

# Joy and Rigor in Behavioral Science

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

Working Paper 20-090



# Joy and Rigor in Behavioral Science

Hanne K. Collins  
Harvard Business School

Ashley V. Whillans  
Harvard Business School

Leslie K. John  
Harvard Business School

**Working Paper 20-090**

Copyright © 2020, 2021 Hanne K. Collins, Ashley V. Whillans, and Leslie K. John..

Working papers are in draft form. This working paper is distributed for purposes of comment and discussion only. It may not be reproduced without permission of the copyright holder. Copies of working papers are available from the author.

Funding for this research was provided in part by Harvard Business School.

## Joy and Rigor in Behavioral Science

Hanne K. Collins<sup>1</sup>

Ashley V. Whillans<sup>1</sup>

Leslie K. John<sup>1</sup>

<sup>1</sup>Harvard Business School

**Corresponding Author:**

Leslie K. John  
Harvard Business School  
ljohn@hbs.edu

**Acknowledgements:**

We would like to thank Shannon Sciarappa for assistance in conducting this research. We also thank our respondents – peer behavioral scientists who took the time to participate – especially participants in Studies 3 and S1 which were conducted during the global COVID-19 pandemic.

## Abstract

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 on 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 ( $N=404$  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.

*Keywords:* Open Science; Pre-Registration; Exploration; Confirmation; False Positives; Career Satisfaction; Diversity

“The most exciting phrase to hear in science, the one that heralds new discoveries, is not ‘Eureka!’ but ‘That’s funny...’ ” – Isaac Asimov

Exploration is the wellspring of discovery. One might say it is fueled by a kind of open-mindedness, or even a sense of excitement, about the unexpected. Though, to be sure, moments of finding the expected are no less crucial to the scientific enterprise. Indeed, exploration and confirmation are mutually reinforcing (e.g., de Groot, 2014; Gutting, 1980; Rothchild, 2006)—their very interplay is the hallmark of empiricism (Laudan, 1968; Patterson, 2002). Accordingly, philosophers of science have long underscored two “directions” of scientific inquiry: one moving from observation to general principles; the other from general principles to specific observations—commonly referred to as the inductive and deductive method. This philosophical distinction is apparent in how behavioral scientists conduct research—in exploratory research, they seek to learn from specific observations and generate theory; in confirmatory research, they seek to test those theories with specific observations.

However, about ten years ago, behavioral scientists became concerned over an imbalance in these two activities. Scholars warned of the prevalence of false positive results (Simmons, Nelson, & Simonsohn, 2011) and pointed to exploration *absent* adequate confirmation, and exploration *described* as confirmation, as likely culprits (John, Loewenstein, & Prelec, 2012; Wagenmakers, Wetzels, Borsboom, vander Maas, & Kievit, 2012). 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; Crandall & 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, 1998).

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 true positives)—plausible, given other work pointing to 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, Shrirra, & 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.

## **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 discussed its correlates (Study 3). All data and materials are available through the Open Science Framework [here](#).

### **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.

## Method

### *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),<sup>1</sup> 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;<sup>2</sup> 14% unspecified).<sup>3</sup> 79% of participants completed all primary outcome measures.

### *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*;

---

<sup>1</sup> 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.

<sup>2</sup> 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.

<sup>3</sup> 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.

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 relative to those who 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 screen shot 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 (Figure 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 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).

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 (Figure 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!”

**Figure 1**

*Pre-registration simulation presented to participants in the confirmation and hybrid conditions*

<p>You enter the following pre-registration information into the online pre-registration tool:</p> <p style="text-align: center;"><b>Pre-Registration for Study: “Yoga &amp; Happiness”</b></p> <p><b>1) What’s the main question being asked or hypothesis being tested in this study?</b> We hypothesize that people who do yoga on a weekly basis are happier than people who do not do yoga on a weekly basis.</p> <p><b>2) Describe the key dependent variable(s) specifying how they will be measured.</b> Our key dependent variable is self-reported happiness on a scale from 1=Not at all to 7=Extremely.</p> <p><b>3) How many and which conditions will participants be assigned to?</b> N/A</p> <p><b>4) Specify exactly which analyses you will conduct to examine the main question/hypothesis.</b> To examine our main hypothesis, we will conduct a t-test investigating whether people who do yoga on a weekly basis report higher levels of happiness than those who do not do yoga on a weekly basis (0=no; 1=yes).</p> <p><b>5) How many observations will be collected or what will determine sample size?</b> We will survey 1,000 people.</p> <p><b>6) Anything else you would like to pre-register? (e.g., data exclusions, variables collected for exploratory purposes, unusual analyses planned?)</b> N/A</p> <p><b>7) Have any data been collected for this study already?</b> No, no data have been collected for this study yet</p> <p style="color: red; font-weight: bold;">Please click on the pre-registration icon below to pre-register this study!</p> <div style="text-align: center; margin: 10px 0;">  </div> <div style="text-align: right; margin-top: 20px;">  </div>	<p style="text-align: center;"><b>Pre-Registration for Study: “Yoga &amp; Happiness”</b></p> <p><b>1) What’s the main question being asked or hypothesis being tested in this study?</b> We hypothesize that people who do yoga on a weekly basis are happier than people who do not do yoga on a weekly basis.</p> <p><b>2) Describe the key dependent variable(s) specifying how they will be measured.</b> Our key dependent variable is self-reported happiness on a scale from 1=Not at all to 7=Extremely.</p> <p><b>3) How many and which conditions will participants be assigned to?</b> N/A</p> <p><b>4) Specify exactly which analyses you will conduct to examine the main question/hypothesis.</b> To examine our main hypothesis, we will conduct a t-test investigating whether people who do yoga on a weekly basis report higher levels of happiness than those who do not do yoga on a weekly basis (0=no; 1=yes).</p> <p><b>5) How many observations will be collected or what will determine sample size?</b> We will survey 1,000 people.</p> <p><b>6) Anything else you would like to pre-register? (e.g., data exclusions, variables collected for exploratory purposes, unusual analyses planned?)</b> N/A</p> <p><b>7) Have any data been collected for this study already?</b> No, no data have been collected for this study yet</p> <hr/> <p>Pre-registration confirmed. The study is now pre-registered. Now it's time to collect the data.</p> <p>Click '-&gt;' to collect the data!</p> <div style="text-align: right; margin-top: 20px;">  </div>
---	--

## Measures

**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.

**Selecting Analyses to Report.** On the next page (i.e., after the page on which they selected which analyses to run), we presented the output of participants’ requested analyses in their preferred format (SPSS or R).<sup>4</sup> The results supported the hypothesis; the t-test of happiness as a function of yoga status revealed that those who did yoga reported significantly greater happiness than those who did not. However, participants who opted to view the 2x2 ANOVA also learned that this main effect was qualified by a statistically significant interaction ( $p < .001$ ): the effect of yoga on happiness depended on gender; it was more pronounced for men. On a separate screen, participants were then asked to choose the results that they would like to report in a final manuscript (see Figure 2).

**Research Demographics.** Participants reported 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 (lab experiments, field experiments, non-experimental field data, non-experimental survey research, modelling, qualitative research, other).

**Personal Demographics.** Participants also reported various demographic characteristics, including year of PhD (actual or expected), role (graduate student, post-doctoral student, assistant professor, associate professor-untentured, associate professor-tenured, full professor, or other), sub-discipline (consumer behavior, social psychology, cognitive psychology, personality psychology, behavioral economics, experimental economics, micro-organizational behavior, macro-organizational behavior, other), and their gender.

---

<sup>4</sup> Participants who had selected “Other: Describe” were not shown the results of their requested analyses (as they were for the other response options).

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.

This study was pre-registered through AsPredicted [here](#).

**Figure 2**  
*Screenshot of the analysis selection and viewing procedure*

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.

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; DV=Happiness

Other. Describe:

Below are the results of the analyses you ran on the data.

T-test: IV = Yoga; DV = Happiness

```
> t.test(happy ~ yoga, data=dtstudy_load)
Welch Two Sample t-test

data: happy by yoga
t = -22.237, df = 438.84, p-value < 2.2e-16
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -1.585191 -1.268730
sample estimates:
mean in group no mean in group yes
 2.874439      4.257400
```

T-test: IV = Gender; DV = Happiness

```
> t.test(happy ~ male, data=dtstudy_load)
Welch Two Sample t-test

data: happy by male
t = -11.193, df = 995.28, p-value < 2.2e-16
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.8332992 -0.5846942
sample estimates:
mean in group female mean in group male
 3.551253      4.268258
```

Which results would you include in your final write up of this study? Select all that apply.

We find that people who report doing yoga weekly report greater happiness.

We find that men report greater happiness than women.

*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.

## Results

We tested whether, relative to those in the exploratory condition, participants in the confirmation condition viewed fewer analyses overall and were less likely to view and report the results of the gender interaction. We also tested whether the reminder to explore might mitigate such effects.

### ***Selecting Analyses to Run***

Across conditions, participants viewed an average of 3.57 (out of 6) analyses. A negative binomial logistic regression (Hilbe, 2011) revealed no difference between conditions in the number of analyses participants viewed ( $M_{\text{exploration}} = 3.48$ ,  $SD_{\text{exploration}} = 2.08$ ;  $M_{\text{confirmation}} = 3.79$ ,  $SD_{\text{confirmation}} = 1.99$ ;  $M_{\text{hybrid}} = 3.67$ ,  $SD_{\text{hybrid}} = 2.19$ ; all  $ps \geq 0.45$ ). Of particular interest, we assessed between condition differences in the propensity to view the results of an exploratory interaction using binary logistic regressions. In the confirmation condition, 53% of participants viewed the results of the interaction compared with 69% in the exploration condition,  $b = 0.70$ ,  $SE = 0.24$ ,  $p = .01$ . This translates to an odds ratio of 2.01, indicating that researchers were twice as likely to view the interaction results in the exploration rather than confirmation condition. The exploration reminder did not seem to mitigate this effect: 57% of participants in the hybrid condition requested the interaction—this percentage was statistically equivalent to the confirmatory condition,  $b = -0.19$ ,  $SE = 0.23$ ,  $p = .70$  (0.83 times as likely to view interaction in the confirmation condition), and marginally lower than the exploratory condition,  $b = 0.51$ ,  $SE = 0.24$ ,  $p = .09$  (1.67 times more likely to view interaction in the exploration condition).

### ***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,  $b = 0.73$ ,  $SE = 0.24$ ,  $p = .006$ , and marginally more so than the hybrid condition,  $b = 0.54$ ,  $SE = 0.24$ ,  $p = .06$ . Unsurprisingly, this interaction 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%;  $ps \geq 0.53$ ).

### ***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.

### **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 “Remember: You can explore your data as long as you clearly report post-hoc analyses.” It is also possible that regardless of the wording, explicit reminders to explore may be ineffective. As we discuss in the General Discussion section, stronger, institutional interventions may be needed.

Study 1 focused on the *act* of exploration; future research could explore whether reforms such as pre-registration could make researchers more reticent to *report* exploratory results (because exploratory results are more likely to be false positives than confirmatory results). Study 1 hints that this is not the case because nearly all participants who viewed the interaction indicated that they would report it. Further, the potential for a false positive should not affect whether researchers run an analysis in the first place—rather, it concerns how they interpret and report results. Nonetheless, future work could investigate how pre-registration shapes reporting; for example, by randomizing whether participants engage in pre-registration as we did here, but then exposing all participants to the new discovery and observing how cautiously they report the interaction when asked to do so. Such research could also test whether reminding researchers that they can report results as exploratory, may promote reporting.

**Table 1****Sample Demographics**

	Study 1 (N = 449)	Study 2 (N = 400)	Study 3 (N = 314)
<b>PhD Year (Actual or Expected)</b>			
	<i>M</i> = 2011 ( <i>SD</i> = 15), Median = 2018	<i>M</i> = 2015 ( <i>SD</i> = 10), Median = 2020	<i>M</i> = 2004 ( <i>SD</i> = 15), Median = 2006
<b>Use Experimental Methods</b>			
Yes	86%	94%	84%
No	14%	6%	16%
<b>Gender</b>			
Man	42%	50%	52%
Woman	56%	50%	47%
Non-Binary	2%	0%	<1%
<b>Role</b>			
Grad Student	47%	52%	12%
Post-Doc	1%	11%	6%
Assistant Prof	14%	13%	13%
Associate Prof, Untenured	3%	3%	3%
Associate Prof, Tenured	14%	9%	23%
Full Professor	14%	12%	33%
Other	6%	0%	11%
<b>Pre-Registration</b>			
Yes	54%	60%	52%
No	46%	40%	48%
<b>Sub-Discipline</b>			
Cognitive Psychology	9%	25%	19%
Consumer Behavior	20%	18%	8%
Social Psychology	41%	12%	18%
Behavioral Economics	<1%	9%	2%
Micro-Organizational Behavior	4%	5%	5%
Experimental Economics	0%	2%	<1%
Personality Psychology	4%	1%	2%
Macro-Organizational Behavior	<1%	<1%	<1%
Interdisciplinary	NA <sup>5</sup>	16%	11%
Other	21%	11%	35%
<b>Time Allocation (%)</b>			
Exploratory (vs. Confirmatory) Research	N/A	<i>M</i> = 49% ( <i>SD</i> = 23%), Median = 50%	N/A

<sup>5</sup> The option to select multiple sub-disciplines was not provided in Study 1.

## **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.

### **Methods**

#### ***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-Methods 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-Methods 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.

#### ***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 in detail 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 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=*Strongly Disagree* to 7=*Strongly Agree*. Item order was randomized between-subjects.

Respondents also read a description of confirmatory research (derived from the identical 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 reported 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 on the topic 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 explicated 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=*Not at all* to 7=*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, or 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.

**Research & 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.

## 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. We refer the interested reader to the SOM for additional results, and to the dataset posted online should they wish to explore it themselves.

### *Subjective Experience of Exploration versus Confirmation*

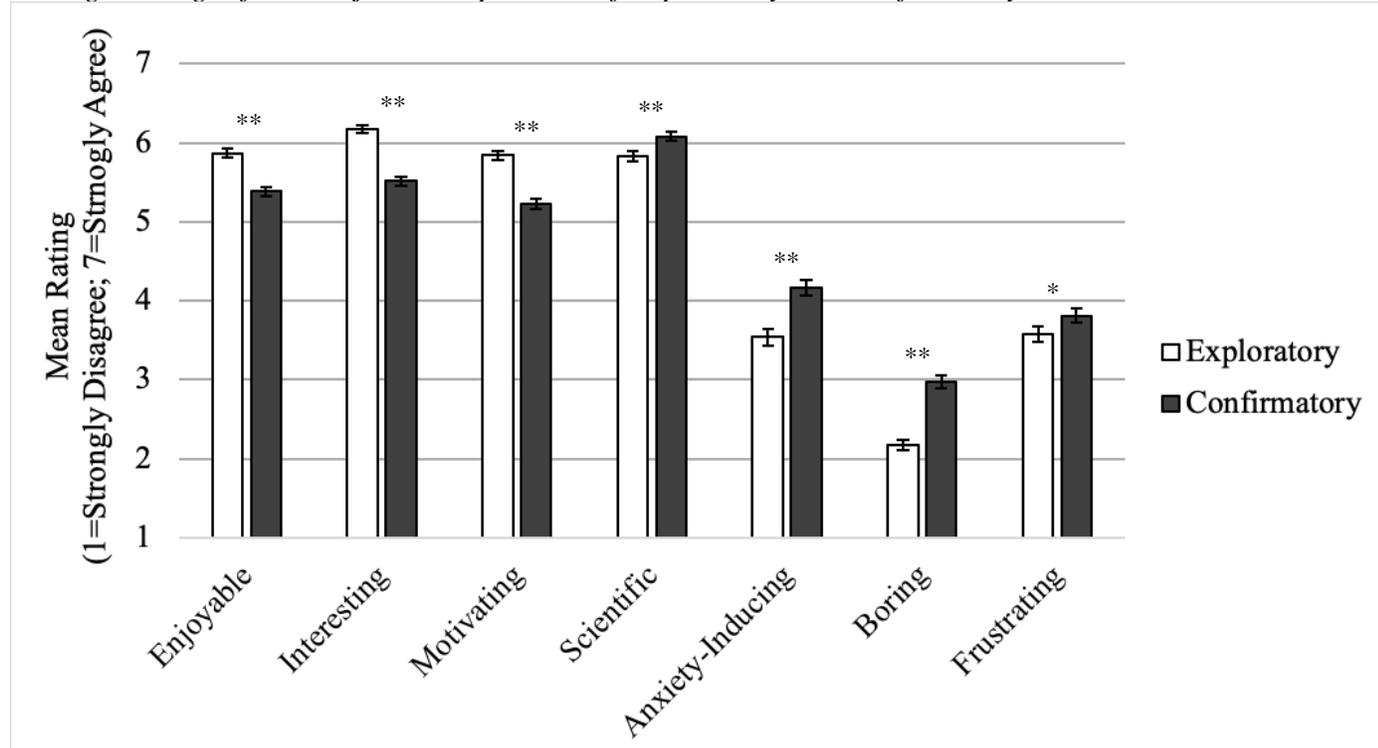
Descriptive statistics and statistical significance tests are reported in Table 2. Relative to confirmatory research, respondents reported that exploratory research was more enjoyable, motivating, and interesting,  $ps < .001$ . Respondents also reported that, relative to confirmatory research, exploratory research was less frustrating, anxiety-inducing, boring, and scientific,  $ps \leq .03$ . 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:  $M_{\text{diff\_exploratory\_vs\_confirmatory}} = 0.60$ ,  $t(338) = 8.47$ ,  $p < .001$ ; negative:  $M_{\text{diff\_exploratory\_vs\_confirmatory}} = -0.56$ ,  $t(342) = -7.86$ ,  $p < .001$ ; Figure 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.

**Table 2**  
***Subjective Experience of Exploratory Versus Confirmatory Research***

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

**Figure 3**

*Average Ratings of the Subjective Experience of Exploratory and Confirmatory Research*



*Note.* Bars represent means and error bars represent standard errors.

\*  $p < .05$ ; \*\*  $p < .001$

### ***Most and Least Enjoyable Tasks***

As depicted in Figure 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 enjoyable task as both exploratory and confirmatory; 14% categorized this task as exploratory, 10% as confirmatory, and 18% as neither.



**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 ( $b = 0.19$ ,  $SE = 0.10$ ,  $p < .001$ ), and the negative experiences being negatively related ( $b = -0.12$ ,  $SE = 0.07$ ,  $p = .03$ ).

## **Discussion**

As Study 1 attests, emphasizing confirmation can shift researchers away from exploration. In Study 2, researchers reported more positive and fewer negative experiences when conducting exploratory research. In turn, these positive experiences were linked to greater interest in staying 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 and least enjoyable tasks as both exploratory and confirmatory—many research tasks such as data analysis contain elements of both. Alternatively, this finding could simply be an artifact of how these

tasks were elicited—the item noted that “researchers often toggle back and forth between confirmatory and exploratory research throughout the course of any given research project.”

### **Study 3: Prediction Preoccupation**

Building on Study 2, Study 3 explored the novel construct of “prediction preoccupation”—the extent to which researchers feel heightened concern over, and fixation with, confirming predictions. As noted above, Studies 1 and 2 piloted initial versions of this scale (see SOM for psychometric properties of these earlier versions). Here, we present a refined version of the scale, discuss its psychometric properties, and offer initial evidence of its validity. Consistent with best practices for scale development (Clark & Watson, 1995), we assessed convergent and discriminant validity by measuring whether the scale is related to similar constructs and unrelated to dissimilar constructs; and predictive validity by assessing whether the scale predicts researchers’ felt anxiety over various aspects of the research process, especially those explicitly entailing prediction (and whether it does so even after controlling for conceptually similar measures).

According to our theorizing, prediction preoccupation should be associated with anxiety over various aspects of the research process, including those explicitly related to making and confirming predictions. It should predict anxiety over pre-registration activities; in particular, while conducting analyses to see if one’s predictions are supported. It may also predict anxiety over conducting exploratory analyses, particularly when these analyses are completed in the context of confirmatory research. Further, if a researcher is fixated on confirming predictions, it may feel “wrong” to conduct analyses outside of these predictions, or even to conduct

exploratory research at all; thus, we test whether prediction preoccupation is associated with anxiety with these activities as well.

This scale is rooted in, and builds on, the growing literature on workplace rumination (Cropley & Zijlstra, 2011; Cropley, Michalianou, Pravettoni, & Millward, 2012; Vahle, Mauno, deBloom, & Kinnunen, 2017). This research has found that work-related rumination—often referred to as “mental preoccupation with work” (Eib, Bernhard-Oettel, Magnusson Hanson, & Leineweber, 2018; Siegrist, 1996; von Thiele Schwarz, 2011)—is linked to emotional exhaustion and lower work engagement (Sonnentag, Binnewies, & Mojza, 2010). Job-anxiety has also been reliably associated with reduced workplace satisfaction (e.g., Newbury-Birch & Kamali, 2001), job performance (e.g., Srivastava & Krishna, 1980), and retention (e.g., Batlis, 1980). Given that prediction is a core element of a behavioral researcher’s job (especially in light of reforms), it is possible that prediction preoccupation could shape researchers’ emotional experiences at work. Thus, we sought to explore correlates of prediction preoccupation including researcher’s career dissatisfaction.

## **Method**

### ***Sample***

We recruited behavioral scientists to complete a brief survey about their subjective experiences with research in exchange for a \$5.00 donation to COVID-19 relief—participants could direct their donation to one of three relief programs.<sup>7</sup> Participants were recruited via: 1) an email sent to academic psychologists at major U.S. universities ( $n = 5,186$ , from an updated version of the list used in John et al., 2012, contacting only those not contacted in Study S1); 2) the SJDM list-serv; and 3) snowball sampling (our survey concluded with a request to forward it

---

<sup>7</sup> See Appendix B for donation receipts.

to colleagues, and offered an additional \$1,000 donation to the most popular relief program if we obtained at least 100 respondents). These efforts garnered 342 respondents.<sup>8</sup> We analyzed responses from the 314 (92% of) participants who reported conducting behavioral research. We used all available responses from the 250 respondents who did not complete the entire survey. 84% of participants completed all primary outcome measures.

### ***Measures***

**Prediction Preoccupation Scale.** The scale consisted of six items (Table 3; Appendix A) designed to capture researchers' preoccupation with prediction—the extent to which they feel heightened concern over, 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

---

<sup>8</sup> 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.

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-valanced 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. General anxiety was assessed with the 6-item Brief State Anxiety Inventory

(Berg, Shapiro, Chambless, & Ahrens, 1998). 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](#).

## 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. We refer interested readers to the SOM for additional exploratory results.

### *Psychometric Properties*

We calculated prediction preoccupation scores by computing the average of the six items, which showed acceptable reliability ( $\alpha=0.71$ ; 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, 1990). 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).

**Table 3**  
***Prediction Preoccupation Scale Items and Factor Loadings***

Prediction Preoccupation Scale	Factor 1	Factor 2
1. I feel nervous running replication studies.	0.72	
2. I'd feel anxious pre-registering a study when I don't have a strong sense of what the result will be.	0.49	
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.	0.46	
4. I feel stress when the results of a study do not confirm my predictions.	0.70	
5. I feel apprehensive exploring data without a specific hypothesis.		0.86
6. I feel uncomfortable running a study without a strong prior (i.e., simply to "see what happens").		0.71

*Note.* Factor loadings less than 0.30 are not included in the table (Field, 2013)

### ***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 ( $b = 0.62$ ,  $SE = 0.06$ ,  $p < .001$ , pre-registered one-tailed test), and with running additional exploratory analyses on direct replication data ( $b = 0.38$ ,  $SE = 0.06$ ,  $p < .001$ , pre-registered one-tailed test).

The scale was also positively associated with anxiety during the four other activities: designing a direct replication study ( $b = 0.54$ ,  $SE = 0.05$ ,  $p < .001$ , pre-registered two-tailed test); writing up a pre-registration for a direct replication study ( $b = 0.52$ ,  $SE = 0.06$ ,  $p < .001$ , pre-registered two-tailed test); designing an exploratory study ( $b = 0.41$ ,  $SE = 0.05$ ,  $p < .001$ , pre-registered two-tailed test); and analyzing exploratory data for possible effects of interest ( $b = 0.41$ ,  $SE = 0.06$ ,  $p < .001$ , pre-registered two-tailed test). The scale was not associated with excitement during any of the activities (designing replication study:  $b = -0.006$ ,  $SE = 0.06$ ,  $p = .91$ ; writing pre-registration:  $b = 0.03$ ,  $SE = 0.06$ ,  $p = .67$ ; running pre-registered analyses:  $b = 0.03$ ,  $SE = 0.06$ ,  $p = .60$ ; running exploratory analyses on replication data:  $b = -0.02$ ,  $SE = 0.05$ ,  $p = .73$ ; designing exploratory study:  $b = 0.03$ ,  $SE = 0.05$ ,  $p = .55$ ; running exploratory analyses:  $b = 0.03$ ,  $SE = 0.05$ ,  $p = .54$ , pre-registered two-tailed tests). These results suggest that the predictive ability of the prediction preoccupation scale does not extend to feelings of excitement.

### ***Convergent and Discriminant Validity***

Attesting to convergent validity, the scale was moderately correlated with the related constructs of trait anxiety ( $r = 0.33$ ,  $p < .001$ , 95% CI [0.22, 0.44]) and fear of negative evaluation ( $r = 0.41$ ,  $p < .001$ , 95% CI [0.30, 0.51]). Attesting to discriminant validity, the scale was not correlated with the unrelated construct of social desirability ( $r = -0.02$ ,  $p = .79$ , 95% CI [-

0.14, 0.10]). As expected, the scale was not associated with feeling alert or inspired (see SOM for analyses). Most importantly, prediction preoccupation was significantly associated with anxiety during all six research activities, even when simultaneously controlling for trait anxiety, fear of negative evaluation, and social desirability (Table 4). Moreover, for every research activity, the regression coefficients for prediction preoccupation were significantly larger than those for fear of negative evaluation and social desirability. They were also significantly larger than those for trait anxiety for the majority of research activities (Table 4, see subscripts).

**Table 4**  
*Associations Between Prediction Preoccupation and Other Variables*

Research Activity	Prediction Preoccupation	Trait Anxiety	Fear of Negative Evaluation	Social Desirability
Designing direct replication	$b = 0.43,$ $SE = 0.06^{**}$	$b = 0.20,$ $SE = 0.10^*$	$b = 0.14,$ $SE = 0.05^*_a$	$b = 0.02,$ $SE = 0.03_a$
Pre-registering confirmatory study	$b = 0.43,$ $SE = 0.06^{**}$	$b = 0.11,$ $SE = 0.11_a$	$b = 0.12,$ $SE = 0.06^*_a$	$b = -0.01,$ $SE = 0.03_a$
Running pre-registered analyses	$b = 0.50,$ $SE = 0.06^{**}$	$b = 0.14,$ $SE = 0.11_a$	$b = 0.16,$ $SE = 0.06^*_a$	$b = -0.02,$ $SE = 0.03_a$
Running additional exploratory analyses	$b = 0.35,$ $SE = 0.06^{**}$	$b = 0.14,$ $SE = 0.11$	$b = -0.001,$ $SE = 0.06_a$	$b = 0.004,$ $SE = 0.03_a$
Designing exploratory study	$b = 0.33,$ $SE = 0.06^{**}$	$b = 0.17,$ $SE = 0.10$	$b = 0.08,$ $SE = 0.05_a$	$b = 0.03,$ $SE = 0.03_a$
Running exploratory analyses	$b = 0.34,$ $SE = 0.06^{**}$	$b = 0.30,$ $SE = 0.11^*$	$b = 0.02,$ $SE = 0.06_a$	$b = 0.02,$ $SE = 0.03_a$

\* $p < .05$ ; \*\* $p < .001$

<sub>a</sub> Regression coefficient is significantly smaller compared to that of prediction preoccupation.

### *Relationship to Career Satisfaction*

Given the relatively low correlations between each outcome measure ( $r_s \leq 0.49$ ), we analyzed the results for each outcome separately, and on an exploratory basis. Prediction preoccupation was negatively associated with researchers' satisfaction with their role,  $b = -0.22$ ,  $SE = 0.07$ ,  $p < .001$ . Prediction preoccupation was not significantly related to researchers'

satisfaction with the field ( $b = 0.06$ ,  $SE = 0.07$ ,  $p = .40$ ) or their interest in staying in the field ( $b = -0.05$ ,  $SE = 0.07$ ,  $p = .51$ ).

### ***Who Experiences Prediction Preoccupation?***

Thirty percent of respondents scored above the scale midpoint, suggesting that one third of respondents experienced prediction preoccupation. Prediction preoccupation was higher: in females ( $M = 3.20$ ,  $SD = 1.02$ ) than males ( $M = 3.82$ ,  $SD = 1.14$ ),  $t(245.66) = -4.54$ ,  $p < .001$ ; in those who received their PhDs after 2011 when reforms began ( $M = 3.87$ ,  $SD = 1.10$ ) 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 ( $M = 2.54$ ,  $SD = 1.18$ ),  $t(255.7) = 0.84$ ,  $p = .40$ .

### **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.

### **General Discussion**

We are proponents of the reform movement. Yet 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 rigorous 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 to normative claims about how much time researchers should spend engaging in exploratory or confirmatory research process or what the ideal state of the field should be—additional research should further explore these critical 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 crossover 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 a significant 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; Batlis, 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 all together. 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, Ramsay, & 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 mid-point 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 rumination 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, & Krueger, 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 & McGuinn, 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.

## References

- Amabile, T. M., Barsade, S. G., Mueller, J. S., & Staw, B. M. (2005). Affect and creativity at work. *Administrative science quarterly*, 50(3), 367-403.
- Batlis, N. C. (1980). The effect of organizational climate on job satisfaction, anxiety, and propensity to leave. *The Journal of Psychology*, 104(3-4), 233-240.
- Baumeister, R. F. (2016). Charting the future of social psychology on stormy seas: Winners, losers, and recommendations. *Journal of Experimental Social Psychology*, 66, 153-158.
- Bowles, H. R., Babcock, L., & McGinn, K. L. (2005). Constraints and triggers: situational mechanics of gender in negotiation. *Journal of personality and social psychology*, 89(6), 951.
- Bellet, C., De Neve, J. E., & Ward, G. (2019). Does employee happiness have an impact on productivity? *Saïd Business School WP*, 13.
- Berg, C. Z., Shapiro, N., Chambless, D. L., & Ahrens, A. H. (1998). Are emotions frightening? II: An analogue study of fear of emotion, interpersonal conflict, and panic onset. *Behaviour Research and Therapy*, 36(1), 3-15.
- Camerer, C. F., Dreber, A., Forsell, E., Ho, T. H., Huber, J., Johannesson, M., ... & Heikensten, E. (2016). Evaluating replicability of laboratory experiments in economics. *Science*, 351(6280), 1433-1436.
- Chambers, C. (2019). What's next for Registered Reports? *Nature*, 573(7773), 187.
- Charmaz, K. (2014). Chapter 5: “The logic of grounded theory coding practices and initial coding,” and Chapter 6: “Focused coding and beyond,” in *Constructing Grounded Theory*. Thousand Oaks, CA: Sage Publications: 109-160.

- Ciarocco, N. J., Vohs, K. D., & Baumeister, R. F. (2010). Some good news about rumination: Task-focused thinking after failure facilitates performance improvement. *Journal of social and clinical psychology, 29*(10), 1057-1073.
- Clark, L., & Watson, D. (1995). Constructing validity: basic issues in objective scale development. *Psychological Assessment, 7*(3), 309-319.
- Coffman, L. C., & Niederle, M. (2015). Pre-analysis plans have limited upside, especially where replications are feasible. *Journal of Economic Perspectives, 29*(3), 81-98.
- Crandall, C. S., & Sherman, J. W. (2016). On the scientific superiority of conceptual replications for scientific progress. *Journal of Experimental Social Psychology, 66*, 93-99.
- Cropley, M., Michalianou, G., Pravettoni, G., & Millward, L. J. (2012). The relation of postwork ruminative thinking with eating behaviour. *Stress and Health, 28*, 23–30.  
<http://doi.org/10.1002/smi.1397>
- Cropley, M., & Zijlstra, F. R. (2011). Work and rumination. *Handbook of stress in the occupations, 487*, 503.
- de Groot, A. D. (2014). The meaning of “significance” for different types of research [translated and annotated by Eric-Jan Wagenmakers, Denny Borsboom, Josine Verhagen, Rogier Kievit, Marjan Bakker, Angelique Cramer, Dora Matzke, Don Mellenbergh, and Han LJ van der Maas]. *Acta Psychologica, 148*, 188-194.
- De Neve, J. E., Krekel, C., & Ward, G. (2018). Work and well-being: A global perspective. *Global happiness policy report, 74-128*.
- Dougherty, M. R., Slevc, L. R., & Grand, J. A. (2019). Making research evaluation more transparent: Aligning research philosophy, institutional values, and reporting. *Perspectives on Psychological Science, 14*(3), 361-375.

- Duke, D., Krishnan, M., Faith, M., & Storch, E. A. (2006). The psychometric properties of the brief fear of negative evaluation scale. *Journal of Anxiety Disorders, 20*(6), 807-817.
- Eib, C., Bernhard-Oettel, C., Magnusson Hanson, L. L., & Leineweber, C. (2018). Organizational justice and health: Studying mental preoccupation with work and social support as mediators for lagged and reversed relationships. *Journal of Occupational Health Psychology, 23*(4), 553.
- Firoozabadi, A., Uitdewilligen, S., & Zijlstra, F. R. (2018). Should you switch off or stay engaged? The consequences of thinking about work on the trajectory of psychological well-being over time. *Journal of occupational health psychology, 23*(2), 278.
- Fresco, D. M., Frankel, A. N., Mennin, D. S., Turk, C. L., & Heimberg, R. G. (2002). Distinct and overlapping features of rumination and worry: The relationship of cognitive production to negative affective states. *Cognitive Therapy and Research, 26*(2), 179-188.
- Gagné, M., Forest, J., Gilbert, M. H., Aubé, C., Morin, E., & Malorni, A. (2010). The motivation at work scale: Validation evidence in two languages. *Educational and Psychological Measurement, 70*(4), 628-646.
- Gagné, M., Forest, J., Vansteenkiste, M., Crevier-Braud, L., Van den Broeck, A., Aspel, A. K., ... & Halvari, H. (2015). The Multidimensional Work Motivation Scale: Validation evidence in seven languages and nine countries. *European Journal of Work and Organizational Psychology, 24*(2), 178-196.
- Goldin-Meadow, S. (2016). Why preregistration makes me nervous. *APS Observer, 29*(7).
- Gutting, G. (1980). Science as discovery. *Revue Internationale de Philosophie, 26*-48.
- Hilbe, J. M. (2011). *Negative binomial regression*. Cambridge University Press.

- Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1-55.
- Harmon-Jones, E. (2003). Anger and the behavioral approach system. *Personality and Individual Differences*, 35(5), 995-1005.
- Ioannidis, J. P. (2012). Why science is not necessarily self-correcting. *Perspectives on Psychological Science*, 7(6), 645-654.
- Jackson, S. E., & Schuler, R. S. (1985). A meta-analysis and conceptual critique of research on role ambiguity and role conflict in work settings. *Organizational behavior and human decision processes*, 36(1), 16-78.
- Jaeger, R. G., & Halliday, T. R. (1998). On confirmatory versus exploratory research. *Herpetologica*, S64-S66.
- John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological Science*, 23(5), 524-532.
- Kaplan, R. M., & Irvin, V. L. (2015). Likelihood of null effects of large NHLBI clinical trials has increased over time. *PloS one*, 10(8), e0132382.
- Klein, O., Hardwicke, T. E., Aust, F., Breuer, J., Danielsson, H., Hofelich Mohr, A., ... & Frank, M. C. (2018). A practical guide for transparency in psychological science. *Collabra: Psychology*, 4(1), 1-15.
- Laudan, L. (1968). Theories of scientific method from Plato to Mach: A bibliographical review. *History of Science*, 7(1), 1-63.

- Lewis, M., Sullivan, M. W., Ramsay, D. S., & Alessandri, S. M. (1992). Individual differences in anger and sad expressions during extinction: Antecedents and consequences. *Infant Behavior and Development, 15*(4), 443-452.
- LeBel, E. & John, L.K. (2016). Toward Transparent Reporting of Psychological Science, in S.O. Lilienfeld and I.D. Waldman (Eds.), *Psychological Science Under Scrutiny: Recent Challenges and Proposed Solutions*, New York: John Wiley & Sons.
- Martin, L. L., & Tesser, A. (1996). Some ruminative thoughts. *Advances in Social Cognition, 9*, 1-47.
- Martin, L. L., Shrira, I. L. A. N., & Startup, H. M. (2004). Rumination as a function of goal progress, stop rules, and cerebral lateralization. *Depressive rumination: Nature, theory and treatment*, 153-176.
- Munafò, M. R., Nosek, B. A., Bishop, D. V., Button, K. S., Chambers, C. D., Du Sert, N. P., ... & Ioannidis, J. P. (2017). A manifesto for reproducible science. *Nature Human Behaviour, 1*(1), 0021.
- Nelson, L. D., Simmons, J., & Simonsohn, U. (2018). Psychology's renaissance. *Annual Review of Psychology, 69*, 511-534.
- Newbury-Birch, D., & Kamali, F. (2001). Psychological stress, anxiety, depression, job satisfaction, and personality characteristics in preregistration house officers. *Postgraduate Medical Journal, 77*(904), 109-111.
- Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., ... & Contestabile, M. (2015). Promoting an open research culture. *Science, 348*(6242), 1422-1425.

- Nosek, B. A., Beck, E. D., Campbell, L., Flake, J. K., Hardwicke, T. E., Mellor, D. T., ... & Vazire, S. (2019). Preregistration is hard, and worthwhile. *Trends in Cognitive Sciences*, 23(10), 815-818.
- Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, 349(6251), aac4716.
- Oswald, A. J., Proto, E., & Sgroi, D. (2015). Happiness and productivity. *Journal of Labor Economics*, 33(4), 789-822.
- Patterson, R. (2002). *Aristotle's modal logic: essence and entailment in the Organon*. Cambridge University Press.
- Platt, J. R. (1964). Strong inference. *Science*, 146(3642), 347-353.
- Popper, K. R. (1959). The propensity interpretation of probability. *The British Journal for the Philosophy of Science*, 10(37), 25-42.
- Quoidbach, J., Gruber, J., Mikolajczak, M., Kogan, A., Kotsou, I., & Norton, M. I. (2014). Emodiversity and the emotional ecosystem. *Journal of Experimental Psychology: General*, 143(6), 2057.
- Rothchild, I. (2006). Induction, deduction, and the scientific method. *Society for the Study of Reproduction*.
- Sakaluk, J. K. (2016). Exploring small, confirming big: An alternative system to the new statistics for advancing cumulative and replicable psychological research. *Journal of Experimental Social Psychology*, 66, 47-54.
- Siegrist, J. (1996). Adverse health effects of high-effort/low-reward conditions. *Journal of Occupational Health Psychology*, 1, 27– 41. [http:// dx.doi.org/10.1037/1076-8998.1.1.27](http://dx.doi.org/10.1037/1076-8998.1.1.27)

- Smaldino, P. E., & McElreath, R. (2016). The natural selection of bad science. *Royal Society Open Science*, 3(9), 160384.
- Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014). p-curve and effect size: Correcting for publication bias using only significant results. *Perspectives on Psychological Science*, 9(6), 666-681.
- Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, 22(11), 1359-1366.
- Smith, J. M., & Alloy, L. B. (2009). A roadmap to rumination: A review of the definition, assessment, and conceptualization of this multifaceted construct. *Clinical Psychology Review*, 29(2), 116-128.
- Sonnentag, S., & Krueel, U. (2006). Psychological detachment from work during off-job time: The role of job stressors, job involvement, and recovery-related self-efficacy. *European Journal of Work and Organizational Psychology*, 15(2), 197-217.
- Spellman, B. A. (2015). A short (personal) future history of revolution 2.0. *Perspectives on Psychological Science*.
- Srivastava, A. K., & Krishna, A. (1980). The effect of job-anxiety on job performance. *The Indian Journal of Social Work*, 41, 255-260.
- Strahan, R., & Gerbasi, K. C. (1972). Short, homogeneous versions of the Marlowe-Crowne social desirability scale. *Journal of Clinical Psychology*.
- Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education*, 2, 53.

- von Thiele Schwarz, U. (2011). Inability to withdraw from work as related to poor next-day recovery and fatigue among women. *Applied Psychology: An International Review*, 60, 377–396. <http://dx.doi.org/10.1111/j.1464-0597.2011.00440.x>
- Uhlmann, E. L., Ebersole, C. R., Chartier, C. R., Errington, T. M., Kidwell, M. C., Lai, C. K., ... & Nosek, B. A. (2019). Scientific utopia III: Crowdsourcing science. *Perspectives on Psychological Science*, 14(5), 711-733.
- Vahle-Hinz, T., Mauno, S., de Bloom, J., & Kinnunen, U. (2017). Rumination for innovation? Analyzing the longitudinal effects of work-related rumination on creativity at work and off-job recovery. *Work & Stress*, 31(4), 315-337.
- Van't Veer, A. E., & Giner-Sorolla, R. (2016). Pre-registration in social psychology—A discussion and suggested template. *Journal of Experimental Social Psychology*, 67, 2-12.
- Vazire, S. (2016). Editorial. *Social Psychological and Personality Science*, 7 (1), 3-7.
- Wagenmakers, E. J., Wetzels, R., Borsboom, D., van der Maas, H. L., & Kievit, R. A. (2012). An agenda for purely confirmatory research. *Perspectives on Psychological Science*, 7(6), 632-638.
- Weston, S. J., Ritchie, S. J., Rohrer, J. M., & Przybylski, A. K. (2019). Recommendations for increasing the transparency of analysis of preexisting data sets. *Advances in Methods and Practices in Psychological Science*, 2(3), 214-227.
- Whillans, A., Macchia, L., & Dunn, E. (2019). Valuing time over money predicts happiness after a major life transition: A preregistered longitudinal study of graduating students. *Science Advances*, 5(9), eaax2615.
- Wiseman, R., Watt, C., & Kornbrot, D. (2019). Registered reports: an early example and analysis. *PeerJ*, 7, e6232.

Woolley, K., & Fishbach, A. (2015). The experience matters more than you think: People value intrinsic incentives more inside than outside an activity. *Journal of Personality and Social Psychology, 109*(6), 968.

Woolley, K., & Fishbach, A. (2016). For the fun of it: Harnessing immediate rewards to increase persistence in long-term goals. *Journal of Consumer Research, 42*(6), 952-966.

### 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:

Strongly disagree	Disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Agree	Strongly agree
1	2	3	4	5	6	7

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: Donation Receipts**

[RECEIPTS REMOVED FOR BLINDED PEER REVIEW]

## Appendix S1: Defining Exploration and Confirmation

In Study S1, we examined how behavioral scientists define confirmatory and exploratory research and whether researchers' intuitive definitions of these constructs aligned with those used in Study 2. First, respondents defined the constructs of exploration and confirmation in their own words. Next, we administered two sets of closed-ended questions. In the first question, participants were presented with statements capturing various research activities and indicated the extent to which each activity represented confirmatory versus exploratory research. Then, we presented respondents with the definitions (used in Study 2) and asked them to rate the extent to which each captured exploratory or confirmatory research. Thus, in Study S1 we used a multi-method approach to examine whether the themes that emerged in respondents' own definitions were consistent with the definitions used in Study 2.

### Methods

#### *Sample*

We recruited behavioral scientists to complete a brief survey about their research experiences in exchange for a \$5.00 donation to COVID-19 relief—participants could direct their donation to one of three relief programs. Participants were recruited via emails sent to: 1) the authors' personal networks ( $n = 151$ ); and 2) academic psychologists at major U.S. universities ( $n = 2,300$ ; from an updated version of the list used in John et al., 2012, contacting only those not contacted in Study 3). These efforts garnered 186 respondents. We analyzed data from the 168 (90% of) respondents who reported conducting behavioral research.<sup>9</sup> We used all

---

<sup>9</sup> We pre-registered collecting 150 responses but were able to collect a total of 186 responses due to snowball sampling and rolling recruitment. Results are substantively equivalent when we restrict the sample to the first 150 responses.

available responses from the 129 respondents who did not complete the entire survey. 72% of participants completed all primary outcome measures.

### *Measures*

**Open-Ended Definitions.** First, participants were asked, separately for exploratory and confirmatory research (order counter-balanced) to “take a moment to think about what you believe is meant by “confirmatory [exploratory] research,” and to “write a definition in your own words in the text box below.”

**Research Activity Ratings.** Next, participants were presented with six common research activities and rated “how exploratory and/or confirmatory” each was, selecting from the response options: “Completely Exploratory,” (coded as -2 for data analysis) “More Exploratory than Confirmatory” (coded as -1 in data analysis), “Equally Exploratory and Confirmatory” (0), “More Confirmatory than Exploratory” (+1), and “Completely Confirmatory” (+2). The activities were: (1) “Conducting a study with a directional hypothesis (e.g., testing whether X increases Y),” (2) “Conducting a study with a non-directional hypothesis (e.g., testing whether X affects Y),” (3) “Conducting a study without a specific hypothesis,” (4) “Pre-registering a study,” (5) “Deciding a priori how much data to collect,” and (6) “Deciding a priori how data will be analyzed.”

**Definition Validation.** Finally, we presented definitions of exploratory and confirmatory research (order counterbalanced) that we had developed based on a literature review of how scholars have thought of these constructs, and iterative input from fellow behavioral scientists.

Exploratory research was defined as:

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 have a prediction, here it would be based more on theory or even intuition, than on prior empirical research. This type of research is common in the early stages of a research project. But it's not limited to the early stages. We consider this type of 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.

Confirmatory research was defined as:

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. This type of research is common in the later stages of a research project. But it's not limited to the later stages. We consider this type of research to include activities such as: conceptual and direct replications, and extending known effects (e.g., identifying boundary conditions or moderators).

For each definition, participants answered: "How exploratory and/or confirmatory is the research?" on the same 5-point scale as the research activity ratings measure.

**Demographics.** As in Studies 1 to 3, participants reported their current role, sub-discipline, primary research methods, gender, and the percentage of studies they had pre-registered in the last 12 months.

This study was pre-registered through AsPredicted [here](#).

## **Results**

### ***Demographics***

See Table S1.

### **Table S1**

### *Sample Demographics*

Appendix S1 (N = 168)	
PhD Year (Actual or Expected)	
	N/A
Use Experimental Methods	
Yes	92%
No	8%
Gender	
Man	48%
Woman	50%
Non-Binary	2%
Role	
Grad Student	10%
Post-Doc	6%
Assistant Prof	13%
Associate Prof, Untenured	3%
Associate Prof, Tenured	19%
Full Professor	41%
Other	7%
Pre-Registration	
Yes	59%
No	41%
Sub-Discipline	
Cognitive Psychology	14%
Consumer Behavior	13%
Social Psychology	20%
Behavioral Economics	4%
Micro-Organizational Behavior	12%
Experimental Economics	<1%
Personality Psychology	2%
Macro-Organizational Behavior	<1%
Interdisciplinary	8%
Other	27%
Time Allocation (%)	
Exploratory (vs. Confirmatory) Research	N/A

### *Open-Ended Definitions*

First, on an exploratory basis, we developed separate coding schemes to capture the dominant themes that emerged in response to the two open-ended questions in which respondents define exploratory and confirmatory research. Following Charmaz (2014), we first

conducted open-coding to see what themes emerged from the data and followed this up with focused-coding to sharpen our coding scheme. One of the authors read the responses and identified the most common themes that emerged in response to each question. Four primary themes emerged for each type of research. The emergent themes for exploratory research were: (1) research for which one does not have an a priori prediction; (2) research in which one learns from the data; (3) research in which there is little to no prior research in the area; and (4) research aimed at generating ideas of hypotheses to test. The emergent themes for confirmatory research were: (1) research driven by an a priori prediction about what the results will be; (2) research informed by previous theory or research; (3) research aimed at replicating previously-documented effects; (4) research confirming hypotheses found in exploratory work. Once these themes were identified, another author used it as a coding scheme to assess whether each of the themes was present or absent from each response.

Respondents were most apt to define exploratory research as research for which one does not have an a priori prediction, with this theme being present in 71% of definitions. Respondents were most apt to define confirmatory research as research for which one has an a priori prediction, with this theme being present in 86% of definitions. See Table S2 for the full qualitative coding results, which displays how frequently each of the eight themes (four each for exploratory and confirmatory research) was detected in respondents' definitions.

**Table S2**

*Qualitative coding results*

Theme	Frequency
Exploratory Research	
Research for which one does not have an a priori prediction	71%
Research aimed at generating ideas of hypotheses to test	39%
Research in which there is little to no prior research in the area	36%

Research in which one learns from the data	13%
<hr/>	
Confirmatory Research	
<hr/>	
Research driven by an a priori prediction about what the results will be	86%
Research informed by previous theory or research	36%
Research aimed at replicating previously-documented effects	18%
Research confirming hypotheses found in exploratory work	2%
<hr/>	

*Note.* Shows the percent of responses that mentioned the given theme.

### ***Research Activity Ratings***

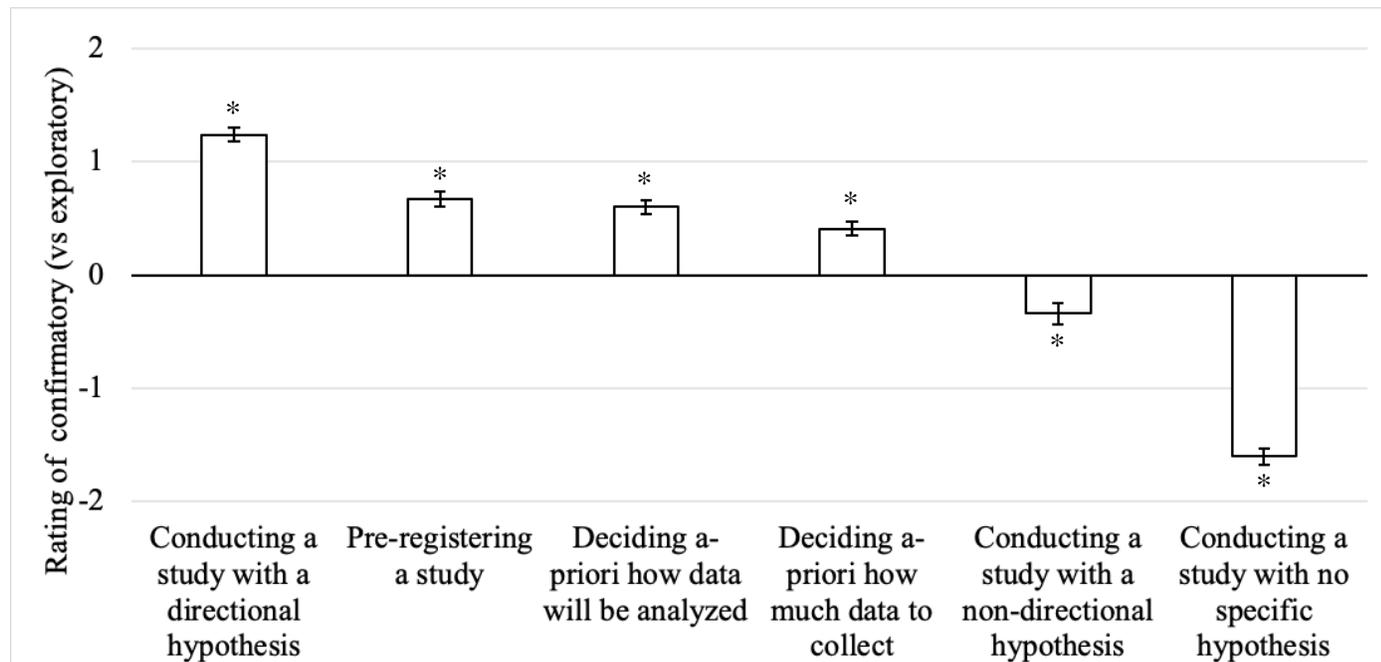
Figure S1 lists the six activities in order of descending average rating. Respondents classified the following activities as confirmatory research (as evidenced by average scores significantly higher than the midpoint): conducting a study with a directional hypothesis ( $M = 1.24$ ,  $SD = 0.69$ ,  $t(130) = 20.55$ ,  $p < .001$ ; pre-registered one-tailed test), pre-registering a study ( $M = 0.67$ ,  $SD = 0.67$ ,  $t(130) = 10.61$ ,  $p < .001$ ), deciding a priori how much data to collect ( $M = 0.41$ ,  $t(130) = 6.96$ ,  $p < .001$ ), and deciding a priori how data will be analyzed ( $M = 0.60$ ,  $SD = 0.41$ ,  $t(130) = 9.60$ ,  $p < .001$ ). Respondents classified the following activities as exploratory research (as evidenced by average scores significantly lower than the midpoint): conducting a study with a non-directional hypothesis ( $M = -0.34$ ,  $SD = 1.06$ ,  $t(130) = -3.72$ ,  $p < .001$ ); and conducting a study without a specific hypothesis ( $M = -1.60$ ,  $SD = 0.76$ ,  $t(130) = -23.95$ ,  $p < .001$ ; pre-registered one-tailed test).

Conducting a study with a directional hypothesis received the highest average rating, suggesting that, of the activities we assessed, it was perceived as the most characteristic of confirmatory research. Its average rating, of 1.24, was significantly higher than the second-highest-rated activity: pre-registering a study, which garnered an average rating of 0.67,  $t(130) = 7.02$ ,  $p < .001$ . Similarly, conducting a study without a specific hypothesis received the lowest average rating, suggesting that, of the activities we assessed, it was deemed most characteristic of exploratory research. Its average rating, of -1.60, was significantly lower than the second-

lowest-rated activity: conducting a study with a non-directional hypothesis, which garnered an average rating of -0.34,  $t(130) = -12.02$ ,  $p < .001$ .

### Figure S1

#### *Confirmatory vs. Exploratory Research Experiences*



*Note.* Researcher ratings of how “confirmatory” versus “exploratory” various research experiences were. Zero represents the scale midpoint (“Equally confirmatory and exploratory”), +2 represents “Completely Confirmatory,” and -2 represents “Completely Exploratory.”

#### **Definition Validation**

As expected, participants rated our definition of confirmatory research as significantly more confirmatory than our definition of exploratory research ( $M_{diff} = 2.35$ ,  $t(130) = 26.33$ ,  $p < .001$ ; pre-registered one-tailed test). Furthermore, participants rated our definition of confirmatory research as significantly more confirmatory compared to the scale midpoint ( $M = 1.19$ ,  $SD = 0.60$ ,  $t(130) = 22.86$ ,  $p < .001$ ; pre-registered one-tailed test), and our definition of

exploratory research as significantly more exploratory compared to the midpoint ( $M = -1.16$ ,  $SD = 0.79$ ,  $t(130) = -16.76$ ,  $p < .001$ ; pre-registered one-tailed test).

## **Discussion**

Study S1 suggests that our definitions of exploratory and confirmatory research, which we used in Study 2, are consistent with how behavioral scientists think of these constructs. When asked to spontaneously define confirmatory research, the possession of an a priori prediction featured prominently in respondents' free responses. Notably, predictions can be directional or non-directional, and respondents' open-ended responses typically did not make this distinction. However, as we discuss below, our second measure delved into this distinction. Respondents' definitions also hit on several other key themes featured in our definitions. For example, 36% of respondents' definitions of confirmatory research captured the notion of being informed by previous theory or research. Similarly, our confirmatory research definition featured having "a sense from prior research [...] of what the result will be." Eighteen percent of confirmatory definitions featured replication; similarly, our definition listed "conceptual and direct replications, and extending known effects." As for exploratory research, 36% of respondents' definitions captured the notion of conducting research in a new space, where there is little prior work; 39% noted generating and testing new ideas. Similarly, our exploratory research definition featured "brainstorming research ideas that have not yet been addressed," and "designing and conducting the first tests of those research ideas, and finding out what the results of those first tests show." Thus, our definitions included many of the most common themes in respondents' own definitions of these terms. We note however, that our definitions did not feature all themes

mentioned by respondents. For example, our definition of exploratory research did not capture the idea of “learning from one’s data,” which was captured in 13% of respondents’ definitions.

Both sets of closed-ended questions also point to the conclusion that our definitions of exploratory and confirmatory research reasonably reflected how behavioral scientists think of these constructs. Knowing aspects of the research process upfront was deemed characteristic of confirmatory research; of the items that we assessed, having an a priori *directional* prediction was deemed particularly emblematic. By contrast, conducting a study without a specific prediction was deemed particularly characteristic of exploratory research. Notably however, conducting a study with a non-directional hypothesis was also viewed as being more indicative of exploratory research than confirmatory research. Fittingly, our definition of exploratory research does not preclude the possibility of having a prediction (i.e., it states that “you may or may not have a prediction”). Finally, participants rated our definition of confirmatory research as being significantly more reflective of confirmatory research than exploratory research; and vice versa for our definition of exploratory research.

In sum, based on this evidence, we conclude that our definitions in Study 2 were reasonably representative of, and consistent with, how behavioral scientists think of these constructs—though not necessarily exhaustive.

## Supplemental Online Material

### Study 1 Supplemental Information

#### Supplemental Methods

##### *Prediction Preoccupation*

Respondents completed a preliminary version of our six-item prediction preoccupation scale ( $\alpha=0.67$ ; Table S1). This scale was later finalized and validated in Study 3 of the main manuscript. See the supplemental information for Study 3, later in this supplement for more information on the iterative scale construction process.

#### Table S1

##### *Preliminary Version of the Prediction Preoccupation Scale*

Prediction Preoccupation Pilot Scale Items
1. I would consider avoiding running a study if I was worried that the results weren't going to confirm my predictions.
2. It's risky to run studies when you don't have a strong sense of what the result will be.
3. I sometimes feel "stuck," in the sense that I can't run a study unless I know what the result will be, but to know what the result will be, I need to run a study.
4. I feel relieved when the results of a study confirm my prediction.
5. I feel guilty when I run studies that aren't pre-registered.
6. It feels wrong to run a study without the intent to publish (i.e., simply to "see what happens").

#### Supplemental Results

##### *Demographic Differences in Condition Effects*

We conducted additional exploratory analyses to investigate whether the effects of experimental condition in Study 1 were moderated by respondent characteristics.

##### *Moderation by Pre- vs. Post-Bem*

We examined whether the effects of the confirmatory and hybrid conditions differed between participants who received their PhD before/during vs. after 2011, when significant

scientific reforms began. There was no effect on the likelihood that participants viewed the interaction,  $b = -0.02$ ,  $SE = 0.55$ ,  $p = .96$  and no significant interaction with either the confirmation or hybrid conditions versus the exploration condition ( $ps \geq .15$ ). We found no effect of PhD graduation year on the likelihood of reporting the results of the interaction ( $p \geq .99$ ) and no interaction with the confirmation or hybrid conditions versus the exploration condition ( $ps \geq .99$ ). Finally, we found no main effect of PhD graduation year on the number of analyses participants selected to view,  $b = -0.11$ ,  $SE = 0.10$ ,  $p = .27$  and no interaction with the confirmation or hybrid conditions versus the exploration condition ( $ps \geq .62$ ).

### ***Other Demographic Effects***

We found no main effects or interactions with other respondent characteristics including gender, whether or not participants pre-register, tenure status, or whether or not participants used experimental methods.

## **Study 2 Supplemental Information**

On an exploratory basis, we investigated whether there were differences in respondents' subjective experiences of research as a function of various respondent characteristics.

### **Supplemental Results**

#### ***Subjective Experience of Exploration***

We tested for gender differences in researchers' subjective experiences of exploratory research using a Bonferroni correction for multiple t-tests. Using a revised alpha cutoff of  $p < .007$  (as we conducted seven t-tests—one for each adjective), women reported exploratory research to be more frustrating ( $M_{\text{men}} = 3.30$ ,  $SD = 3.00$ ;  $M_{\text{women}} = 3.85$ ,  $SD = 1.89$ ;  $t(331.58) = 3.86$ ,  $p < .001$ ) and anxiety-inducing ( $M_{\text{men}} = 3.24$ ,  $SD = 1.75$ ;  $M_{\text{women}} = 3.84$ ,  $SD = 1.99$ ;

$t(333.43) = 2.95, p < .001$ ) than men. All other comparisons between men and women were not statistically significant ( $ps \geq 0.28$ ).

Further, we investigated whether researchers' subjective experiences of exploratory research differed as a function of whether they graduated before vs after 2011 (when significant reforms began), whether they were pre vs post-tenure, and whether they did vs. not engage in pre-registration practices. Again, we applied a Bonferroni correction to account for multiple t-tests and found that all comparisons failed to reach statistical significance ( $ps \geq .03$ ).

### ***Subjective Experience of Confirmation***

As above, we tested for differences in researchers' subjective experiences of confirmatory research using a Bonferroni correction for multiple t-tests. With an alpha cutoff level of  $p < .007$  (as we conducted seven t-tests for each demographic comparison), researchers who received their PhD after 2011 (vs. before) reported confirmatory research to be significantly more anxiety-inducing ( $M_{pre} = 3.66, SD = 1.81; M_{post} = 4.34, SD = 1.72; t(106.56) = 2.81, p = 0.006$ ). All other comparisons by gender, tenure status and engagement with pre-registration failed to reach significance ( $ps \geq .01$ ).

### ***Time Allocation (Exploratory vs. Confirmatory)***

There were no differences in the amount of time that researchers spent on exploratory vs confirmatory research as a function of researchers' gender, year of PhD, tenure status, or engagement with pre-registration practices ( $ps \geq .07$ ).

### ***Satisfaction with the Field***

We found no significant differences in researchers' satisfaction with the field of behavioral science as a function of their gender, year of PhD, tenure status, or engagement with pre-registration practices. See Table S2.

### ***Satisfaction with Role***

The year that respondents received their PhD was significantly negatively associated with role satisfaction,  $b = -0.02$ ,  $SE = 0.008$ ,  $p = 0.002$ . Researchers who received their PhD post-2011 were significantly less satisfied with their current role ( $M_{pre} = 5.14$ ,  $SD = 1.31$ ;  $M_{post} = 4.65$ ,  $SD = 1.43$ ),  $t(117.97) = 2.72$ ,  $p = 0.007$ . Further, participants with tenure ( $M = 5.31$ ,  $SD = 1.17$ ) were significantly more satisfied with their current role than those without tenure ( $M = 4.63$ ,  $SD = 1.42$ ),  $t(116.96) = 3.97$ ,  $p < .001$ . There were no significant differences in satisfaction with role as a function of gender or whether or not researchers engaged in pre-registration. See Table S2.

### ***Interest in Staying in Field***

The year participants received their PhD was significantly negatively associated with their interest in staying in the field,  $b = -0.02$ ,  $SE = .007$ ,  $p < .001$ . Researchers who received their PhD before/during 2011 ( $M = 6.28$ ,  $SD = 1.04$ ) were more interested in staying in the field than those who received their PhD during/after 2011 ( $M = 5.76$ ,  $SD = 1.38$ ),  $t(141.58) = 3.47$ ,  $p < .001$ . Researchers with tenure ( $M = 6.36$ ,  $SD = 0.93$ ) were significantly more interested in staying in the field than those without tenure ( $M = 5.79$ ,  $SD = 1.39$ ),  $t(145.55) = 3.89$ ,  $p < .001$ . Finally, those who did preregister ( $M = 6.00$ ,  $SD = 1.20$ ) reported higher interest in staying in the field than those who did not ( $M = 5.69$ ,  $SD = 1.50$ ),  $t(256.20) = 2.03$ ,  $p = 0.04$ . Men and women did not differ in the extent to which they were interested in staying in the field. See Table S2.

### **Table S2**

#### *Demographic Differences in Career Satisfaction*

	Men	Women	Comparison
Satisfaction with field	$M = 4.67$ , $SD = 1.29$	$M = 4.84$ , $SD = 1.13$	$t(343.80) = 1.32$ , $p = 0.20$
Satisfaction with role	$M = 4.86$ , $SD = 1.42$	$M = 4.64$ , $SD = 1.42$	$t(350.00) = 1.43$ , $p = .15$

Interest in staying	$M = 5.98, SD = 1.27$	$M = 5.76, SD = 1.41$	$t(346.39) = 1.51, p = .13$
	Pre-Bem	Post-Bem	Comparison
Satisfaction with field	$M = 5.01, SD = 1.28$	$M = 4.70, SD = 1.20$	$t(104.93) = 1.87, p = 0.06$
Satisfaction with role	$M = 5.14, SD = 1.31$	$M = 4.65, SD = 1.43$	$t(117.97) = 2.72, p = 0.007$
Interest in staying	$M = 6.28, SD = 1.04$	$M = 5.76, SD = 1.38$	$t(141.58) = 3.47, p < .001$
	No Tenure	Tenure	Comparison
Satisfaction with field	$M = 4.68, SD = 1.19$	$M = 4.89, SD = 1.38$	$t(89.32) = 1.12, p = 0.27$
Satisfaction with role	$M = 4.63, SD = 1.42$	$M = 5.31, SD = 1.17$	$t(96.92) = 0.28, p = 0.78$
Interest in staying	$M = 5.79, SD = 1.39$	$M = 6.36, SD = 0.93$	$t(145.55) = 3.89, p < .001$
	Do not pre-register	Pre-register	Comparison
Satisfaction with field	$M = 4.87, SD = 1.05$	$M = 4.69, SD = 1.31$	$t(340.70) = 1.49, p = 0.14$
Satisfaction with role	$M = 4.65, SD = 1.45$	$M = 4.82, SD = 1.40$	$t(294.45) = 1.04, p = 0.30$
Interest in staying	$M = 5.69, SD = 1.50$	$M = 6.00, SD = 1.20$	$t(256.20) = 2.03, p = 0.04$

### ***Prediction Preoccupation***

Respondents completed a preliminary version of our six-item prediction preoccupation scale as in Study 1 ( $\alpha=0.66$ ; Table S1). This scale was later finalized and validated in Study 3 of the main manuscript. See the supplemental information for Study 3, later in this supplement for more information on the iterative scale construction process.

On an exploratory basis, we examined associations between this preliminary version of the scale and career satisfaction. However, given that this scale was later finalized and validated in Study 3 of the main manuscript, we do not draw strong conclusions from these results.

**Demographic Associations.** On an exploratory basis, we were interested in *who* experienced prediction preoccupation. Prediction preoccupation was higher among female ( $M = 3.80$  out of 7,  $SD = 0.98$ ) than male respondents ( $M = 3.50, SD = 1.06$ ),  $t(347.66) = 2.76, p =$

.006, Cohen's  $d = 0.29$ , 95%  $CI [0.08, 0.51]$ . People who received their PhD after 2011, when major scientific reforms started to take place, reported significantly higher levels of prediction preoccupation than those who obtained their PhD in or before 2011 ( $M_{\text{post}_{2011}} = 3.79$ ,  $SD = 1.01$ ;  $M_{\text{pre}_{2011}} = 3.18$ ,  $SD = 0.97$ ),  $t(114.21) = 4.75$ ,  $p < .001$ , Cohen's  $d = 0.62$ , 95%  $CI [0.35, 0.88]$ . Similarly, untenured respondents reported significantly higher feelings of prediction preoccupation than tenured respondents ( $M_{\text{untenured}} = 3.84$ ,  $SD = 1.03$ ;  $M_{\text{tenured}} = 3.11$ ,  $SD = 0.89$ ),  $t(111.94) = 5.67$ ,  $p < .001$ ,  $d = 0.73$ , 95%  $CI [0.45, 1.01]$ . Lastly, respondents who reported pre-registering at least one study in the past year reported similar levels of prediction preoccupation as compared to those who did not report pre-registering ( $M_{\text{no\_preregister}} = 3.57$ ,  $SD = 1.06$ ;  $M_{\text{preregister}} = 3.69$ ,  $SD = 1.01$ ),  $t(290.39) = 1.03$ ,  $p = 0.30$ , Cohen's  $d = -0.11$ , 95%  $CI [-0.33, .10]$ .

**Satisfaction with Field.** First, we explored whether prediction preoccupation, subjective experience, and time-use predicted satisfaction with current role using ordinal logistic regression models. We found no significant relationship between prediction preoccupation,  $b = -0.04$ ,  $SE = 0.10$ ,  $p = 0.66$ ), time spent on exploratory vs. confirmatory research,  $b = -0.005$ ,  $SE = 0.25$ ,  $p = 0.25$ , and subjective experiences (enjoyable, motivating, interesting, frustrating, anxiety inducing, boring, scientific) of both exploratory and confirmatory research (all  $ps \geq 0.07$ ).

**Satisfaction with Role.** Next, we explored whether prediction preoccupation, subjective experience, and time-use predicted satisfaction with current role using ordinal logistic regression models. Prediction preoccupation was a significant negative predictor of satisfaction with role,  $b = -0.37$ ,  $SE = 0.09$ ,  $p < .001$  (this relationship held controlling for gender, tenure, and whether participants received their PhD before or after 2011:  $b = -0.32$ ,  $SE = 0.11$ ,  $p = .003$ ). Role satisfaction predicted anxiety with exploratory research,  $b = -0.17$ ,  $SE = 0.07$ ,  $p = 0.01$ , and confirmatory research,  $b = -0.22$ ,  $SE = 0.07$ ,  $p = 0.002$ . In other words, people who felt that

research was “anxiety-inducing” were also less satisfied with their role (these links held controlling for gender, tenure, and whether participants received their PhD before or after 2011; exploratory anxiety-inducing:  $b = -0.18$ ,  $SE = .07$ ,  $p = .01$ ; confirmatory anxiety-inducing:  $b = -0.20$ ,  $SE = 0.08$ ,  $p = .01$ ). Time spent on exploratory (vs. confirmatory) research did not significantly relate to participants experienced satisfaction with their role,  $b = -0.004$ ,  $SE = 0.003$ ,  $p = 0.37$ .

**Interest in Staying in the Field.** Finally, we explored whether prediction preoccupation, subjective experience, and time-use predicted the likelihood of staying in the field of behavioral science using ordinal logistic regression models. Prediction preoccupation negatively predicted interest in staying in the field,  $b = -0.24$ ,  $SE = 0.10$ ,  $p = 0.01$ . This relationship did not hold controlling for gender, tenure and whether participants received their degree before or after 2011,  $b = -0.12$ ,  $SE = 0.11$ ,  $p = .29$ . Respondents’ who reported that exploratory research was “interesting” were more interested in staying the field,  $b = 0.40$ ,  $SE = 0.18$ ,  $p = 0.03$ . A similar pattern emerged for “motivating,”  $b = 0.24$ ,  $SE = 0.12$ ,  $p = .04$ . Once again, these results did not hold controlling for gender, tenure, and whether participants received their PhD before/during or after 2011  $ps \geq 0.10$ . Time spent on exploratory (vs. confirmatory) research did not predict interest in staying in the field,  $b = -0.0004$ ,  $SE = .004$ ,  $p = 0.92$ .

### Study 3 Supplemental Information

#### Supplemental Methods

##### *Prediction Preoccupation Scale Construction*

We piloted our prediction preoccupation scale in Study 1 ( $\alpha=0.67$ ) and Study 2 ( $\alpha=0.66$ ) using the items presented in Table S1. Exploratory factor analysis in both of these studies failed to produce a factor solution in which all items loaded reasonably well onto at least one factor

(see Table S3 for factor loadings with a three-factor solution which was determined to be the best fit to the data using a scree plot and eigenvalues).

**Table S3**

*Prediction Preoccupation Scale Items and Factor Loadings for Studies 1 and 2*

	Study 1			Study 2		
	Confirmation factor	Exploration factor	Factor 3	Confirmation factor	Exploration factor	Factor 3
Item 1			0.77	0.81		
Item 2		>0.99				>0.99
Item 3			0.44	0.44		
Item 4			0.47	0.41		
Item 5	>0.99					
Item 6					>0.99	

*Note.* Factor loadings less than 0.30 are not included in table (Field, 2013).

Based on these results, in Study 3 we sought to revise our prediction preoccupation scale to improve its psychometric properties. We refined our construct definition, to refer to “prediction preoccupation”—the extent to which researchers feel heightened concern over, and fixation with, confirming predictions. Our revised items correspond to this refined definition. Specifically, we revised the items: (1) to focus on feelings as opposed to behavior; and (2) to establish a clear two-factor solution (with items pertaining to confirmatory contexts loading on one factor, and those pertaining to exploratory contexts loading onto another). We also honed the wording of the items, in the interest of maximizing clarity and hence, reliability. The resultant scale exhibited reliability ( $\alpha=0.71$ ) and a clear two-factor solution (Confirmation factor: items 1-4; Exploration factor: items 5 & 6). Detailed results are presented in the main manuscript. See Table S3 for a comparison of the two scale versions.

**Supplemental Results**

*Prediction Preoccupation Predicting Filler Items*

Researchers reported how ‘alert’ and ‘inspired’ they would feel when engaging in various research activities. These were included as filler items. As expected, prediction preoccupation was not significantly related to alertness or inspiration, with one exception: Prediction preoccupation was positively associated with anticipated alertness when writing up a pre-registration for a direct replication study ( $b = 0.14$ ,  $SE = 0.06$ ,  $p = .02$ ). However, our scale showed no relationship to alertness for any other research activity,  $ps \geq 0.13$ . Further, prediction preoccupation was unrelated to anticipated inspiration for every research activity,  $ps \geq 0.58$ .

### ***Predictive Validity of Prediction Preoccupation by Factor***

In our exploratory factor analysis, we found evidence for a two-factor solution (the confirmation factor was comprised of items 1-4; the exploration factor was comprised of items 5-6). We examined whether these two factors were differentially linked to respondents expected “anxiety” and “excitement” when engaging in various research experiences. The confirmation factor was more strongly positively related to anxiety than the exploration factor when designing a direct replication study (confirmation factor:  $b = 0.51$ ,  $SE = 0.05$ ,  $p < .001$ ; exploration factor:  $b = 0.04$ ,  $SE = 0.04$ ,  $p = .34$ ; comparison of betas:  $F(1, 261) = 47.30$ ,  $p < .001$ ), writing a pre-registration for a direct replication study (confirmatory factor:  $b = 0.42$ ,  $SE = 0.05$ ,  $p < .001$ ; exploration factor:  $b = 0.10$ ,  $SE = 0.04$ ,  $p = .01$ ; comparison of betas:  $F(1, 261) = 18.67$ ,  $p < .001$ ), and conducting the pre-registered analyses (confirmation factor:  $b = 0.56$ ,  $SE = 0.05$ ,  $p < .001$ ; exploration factor:  $b = 0.06$ ,  $SE = 0.04$ ,  $p = .12$ ; comparison of betas:  $F(1, 261) = 44.92$ ,  $p < .001$ ), but were similarly related to anxiety when conducting additional exploratory analyses on these data (confirmation factor:  $b = 0.20$ ,  $SE = 0.05$ ,  $p < .001$ ; exploration factor:  $b = 0.19$ ,  $SE = 0.04$ ,  $p < .001$ ; comparison of betas:  $F(1, 261) = 0.01$ ,  $p = .91$ ). The confirmation factor and exploration factor were similarly positively related to predicted anxiety when designing an

exploratory study (confirmation factor:  $b = 0.25$ ,  $SE = 0.05$ ,  $p < .001$ ; exploration factor:  $b = 0.16$ ,  $SE = 0.04$ ,  $p < .001$ ; comparison of betas:  $F(1, 261) = 1.47$ ,  $p = .24$ ) or running exploratory analyses (confirmation factor:  $b = 0.15$ ,  $SE = 0.05$ ,  $p < .001$ ; exploration factor:  $b = 0.26$ ,  $SE = 0.04$ ,  $p < .001$ ; comparison of betas:  $F(1, 261) = 2.62$ ,  $p = .11$ ).

### ***Prediction Preoccupation and Researcher Experiences***

Prediction preoccupation was not significantly related to researchers' satisfaction with the field ( $b = 0.06$ ,  $SE = 0.07$ ,  $p = .40$ ), however when we separated the measure into two factors, the confirmation factor was not related to researchers' satisfaction ( $b = -0.06$ ,  $SE = 0.07$ ,  $p = .36$ ), but the exploration factor was significantly positively related ( $b = 0.11$ ,  $SE = 0.05$ ,  $p = .21$ ).

Prediction preoccupation was significantly negatively related to researchers' satisfaction with their current role ( $b = -0.23$ ,  $SE = 0.07$ ,  $p < .001$ ). When we separate out the two factors, the confirmation factor was significantly related ( $b = -0.25$ ,  $SE = 0.06$ ,  $p < .001$ ) while the exploration factor was not ( $b = 0.02$ ,  $SE = 0.05$ ,  $p = .71$ ).

Finally, prediction preoccupation was not significantly related to the extent researchers reported intentions to stay in the field ( $b = -0.05$ ,  $SE = 0.07$ ,  $p = .51$ ), and neither factor was significantly related (confirmation factor:  $b = -0.0002$ ,  $SE = 0.07$ ,  $p = .998$ ; exploration factor:  $b = -0.05$ ,  $SE = 0.05$ ,  $p = .37$ ). Finally, predicting the number of studies run in the last 12 months from both prediction preoccupation factors simultaneously, the confirmation factor was negatively related to number of studies run ( $b = -1.41$ ,  $SE = 0.48$ ,  $p = .004$ ) while the exploration factor was not significantly related ( $b = 0.95$ ,  $SE = 0.63$ ,  $p = .13$ ). Future research is needed to understand the differential predictive validity of these factors.