Joy and rigor in behavioral science

Hanne K. Collins, Ashley V. Whillans, Leslie K. John *

Harvard Business School, United States

A R T I C L E   I N F O
Edited by Don Moore and Stefan Thau

Keywords:
Open science
Pre-registration
Exploration
Confirmation
False positives
Career satisfaction
Diversity

A B S T R A C T
In the past decade, behavioral science has seen the introduction of beneficial reforms to reduce false positive results. Serving as the motivational backdrop for the present research, we wondered whether these reforms might have unintended negative consequences for researchers’ behavior and emotional experiences. In an experiment simulating the research process, Study 1 (N = 449 researchers) suggested that engaging in a pre-registration task impeded the discovery of an interesting but non-hypothesized result. Study 2 (N = 400 researchers) indicated that relative to confirmatory research, researchers found exploratory research more enjoyable, motivating, and interesting; and less anxiety-inducing, frustrating, boring, and scientific. These studies raise the possibility that emphasizing confirmation can shift researchers away from exploration, and that such a shift could degrade the subjective experience of conducting research. Study 3 (N = 314 researchers) introduced a scale to measure “prediction preoccupation”—the feeling of heightened concern over, and fixation with, confirming predictions.

Since then, a growing number of researchers and academic journals have been embracing welcome and much-needed reforms: pre-registration, running replication studies, and transparently reporting methods and results (Camerer et al., 2016; Gr Caldwell & Sherman, 2016; Dougherty, Slevc, & Grand, 2019; Klein et al., 2018; LeBel & John, 2016; Nosek et al., 2015; Nosek et al., 2019; Open Science Collaboration, 2015; Uhlmann et al., 2019; Van’t Veer & Giner-Sorolla 2016; Vazire, 2016; Weston, Ritchie, Rohrer, & Przybylski, 2019; see Nelson, Simmons, & Simonsohn, 2018 for a review). In their focus on reducing false positives (i.e., Type I error), these reforms emphasize confirmation; in particular, they encourage specifying and testing predictions (Popper, 1959; Platt, 1964; Jaeger & Halliday, 1990). We believe—like many other researchers (e.g., Munafò et al., 2017; Smaldino & McElreath, 2016; Spellman, 2015)—that these reforms are instrumental to improving the quality of research. Although we are not aware of any direct causal evidence of reform efficacy, one study reported a dramatic drop in positive results—from 57% to 8%—following the introduction of a pre-registration requirement (Kaplan & Irvin, 2015). Another study reported a decrease in positive results following the introduction of a registered report manuscript category—from 24% for non-registered reports to 8% for registered reports (Wiseman, Watt, & Kornbrot, 2019; see Chambers, 2019 for a review). If one assumes these reductions are driven by decreases in false positives (as opposed to

☆ This article is an invited submission. It is part of the special issue “Best Practices in Open Science,” Edited by Don Moore and Stefan Thau.

* Corresponding author.
E-mail address: ljohn@hbs.edu (L.K. John).

https://doi.org/10.1016/j.obhdp.2021.03.002
Received 30 December 2019; Received in revised form 25 February 2021; Accepted 12 March 2021
Available online 10 April 2021
0749-5978/© 2021 Elsevier Inc. All rights reserved.
true positives)—plausible, given the prevalence of false positives pre-reform (John et al., 2012; Ioannidis, 2012; Simonsohn, Nelson, & Simmons, 2014)—such results are suggestive of the effectiveness of reforms.

However, alongside the benefits of reforms, there may be unintended negative consequences. In particular, although confirmation and exploration are mutually reinforcing—both are required for the scientific process—we wondered whether the reform movement’s emphasis on confirmation may cause researchers to act as if confirmation precludes exploration. As a President of the American Psychological Association remarked, “I fear that pre-registration will stifle discovery. Science isn’t just about confirming hypotheses” (Goldin-Meadow, 2016).

With the reform movement’s emphasis on confirmation as a motivating backdrop, first, we sought to test whether emphasizing confirmation reduced researchers’ propensity to explore. Next, we explored the question: to the extent that reforms have shifted researchers’ behavior toward confirmation, what (if any), impact has there been on the subjective experience of conducting research? If researchers differentially experience confirmatory and exploratory research, the reform movement could influence how behavioral scientists experience their work, potentially affecting job satisfaction. Thus, our second goal was to explore behavioral scientists’ subjective experience of conducting research, with an emphasis on how their experiences vary in confirmatory versus exploratory settings.

Our third goal was to focus on one subjective experience in particular. Namely, we assessed the extent to which researchers felt anxious fixation over making and confirming predictions. Given the (appropriately) heightened importance of confirmation in today’s climate of research reform, we wondered whether prediction might feel particularly high stakes, and hence, be a source of ruminative preoccupation. Therefore, we sought to develop and validate a scale to measure the construct of “prediction preoccupation,” which refers to experiencing heightened concern over, and fixation with, confirming predictions.

The construct of prediction preoccupation is rooted in prior work in clinical psychology and organizational behavior. Central to the construct is the notion of rumination—the presence of recurrent conscious thoughts about a common theme (Ciarocco, Vohs, & Baumeister, 2010; Martin & Tesser, 1996). Such thoughts are preoccupying—because they recur, even when they are no longer contextually relevant—and tend to be negative in tone (Ciarocco et al., 2010; Fresco, Frankel, Mennin, Turk, & Heimberg, 2002; Smith & Alloy, 2009). Prior work also indicates that ruminative thoughts are most likely to emerge when people feel as though they are not progressing satisfactorily towards personally important goals (Martin & Tesser, 1996; Martin, Shrira, & Startup, 2004). Today’s climate of reform positions the acts of making and confirming predictions as important goals, while also highlighting researchers’ shortcomings in attaining them. As a result, we propose that researchers may experience prediction preoccupation.

1. Overview

We present three studies. First, we provide an initial test of the potential impact of one aspect of the reform movement—an emphasis on confirmation—on behavioral scientists’ behavior; namely, their propensity to explore. In a study simulating the research process, we tested whether engaging in a pre-registration task could impede exploration (Study 1). Next, we explored possible implications for how behavioral scientists experience the research process. Specifically, we assessed their subjective experience with conducting research, and whether it differed within confirmatory versus exploratory research settings (Study 2). Finally, we developed and validated a scale to measure prediction preoccupation and discuss its correlates (Study 3). All data and materials are available through the Open Science Framework here https://bit.ly/3eE3HJb.

2. Study 1: Exploration inhibition

Although confirmatory and exploratory research are not mutually exclusive, Study 1 tested whether confirmatory research settings can unintentionally stifle exploration. We also tested whether a simple reminder to explore could mitigate this hypothesized effect. Study 1 assessed these ideas in a simulated research task in which participants in the experimental conditions were placed in a confirmatory context by pre-registering a directional prediction, sample size, and data analysis plan. We examined these participants’ likelihood of discovering an interesting but non-predicted interaction relative to those placed in an exploratory context.

2.1. Method

2.1.1. Sample

We recruited behavioral scientists to participate in a brief survey in which they would simulate the research process. We recruited participants via email and listservs, offering a chance to win a $50 Amazon gift card. We invited 6,778 academic psychologists at major U.S. universities via email (using an updated version of the list used in John et al., 2012), and by posting the survey link to the SPSP Student Group, ACR listserv, and AOM OB Student Network. Our response rate is unknown as we do not know the number of people belonging to these listservs. Our recruitment efforts garnered 449 respondents (approximate sample composition by channel: 30% via email list; 39% via SPSP student group; 17% via ACR listserv; 0% via AOM OB; 14% unspecified). 79% of participants completed all primary outcome measures.

2.1.2. Procedure

Participants were randomly assigned to one of three conditions: an exploration condition (N = 149), a confirmation condition (N = 150), or a hybrid condition (N = 150) in which we sought to highlight both confirmation and exploration. Participants in the exploration condition imagined that they had collected a dataset of 1,000 responses to three questions: (1) “Do you do yoga on a weekly basis?” (0 = No; 1 = Yes); (2): “How happy are you today?” (1 = Not at all; 7 = Extremely); and (3): “What is your gender?” (0 = Male; 1 = Female). On the next page, they were asked which analyses they would run on the data (described below, in the Measures subsection).

For participants in the confirmation and hybrid conditions, we situated this research project within the context of confirmatory research by simulating the pre-registration of a hypothesis. Prior to being asked which analyses they would run (on the same three measures as in the exploration condition—i.e., yoga, happiness, gender), these participants were first told: “Suppose you had a prediction that people who reported doing yoga on a weekly basis would report significantly greater happiness and who also did not report doing yoga on a weekly basis. Therefore, before collecting or analyzing any data, you decide to pre-register this hypothesis. Click ‘>’ to be taken to the pre-registration page.” Next, participants in the confirmatory and hybrid conditions were shown a screenshot of a completed pre-registration form and asked to click on a button (an image of Psychological Science’s pre-registration badge) to pre-register the study (Fig. 1). This form stipulated the target sample size and the key dependent variables, as well as the directional prediction and statistical analysis proposed to test that prediction. In

1 We thank John McCoy, Assistant Professor at the Wharton School, and Nick Fox, Research Scientist at the Center for Open Science for updating this list.

2 We surmise that we did not obtain any respondents from this channel because the post was not very salient. To view it, one had to a) be a member of AOM, b) be a member of this particular discussion group through Connect@AOM, and c) either explicitly log in to look for messages or have signed up for email updates for this specific discussion group.

3 Because we recruited through multiple channels, at the end of the survey, we included a question asking participants whether they had completed the survey before. Two people responded “yes” and four responded “maybe.” Everyone else reported that they had not completed the survey before. To be conservative, we included all participants in our final analyses. However, results were substantively equivalent when we included these six respondents.
support of this operationalization of a “confirmatory research context,” a separate study (Appendix S1) confirmed that researchers consider pre-registration and having an a priori sense of various aspects of the research—including how much data to collect, how the data will be analyzed, and what the result is likely to be—to be key components of confirmatory research. This operationalization is also consistent with how other scholars have thought of the constructs of exploratory versus confirmatory research (e.g., Wagenmakers et al., 2012).

Fig. 1. Pre-registration simulation presented to participants in the confirmation and hybrid conditions.

Fig. 2. Screenshot of the analysis selection and viewing procedure. Note. In this example, the participant has selected to view only the results of the ‘t-test: IV = Gender, DV = Happiness’ and ‘t-test: IV = Yoga, DV = Happiness’ analyses. Further, this participant selected to view the results in ‘R’ format.
After participants in the confirmation and hybrid conditions had clicked on the pre-registration button, the next screen they encountered displayed the pre-registration form, with the pre-registration badge added as a watermark (Fig. 1). These participants were next asked which analyses they would run on the data (described below, in the Measures subsection). For participants in the hybrid condition, the following prompt also appeared, in bolded green font, at the top of the page: “REMEMBER: Pre-registering doesn’t mean you can’t explore the data!”

2.2.1. Selecting analyses to run

Participants were told: “Now it’s time to analyze the data! Below are various analyses that could be run on these data. Which, if any, of the following analyses would you run on these data? Select all that apply. On the next page, we will display the results of any of the analyses you choose.” Participants indicated which analyses they wanted to view from the following list: Descriptive statistics: Gender; Descriptive statistics: Yoga; Descriptive statistics: Happiness; t-test: IV = Gender, DV = Happiness; t-test: IV = Yoga, DV = Happiness; 2x2 ANOVA: IVs = Yoga, Gender and DVs = Happiness. An “Other: Describe” option captured any additional analyses respondents were interested in seeing (see Fig. 2).

2.2.2. Reporting the interaction

An intent-to-treat analysis (i.e., including all participants regardless of whether they chose to view the interaction) indicated that participants in the exploration condition were more likely to report the interaction relative to those in the confirmation condition, $b = -0.19$, $SE = 0.23$, $p = .006$, and marginally more so than the hybrid condition, $b = 0.54$, $SE = 0.24$, $p = .06$. Unsurprisingly, this difference was driven by a larger proportion of participants in the exploration condition choosing to view the interaction. Among the participants who chose to view the results of the interaction, 92% decided to report it in their final write-up; this tendency did not differ by condition (exploration: 94%, confirmation: 90%, hybrid: 91%; $p$s $\geq$ 0.53).

2.2.3. Demographics

See Table 1 for full sample demographics. All results held when controlling for demographic characteristics (SOM). On an exploratory basis, we assessed whether treatment effects were moderated by demographics (e.g., gender, PhD year, tenure, engagement in pre-registration, use of experimental methods) and found no consistent patterns. See SOM for details.

2.3. Discussion

In a simulated research task, researchers randomly assigned to participate in a confirmatory research experience were significantly less likely to discover an interesting, but non-hypothesized interaction relative to those assigned to the exploration condition. A reminder to explore did not seem to mitigate this effect. To induce a confirmatory context, our pre-registration task prominently featured the act of making a directional prediction. That said, we note that pre-registration does not require making a formal prediction; rather, the essential activity of a pre-registration is the delineation of design and analysis choices, and which research questions are confirmatory versus exploratory (e.g., Nosek et al., 2019; Van’t Veer & Giner-Sorolla 2016; Wagenmakers et al., 2012).

Study 1 provides an “existence proof” that a focus on confirmation can impede exploration—concerning, considering that both of these activities are crucial to scientific discovery. One could argue that demand characteristics contributed to this effect—participants in the confirmation and hybrid conditions might have avoided the interaction because they thought the experimenters wanted them to only select the analyses required for testing the pre-registered hypothesis. By this logic, the hybrid condition should have fostered more exploration because it featured an explicit encouragement to explore. Instead, participants were just as unlikely to discover the interaction in the hybrid as in the confirmation condition, suggesting that demand effects did not drive exploration inhibition. Thus, we posit that our results are symptomatic of a heightened emphasis on confirmation as opposed to a product of demand.
In the hybrid condition, participants were four percentage points more likely to request the interaction relative to the confirmation condition; however, this effect was so small that it did not reach statistical significance. If this result is a true null, it highlights the need to test other ways of fostering exploration within confirmatory contexts. One starting point could be to test other, more directive reminders, such as remembering to explore researchers run an analysis in the first place — that is, you do not have a strong prediction, but if you do have a prediction, here it would be based more on theory or even intuition, than on prior empirical research. Exploratory research is common in the early stages of a research project. But it is not limited to the early stages — researchers often toggle back and forth between the subjective experience of conducting research? The goal of Study 2 was to explore researchers’ current subjective experience of conducting research, with an emphasis on how this experience may differ as a function of participation in exploratory versus confirmatory research.

3. Study 2: Researching researcher experiences

Study 1 suggested that an emphasis on confirmation can reduce exploration. Stemming from this result, we wondered: to the extent that reforms have shifted researchers’ behavior toward confirmation and away from exploration, what, if any, impact might such a shift have on the subjective experience of conducting research? The goal of Study 2 was to explore researchers’ current subjective experience of conducting research, with an emphasis on how this experience may differ as a function of participation in exploratory versus confirmatory research.

3.1. Methods

3.1.1. Sample

We recruited participants in-person at the 2019 meeting of the Society for Judgment and Decision Making (SJDM; n = 298) and online through the SJDM list-serv (containing approximately 3,000 members; n = 91) and Psych-Method-list-serv (n = 38), for a total of 431 respondents. The response rate is unknown as we do not know the exact number of individuals exposed to our recruitment sign (SJDM conference) or emails (SJDM/Psych-Method-list-servs). We analyzed the data from the 400 (93% of) respondents who reported conducting behavioral research. We used all available responses from the 352 respondents (out of 400 eligible respondents) who did not complete the entire survey. 89% of participants completed all primary outcome measures.

3.1.2. Procedure and measures

Subjective Experience of Exploration versus Confirmation. Respondents read a description of exploratory research (that was devised based on an iterative process described below):

"By exploratory research, we are referring to research for which you do not have a strong prior – that is, you do not have a strong sense from prior research (be it your own or others’) of what the result will be. You may or may not have a prediction, but if you do not have a prediction, here it would be based more on theory or even intuition, than on prior empirical research. Exploratory research is common in the early stages of a research project. But it’s not limited to the early stages – researchers often toggle back and forth between..."
exploratory and confirmatory research throughout the course of any given research project. We consider exploratory research to include activities such as: brainstorming research ideas that have not yet been addressed in the literature, thinking through how to operationalize those ideas you choose to pursue, designing and conducting the first tests of those research ideas, and finding out what the results of those first tests show."

Respondents then rated the extent to which they found exploratory research: enjoyable, motivating, interesting, frustrating, anxiety-inducing, boring, and scientific, on a scale from 1 = \textit{Strongly Disagree} to 7 = \textit{Strongly Agree}. Item order was randomized between-subjects.

Respondents also read a description of confirmatory research (derived from the iterative process described below):

"By confirmatory research, we are referring to research for which you DO have a prior – that is, you have a sense from prior research (be it your own or others’) of what the result will be. Confirmatory research is common in the later stages of a research project. But it’s not limited to the later stages – researchers often toggle back and forth between confirmatory and exploratory research throughout the course of any given research project. We consider confirmatory research to include activities such as: conceptual and direct replications, and extending known effects (e.g., identifying boundary conditions or moderators)."

Respondents then rated the same dimensions as for exploratory research. We counterbalanced the presentation order of the exploratory and confirmatory questions.

In choosing which dimensions to measure, we selected a balance of positively-valenced (enjoyable, motivating, and interesting) and negatively-valenced adjectives (frustrating, anxiety-inducing, boring) that would likely correlate with job satisfaction given their similarity to items used in two well-validated workplace motivation scales ([Gagné, Forest, Gilbert, Aubé, Morin, & Malorni, 2010; Gagné et al., 2015]). We included the term “scientific” to assess whether researchers viewed one type of research as more integral to the scientific process than the other. We report the results for each adjective separately, as well as for two composites (positive versus negatively-valenced adjectives).

To ensure that our definitions for exploratory versus confirmatory research fit with respondents’ perceptions of these concepts, we created our descriptions iteratively, pulling from previous literature and input from other behavioral researchers. Moreover, we conducted a validation study of 168 behavioral researchers (Appendix S1) to assess whether our target population found our descriptions to reasonably describe exploratory versus confirmatory research; results suggest that they did. Having an a priori sense of various aspects of the research—including how much data to collect, how those data would be analyzed, and what the result was likely to be—featured prominently into respondents’ sense of confirmatory research. By contrast, not having a specific prediction and conducting research in a novel area featured prominently into respondents’ sense of exploratory research. Our definitions were consistent with these defining features.

Our definitions are also consistent with scholarly writing that characterizes confirmatory research as “hypothesis-testing” ([de Groot, 2014], and research in which “the entire analysis plan has been explained before the first participant is tested” ([Wagenmakers et al., 2012]). Central to these definitions is the possession of an a priori hypothesis—as [Wagenmakers et al. (2012)] note, pre-specifying one’s analysis plan entails denoting the hypotheses of interest. Our definition of exploratory research is also consistent with how other scholars have characterized this construct; for example, [Wagenmakers et al. (2012)] describe exploratory work as particularly appropriate “in the first stage of a research program” wherein “researchers find their hypothesis in the data.”

**Most and Least Enjoyable Tasks.** Respondents described which research tasks they found most enjoyable and categorized the task as: (1) Exploratory research, (2) Confirmatory research, (3) Neither or (4) Both. Similarly, respondents described and categorized the task they found least enjoyable. Order was counterbalanced.

**Career Satisfaction.** Respondents answered three questions assessing their career satisfaction on a scale from 1 = \textit{Not at all} to 7 = \textit{Extremely}: (1) “Overall, how satisfied are you with the field of behavioral science?”, (2) “Overall, how satisfied are you with your current role in the field of behavioral science?”, and (3) “Overall, how interested are you in staying in the field of behavioral science (vs. finding a job elsewhere)?”

**Research Demographics.** We asked respondents several questions about their research: how they allocated their time across exploratory versus confirmatory research (answered on a scale from 0% exploratory to 100% exploratory, 0% confirmatory to 100% confirmatory; randomly assigned); and, as in Study 1, the number of studies they had run in the last 12 months, the percent of these studies that were pre-registered, and their primary research method.

**Personal Demographics.** As in Study 1, respondents reported: the year (actual or expected) of their PhD; their current role; their sub-discipline; and their gender. See Table 1 for full sample demographics.

Participants also completed an earlier, exploratory version of our prediction preoccupation scale (see SOM for items and results), which we later refined in Study 3.

### 3.2. Results

Here, we focus on the results of primary interest, namely those pertaining to the subjective experience of doing research, how that might differ as a function of exploratory versus confirmatory research, and how it might be related to satisfaction with behavioral science. Additional results are in the SOM.

#### 3.2.1. Subjective experience of exploration versus confirmation

Descriptive statistics and statistical significance tests are reported in Table 2. Relative to confirmatory research, respondents reported that exploratory research was more enjoyable, motivating, and interesting, \(p < 0.001\). Respondents also reported that, relative to confirmatory research, exploratory research was less frustrating, anxiety-inducing, boring, and scientific, \(ps < 0.03\). Respondents reported experiencing the positive adjectives to a greater extent, and the negative adjectives to a lesser extent, when engaging in exploratory research relative to confirmatory research (positive: \(M_{\text{exploratory vs confirmatory}} = 0.60, t(338) = 8.47, p < .001; \) negative: \(M_{\text{exploratory vs confirmatory}} = -0.56, t(342) = -7.86, p < .001\); Fig. 3). Consistent with our expectations, these items were also correlated with respondents’ satisfaction with their role and interest in staying in the field. See SOM for exploratory analyses that break down these results by demographic characteristics.

#### 3.2.2. Most and least enjoyable tasks

As depicted in Fig. 4, idea generation and data analysis were commonly viewed as the most enjoyable research tasks, while writing and the peer review process were commonly viewed as the least enjoyable research tasks. Most respondents (55%) categorized their most enjoyable task as both exploratory and confirmatory; 14% of respondents categorized this task as exploratory, 9% as confirmatory, and 22% as neither. Similarly, most respondents (58%) categorized their least
Table 2
Subjective Experience of Exploratory Versus Confirmatory Research.

<table>
<thead>
<tr>
<th></th>
<th>Exploratory Research</th>
<th>Confiratory Research</th>
<th>Mean Difference</th>
<th>t-test, Cohen’s d</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enjoyable</td>
<td>5.87 (1.04)</td>
<td>5.39 (1.10)</td>
<td>t(340) = 6.18, p &lt; .001, d = 0.33</td>
<td></td>
</tr>
<tr>
<td>Motivating</td>
<td>5.84 (1.10)</td>
<td>5.22 (1.27)</td>
<td>t(342) = 7.20, p &lt; .001, d = 0.39</td>
<td></td>
</tr>
<tr>
<td>Interesting</td>
<td>6.17 (0.99)</td>
<td>5.51 (1.14)</td>
<td>t(340) = 8.42, p &lt; .001, d = 0.46</td>
<td></td>
</tr>
<tr>
<td>Positive Composite</td>
<td>5.96 (0.91)</td>
<td>5.37 (1.01)</td>
<td>t(338) = 8.47, p &lt; .001, d = 0.46</td>
<td></td>
</tr>
<tr>
<td>Frustrating</td>
<td>3.57 (1.78)</td>
<td>3.81 (1.63)</td>
<td>t(342) = -2.40, p = .02, d = -0.13</td>
<td></td>
</tr>
<tr>
<td>Anxiety-Inducing</td>
<td>3.54 (1.88)</td>
<td>4.16 (1.75)</td>
<td>t(342) = -4.96, p &lt; .001, d = -0.27</td>
<td></td>
</tr>
<tr>
<td>Boring</td>
<td>2.18 (1.23)</td>
<td>2.98 (1.45)</td>
<td>t(342) = -8.32, p &lt; .001, d = -0.45</td>
<td></td>
</tr>
<tr>
<td>Negative Composite</td>
<td>3.10 (1.24)</td>
<td>3.65 (1.17)</td>
<td>t(342) = -7.86, p &lt; .001, d = -0.42</td>
<td></td>
</tr>
<tr>
<td>Scientific</td>
<td>5.83 (1.14)</td>
<td>6.08 (0.97)</td>
<td>t(341) = -3.82, p &lt; .001, d = -0.21</td>
<td></td>
</tr>
</tbody>
</table>

Fig. 3. Average Ratings of the Subjective Experience of Exploratory and Confirmatory Research. Bars represent means and error bars represent standard errors. * p < .05; ** p < .001.

Table 2: Subjective Experience of Exploratory Versus Confirmatory Research.

Fig. 4. Word clouds representing open-ended text responses describing respondents’ most and least enjoyed research tasks.
3.2.3. Career satisfaction

**Satisfaction with the field.** Average satisfaction with the field of behavioral science was significantly higher than the scale midpoint ($M = 4.76$, $SD = 1.21$), $t(354) = 11.82, p < .001$, Cohen’s $d = 0.63$, 95% CI $[0.51, 0.74]$.—68% reported satisfaction above the scale midpoint.

**Satisfaction with role.** Average satisfaction with one’s role was also significantly higher than the scale midpoint ($M = 4.75$, $SD = 1.42$), $t(354) = 9.96, p < .001$, Cohen’s $d = 0.53$, 95% CI $[0.42, 0.64]$—61% reported satisfaction above the scale midpoint. Role satisfaction was related to the subjective experience items. Specifically, the positive adjectives composite collapsed across research type (i.e., the exploration versus confirmation distinction) was positively related to role satisfaction ($r = 0.12, p = .03$, 95% CI $[0.01, 0.22]$). Similarly, the negative composite was negatively related to role satisfaction ($r = -0.24, p < .001$, 95% CI $[-0.34, -0.14]$).

**Interest in staying in field.** Average interest in staying in the field was well above the scale midpoint ($M = 5.87$, $SD = 1.34$), $t(354) = 26.39, p < .001$, Cohen’s $d = 1.40$, 95% CI $[1.25, 1.54]$—84% reported interest above the scale midpoint. Interest in staying in the field was related to the subjective experience items, with the positive composite being positively related ($r = 0.19, p < .001$, 95% CI $[0.08, 0.29]$), and the negative experiences being negatively related ($r = -0.12, p = .03$, 95% CI $[-0.22, -0.01]$).

3.3. Discussion

As Study 1 attests, emphasizing confirmation can shift researchers away from exploratory. In Study 2, researchers reported more positive and fewer negative experiences when conducting exploratory compared to confirmatory research. In turn, these positive experiences were linked to greater interest in staying in the field. Together, these results raise the possibility that a focus on confirmation could degrade the subjective experience of conducting research and reduce career satisfaction over time. However, additional research is needed to definitively make such a claim. For example, future work could assess additional sentiments beyond those we measured here or conduct experimental or longitudinal research to provide causal evidence for these ideas. In addition, it is worth noting that, fortunately, career satisfaction was fairly high.

Although we measured subjective experiences separately for confirmatory versus exploratory research tasks, researchers do not always engage in these activities separately. For example, a researcher may include a few exploratory variables in the context of a replication study. This co-occurrence could explain why participants categorized their most enjoyable task as both exploratory and confirmatory (as opposed to feelings associated specifically with either exploratory or confirmatory work).

We collapsed across the exploratory versus confirmatory distinction because we observed the same associations with career satisfaction within each type of research. In other words, it seems that it is positive and negative experiences of research in general that are correlated with satisfaction (as opposed to feelings associated specifically with either exploratory or confirmatory work).

We pre-registered collecting 250 responses but were able to collect 342 responses due to snowball sampling and rolling recruitment. Results are substantively equivalent when we restrict the sample to the first 250 responses.
and fixation with, confirming predictions. Because such concerns could manifest while conducting both exploratory and confirmatory research, we included items designed to capture this feeling within both contexts (items 1 to 4 focused on confirmatory contexts such as designing a pre-registered study; items 5 and 6 focused on exploratory contexts such as designing an exploratory study). Participants reported the extent to which they agreed or disagreed with each statement on a scale from 1 = Strongly Disagree to 7 = Strongly Agree.

**Predictive Validity Items.** To assess the scale’s predictive validity, participants indicated their anxiety associated with each of six common behavioral scientific research experiences on a scale from 1 = Not at all to 5 = Extremely. These items were administered prior to the prediction preoccupation scale. The first four experiences pertained to confirmatory research contexts. Participants read:

> “Please imagine the following: You have a hypothesis: X increases Y. You run a study to test this hypothesis. The results support the hypothesis. You think this finding contributes to the literature, so you plan to write a paper about it, and submit this paper for publication. As part of this process, you are going to run a direct replication of the study. Before running the direct replication, you pre-register the methods and predicted hypotheses.”

Participants reported how anxious they would feel: (1) “designing this direct replication study,” (2) “writing up the pre-registration,” (3) “running the pre-registered analyses to see if your results confirm your hypotheses,” and (4) running “additional, exploratory, analyses to see if you could ‘learn’ anything else from the data.” The next two experiences pertained to exploratory research contexts. Participants read:

> “Please imagine the following: You are interested in studying topic X. You don’t have any specific hypotheses. You decide to conduct an exploratory study in which you collect a dataset on topic X.”

Participants reported how anxious they would feel: (5) “designing this exploratory study,” and (6) “analyzing the data to explore possible effects of interest.”

To explore the breadth of our scale’s predictive validity, we also assessed whether prediction preoccupation negatively predicted excitement—because, according to the affective circumplex model of emotion (Russell, 1980), excitement can be thought of as anxiety’s positively-valenced counterpart (like anxiety, it is high in arousal). To mask the study’s purpose, participants also reported how “alert” and “inspired” they would feel for each of the six tasks.

To summarize: for each of the six tasks, respondents rated the extent to which they would feel: anxious, excited, alert and inspired.

**Convergent and Discriminant Validity.** We assessed whether prediction preoccupation was correlated with, but distinct from, two related constructs: general anxiety and aversion to negative evaluation. Anxiety was assessed with the 6-item Brief State Anxiety Inventory (Berg, Shapiro, Chambless, & Ahrens, 1998), which asked participants to report generally on their anxious feelings in daily life. Aversion to negative evaluation was assessed with the three highest loading items of the Brief Fear of Negative Evaluation scale (Duke, Krishnan, Faith, & Storch, 2006). We also assessed whether prediction preoccupation was uncorrelated with social desirability, a distinct construct outside of the scale’s nomological network (assessed by the 10-item Social Desirability Scale; Strahan & Gerbasi, 1972). Participants’ reports of being alert and inspired, although primarily used as distractors, also provided a test of discriminant validity—our scale should not predict these experiences.

**Career Satisfaction.** Respondents answered the same three items as in Study 2.

**Research Demographics.** Respondents indicated the extent to which conducting open science was part of their identity as a researcher (1 = Not at all; 5 = Very much so); and, as in Studies 1 and 2, the number of studies they had run in the last 12 months, the percent of these studies that were pre-registered, and their primary research method.

**Personal Demographics.** As in Studies 1 and 2, respondents indicated: the year (actual or expected) of their PhD; their current role; their sub-discipline; and their gender.

This study was pre-registered through AsPredicted here https://aspredicted.org/7sr7d.pdf.

4.2. Results

First, we assessed the psychometric properties of the prediction preoccupation scale. Next, we assessed how the scale correlated with researchers’ anxiety with various research experiences. Finally, we investigated who experienced prediction preoccupation and explored its correlates. The SOM contains additional exploratory results.

4.2.1. Psychometric properties

We calculated prediction preoccupation scores by computing the average of the six items, which showed acceptable reliability (α = 0.71; Tavakol & Dennick, 2011). Exploratory factor analysis revealed that a two-factor solution best fit the data (Root Mean Squared Error of Approximation (RMSEA) = 0.02; Standardized Root Mean Square Residual (SRMR) = 0.04; Hu & Bentler, 1999). As expected, items 1 to 4, which pertained to confirmatory contexts, loaded onto one factor, and items 5 and 6, which pertained to exploratory contexts, loaded onto a separate factor (Table 3). For simplicity, we report the predictive validity results using the single-factor solution because the results were substantively equivalent to the two-factor solution (see SOM).

4.2.2. Predictive validity

We tested our scale’s capacity to predict the anxiety associated with each of the six activities, and had two key pre-registered predictions. First, given the reform movement’s emphasis on confirmation, we thought that prediction preoccupation would be particularly likely to predict anxiety when running pre-registered analyses. This is the “moment of truth” in the sense that the researcher discovers whether they have accomplished what reforms emphasize: confirmation. Second, given that Study 1 suggested that pre-registration can impede exploration, we hypothesized that prediction preoccupation would predict anxiety when conducting exploratory analyses within a confirmatory context. Both of these pre-registered predictions were supported: prediction preoccupation was significantly positively associated with anxiety when running pre-registered analyses.
The scale was also positively associated with anxiety during the four other activities: designing a direct replication study ($b = 3.87, SD = 1.10$) compared with pre-2011 graduates ($M = 3.28, SD = 1.08$), $t(174.84) = 4.18, p < .001$; and in researchers without tenure ($M = 3.88, SD = 1.11$) than with tenure ($M = 3.24, SD = 1.07$), $t(169.42) = 4.31, p < .001$. Prediction preoccupation was similar between researchers who reported engaging in pre-registration ($M = 3.42, SD = 1.07$) and those who did not ($M = 2.54, SD = 1.18$), $t(255.7) = 0.84, p = .40$.

4.3. Discussion

Study 3 introduced a scale to measure the construct of prediction preoccupation. Attesting to predictive validity, individual scores on this scale were correlated with the extent to which researchers experienced anxiety with a variety of research activities. Attesting to convergent and discriminant validity, the scale was moderately associated with trait anxiety and fear of negative evaluation and unrelated to social desirability. Importantly, the scale predicted researcher anxiety above and beyond these constructs. Consistent with the findings of Study 2, prediction preoccupation was significantly negatively associated with researchers’ satisfaction with their current role. It was also more likely to affect women and more junior scholars.

Study 3 assessed prediction preoccupation, and its correlates, at a static point in time. Although the reform movement’s emphasis on confirmation spurred us to investigate this construct, our data do not speak to whether reforms induced such feelings. We leave it to future work to contextualize prediction preoccupation scores and test whether the observed relationships are causal. Research could also delve further into the scale’s predictive validity; for example, by conducting a prospective study that tracks prediction preoccupation and examines the long-term effects of higher scores on sustained career satisfaction and subsequent retention.

5. General discussion

We are proponents of the reform movement. However, we wondered whether reforms might have negative unintended consequences on researchers’ behavior and subjective experiences. With respect to behavior, we worried that reforms, with their (justified) emphasis on confirmation, could impede exploration—unfortunate, given that exploration followed by confirmation is integral to scientific discovery (Laudan 1968; Patterson, 2002). Relatedly, we wondered whether a focus on confirmation may impact the subjective experience of conducting research.

With these reflections as a motivating backdrop, we conducted three studies. In Study 1, researchers who were randomly assigned to pre-register a prediction were less likely to discover an interesting, non-hypothesized result. In Study 2, researchers reported more positive and fewer negative experiences when engaged in exploration versus confirmation. In Study 3, we developed and validated the prediction preoccupation scale and found significant links between scores on this scale and researchers’ experience of anxiety while conducting relevant research activities, such as when conducting exploratory analyses in the context of a pre-registered study. We note that our work does not speak

### Table 4

<table>
<thead>
<tr>
<th>Research Activity</th>
<th>Prediction Preoccupation</th>
<th>Trait Anxiety</th>
<th>Fear of Negative Evaluation</th>
<th>Social Desirability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Designing direct replication</td>
<td>$b = 0.43, SE = 0.06$</td>
<td>$b = 0.20, SE = 0.10$</td>
<td>$b = 0.14, SE = 0.05$</td>
<td>$b = 0.02, SE = 0.03$</td>
</tr>
<tr>
<td>Pre-registering confirmatory study</td>
<td>$b = 0.43, SE = 0.06$</td>
<td>$b = 0.11, SE = 0.11$</td>
<td>$b = 0.12, SE = 0.06$</td>
<td>$b = 0.01, SE = 0.03$</td>
</tr>
<tr>
<td>Running pre-registered analyses</td>
<td>$b = 0.50, SE = 0.06$</td>
<td>$b = 0.14, SE = 0.11$</td>
<td>$b = 0.16, SE = 0.06$</td>
<td>$b = 0.02, SE = 0.03$</td>
</tr>
<tr>
<td>Running additional exploratory analyses</td>
<td>$b = 0.35, SE = 0.06$</td>
<td>$b = 0.14, SE = 0.11$</td>
<td>$b = 0.001, SE = 0.06$</td>
<td>$b = 0.004, SE = 0.03$</td>
</tr>
<tr>
<td>Designing exploratory study</td>
<td>$b = 0.33, SE = 0.06$</td>
<td>$b = 0.17, SE = 0.10$</td>
<td>$b = 0.06, SE = 0.05$</td>
<td>$b = 0.03, SE = 0.03$</td>
</tr>
<tr>
<td>Running exploratory analyses</td>
<td>$b = 0.34, SE = 0.06$</td>
<td>$b = 0.30, SE = 0.11$</td>
<td>$b = 0.02, SE = 0.06$</td>
<td>$b = 0.02, SE = 0.03$</td>
</tr>
</tbody>
</table>

* $p < .05$  
* $p < .001$  
* Regression coefficient is significantly smaller compared to that of prediction preoccupation.
of career-relevant outcomes including research rigor, productivity, retention, fueling renewed commitment, e.g., Harmon-Jones, 2003; Lewis, motivating higher quality work (as when, for example, one researcher only discovered the main effect, they might simply conclude can at least sometimes undermine researchers engaging in exploratory or confirmatory research or what the ideal state of the field should be—additional research should further explore these questions. Study 1 provided initial evidence in support of the claim that reforms can at least sometimes undermine researchers’ willingness to engage in exploration. In Study 1, the non-discovery of the interaction was rather benign given that it could not lead to egregiously inaccurate conclusions. If, like most participants in the confirmation and hybrid conditions, a researcher only discovered the main effect, they might simply conclude that there was a positive relationship between yoga and happiness (and not that it was particularly strong among men). Conceivably, however, there are circumstances where non-discovery is more problematic; for example, a disordinal interaction indicating a positive relationship among men and a negative (though somewhat weaker) relationship among women. In such cases, non-discovery could prevent a researcher from placing appropriate boundary conditions on their conclusions. In certain cases, non-discovery could be especially harmful; for example, in drug trials where failing to detect an interaction masks an adverse effect in a subgroup.

Given the importance of exploration, future research should test the generalizability of the findings from Study 1 and probe the psychological processes that underlie it. For example, is the apparent inhibition of exploration driven by an attentional mechanism whereby confirmatory contexts divert researchers’ attention away from exploration? Or could flawed mental models be at work whereby researchers erroneously believe that exploration in the context of confirmation is, ipso facto, a questionable research practice? If so, has the reform movement (inadvertently) induced such inaccurate beliefs? Future work could investigate such possibilities.

With respect to subjective experiences, reforms may prompt behavioral scientists to derive less pleasure from their primary job task: conducting research. Consistent with this concern, our results suggest that emphasizing confirmation can shift researchers away from exploration (Study 1) and that such a shift may degrade the subjective experience of doing research. In Study 2, researchers reported more positive and fewer negative experiences when engaged in exploration versus confirmation. While these results cannot directly speak to researchers’ job outcomes, a large body of work would suggest a negative externality from such a shift away from exploration: lower job performance, creativity, motivation, satisfaction and retention (e.g., Amabile, Barsade, Mueller, & Staw, 2005; Batliss, 1980; Bellet, DeNeve & Ward, 2019; Newbury-Birch & Kamali, 2001; Oswald, Proto & Sgroi, 2015; Srivastava & Krishna, 1980; Whillans, Macchia & Dunn, 2019; Woolley & Fishbach, 2015, 2016). Future work should directly test these relationships.

Of course, we are not arguing for avoiding negative sentiment altogether. Research on “emodiversity” attests to the benefits of experiencing a variety of both positive and negative emotions (Quoidbach, Gruber, Mikolajczak, Kogan, Kotsou, & Norton, 2014). Undoubtedly, rigorous science can spur negative emotions, and such feelings may be useful in motivating higher quality work (as when, for example, one’s manuscript is rejected, fueling renewed commitment, e.g., Harmon-Jones, 2003; Lewis, Sullivan, Rumsey, & Alessandri, 1992). Future research should explore the emotional consequences of exploration and confirmation on a broader set of career-relevant outcomes including research rigor, productivity, retention, and long-term career satisfaction. Study 2 also indicated that researchers considered confirmation to be more scientific than exploration—a perspective we find worrisome, given that both activities are critical to the scientific enterprise (de Groot, 2014; Gutting, 1980; Rothchild, 2006). If researchers view exploration as less scientific than confirmation, might they be disinclined to partake in it? Are researchers who do quality exploratory work in danger of being sidelined?

To the extent that these possibilities are founded, institutional changes may be needed to address them. Journals could play a crucial role in legitimizing and incentivizing exploration. For example, journals could establish article categories reserved for exploratory work (that meet rigorous empirical standards such as robustness to overfitting) and continue to welcome smaller-scale exploratory research alongside large sample confirmatory work (Baumeister, 2016; Sakaluk, 2016), perhaps by offering incentives (Coffman & Niederle, 2015). Doctoral education could be reviewed to ensure that in addition to confirmatory research methods training, students also learn how to conduct exploratory research in a rigorous way (e.g., by using the tools of machine learning). Open science platforms could also contribute by promoting exploration even within the context of confirmation. Many platforms feature pre-registration forms, some of which include explicit sections for pre-registering exploratory analyses. Research could test whether such sections spur or prevent exploration. Such elements may spur exploration by serving as a reminder, or even pre-commitment, to explore. They may also, ironically, inhibit exploration, if they lead researchers to feel as though they cannot conduct additional exploratory analyses that were not pre-registered. These possibilities underscore the need for testing whether institutional changes achieve their intended effects.

In Study 3, over 30 percent of behavioral scientists scored above the midpoint on the prediction preoccupation scale, suggesting that many researchers suffer from anxiety associated with conducting research—particularly when making and confirming predictions. Whether the reform movement has increased such feelings—especially among researchers who received their PhD after 2011, women, and untenured professors—remains an important open question. These findings are consistent with previous research showing that experiencing ruminations and negative emotion at work are linked to reduced job satisfaction (e.g., De Neve, Krekel, & Ward, 2018). Further, this research extends prior work on workplace rumination by showing that individual differences related to specific experiences in the context of daily work can also significantly shape job satisfaction. However, this research is only the first step in understanding prediction preoccupation.

Future research could investigate the effects of prediction preoccupation for other outcomes, such as creativity—because rumination has been linked to reduced innovation (Vahle-Hinz, Mauno, de Bloom, & Kinnunen, 2017). Researchers could also explore the potential long-term health consequences of prediction preoccupation. Consistent with prior research on workplace rumination (Firoozabadi, Uitdewilligen, & Zijlstra, 2018), researchers who experience greater prediction preoccupation—women, untenured professors and those who received their PhD after 2011—may experience persistent decreases in psychological and physical well-being. Finally, considering that ambiguity is a critical predictor of rumination at work (Jackson & Schuler, 1985; Sonnentag, & Krue, 2006), future research could also investigate whether clarity in methodological reforms might alleviate prediction preoccupation—particularly among women and junior scholars who may be less confident and experience greater anxiety in response to ambiguity (e.g., Bowles, Babcock & McGinn, 2005).

We performed this research partly to better understand the subjective experience of conducting research. We wondered: how do behavioral scientists experience exploration and confirmation amidst methodological reforms? Our results suggest that some researchers hesitate to explore in the context of confirmation; some experience exploration more positively than confirmation despite believing that confirmation is more scientific; and some experience heightened anxiety about making and confirming predictions. Broadly, our results point to the important interplay between exploration and confirmation. Like exploration, confirmation is integral to the research process, yet, more so than exploration, it seems to spur negative sentiment. We suggest that although both exploration and confirmation are essential to rigorous scientific research, in practice, confirmation may preclude exploration, and hence, rigor might come at the expense of joy. Because, as it turns out, Asimov was right: noticing ‘funny’ things is fun! We hope that this investigation reminds us of the vital, and mutually-reinforcing, functions of confirmatory and exploratory research. Just as confirmation and exploration ought to co-exist, so, too, can joy and rigor.
Appendix A. Prediction preoccupation scale

Thinking about your feelings toward research overall, please rate the extent to which you agree or disagree with each of the following statements:

<table>
<thead>
<tr>
<th>Strongly disagree</th>
<th>Disagree</th>
<th>Somewhat disagree</th>
<th>Neither agree nor disagree</th>
<th>Somewhat agree</th>
<th>Agree</th>
<th>Strongly agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
</tr>
</tbody>
</table>

1. I feel nervous running replication studies.
2. I’d feel anxious pre-registering a study when I don’t have a strong sense of what the result will be.
3. I sometimes feel stuck. I worry that I can’t run a study unless I know what the result will be, and I don’t know what the result will be until I run a study.
4. I feel stress when the results of a study do not confirm my predictions.
5. I feel apprehensive exploring data without a specific hypothesis.
6. I feel uncomfortable running a study without a strong prior (i.e., simply to “see what happens”).

Appendix B. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.obhdp.2021.03.002.

References

Coffman, L. C., & Niederle, M. (2015). Pre-analysis plans have limited upside, especially participants in Studies 3 and S1 which were conducted during the global COVID-19 pandemic.
Kaplan, R. M., & Irvin, V. L. (2015). Likelihood of null effects of large NHLBI clinical trials has increased over time. PloS one, 10(8), Article e0132382.