bias any attempts to do a systematic review or to derive valid estimates by pooling data from multiple studies ([Thornton & Lee, 2000](#)). Also, because published studies may systematically differ from unpublished ones, meta-analysis based only on published data may reach misleading conclusions. This holds major practical implications because meta-analysis aggregates data from many studies and is thus a good reference source for practitioners, policy makers and the general public.

The consequences of ignoring negative studies, however, are not only constrained to the meta-analysis at hand and the validity of its conclusions. When a meta-analysis that ignores relevant negative studies is published, and the conclusions are later found to be misleading or incorrect, the perception is fostered that meta-analysis cannot be trusted ([Rothstein, Sutton, & Borenstein, 2005a](#)). Another problem with a misleading meta-analysis is that it slows scientific progress and it may give us the false illusion that conclusions are supported by data from multiple studies when, in fact, they are biased or overstated due the excessive number of positive results. These consequences probably don't happen often, but when they do, they can have lasting effects on the field.

Although different methods for overcoming this bias have been proposed (see below), [Thornton and Lee \(2000\)](#) argue that the best way to eliminate its effects on meta-analysis is to stop it from happening in the first place. For this reason, negative studies play an important role in systemizing scientific knowledge and the trust that we have in the validity of our conclusions.

The Pressure to Publish Positive Results

A second problem with the excessive amount of positive results in the literature is that they can influence research decisions. Because negative results are often difficult to publish, this can indirectly encourage questionable research practices that are aimed at obtaining publishable results ([Nosek et al., 2012](#)). In the publish-or-perish world of academia, publishing as many papers as possible is essential for one's career. Yet, researchers eventually find themselves in a situation in which they fail to obtain results that reach the conventional significance level ($p < 0.05$). Such a situation can be tempting in itself and encourage research decisions that vary in their degree of scientific misconduct but that are ultimately aimed at getting one's results published (e.g., see [Neuroskeptic, 2012](#)).

The prevalence of questionable research practices was aptly demonstrated by a recent survey of over 2 000 psychologists ([John, Loewenstein, & Prelec, 2012](#)). It measured the extent to which they engage in ten such

practices that range from milder ones (e.g., not reporting all conditions or dependent variables) to more severe ones, such as falsifying data or deciding to exclude data after looking at their effect on the results. The survey revealed that 94% of them admitted to having engaged in at least one questionable research practice. Moreover, 66% of the surveyed researchers admitted that they have failed to report all dependent variables and 58% admitted that they have checked for significant results before deciding whether to collect more data.

The results from this survey are also in line with recent studies that have shown how “unacceptably” easy it can be to obtain and report positive results (Simmons, Nelson, & Simonsohn, 2011). Because psychologists normally have great flexibility during the process of data collection and analysis, the rate of discovering false positive results is inflated and researchers can easily accumulate significant findings that they can later report. Thus, it can be argued that the pressure to come up with positive findings stimulates questionable research practices among psychologists.

Negative Results and Replications

The lack of negative results in the literature is also closely related to another problem in empirical psychology – the low rate of replications (Makel, Plucker, & Hegarty, 2012; Pashler & Wagenmakers, 2012). The lack of replications of research findings is a concern in many other scientific disciplines (Ioannidis, 2005), but the situation in psychology seems to be further complicated by the high prevalence of positive results. Currently, psychologists have little to gain from doing replications, but they do even less so if they cannot publish unsuccessful replications due to the lack of significant results. A single failed replication is often difficult to publish (Aldhous, 2011), and peer-reviewed journals are likely to accept failed replications only after a few attempts have been made. In this sense, the bias against negative findings is actually hindering replication studies because failed replications are often denied publication.

The problem with unpublished failed replications can be viewed in at least two ways. First, on a more pragmatic level, when an effect fails to be replicated, scientists from other labs may be largely unaware of this, especially if the study is a conceptual replication (see Pashler & Harris, 2012). For example, if there are a few unsuccessful attempts to replicate an effect, but nobody knows about them because they were not published, researchers from other labs may also go on to pursue effects that don't exist. This could potentially lead to wasted research funds, time and resources that could be used to pursue other avenues (Knight, 2003).

Second, unpublished replications may also have wider implications for the field. Currently, replications in psychology are very rare (about 1% of all publications; Makel et al., 2012), and it could be speculated that a large number of replications that are actually carried out are never published. Unsuccessful replications may be difficult to publish, but successful ones may ironically share the same fate, as some journals reject them on the grounds that they do not contribute anything new beyond what is already known (Spellman, 2012). This further exacerbates the problem because incorrect results from previous studies, however well-intentioned, remain unchallenged. If replications (successful or not) and negative studies are hard to publish, then this leads to a situation in which psychological theories are not rigorously evaluated and may even become unfalsifiable (Ferguson & Heene, 2012). For this reason, replications and negative studies have implications that are not restricted to the respective field of research, but that are also important for maintaining the scientific integrity of psychology.

Prevalence of Negative Results: Previous Studies

Considering the fact that these issues hold major implications for psychology as a science, it is important to know how prevalent negative studies are. At present, however, not much is known about the extent to which they exist in the literature. The main reason for this is that the studies that have investigated the problem are few and far in between (Bozarth & Roberts, 1972; Spence & Blanchard, 2001; Sterling, 1959; Sterling, Rosenbaum, & Weinkam, 1995).

Sterling (1959) was first to systematically analyze articles published in four psychology journals. He found that 97.2% of all articles using significance tests rejected the main stated null hypothesis. Three decades later, Sterling et al. (1995) reviewed eight psychological journals using the same criteria and concluded that publication practices have not changed over the past 30 years. These results were also confirmed by Bozarth and Roberts (1972), who found that 94% of articles published in three psychology journals rejected the null hypothesis. More recently, Spence and Blanchard (2001) did a cross-sectional analysis of five journals in sport and exercise psychology that spans over 10 years and found that 97.5% of all articles using significance tests rejected at least one null hypothesis and that 80.1% rejected the main stated null hypothesis.

The Present Study

Most of these studies were conducted decades ago and this would imply that the problem has since been forgotten. However, in recent years there has been a mounting interest in the lack of negative results and replicability of psychological research both among professionals and European psychology students (Flis, 2012, January 1; Pashler & Wagenmakers, 2012). This calls for a more thorough investigation of the problem, which will allow psychologists to determine the extent to which negative findings are published in the literature. Such an estimate is important because the information can be used to inform future meta-analytic studies and systematic reviews.

Traditionally, efforts to detect and correct for excessive positive findings have been done in meta-analytic studies within the framework of publication bias – defined as the tendency of scientific journals to preferentially publish papers that report significant results (Pautasso, 2010; Rosenthal, 1979). Nowadays, the majority of published meta-analytic studies in psychology make some efforts to analyze for publication bias (Ferguson & Brannick, 2012) and authors are routinely advised to take it into account when doing a meta-analysis (e.g., see Field & Gillett, 2010). There are several statistical methods that are normally used to control for publication bias: Failsafe N, Funnel plot, Trim and Fill, correlation and regression-based methods, and selection models (Kepes, Banks, McDaniel, & Whetzel, 2012). However, these methods are not without their limitations and the wisdom of their use lies with the researchers. For example, funnel plots, which are often used to detect publication bias, can be distorted by heterogeneity across studies or by the number of studies included in the plot (Daya, 2006; Lau, Ioannidis, Terrin, Schmid, & Olkin, 2006).

Because these methods are more suitable for controlling for publication bias when doing a meta-analysis, they were not used in the present study. The aim of this study was to estimate the proportion of positive and negative results in European psychology journals, as this is more closely related to the problems discussed in this paper. This was done by reviewing published articles and coding for whether the tested hypotheses are supported or not supported by evidence. While the current method estimates the prevalence of positive results, it does not in itself demonstrate evidence for publication bias, as no expected percentage of positive studies exists (Thornton & Lee, 2000). Therefore, the results will only be discussed in terms of hypotheses that are supported or not supported by evidence.

The present study utilized a design that is similar to Spence and Blanchard's (2001), Sterling's (1959), and Sterling et al.'s (1995) studies. However, in contrast to them, articles were not coded according to whether a particular null hypothesis was accepted or rejected, but according to whether all tested research hypotheses are supported or not supported by evidence. This approach provides wealthier data about published articles and avoids certain technical problems associated with coding articles in terms of accepted or rejected null hypotheses (e.g., some research hypotheses are tested with multiple significance tests).

Method

Data

All research articles from 2001, 2006 and 2011 were downloaded from the following five journals: European Journal of Psychology of Education, European Journal of Social Psychology, Scandinavian Journal of Psychology, British Journal of Clinical Psychology and British Journal of Developmental Psychology. The main inclusion criteria were that journals had to: 1) be intended as an outlet for European psychologists 2) publish at least 4 issues per year and 3) have a publication record that goes back to at least 2000.

From all eligible journals, the present five were chosen so that they represent geographically diverse areas and serve as research outlets of different fields in psychology. Furthermore, preference was given to journals that publish more research articles. The first two journals were selected because they meet the criteria of representing different fields and European psychologists at the same time. The Scandinavian Journal of Psychology was chosen because it publishes papers in all areas of psychology and is intended as an outlet for psychologists from Scandinavia. The last two journals were selected in order to acknowledge the fact that the majority of European English-speaking journals are based in the United Kingdom; another reason was the above-mentioned aim to include journals that publish papers in different psychological areas. The temporal choice of articles was motivated by the fact that five years should be enough to detect possible changes over time. A wider cross-sectional span could make results difficult to interpret and would overlap with previous studies that covered articles published in the previous century.

Coding of Articles

All downloaded articles ($N = 639$) were reviewed manually and the ones not relevant to the current study were excluded. First, all articles that did not test any hypotheses were removed. These included brief reports, pilot studies, validation studies, and reliability and factor analysis studies. Second, in order to reduce bias, only papers with explicitly stated hypotheses were included. Hypotheses were coded if they were clearly stated, appeared before the 'Method' section and contained one of the following keywords – “hypothesize”, “expect”, “predict”, “assume”, “anticipate”, “propose”, “postulate”. The articles were screened for these keywords with the search tool if authors hadn't differentiated the hypotheses from the introduction part. Moreover, hypotheses that were stated as a question (e.g., ‘Is X positively associated with Z?’), contained “whether” (e.g., ‘We wanted to see whether X predicts Z’), or were otherwise not clearly formulated were also excluded. Of all the articles that had explicitly stated hypotheses ($N = 393$), 14 were excluded because it wasn't possible to determine whether one or more hypotheses were supported or not supported by evidence without prior knowledge in the area. Furthermore, 24 articles were excluded because all tested hypotheses were partially supported. Finally, two articles were excluded because they did not use Null Hypothesis Significance Testing (NHST) for testing hypotheses. This left a total of 353 articles that were used in the data analysis. Table 1 shows the number of articles per year for all five journals.

Table 1

Coded Articles per Year for All Five Journals

Journal	2001	2006	2011
EJPE	13	6	14
EJSP	34	39	63
SJP	7	22	39
BJCP	11	12	13
BJDP	19	25	36
Total	84	104	165

All papers were coded according to how many hypotheses were 1) tested, 2) supported, 3) not supported, 4) partially supported and 5) whether at least one hypothesis was supported by evidence. Articles that reported partially supported hypotheses ($N = 97$) were coded as usual, but these hypotheses were not included in the analysis. While hypotheses in the sample were tested using NHST, not every hypothesis was tested with just one significance test. Therefore, the number of tested research hypotheses does not necessarily correspond to the number of significance tests that were carried out by the authors.

To calculate coding reliability, 30 randomly chosen articles were coded by an independent rater. The overall inter-rater agreement was 75.3% for all five coded variables. Cohen's kappa for the "at least one supported hypothesis" variable was $k = .59$.

Methodological Issues

Investigating positive and negative results in psychology is intimately related to the NHST procedure, a method for statistical inference that is widely used in psychology and other social sciences. Despite its enormous popularity, NHST has generated a considerable amount of controversy (Nickerson, 2000) and has been criticized for its incorrect use by researchers (e.g., see Cohen, 1994; Gigerenzer, 2004).

All articles that were analyzed in this study used NHST for statistical inference. Although alternative methods for testing hypotheses, such as the Bayesian factor, are becoming more popular in psychology (Andrews & Baguley, 2013), NHST continues to dominate the field overwhelmingly (Cumming et al., 2007). This is further supported by the fact that out of all eligible articles, only two did not use NHST for testing hypotheses.

Due to the controversial nature of NHST, alternative methods for data analysis, such as effect size estimation and interval estimation, have been suggested (American Psychological Association, 2009; Kline, 2004). However, they were not taken into account as separate methods of inference as a lot of articles were published when these practices were not so common. Moreover, the implementation of these alternative methods has been happening slowly and even authors that report confidence intervals rarely use them as a method of inference (Cumming et al., 2007; Fidler, Thomason, Cumming, Finch, & Leeman, 2004).

Results

The outcome of data analysis is displayed in Table 2. The data show that 95.4% of all coded articles found support for at least one tested hypothesis. Interestingly, the percentage of articles supporting at least one hypothesis was 95.3% in 2001 and then it dropped to 92.3% in 2006. However, this tendency did not persist into 2011, when the

percentage of articles supporting at least one hypothesis was 97.5%. Articles that tested only one hypothesis ($N = 68$) yielded a similar result – 92.6% of them found support for it.

Table 2

Number of Hypotheses Supported and not Supported by Evidence

Year of Publication	Number of tested hypotheses	At least one hypothesis supported by evidence (%)	All hypotheses supported by evidence (%)	Average number of hypotheses supported by evidence (SD)	Average number of hypotheses not supported by evidence (SD)
2001	211	95.2	80.9	2.05 (1.34)	0.22 (0.49)
2006	289	92.3	66.3	2.02 (1.47)	0.44 (0.69)
2011	509	97.5	73.3	2.41 (1.68)	0.40 (0.77)
Total	1009	95.4	73.0	2.21 (1.55)	0.37 (0.69)

The results also demonstrate that 73% of all articles found support for all tested hypotheses. Similarly, there was a slight decrease in this number in 2006, but then it went up again in 2011.

Overall, the average paper supported 2.21 hypotheses and failed to support 0.37 hypotheses. This means that the average hypothesis was 5.9 times more likely to be supported by evidence than not to. The average hypothesis from articles in 2001, 2006 and 2011 was 9.3, 4.5 and 6 times more likely to be supported than not to be supported, respectively.

Temporal Trends

A chi-square test showed no significant relationship between the year in which papers were published and the number of papers that supported all tested hypotheses, $\chi^2(2) = 5.050$, $p = .080$, $\phi = .120$. The effect of the year of publication on whether at least one hypothesis was supported was not possible to calculate because the expected counts of 2 cells were less than 5, thus violating the assumptions of the chi-square test statistic (the minimum expected count was 3.81).

Kruskal–Wallis tests revealed that the year in which the papers were published does not have a significant effect on the number of hypotheses that were supported ($H(2) = 5.638$, $p = .060$) or not supported ($H(2) = 5.174$, $p = .075$) by evidence. In other words, no statistically significant change in the number of hypotheses supported or not supported by evidence was observed.

Even though no significant differences emerged in the time period that was studied, it is interesting to note that from 2001 to 2011 there was a 2.4% increase in the number of studies that found support for at least one hypothesis. In contrast, the percentage of papers supporting all hypotheses decreased by 9.3%. Mixed results were also observed in the average number of tested hypotheses. The average number of supported hypotheses increased by 17.5%, but the average number of hypotheses not supported by evidence also increased by 81.8%. In other words, in 2011 there were 81.8% more hypotheses that failed to receive support than there were in 2001.

Discussion

The present study set out to investigate the extent to which negative results exist in five European psychology journals. To achieve this, it measured four variables that can be used to infer this: the number of articles that

support at least one hypothesis, the number of articles that support all tested hypotheses, and the average number of hypotheses that are supported and not supported by evidence. Taken together, these findings demonstrate that negative results are rare and that positive results still largely dominate all five journals. Because articles were not coded in terms of separate null hypotheses, it is difficult to make straightforward comparisons with previous studies. Nevertheless, the present study obtained results that are very similar to those reported by [Spence and Blanchard \(2001\)](#) in that 95.4% of all articles found support for at least one tested hypothesis.

The strength of the current study is that it provides more information that can be used to estimate the prevalence of negative results in European psychology journals. It shows that even articles that tested one hypothesis found support for it in almost all cases. Also, the majority (73%) of published papers supported all tested hypotheses. Moreover, the average hypothesis was almost six times more likely to be supported than not to. This suggests that studies testing multiple hypotheses may be less likely to be published if the majority of research hypotheses are not supported by evidence. Even though no expected percentage of negative results exists, the fact that only 4.6% of all articles failed to find support for any of the tested hypotheses suggests that studies reporting entirely negative results may be difficult to publish.

While the present data cannot give an explanation as to why this happens, it could be speculated that some authors who obtain negative results may conduct other studies, perhaps with significant results, in order to increase the publication “value” of their paper. Another possibility is that authors may report only what “worked” and leave out negative findings that may make the paper more difficult to publish. The latter argument is in line with the results from [John et al.'s \(2012\)](#) survey, which found that a fair number of psychologists admitted that they have not always reported all conditions or dependent variables.

The statistical analyses that were carried out revealed that the year of publication did not have a significant effect on the variables of interest. In other words, the number of positive findings has not changed significantly over the time period that was studied. This conclusion is further supported by the average number of supported hypotheses. While the number of supported hypotheses increased from 2001 to 2011, the increase in the number of hypotheses that were not supported by evidence was more than four times as much. This suggests that although the number of articles reporting positive results remained high, more articles that reported hypotheses not supported by evidence were published in 2011 compared to 2001.

These results seem to contradict a recent survey by [Fanelli \(2012\)](#), who reported that negative results are disappearing from most sciences. According to [Fanelli \(2012\)](#), articles that report full or partial support for hypotheses have increased by more than 20% across multiple disciplines from 1990 to 2007 and the trend is “significantly stronger in the social sciences (p. 895).” This difference may be due to the fact that the present study was concentrated on cross-sectional data that covers two 5-year periods. Also, [Fanelli \(2012\)](#) coded hypotheses that received both full and partial support, whereas the present study did not take into account partially supported hypotheses.

The results of the present study, however, are in line with [Spence and Blanchard's \(2001\)](#) study, which also did not find any clear temporal change in the percentage of positive results. Moreover, [Pautasso \(2010\)](#) reviewed abstracts of published papers but also failed to find evidence that positive results in psychology are increasing. This could suggest that positive results may fluctuate slightly from year to year, but that they do not change significantly. In this sense, although the data from the present study shows that positive findings are high in all five

journals, there is no evidence to suggest that they have increased (or decreased) significantly over the last 10 years.

Limitations of the Study

The present study, of course, is not without its limitations and the results should be considered in the light of its shortcomings. First, the collected data is not representative for all European psychology journals. Therefore, no attempts are made to generalize the findings to all European psychology journals. Second, inter-rater agreement was only marginally substantial (Landis & Koch, 1977). However, this calculation was based on only 8% of all articles and inter-rater agreement could be different if all papers were independently coded. Furthermore, all variables are interdependent and coding one variable differently results in differences in most others variables, thus making nearly perfect agreement impossible to achieve.

The results of the study may also be potentially influenced by the articles published in the chosen journals or the editorial policies at that time. For example, some articles may be unique in their research hypotheses or studied populations. Also, journal turnover rate of editors and reviewers could be an influence because it is not known whether or how many of them had stayed with the journals for the whole time period. Finally, the empirical evidence presented here cannot be used to give an explanation why this bias against negative results exists.

The Importance of Negative Results in Psychological Research

The present article has tried to argue that the low number of negative results indirectly encourages questionable research practices and can have a detrimental effect on systematic reviews, meta-analyses, and replications in psychology. Even though publishing more negative results will not eradicate all the problems associated with these areas, it can nonetheless do a lot to ameliorate them.

It is difficult to draw any causal relationships between negative results, questionable research practices and doing replications, but it is logical to assume that an increase in the proportion of negative findings in the literature will also reduce the incentives that researchers have to engage in questionable research practices. This is likely to happen because the difficulty of publishing papers with negative results is what tempts researchers to resort to them in the first place. Moreover, such an increase in the published studies that report negative results will also help design better systematic reviews and meta-analyses that will account for data that is not otherwise accessible. Finally, publishing more studies with negative results will also stimulate researchers to conduct more replications, not fearing that failed attempts will not get published.

The long-term benefits of publishing more negative findings are hard to ignore, but the shift towards this change is more difficult. Some attempts have already been made to stimulate the replicability of psychological research and to make negative results more easily accessible. For example, the PsychDrawer depository (<http://www.psych-filedrawer.org/>) and the Journal of Articles in Support of the Null Hypothesis both aim to make replications and studies with negative results publicly available. The viability of these research outlets, however, is questionable in the long term as few researchers to date have uploaded their unpublished manuscripts to PsychDrawer and a journal publishing entirely negative findings is bound to face different obstacles. Nevertheless, they carry the important message that the lack of negative results in the literature is a problem that hinders scientific progress and that it should be addressed in a more substantial manner by psychologists.

Another intuitive but more difficult to implement solution would be to change editorial practices of peer-reviewed journals that lead to this problem. If journal editors put more emphasis on the contribution of a given study instead

of whether most (or all) significance tests yielded positive results, the problems associated with excessive positive findings are likely to be reduced. More negative findings in the research literature will not only benefit systematic reviews and meta-analyses, but they will also show if previous findings hold up; in the case that they don't, this will help identify areas that require more attention from researchers.

In fairness to journal editors, it should be noted that authors may also contribute to the problem. For example, they may be more reluctant to submit a manuscript for publication if it reports negative results (Coursol & Wagner, 1986; Møller & Jennions, 2001). One reason why this could happen is that the increasingly competitive world of academia puts pressure on researchers to come up with positive findings (Fanelli, 2010). For this reason, researchers who obtain negative results may not generally consider it worthwhile to pursue their publication, perhaps fearing that they will waste precious time and resources.

It could be further argued that some researchers who do submit their negative studies for publication may not make much effort to explain why their negative results are important and how they fit in the scientific body of knowledge. Ultimately, it is the author's job to explain in a comprehensive way why their findings are important and to place them in the context of past and future research; likewise, it is the editor's and reviewers' job to evaluate the contribution of each article to existing knowledge. If authors reporting negative results are not motivated to make a strong case for their paper, then perhaps reviewers and editors may not be entirely responsible for preferring positive studies that may explain their findings better.

In this sense, it is an oversimplification to say that journal editors are entirely responsible for the high prevalence of positive results in the literature. However, they also hold the key to changing the current situation. Just like the authors who submit their papers, journal editors are also researchers themselves, and as such, they should live up to the challenges that psychology faces today (e.g., see Nosek et al., 2012; Pashler & Harris, 2012). By developing journal policies that are more accepting towards negative results, editors will also motivate researchers to submit more negative studies. This will stimulate an environment in which editors and reviewers are less likely to consider negative findings as inferior to positive ones, and authors are less likely to not submit them for publication.

While this change would probably happen slowly, it has great potential to benefit empirical psychology. The lack of negative results is part of a wider context and it is not by any means the ultimate solution to all other problems. However, publishing more papers with negative results will stimulate the replicability of research findings and the trust that we have in their validity.

Conclusion

Bias against negative findings exists in many scientific disciplines and psychology is no exception (Fanelli, 2012). However, as this and previous studies have demonstrated, the percentage of papers reporting negative results in psychological research is alarmingly low. This leads to a situation in which studies are virtually excluded from the scientific body of knowledge if the results failed to reach the "holy grail of $p < .05$ " (Glaser, 2010, p.330). As a consequence, there is tremendous pressure on researchers to obtain and report positive results, which could potentially lead to dishonest research practices aimed at producing publishable results. The difficulty in publishing negative results is also part of a wider framework, because it contributes to the problem of replicability and validity of research findings in psychology.

While there is no easy solution to the problem, it is important to remember that scientific journals are the gatekeepers of research output, and as such, they should strive to publish research findings that represent reality more closely.

Reporting findings that support the tested hypotheses is indispensable to the advancement of psychology as a science because it tells us what works. However, we must not forget that often it is just as important to know what *does not* work (Nippold, 2012). Putting away studies in file drawers only because they failed to reach the conventional significance level leads to a unilateral view of reality. When we cannot see the bigger picture, it is difficult to estimate the validity of research findings, to inform practice and to translate research findings into psychological interventions.

Acknowledgements

I would like to thank Dr. Nikolay Ratchev for helping me with the independent coding of articles and for his insightful comments during informal discussions of this research project.

References

- Aldhous, P. (2011, May 5). Journal rejects studies contradicting precognition. *New Scientist*. Retrieved from <http://www.newscientist.com/article/dn20447-journal-rejects-studies-contradicting-precognition.html>
- American Psychological Association. (2009). *Publication manual of the American Psychological Association* (6th ed.). Washington, DC: Author.
- Andrews, M., & Baguley, T. (2013). Prior approval: The growth of Bayesian methods in psychology [Editorial]. *The British Journal of Mathematical and Statistical Psychology*, 66(1), 1-7. doi:10.1111/bmsp.12004
- Banks, G. C., Kepes, S., & Banks, K. P. (2012). Publication bias: The antagonist of meta-analytic reviews and effective policymaking. *Educational Evaluation and Policy Analysis*, 34(3), 259-277. doi:10.3102/0162373712446144
- Bozarth, J. D., & Roberts, R. R. (1972). Signifying significant significance. *The American Psychologist*, 27(8), 774-775. doi:10.1037/h0038034
- Bradley, M. T., & Gupta, R. D. (1997). Estimating the effect of the file drawer problem in meta-analysis. *Perceptual and Motor Skills*, 85, 719-722. doi:10.2466/pms.1997.85.2.719
- Cohen, J. (1994). The earth is round ($p < .05$). *The American Psychologist*, 49(12), 997-1003. doi:10.1037/0003-066X.49.12.997
- Coursol, A., & Wagner, E. E. (1986). Effect of positive findings on submission and acceptance rates: A note on meta-analysis bias. *Professional Psychology, Research and Practice*, 17(2), 136-137. doi:10.1037/0735-7028.17.2.136
- Cumming, G., Fidler, F., Leonard, M., Kalinowski, P., Christiansen, A., Kleinig, A., . . . Wilson, S. (2007). Statistical reform in psychology: Is anything changing? *Psychological Science*, 18(3), 230-232. doi:10.1111/j.1467-9280.2007.01881.x
- Dalton, D. R., Aguinis, H., Dalton, C. M., Bosco, F. A., & Pierce, C. A. (2012). Revisiting the file drawer problem in meta-analysis: An assessment of published and nonpublished correlation matrices. *Personnel Psychology*, 65, 221-249. doi:10.1111/j.1744-6570.2012.01243.x
- Daya, S. (2006). Funnel plots and publication bias: Work in progress? [Editorial]. *Evidence-Based Obstetrics & Gynecology*, 8, 71-72. doi:10.1016/j.ebogyn.2006.10.001

- Dwan, K., Altman, D. G., Arnaiz, J. A., Bloom, J., Chan, A.-W., Cronin, E., . . . Williamson, P. R. (2008). Systematic review of the empirical evidence of study publication bias and outcome reporting bias. *PLoS ONE*, 3(8), e3081. doi:10.1371/journal.pone.0003081
- Egger, M., & Smith, G. D. (1998). Meta-analysis bias in location and selection of studies. *British Medical Journal*, 316, 61-66. doi:10.1136/bmj.316.7124.61
- Fanelli, D. (2010). Do pressures to publish increase scientists' bias? An empirical support from US states data. *PLoS ONE*, 5(4), e10271. doi:10.1371/journal.pone.0010271
- Fanelli, D. (2012). Negative results are disappearing from most disciplines and countries. *Scientometrics*, 90, 891-904. doi:10.1007/s11192-011-0494-7
- Ferguson, C. J., & Brannick, M. T. (2012). Publication bias in psychological science: Prevalence, methods for identifying and controlling, and implications for the use of meta-analyses. *Psychological Methods*, 17(1), 120-128. doi:10.1037/a0024445
- Ferguson, C. J., & Heene, M. (2012). A vast graveyard of undead theories: Publication bias and psychological science's aversion to the null. *Perspectives on Psychological Science*, 7(6), 555-561. doi:10.1177/1745691612459059
- Fidler, F., Thomason, N., Cumming, G., Finch, S., & Leeman, J. (2004). Editors can lead researchers to confidence intervals, but can't make them think: Statistical reform lessons from medicine. *Psychological Science*, 15(2), 119-126. doi:10.1111/j.0963-7214.2004.01502008.x
- Field, A. P., & Gillett, R. (2010). How to do a meta-analysis. *The British Journal of Mathematical and Statistical Psychology*, 63, 665-694. doi:10.1348/000711010X502733
- Fliis, I. (2012, January 1). What happens to studies that accept the null hypothesis? *JEPS Bulletin*. Retrieved from <http://jeps.efpsa.org/blog/2012/01/01/what-happens-to-studies-that-accept-the-null-hypothesis/>
- Gigerenzer, G. (2004). Mindless statistics. *Journal of Socio-Economics*, 33, 587-606. doi:10.1016/j.socec.2004.09.033
- Glaser, D. (2010). When interpretation goes awry: The impact of interim testing. In D. L. Streiner & S. Sidani (Eds.), *When research goes off the rails: Why it happens and what you can do about it* (pp. 327-333). New York, NY: The Guilford Press.
- Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLoS Medicine*, 2(8), e124. doi:10.1371/journal.pmed.0020124
- 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. doi:10.1177/0956797611430953
- Kepes, S., Banks, G. C., McDaniel, M., & Whetzel, D. L. (2012). Publication bias in the organizational sciences. *Organizational Research Methods*, 15(4), 624-662. doi:10.1177/1094428112452760
- Kline, R. B. (2004). *Beyond significance testing: Reforming data analysis methods in behavioral research*. Washington, DC: American Psychological Association.
- Knight, J. (2003). Negative results: Null and void. *Nature*, 422, 554-555. doi:10.1038/422554a
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33, 159-174. doi:10.2307/2529310

- Lau, J., Ioannidis, J. P. A., Terrin, N., Schmid, C. H., & Olkin, I. (2006). Evidence based medicine: The case of the misleading funnel plot. *British Medical Journal*, 333, 597-600. doi:10.1136/bmj.333.7568.597
- Makel, M. C., Plucker, J. A., & Hegarty, B. (2012). Replications in psychology research: How often do they really occur? *Perspectives on Psychological Science*, 7(6), 537-542. doi:10.1177/1745691612460688
- McNemar, Q. (1960). At random: Sense and nonsense. *The American Psychologist*, 15(5), 295-300. doi:10.1037/h0049193
- Møller, A. P., & Jennions, M. D. (2001). Testing and adjusting for publication bias. *Trends in Ecology & Evolution*, 16(10), 580-586. doi:10.1016/S0169-5347(01)02235-2
- Neuroskeptic. (2012). The nine circles of scientific hell. *Perspectives on Psychological Science*, 7, 643-644. doi:10.1177/1745691612459519
- Nickerson, R. S. (2000). Null hypothesis significance testing: A review of an old and continuing controversy. *Psychological Methods*, 5(2), 241-301. doi:10.1037/1082-989X.5.2.241
- Nippold, M. A. (2012). The power of negative findings [Editorial]. *Language, Speech, and Hearing Services in Schools*, 43, 251-252. doi:10.1044/0161-1461(2012)ed-03
- Nosek, B. A., Spies, J. R., & Motyl, M. (2012). Scientific utopia: II. Restructuring incentives and practices to promote truth over publishability. *Perspectives on Psychological Science*, 7, 615-631. doi:10.1177/1745691612459058
- Pashler, H., & Harris, C. R. (2012). Is the replicability crisis overblown? Three arguments examined. *Perspectives on Psychological Science*, 7, 531-536. doi:10.1177/1745691612463401
- Pashler, H., & Wagenmakers, E.-J. (2012). Editors' introduction to the special section on replicability in psychological science: A crisis of confidence? *Perspectives on Psychological Science*, 7, 528-530. doi:10.1177/1745691612465253
- Pautasso, M. (2010). Worsening file-drawer problem in the abstracts of natural, medical and social science databases. *Scientometrics*, 85, 193-202. doi:10.1007/s11192-010-0233-5
- Rosenthal, R. (1979). The "file drawer problem" and tolerance for null results. *Psychological Bulletin*, 86(3), 638-641. doi:10.1037/0033-2909.86.3.638
- Rothstein, H. R., Sutton, A. J., & Borenstein, M. (2005a). Publication bias in meta-analysis. In H. R. Rothstein, A. J. Sutton, & M. Borenstein (Eds.), *Publication bias in meta-analysis: Prevention, assessment and adjustments* (pp. 1-8). West Sussex, England: John Wiley & Sons.
- Rothstein, H. R., Sutton, A. J., & Borenstein, M. (2005b). *Publication bias in meta-analysis: Prevention, assessment and adjustments*. West Sussex, England: John Wiley & Sons.
- 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. doi:10.1177/0956797611417632
- Smith, M. B. (1956). Editorial. *Journal of Abnormal and Social Psychology*, 52(1), 1-4. doi:10.1037/h0039152
- Spellman, B. A. (2012). Introduction to the special section: Data, data, everywhere . . . especially in my file drawer. *Perspectives on Psychological Science*, 7(1), 58-59. doi:10.1177/1745691611432124

- Spence, J. C., & Blanchard, C. (2001). Publication bias in sport and exercise psychology: The games we play. *International Journal of Sport Psychology*, 32, 386-399.
- Sterling, T. D. (1959). Publication decisions and their possible effects on inferences drawn from tests of significance - or vice versa. *Journal of the American Statistical Association*, 54, 30-34. doi:[10.2307/2282137](https://doi.org/10.2307/2282137)
- Sterling, T. D., Rosenbaum, W. L., & Weinkam, J. J. (1995). Publication decisions revisited: The effect of the outcome of statistical tests on the decision to publish and vice versa. *The American Statistician*, 49(1), 108-112.
- Thornton, A., & Lee, P. (2000). Publication bias in meta-analysis: Its causes and consequences. *Journal of Clinical Epidemiology*, 53, 207-216. doi:[10.1016/S0895-4356\(99\)00161-4](https://doi.org/10.1016/S0895-4356(99)00161-4)
- Torgerson, C. J. (2006). Bias: The Achilles' heel of systematic reviews? *British Journal of Educational Studies*, 54(1), 89-102. doi:[10.1111/j.1467-8527.2006.00332.x](https://doi.org/10.1111/j.1467-8527.2006.00332.x)
- Tulloch, G. (1959). Publication decisions and tests of significance: A comment. *Journal of the American Statistical Association*, 54(287), 593. doi:[10.2307/2282539](https://doi.org/10.2307/2282539)

About the Author

Martin Rachev Vasilev is a recent psychology graduate (B.A.) from Sofia University "St. Kliment Ohridski" who has a wider interest in methodology and scientific publishing.