Science and Engineering Ethics
https://doi.org/10.1007/s11948-018-0065-x

ORIGINAL PAPER

Grounds for Ambiguity: Justifiable Bases for Engaging in Questionable Research Practices

Donald F. Sacco[1] · **Mitch Brown**[1] · **Samuel V. Bruton**[2]

Received: 29 March 2018 / Accepted: 5 September 2018

Abstract

The current study sought to determine research scientists' sensitivity to various justifications for engaging in behaviors typically considered to be questionable research practices (QRPs) by asking them to evaluate the appropriateness and ethical defensibility of each. Utilizing a within-subjects design, 107 National Institutes of Health principal investigators responded to an invitation to complete an online survey in which they read a series of research behaviors determined, in prior research, to either be ambiguous or unambiguous in their ethical defensibility. Additionally, each behavior was paired with either an ostensibly sound or unsound reason for the behavior. Consistent with hypotheses, the results indicated that scientists perceived QRPs as more appropriate and defensible when paired with a justifiable motive relative to when paired with a clearly unethical motive, particularly for QRPs that are more ambiguous in their ethicality. In fact, ambiguous QRPs were perceived as categorically defensible when given a justifiable motive. This suggests scientists are sensitive to contextual factors related to QRPs' appropriateness, which could inform how institutions develop appropriate training modules for research integrity.

Keywords Questionable research practices · Ethics · Integrity · Motives

Introduction

Scientific research has come under increasing scrutiny in recent years, both from society and the scientific community. Although such scrutiny is due to a multitude of factors, one of the more salient sources of concern is related to what some have called a crisis in reproducibility (Ioannidis 2012; Loken and Gelman 2017).

✉ Donald F. Sacco
Donald.Sacco@usm.edu

[1] School of Psychology, The University of Southern Mississippi, Owings-McQuagge Hall, 118 College Drive #5025, Hattiesburg, MS 39406, USA

[2] School of Humanities, The University of Southern Mississippi, 118 College Drive #5037, Hattiesburg, MS 39406, USA

Published online: 26 September 2018

A sizeable proportion of findings in social and biomedical sciences have produced numerous failed attempts at replication by independent labs, including numerous failures in replicating highly publicized findings (Lynott et al. 2014; Open Science Collaboration 2015; Ritchie et al. 2012; Steele et al. 1999). While rigorous experimental design cannot eliminate the possibility that positive results are merely Type-I errors (false positives), it has been argued that the reproducibility problem has less to do with overt research misconduct than with more ethically ambiguous research practices, referred to as questionable research practices (QRPs; John et al. 2012; Simmons et al. 2011; Wicherts et al. 2016).[1]

These practices are "ambiguous" in their ethicality because certain research contexts may warrant their utilization. While only about 2% of researchers admit to research misconduct in the technical sense of fabrication, falsification and plagiarism, more than 70% of researchers indicate having engaged in behaviors that would likely qualify as QRPs (Fanelli et al. 2015). Such findings are particularly problematic when considering the documented 22% increase in published positive results from 1990 to 2007 (Fanelli 2010, 2011). Thus, based on self-reported behavior alone, it appears that most scientists find QRPs defensible under certain circumstances, suggesting the underlying reasons for engaging in these "questionable" behaviors may be an important determinant of their defensibility. However, little discussion in research ethics has explicitly focused on possible reasons for engaging in these practices and whether scientists are sensitive to them in ways that influence determinations of their ethical defensibility. The current study focuses on how various motives for QRPs influence the extent to which research scientists find these behaviors ethically defensible and appropriate.

Questionable Research Practices

In their seminal paper, Simmons et al. (2011) documented the impact that various methodological decisions, to which they refer as "researcher degrees of freedom," can have on the probability of Type-I errors in research. Using statistical simulations and a standard significance cutoff value of $p = 0.05$, they identified the impact of several common practices on alpha inflation: choosing among different dependent variables, choosing sample size, using covariates, reporting subsets of experimental conditions, and combinations of these practices. Specifically, they generated samples in which each observation was independently drawn from a normal distribution. Analyses were then conducted on each of these samples to calculate how often at least one p value was below standard significance levels. For example, if one utilizes two moderately correlated dependent measures, one could analyze each separately as well as the combination of the two variables; the probability of one of these tests producing a significant result is greater than 0.05. The added value of this approach

[1] A recent report by the National Academies of Science, Engineering, and Medicine refer to QRPs as Detrimental Research Practices, thereby emphasizing the potential deleterious consequences of these behaviors on scientific research (NASEM 2017).

is that it quantifies how much greater than 0.05, and therefore how much alpha inflation, each behavior produces.

Any single practice has a notable impact on alpha inflation (range of alphas: 7.7–12.6%, thereby increasing the probability of Type-I errors between 2.7 and 7.6%; Simmons et al. 2011). Additionally, combinations of these practices have an even more noticeable impact on alpha inflation, with adjusted alphas ranging from 14.4% when using a combined strategy of choosing among multiple dependent measures and adding participants to one's sample, to an alpha of 60.7% when utilizing all of these behaviors simultaneously. These results raise concerns for the reproducibility of science because researchers have historically relied on conventional alpha cutoffs (e.g., 0.05, 0.01) to ensure results are reported with the likelihood of false-positive errors being low, compared to false-negative errors (failing to identify a true effect as statistically significant). However, little attention had been paid to the impact of researcher methodological decisions on Type I error rates. That such behaviors may have a deleterious impact on rates of Type I errors is potentially even more problematic when considering the self-reported frequency with which scientists report engaging in potential QRPs. For example, more than 25% of sampled scientists indicate having failed to report all of a study's conditions in a paper (John et al. 2012).

On the face of it, QRPs are ethically questionable in part because they can potentially increase the probability of generating false-positive findings, as the above example illustrates. Other QRPs can lead to different sorts of problems. For example, failure to publish negative results, whether to please a sponsor or because of the difficulty of getting negative results published (i.e., the "file drawer" problem), can lead to biased estimates of intervention effect sizes in meta-analytic studies (Franco et al. 2014). Strategically delaying publication of results in an attempt to position them for publication in a higher impact journal can impede scientific progress. However, there are contextual factors that may influence the ethics of many behaviors that have been labeled questionable, and previous research asked participants only whether they had ever engaged in various questionable research practices but did not ask for the rationale for these decisions (John et al. 2012). Without information regarding the underlying motives for a particular research behavior, it is challenging to determine the extent to which purported QRPs are more or less ethically defensible; that they are referred to as "questionable" practices suggests the underlying reasons for engaging in them is an important factor when attempting to determine the extent to which the research context may have warranted such behavior.

It is possible to identify numerous situations that would not only justify, but perhaps even necessitate, utilization of QRPs. For example, it can be argued that inclusion of covariates in statistical models has a place. When random assignment is not possible and experimental groups can differ on important dimensions prior to the introduction of an intervention, researchers must assess and control for pretest scores on critical variables to get a more accurate estimate of the efficacy of an intervention; failure to covary out, or control for, these potential pre-existing group level differences can bias the estimate of treatment effects (May 2012). In another example, participants may be justifiably excluded from analyses because they fail to complete all study procedures, for failing attentional check items embedded in a

survey instrument, or because their responses are revealed to be outliers based on certain statistical analyses.

These behaviors are at the discretion of the researcher, hence their label as "researcher degrees of freedom," but unfortunately, the motivations for these behaviors are not always clear to those evaluating research, as the researcher could at best fail to provide a rationale for the behavior or at worst, could simply be dishonest. Additionally, researchers may not report these practices when disseminating their work. For example, a researcher may make a motivated decision to exclude participants but misrepresent the sample size in the paper, thereby precluding evaluation of the decision's merits. Because researchers' own integrity is the primary safeguard against QRPs and because sound research ethics often depend on highly contextual aspects of projects that are difficult for those unaffiliated with the research to evaluate, such self-policing may be of limited effectiveness. Additionally, many have questioned the efficacy of self-correction in science for numerous reasons, including the challenge of getting null results published in academic journals (Ioannidis 2012).

The Current Study: Awareness of and Sensitivity to QRP Motives

To date, most of the research on QRPs has focused on their prevalence, their potential impact on scientific findings, and the extent to which scientists find such behaviors ethically defensible (e.g., John et al. 2012; Sacco et al. 2018; Simmons et al. 2011). However, such research has not explicitly tested the extent to which researchers *are* aware of and sensitive to underlying motives for QRPs when assessing their ethical defensibility. The benefit of knowing such information is twofold. Foremost, improved understanding of situation-specific factors relevant to QRP use could lead to better educational tools. Additionally, such information may add broader context to the significance of researchers' high frequency of self-reported engagement in QRPs.

To this end, the current study utilized a cohort of research scientists to determine the extent to which they are sensitive to situations that may influence QRPs' justifiability. Importantly, participants were presented with two categories of behaviors validated in previous research: behaviors that were more clearly unethical whereas the other half were ambiguous in their ethicality (Sacco et al. 2018). Additionally, each behavior was accompanied by a motive intended to be viewed as clearly unjustifiable or justifiable. Upon reading each statement, participants indicated how ethically defensible and appropriate each behavior was. There were several a priori hypotheses for this study. The first prediction was that participants would find clearly unethical research practices less defensible and less appropriate than ambiguously unethical research practices. The second prediction was that clearly unethical research practices would be considered ethically indefensible and inappropriate, regardless of the justification (i.e., clearly defensible vs. clearly indefensible justification). That is, clearly unethical research practices are unethical, regardless of the justifications for their implementation.

Conversely, it was expected that the justification provided in the scenarios would qualify participants' perceptions of the ethical defensibility and appropriateness of

ambiguously unethical practices, such that participants would find these behaviors more defensible and appropriate when accompanied by a justifiable reason for their implementation. Thus, it was predicted that research scientists would be sensitive to the variations in research contexts that would indicate the relative ethical defensibility of a particular research behavior. If supported, these predictions would suggest that overly general instructional guidance about QRPs may have limited impact, given that researchers are responsive to the contextual (in)appropriateness of such behaviors.

Method

Participants and Procedure

A power analysis indicated 90 participants would sufficiently detect small to medium effects (Cohen's $f = 0.15$, $\beta = 0.80$). Potential participants were contacted through an email listserv generated from emails of 1500 NIH-funded research scientists from 30 different research institutions across the U.S. in multiple waves of emails (for more detail on this procedure, see Sacco et al. 2018). Waves included 100–200 emails at a time and were sent over the course of a month. Of those contacted, 135 scientists consented and provided responses to the survey. However, twenty-eight participants provided incomplete responses and were therefore excluded from statistical analyses, resulting in a final sample of 107 participants ($M_{Age} = 50.60$ years, SD 8.97; 58 women, 46 men, 1 other, 2 did not specify gender; 82.2% White). Participants represented a variety of different scientific fields, including medicine, psychology, and genetics, and had spent, on average, $M = 22.82$ (SD 8.51) years in the field. Respondents were compensated with a redemption code for an Amazon gift card worth $10 (US).

After reading an invitation email, interested participants clicked on the available link containing the informed consent document. Consenting participants were redirected to a screen to begin the survey; participants not wishing to consent to the study were instructed to close their browser window and were unable to access the survey. Following consent, participants read each scenario in a randomized and counterbalanced order. Then, participants provided demographics information before being debriefed and provided with instructions to receive their Amazon gift card code. Importantly, participation in this study was confidential and participants' responses were dissociated from their identities.[2]

[2] This research project was reviewed and approved by the University of Southern Mississippi Institutional Review Board (Protocol Number: CH2-16110904).

Materials

QRP Scenarios

Participants read vignettes of research scientists engaging in various research-related QRPs. These vignettes featured QRPs derived from previous work (e.g., John et al. 2012; Tijdink et al. 2014) as well as the research team's own knowledge of research ethics. In prior work, a factor analysis revealed a two-factor structure associated with these QRPs such that some behaviors were determined to be unambiguously unethical and others as ambiguously unethical (Sacco et al. 2018). The unambiguously unethical QRPs (UU-QRP) included 6 different decisions that included items such as refusing to share data and statistical analysis output with all coauthors, and not clearly reporting a corporate sponsor's involvement in the preparation of a manuscript. These practices are typically perceived by research scientists as being *unequivocally* unethical and adverse to good science. Conversely, the 12 ambiguously unethical QRPs (AU-QRP) referred to those related to data analytic strategies and results reporting that are often in some sense technically accurate but misleading (e.g., collecting additional data following an analysis finding results trending toward significance; Sacco et al. 2018).

Twice as many ambiguous items were generated for two reasons. First, it was inherently more challenging to generate sound motives for unambiguously unethical research practices, thereby limiting the number of items in this category. Second, it was hypothesized that the two different motives offered for the ambiguous behaviors would moderate judgments of their ethical defensibility and appropriateness; it was not expected that the differing motives for unambiguously unethical research behaviors would have this effect. In general, AU-QRPs are perceived as being more defensible than UU-QRPs (Sacco et al. 2018).

Participants viewed two versions of these vignettes that varied in terms of the motive of each QRP. Motives for QRPs were classified as either theoretically justifiable or not. For example, for the QRP involving collecting additional data following an initial analysis, a justifiable motive was having to exclude 10 participants from the initial analysis for failing a critical attention check, thus necessitating the collection of additional data to attain minimum statistical power to detect effects. Conversely, an unjustifiable motive for this QRP would be simply to find the effect and ultimately not reporting this decision in the manuscript. In the latter case, simply adding participants because a statistical test has not yet reached conventional significance in the hopes that additional data collection will result in statistical significance is an unwarranted "researcher degree of freedom," and has been identified as a practice directly associated with increased Type I error rates (Simmons et al. 2011). All behaviors and motives used in the scenarios can be found in "Appendix".

Following each scenario, participants responded to face-valid questions about each QRP scenario. Participants indicated the extent to which they found the researcher's action appropriate in each vignette (i.e., "How appropriate was this researcher's action in this scenario?", 1 = *Very Inappropriate*; 4 = *Neither Appropriate nor Inappropriate*; 7 = *Very Appropriate*) and how defensible the decision was (i.e., "How defensible was this researcher's decision?", 1 = *Completely Indefensible*;

4 = *Neither Defensible nor Indefensible*; 7 = *Completely Defensible*). As each attitude was assessed with a single item, we calculated participants' mean ethical defensibility and appropriateness ratings for unambiguously unethical QRPs paired with a justifiable motive, unambiguously unethical QRPs paired with a justifiable motive, ambiguously unethical QRPs paired with a justifiable motive, and ambiguously unethical QRPs paired with an unjustifiable motive.[3]

Results

Behavioral Appropriateness

To ensure that participants found the various categories of behaviors appropriate or inappropriate as predicted, a 2 (QRP: Ambiguously Unethical vs. Unambiguously Unethical) × 2 (Motive: Justifiable vs. Unjustifiable) repeated-measures ANOVA was conducted. There was a main effect of QRP, such that UU-QRPs (e.g., principal investigator refusing to share data with collaborators; $M = 2.90$, SD 0.63) were perceived as less appropriate than AU-QRPs (e.g., reporting results from 2 of the 4 tested dependent variables tested; $M = 3.73$, SD 0.67), $F(1, 106) = 293.09$, $p < 0.001$, $\eta_p^2 = 0.734$.

Another main effect indicated that QRPs in general were perceived as more inappropriate when paired with an unjustifiable motive (e.g., reporting unexpected results as predicted from the beginning; $M = 2.58$, SD 0.63) than paired with a more justifiable motive (e.g., providing a tentative post hoc explanation for unexpected results; $M = 4.06$, SD 0.66), $F(1, 106) = 1245.13$, $p < 0.001$, $\eta_p^2 = 0.922$. Effects were not qualified by an interaction, $F(1, 106) = 2.46$, $p = 0.119$, $\eta_p^2 = 0.023$. Importantly, these omnibus means were either not significantly different from the midpoint or significantly below the midpoint on the scale (significant ps < 0.001, non-significant $p = 0.298$). This means that these QRPs are either categorically inappropriate or neither appropriate nor inappropriate. This suggests that participants were cautious in reporting any QRPs as appropriate, regardless of category or motive.

[3] Reliabilities were computed for both types of motives for UU- and AU-QRPs for the appropriateness and defensibility scales. For UU-QRPs' appropriateness, both justifiable ($\alpha = 0.33$) and unjustifiable motives ($\alpha = 0.33$) produced low reliabilities. The reliabilities were also low in defensibility for justifiable ($\alpha = 0.41$) and unjustifiable motives ($\alpha = 0.41$). With AU-QRPs' appropriateness, justifiable motives elicited a low reliability ($\alpha = 0.58$), whereas unjustifiable was acceptable ($\alpha = 0.70$). For defensibility, unjustifiable ($\alpha = 0.74$) and justifiable motives ($\alpha = 0.64$) were acceptable. The low reliabilities of the UU-QRP items were not substantially improved when removing less reliable items from the analyses (αs < 0.50) and the justifiable motives reliability for AU-QRPs did not improve following removal of any single item (αs < 0.59). Because reliabilities did not improve into an acceptable range for any of the low-reliability subscales (Cronbach and Meehl 1955), and we wanted to eliminate as much of an imbalance between responses for subscales as possible, we opted to report analyses including all tested items. These low alphas, particularly for ostensibly justifiable motives, may reflect the actual variability in perceptions of how acceptable certain research practices are, even when considering these motives.

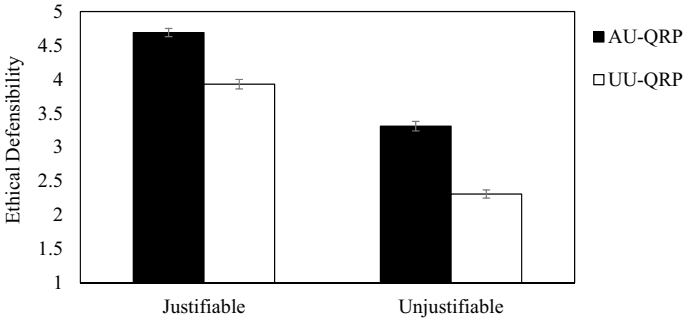

Fig. 1 Perceived ethical defensibility of QRPs with justifiable and unjustifiable motives; Ambiguously unethical QRPs are referred to as AU-QRPs and Unambiguously unethical QRPs are referred to as UU-QRPs

Ethical Defensibility

A similar 2 × 2 repeated-measures ANOVA was conducted to determine the perceived defensibility of the QRPs. There was a main effect of QRP, such that AU-QRPs ($M = 3.99$, SD 0.70) were perceived as more defensible than UU QRPs ($M = 3.12$, SD 0.69), $F(1, 106) = 247.00$, $p < 0.001$, $\eta_p^2 = 0.734$. Another main effect indicated that both types of QRPs were perceived as more ethically defensible when paired with justifiable motives ($M = 4.31$, SD 0.68) than unjustifiable motives ($M = 2.81$, SD 0.68), $F(1, 106) = 1449.22$, $p < 0.001$, $\eta_p^2 = 0.932$.

Effects were qualified by an interaction, $F(1, 106) = 9.07$, $p = 0.003$, $\eta_p^2 = 0.079$ (see Fig. 1). Simple effects tests indicated that AU-QRPs with a justifiable motive (e.g., excluding participants from analyses after determining they are statistical outliers; $M = 4.69$, SE 0.06) were perceived as more defensible than AU-QRPs with unjustifiable motives (e.g., excluding participants from analyses without determining they are outliers; $M = 3.31$, SE 0.07), $F(1, 106) = 778.22$, $p < 0.001$, $\eta_p^2 = 0.880$. Whereas the AU-QRPs with unjustifiable motives (e.g., stopping data collection early when significant results are obtained and immediately attempting to publish work) were categorically indefensible when weighted against the midpoint, $t(106) = -9.57$, $p < 0.001$, $d = 0.65$, AU-QRPs with justifiable motives (e.g., stopping data collection early for significant results but immediately conducting a follow-up study to determine the reliability of its findings) elicited a score significantly above the midpoint, $t(106) = 10.83$, $p < 0.001$, $d = 0.65$. UU-QRPs with justifiable motives (e.g., refusing to share data with collaborators to ensure confidentiality of participants; $M = 3.93$, SE 0.07) were perceived as more defensible than those with an unjustifiable motive (e.g., refusing to share data with collaborators to ensure only the principal investigator has full access to the data; $M = 2.31$, SE 0.06), albeit at a reduced magnitude, $F(1, 106) = 664.56$, $p < 0.001$, $\eta_p^2 = 0.862$. Further, whereas UU-QRPs with unjustifiable motives were categorically indefensible, $t(106) = -27.44$, $p < 0.001$, $d = 0.65$, those with justifiable motives were not significantly different from the midpoint, $t(106) = -0.83$,

$p = 0.487$, $d = 0.13$, suggesting that even defensible motives do not lead scientists to regard clearly unethical research practices as ethically defensible.

Discussion

Much research over the past decade has identified various researcher "degrees of freedom," their prevalence, and their impact on science (John et al. 2012). In part, this research has been a response to science's reproducibility crisis (Simmons et al. 2011). Despite consensus condemnation of research practices such as fabrication, falsification, and plagiarism, most research ethicists describe various other behaviors as ethically questionable (i.e., QRPs). Implicit in this language is the idea that these behaviors are not a priori categorically inappropriate. The motive guiding the decision to engage in QRPs, and the context in which those QRPs happen, appear to influence their defensibility. Some suggest educational interventions to inform scientists about these practices and how they may be detrimental to science as a possible means of reducing inappropriate utilization of these various behaviors (Anderson et al. 2007; John et al. 2012). However, many educational initiatives have limited efficacy for fostering early-career scientists' engagement in best research practices (Ramalingam et al. 2014; Todd et al. 2017; Watts et al. 2017).

The current study suggests that one reason educational interventions may have limited effectiveness is because they are redundant with scientists' typical knowledge of research integrity. Scientists' extensive training in methodology and statistical analysis have also made them adept at knowing when a particular research practice is more or less defensible. For example, although many governing bodies (e.g., Offices of Research Integrity) discourage "salami-slicing," and the re-use of previously published data for a different analysis and publication (e.g., Roig 2015), such practices may be defensible, or even necessary, to address variables that were previously unconsidered or to employ a different analytic strategy that would be more appropriate. Thus, it may not be primarily limited knowledge or expertise responsible for the prevalence of these behaviors.

Several findings supported this study's hypotheses that scientists are highly sensitive to best practices as they relate to methodology and statistical analysis. First, regardless of the motive provided, participants were aware of which behaviors were clearly inappropriate and ethically indefensible and which evoked ambiguity. Second, participants were sensitive to which kinds of motives determine the appropriateness of a potential questionable research practice. Both categories of QRPs were perceived as more defensible and appropriate when accompanied by a justifiable motive. Most importantly, however, participants were especially sensitive to ambiguously unethical QRPs rooted in ostensibly justifiable motives, as this was the only category in which participants defensibility ratings were categorically defensible. In all other cases, certain motives may have made the behavior more defensible than it otherwise might be, but the behaviors were still considered largely indefensible. Thus, participants seem to have a relatively clear understanding of the types of research behaviors that are clearly unethical and those for which ethicality is more ambiguous. Furthermore, even in the case of more ambiguous behaviors,

participants are familiar with the motive that would make these ambiguous behaviors ethically defensible, versus the motive that would preclude ethical implementation of these behaviors. As such, interventions designed to educate scientists about the negative consequences of QRPs may be more efficacious when supplemented with information regarding the contexts in which these various behaviors may be more or less ethically defensible.

The implications of these findings may offer insight into the reported high frequency of self-reported engagement in questionable research practices by scientists. Foremost, it suggests that the context in which these behaviors occur is critical. It is apparent from these findings that scientists are well-aware of contexts in which these behaviors are more or less appropriate. Thus, the fact that a significant percentage of scientists report having engaged in these behaviors in their own research may simply be due to the fact that the researcher perceived that the research situation necessitated their implementation, rather than specific endorsement of these practices in contexts where their utilization would be considered ethically questionable.

Nonetheless, it is impossible to deduce from this and prior work whether or not scientists actually engage in these practices only in situations in which such behaviors are perceived to be appropriate. It may be the case that even though scientists know when these ambiguous research practices are inappropriate, there is nonetheless motivation to engage in the behavior. Specifically, the goal to produce statistically "significant" results may inform decisions to engage in questionable research practices, which itself may be motivated by institutional pressures or professional goals. For example, past research finds that to the extent that scientists believe that such behaviors are necessary to remain competitive in their discipline or that the behaviors are normatively acceptable because their peers engage in such behaviors for professional advancement, the more strongly they believe that questionable research practices are ethically defensible and the more they indicate willingness to engage in such practices (Sacco et al. 2018). Given research showing that the inappropriate implementation of questionable research practices is detrimental to science, and the fact that scientists seem to be aware of the contexts in which these behaviors are more or less appropriate, interventions designed to reduce the misuse of these behaviors through traditional/standard/formal RCR education alone may be of limited effectiveness if the underlying motive for engaging in the behaviors is unrelated to a lack of knowledge regarding their appropriateness.

In cases where implementation of questionable research practices are motivated not by their necessity to facilitate rigorous science, but rather due to institutional or professional pressure, research would benefit by developing intervention strategies that circumvent the logic that leads scientists to endorse behaviors detrimental to science. For example, to counter the belief that these behaviors are normatively necessary, perhaps an educational intervention designed to make salient to participants the value they place on ethical science and how these behaviors are inconsistent with their value system as an ethical researcher could leverage findings from the cognitive dissonance scientific tradition to reduce support for QRPs, which would align with previous research suggesting dissonance-reduction motives ultimately foster prosocial behavior (Festinger and Carlsmith 1959; Goldstein et al. 2007; LaRose and Kim 2006).

It is worth noting that, by and large, participants found the majority of unambiguously and ambiguously unethical behaviors to be more inappropriate than appropriate and less ethically defensible than defensible. Only one category of behaviors—ambiguously unethical behaviors accompanied by defensible justifications—produced mean ethical defensibility ratings and appropriateness ratings above the midpoint of the scale, and thus, perceived to be more defensible and appropriate than indefensible and inappropriate. Thus, it seems that scientists are highly cautious in their evaluations of these behaviors, more frequently erring on the side of viewing them as inappropriate and indefensible, regardless of the provided rationale. However, this could also represent an artefact of socially desirable responses (Leary and Kowalski 1990). That is, participants may have deduced, at least to some extent, what the study was about and might have responded in a manner that would generate a favorable view of them (Paulhus 2002). Relatedly, those who participated in the study could represent a response bias for research scientists whose research practices generally conform to recommendations for best practices and the data may not necessarily assess responses of those who could actually find certain QRPs appropriate or defensible. One possible way to circumvent these issues could include using a "Bayesian truth serum," an algorithmic technique designed to provide incentives for truthful responding (Prelec 2004). Specifically, this technique creates a scenario in which participants would be rewarded for providing truthful answers about their own behavior. Previous research in estimating research misconduct has utilized this technique by amplifying the moral stakes of research scientists' responses by indicating that an increased charitable donation would occur for those most willing to respond truthfully, compared to a control technique that simply indicated a charitable donation would be made for each participant with no chance of increasing the amount (John et al. 2012).

Limitations

In addition to the above-noted potential and specific limitation, the current study presents various other limitations. Foremost, although the ostensibly justifiable and unjustifiable motives for the various QRPs varied in the predicted capacities, the basis of such justifiability remains unclear. Perhaps justifiable motives were perceived as being part of best practices for science (e.g., replacing excluded data to ensure full statistical power) or necessary to ensure the most accurate representation of the population (e.g., eliminating statistical outliers following the appropriate analyses). Future research would benefit from considering the basis of justifiability for QRPs, which could then form the basis of nuanced ethics training modules that address when these ambiguous QRPs are not questionable in their implementation, but necessary. Furthermore, identification of the basis of justifiability may ultimately predict behavioral intentions for QRP engagement itself. For example, recognizing a QRP as having a legitimate use may elicit a greater willingness to use it. Conversely, if motivations to engage in QRPs are unjustifiable, this may ultimately mitigate the likelihood one engages in such practices.

Participants also had the opportunity to provide feedback following completion of the study, which resulted in several scientists indicating that they were not necessarily familiar enough with certain QRPs to feel confident in their answers. For example, those conducting qualitative research may not be as familiar with the statistical component of QRPs and may not readily understand the ethical implications of certain behaviors compared to individuals whose work is exclusively quantitative. Furthermore, different disciplines rely on different methodologies, which could preclude some scientists from knowing the relevant ethical nuances. Future research could consider creating QRPs specific to each field. For example, although learning about responsible conduct for research with animals would contribute to any scientist's consummate knowledge of ethicality, scientists who exclusively conduct research with humans may not benefit from training modules addressing animal research, whereas discussing ambiguous scenarios common in their respective field could be a more effective use of the limited time allotted for training.

Finally, while the reliability of appropriateness and defensible ratings of ambiguously unethical behaviors were acceptable (ranging from 0.58 to 0.74), the reliabilities for unambiguously unethical behaviors were quite low (ranging from 0.33 to 0.41). Furthermore, the removal of any single "poor" item from any variable category did not significantly improve reliability. There are potential reasons for these low reliabilities for UU QRPs. Given that they were previously found to be clearly unethical, providing a context that attempts to make them appear more defensible or indefensible (or more appropriate or inappropriate) may present as an unrealistic task for participants, thereby adding error variance to the assessment of these behaviors. That is, participants' responses may be highly variable for these behaviors because it may be confusing to evaluate the appropriateness and defensibility of clearly inappropriate and indefensible behaviors based on provided reasons for said behaviors. Nonetheless, participants did perceive the unambiguously unethical behaviors as more defensible when ostensibly justifiable reasons were provided for the behavior compared to when the behaviors were paired with unjustifiable reasons.

Conclusion

Questionable research practices pose a pervasive threat to science, in part through their ability to inflate the rate of Type-I Error. At times, however, QRPs can be used legitimately. The current research sought to identify potential scenarios where QRPs could justifiably occur, which could inform mentors as to how to improve science education while acknowledging the necessity of certain "researcher degrees of freedom." Ultimately, this work could contribute to a greater consensus among scientists for the acceptability of certain behaviors, which could foster more transparency and therefore reproducible results.

Funding The authors disclose that this research was funded by grants awarded to the first and third author from the Department of Health and Human Services' Office of Research Integrity (Grant Nos. 1 ORIIR170035-01-00 and 1 ORIIR160021-01-00).

Appendix

UU-QRPs with Unjustifiable Motives

1. A researcher is coauthoring a manuscript using data he collected and used previously for another study and separate paper. He does not indicate the origin of his data, because the hypotheses and variables analyzed in the current manuscript differ from those in the prior work.
2. A researcher submits a paper for publication without receiving direct permission from all coauthors, since as the lead author, she believed she could speak for all of her coauthors.
3. A researcher investigating the preliminary effects of a new pharmaceutical withholds journal submission of a manuscript at the request of the corporate sponsor, because the sponsor wants the paper's publication to coincide with the release of the new drug.
4. A researcher's coauthor wants to conduct his own analyses on some data recently collected by the research team, but the Principal Investigator refuses to share the raw dataset to ensure that he alone has access to the findings.
5. A researcher fails to replicate a finding that she found recently in another study. She decides to publish the results from the first study alone, without reporting the failed replication attempt, since she is confident that the first study represents the truth.
6. An employee of a corporate sponsor of the research analyzes the data for publication and is listed as one of the study's co-authors. Though the sponsor's involvement in the project is clearly indicated in the manuscript, the employee's relationship to the sponsor is not disclosed.

UU-QRPs with Justifiable Motives

1. A researcher is coauthoring a manuscript using data he collected and used previously for another study and separate paper. He does not indicate the origin of his data, because his graduate school mentor used to do the same thing, leading him to infer that this is an acceptable practice.
2. A researcher submits a paper for publication without first receiving direct permission from all coauthors, because the coauthors had approved earlier, similar drafts.
3. A researcher investigating the preliminary effects of a new pharmaceutical withholds journal submission of a manuscript at the request of the corporate sponsor, because the sponsor wants additional evidence that the drug has no significant side effects.
4. A researcher's coauthor wants to conduct his own analyses on some data recently collected by the research team, but the Principal Investigator refuses to share the raw dataset in order to protect participants' confidential information.
5. A researcher fails to replicate a finding that she found recently in three previous studies. She decides to publish the results from the original three studies alone,

without indicating the failed replication attempt, since she is confident that the first three studies represent the truth.

6. An employee of a corporate sponsor of the research analyzes the data for publication and is listed as one of the study's co-authors. The sponsor's involvement in the project is clearly indicated in the manuscript, and the employee's relationship to the sponsor is disclosed.

AU-QRPs with Unjustifiable Motives

1. The results from a researcher's most recent study are trending toward significance, which prompts the researcher to collect data from an additional ten participants beyond what was estimated from an a priori power analysis. The original data set was sufficiently powered to detect effects and no participants were excluded from analyses. Following this extra data collection, the results become conventionally significant and are then published, but without disclosing the post hoc modification of the research plan.
2. Upon a preliminary analysis of a data set, a researcher observes that she is about 20 observations away from what her original study design called for, but since her hypothesis is already supported by the existing data, she stops data collection to write up the results for publication immediately.
3. So as to meet a target journal's standards for statistical significance, and because the journal requires reporting p values only to the hundredth place, a researcher reports the p value of a finding as $p = 0.05$, rounding down the actual significance observed, which was $p = 0.053$.
4. In a submitted manuscript, a researcher acknowledges a statistician for her technical assistance without asking the statistician for permission to do so, simply assuming the statistician would appreciate the acknowledgement.
5. A researcher publishes statistically significant results, but statistical significance was attained only by excluding data from two participants after the researcher becomes convinced that these participants probably should not have been included in the study to begin with.
6. A researcher conducts a study in which two outcome measures out of four yield significant results. She reports the results of only the two significant measures, and fails to mention the two others, so as to present the results as more consistent and impressive than the entirety of her data supports.
7. A researcher obtains unexpected results from what she predicted for a study and drafts a manuscript reporting these results as having been predicted from the start, reporting the results in the discussion as support for predicted relationships.
8. A researcher conducts a study with two independent variables, with an a priori prediction that there should be a statistical interaction between the two factors. However, the statistical interaction is not conventionally significant. He omits the results of this analysis in his paper in order to conduct post hoc tests, which provide support for more specific predictions, and does not disclose that he predicted a statistical interaction between the two variables.

9. Several of a researcher's colleagues have questioned why she has not yet submitted for publication a paper they had all recently coauthored. Because she is under pressure from her department chair to publish in high-impact journals, she is delaying submission because she hopes to discuss the paper at an upcoming conference with the editor of a journal she has in mind. She hopes that by delaying publication, it will increase the likelihood of publication in a high-impact journal.
10. A researcher conducts two different but conceptually similar studies that she intends to make into two different papers. For convenience, she re-uses the literature review from the first-drafted paper for the second without disclosing the reuse of this material.
11. So as to garner more publications, a researcher separates out several aspects of a large study he recently conducted, making each an independent manuscript, but without disclosing the relationship of the papers to each other in each manuscript.
12. At the suggestion of a corporate sponsor of the research, a researcher changes her planned methods of statistical analysis, even though she believes her original methods more accurately represent the data.

AU-QRPs with Justifiable Motives

1. The results from a researcher's most recent study are trending toward significance, which prompts the researcher to collect data from an additional ten participants beyond what was estimated from an a priori power analysis. Ten participants from the initial sample were excluded from analysis after failing an attention check, thus necessitating the additional ten participants for the study to attain adequate statistical power. The results then become conventionally significant and are published; the researcher discloses details regarding post hoc modification of the research plan.
2. Upon a preliminary analysis of a data set, a researcher observes that she is about 20 observations away from what her original study design called for, but since her hypothesis is already supported by the existing data, she stops data collection to preserve unused research funds for a follow-up study that will determine the reliability of the original findings.
3. Because a target journal limits reporting p values to the hundredth place, a researcher reports the p value of a finding as $p = 0.05$, rounding down the actual significance observed, which was $p = 0.053$. He acknowledges the finding was marginal but did not reach conventional statistical significance.
4. In a submitted manuscript, a researcher acknowledges a statistician for her technical assistance without asking the statistician for permission to do so, unaware that it is impolite to acknowledge someone without her permission.
5. A researcher publishes statistically significant results, but statistical significance was attained only by excluding data from two participants after a statistical test for outliers confirmed the appropriateness of removing these participants' data.

6. A researcher conducts a study in which two outcome measures out of four yield significant results. She reports the results of only the two significant measures, but mentions the other two in a description of her methods, despite not reporting findings from them, because these other measures had insufficient reliability.
7. A researcher obtains unexpected results from what she predicted for a study and drafts a manuscript reporting these results as not having been predicted from the start, offering a tentative post hoc explanation for the findings in the paper's discussion.
8. A researcher conducts a study with two independent variables, with an a priori prediction that there should be a statistical interaction between the two factors. However, the statistical interaction is not conventionally significant. He reports in the manuscript that although the statistical interaction was not significant, he conducted post hoc tests, which supported his more specific predictions, but indicated that such findings should be interpreted cautiously in light of the non-significant statistical interaction.
9. Several of a researcher's colleagues have questioned why she has not yet submitted for publication a paper they had all recently coauthored. She hopes to collect data from a few more studies to confirm the initial results, which she hopes would help place the paper in a high-impact journal.
10. A researcher conducts two different but conceptually similar studies that she intends to make into two different papers. As she had seen her graduate school mentor do several times, she re-uses the literature review from the first-drafted paper for the second without disclosing the reuse of this material.
11. Consistent with what often occurs in his field, a researcher separates out several aspects of a large study he recently conducted, making each an independent manuscript, but without disclosing the relationship of the papers to each other.
12. At the suggestion of a journal manuscript peer reviewer, a researcher changes her planned methods of statistical analysis, even though she believes her original methods more accurately represent the data.

References

Anderson, M. S., Horn, A. S., Risbey, K. R., Ronning, E. A., De Vries, R., & Martinson, B. C. (2007). What do mentoring and training in the responsible conduct of research have to do with scientists' misbehavior? Findings from a national survey of NIH-funded scientists. *Academic Medicine, 82,* 853–860.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52*(4), 281–302.

Fanelli, D. (2010). Do pressures to publish increase scientists' bias? An empirical support from US States Data. *PLoS ONE, 5,* e10271.

Fanelli, D. (2011). Negative results are disappearing from most disciplines and countries. *Scientometrics, 90,* 891–904.

Fanelli, D., Costas, R., & Larivière, V. (2015). Misconduct policies, academic culture and career stage, not gender or pressures to publish, affect scientific integrity. *PLoS ONE, 10,* e0127556.

Festinger, L., & Carlsmith, J. M. (1959). Cognitive consequences of forced compliance. *The Journal of Abnormal and Social Psychology, 58,* 203–210.

Franco, A., Malhotra, N., & Simonovits, G. (2014). Publication bias in the social sciences: Unlocking the file drawer. *Science, 345,* 1502–1505.

Goldstein, N. J., Griskevicius, V., & Cialdini, R. B. (2007). Invoking social norms: A social psychology perspective on improving hotels' linen-reuse programs. *Cornell Hotel and Restaurant Administration Quarterly, 48,* 145–150.

Ioannidis, J. P. (2012). Why science is not necessarily self-correcting. *Perspectives on Psychological Science, 7,* 645–654.

John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological Science, 5,* 524–532.

LaRose, R., & Kim, J. (2006). Share, steal, or buy? A social cognitive perspective of music downloading. *CyberPsychology & Behavior, 10,* 267–277.

Leary, M. R., & Kowalski, R. M. (1990). Impression management: A literature review and two-component model. *Psychological Bulletin, 107,* 34–47.

Loken, E., & Gelman, A. (2017). Measurement error and the replication crisis. *Science, 355,* 584–585.

Lynott, D., Corker, K. S., Wortman, J., Connell, L., Donnellan, M. B., Lucas, R. E., et al. (2014). Replication of "Experiencing physical warmth promotes interpersonal warmth" by Williams and Bargh (2008). *Social Psychology, 45,* 216–222.

May, H. (2012). Nonequivalent comparison group designs. In H. Cooper, P. M. Camic, D. L. Long, A. T. Panter, D. Rindskopf, & K. J. Sher (Eds.), *APA handbook of research methods in psychology, Vol 2: Research designs—Quantitative, qualitative, neuropsychological, and biological* (pp. 489–509). Washington, DC: American Psychological Association.

National Academies of Science, Engineering, and Medicine. (2017). *U.S. scientific research enterprise should take action to protect integrity in research; New advisory board on research integrity should be established.* http://www8.nationalacademies.org/onpinews/newsitem.aspx?RecordID=21896. Accessed 23 July 2018.

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science, 34,* aac4716.

Paulhus, D. L. (2002). Socially desirable responding: The evolution of a construct. *The Role of Constructs in Psychological and Educational Measurement, 4969,* 49–69.

Prelec, D. (2004). A Bayesian truth serum for subjective data. *Science, 306,* 462–466.

Ramalingam, S., Bhuvaneswari, S., & Sankaran, R. (2014). Ethics workshops-are they effective in improving the competencies of faculty and postgraduates? *Journal of Clinical and Diagnostic Research: JCDR, 8,* XC01.

Ritchie, S. J., Wiseman, R., & French, C. C. (2012). Failing the future: Three unsuccessful attempts to replicate Bem's 'Retroactive Facilitation of Recall' effect. *PLoS ONE, 7,* e33423.

Roig, M. (2015). *Avoiding plagiarism, self-plagiarism, and other questionable writing practices: A guide to ethical writing*. Office of Research Integrity. https://ori.hhs.gov/education/products/roig_st_johns/Salami%20slicing.html. Accessed 7 Mar 2018.

Sacco, D. F., Bruton, S. V., & Brown, M. (2018). In defense of the questionable: Defining the basis of research scientists' engagement in questionable research practices. *Journal of Empirical Research on Human Research Ethics, 13,* 101–110.

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,* 1359–1366.

Steele, K. M., Bass, K. E., & Crook, M. D. (1999). The mystery of the Mozart effect: Failure to replicate. *Psychological Science, 10,* 366–369.

Tijdink, J. K., Verbeke, R., & Smulders, Y. M. (2014). Publication pressure and scientific misconduct in medical scientists. *Journal of Empirical Research on Human Research Ethics, 9,* 64–71.

Todd, E. M., Torrence, B. S., Watts, L. L., Mulhearn, T. J., Connelly, S., & Mumford, M. D. (2017). Effective practices in the delivery of research ethics education: A qualitative review of instructional methods. *Accountability in Research, 24,* 297–321.

Watts, L. L., Mulhearn, T. J., Medeiros, K. E., Steele, L. M., Connelly, S., & Mumford, M. D. (2017). Modeling the instructional effectiveness of responsible conduct of research education: A meta-analytic path-analysis. *Ethics and Behavior, 27,* 632–650.

Wicherts, J. M., Veldkamp, C. L., Augusteijn, H. E., Bakker, M., Van Aert, R., & Van Assen, M. A. (2016). Degrees of freedom in planning, running, analyzing, and reporting psychological studies: A checklist to avoid p-hacking. *Frontiers in Psychology, 7,* 1832.