

Exploration in Behavioral Science

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

Working Paper 20-090



Exploration 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 by 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.

Exploration in Behavioral Science

Hanne K. Collins

Ashley V. Whillans

Leslie K. John

Abstract

In the past decade, behavioral science has undergone beneficial shifts towards greater transparency and more rigorous research. However, we wondered whether this heightened focus on confirmation might make researchers feel inhibited towards doing research, and reduce the propensity to explore. Study 1 ($N=404$), a descriptive survey, indicated that relative to confirmatory research, researchers found exploratory research more enjoyable, motivating, and interesting; and less anxiety-inducing, frustrating, boring, and scientific. It also introduced a scale to measure research inhibition; higher scores were linked to running fewer studies and were particularly common among women, untenured professors, and receiving one's PhD after 2011—when reforms began. Study 2 ($N=463$), a pre-registered experiment, indicated that a confirmatory mindset made researchers less likely to find an interesting (non-hypothesized) interaction. A reminder to explore did not mitigate this effect. These studies suggest that work is needed to ensure that rigorous scientific reforms and exploration can coexist.

Keywords: Open Science; Pre-Registration; Exploration; Career Satisfaction; Diversity

A fateful 2011 *JPSP* article claimed to have documented evidence of pre-cognition (Bem, 2011). Yes, it was an outlandish claim. But the article had the trappings of robust evidence, offering nine experimental demonstrations of the phenomenon across different paradigms. Perhaps because of the article's bold claims, it underwent particularly close scrutiny. Observers noted how the methods left room for exploitation of researcher degrees of freedom, magnifying the possibility for Type I error (Carey, 2011; Wagenmakers, Wetzels, & Borsboom, 2011). Indeed, when those liberties were not taken, the effect did not replicate (Galak, LeBoeuf, Nelson, & Simmons, 2012). Although probably not its intended contribution, Bem (2011) helped spark a movement designed to improve the credibility of research in social psychology and allied fields (John, Loewenstein, & Prelec, 2012; Simmons, Nelson, & Simonsohn, 2011). Together with failures to replicate famous effects and detection of data fabrication, the Bem (2011) paper was a proverbial straw that broke the camel's back (see Nelson, Simmons & Simonsohn, 2018 for an overview of precipitating events).

Since then, many positive reforms to behavioral science have taken place (Camerer et al., 2016; Crandall & Sherman, 2016; Dougherty, Slevic, & Grand, 2019; Klein et al., 2018; Lebel & John, 2016; Nelson, Simmons, & Simonsohn, 2018; Nosek et al., 2015; Nosek et al., 2019; Open Science Collaboration, 2015; Simmons, Nelson, & Simonsohn, 2011; Uhlmann et al., 2019; Van't Veer & Giner-Sorolla 2016; Weston, Ritchie, Rohrer, & Przybylski, 2019). For example, there has been a movement towards pre-registration and registered replications. Whereas ten years ago it was virtually unheard of for behavioral scientists to pre-register their studies, 2,000 independent researchers did so within the first 14 months of AsPredicted.com's inception (Nelson, Simmons & Simonsohn, 2018). For its part, the Open Science Framework had more than 8,000 documented preregistrations as of 2017 (Nosek, Ebersole, DeHaven, & Mellor,

2018). In addition, leading journals such as *Nature*, *Nature Human Behavior*, and most recently, *Science*, have introduced a registered report manuscript category, whereby peer review is conducted prior to data collection, and accepted plans are published regardless of the findings.

Emblematic of these changes, behavioral science has seen a shift towards greater transparency. Publicly posting one's research stimuli and data is standard practice for a growing number of researchers (McKiernan et al., 2016; Nosek et al., 2018). Major journals now require authors to explicitly state how they determined their sample sizes, and whether they have reported all of their manipulations, measures, and observations (Asendorpf, Conner, De Fruyt, De Houwer, & Denissen, 2016; Cumming, 2014; Dexter & Shafer, 2017; Eich, 2014; Lindsay, 2015; Lindsay, 2017; Simonsohn, 2013; Trafimow & Marks, 2015; Vazire, 2016; Veldkamp, Nuijten, Dominguez-Alvarez, van Assen, & Wicherts, 2014; Wicherts, Borsboom, Kats, & Molenaar, 2006; Giofrè, Cumming, Fresc, Boedker, & Tressoldi, 2017 for a review of recent journal practices). In 2014, *Psychological Science*, psychology's flagship journal, introduced "open practices badges" – icons published on the first page of the article, denoting that the authors have shared their data, made their materials public, and preregistered their hypotheses (Eich, 2014). In the first year of these badges alone, the percent of articles with open data increased from 3% to 36%, and the percent of articles with open materials jumped from 13% to 41% (Kidwell et al., 2016; see also: Giofrè et al., 2017; Nuijten et al., 2017). Debates over whether findings replicate are now being held in the open, as opposed to being only the stuff of conference gossip. And now, more than ever, behavioral scientists are engaged in healthy debate over what constitutes sound versus unsound scientific practice (Ioannidis, Fanelli, Dunne, & Goodman, 2015).

We believe—like many other researchers (e.g., Munafò et al., 2017; Smaldino & McElreath, 2016; Spellman, 2015)—that such changes are instrumental to improving the quality of research. However, alongside the benefits of reform, some scholars have expressed concerns over negative unintended consequences. Some have suggested that the reform movement has led to a reliance on online studies, and a corresponding (over)reliance on self-report measures over consequential behavioral outcomes (Sassenberg & Ditrich, 2019). Of particular relevance to our inquiry, there is a worry that these reforms may be hindering exploration. Some have expressed concerns over increases in the Type 2 error rate (Gilbert, King, Pettigrew, & Wilson, 2016). As the 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). Might the reform movement, and an emphasis on pre-registration in particular, be making researchers feel a kind of anxiety over confirming their predictions, a sense malaise over conducting research, or a reticence to engage with the process?

We find ourselves having a recurrent conversation. Sometimes it is a conversation with a colleague or student. But sometimes it is a conversation with ourselves, in our own minds. It goes something like this:

Voice 1 (with consternation): I’m reluctant to run this study because I’m not sure what the result will be.

Voice 2: How might you get a sense of what the result will be?

Voice 1: By running the study.

Voice 2: Why don’t you run the study, then?

Voice 1. Because I don’t know what the result will be.

To be sure, the exchange usually has more subtlety to it. But by paraphrasing its gist here, the inherent circularity is striking: researchers who feel this way may find themselves in a kind of perpetual “wheel-spinning” mode. Indeed, we have had colleagues and doctoral students tell us that they have deliberated for months over designing the “perfect” paradigm – all without collecting a single data point. Another factor that strikes us about this conversation is its tone; it often has a kind of anxious listlessness to it.

As a result of these observations, we wondered whether the reform movement, for all its laudable qualities, may be producing some unintended (but repairable) negative consequences. One such consequence, as others have alluded to, is that it might shift the nature of what is discovered – or perhaps more aptly, what is *not* discovered – a movement towards “safer,” less novel topics and perhaps a movement away from exploration. If the above exchange is any indication, feeling that one cannot run a study without knowing its outcome in advance seems like a surefire way to stifle exploration. Specifically, we wondered whether our field’s increased focus on confirmation might make researchers feel inhibited towards doing research, and reduce the propensity to explore. We also wondered what implications such a reduction might have on the subjective experience of doing research – if exploration is more enjoyable than confirmation, then doing less of it might have negative implications for overall satisfaction with research.

Recent meta-scientific research focuses on assessing the replicability of published findings, documents the problematic aspects of the research process, and offers helpful prescriptions. However, out of the nearly 2,000 articles published each year on meta-scientific concerns (Ioannidis et al., 2015), no research to our knowledge has explored factors that predict the satisfaction that researchers derive from various aspects of the research process—and how recent scientific reforms might shape researchers’ satisfaction. We believe that this question

matters in light of research showing that the experience of satisfaction, happiness, and joy at work is linked to greater productivity (Bellet, DeNeve & Ward, 2019; Oswald, Proto & Sgroi, 2015); as well as research showing that greater task enjoyment predicts intrinsic motivation and long-term persistence on difficult and challenging activities (Woolley & Fishbach, 2015; Woolley & Fishbach, 2016). Understanding how scientific reforms might shape researchers' subjective experiences and satisfaction could also contribute to the broader discussion of how such reforms shape *who* decides to stay in academia and why (Yarkoni, 2016).

The current research has two primary aims. First, we sought to assess researchers' current subjective experience in conducting research. We explored this topic by surveying behavioral researchers and understanding the aspects of the research process they find more-versus-less enjoyable, and whether researchers feel more apprehensive about, and derive less satisfaction from, confirmatory—vs exploratory—research. We also designed a scale to measure 'research inhibition' – the kind of apprehension towards doing research exemplified by the “recurrent conversation” we described above. Second, we assessed whether a focus on confirmation might impede exploration. We tested this idea in a pre-registered experiment in which we tested whether inducing a confirmatory mindset (via a simulated pre-registration task) reduced researchers' propensity to explore, and if so, whether reminding people to explore could mitigate such an effect. Ultimately, we hope that by gaining a better understanding of the subjective experiences of conducting research in light of scientific reform, we can find a way for rigorous open-science practices and exploration to coexist—and for researchers to see open-science practices as empowering and exciting (Fanelli, 2018) as opposed to anxiety-inducing.

Study 1: Survey

We conducted a descriptive survey to gain an understanding of researchers' current subjective experience of doing research.

Methods

We recruited behavioral science researchers to participate in a brief 7-10 minute survey. Prior to running the study, we conducted a small pilot with $n = 10$ colleagues and doctoral students, adjusting our survey according to their feedback. Participants were recruited in-person at the 2019 meeting of the Society for Judgement and Decision-Making (SJDM; $n = 277$) and online through the SJDM ($n = 89$) and Psych-Methods ($n = 38$) email list-servs, for a total of 431 respondents. We have no way of knowing the exact number of people exposed to our recruitment sign (SJDM) or emails (SJDM/Psych-Methods list-servs). Thus, the response rate is unknown.

We analyzed the data from the 400 (93%) of respondents who reported conducting behavioral research; 356 (88%) of these respondents completed the full survey. We used all available responses from respondents who did not complete the entire survey.

Measures

Research Inhibition. We developed a 6-item measure of research inhibition. Respondents were asked to report the extent to which they agreed or disagreed with six statements, including "It's risky to run studies when you don't have a strong sense of what the result will be," and "I would consider avoiding running a study if I was worried that the results weren't going to confirm my predictions" on a scale from 1 = *Strongly Disagree* to 7 = *Strongly Agree* (see Appendix A for full scale).

Subjective Experiences. Next, respondents read a description of exploratory research: "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," and 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: "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," and responded using the identical scale. The presentation order of exploratory and confirmatory questions was counterbalanced.

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

Time Allocation. Respondents reported how they allocated their time "between exploratory versus confirmatory research" on a scale from 0% exploratory to 100% exploratory, or 0% confirmatory to 100% confirmatory (randomly assigned).

Satisfaction with Behavioral Science. Respondents answered three questions assessing their satisfaction with behavioral science 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)?"

Demographics. Respondents reported the number of studies that they had run in the past 12 months, which provided us with a conceptual validation of our research inhibition measure. Consistent with the anecdotal evidence presented above, respondents who report higher research

inhibition should run fewer studies. Respondents then reported the percentage of these studies that were pre-registered. Respondents also reported the year (actual or expected) of their PhD, their current role (i.e., graduate student, post-doctoral student, assistant professor, associate professor-untentured, associate professor-tentured, full professor, or other), sub-discipline, primary research methods (e.g., experimental research), and gender.

Results

First, we present descriptive statistics about the extent to which respondents reported research inhibition, and whether research inhibition predicts number of studies run. Next, we describe who is most likely to report higher levels of research inhibition. Then, we assess the subjective experience of conducting exploratory vs. confirmatory research. Our full data set and survey instrument are available here:

(https://osf.io/9w8a3/?view_only=7392d5f3be7f4f7cb4313330d29992a9).

Demographics. Our final sample consisted primarily of graduate students and post-docs (63%), was 50% male and on average, had graduated in 2015 ($N = 400$); See Table 1 for full sample demographics. Nearly all of the respondents we surveyed (94%), reported that they primarily conducted experiments (either lab or field), and the majority of respondents identified their subdiscipline as cognitive psychology (25%), consumer behavior (18%), or social psychology (12%). On average, respondents reported running 11.40 studies in the last year ($SD = 14.26$). Overall, about 40% of our respondents reported not pre-registering. Among those who did pre-register, they pre-registered about half of their studies (46.31%; $SD = 35.77\%$).

Table 1. Sample Demographics

	Study 1	Study 2
PhD Year (Actual or Expected)		
	<i>M</i> = 2015 (<i>SD</i> = 10), Median = 2020	<i>M</i> = 2011 (<i>SD</i> = 15), Median = 2018
Use Experimental Methods		
Yes	94%	86%
No	6%	14%
Gender		
Man	50%	42%
Woman	50%	56%
Non-Binary	0%	2%
Role		
Grad Student	52%	47%
Post-Doc	11%	1%
Assistant Prof	13%	14%
Associate Prof, Untenured	3%	3%
Associate Prof, Tenured	9%	14%
Full Professor	12%	14%
Other	0%	6%
Pre-Registration		
Yes	60%	54%
No	40%	46%
Sub-Discipline		
Cognitive Psychology	25%	9%
Consumer Behavior	18%	20%
Social Psychology	12%	41%
Behavioral Economics	9%	<1%
Micro-Organizational Behavior	5%	4%
Experimental Economics	2%	0%
Personality Psychology	1%	4%
Macro-Organizational Behavior	<1%	<1%
Interdisciplinary	16%	NA ¹
Other	11%	21%

Research Inhibition

Psychometric properties. We took the average of the six items to calculate an index of research inhibition. Cronbach's alpha was acceptable ($\alpha = 0.67$; Tavakol & Dennick, 2011). As

¹ The option to select multiple sub-disciplines was not provided in Study 2.

expected, the more apprehensive someone was about conducting research (i.e., the higher the score on the research inhibition scale), the fewer number of studies they had run in the past 12 months, $B = -2.24$, $SE = 0.73$, $p = 0.002$. This negative relationship held controlling for gender, tenure, PhD year, and pre-registration rates ($B = -3.04$, $SE = 0.85$, $p < .001$).

Who experiences research inhibition? Research inhibition was higher among female respondents ($M = 3.80$ out of 7, $SD = 0.98$) than male respondents ($M = 3.50$, $SD = 1.06$), $t(347.66) = 2.76$, $p = 0.006$, $d = 0.29$, 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 research inhibition than those who obtained their PhD in or before 2011 ($M_{\text{post}_{2011}} = 3.79$, $SD = 1.01$; $M_{\text{pre}_{2011}} = 3.18$, $SD = 0.97$), $t(114.21) = 4.75$, $p < .001$, $d = 0.62$, 95% $CI [0.35, 0.88]$. Similarly, tenured respondents reported significantly lower feelings of research inhibition than untenured respondents ($M_{\text{tenured}} = 3.11$, $SD = 0.89$; $M_{\text{untenured}} = 3.84$, $SD = 1.03$), $t(111.94) = 5.67$, $p < .001$, $d = 0.73$, 95% $CI [0.45, 1.01]$. Lastly, respondents who reported pre-registering at least one study in the past year reported similar levels of research inhibition as those who did not report pre-registering ($M_{\text{no_preregister}} = 3.57$, $SD = 1.06$; $M_{\text{preregister}} = 3.69$, $SD = 1.01$), $t(290.39) = 1.03$, $p = 0.30$, $d = -0.11$, 95% $CI [-0.33, .10]$.

Importantly, we also found that research inhibition was a significant *negative* predictor of respondents' satisfaction with their current role ($b = -0.28$, $SE = 0.07$, $p < .001$), and this relationship held controlling for gender, tenure, and whether participants received their PhD before or after 2011: $b = -0.24$, $p = .002$; see SOM for more detailed exploratory analyses).

Subjective Experience of Exploration versus Confirmation

Descriptives and statistical significance tests are reported in Table 2. Relative to confirmatory research, respondents reported that exploratory research was more enjoyable,

motivating, and interesting, $ps < 0.001$. Respondents also reported that, relative to confirmatory research, exploratory research was less frustrating, anxiety-inducing, boring, and scientific, $ps < .03$. Overall, respondents reported more positive subjective experiences when engaging in exploratory vs. confirmatory research processes (Figure 1). See SOM for exploratory analyses that break down these results by various demographic characteristics.

Table 2. Comparisons between subjective experience of exploratory and confirmatory research.

	Confirmatory Research <i>M (SD)</i>	Exploratory Research <i>M (SD)</i>	Mean Difference <i>t</i>-test, Cohen's <i>d</i>, [95% <i>CI</i>]
Enjoyable	5.29 (1.10)	5.87 (1.04)	$t(340) = 6.18, p < .001, d = -0.45, [-0.60, -0.30]$
Motivating	5.23 (1.27)	5.85 (1.05)	$t(342) = 7.20, p < .001, d = -0.52, [-0.67, -0.37]$
Interesting	5.52 (1.14)	6.18 (0.99)	$t(340) = 8.42, p < .001, d = 0.62, [-0.77, -0.46]$
Frustrating	3.82 (1.64)	3.58 (1.78)	$t(340) = 2.40, p = .02, d = 0.62, [-0.77, -0.46]$
Anxiety- Inducing	4.17 (1.76)	3.55 (1.89)	$t(342) = 4.96, p < .001, d = 0.34, [0.20, 0.49]$
Boring	2.99 (1.47)	2.19 (1.26)	$t(342) = 8.32, p < .001, d = 0.59, [0.44, 0.74]$
Scientific	6.09 (0.97)	5.82 (1.14)	$t(341) = 3.82, p < .001, d = 0.24, [0.09, 0.37]$

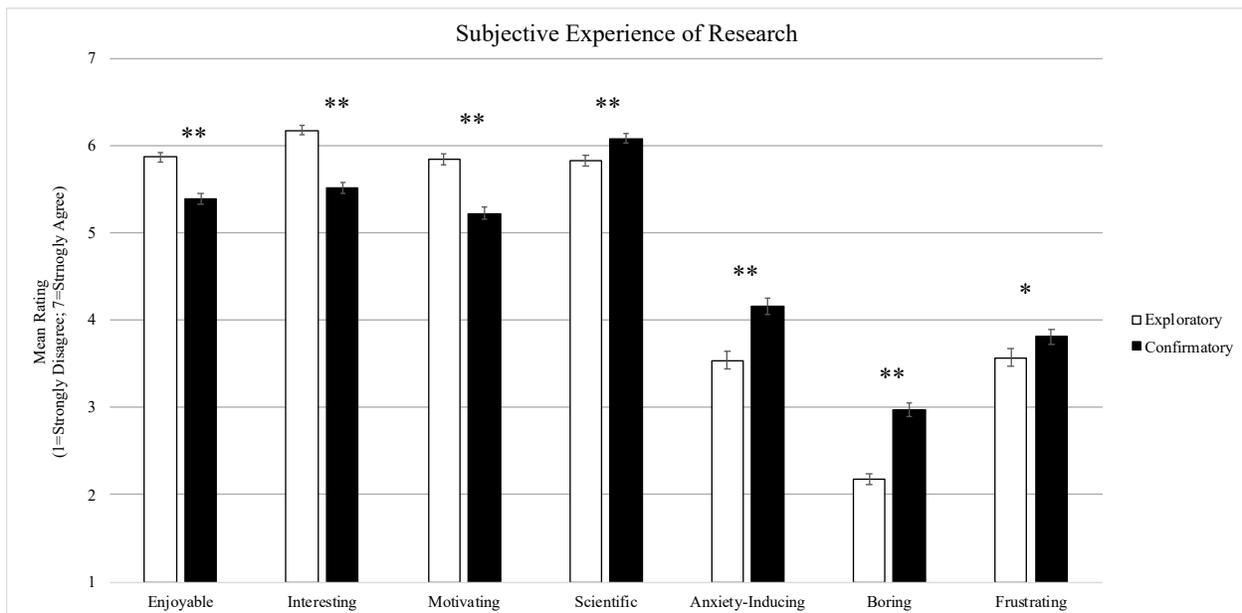


Figure 1. Respondent ratings of the subjective experience of confirmatory and exploratory research. Bars represent means and error bars represent standard errors. * $p < .05$; ** $p < .001$

time on confirmatory research, and 15% reporting that they spent at least 75% of their time on exploratory research.

Career Satisfaction Descriptive Statistics

Satisfaction with the field. Average satisfaction with the field of behavioral science was significantly higher than the scale midpoint ($M = 4.76$, $SD = 1.21$), $t(354) = 11.82$, $p < .001$, Cohen's $d = 0.63$, 95% $CI [0.51, 0.74]$ —67% of respondents reported satisfaction that was above the scale midpoint.

Satisfaction with role. Average satisfaction with one's role was also significantly higher than the scale midpoint ($M = 4.75$, $SD = 1.42$), $t(354) = 9.96$, $p < .001$, Cohen's $d = 0.53$, 95% $CI [0.42, 0.64]$ —61% of respondents reported satisfaction with their role that was above the scale midpoint.

Interest in staying in field. Average interest in staying in the field was well above the scale midpoint ($M = 5.86$, $SD = 1.34$), $t(354) = 26.39$, $p < .001$, Cohen's $d = 1.40$, 95% $CI [1.25, 1.54]$ —83% of respondents reported interest in staying that was above the scale midpoint.

Study 1 Discussion

In Study 1, researchers who felt greater research apprehension reported running fewer studies. Relatedly, researchers reported that exploration was more satisfying than confirmation. However, the researchers we studied did not necessarily spend more time on exploration. Building on these results, we conducted an experiment to assess the causal impact of confirmatory mindsets on exploratory behavior.

Study 2: Experiment

In Study 2, we conducted a pre-registered experiment to examine whether inducing a confirmatory mindset reduced researchers' propensity to explore, and if, so, whether reminding participants to explore their data could mitigate this effect.

All participants first imagined they had collected a data set. Two thirds of participants were told that they had pre-registered the prediction: people who did yoga are happier. In one condition, participants were reminded that they could still explore their data set—even though they had pre-registered their main prediction. Critically, we constructed the data set such that the predicted main effect was qualified by an interesting, albeit unpredicted, interaction: yoga was particularly beneficial for men's happiness. Thus, this study allowed us to examine whether a confirmation mindset undermined participants' likelihood of detecting and reporting the unpredicted interaction, and whether an exploration reminder mitigated these effects.

Method

Sample. We recruited behavioral science researchers to participate in a brief 7-10 minute survey in which they would simulate the research process. We recruited participants via email and list serves. Specifically, we invited 5,964 academic psychologists at major U.S. universities via email (using an updated contact list from John, Loewenstein, & Prelec, 2012),² and we posted the survey link to the SPSP Student Group, ACR listserv, and AOM OB Student Network. These efforts garnered 458 respondents, 84% of whom completed at least one of our key DVs. We analyzed all available data. An item at the end of the survey which asked people to optionally indicate how they had been recruited told us which source some of the respondents had come from (not all respondents answered this question): email list ($n = 134$ respondents),

² 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.

SPSP Student Group ($n = 173$ respondents), ACR listserv ($n = 78$ respondents), and AOM OB Student Network ($n = 0$ respondents).³

Procedure. Participants completed a simulated research process. First, participants imagined that they had collected a dataset consisting of a 1,000 people's answers to three questions: (1) "Do you do yoga on a weekly basis?" (0 = *No*; 1 = *Yes*); (2): "How happy are you today?" (1 = *Not at all*; 7 = *Extremely*); and (3): "What is your gender?" (0 = *Male*; 1 = *Female*).

Participants were randomized to one of three conditions: a control condition, a confirmation prime condition, or a hybrid prime condition (in which we sought to prime both the notions of confirmation *and* exploration). Participants in the confirmation prime and hybrid prime conditions were 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 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. After doing so, the next screen displayed the pre-registration form, with the pre-registration badge added as a watermark (Figure 3).

³ Since we recruited participants through multiple channels, we included a question asking participants whether they had completed the survey before. Only 2 people responded "yes" and 4 people responded "maybe." Everyone else reported that they had not completed the survey before. To provide a conservative test of our analyses, we included all participants in our final analyses. Results are statistically equivalent whether we include or exclude these cases.

<p>You enter the following pre-registration information into the online pre-registration tool:</p> <p>Pre-Registration for Study: "Yoga & Happiness"</p> <p>1) What's the main question being asked or hypothesis being tested in this study? 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>2) Describe the key dependent variable(s) specifying how they will be measured. Our key dependent variable is self-reported happiness on a scale from 1=Not at all to 7=Extremely.</p> <p>3) How many and which conditions will participants be assigned to? N/A</p> <p>4) Specify exactly which analyses you will conduct to examine the main question/hypothesis. 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>5) How many observations will be collected or what will determine sample size? We will survey 1,000 people.</p> <p>6) Anything else you would like to pre-register? (e.g., data exclusions, variables collected for exploratory purposes, unusual analyses planned?) N/A</p> <p>7) Have any data been collected for this study already? No, no data have been collected for this study yet</p> <p>Please click on the pre-registration icon below to pre-register this study!</p>  <p style="text-align: right;"></p>	<p>Pre-Registration for Study: "Yoga & Happiness"</p> <p>1) What's the main question being asked or hypothesis being tested in this study? 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>2) Describe the key dependent variable(s) specifying how they will be measured. Our key dependent variable is self-reported happiness on a scale from 1=Not at all to 7=Extremely.</p> <p>3) How many and which conditions will participants be assigned to? N/A</p> <p>4) Specify exactly which analyses you will conduct to examine the main question/hypothesis. 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>5) How many observations will be collected or what will determine sample size? We will survey 1,000 people.</p> <p>6) Anything else you would like to pre-register? (e.g., data exclusions, variables collected for exploratory purposes, unusual analyses planned?) N/A</p> <p>7) Have any data been collected for this study already? 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 '->' to collect the data!</p> <p style="text-align: right;"></p>
---	--

Figure 3. Confirmation prime presented to participants in the 'confirmation prime' and 'hybrid prime' conditions.

Next, all participants in the confirmation and hybrid prime conditions 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.” For participants in the hybrid prime condition, the following text also appeared, in bold green font, at the top of the page: “REMEMBER: Pre-registering doesn't mean you can't explore the data!”

Analyses Viewed. All participants then 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.

Analyses Reported. On the next page, we presented the output of participants’ requested analyses in their preferred format (SPSS or R).⁴ The results of these analyses revealed that people who did yoga reported significantly greater happiness than those who did not. However, the effect of yoga on happiness depended on gender; it was more pronounced for men. Finally, participants chose the results that they would report in a final manuscript (see Figure 4).⁵

At the end of the survey, participants reported various demographic characteristics. Participants completed our 6-item research inhibition scale and the identical demographics items from Study 1 as follows: the number of studies run in the last 12 months and the percent pre-registered, year of PhD, role, sub-discipline, primary methods, and gender.

⁴ Participants who had selected “Other: Describe” were not able to view their additional requested analyses.

⁵ Participants were asked whether they wanted to view the results of the analyses in R or SPSS format and were shown the results accordingly.

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=dstudy_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:
 -1.505191 -1.260730
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=dstudy_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      4.268250
```

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.

Figure 4. Screenshot of the analysis selection and viewing procedure. 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 has selected to view the results in ‘R’ format.

Results

We first examine whether participants in the confirmation prime condition (vs. control) view fewer analyses overall and are less likely to view and report the results of the gender interaction. We also explore whether reminders of exploration mitigate such effects. Finally, we investigate whether these effects are moderated by participants’ level of research inhibition, and other important demographics (e.g., gender, post-bem, tenure, whether or not they pre-register). This study was pre-registered (<https://aspredicted.org/blind2.php>), and all data and materials are available online (https://osf.io/9w8a3/?view_only=7392d5f3be7f4f7cb4313330d29992a9).

Demographics. The majority of participants were graduate students and post-docs (48%), 42% were male, and 2011 was the average year of PhD completion (See Table 2 for

sample demographics). The majority of participants reported primarily conducting experiments (86%) and identified their sub-discipline as social psychology (41%) or consumer behavior (20%). On average, participants reported running 8.74 studies in the last 12 months ($SD = 11.75$). Overall, 46% reported not pre-registering any of their studies. Among those who did, they pre-registered an average of 43.14% ($SD = 35.81\%$) of their studies.

DV: Number of Analyses Viewed. Across conditions, participants viewed an average of 4.23 (out of 6) analyses. Using a negative binomial logistic regression (Hilbe, 2011), there was no difference across conditions in the number of analyses participants viewed ($ps > 0.48$).

DV: Viewing Results of Interaction. We assessed differences between conditions 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 as compared to 69% in the control condition, $b = -0.70$, $SE = 0.24$, $p = .01$. However, the exploration reminder did not dampen this effect: 57% of participants in the hybrid condition requested the interaction—which did not significantly differ from the confirmatory condition, $b = 0.19$, $SE = 0.23$, $p = .70$.

DV: Reporting Results of 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 control condition were more likely to report the interaction relative to those in the confirmatory condition, $b = -0.73$, $SE = 0.24$, $p = .002$, as well as the hybrid condition, $b = -0.54$, $SE = 0.24$, $p = .02$. Not surprisingly, this was driven by a larger proportion of participants in the control condition choosing to view the interaction. Indeed, among the participants who chose to view the results of the interaction, 92% of them decided to report the gender interaction in their final write-up, and this tendency did not differ by condition, $ps > 0.50$.

All of the results reported above held controlling for demographics (SOM). We also examined whether our results differed across condition depending on position, PhD year, and research inhibition and found no consistent pattern of results. See SOM.

Study 2 Discussion

In Study 2, a confirmatory mindset made researchers less likely to view a qualifying interaction effect in their data. The inclusion of a simple reminder to explore did not mitigate this effect. These results provide further evidence that, to the extent that open science reforms are focusing researchers on confirmation, they may be (unintentionally) inhibiting exploration.

General Discussion

Against the backdrop of today's emphasis on open science practices, we examined a diverse group of behavioral scientists' subjective experiences doing research (Study 1), as well as their propensity to engage in exploration (Study 2).

In Study 1, researchers derived greater enjoyment from exploration—despite spending an equivalent amount of time engaged in exploratory (vs. confirmatory) research processes. And although both confirmation *and* exploration are integral scientific activities, researchers deemed exploration to be less scientific than confirmation. Though fortunately, with its mean rating of 5.82 out of 7, people generally find exploration to be scientific, albeit not quite as scientific as confirmation, which had a mean rating of 6.09. In Study 1, we also developed and administered a scale designed to measure research inhibition. Attesting to the scale's predictive validity, researchers who felt greater apprehension conducted fewer studies than their less apprehensive counterparts—and, as one might expect, this apprehension was most common among women, untenured professors, and researchers who received their PhDs after 2011—when numerous scientific reforms were taking place. Research inhibition also predicted academic satisfaction:

those who experienced less research inhibition felt more satisfied with their current roles.

However, given that it did not predict the other two satisfaction measures, more research is needed to understand when and how research inhibition shapes people's career satisfaction.

In Study 2, we followed up on this initial survey by examining whether open science practices focused on confirmation (e.g., pre-registration) had a causal effect on exploration. Consistent with this hypothesis, participants assigned to the confirmation condition were significantly *less* likely to find an unpredicted interaction effect in their data—which fundamentally changed the interpretation of their results. Those in the confirmation condition who did ask to see the interaction were just as likely to report it compared to those in the control condition, suggesting that confirmation specifically impedes the act of discovery, as opposed to instilling reservations about reporting those discoveries. The effect of the confirmation prime was unchanged when participants were shown a brief reminder that pre-registration does not preclude exploration.

It is possible that the exploration reminder does have an effect—with it, participants were five percentage points more likely to ask for the interaction than those in the confirmation condition—but that its effect was so small that it did not reach statistical significance in our data. If however, it is a true null result, it implies that strong norms about confirmation may undermine scientific discovery, raising the question of how to foster exploration in the face of rigorous open science practices. In that case, research could test whether different types of exploration reminders may be more effective than the one we tested. For example, research could test a positively-framed reminder, such as “Reminder: Although you pre-registered hypotheses, you can still explore your data!” or a very simple one, such as: “Remember to explore your data!” On the other hand, it is possible that these explicit reminders—such as the ‘exploratory analyses’

section on many pre-registration forms might not be enough. Speaking of which, these sections could potentially be counterproductive. Conducting exploratory analyses—by definition—means conducting analyses that could lead to unpredicted subsequent analyses. By having an explicit section for exploratory analyses, people might feel as though they cannot conduct exploratory analyses that were not pre-registered. Future research should therefore further examine how to increase exploration in the context of pre-registered studies.

A number of other interesting patterns emerged from our data that are worthy of further exploration. In Study 1, the percentage of studies that our respondents pre-registered did not significantly predict research inhibition. This result suggests that engagement in open science practices such as pre-registration may not, in and of themselves, foster research inhibition. In this vein, we wonder whether it may be the way we discuss and communicate scientific reforms that fosters research inhibition and anxieties more so than the reforms themselves. Indeed, anyone—even the most upstanding researcher—who has received one of those litigiously-worded inquiries from large-scale replication projects can surely relate to this point. Indeed, some worry that the narrative around open science can be adversarial in tone, and, moreover, that this negative tone can be particularly detrimental to scholars of color, women, and junior academics (Fiske, 2017). Consistent with this line of thinking, we found research inhibition to be stronger among women and junior scholars. More work is needed to understand factors that protect all scientists from the negative effects of research inhibition (e.g., running fewer studies, deriving less satisfaction from one's role).

Our results show—unsurprisingly to us—that researchers enjoy exploration. In fact, exploration is why most of us became scientists in the first place. Of course, researchers also enjoy confirming their hypotheses and methodological rigor—we want our scientific results to

hold to the highest level of scrutiny. Yet, as we have observed, our field's increased focus on confirmation may have negative unintended consequences for researchers' enjoyment of the research process, willingness to explore data, and self-reported satisfaction with their careers.

We hope this investigation reminds us of the vital, and mutually-reinforcing nature of confirmatory *and* exploratory research. The next time your colleague expresses concern over conducting a study with a new and "imperfect" paradigm, encourage them to do so—they can even pre-register their new foray. Remind them that, in time, this exploratory work will become confirmatory and they need not compromise joy for rigor.

References

- Asendorpf, J., Conner, M., De Fruyt, F., De Houwer, J., & Denissen, J. (2016). Recommendations for increasing replicability in psychology. *Methodological issues and strategies in clinical research* (pp. 607-622). American Psychological Association.
- Bellet, C. S., De Neve, J. E., & Ward, G. (2019). Does Employee Happiness Have an Impact on Productivity? *Working Paper, University of Oxford*.
- Bem, D. J. (2011). Feeling the future: experimental evidence for anomalous retroactive influences on cognition and affect. *Journal of Personality and Social Psychology, 100*(3), 407.
- 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.
- Carey, B. (2011). Journal's paper on ESP expected to prompt outrage. *The New York Times A, 1*.
- Chambers, C. (2019). What's next for Registered Reports?. *Nature, 573*(7773), 187.
- 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.
- Cumming, G. (2014). The new statistics: Why and how. *Psychological Science, 25*(1), 7-29.
- Dexter, F., & Shafer, S. L. (2017). Narrative review of statistical reporting checklists, mandatory statistical editing, and rectifying common problems in the reporting of scientific articles. *Anesthesia & Analgesia, 124*(3), 943-947.
- 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.

- Eich, E. Business not as usual. (2014). *Psychological Science*, 3-6.
- Fanelli, D. (2018). Opinion: Is science really facing a reproducibility crisis, and do we need it to? *Proceedings of the National Academy of Sciences*, 115(11), 2628-2631.
- Fiske S. T. (2017). Going in Many Right Directions, All at Once. *Perspectives on Psychological Science*, 12(4), 652–655. doi:10.1177/1745691617706506
- Galak, J., LeBoeuf, R. A., Nelson, L. D., & Simmons, J. P. (2012). Correcting the past: Failures to replicate psi. *Journal of Personality and Social Psychology*, 103(6), 933.
- Gilbert, D. T., King, G., Pettigrew, S., & Wilson, T. D. (2016). Comment on “Estimating the reproducibility of psychological science”. *Science*, 351(6277), 1037-1037.
- Giofrè, D., Cumming, G., Fresc, L., Boedker, I., & Tressoldi, P. (2017). The influence of journal submission guidelines on authors' reporting of statistics and use of open research practices. *PLoS One*, 12(4), e0175583.
- Goldin-Meadow, S. (2016). Why preregistration makes me nervous. *APS Observer*, 29(7).
- Ioannidis, J. P., Fanelli, D., Dunne, D. D., & Goodman, S. N. (2015). Meta-research: evaluation and improvement of research methods and practices. *PLoS Biology*, 13(10), e1002264.
- 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.
- Kidwell, M. C., Lazarević, L. B., Baranski, E., Hardwicke, T. E., Piechowski, S., Falkenberg, L. S., ... & Errington, T. M. (2016). Badges to acknowledge open practices: A simple, low-cost, effective method for increasing transparency. *PLoS biology*, 14(5), e1002456.

- 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.
- Lindsay, D. S. (2015). Replication in psychological science. 1827-1832. *Psychological Science*.
- Lindsay, D. S. (2017). Sharing data and materials in Psychological Science. *Psychological Science*.
- 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.
- McKiernan, E. C., Bourne, P. E., Brown, C. T., Buck, S., Kenall, A., Lin, J., ... & Spies, J. R. (2016). Point of view: How open science helps researchers succeed. *Elife*, 5, e16800.
- 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.
- 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.

- Nosek, B. A., Ebersole, C. R., DeHaven, A. C., & Mellor, D. T. (2018). The preregistration revolution. *Proceedings of the National Academy of Sciences, 115*(11), 2600-2606.
- Nuijten, M. B., Borghuis, J., Veldkamp, C. L., Dominguez-Alvarez, L., Van Assen, M. A., & Wicherts, J. M. (2017). Journal data sharing policies and statistical reporting inconsistencies in psychology. *Collabra: Psychology, 3*(1).
- 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.
- Sassenberg, K., & Ditrich, L. (2019). Research in social psychology changed between 2011 and 2016: larger sample sizes, more self-report measures, and more online studies. *Advances in Methods and Practices in Psychological Science, 2515245919838781*.
- Smaldino, P. E., & McElreath, R. (2016). The natural selection of bad science. *Royal Society Open Science, 3*(9), 160384.
- Simonsohn, U. (2013). Just post it: The lesson from two cases of fabricated data detected by statistics alone. *Psychological Science, 24*(10), 1875-1888.
- 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.
- Spellman, B. A. (2015). A short (personal) future history of revolution 2.0. *Perspectives on Psychological Science*.
- Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education, 2*, 53.

- Trafimow, D., & Marks, M. (2015). Editorial in Basic and Applied Social Psychology. *Basic and Applied Social Psychology, 37*, 1-2.
- 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.
- 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.
- Veldkamp, C. L., Nuijten, M. B., Dominguez-Alvarez, L., van Assen, M. A., & Wicherts, J. M. (2014). Statistical reporting errors and collaboration on statistical analyses in psychological science. *PloS one, 9*(12), e114876.
- Wagenmakers, E. J., Wetzels, R., Borsboom, D., & van der Maas, H. L. (2011). Why psychologists must change the way they analyze their data: the case of psi: comment on Bem (2011). *Journal of Personality and Social Psychology, 100*(3), 426.
- 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.
- Wicherts, J. M., Borsboom, D., Kats, J., & Molenaar, D. (2006). The poor availability of psychological research data for reanalysis. *American Psychologist, 61*(7), 726.
- 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.

Yarkoni, T. (2016). There is no “tone” problem in psychology. Retrieved December 25, 2017.

<https://www.talyarkoni.org/blog/2016/10/01/there-is-no-tone-problem-in-psychology/>

Appendix A: Research Inhibition 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 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").