

# Journal of Experimental Psychology: Applied

## Do Publications in Low-Impact Journals Help or Hurt a CV?

Kristin Donnelly, Craig R. M. McKenzie, and Johannes Müller-Trede

Online First Publication, May 20, 2019. <http://dx.doi.org/10.1037/xap0000228>

### CITATION

Donnelly, K., McKenzie, C. R. M., & Müller-Trede, J. (2019, May 20). Do Publications in Low-Impact Journals Help or Hurt a CV?. *Journal of Experimental Psychology: Applied*. Advance online publication. <http://dx.doi.org/10.1037/xap0000228>

## Do Publications in Low-Impact Journals Help or Hurt a CV?

Kristin Donnelly, Craig R. M. McKenzie, and Johannes Müller-Trede  
University of California, San Diego

Using psychology professors as participants, the present study investigates how publications in low-impact psychology journals affect evaluations of a hypothetical tenure-track psychology job applicant. Are “weak” publications treated as evidence for or against a candidate’s ability? Two experiments revealed that an applicant was rated as stronger when several weak publications were added to several strong ones and was rated as weaker when the weak publications were removed. A third experiment showed that the additional weak publications were not merely viewed as a signal of additional strong publications in the future; instead, the weak publications themselves appear to be valued. In a fourth and final experiment, we found that adding a greater number of weak publications also strengthened the applicant, but not more so than adding just a few. The study further suggests that the weak publications may signal ability, as applicants with added weak publications were rated as both more hardworking and more likely to generate innovative research ideas. Advice for tenure-track psychology applicants: Do not hesitate to publish in even the weakest journals, as long as it does not keep you from publishing in strong journals. Implications of the market rewarding publications in low-impact journals are discussed.

### **Public Significance Statement**

Publications are the primary means for academics to establish reputation, communicate value, and advance their careers. This study investigates whether and how publications in low-impact journals affect professors’ evaluations of hypothetical tenure-track job candidates in psychology.

*Keywords:* belief updating, averaging, hiring

Imagine evaluating the curriculum vitae (CV) of an advanced psychology graduate student or a recent Ph.D. seeking a tenure-track or postdoctoral position. The applicant has a few publications, which have all appeared in very strong journals. Now imagine the same CV, but with a few additional publications that appeared in psychology journals that you know have very low impact. How would your opinion of the applicant change? Would you have a higher or lower opinion of the applicant? Or would the weaker publications make no difference at all?

This question has obvious importance for job applicants in psychology. Is it worthwhile to publish as much as possible, even in low-impact journals? The answer is obviously “no” if such publications only hurt a CV. They would even be a mistake to pursue if they have no effect on a CV because every publication is time consuming and time would be better spent on more promising

projects. But if “weak” publications help a CV, then writing up and publishing incremental, or perhaps flawed, studies is worthwhile, provided it does not take too much time away from publishing more impactful work. The question also has implications for psychology as a field, because the answer will reveal whether the market is encouraging or discouraging weak publications.

Despite the importance of the issue, there was no empirical work directly addressing it when we first started this research, and an informal polling of colleagues’ (and our own) intuitions resulted in mixed predictions. In terms of theory, we looked to the literature on belief updating because we are interested in how judgments of applicant quality change when new evidence—in the form of weak publications—is introduced. Consider two simple models in which publication strength is a cue to the applicant’s ability. An “adding” model of belief updating makes different predictions depending on whether publishing in a low-impact journal is seen as evidence against the applicant’s ability, or as weak evidence *for* it (Carlson & Dulany, 1988; Edwards, 1968; Wallsten & Manley Sapp, 1977; see also Hogarth & Einhorn, 1992; McKenzie, Lee, & Chen, 2002). In the former case, weak publications will hurt; in the latter case, weak publications will help. By contrast, an “averaging” model makes the unequivocal prediction that adding relatively weak publications will always hurt a candidate because they will lower the average publication strength. Considerable evidence indicates that people often average when updating their opinions, regardless of whether the new information is quantitative or qualitative (Anderson, 1981; Lopes, 1985, 1987; Meyvis & Janiszewski,

---

Kristin Donnelly, Craig R. M. McKenzie, and Johannes Müller-Trede, Department of Psychology, University of California, San Diego.

Craig R. M. McKenzie is now at the Rady School of Management and Department of Psychology, University of California, San Diego. Johannes Müller-Trede is now at IESE Business School, University of Navarra.

Correspondence concerning this article should be addressed to Kristin Donnelly, who is now at Haas School of Business, University of California, Berkeley, 545 Student Services Building #1900, Berkeley, CA 94720. E-mail: kristin\_donnelly@berkeley.edu

wski, 2002; Nisbett, Zukier, & Lemley, 1981; Shanteau, 1970, 1972, 1975). For example, people have been found to average the information contained in a product's distinct attributes (Troutman & Shanteau, 1976) or in an individual's personality traits (Anderson & Alexander, 1971).

A recent, unpublished study addressed this question using economists as participants, and the results showed that adding weak articles to a list of economics publications led to a more negative evaluation (Powdthavee, Riyanto, & Knetsch, 2017). These results are consistent with both an averaging model and an adding model in which weak publications are seen as evidence against a candidate's ability. For at least two reasons, however, these results might differ from the results we report in this article. First, Powdthavee et al. used economists as participants. There might be important between-discipline differences in terms of how publications in low-impact journals are perceived. Relative to psychologists, economists may value quality of publications more and quantity of publications less (cf. Conroy, Dusansky, Drukker, & Kildegaard, 1995; McPherson, 2012). Second, we explicitly focus on situations that are representative of a recent Ph.D. graduate with a degree in psychology competing for a tenure-track position. Powdthavee et al.'s design may be less representative of such situations than ours because of several methodological choices that we examine in more detail in the General Discussion. Our question is simple: If you have a solid CV, is it improved or diminished by adding weak publications? The data we report below indicate that it is improved.

In a series of four experiments, we recruited professors to evaluate a hypothetical tenure-track applicant based on the applicant's list of publications in psychology journals. Most of the participants were professors in psychology departments, though there were roughly equally many business school professors in our first experiment. Participating professors saw a list with either three strong publications, three weak publications, and/or three strong interspersed with additional weak publications. This allowed us to assess how the weak and strong publications are viewed independently, as well as how evaluations change when some, or many, weak publications are added to the strong ones.

## Experiment 1

### Method

**Participants.** Ninety-two academics (23 females) were recruited from the listserv of the Society for Judgment and Decision Making.<sup>1</sup> This sample represents the complete collection of responses to our recruitment emails. Partly due to a lack of prior literature, we did not have clear expectations concerning effect sizes and did not conduct formal power analyses. Instead, we drew on our experience to decide on a minimum sample size before beginning data collection. Once we reached the minimum sample size, our stopping rule was to keep gathering data until responses seemed to have come to a stop. This process generally involved an initial recruitment e-mail as well as one or two reminders.

Our recruitment e-mail explicitly stated that the study was restricted to professors. We targeted professors both because they are typically the ones who evaluate and make decisions regarding tenure-track job applicants, and because they presumably know which psychology journals are low impact and which are high

impact. A final sample of 87 remained after removing respondents who started but did not complete the survey. Of these respondents, 37% reported that they were in a psychology department, 44% were in a business school, and 19% were in another department. Seven respondents (8%) identified as "lecturers" or "post-docs," and three respondents (3%) failed an attention check. In this and all subsequent studies, we report analyses that include all respondents, but all results and their statistical significance are robust to the exclusion of respondents who self-identified as anything other than "professors" in the postexperimental survey, and to the exclusion of respondents who failed an attention check (modified from Oppenheimer, Meyvis, & Davidenko, 2009).

Each study reported here was approved by the IRB committee at the authors' institution. Experiment 1 and all other studies consisted of web forms programmed in Qualtrics (Qualtrics, Provo, UT). Neither Experiment 1 nor any of the other studies included any measures or conditions that are not reported. The materials for all studies can be found on the Open Science Framework page for this project: <https://osf.io/jfazm/>.

**Materials.** Each participant was presented with two hypothetical publication lists, one with three publications (manipulated between subjects: all strong vs. all weak) and one with six publications (the same three weak and the same three strong publications). The lists resembled excerpts from a CV in which publication title and coauthors had been redacted, leaving only the applicant's name, authorship position, year of publication, and journal outlet (see Figure 1). All six publications were either in press or had been published in the previous two years, and the strong and weak publications were equally distributed across that time period. In addition, the applicant was either first or second author on all the publications, with first authorships on two strong and two weak publications. Finally, two strong and two weak publications appeared in general psychology journals, and one strong and one weak publication appeared in specialty journals (in social psychology).

To measure the impact of a journal, we used the "Article Influence" (AI) score from eigenfactor.org, which is comparable to a journal's impact factor. A journal's AI score "is a measure of the average influence of each of its articles over the first five years after publication" (<http://eigenfactor.org/about.php>). At the time of running this experiment (mid 2016), the range of percentile scores for the high-impact journals we used (*Psychological Science*; *Journal of Experimental Psychology: General*; and *Journal of Personality and Social Psychology*) was 97 to 98 (impact factors ranged between 4.7 and 5.9). The range for the low-impact journals (*Psychological Reports*; *New Ideas in Psychology*; and *Social Behavior & Personality*) was 15 to 49 (impact factors ranged between 0.4 and 1.8).

**Procedure and design.** Participants began by reading the following:

<sup>1</sup> From the society's website ([www.sjdm.org](http://www.sjdm.org)): "The Society for Judgment and Decision Making is an interdisciplinary academic organization dedicated to the study of normative, descriptive, and prescriptive theories of judgments and decisions. Its members include psychologists, economists, organizational researchers, decision analysts, and other decision researchers."

Imagine that you are evaluating a candidate for a tenure-track Assistant Professor position in your department. The candidate is finishing up his doctoral work at a major research university with a good reputation. We've eliminated some information (e.g., publication title, coauthors). Please look over the list of the candidate's publications below.

Participants then saw either a list with the three strong publications, the three weak publications, or all six publications, and were asked, "Based on this list of publications, how strong do you find this candidate?" They responded on a sliding scale from 0 to 100, with 0 = "extremely weak" and 100 = "extremely strong."

On the next screen, participants read, "Imagine the candidate's list of publications looked like this instead:" and were presented with a modified list. Those who first rated either three strong or three weak publications now saw all six publications. Those who first rated all six publications now saw either only the three strong publications or only the three weak publications. They again rated the applicant's strength on the sliding scale.

Thus, each participant rated the strength of the applicant based on two publication lists, one with three, and one with 6 publications. We manipulated between subjects whether the list with three publications was "strong only" or "weak only" and whether the list with three publications or the list with six publications was rated first.

## Results and Discussion

Mean strength ratings of the applicant, as a function of the composition of the short list (weak vs. strong), list length, and list order, are shown in Figure 2. The bottom panel shows that the weak short list (three weak publications) was rated as quite weak and, not surprisingly, that the additional three strong publications had a large positive effect. The top panel, on the other hand, reveals that adding the three weak publications to the three strong ones did not have an effect on rated applicant strength (left pair of columns;  $t(19) = .33, p = .75, d = .05$ ), and that removing the three weak publications from the list of three strong and three weak publications hurt the applicant (right columns;  $t(23) = 3.46, p = .002, d = .52$ ). A  $t$  test collapsing across order revealed higher strength ratings for the candidate when the CV listed both strong and weak publications rather than just strong ones ( $M_s = 79.5, 75.3; SD_s = 14.4, 17.2; t(43) = 2.55, p = .01, d = .27$ ).<sup>2</sup> This pattern held up at the individual level as well: 66% of participants rated the candidate as stronger when the weak publications were included, 20% rated him as weaker, and 14% as equally strong ( $\chi^2(1, N = 35) = 4.45, p = .03, V = .32$  for stronger vs. weaker and equally strong combined). Thus, there is no evidence that the weak publications hurt the applicant, and some evidence that they helped, which indicates that the participants were not "averaging" publication strength when evaluating the job candidate.

## Experiment 2

The results of Experiment 1 indicate that adding weak publications to a CV has either no effect or a small positive one. The sample was relatively small (24 or fewer participants per cell), however, and whether participants in the strong Short List conditions first evaluated the list with or without the added weak publications seemed to affect the impact of those publications. In Experiment 2, we increased the sample of participants and ran only

the strong Short List conditions in order to focus on the effect of adding weak publications to a CV.

## Method

**Participants.** One hundred and 73 academics (90 females) were recruited from the Society for Personality and Social Psychology listserv.<sup>3</sup> For this and all subsequent studies, we enabled a filter in the experimental software Qualtrics to prevent respondents from participating in more than one of our studies. A final sample of 161 remained after removing incomplete responses. Eighty-five percent reported that they were in a psychology department, 10% in a business school, and 5% in another department. Ten respondents did not identify as "professor" (6%), and seven (4%) failed the attention check.

**Materials and procedure.** Just as in the strong Short List conditions in Experiment 1, participants rated the strength of the applicant based on the list of three strong publications and the list of six (three strong + three weak) publications. List order was again counterbalanced.

## Results and Discussion

The results are illustrated in Figure 3. Regardless of whether the list with only three strong publications or the list with all six publications was presented first, the longer list led the applicant to be rated as stronger; that is, the weak publications helped. A 2 (List Order: short first, long first)  $\times$  2 (List Length: short, long) mixed-model ANOVA was conducted on strength ratings using List Length as a within-subjects variable.<sup>4</sup> The only significant effect

<sup>2</sup> We also analyzed the full design. A 2 (Short List: weak, strong)  $\times$  2 (List Order: short first, long first)  $\times$  2 (List Length: short, long) mixed-model ANOVA was conducted on strength ratings, with List Length as a within-subjects variable. There was a significant effect of Short List, with ratings in the strong Short List condition higher than those in the weak Short List condition ( $M_s = 77.4, 60.3; SD_s = 15.9, 27.9; \chi^2(1, N = 87) = 36.58, p < .0001, d = .75$ ). The strong publications thus led to higher overall ratings than the weak publications, but note that this analysis averaged across ratings for the short and long lists. The difference between ratings for just the three strong vs. three weak publications can be seen by comparing the two dark bars in the top panel of Figure 2 to the two dark bars in the bottom panel. Collapsing across List Order, the respective means were 75.3 ( $SD = 17.2$ ) and 37.8 ( $SD = 20.9$ ),  $t(85) = 9.13, p < .0001, d = 1.96$ . There was also an effect of List Length, such that the list with six publications received a higher average rating than the lists with three publications ( $M_s = 81.1, 56.8; SD_s = 12.6, 26.7; \chi^2(1, N = 87) = 188.58, p < .0001, d = 1.17$ ). This effect occurred because the long list always includes the three strong publications, whereas the short list is comprised of only the weak publications half the time. There was also a significant interaction between Short List and List Length,  $\chi^2(1, N = 87) = 131.85, p < .0001, d = 2.40$ : Adding the three strong publications to the three weak ones led to a large increase in rated strength, but adding the three weak ones to the three strong ones led to only a small increase. No other effects were significant.

<sup>3</sup> From the society's website ([www.spsp.org/about-spsp](http://www.spsp.org/about-spsp)): "... the Society for Personality and Social Psychology is the largest organization of social psychologists and personality psychologists."

<sup>4</sup> Throughout this article, we report Type 3 mixed-model ANOVAs and likelihood ratio tests associated with them for statistical hypothesis testing (see Pinheiro & Bates, 2009). We estimated these models via maximum likelihood with the R package *lme4* (Bates, Maechler, Bolker, & Walker, 2018), and we used the packages *car* (Fox, Weisberg, & Price, 2018) and *lmerTest* (Kuznetsova, Brockhoff, & Christensen, 2018) for hypothesis testing.



Daggert, J.A., [REDACTED], [REDACTED], & [REDACTED] (in press). [REDACTED] *Social Behavior & Personality*.

Daggert, J.A., & [REDACTED] (in press). [REDACTED] *Journal of Personality and Social Psychology*.

Daggert, J.A., & [REDACTED] (2016). [REDACTED] *New Ideas in Psychology*.

Daggert, J.A., [REDACTED], [REDACTED], & [REDACTED] (2016). [REDACTED] *Journal of Experimental Psychology: General*.

[REDACTED] Daggert, J.A., & [REDACTED] (2015). [REDACTED] *Psychological Science*.

[REDACTED] Daggert, J.A., & [REDACTED] (2015). [REDACTED] *Psychological Reports*.

Figure 1. List of six publications.

was that of List Length: The list with three strong and three weak publications was rated as stronger ( $M = 83.0, SD = 13.6$ ) than the list with only three strong publications ( $M = 79.6, SD = 13.3, \chi^2(1, N = 163) = 18.98, p < .0001, d = .26$ ). Candidate strength was increased by adding three weak publications,  $t(80) = 2.70, p = .008, d = .20$  and was decreased by removing them,  $t(79) = 3.37, p = .001, d = .31$ . At the individual level, 62% of participants rated the candidate as stronger when the weak publications were included, while 24% rated him as weaker and 14% rated him as equally strong ( $\chi^2(1, N = 161) = 9.45, p = .002, V = .24$  for stronger vs. weaker and equally strong combined). These results are consistent with those of Experiment 1 and show that adding weak publications strengthens a CV.

### Experiment 3

The results of Experiments 1 and 2 could be interpreted as indicating that weak publications make an applicant appear stronger because weak publications are considered valuable in themselves. Perhaps each publication, no matter the outlet, has value and reflects positively on its authors. As we explain below, however, the results are also consistent with not valuing weak publications per se, but with caring only about the number of strong publications—in the future.

Evaluating a junior applicant is in large part about predicting future performance. Assume that the evaluator cares only about the expected number of future strong publications. One can think of the applicant's publication list as providing two cues regarding the number of future strong publications: current number of publications, and current percentage of strong publications. In Experiment

2 (and in the strong Short List conditions in Experiment 1), one list had twice as many publications (six vs. three), but only half the percentage of strong publications (50% vs. 100%). If these trends were to continue in the future—twice as many publications in one case, but also only half as many strong publications—the number of strong publications in each CV would increase at the same rate. But these trends are unlikely to continue. A new Ph.D. with 100% strong publications is unlikely to publish every future article in a high-impact journal. Similarly, an applicant with twice the number of publications as another may not continue to publish at twice the rate. Instead, the outcomes are likely to be regressive, or less extreme in the future. Importantly, if predictions regarding the percentage of strong publications are more regressive than predictions regarding the number of publications, then the applicant with more current publications (which includes the added weak ones) would be expected to have more strong publications in the future.

To take an extreme example, if the future percentage of strong publications is completely regressive—that is, is not at all predictable from the current percentage—then the prediction for each applicant's future percentage would be the same (say, 50%). And if the future number of publications is not regressive at all—that is, current extreme values are equally extreme in the future—then predictions about the future would be the same as the current trend (e.g., an applicant who currently has twice as many publications as another applicant would also be expected to have twice as many in the future). To the extent that this is the case, the applicant with more current publications will not only be expected to have more publications in the future, but will also be expected to have more strong publications—twice as many, in fact, in the above scenario.

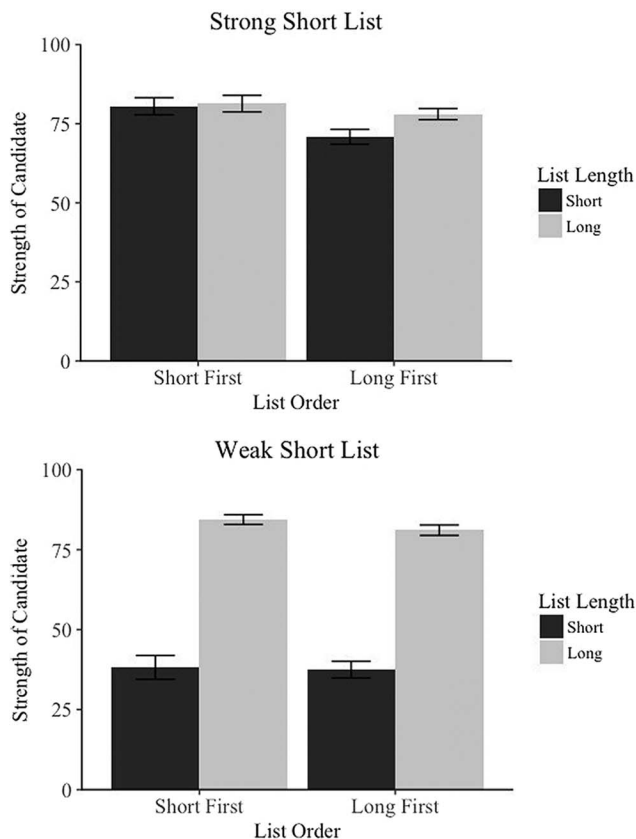


Figure 2. Experiment 1: Mean strength ratings as a function of short list, list length, and list order. Standard error bars are shown.

Intuitively, the pattern in which future percentage of strong publications is more regressive than future number of publications may be more compelling than the opposite pattern. Publishing in only top journals is very extreme; arguably, researchers have less control over whether these journals accept their work. Frequency of publishing—divorced from journal outlet—might be more under the author's control and thus less regressive.

Experiment three tests whether participants view the list with the added weak publications as stronger because they believe that having more current total publications predicts having more future strong publications. We provided the strong Short List to one group and the list of six publications to another, and asked them to predict how many new publications they expected the applicant to have in 5 years, as well as what percentage they expected to be published in high-impact journals. This allowed us to calculate the expected number of future strong publications for each list.

## Method

**Participants.** We recruited 200 academics (84 females) from the Society for Personality and Social Psychology listserv. This sample size represents all of the data obtained in response to our recruitment e-mail; a final sample of 164 remained after removing incomplete responses. Eighty-one percent of respondents reported being from a psychology department, 8% from a business school,

and 13% from a different department. Eleven respondents (7%) did not identify as professors, and six failed the attention check (4%).

**Materials and procedure.** As in Experiment 2, there were just two publication lists (three strong or three strong + three weak), but now each list was presented to only one group of participants (i.e., list length was a between-subjects variable). Furthermore, rather than rating candidate strength, participants predicted both the quality and quantity of publications they expected the applicant to produce in the near future. To assess expected publication quality, they were asked, "Given this publication list, what percentage of the candidate's new publications over the next 5 years would you predict to be in top-tier (A-level) journals?". Participants responded on a scale with percentages in intervals of 10 (i.e., 0%, 10%, 20% . . .100%). For publication quantity estimates, participants were asked, "Given this publication list, what is your best estimate of the total number of new journal publications this candidate will have over the next 5 years?" They provided their responses on a 21-point scale ranging from "0" to "20+".

## Results and Discussion

Participants predicted that the applicant with six publications (three strong + three weak) would have more new publications in 5 years than the applicant with three strong publications:  $M_s = 10.0$ ,  $SD = 4.2$  versus  $M = 8.4$ ,  $SD = 3.7$ ,  $t(162) = 2.53$ ,  $p = .012$ ,  $d = .40$ . They also predicted that the applicant with three strong publications would have a higher percentage of strong publications than the applicant with three strong and three weak publications:  $M = 60.5$ ,  $SD = 21.1$ , versus  $M = 48.3$ ,  $SD = 20.9$ ,  $t(162) = 3.70$ ,  $p = .0003$ ,  $d = .58$ . Note that the predictions for both variables are highly regressive. The applicant currently with twice the publications of the other applicant is predicted to have only 19% more publications in the future, and the applicant currently with twice the percentage of strong publications is predicted to have only a 25% higher percentage of strong publications in the future.

For each participant, we calculated the expected number of future strong publications by multiplying the predicted number of future publications by the predicted percentage of future strong

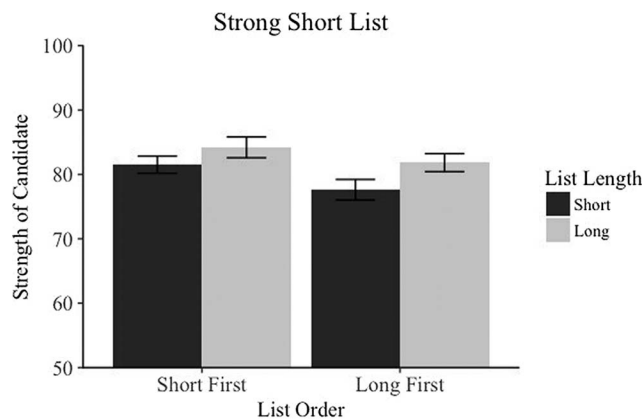


Figure 3. Experiment 2: Mean strength ratings as a function of list length and list order. Standard error bars are shown.

publications (divided by 100). The resulting values were virtually identical for the list with three strong publications and for the list with three strong and three weak publications, with both  $M_s = 5.2$ ;  $SD_s = 3.5$  and  $3.1$ ,  $t(162) = .05$ ,  $p = .96$ ,  $d = .01$ . These results provide evidence against the hypothesis that adding weak publications strengthens a CV because the current number of publications is seen as a more reliable signal of future performance than is the current percentage of strong publications, and that additional weak publications are only viewed positively because they increase the expected number of future strong publications. Indeed, the results showed at least as much regression for the predicted number of publications as for the predicted percentage of strong publications, and further revealed that the two candidates were expected to publish an equal number of strong publications in the future. It does not seem to be the case that the results of Experiments 1 and 2 can be explained by evaluators' caring only about future strong publications. Instead, it appears that the added weak publications are themselves valued.

### Experiment 4

Experiments 1 and 2 found that the addition of weak publications increased ratings of a hypothetical job candidate. The number of added weak publications was small, however, and always equal to the number of strong publications. It is unclear whether adding a larger number of weak publications would have a different effect. Powdthavee et al. (2017), for instance, found that adding weak publications decreased economists' evaluations, and a methodological difference between their study and ours was that they added a large number of weak publications (to a list that already had several weak publications). Perhaps a small number of weak publications helps, but a large number of weak publications overwhelms the strong ones and lowers evaluations. Such a finding would be intriguing because it would suggest that evaluators do not treat each additional weak publication as an independent piece of new information, and would thus contradict both of the simple models of belief updating that we have been discussing (averaging vs. adding).

Experiment 4 thus tests whether adding even more weak publications—nine rather than three—indeed hurts the candidate. In Experiment 4, we also elicit judgments on two dimensions that may help us understand any negative or positive effects of added weak publications: How hardworking the applicant is, and whether the applicant generates innovative research ideas. It could be, for example, that weak publications help because they signal a hardworking applicant, but a large proportion of weak publications could signal an inability to generate good ideas. Finally, Experiment 4 generalizes our previous studies by relying on a substantially different list of journals and, furthermore, by targeting cognitive (rather than social) psychology. We also conducted a pretest to examine how often the journals were recognized and their perceived quality.

### Method

**Pretest.** The previous studies relied solely on objective ratings of journal impact to select high and low-tier publication outlets. In Experiment 4, we again selected journals based on their AI percentile score, but also verified that perceptions of representative

faculty members aligned with these ratings. Thus, we asked 27 professors (19 responded) at a large California research university's psychology department to rate 14 journals (11 were low-tier, three high-tier). For each journal, they indicated whether they recognized its name (yes, no, or not sure) and provided their best estimate of its quality on a 7-point scale (1 = Very low quality, 7 = Very high quality). As expected, the weak journals were much less likely to be recognized (29% vs. 86% for strong journals) and were rated far lower on average than the strong journals (2.7 vs. 5.6). The six that received the lowest ratings served as the weak journals in this experiment. It is worth noting that unrecognized journals received low quality ratings. Thus, it is not crucial that our participants know that a particular journal is low impact; a lack of recognition suffices (Goldstein & Gigerenzer, 2002).

**Participants.** We recruited 235 academics (131 females) through the listservs of three cognitive psychology-oriented organizations: Cognitive Science Society, Vision Science Society, and Cognitive Development Society. This sample represents all responses to our recruitment e-mail; 206 participants remained after removing those who failed to finish. Seventy-four percent reported being from a psychology department, 10% from a cognitive science department, and 16% from a different department. Fifteen respondents (7%) said that they were not professors. We did not include an attention check for this study.

**Materials.** As before, each participant was presented with two hypothetical publication lists—one with three strong publications and one with added weak publications. However, we now manipulated between-subjects whether three or 9 weak publications were added. To otherwise reduce differences between two lists with the weak publications, the list with 9 used the three weak journals from the shorter list, but used each of them twice. Thus, only three journals on the list with 9 weak publications differed from those on the list with three weak publications. All publications were either in press or had been published in the last two years, with equal numbers of weak publications for each time period. The applicant was first, second, or third author on each publication, and held the same proportion of those authorships for each publication list.

For the high-impact publication outlets, we used the same general psychology journals as before (*Psychological Science*; *Journal of Experimental Psychology: General*) and replaced the high-impact social psychology journal with a cognitive counterpart (*Cognitive Psychology*). Our pretest confirmed that these journals are indeed regarded as high quality (mean ratings ranged from 5.2 to 6 out of 7). When we conducted both the pretest and the experiment (May 2018), AI percentile scores for these journals ranged from 96 to 98 (impact factors were between 4.4 and 5.7). The lowest-rated journals from the pretest served as our low-impact journals (*The Psychological Record*; *Memory Studies*; *Pratiques Psychologiques*; *Behavioral Psychology/Psicología Conductual*; *Biologically Inspired Cognitive Architectures*; *Journal of Psychology and Theology*). Their percentile scores ranged from 7 to 34 (impact factors ranged between 0.3 and 1.1).

**Procedure.** Participants evaluated the strength of the applicant two times, once based on a list of three strong publications, and once when that list contained added weak publications. The number of added weak publications varied between subjects (three strong + three weak or three strong + 9 weak). Evaluation order was counterbalanced.

After making their evaluations, participants compared the two publication lists in terms of hard work and innovative research ideas. Participants were asked, “Which CV do you think indicates more *hard work*, the first one you saw or the second one?” and “Which CV do you think indicates greater *ability to generate innovative research ideas*, the first one you saw or the second one?”. They responded on 7-point scales in which the endpoints of 1 and 7 corresponded to the first and second CV, respectively (e.g., 1 = “First CV indicates far more hard work;” 7 = “Second CV indicates far more hard work”). For both questions, to ensure that it was clear which CV was which, the CVs were reproduced below their respective anchors, in reduced size. The questions’ presentation order was counterbalanced.

## Results

Mean strength ratings are displayed in Figure 4. We performed a 2 (List Order: short first vs. long first)  $\times$  2 (List Length: short vs. long)  $\times$  2 (Long List Type: three strong + three weak vs. three strong + 9 weak) mixed-model ANOVA on strength ratings, specifying List Length as a within-subjects variable. A main effect of List Length revealed that the candidate was rated as stronger when the publication list contained (three or 9) weak publications ( $M = 68.6$ ,  $SD = 20.1$ ) compared to when it did not ( $M = 59.4$ ,  $SD = 22.7$ ),  $\chi^2(1, N = 206) = 69.85$ ,  $p < .0001$ ,  $d = .43$ . Ratings increased when (three or nine) weak publications were added,  $t(104) = 5.60$ ,  $p < .0001$ ,  $d = .33$ , and decreased when they

were removed,  $t(100) = -6.23$ ,  $p < .0001$ ,  $d = .54$ . These results replicate the general finding of Experiments 1 and 2: Adding weak publications helps. Furthermore, a significant main effect of Long List Type showed that participants gave higher overall ratings when the long list contained 9 additional weak publications ( $M = 68.4$ ,  $SD = 20.18$ ) instead of three ( $M = 59.2$ ,  $SD = 22.74$ ),  $\chi^2(1, N = 206) = 12.00$ ,  $p < .001$ ,  $d = .43$ . But the interaction between List Length and Long List Type was not significant ( $p = .25$ ), so the main effect of Long List Type does not imply that participants viewed adding 9 weak publications more positively than adding three weak publications. Instead, participants who saw the long list with 9 weak publications provided higher ratings not only of the long list with the added weak publications, but also of the short list with only three strong publications (which was the same for all participants).<sup>5</sup>

Importantly, an individual-level analysis confirmed both that the additional weak publications helped, and that adding nine of them did not yield a greater benefit over adding just three. When three weak publications were added, 73% of participants rated the candidate as stronger, 17% as weaker, and 9% as equally strong,  $\chi^2(1, N = 98) = 21.52$ ,  $p < .0001$ ,  $V = .47$ , for stronger versus weaker and equally strong combined. When nine weak publications were added, a virtually identical result obtained: 73% rated the candidate as stronger, 19% as weaker, and 7% as equally strong,  $\chi^2(1, N = 108) = 23.15$ ,  $p < .0001$ ,  $V = .46$ . The distributions of individual-level responses were not statistically different from one another ( $p = .72$ ). Thus, adding weak publications helped, but it did not matter if three or nine were added.

We next examined how adding weak publications affected inferences about the candidate’s ability to work hard and to generate innovative research ideas. We subtracted 4 from each response, so that 0 corresponded to the midpoint on the scale, negative values corresponded to ratings favoring the short list with three strong publications, and positive values corresponded to ratings favoring the longer list with added weak publications (either three or nine). Again, weak publications helped: Participants rated the candidate with the longer publication list as both harder working ( $M = 1.37$ ,  $SD = 1.18$ ,  $t(205) = 16.63$ ,  $p < .0001$ ,  $d = 1.16$ ) and better able to generate innovative research ideas ( $M = .84$ ,  $SD = 1.17$ ,

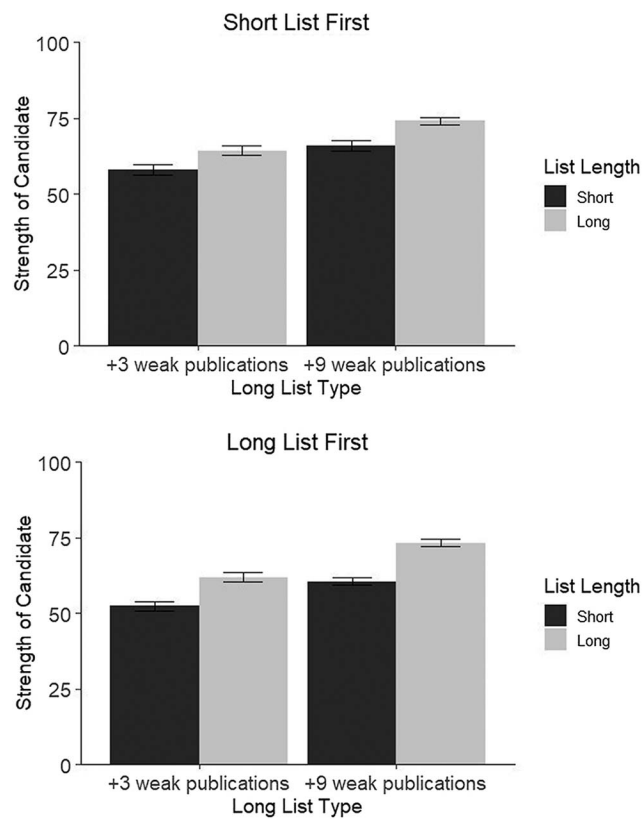


Figure 4. Experiment 4: Mean strength ratings as a function of list order, list length, and long list type. Standard error bars are shown.

<sup>5</sup> At the suggestion of a reviewer, we also conducted a set of post-hoc, between-subject analyses restricted to each participant’s first rating. In line with the results of the full ANOVA, the same short list with three strong publications was rated somewhat stronger by participants who subsequently evaluated the long list with nine weak publications ( $M = 65.9$ ,  $SD = 23.3$ ) than by their counterparts who subsequently evaluated the long list with only three weak publications ( $M = 58.0$ ,  $SD = 24.4$ ),  $t(103) = 1.69$ ,  $p = .095$ ,  $d = .33$ . Collapsing across this unpredicted difference, the average rating for the short list when presented first was 62.0 ( $SD = 24.1$ ), which was about equal to that for the long list with three weak publications when presented first ( $M = 61.8$ ,  $SD = 22.4$ ),  $p = .97$ , and significantly smaller than that for the long list with 9 weak publications when presented first ( $M = 73.2$ ,  $SD = 17.6$ ),  $t(158) = 3.05$ ,  $p < .01$ ,  $d = .50$ . Finally, the long list with 9 weak publications was rated significantly stronger than that with three weak publications when presented first,  $t(99) = 2.86$ ,  $p < .01$ ,  $d = .55$ . This first-list-only analysis thus suggests that adding three weak publications does not help, but adding 9 does, whereas both the full ANOVA and an individual-level analysis (subsequently reported in the main text) indicate that adding three helps but adding nine is no better than adding three. Importantly, however, and in marked contrast with the results reported by Powdthavee et al., all three analyses confirm that adding (three or nine) weak publications does not hurt the candidate.



$t(205) = 10.32, p < .0001, d = .72$ ). Ratings of hard work did not differ based how many weak publications were added (three vs. nine,  $M_s = 1.31$  vs.  $1.14, t(205) = .78, p = .44$ ), and neither did ratings of the ability to generate innovative research ideas ( $M_s = .83$  vs.  $.86, t(204) = .21, p = .84$ ). Thus, similar to what we found with strength ratings, adding a few weak publications led to more positive inferences about the applicant, and adding even more weak publications did not have a negative effect.

### General Discussion

A series of experiments with psychology professors as participants showed that publications in low-impact journals increased the perceived strength of a hypothetical tenure-track job candidate in psychology. Experiments 1 and 2 found that adding a few such “weak” publications to a list of “strong” ones increased the reported strength of the candidate, and that eliminating the weak ones decreased the candidate’s strength. This is inconsistent both with an averaging model of candidate ability based on the strength of publications, and with an adding model of belief updating that treats weak publications as negative evidence of a candidate’s ability. The data are consistent, however, with an adding model of belief updating in which publications in low-impact journals are seen as weak evidence in favor of a candidate’s ability. Experiment 3 provided evidence against the hypothesis that additional weak publications strengthen a CV because more current publications predict more strong publications in the future. Weak publications themselves appear to be valued. Experiment 4 showed that adding even more weak publications did not lead to lower evaluations. Furthermore, the data indicated that adding weak publications led the applicant to be seen as both harder working and having more innovative research ideas and, again, adding even more weak publications did not hurt.

Our findings differ from related work in economics. Powdthavee et al. (2017) found that adding publications in “weak” economics journals to a hypothetical economist’s publication record led economists to evaluate the record less favorably. It is worth noting that academic areas can differ substantially in their standards for research output. This could in principle explain the difference between our results and those of Powdthavee et al. In economics, where a single “Top 5” publication often makes a critical difference for the career path of Ph.D. candidates and junior faculty (Card & DellaVigna, 2013; Conley, Crucini, Driskill, & Önder, 2013), quality may matter much more than quantity. Psychologists, on the other hand, may value quantity relatively highly.

But there are also important methodological differences between Powdthavee et al.’s study and ours that may contribute to the discrepancy in the results. For instance, Powdthavee et al. did not mention the career stage of the candidate, the purpose of the evaluation, or the time period of the publication record. In contrast, we made these aspects explicit, which allowed us to hold them constant across conditions: Professors rated the strength of an individual at a specified career stage (new Ph.D.) for a specified purpose (hiring for a tenure-track position), based on a publication record that spanned a specified amount of time (three years). This approach limited additional inferences that could shift participant’s standards for evaluation (e.g., inferring that the candidate with

more publications was more senior, and judging them accordingly).

We have explored different subfields within psychology and different numbers of added weak publications. Our results were remarkably consistent, which suggests that they may also generalize further, at least for tenure-track applicants in psychology. Given the limited and nonreplenishing nature of our subject pool, however, our experiments focused on what we considered to be plausible publication records for a competitive, tenure-track job candidate in psychology. We did not explore a number of related questions, such as whether adding even more weak publications would affect evaluations, whether changing the number of strong publications matters, or whether a longer track record (e.g., 6 years) would lead to less regressive predictions of performance compared to the 3-year record we examined. In addition, expected level of productivity may not be equal for different career stages—the first 3 years, for instance, may be less productive than the next three—and we did not examine whether evaluations of more senior candidates (e.g., Full Professors) would be similarly boosted by weak publications. Finally, we note that added weak publications may not have as strong an impact on actual hiring decisions, where much richer information is available than just the publication lists used in our studies. Nevertheless, initial invitations to interview are often heavily influenced by a candidate’s CV, with special attention paid to the publication record.

Our results suggest that publishing at least some additional articles in low-impact psychology journals is being reinforced by the market. The advice for junior psychology candidates is clear: Assuming that it is not taking away (too much) time from work that would result in a strong publication, publish your less impressive or less interesting work. This, in turn, could be good or bad for the field of psychology, depending on how articles in low- and high-impact journals differ (besides impact). If the primary difference is that articles in low-impact journals tend to make incremental contributions, while those in high-impact journals tend to break new ground, then weak articles are beneficial, just less so. However, if articles appear in low-impact journals because they do not meet the quality standards of high-impact journals, and contain methods, analyses, or reasoning that is flawed, then reinforcing these publications may impede progress.

Interestingly, the recent “replicability crisis” in psychology has focused on findings published in prestigious, high-impact journals, and their reproducibility has been called into question (Open Science Collaboration, 2015). Indeed, the emphasis that some high-impact journals place on surprising, “sexy” findings might contribute to the crisis (Nosek, Spies, & Motyl, 2012). Holding strength of evidence constant, effects that are *a priori* unlikely are more likely to be false. Less is known about the reproducibility of findings in low-impact journals. How their reproducibility compares to that of high-impact journals might again hinge on how articles in these two types of journals differ. If the main difference is that articles in low-impact journals tend to be incremental, their effects may be *more* likely to replicate. If instead the main difference is that articles in low-impact journals tend to be scientifically less sound, their effects will be less likely to replicate. It is therefore unclear whether additional publications in low-impact journals benefit the field. Our results make clear, however, that such publications do benefit tenure-track psychology job candidates.

## References

- Anderson, N. H. (1981). *Foundations of information integration theory*. New York, NY: Academic Press.
- Anderson, N. H., & Alexander, G. R. (1971). Choice test of the averaging hypothesis for information integration. *Cognitive Psychology*, 2, 313–324. [http://dx.doi.org/10.1016/0010-0285\(71\)90017-X](http://dx.doi.org/10.1016/0010-0285(71)90017-X)
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2018). *lme4: Linear mixed-effects models using Eigen and S4* (R package version 1.1–19). Retrieved from <http://CRAN.R-project.org/package=lme4>
- Card, D., & DellaVigna, S. (2013). Nine facts about top journals in economics. *Journal of Economic Literature*, 51, 144–161. <http://dx.doi.org/10.1257/jel.51.1.144>
- Carlson, R. A., & Dulany, D. E. (1988). Diagnostic reasoning with circumstantial evidence. *Cognitive Psychology*, 20, 463–492. [http://dx.doi.org/10.1016/0010-0285\(88\)90013-8](http://dx.doi.org/10.1016/0010-0285(88)90013-8)
- Conley, J., Crucini, M. J., Driskill, R., & Önder, A. S. (2013). The effects of publication lags on life cycle research productivity in economics. *Economic Inquiry*, 51, 1251–1276. <http://dx.doi.org/10.1111/j.1465-7295.2012.00480.x>
- Conroy, M., Dusansky, R., Drukker, D., & Kildegaard, A. (1995). The Productivity of Economics Departments in the U. S.: Publications in the Core Journals. *Journal of Economic Literature*, 33, 1966–1971.
- Edwards, W. (1968). Conservatism in human information processing. In B. Kleinmuntz (Ed.), *Formal representation of human judgment* (pp. 17–52). New York, NY: Wiley.
- Fox, S., Weisberg, S., & Price, B. (2018). *car: Companion to Applied Regression* (R package version 3.0–2). Retrieved from <http://CRAN.R-project.org/package=car>
- Goldstein, D. G., & Gigerenzer, G. (2002). Models of ecological rationality: The recognition heuristic. *Psychological Review*, 109, 75–90. <http://dx.doi.org/10.1037/0033-295X.109.1.75>
- Hogarth, R. M., & Einhorn, H. J. (1992). Order effects in belief updating: The belief-adjustment model. *Cognitive Psychology*, 24, 1–55. [http://dx.doi.org/10.1016/0010-0285\(92\)90002-J](http://dx.doi.org/10.1016/0010-0285(92)90002-J)
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2018). *lmerTest: Tests for random and fixed effects for linear mixed effect models* (R package version 3.0–1). Retrieved from <http://CRAN.R-project.org/package=lmerTest>
- Lopes, L. L. (1985). Averaging rules and adjustment processes in Bayesian inference. *Bulletin of the Psychonomic Society*, 23, 509–512. <http://dx.doi.org/10.3758/BF03329868>
- Lopes, L. L. (1987). Procedural debiasing. *Acta Psychologica*, 64, 167–185. [http://dx.doi.org/10.1016/0001-6918\(87\)90005-9](http://dx.doi.org/10.1016/0001-6918(87)90005-9)
- McKenzie, C. R. M., Lee, S. M., & Chen, K. K. (2002). When negative evidence increases confidence: Change in belief after hearing two sides of a dispute. *Journal of Behavioral Decision Making*, 15, 1–18. <http://dx.doi.org/10.1002/bdm.400>
- McPherson, M. (2012). Ranking U.S. Economics Programs by Faculty and Graduate Publications: An Update Using 1994–2009 Data. *Southern Economic Journal*, 79, 71–89. <http://dx.doi.org/10.4284/0038-4038-79.1.71>
- Meyvis, T., & Janiszewski, C. (2002). Consumers' beliefs about product benefits: The effect of obviously irrelevant product information. *Journal of Consumer Research*, 28, 618–635. <http://dx.doi.org/10.1086/338205>
- Nisbett, R. E., Zukier, H., & Lemley, R. E. (1981). The dilution effect: Nondiagnostic information weakens the implications of diagnostic information. *Cognitive Psychology*, 13, 248–277. [http://dx.doi.org/10.1016/0010-0285\(81\)90010-4](http://dx.doi.org/10.1016/0010-0285(81)90010-4)
- 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. <http://dx.doi.org/10.1177/1745691612459058>
- Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, 349(6251), aac4716. <http://dx.doi.org/10.1126/science.aac4716>
- Oppenheimer, D. M., Meyvis, T., & Davidenko, N. (2009). Instructional manipulation checks: Detecting satisficing to increase statistical power. *Journal of Experimental Social Psychology*, 45, 867–872. <http://dx.doi.org/10.1016/j.jesp.2009.03.009>
- Pinheiro, J., & Bates, D. (2009). *Mixed-Effects models in S and S-Plus*. New York, NY: Springer.
- Powdthavee, N., Riyanto, Y. E., & Knetsch, J. L. (2017, April). *Impact of lower-rated journals on economists' judgments of publication lists: Evidence from a survey experiment* (IZA Discussion Paper No. 10752). Retrieved from <https://ssrn.com/abstract=2964161>
- Shanteau, J. (1972). Descriptive versus normative models of sequential inference judgment. *Journal of Experimental Psychology*, 93, 63–68. <http://dx.doi.org/10.1037/h0032509>
- Shanteau, J. C. (1970). An additive model for sequential decision making. *Journal of Experimental Psychology*, 85, 181–191. <http://dx.doi.org/10.1037/h0029552>
- Shanteau, J. (1975). Averaging versus multiplying combination rules of inference judgment. *Acta Psychologica*, 39, 83–89. [http://dx.doi.org/10.1016/0001-6918\(75\)90023-2](http://dx.doi.org/10.1016/0001-6918(75)90023-2)
- Troutman, C. M., & Shanteau, J. (1976). Do consumers evaluate products by adding or averaging attribute information? *Journal of Consumer Research*, 3, 101–106. <http://dx.doi.org/10.1086/208657>
- Wallsten, T. S., & Manley Sapp, M. (1977). Strong ordinal properties of an additive model for the sequential processing of probabilistic information. *Acta Psychologica*, 41, 225–253. [http://dx.doi.org/10.1016/0001-6918\(77\)90037-3](http://dx.doi.org/10.1016/0001-6918(77)90037-3)

Received October 12, 2017

Revision received March 16, 2019

Accepted March 20, 2019 ■