

Methodology and Research Practice

# How Do Researchers in Psychology Perceive the Field? A Qualitative Exploration of Critiques and Defenses

Jacob F. Miranda<sup>1</sup> <sup>a</sup>, Cassie M. Whitt<sup>1</sup>, Alexander McDiarmid<sup>1</sup>, Jeremy E. Stephens<sup>1</sup>, Dillon Purdue<sup>1</sup>, Clayton Hall<sup>1</sup>, Alexa M. Tullett<sup>1</sup>

<sup>1</sup> Department of Psychology, The University of Alabama, Tuscaloosa, AL, US

Keywords: replication, reproducibility, metascience, open data, inductive coding, justification

<https://doi.org/10.1525/collabra.35711>

---

## Collabra: Psychology

Vol. 8, Issue 1, 2022

---

As awareness of the replication crisis in psychology has become increasingly widespread, several meta-scientific investigations have focused on the research practices and attitudes of researchers in psychology. Here, we aimed to add to this body of work by exploring academic psychologists' perceptions of the state of the field using both quantitative and qualitative approaches. As part of a larger project, psychological researchers ( $N = 548$ ) used 3-point scales to rate their perceptions of: 1) the rate of false positive findings in psychology and 2) the quality of research practices in psychology. They then wrote about the reasons for their ratings. Using a qualitative approach, we assessed the prevalence of criticisms and defenses of the field, as well as subtypes of each. Overall, these data shed light on the extent, and nature, of concerns about false positives and research practices within the psychological community.

In 2010, social psychologist Daryl Bem published a paper claiming that human beings could see the future. Perhaps unsurprisingly, the paper invoked a significant amount of criticism and doubt (Bhattacharjee, 2012; Engber, 2017; French, 2012; Wagenmakers et al., 2011). Soon after, Simmons et al. (2011) identified data collection and analysis practices that can vastly inflate the risk of false positives (*p*-hacking), and John et al. (2012) showed that the use of such practices was frequent among psychologists. Large scale replication projects began to yield disappointing results, reinforcing concerns that common statistical practices may be yielding unreplicable findings (Camerer et al., 2018; Ebersole et al., 2020; Klein et al., 2018; Open Science Collaboration, 2015). These events foreshadowed what would become commonly referred to as the “replication crisis.”

Some scientists responded to these events, and the metascientific investigations that followed, by concluding that the trustworthiness of the psychological literature should be seriously questioned. For instance, Simmons and colleagues (2011) stated that “in many cases, a researcher is more likely to falsely find evidence that an effect exists than to correctly find evidence that it does not” (p. 1359). They argued that without changes in reporting practices, the credibility of the profession would be at stake. The *Reproducibility Project: Psychology*, which documented replications of 100 empirical studies, found that 36% of replications (versus 97% of original studies) reported statistically significant findings, leading the authors to conclude that

“there is room to improve reproducibility in psychology” (pp. aac4716-7). In slightly less diplomatic terms, Srivastava (2016) wrote a blog post detailing a syllabus for a disconcertingly plausible college psychology course: “Everything is fucked.”

Others, however, have argued that there is no need to worry about the reproducibility of psychological research (Gilbert et al., 2016) and that the crisis has put some popular findings under “undeservedly heavy attack” (Buffalmano, 2019). Dr. Leonard Martin warned that sloppy replicators may create psychology’s “version of the McCarthy era,” in which the replicators tear down established findings and researchers (as quoted in Engber, 2016). In her op-ed for *The New York Times*, Lisa Feldman-Barrett claimed that “psychology is not in crisis” and that “contrary to the implication of the Reproducibility Project, there is no replication crisis in psychology. The ‘crisis’ may simply be the result of a misunderstanding of what science is” (Feldman-Barrett, 2015). Gilbert and colleagues (2016) conducted a reanalysis of the *Reproducibility Project: Psychology* and concluded that “OSC’s [Open Science Collaboration] data clearly provide no evidence for a ‘replication crisis’ in psychological science” (p. 1037b). Thus, alongside psychologists who have raised serious criticisms of the field are those who have offered up vigorous defenses.

The primary aim of this project is to conduct a qualitative analysis of academic psychologists’ perceptions of the field, with specific reference to the rate of false positives in the literature and the quality of research practices. Our goal

---

a jfmiranda@crimson.ua.edu

is descriptive, rather than experimental; gaining a better understanding of the frequency and kind of concerns and commendations seems fundamental to the broader goal of improving the field (Rozin, 2001; Yarkoni, 2020). In sum, our work was guided by the following research question: what are the major themes that emerge when psychology researchers verbally explain their views about false positives and research practices, and what is the frequency of those themes?

We believe the knowledge gained through probing these attitudes can help improve psychological research by bringing the concerns of researchers to light. Specifically, it is difficult to fix a problem if a field cannot agree on what the problem is to begin with. Our data provides information about the prevalence of different types of criticisms among academic psychologists, allowing the field to focus more time and resources on more widespread concerns.

Our work fits within an emerging tradition of metascientific research that examines psychologists' research practices and perceptions of the field. For example, researchers have turned the scientific lens on psychological scientists to explore barriers to data sharing (Houtkoop et al., 2018), strategies for predicting replication results (Dreber et al., 2015), and attitudes towards researchers whose results fail to replicate (Ebersole et al., 2016; Fetterman & Sassenberg, 2015). By scientifically examining the operation of the field, these projects shed light on the most effective strategies for improving research practices.

Most relevant to the present project, Washburn et al. (2018) used quantitative and qualitative methods to explore social and personality psychologists' resistance to proposed reforms to research practices (PRRPs). They observed that there is significant disagreement about the risks and utility of such practices. Similarly, Baker (2016) conducted a quantitative survey of *Nature* members and asked questions about open science practices and the existence of a "replication crisis". This survey was one of the first to identify widespread concern about replicability among scientists. Here, we adopted, and built upon, elements of both aforementioned projects; we used quantitative and qualitative approaches to examine psychologists' perceptions of the state of the field. In addition to capturing offered criticisms and defenses of the field, our dataset includes a diverse sample of academic psychologists including those from social, personality, developmental, clinical, and cognitive areas. Thus, these data provide new information about how psychologists feel about the quality of research practices and published findings in their field.

## Disclosures

### Ethical approval

The research reported in this paper was conducted under the IRB of the Office for Research Compliance at the University of Alabama (Protocol: 16-OR-273-ME-R2). The data for this study is coming from a larger research project funded by the National Science Foundation, Award #: 1728332 (See [link](#)).

## Method

### Participants

Participants were practicing psychological researchers (i.e., current graduate students or individuals holding a graduate degree in psychology) who were recruited at psychology conferences and via email listservs as part of a larger project in which the goal was to investigate academic psychologists' belief updating in light of new evidence (McDiarmid et al., 2021). The conferences were in the United States and Canada and the listservs have a predominantly North American membership (see Supplementary Materials for specific recruitment information). This larger study included data from two time-points: Phase I (November 2017 - March 2018) and Phase II (February 2019 - April 2019). Participants were compensated \$10 for completing Phase I and \$20 for completing Phase II. Here, we focus our analysis on three items from Phase II that assessed participants' attitudes regarding the state of the field of psychology. Our sub-sample consisted of the 548 participants who responded to all three items at Phase II. See [Table 1](#) for demographic information. The analyses conducted were exploratory in nature and thus do not support any confirmatory claims.

### Measures

Two questions addressed participants' perceptions of the state of the field. The first question asked participants to choose the most accurate statement regarding the rate of false positive findings: "*The rate of false positive findings in the published psychology literature is acceptably low*" (Choice 1), "*The rate of false positive findings in the published psychology literature is somewhat higher than it should be, and we should try to lower it*" (Choice 2), or "*The rate of false positive findings in the published psychology literature is much higher than it should be and we should take major steps to lower it*" (Choice 3).

The second question asked participants to choose the most accurate statement regarding research practices in psychology: "*Our research practices were fine before the crisis, and we should continue employing more or less the same practices*" (Choice 1), "*Our research practices have been shown to be somewhat worse than we thought, and we should make some incremental changes*" (Choice 2), "*Our research practices have been shown to be extremely problematic, and we should make profound changes.*" (Choice 3).

The third question asked participants to elaborate on their responses to the previous two questions in an open-ended fashion.

### Qualitative Approach: Combination of Thematic and Content Analysis

Our approach to coding aligns with recommendations for thematic coding (Braun & Clarke, 2006), and for content analysis (Hsieh & Shannon, 2005). We took a blended approach between the two to capitalize on the advantages of both approaches. Braun and Clarke (2006) write that the biggest advantage of thematic coding is its flexibility. Simultaneously, they note that this flexibility can make it

**Table 1. Demographic Comparison of Phase II Sub-Sample and Overall Sample**

Demographic Variable	Phase II sub-sample		Phase II overall sample			
	<i>n</i>	%	<i>N</i>	%		
<b>Gender</b> ( <i>n</i> = 548, <i>N</i> = 1240)						
Female	351	64.1	843	68		
Male	187	34.1	379	30.6		
Non-Binary or 'Other'	10	1.8	18	1.5		
<b>Racial identity</b> ( <i>n</i> = 545, <i>N</i> = 1233)						
African American or Black	6	1.1	22	1.8		
American Indian or Alaskan Native <sup>a</sup>	0	0	0	0		
Asian American or Pacific Islander	57	10.5	134	10.9		
White	441	80.9	982	79.6		
'Other' or selected more than 1 identity	41	7.5	95	7.7		
<b>Hispanic/LatinX identity</b> ( <i>n</i> = 545, <i>N</i> = 1238)						
Yes	36	6.6	90	7.3		
<b>Political Ideology</b> ( <i>n</i> = 547, <i>N</i> = 1240)						
Very conservative	3	0.5	7	0.6		
Conservative	14	2.6	32	2.6		
Moderate	64	11.7	148	11.9		
Liberal	261	47.7	575	46.4		
Very liberal	185	33.8	435	35.1		
None	20	3.7	43	3.5		
<b>Career stage</b> ( <i>n</i> = 548, <i>N</i> = 1242)						
Full Professor	18	3.3	44	3.5		
Associate Professor	35	6.4	75	6		
Assistant Professor	77	14.1	169	13.6		
Lecturer	12	2.2	23	1.9		
Post-doc	66	12	146	11.8		
Graduate student	292	53.3	661	53.2		
Undergraduate student or 'Other'	48	8.8	124	10		
<b>Psychology sub-field</b> ( <i>n</i> = 548, <i>N</i> = 1242)						
Social Psychology	185	33.8	388	31.2		
Personality Psychology	9	1.6	31	2.5		
Cognitive Psychology	132	24.1	271	21.8		
Developmental Psychology	77	14.1	200	16.1		
Clinical Psychology	77	14.1	229	18.4		
Quantitative Psychology	6	1.1	11	0.9		
Industrial and Organizational Psychology	11	2	13	1		
Not applicable or 'Other'	51	9.3	99	7.9		
	Phase II sub-sample			Phase II overall sample		
	Range	<i>M</i>	<i>SD</i>	Range	<i>M</i>	<i>SD</i>
Age ( <i>n</i> = 542, <i>N</i> = 1226)	22-74	31.5	7.7	21-74	31.5	7.8

*Note.* The lowercase '*n*' represents the subset of people who responded to the free response question, out of the total '*N*' of participants in the broader dataset. When reporting age, data from 2 participants were excluded for reporting impossible ages of 321 and 443. <sup>a</sup>All 7 Native American participants reported multiple racial identities.

Downloaded from [http://online.sagepub.com/journals/colab/colab/135711/706571/colab/2022\\_8\\_1\\_35711.pdf](http://online.sagepub.com/journals/colab/colab/135711/706571/colab/2022_8_1_35711.pdf) by guest on 10 August 2022

difficult to develop guidelines for researchers, resulting in inconsistencies. Hsieh and Shannon (2005) write that the biggest advantage of content analysis is that it allows for rules based on the words used by participants. Meanwhile, this makes it slightly more rigid. In this blended approach, we believe we have been able to capitalize on the advantages of the approaches while mitigating their disadvantages.

Braun and Clarke (2006) lay out six phases of thematic analysis: 1) become familiar with the data; 2) generate initial codes; 3) search for themes among the codes; 4) review themes; 5) define and name themes; and 6) produce a report. These steps guided our approach, as follows. For phase 1, we sampled 100 initial responses from our data set and had three research assistants independently read (and re-read) the open-ended responses following recommendations (e.g., Hsieh & Shannon, 2005). These qualitative responses were presented in isolation from any other information, including the quantitative responses to the perceived state of the field questions.

Our research group then met with the three coders and began phase 2. In line with recommendations by Hsieh and Shannon (2005), coders shared their first impressions, and our research group used those impressions to form initial codes after several conversations. For phase 3, we followed the recommendation of using a tree diagram to arrange lower order codes and higher-order themes (e.g., Hsieh & Shannon, 2005). Two higher-order themes emerged: 1) criticisms of the field and 2) defenses of the field. The higher-order theme of criticisms included four sub-types: 1.1) statistics and/or research methods are misused and/or misunderstood, 1.2) problematic incentives in the field, 1.3) need for more transparency (e.g., open data, practices, pre-registration) and 1.4) other criticisms. The higher-order theme of defenses included five sub-types: 2.1) criticisms of replications (e.g., contextual differences), 2.2) criticisms that people are overreacting, 2.3) emphasizing psychology's safeguards that protect the field (e.g., self-correcting nature of science), 2.4) criticisms of open science advocates and/or replicators, and 2.5) other defenses.

For phases 4 and 5, our research group discussed a subset of responses in detail to clarify theme definitions and the coding process. A list of open science terminology was created for coders and frame of reference training was utilized to improve inter-rater reliability (Jackson et al., 2005). If a sub-theme was coded as present, the higher-order theme was automatically coded as present. The categories were not considered mutually exclusive, and some statements included both defenses and criticisms. As an example, one response said: "There are definitely issues in research practices used in the past that we should correct...definitely significant improvements can be made...", and "However, I don't believe these issues represent a 'crisis'... most researchers are already doing many things well." Some responses received no codes as they were not germane and/or were too ambiguous to be interpretable. For example, one response simply stated, "Replication studies are increasingly important." In this case, it was not clear if the participant was criticizing or defending the field.

At this point, two researchers coded the entire sample of responses. Inter-rater reliability (IRR) was calculated in

two ways. The first, more conservative approach, was to use the proportion of responses for which both coders said the theme was present over the total number of instances where at least one coder said the theme was present. This value was 79.3% for criticisms, and 47.4% for defenses. A more liberal approach was to use the proportion of responses for which both coders said the theme was present *and/or* absent over the total number of responses based off McAlister et al. (2017). This value was 81.8% for criticisms, and 85.4% for defenses. Disagreements in coding were resolved by a third coder. The prevalence of criticisms and defenses are reported in [Table 2](#) and [Table 3](#), respectively. See [Table 4](#) for liberal and conservative IRRs for all sub-themes. The sixth and final phase of thematic analysis is producing the report which manifests within this article.

## Results

### Perceptions Regarding the Rate of False Positives in Psychology

Twenty-three participants (4.2%) selected "*the rate of false positive findings in the psychology literature is acceptably low*," 332 participants (60.6%) selected "*the false positive rate is somewhat higher than it should be and there should be some attempt to lower it*," and 193 participants (35.2%) selected "*the false positive rate is much higher than it should be and major steps should be taken*." See [Table 5](#) for a comparison of the prevalence of the coded themes between the three possible responses to this question.

### Perceptions Regarding Research Practices in Psychology

Eleven participants (2%) selected "*research practices were fine before the crisis and the field should continue past practices as they have traditionally been implemented*," 329 participants (60%) selected "*past research practices are somewhat worse than people once thought, and that incremental changes are needed*," and 208 participants (38%) selected "*past practices are extremely problematic and there is a need for profound change*." See [Table 6](#) for a comparison of the prevalence of the coded themes between the three possible responses to this question. Those who give more critical responses to the quantitative items include more criticisms in their qualitative responses, providing some additional evidence of the construct validity of our coding method.

### Prevalence of Criticisms

Of the 548 responses, 462 (84.3%) included a criticism of the field. The most common criticism was that statistics and/or research methods are misused and/or misunderstood. This theme was present in 218 (39.8%) responses and includes claims that researchers: inappropriately use statistical techniques, lack understanding of validity and/or reliability of measures, make improper claims from their data, and use questionable research practices (e.g., *p*-hacking, hypothesizing after the results are known (HARKING), etc.

The second most common criticism was problematic incentives within the field. This theme was present in 214 (39.1%) responses and includes claims that there are: issues

**Table 2. Frequencies and Examples of Criticisms**

Theme	Prevalence	Inter-Rater Reliability (IRR)	Example quote
Theme 1: Criticisms of the Field (N = 548)	462 (84.3%)	79.3%	
1.1: Statistics and/or research methods are misused and/or misunderstood	218 (39.8%)	66.9%	"Problematic and/or poor research practices seem rampant within the field, and this problem seems to stem from a poor understanding of statistics among many psychological researchers."
1.2: Problematic incentives in the field	214 (39.1%)	82.6%	"The whole atmosphere of the academy, in my opinion, rewards expeditious and fruitful studies. Thus it systematically reinforces lower sample sizes with unknown effect sizes (hiding within statistical significance)"
1.3: There is a need for more transparency	100 (18.2%)	81.4%	"Generally, most research methods used today are fairly sound, but the biggest changes should be transparency (to reduce p-hacking and general fishing). Overall, the culture is more the issue than the methods, in my opinion."
1.4: Other Criticisms	80 (14.6%)	13.2%	"I think there is currently a large problem and in order to fix it, we have to make significant changes or else it will be easy for the changes to get lost and not have much of an effect."

Note. The percentages indicate the number of free responses which contained a certain criticism theme divided by the total number free responses collected.

within the hiring process (prioritizing high numbers of publications), biases for novelty and counterintuitive results in publication, a file-drawer problem that is obscuring a large number of null results, and a lack of incentives for conducting replication research.

The third most common criticism was a need for more transparency. This theme was present in 100 (18.2%) responses and includes recommendations that: data, syntax, and code should be open, methodology should be clear, statistical decisions should be disclosed and justified, and pre-registration and registered reports should be more common.

Lastly, there were responses that communicated some critique of the field that did not fall into the previous categories. These were coded as 'other' criticisms and were present in 80 (14.6%) of responses.

### Prevalence of Defenses

Of the 548 responses, 124 (22.6%) included a defense of the field. Tied as the most common sub-themes of defenses were criticisms of replications, and claims that people are overreacting in the field. These defenses were both present in 40 (3%) responses. The former includes claims that replications disregard context and fail to measure hidden moderators. The latter includes claims that replication failures are exaggerated and throwing away the credibility of the whole field is not appropriate.

The third most common defense of the field was that psychology has safeguards to protect the field. This theme was present in 13 (2.4%) responses and include claims that science

is naturally self-correcting process, statistical techniques (e.g., meta-analyses) aid in self-correction, and the detection of false-positives in our field is evidence of self-correction in action.

The fourth most common sub-theme of defenses were those that included criticisms of open science advocates and/or replicators. This theme was present in 10 (1.8%) responses and includes claims that replication advocates are spot-lighting replication failures thus hurting trust in the field, that there may be some political agenda to go after the original authors (e.g., 'predatory nature' of current replication efforts), and that replicators may be less competent scientists for not performing original work.

Lastly, there were responses that communicated a defense of the field that did not fall into the previous categories. These were coded as 'other' defenses and were present in (6.9%) responses.

### Discussion

With this project, we aimed to document academic psychologists' perceptions on the state of the field, in addition to their justifications for those perceptions. The prevalence of criticisms, present in roughly 85% of responses, outweighed the prevalence of defenses, which were present in roughly 22% of responses. This suggests that our sample generally held critical views of the field and that the source of this criticism largely came from concerns that people (mis)understand/(mis)apply statistics and methodology and face problematic incentives. Among the responses that defended the field, no single defense seems to stand out, with none reaching prevalence levels over 10%.

These results should not be interpreted as the final word regarding the prevalence of psychological researchers' criticisms and/or defenses of the field. Our sample provides the benefit of focusing on practicing psychological researchers from a range of sub-disciplines but also possesses real limitations. For people to take our survey, they would have had to attend a conference or have been included on a psychological society's email listserv - both of which may be sub-

**Table 3. Frequencies and Examples of Defenses**

Theme	Prevalence	Inter-rater reliability (IRR)	Example quote
Theme 2: Defenses of the field (n = 548)	124 (22.6%)	47.4%	
2.1: Criticisms of how replications are conducted (e.g., contextual differences)	40 (7.3%)	62.2%	"I think people fail to consider contextual effects enough when talking about the replication crisis. There are so many changes over time and across populations that it is impossible to run an exact replication. Also, based on sheer probability, not all replication studies should be significant, even if the true effect does exist."
2.2: People are overreacting in the field	40 (7.3%)	64.7%	"The replicability crisis doesn't affect every sub-discipline of psychology equally, and so widespread panic is unnecessary. Identifying those sub-disciplines (or even research paradigms) that are especially problematic is a good idea, but over-criticizing the whole discipline casts doubt even in areas that need no doubt (especially for students). [I suspect that many of the areas where replications failed were already known to be problematic by people with frequent attendance at conferences.]"
2.3: Psychology has safeguards to protect the field (e.g., self-correcting nature of science).	13 (2.4%)	46.2%	"There are some positive steps researchers can take to improve basic research practices (e.g., replication, sharing materials), but the scientific method is relatively robust to many of the issues researchers [sic] are focused on."
2.4: Criticisms of open science advocates and/or replicators	10 (1.8%)	21.4%	"Not all researchers are equally competent. People who are going nuts about the so-called "replication crisis" are entirely ignorant of the fact that people who come up with original research are much more competent than people who attempt (and fail) to replicate. That's why original researchers succeed [sic] and replicators fail. Those who can, do science; those who can't, (fail to) replicate."
2.5: Other Defenses	38 (6.9%)	7.0%	"We should continue improving our approach but many of our design elements are already strong."

Note. The percentages indicate the number of free responses which contained a certain defense theme divided by the total number free responses collected.

ject to selection effects (e.g., involvement in the field, financial resources, etc.). We were able to recruit participants from social/personality, developmental, cognitive, and clinical psychology, however, we lack representation among other subfields, such as industrial/organizational psychology. Thus, our results are likely not representative of the entire field of psychology.

It is worth noting that early career researchers are more represented than other groups (e.g., tenured faculty), which may impact the data presented here. Specifically, of the 548 participants, 53.2% identified as graduate students and 23.8% identified as professors. To get a sense of the representativeness of our sample, we compared our sample to 2017 demographic statistics from the American Psychological Association (APA). In APA's 2017 member profile report, 25.4% of members identified as early career researchers. This suggests that our sample overrepresented researchers earlier in their career while simultaneously underrepresenting more established academics.

A related concern is that early career scholars may be more critical and more likely to endorse systemic changes compared to established researchers, and thus our data may convey more negative perceptions than if a representative sample had been collected. To address this possibility, we examined the prevalence of qualitative and quantitative responses as a function of career stage (see [Table 7](#) and [Table 8](#)). These data are not consistent with a trend where criticism is stronger at earlier career stages. Assistant professors had the highest rate of criticisms in their responses (87.0%), lecturers had the lowest, (75.0%), and graduate students fell in between (85.6%). Lecturers expressed the highest rate of support for "profound changes" (58.3%), full professors expressed the lowest (27.8%), and graduate students fell in between (38.0%).

Like our study, Baker's (2016) survey addressed beliefs about the replicability of published scientific findings. Their sample differed from ours in two major ways. First, their sample was predominantly composed of researchers from

**Table 4. Conservative and Liberal Inter-Rater Reliability (IRR) Calculations**

T Total # of Free Responses = 548	A # of agreements that a theme was present	B # of agreements that a theme was present OR absent	C Total # of times at least one judged a theme to be present	D Liberal IRR B/T	E Conservative IRR A/C
Theme 1: Criticisms of the Field	383	448	483	81.8%	79.3%
1.1: Statistics and/or research methods are misused and/or misunderstood	164	467	245	85.2%	66.9%
1.2: Problematic incentives in the field	185	509	224	92.9%	82.6%
1.3: There is a need for more transparency	83	529	102	96.5%	81.4%
1.4: Other criticisms	16	443	121	80.8%	13.2%
Theme 2: Defenses of the Field	72	468	152	85.4%	47.4%
2.1: Criticisms of how replications are conducted (e.g., contextual differences)	28	531	45	96.9%	62.2%
2.2: People are overreacting in the field	33	530	51	96.7%	64.7%
2.3: Psychology has safeguards to protect the field (e.g., self-correcting nature of science).	6	541	13	98.7%	46.2%
2.4: Criticisms of open science advocates and/or replicators	3	537	14	98%	21.4%
2.5: Other Defenses	4	495	57	90.3%	7%

Note. The letters at the top of each column (e.g., A) are meant to act as labels for the various values used to calculate IRR. 'A' represents the number of agreements that a theme was present. The conservative calculation used the proportion of responses for which both coders said the theme was present over the total number of instances where at least one coder said the theme was present.

biology, chemistry, earth sciences, medicine, physics, and engineering disciplines, while ours was composed exclusively of psychological researchers. Second, their survey was conducted in October through November of 2015, whereas ours was conducted in February through April of 2019. Nevertheless, considering both studies in tandem could give a more holistic picture of perceptions of reproducibility in the sciences.

In our study, 4.2% of participants indicated that *"The rate of false positive findings in the published psychology literature is acceptably low"* and 2% indicated that *"Our research practices were fine before the crisis, and we should continue employing more or less the same practices."* By comparison, in Baker's survey they found that 11.5% of participants disagreed with the statement *"I think that the failure to reproduce scientific studies is a major problem for all fields"* and 3% indicated that *"there is no significant crisis of reproducibility."* Thus, in both studies, the proportion of surveyed scientists who don't perceive a problem with replicability in science was quite low.

At the other end of the spectrum, 20.2% of Baker's participants chose the most extreme position (*strongly agree*) with respect to the statement *"I think that the failure to reproduce scientific studies is a major problem for all fields,"* and 35.2% of our participants choose the most extreme position *"The rate of false positive findings in the published literature is*

*much higher than it should be, and we should take major steps to lower it."* Overall, there is a sizable group of participants in both studies who see serious problems with replicability in the sciences.

One challenge for interpreting our quantitative results is that some of our items are 'double-barreled.' For instance, participants may agree that *"the rate of false positive findings in the published psychology literature is somewhat higher than it should be,"* yet disagree with the conjunction *"...and we should try to lower it"*. The structure of the question could confuse participants and made it difficult for them to choose a numerical response that accurately captured their perceptions.

Relatedly, the first question asked about the '*published psychology literature*' as a whole, whereas the second question asked about '*our practices*'. The former may have focused survey takers on the broad category of the field and the latter is ambiguous to the extent that it allows for personalized interpretation in a narrow sense (e.g., the participant's subfield).

While recognizing the limitations of our descriptive work, we believe it has allowed us to learn more about how academic psychologists, especially early career psychologists, perceive the field. Recurring critiques are statistics and methodology are misused/misunderstood, the current incentive structure is problematic, and there is a greater

**Table 5. Theme Prevalence as a Function of Perceptions Regarding the Rate of False Positives in Psychology**

	Acceptably low	Somewhat higher than it should be, and we should try to lower it	Much higher than it should be, and we should take major steps to lower it
Total frequency (N = 548)	23	332	193
Theme 1: Criticisms of the field	15 (65.2%)	268 (80.7%)	179 (92.7%)
1.1: Statistics and/or research methods are misused and/or misunderstood	6 (26.1%)	119 (35.8%)	93 (48.2%)
1.2: Problematic incentives in the field	8 (34.8%)	119 (35.8%)	87 (45.1%)
1.3: There is a need for more transparency	3 (13%)	60 (18.1%)	37 (19.2%)
1.4: Other Criticisms	2 (8.7%)	49 (14.8%)	29 (15%)
Theme 2: Defenses of the Field	14 (60.9%)	90 (27.1%)	20 (10.4%)
2.1: Criticisms of how replications are conducted (e.g., contextual differences)	3 (13%)	32 (9.6%)	5 (2.6%)
2.2: People are overreacting in the field	4 (17.4%)	33 (9.9%)	3 (1.6%)
2.3: Psychology has safeguards to protect the field (e.g., self-correcting nature of science).	1 (4.3%)	11 (3.3%)	1 (0.52%)
2.4: Criticisms of open science advocates and/or replicators	2 (8.7%)	7 (2.1%)	1 (0.52%)
2.5: Other Defenses	5 (21.7%)	21 (6.3%)	12(6.2%)

Note. The percentages indicate the number of people whose response contained a certain theme divided by the total number of group members with the same quantitative response.

need for transparency. Broadly, the first insight our data provides is that criticisms seem to be heavily present compared to defenses (i.e., 84.3% criticisms vs 22.6% defenses). This suggests that many academic psychologists believe that shifting away from the status quo is, at least to some degree, desirable. Fortunately, these critiques have not gone unnoticed and there are notable efforts addressing these issues at multiple levels.

We observed the greatest amount of concern regarding statistical awareness, suggesting that efforts to improve the field should place a particular emphasis on statistical training and literacy. With respect to existing resources, recent statistics textbooks, at both the introductory and graduate level, have begun adopting an open science perspective (Cummings, 2012; Cummings & Calin-Jageman, 2016) allowing individuals to teach themselves and others. The Framework for Open and Reproducible Research Training organization provides a consolidated database of teaching resources (*Framework for Open and Reproducible Research Training*, 2022). Creators of the FORRT database have scoured the literature in multiple fields, collecting and summarizing over 100 articles to help teachers, mentors, and students have more accessible avenues for training and self-education. In addition, Daniel Lakens offers a free Massive Open-Online Classroom (MOOC) ‘Improving your statistical inferences’ and has enrolled over 61,000 students for free (*Improving Your Statistical Inferences*, n.d.). The MOOC covers fundamentals of statistics, from correcting misconceptions about p-values to introducing Bayesian statistics. We believe our work suggests that these projects should continue to expand and help improve statistical literacy within the field of psychology.

A second prominent theme from our results is criticism of incentive structures. At an individual level we see academics having discussions about open science and incentives with a large audience. For example, several podcasts (e.g., Everything Hertz, Two Psychologists Four Beers, The Black Goat), blogs (e.g., The 20% Statistician, Data Colada, Replication Index), and journal clubs (e.g., ReproducibiliTea) frequently discuss the publication process, funding, null-results, and related topics with an open-science lens. There are also communities on social media platforms such as Twitter (e.g., @OSFramework, #openscience, #opendata), Reddit (e.g., r/Open\_Science), and Discord (e.g., Git Gud Science), that maintain ongoing conversations about these topics. The adoption of Registered Reports at over 300 journals is another example of shifts in incentive structures; this publication format allows research projects to be evaluated on their methods rather than their results, reducing the incentives to p-hack (Center for Open Science, n.d.). These types of changes may slowly chip away at a shaky incentive structure.

The third prominent critique observed within our data was a call for greater transparency. Not only are the previously mentioned communities having conversations about systemic issues, but they are also having conversations about open data, open materials, and pre-registration. Fortunately, organizations like the Open Science Framework provide many tools to aid researchers by increasing accessibility and reducing the time burden (*Open Science Framework*, n.d.). OSF provides a single location in which a user can submit pre-registrations, their data, funding checklists and can be used for free. Another website, AsPredicted, speeds up pre-registration, can work within the OSF, and

**Table 6. Theme Prevalence as a Function of Perceptions Regarding Research Practices in Psychology**

	Were fine before the crisis, and we should continue employing more or less the same practices	Have been shown to be somewhat worse than we thought, and we should make some incremental changes	Have been shown to be extremely problematic, and we should make profound changes
Total frequency (n = 548)	11	329	208
Theme 1: Criticisms of the field	5 (45.5%)	263 (79.9%)	194 (93.3%)
1.1: Statistics and/or research methods are misused and/or misunderstood	1 (9.1%)	110 (33.4%)	107 (51.4%)
1.2: Problematic incentives in the field exist	5 (45.5%)	114 (34.7%)	95 (45.7%)
1.3: There is a need for more transparency	0	56 (17%)	44 (21.2%)
1.4: Other criticisms	0	55 (16.7%)	25 (12%)
Theme 2: Defenses of the field	7 (63.6%)	99 (30.1%)	18 (8.7%)
2.1: Criticisms of how replications are conducted (e.g., contextual differences)	3 (27.3%)	31 (9.4%)	6 (3%)
2.2: People are overreacting in the field	2 (18.2%)	33 (10%)	5 (2.4%)
2.3: Psychology has safeguards to protect the field (e.g., self-correcting nature of science)	2 (18.2%)	9 (2.7%)	2 (1%)
2.4: Criticisms of open science advocates and/or replicators	1 (9.1%)	9 (2.7%)	0
2.5: Other defenses	2 (18.2%)	28 (8.5%)	8 (4%)

Note. The percentages indicate the number of people whose response contained a certain theme divided by the total number of group members with the same quantitative response.

is suitable for getting co-authors to approve the document before making it official (AsPredicted, n.d.). Moreover, journals like *Advances in Methods and Practices in Psychological Science* publish articles that focus on improving research practices (*Advances and Methods in Psychological Science*, n.d.). Christensen and his co-authors (2019) have also recently dedicated a text on how to do transparent research, covering topics such as reporting standards, data sharing, and creating reproducible workflows.

Our findings show that psychologists are willing to identify and acknowledge weaknesses and shortcomings within the field; optimistically, this practice may be a necessary step towards improvement."

#### Author Contributions

- Substantial contributions to conception and design: J. F. Miranda, C. M. Whitt, A. McDiarmid, J. E. Stephens, & A. M. Tullett
- Acquisition of data: J. F. Miranda, C. M. Whitt, A. McDiarmid, D. Purdue, & A. M. Tullett

- Analysis and interpretation of data: J. F. Miranda, C. M. Whitt, A. McDiarmid, & A. M. Tullett
- Drafted and/or Revision of the article: J. F. Miranda, C. M. Whitt, A. McDiarmid, D. Purdue, C. Hall, & A. M. Tullett
- Approved the submitted version for publication: J. F. Miranda, C. M. Whitt, A. McDiarmid, J. E. Stephens, D. Purdue, C. Hall, & A. M. Tullett

#### Competing Interests

The authors declare that there were no conflicts of interest with respect to the authorship or the publication of this article. Alexa Tullett is an author on this manuscript and is an editor for *Collabra: Psychology*. She was not involved in the review process of this article.

#### Funding

The data for this study is coming from a larger research project funded by the National Science Foundation, Award #:1728332 (See [link](#)).

**Table 7. Theme Prevalence as a Function of Career Position**

	Graduate student	Post-doc	Lecturer	Assistant Professor	Associate Professor	Full Professor	Other career position
Frequency of each career position	292	66	12	77	35	18	47
Theme 1: Criticisms of the field	250 (85.6%)	53 (80.3%)	9 (75.0%)	64 (87.0%)	27 (77.1%)	15 (83.3%)	43 (91.5%)
1.1: Statistics and/or research methods are misused and/or misunderstood	122 (41.8%)	22 (33.3%)	4 (33.3%)	28 (36.4%)	14 (40.0%)	7 (38.9%)	20 (42.6%)
1.2: Problematic incentives in the field exist	118 (40.4%)	21 (31.8%)	5 (41.7%)	31 (40.3%)	12 (34.3%)	6 (33.3%)	20 (42.6%)
1.3: There is a need for more transparency	54 (18.5%)	8 (12.1%)	2 (16.7%)	16 (20.8%)	5 (14.3%)	6 (33.3%)	9 (19.1%)
1.4: Other criticisms	38 (13.0%)	15 (22.7%)	2 (16.7%)	11 (14.3%)	5 (14.3%)	3 (16.7%)	6 (12.8%)
Theme 2: Defenses of the field	64 (21.9%)	20 (30.3%)	2 (16.7%)	13 (16.9%)	13 (37.1%)	5 (27.8%)	7 (14.9%)
2.1: Criticisms of how replications are conducted (e.g., contextual differences)	24 (8.2%)	5 (7.6%)	1 (8.3%)	1 (1.3%)	3 (8.6%)	3 (16.7%)	3 (6.4%)
2.2: People are overreacting in the field	19 (6.5%)	8 (12.1%)	0	6 (7.8%)	3 (8.6%)	1 (5.6%)	3 (6.4%)
2.3: Psychology has safeguards to protect the field (e.g., self-correcting nature of science)	7 (2.4%)	3 (4.5%)	0	3 (3.9%)	0	0	0
2.4: Criticisms of open science advocates and/or replicators	4 (1.4%)	2 (3.0%)	0	1 (1.3%)	2 (5.7%)	0	1 (2.1%)
2.5: Other defenses	18 (6.2%)	3 (4.5%)	1 (8.3%)	6 (7.8%)	7 (2.0%)	1 (5.6%)	2 (4.3%)

Note. Only one undergraduate responded to the free response question and was not included in this table. The percentages indicate the number of people whose response contained a certain theme divided by the total number of group members with the same career position.

### Data, Materials, and Online Resources

Access to all presented data (both raw and clean), SPSS code, and other relevant materials can be found [here](#).

Submitted: November 08, 2021 PDT, Accepted: April 28, 2022  
PDT

**Table 8. Perceptions Regarding the Rate of False Positives and Research Practices as a Function of Career Position**

	Graduate student	Post-doc	Lecturer	Assistant Professor	Associate Professor	Full Professor	Other career position
Frequency of each career position	292	66	12	77	35	18	47
The rates of false positives in psychology are...							
... Acceptably low	8 (2.7%)	4 (6.1%)	0	7 (9.1%)	2 (5.7%)	1 (5.6%)	1 (2.1%)
... Somewhat higher than it should be, and we should try to lower it	181 (62.0%)	42 (63.6%)	5 (41.7%)	43 (55.8%)	21 (60.0%)	12 (66.7%)	28 (59.6%)
... Much higher than it should be, and we should take major steps to lower it	103 (35.3%)	20 (30.3%)	7 (58.3%)	27 (35.1%)	12 (34.3%)	5 (27.8%)	18 (38.3%)
Our research practices...							
... Were fine before the crisis, and we should continue employing more or less the same practices.	6 (2.1%)	2 (3.0%)	0	1 (1.3%)	1 (2.9%)	0	1 (2.1%)
... Have been shown to be somewhat worse than we thought, and we should make some incremental changes.	175 (59.9%)	38 (57.6%)	5 (41.7%)	49 (63.6%)	22 (62.9%)	13 (72.2%)	27 (57.4%)
... Have been shown to be extremely problematic, and we should make profound changes.	111 (38.0%)	26 (39.4%)	7 (58.3%)	27 (35.1%)	12 (34.3%)	5 (27.8%)	19 (40.1%)

Note. Only one undergraduate responded to the free response question and was not included in this table. The percentages indicate the number of people whose response contained a certain theme divided by the total number of group members with the same career position.

Downloaded from [http://online.sagepub.com/journals/135711706571/collabra\\_2022\\_8\\_1\\_35711.pdf](http://online.sagepub.com/journals/135711706571/collabra_2022_8_1_35711.pdf) by guest on 10 August 2022



This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CCBY-4.0). View this license's legal deed at <http://creativecommons.org/licenses/by/4.0> and legal code at <http://creativecommons.org/licenses/by/4.0/legalcode> for more information.

## References

- Advances and methods in psychological science*. (n.d.). Sage Journals. <https://journals.sagepub.com/home/amp>
- AsPredicted. (n.d.). *AsPredicted*. <http://www.aspredicted.org>
- Baker, M. (2016). 1,500 scientists lift the lid on reproducibility. *Nature*, 533(7604), 452–454. <https://doi.org/10.1038/533452a>
- Bhattacharjee, Y. (2012, May 14). Paranormal circumstances: One influential scientist's quixotic mission to prove ESP exists. *Discover*. <http://www.discovermagazine.com/mind/paranormal-circumstances-one-influential-scientists-quixotic-mission-to-prove-esp-exists>
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. <https://doi.org/10.1191/1478088706qp063oa>
- Buffalmano, L. (2019). *Replication crisis: A defense of psychology*. The Power Moves. <https://thepowermove.com/replication-crisis/>
- Camerer, C. F., Dreber, A., Holzmeister, F., Ho, T.-H., Huber, J., Johannesson, M., Kirchler, M., Nave, G., Nosek, B. A., Pfeiffer, T., Altmeld, A., Buttrick, N., Chan, T., Chen, Y., Forsell, E., Gampa, A., Heikensten, E., Hummer, L., Imai, T., ... Wu, H. (2018). Evaluating the replicability of social science experiments in Nature and Science between 2010 and 2015. *Nature Human Behaviour*, 2(9), 637–644. <https://doi.org/10.1038/s41562-018-0399-z>
- Center for Open Science. (n.d.). *Registered Reports: Peer review before results are known to align scientific values and practices*. <https://www.cos.io/initiatives/registered-reports>
- Christensen, G., Freese, J., & Miguel, E. (2019). *Transparent and reproducible social science research: How to do open science*. University of California Press.
- Cummings, G. (2012). *Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis*. Routledge.
- Cummings, G., & Calin-Jageman, R. (2016). *Introduction to the new statistics: Estimation, open science, and beyond*. Routledge.
- Dreber, A., Pfeiffer, T., Almenberg, J., Isaksson, S., Wilson, B., Chen, Y., Nosek, B. A., & Johannesson, M. (2015). Using prediction markets to estimate the reproducibility of scientific research. *Proceedings of the National Academy of Sciences*, 112(50), 15343–15347. <https://doi.org/10.1073/pnas.1516179112>
- Ebersole, C. R., Axt, J. R., & Nosek, B. A. (2016). Scientists' reputations are based on getting it right, not being right. *PLoS Biology*, 14(5), 1002460.
- Ebersole, C. R., Mathur, M. B., Baranski, E., Bart-Plange, D.-J., Buttrick, N. R., Chartier, C. R., Corker, K. S., Corley, M., Hartshorne, J. K., IJzerman, H., Lazarević, L. B., Rabagliati, H., Ropovik, I., Aczel, B., Aeschbach, L. F., Andrighetto, L., Arnal, J. D., Arrow, H., Babincak, P., ... Nosek, B. A. (2020). Many Labs 5: Testing pre-data-collection peer review as an intervention to increase replicability. *Advances in Methods and Practices in Psychological Science*, 3(3), 309–331. <https://doi.org/10.1177/2515245920958687>
- Engber, D. (2016, August 28). Sad Face. *Slate*. [http://www.slate.com/articles/health\\_and\\_science/cover\\_story/2016/08/can\\_smiling\\_make\\_you\\_happier\\_maybe\\_maybe\\_not\\_we\\_have\\_no\\_idea.html](http://www.slate.com/articles/health_and_science/cover_story/2016/08/can_smiling_make_you_happier_maybe_maybe_not_we_have_no_idea.html)
- Engber, D. (2017, June 7). Daryl Bem proved ESP is real. Which means science is Broken. *Slate*. <https://slate.com/health-and-science/2017/06/daryl-bem-proved-esp-is-real-showed-science-is-broken.html>
- Feldman-Barrett, L. (2015, September 1). Psychology is not in crisis. *The New York Times*. <https://www.nytimes.com/2015/09/01/opinion/psychology-is-not-in-crisis.html>
- Fetterman, A. K., & Sassenberg, K. (2015). The reputational consequences of failed replications and wrongness admission among scientists. *PLoS ONE*, 10(12), e0143723. <https://doi.org/10.1371/journal.pone.0143723>
- Framework for open and reproducible research training*. (2022, February 16). FORRT. <https://forrt.org/>
- French, C. (2012, March 15). Precognition studies and the curse of the failed replications. *The Guardian*. <https://www.theguardian.com/science/2012/mar/15/precognition-studies-curse-failed-replications>
- Gilbert, D. T., King, G., Pettigrew, S., & Wilson, T. D. (2016). Comment on “Estimating the reproducibility of psychological science.” *Science*, 351(6277), 1037–1037. <https://doi.org/10.1126/science.aad7243>
- Houtkoop, B. L., Chambers, C., Macleod, M., Bishop, D. V. M., Nichols, T. E., & Wagenmakers, E.-J. (2018). Data sharing in psychology: A survey on barriers and preconditions. *Advances in Methods and Practices in Psychological Science*, 1(1), 70–85. <https://doi.org/10.1177/2515245917751886>
- Hsieh, H.-F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research*, 15(9), 1277–1288. <https://doi.org/10.1177/1049732305276687>
- Improving your statistical inferences*. (n.d.). Coursera. <https://www.coursera.org/learn/statistical-inferences>
- Jackson, D. J. R., Atkins, S. G., Fletcher, R. B., & Stillman, J. A. (2005). Frame of reference training for assessment centers: Effects of interrater reliability when rating behaviors and ability traits. *Public Personnel Management*, 34(1), 17–30. <https://doi.org/10.1177/009102600503400102>

- 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. <https://doi.org/10.1177/0956797611430953>
- Klein, R. A., Vianello, M., Hasselman, F., Adams, B. G., Adams, R. B., Jr., Alper, S., Aveyard, M., Axt, J. R., Babalola, M. T., Bahník, Š., Batra, R., Berkics, M., Bernstein, M. J., Berry, D. R., Bialobrzeska, O., Binan, E. D., Bocian, K., Brandt, M. J., Busching, R., ... Nosek, B. A. (2018). Many Labs 2: Investigating variation in replicability across samples and settings. *Advances in Methods and Practices in Psychological Science, 1*(4), 443–490. <https://doi.org/10.1177/2515245918810225>
- McAlister, A., Lee, D. M., Ehlert, K. M., Kajfez, R. L., & Kennedy, M. S. (2017). Qualitative coding: An approach to assess inter-rater reliability. *ASEE Annual Conference and Exposition*. <https://doi.org/10.18260/1-2--2877>
- McDiarmid, A. D., Tullett, A. M., Whitt, C. M., Vazire, S., Smaldino, P. E., & Stephens, J. E. (2021). Psychologists update their beliefs about effect sizes after replication studies. *Nature Human Behaviour, 5*(12), 1663–1673. <https://doi.org/10.1038/s41562-021-01220-7>
- Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science, 349*(6251), 4716. <https://doi.org/10.1126/science.aac4716>
- Open Science Framework. (n.d.). OSF. <http://www.osf.io>
- Rozin, P. (2001). Social psychology and science: Some lessons from Solomon Asch. *Personality and Social Psychology Review, 5*(1), 2–14. [https://doi.org/10.1207/s15327957pspr0501\\_1](https://doi.org/10.1207/s15327957pspr0501_1)
- 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. <https://doi.org/10.1177/0956797611417632>
- Srivastava, S. (2016, August 11). Everything is fucked: The syllabus. *The Hardest Science*. <https://thehardestscience.com/2016/08/11/everything-is-fucked-the-syllabus/>
- Wagenmakers, E.-J., Wetzels, R., Borsboom, D., & van der Maas, H. L. J. (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), 426–432. <https://doi.org/10.1037/a0022790>
- Washburn, A. N., Hanson, B. E., Motyl, M., Skitka, L. J., Yantis, C., Wong, K. M., Sun, J., Prims, J. P., Mueller, A. B., Melton, Z. J., & Carsel, T. S. (2018). Why do some psychology researchers resist adopting proposed reforms to research practices? A description of researchers' rationales. *Advances in Methods and Practices in Psychological Science, 1*(2), 166–173. <https://doi.org/10.1177/2515245918757427>
- Yarkoni, T. (2020). The generalizability crisis. *Behavioral and Brain Sciences, 45*, 1–37. <https://doi.org/10.1017/s0140525x20001685>

## Supplementary Materials

### Peer Review History

Download: [https://collabra.scholasticahq.com/article/35711-how-do-researchers-in-psychology-perceive-the-field-a-qualitative-exploration-of-critiques-and-defenses/attachment/90229.docx?auth\\_token=AKgCtHPBxPIrdnjstFZV](https://collabra.scholasticahq.com/article/35711-how-do-researchers-in-psychology-perceive-the-field-a-qualitative-exploration-of-critiques-and-defenses/attachment/90229.docx?auth_token=AKgCtHPBxPIrdnjstFZV)

---