Social and Personality Psychology Compass

WILEY

REVIEW ARTICLE OPEN ACCESS

A Credibility Revolution for Relationship Science: Where Can We Step Up Our Game?

Samantha Joel[1] | Paul W. Eastwick[2] | Devinder Khera[1]

[1]Department of Psychology, Western University, London, Canada | [2]Department of Psychology, University of California, Davis, California, USA

Correspondence: Samantha Joel (sjoel2@uwo.ca)

Received: 26 September 2024 | **Revised:** 20 December 2024 | **Accepted:** 3 February 2025

Funding: This work was supported by a Canada Social Sciences and Humanities Research Council (SSHRC).

Keywords: credibility | generalizability | interpersonal relationships | measurement | open science | relationship science | replicability | validity

ABSTRACT

The discipline of psychology is undergoing a credibility revolution whereby researchers are critically evaluating and improving their research practices. In this review, we consider how the field of relationship science could capitalize on this movement in the context of four types of validity. Regarding statistical-conclusions validity, we find that relationship scientists are engaging in open science practices (e.g., preregistration, open data sharing) at similar rates to other fields in the context of personality and social psychology journals. However, journals that are specific to the field (i.e., close relationships journals) could do more to encourage these practices. Meanwhile, new meta-scientific research suggests that the field would benefit greatly from rigorous, widescale measurement validation work (construct validity), novel strategies to account for causal confounds (internal validity), and more diverse representation in our samples and measures (external validity). Overall, the credibility revolution offers several specific, actionable recommendations to improve the validity of research findings, many of which are highly relevant to relationship science.

1 | Introduction

The past decade has been marked by a sustained effort to improve research practices within the field of psychology. This credibility revolution was first kicked off by a series of high-profile events in the early 2010s (e.g., Bem 2011; Open Sci. Collab., 2012; Simmons, Nelson, and Simonsohn 2011), each of which raised serious concerns about how much we should trust psychology's published body of work. Subsequently, researchers developed a suite of new guidelines, tools, and organizations for improving the transparency, integrity, and overall validity of psychological research findings (see Nosek et al. 2022; Vazire, Schiavone, and Bottesini 2022 for recent reviews). This movement is transforming the way that psychologists conduct their science, both with the uptake of open science practices (e.g., Hardwicke et al. 2022), and through the reconsideration of broader research practices, including measurement (e.g., Flake and Fried 2020; Hussey and Hughes 2020), statistical (e.g., Maul 2017; Westfall and Yarkoni 2016), and sampling practices (e.g., Moshontz et al. 2018).

The current review centers on a particular subfield within psychology: the study of close relationships. Relationship science, with its roots in the study of attraction and liking, has close ties to social psychology and personality (Berscheid and Reis 1998). At the same time, relationship science occupies a unique interdisciplinary space, spanning research in human development, communication, and family studies, among other disciplines (see Sharkey, Feather, and Goedeke 2022 for review). Relationship science also brings specific methodological considerations, such as the collection of sensitive data from multiple people (e.g., dyads; Kenny, Kashy, and Cook 2006) who are often studied over time (e.g., Bolger and Laurenceau 2013). Over the past 10 years, the field of relationship science has intersected

with the credibility revolution in unique ways, based on these field-specific features and considerations (Chopik et al. 2020; Finkel, Eastwick, and Reis 2015). In the current review, we seek to showcase recent efforts within relationship science to improve our research practices, and discuss new improvements that the field can work toward in coming years.

Recent reviews of the credibility revolution (e.g., Fabrigar, Wegener, and Petty 2020; Vazire, Schiavone, and Bottesini 2022) have organized the movement according to four types of validity (Cook and Campbell 1979). Statistical conclusion validity—whether a study's statistical conclusions are accurate—has received the most attention to date. However, more recent efforts to improve psychological practices have also focused on construct validity (do measures capture their intended constructs?) internal validity (are causal inferences warranted?), and external validity (do findings generalize)? We have used the same conceptual framework to organize the present review, by considering specific threats to each type of validity that may emerge in the context of close relationship research. Note that some of the methodological issues discussed are of broad importance and therefore transcend these validity categories; nevertheless, we discuss each issue in the context of the type of validity that it affects most strongly. Finally, we consider practical solutions that have, or could, be implemented to help address these validity concerns.

2 | Statistical Conclusion Validity

The most-studied type of validity concerns the accuracy of a study's statistical inferences. For example, imagine a study showing that people who experience more relationship conflict are significantly less satisfied with their romantic partners. This finding is statistically valid to the extent that the study's statistical conclusion—the significant association between conflict and satisfaction—is not a false positive (i.e., a Type 1 error). Statistical-conclusion validity can be threatened by features of the study design (e.g., a lack of statistical power), questionable research practices (e.g., p-hacking, selective reporting), and error (e.g., mistakes in data coding and labeling), among other factors.

Several practices have been developed to combat these problems, including: (1) **Preregistration**. Using websites like the Open Science Framework and AsPredicted.org, researchers can publicly post their study's analytic plan before running any analyses. This practice, when used properly, can distinguish between confirmatory and exploratory analyses. (2) **Power analyses.** Does the study have a sufficiently large sample size to reliably detect (or falsify) the effect of interest? Counterintuitively, underpowered studies are at particular risk of generating false positives (e.g., Button et al. 2013). (3) **Open sharing.** Researchers can publicly post their data, materials and code, allowing other researchers to verify the reproducibility and robustness of their results.

2.1 | Has Relationship Science Adopted Open Science Practices?

All three of these practices are more complicated in the context of close relationships research compared to the standard one-shot laboratory experiments where these practices were honed and developed. Preregistration choices can be complex in the context of large datasets containing many variables, and the effectiveness of different preregistration approaches depend on the goals of the preregistration (e.g., transparency, combating publication bias, theory falsification, Type-I error control; da Silva Frost and Ledgerwood 2020). Power analyses involve more guesswork when couples and repeated observations are involved (e.g., Bolger and Laurenceau 2013). Regarding data sharing, relationships researchers often collect sensitive data that cannot be shared, at least without careful de-identification efforts (Joel, Eastwick, and Finkel 2018). Initial confusion about how to adapt these practices to close relationships research may have slowed their initial uptake.

Nevertheless, today there are new tools and guidelines that can enable relationship scientists to adopt open science practices. For example, user-friendly power analysis tutorials have been released for use with popular relationship modeling techniques, such as multilevel modeling (Lane and Hennes 2018) and the actor-partner interdependence model (Ackerman and Kenny 2016). Greater conceptual clarity around preregistration has highlighted how relationships researchers can achieve the most common goals of preregistration, simply by specifying an as-complete-as-possible set of analyses that test an idea before beginning those analyses (Beer, Eastwick, and Goh 2023). A growing number of services are available to help researchers share sensitive data (e.g., couples data) while still protecting participant confidentiality (Joel, Eastwick, and Finkel 2018). Meanwhile, there have been many high-profile examples of open science practices being successfully applied to close relationships research, such as well-powered, preregistered longitudinal studies (e.g., Bühler and Orth 2024; Gerlach et al. 2019), replications (e.g., Cheung et al. 2016; Floyd and Woo 2020) and registered reports (e.g., Baker et al. 2020; Eastwick et al. 2024; Sakaluk, Fisher, and Kilshaw 2021).

Against this backdrop, what is the adoption rate of open science practices within relationship science? To gauge the field's current open science uptake, we downloaded a total of 1110 articles on both relationship and non-relationship topics from relevant journals published between January 2022 and June 2024. Using R, we scraped all articles for a series of terms related to open science practices (e.g., "Open Science Framework", "AsPredicted.org", "preregistered"), as well as terms related to *not* engaging in these practices (e.g., "not preregistered", "available upon request"). Further, we manually coded a subset of articles (n = 337). Coding scripts and data are available on OSF (https://osf.io/ukcz6/), and a detailed account of the coding process is available in the Supporting Information S1.

The scraping and coding results are displayed on Tables 1 and 2, respectively. These data paint a nuanced picture of open science engagement within relationship science. On one hand, close relationships papers published in social/personality psychology journals appear to have rates of open science engagement that are on-par with papers about other topics that are published in social/personality journals. Notably, these journals tend have guidelines that specifically encourage and incentivize open science practices (e.g., requiring statements about power, preregistration, and data availability). For example, *Journal of Personality and Social*

TABLE 1 | Scraping for open science terms by topic and by journal.

	Journal	*n*	OSF	As-predicted	Not preregistered	Available upon request
Papers in social/Personality journals	JPSP (Section 1)	81	91%	28%	26%	1%
	JPSP (Section 2)	86	92%	31%	24%	5%
	JPSP (Section 3)	94	96%	5%	36%	5%
	Social cognition	59	92%	8%	8%	5%
	JRP	221	72%	1%	45%	4%
	Total	541	84%	12%	33%	4%
Relationship papers in social/Personality journals	JPSP	18	100%	23%	23%	8%
	PSPB	24	83%	4%	21%	4%
	SPPS	21	95%	10%	29%	0%
	Total	63	90%	11%	24%	5%
Papers in relationships journals	JSPR	432	29%	1%	74%	25%
	PR	37	19%	0%	11%	8%
	JMF	55	13%	0%	0%	18%
	Total	524	27%	1%	62%	23%

Note: Note that articles sampled from the Social Cognition and JRP were primarily about non-relationship topics; therefore, these journals were excluded from relationship-specific coding analyses.
Abbreviations: JMF, Journal of Marriage and Family; JPSP, Journal of Personality and Social Psychology; JRP, Journal of Research in Personality; JSPR, Journal of Social and Personal Relationships; PR, Personal Relationships; PSPB, Personality and Social Psychology Bulletin; SPPS, Social Psychological and Personality Science.

TABLE 2 | Manual coding for open science practices by topic and by journal.

	Journal	*n*	Pre-registration	Power analysis	Open data sharing
Papers in social/Personality journals	JPSP (Section 1)	25	76%	96%	96%
	JPSP (Section 2)	50	74%	72%	86%
	JPSP (Section 3)	25	64%	52%	92%
	Social cognition	25	52%	92%	80%
	JRP	25	24%	60%	52%
	Total	150	61%	74%	82%
Relationship papers in social/Personality journals	JPSP	18	78%	72%	94%
	PSPB	24	33%	88%	54%
	SPPS	21	48%	52%	76%
	Total	63	51%	71%	73%
Papers in relationships journals	JSPR	50	14%	12%	28%
	PR	37	5%	19%	16%
	JMF	55	0%	9%	4%
	Total	142	6%	13%	15%

Psychology (JPSP) and *Personality and Social Psychology Bulletin* (PSPB) have both adopted the Center for Open Science's Transparency and Openness Promotion (TOP) guidelines. Journals are awarded TOP scores based on their degree of compliance with transparency guidelines, with a maximum score of 29 (see rubric here). JPSP and PSPB both have relatively high TOP scores (20 and 17, respectively).

In contrast, open science engagement specifically within close relationships journals appears to be quite poor, including when compared to specialty journals from neighboring fields (e.g., social cognition). It is worth noting that the close relationships journals tend to have fewer guidelines encouraging open science adoption. Whereas Social Cognition and JRP have TOP scores of 11 and 21, JSPR, PR, and JMF have TOP scores of 4, 1, and 0,

respectively. The policies that *have* been enacted are reflected in the data. For example, JSPR requires a data availability statement, whereas JMF does not. Of the articles we manually coded, 28% of JSPR papers shared the data from at least one study, compared to 4% of JMF papers. Of course, we cannot draw strong causal conclusions from these exploratory, descriptive results. Nevertheless, the data are consistent with the argument that journal policies have considerable power to shape open science practices within the field. Relationship scientists appear to be willing and able to preregister studies, conduct power analyses, and share data when incentivized to do so.

In sum, our results suggest that relationship science has made meaningful strides toward adopting open science practices (specifically, preregistration, power analyses, and open data sharing). However, adoption of these practices is lagging within the specialty close relationships journals. Policies that explicitly encourage these practices could go a long way toward encouraging their adoption in the context of those journals.

3 | Construct Validity

Whereas psychology's meta-scientific efforts in the 2010s were largely focused on improving the validity of statistical conclusions, the credibility movement has more recently shifted attention to other aspects of the research process, particularly measurement. A growing body of research suggests that measurement problems pose a meaningful threat to the validity of psychological findings (e.g., Elson et al. 2023; Flake and Fried 2020; Hussey and Hughes 2020; Lilienfeld and Strother 2020; Maul 2017; Schimmack 2021). One investigation probed the validity of 15 commonly used measures from social psychology and personality (Hussey and Hughes 2020). In a very well-powered sample ($N = 144{,}496$), many of these measures failed various structural validity tests; only 73% fit their expected factor structures, and only 4% (one of the 15 scales) showed evidence of measurement invariance for age and gender.[1] Further, recent reviews of papers published in *Journal of Personality and Social Psychology* suggest that self-report measures (Flake, Pek, and Hehman 2017) are frequently created "on the fly" for a specific study, with very little validity testing or evidence. Thus, psychology suffers from a proliferation of unstandardized, ad hoc measures, most of which are used only a handful of times (Elson et al. 2023). Even when measures are subjected to validity testing, there is evidence that popular methods of doing so (e.g., by examining the measure's reliability, factor structure, and correlations) do not provide sufficiently rigorous tests of a measure's construct validity (Maul 2017).

3.1 | Construct Validity Threats and Relationship Science

The field of relationship science possesses an enormous wealth of constructs and associated measures pertaining to close relationship functioning (see Finkel, Simpson, and Eastwick 2017 for review). To date, there have been relatively few attempts to interrogate the content, structure, and performance of these measures at large. However, the studies that have done so (e.g., CORE Lab 2024; Delatorre and Wagner 2020; Joel et al. 2020; Kim et al. 2024) raise important concerns about the state of *construct validity* within relationship research: the extent to which these measures are valid representations of their intended constructs. Below, we focus on two specific threats to construct validity and their relevance to relationship science: jingle-jangle fallacies, and shared method biases.

3.1.1 | Jingle-Jangle Fallacies

Jingle-jangle fallacies (Kelley 1927) occur when ostensibly similar measures are in fact representing different constructs (jingle), or when ostensibly distinct measures represent the same construct (jangle). Emerging meta-scientific research in relationship science points to jangle measurement issues in particular regarding the construct of ***relationship quality***. Relationship quality—a person's global evaluation of whether their relationship is good or bad—is a central construct within relationship science. Yet, this construct remains conceptually fuzzy, such that researchers reasonably disagree about what falls within versus outside its bounds. Many of the widely used measures of relationship quality were originally developed atheoretically and without sufficient validity testing (Delatorre and Wagner 2020). There is a lack of consensus about terminology (terms such as satisfaction, adjustment, and quality are used in the literature interchangeably), whether relationship quality is unidimensional or multidimensional, as well as what (if anything) the multiple dimensions of relationship quality might be (CORE Lab 2024; Delatorre and Wagner 2020). Meanwhile, one research project investigated a series of bifactor models conducted on over 30 measures of ostensibly different relationship constructs (trust, conflict, power, etc.) across two representative samples (combined $N = 3439$; Kim et al. 2024). Results suggested that over 70% of the covariance across these measures was captured by a single global evaluative factor, with weak evidence for any additional, substantive factors. In other words, after removing the evaluative core that connected these items, there was no evidence that distinct, coherent factors remained (i.e., items assessing "trust," e.g., were no more related to each other than they were to items assessing "conflict"). Many measures intended to capture distinct relationship domains may in fact be primarily capturing relationship quality.

3.1.2 | Shared Method Biases

Another threat to construct validity is when the measure inadvertently taps into methodological artifacts. Different measures can have overlapping variance not because of their substantive content, but because of the way the measures were collected (Podsakoff et al. 2003, 2024). Shared method biases such as the participant's response style or current mood may lead a person to answer different survey measures in a similar way, artificially inflating the associations between those measures. Researchers in fields such as organizational research (Williams and McGonagle 2016), marketing (Baumgartner, Weijters, and Pieters 2021), advertising (Malhotra, Schaller, and Patil 2017), and management (Jakobsen and Jensen 2015) have demonstrated the seriousness of common method variance as a validity threat,

and developed strategies to address it. Recommended procedural remedies include collecting data at different timepoints, and/or from multiple sources (Podsakoff, MacKenzie, and Podsakoff 2012). Recommended statistical remedies include detecting and isolating method factors with the use of structural equation modeling (e.g., Ding, Chen, and Jane 2023) and marker variables (e.g., Williams and O'Boyle 2015).

Comparatively few papers have attended to method biases specifically within relationship science (see Joel et al. 2024; Orth 2013; Wang and Eastwick 2020 for exceptions). However, the field's strong reliance on self-report methods (Williamson et al. 2022) suggests that its findings should indeed be susceptible to method biases, akin to those that have been uncovered in other self-report-heavy fields (e.g., social desirability bias, Nederhof 1985; halo bias, Nisbett and Wilson 1977; the bloated specific, Rhemtulla, Borsboom, and van Bork 2017). One relationship-specific source of method bias that has received surprisingly little attention is *sentiment override*. When a person is happy with their romantic partner, those positive sentiments can greatly shape their perceptions of more specific aspects of their partnership (Weiss 1980). This phenomenon has been well-documented in the context of couple interaction studies: couples come into the lab with expectations about how their interaction with their partner will go, which in turn shape their perceptions of how the interaction went (Fincham et al. 1995; Hinnekens et al. 2020; McNulty and Karney 2002; Waldinger and Schulz 2006). In the context of self-report surveys, sentiment override may function as a heuristic to help people arrive at specific relationship judgments ("I love my partner a lot, so they probably did the dishes yesterday"). Methodologically, this process may make it difficult for researchers to accurately capture specific aspects of a relationship ("My partner did the dishes yesterday") without inadvertently capturing global sentiments as well ("I love my partner").

We have already discussed recent evidence suggesting that ostensibly different relationship measures are *primarily* tapping into global relationship evaluations, rather than their intended constructs (Kim et al. 2024). However, even if we accept a softened version of this argument—whereby different self-report measures *partially* tap into global relationship sentiments *in addition* to their intended constructs—that shared method effect would nevertheless pose an important problem for relationship science. To demonstrate, researchers recently generated a relationship scale consisting of irrelevant focal words (e.g., "The daffodil in my relationship is close to ideal"; Joel et al., 2024). Despite the items having no clear substantive meaning, this scale was moderately correlated with other relationship measures, and predicted relationship outcomes 3 weeks later. Qualitative and quantitative evidence suggested that this nonsensical scale was in fact tapping into people's current relationship evaluations. Overall, these results suggest that sentiment override—whereby people project their global relationship sentiments onto any relationship-relevant item—can result in inflated or even spurious associations between measures. A relationship measure can appear to predict important outcomes not because of the substantive construct it is intended to represent, but because any measure that asks participants to report on a relationship—however nonsensical—will capture global sentiments.

3.1.3 | Coding the Prevalence of Single-Source Effects

How vulnerable are close relationship findings to common method variance problems? To find out, we re-examined the coding results from Williamson et al. (2022), using the data that the authors made available on OSF. Specifically, we examined how frequently close relationships papers rely on self-report data, collected from the same person at the same timepoint. Common method variance should be particularly high for these effects (Podsakoff et al. 2003), meaning that they should be particularly susceptible to method biases like sentiment override, the halo bias, and the social desirability bias, among others.

Williamson et al. (2022) coded the data collection methods of 771 relationship papers. Self-report was indeed the most common form of measurement (96% of studies), followed by behavioral observation (i.e., video and/or audio; 11%). Other measurement approaches, such as open-ended measures (e.g., interviews; 3%) and physiological measures (2%), were relatively rare. Further descriptive analysis of Williamson et al.'s data (available in their supplemental materials) shows that 83% of papers used self-report as their *only* form of measurement. Recruiting dyads (e.g., both members of a couple) was relatively common (42%), as was the use of longitudinal (24%), experimental (19%), and diary methods (9%). Nevertheless, 30% of papers used only individual, cross-sectional, non-experimental data. Finally, 27% of papers reported on only *self-report measures* collected in the context of individual, cross-sectional, non-experimental data, which are likely to be highly susceptible to shared method variance issues.

We also completed our own descriptive coding for 95 close relationships articles, sampled from JPSP (n = 21) and JSPR (n = 74; see Supporting Information S1 for details). Corroborating the Williamson et al. (2022) data, a sizable number of these papers reported on data collected either from multiple people (38%), from multiple timepoints (46%), with the use of experimental methods (17%), or behavioral observations (15%). Nevertheless, and consistent with the 27% result from Williamson et al. (2022), 35% of the papers we coded used only cross-sectional, non-experimental, self-report evidence collected from one person. Again, such papers should be highly susceptible to the sort of shared method variance problems discussed by Podsakoff and colleagues (2024).

It may be tempting to conclude from these findings that relationship scientists should simply move away from self-report. However, emerging work suggests that alternative measurement options can face even larger validity threats. For example, implicit measures—which aim to assess thoughts and feelings via automatic processes—have gained considerable popularity as an alternative to self-report. Yet, many implicit measures have been shown to have low internal consistency (Greenwald and Lai 2020), low test-retest reliability (Gawronski et al. 2017), and poor discriminant validity (Schimmack 2019), among other issues. A recent review of their relative advantages and disadvantages concluded that "self-reports are most often the better measurement option" (Corneille and Gawronski 2024). As another example, many researchers who study individual differences (e.g., risk-taking preferences, self-control) have attempted to move away from self-report by instead assessing real behavior

in the laboratory. Yet, compared to self-report, these behavioral measures often show weaker convergence across measures and lower test-retest reliability (e.g., Dang, King, and Inzlicht 2020; Frey et al. 2017). Together, these findings underscore the importance of interrogating any measure chosen to represent a given construct, self-report or otherwise.

In sum, there is a pressing need for the field of close relationships to rigorously re-examine the validity of its measures. Like in any field, the validity of close relationship findings hinges upon adequate construct validity, which can be undermined by various measurement problems (e.g., jingle-jangle fallacies, shared method biases). It is time to take greater advantage of the tools and recommendations that have been developed in related research areas to assess and improve our measurement practices. Recent efforts to re-examine close relationship measures have focused on self-report (e.g., CORE Lab 2024; Delatorre and Wagner 2020; Kim et al. 2024); these efforts should be heeded and expanded upon. At the same time, researchers who seek to reduce their reliance on self-report must carefully consider and test the validity of other measurement approaches. Overall, large-scale validity assessments—which have been quite rare within relationship science to date—will be crucial for increasing confidence in the validity of our constructs.

4 | Internal Validity

A third type of validity being re-examined by the credibility revolution is *internal validity*. Are our causal inferences valid, and have plausible alternative explanations been convincingly ruled out?

4.1 | The Challenge of Ruling Out Confounds

Researchers who use correlational designs often attempt to strengthen their causal claims by controlling for third variables. For example, a researcher interested in the effects of conflict *strategy* might control for the related influences of conflict *frequency* by adding it as a covariate to the model. However, recent meta-scientific papers have questioned the effectiveness of this strategy, at least as it is typically used (e.g., Rohrer 2018; Westfall and Yarkoni 2016; Wysocki, Lawson, and Rhemtulla 2022). There are two potential pitfalls to consider. The first is that researchers often include control variables simply because they are related to the construct of interest (Wysocki, Lawson, and Rhemtulla 2022), without careful consideration for their causal role in the model (Rohrer 2018). Poorly justified covariates can introduce bias by parsing out relevant sources of variance (e.g., a person's conflict strategy may have a great deal to do with their conflict frequency). Notably, this bias can go in either direction, meaning that poorly selected covariates have the potential to generate spurious, significant effects.

A second pitfall is measurement error, which renders many attempts at statistical control incomplete. Covariates are often used in the context of regression-based models (e.g., multiple regression models, multilevel models), which assume that each measure perfectly represents its intended construct (i.e., they do not model measurement error). Using simulated data, Westfall and Yarkoni (2016) demonstrated that even modest amounts of measurement unreliability can produce enormously inflated Type I error rates in this kind of model—in some cases even approaching 100%. In other words, even if a predictor in a regression is associated with an outcome above and beyond an appropriately selected covariate, that predictive association is still likely to be spurious unless measurement error has been modeled.

These critiques, lobbied at the field of psychology broadly, are highly relevant to relationship science. Particularly given the prevalence of non-experimental approaches (Williamson et al. 2022), the causal conclusions of close relationships studies rely heavily on the ruling out of potential confounds (see Kenny 2019). Many close relationships studies attempt to rule out confounds with the use of longitudinal or daily experience designs, which separate within-person from between-person variance. Although such designs can certainly be helpful for making causal inferences, they do not necessarily account for confounds that can vary across time along with the variables of interest (e.g., mood, stress, health; Rohrer and Murayama 2023). Statistically controlling for third variables is also common, particularly as a way to test the incremental validity of a new construct (Wang and Eastwick 2020). However, relationship science tends to favor statistical techniques that do not incorporate measurement error, particularly multilevel modeling (see Sakaluk et al. 2024 for review). As Wang and Eastwick (2020) discuss, structural equation modeling may be better suited to address the Type 1 error risks laid out by Westfall and Yarkoni (2016).

4.2 | Experimental Approaches

Although somewhat uncommon, experimental approaches in relationship science can be both creative and informative. For example, the classic "fast friends" paradigm is a highly effective procedure for generating real feelings of closeness between people in a lab environment (Aron et al. 1997). Trained actors can be effective for demonstrating the causal effects of different social behaviors, particularly in the context of interactions with strangers (e.g., Reis et al. 2010). In the context of ongoing close relationships, virtual reality technology has opened avenues for creating immersive yet controlled experimental paradigms. In one study, participants in a virtual environment were asked to walk along a cliff overlooking a canyon while an avatar, ostensibly representing their romantic partner, was randomly assigned to be either attentive (waving, nodding, etc.) or inattentive (oriented away from the participant; Kane et al. 2012). For a more "in vivo" approach, some studies have randomly assigned people to real relationship experiences, such as marital interventions (e.g., the marriage checkup; Cordova et al. 2014), and relationship choices (e.g., the decision to merge finances with a spouse; Olson et al. 2023).

Like other empirical tools (e.g., self-report measures), the validity of an experiment rests partly on the *construct validity* of its manipulation. Experimental manipulations should therefore be carefully validated to ensure that they are indeed manipulating

their construct of interest as intended (Cook and Campbell 1979). Yet, the practice of validating experimental paradigms is lacking within the broader personality and social psychology literature. A review of experimental studies published in *Journal of Personality and Social Psychology* in 2017 found that most experimental manipulations (80%) were created ad hoc for the purposes of a particular paper (Chester and Lasko 2021). Further, many of the studies reviewed (42%) did not include any validation evidence for their manipulation beyond face validity (i.e., no pilot studies, no manipulation checks, and no prior studies that used the manipulation). Going forward, Chester and Lasko (2021) recommend that researchers (a) use previously published manipulations whenever possible, (b) validate their manipulations with pilot studies, and (c) include manipulation checks in pilot testing (which should have ideally been validated in prior work). Researchers should also consider including measures of related yet distinct constructs (i.e., discriminant validity checks), to show that the manipulation only weakly influences related constructs, and does not influence unrelated constructs.

These recommendations are sound advice for experimentalists broadly, including those who are interested in interpersonal processes. A key challenge for experimental designs is that it can be very difficult to manipulate a specific construct without inadvertently manipulating other variables as well (Eronen 2020). Yet, the inclusion of discriminant validity checks—measures that an experimental manipulation should affect weakly or not at all—is rare in psychological experiments (for a study that includes such a check, see da Silva Frost and Eastwick 2024). In the context of close relationships research, it seems particularly worthwhile to ensure that relationship manipulations target their intended constructs and not others, given that many relationship constructs have a high degree of conceptual overlap (see CORE Lab 2024; Kim et al. 2024). For example, imagine that a researcher attempts to manipulate trust by asking participants to recall a time when their romantic partner let them down. The researcher may validate their manipulation with a manipulation check and show that indeed, the manipulation temporarily lowers feelings of trust. However, to further increase confidence in the validity of the manipulation, it would be valuable to *also* include discriminant validity checks (e.g., gratitude, commitment, satisfaction), to show that the manipulation does not influence other, related relationship evaluations to a similar degree. An experimental manipulation intended to manipulate something specific (e.g., trust) that in fact manipulates something broader (e.g., global satisfaction) would pose an important confound for the resulting study, just as it does in other fields (Eronen 2020).

In sum, there has been little meta-science examining the role of causal confounds in relationship science. The relevant data we do have suggest that, together with construct validity threats, internal validity threats are worthy of careful re-examination in the context of close relationships research. As a first step, it is recommended that relationship scientists adopt techniques that are capable of incorporating measurement error, such as structural equation modeling (Sakaluk et al. 2024; Wang and Eastwick 2020). New techniques and packages for this express purpose are increasingly available to relationship scientists, such as dySEM (Sakaluk and Camanto 2024). Partnerships with experimentalists should also be encouraged. Finally, researchers should follow current best practices regarding the validation of experimental paradigms (Chester and Lasko 2021).

5 | External Validity

At last, we turn our attention to external validity: how generalizable are our research findings? Building on discussions around psychology's replicability (e.g., Open Sci. Collab., 2012) and measurement challenges (e.g., Flake and Fried 2020), there are growing concerns that psychology also faces generalizability challenges, whereby findings do not necessarily extend beyond the particular contexts in which they were tested. For example, continuing with the issue of construct validity, researchers often assume that the stimuli used in their studies are perfect stand-ins for their intended constructs (e.g., Yarkoni 2022). If this assumption is violated, it becomes harder to generalize from a given statistical finding (e.g., "participants assigned to Experimental Condition X provided higher scores on Self-Report Measure Y") to a more interesting verbal conclusion (e.g., "conflict decreases relationship satisfaction"). Indeed, a recent ManyLabs investigation showed that different research teams tend to reach very different conclusions about the same research questions, likely due to the heterogeneity in their measurement and design choices (Landy et al. 2020).

Another key generalizability critique is that psychology samples tend to be highly unrepresentative of the broader population (e.g., Brady, Fryberg, and Shoda 2018; Cheon, Melani, and Hong 2020; Gurven 2018; Henrich, Heine, and Norenzayan 2010; Rad, Martingano, and Ginges 2018; Sabik et al. 2021; Thalmayer, Toscanelli, and Arnett 2021).[2] A review of articles published in *Psychological Science* found that most samples were from North American and European countries, particularly the United States (Rad, Martingano, and Ginges 2018). Asia, Africa, and Latin America were all dramatically underrepresented; indeed, 0% of *Psychological Science* articles published in 2017 included samples from either Africa or Latin America. Further, most papers did not provide demographic information such as race or ethnicity (80%), SES (92%), or education level (48%), preventing the authors from even drawing conclusions about sample diversity along these dimensions. Publications are particularly likely to emphasize race or nationality when the sample is not primarily white or American, respectively, suggesting that researchers see white, American samples as more representative and generalizable than other samples (Cheon, Melani, and Hong 2020). Overall, psychology has traditionally focused on the perspectives of a very specific slice of humanity (white, American, educated, heterosexual, able-bodied, affluent people), and assumed that those perspectives generalize globally. This practice seriously undermines the external validity of research findings while also reinforcing racial and cultural power structures (Brady, Fryberg, and Shoda 2018; Remedios 2022; Syed and Kathawalla 2022).

5.1 | How Generalizable Is Close Relationships Research?

There have been numerous efforts to document the diversity, or lack thereof, of samples within relationship science (e.g., Karney,

Kreitz, and Sweeney 2004; McGorray et al. 2023; Pollitt, Blair, and Lannutti 2023; Randall and Curran 2023; Williamson et al. 2022). In one review, Williamson et al. (2022) coded 771 articles from five top relationship journals from 2014 to 2018. In line with the broader psychology literature (e.g., Rad, Martingano, and Ginges 2018), the vast majority of the samples were recruited from the United States (73%), other English-speaking countries (12%), and Europe (10%). In contrast, Asia (3%), Latin America (< 1%), Africa (< 1%) were almost entirely unrepresented. Within the US samples, white participants were oversampled relative to the US population. Meanwhile, the specific racial identities of non-white participants (e.g., Asian, Black, Indigenous) were underreported, creating challenges for the coding team. Relationship journals can help combat this problem by requiring complete, detailed demographic records about participants. Some journals have already adopted this practice; for example, JSPR now strongly encourages researchers to report on relevant details about their samples, such as race/ethnicity, gender identity, sexual orientation, age, and the location of data collection. Guidelines at PSPB (inspired by the ACEMAP task force; Ledgerwood et al. 2024) note how particular cultural contexts render some demographics more relevant and salient than others (e.g., caste or religion may be particularly salient in some world regions).

In another recent review, McGorray et al. (2023) coded 1762 studies from 1084 articles on close relationships, published between 1996 and 2000 and between 2016 and 2020. Some sampling and reported practices had improved over the last 20 years (e.g., the reporting of sexual orientation and race). However, sample diversity remained quite low across time; for example, most samples at both timepoints were recruited from the US, and the median percentage of LGBTQ + participants within samples was 0%. LGBTQ + participants are often intentionally excluded from relationship research (Andersen and Zou 2015); indeed, some analytical approaches are designed specifically for use with other-sex couples (distinguishable dyads). Further, as in psychology more broadly (e.g., Cheon, Melani, and Hong 2020), there was evidence that researchers treated white American participants as the default. For example, mirroring the results of Williamson et al. (2022), race was often reported in a way that centered Whiteness, lumping all other racial categories together ("the sample was 63% white").

Relationship science's lack of sample diversity has implications for the generalizability not only of its findings, but also its measures. Measures are typically assumed to capture the same constructs when administered to different groups. However, this assumption of measurement invariance cannot be directly tested unless the measures are validated with sufficiently diverse samples. For example, observational coding schemes used for couple interaction research have been primarily developed with white, American, heterosexual couples (Friedlander, Lee, and Escudero 2019); their validity in more diverse contexts remains unknown. Assuming measurement invariance where there is none can lead to biased results and false positives (e.g., Jeong and Lee 2019). Meanwhile, results of a recent registered report on dyadic measurement invariance suggest that some important relationship measures (e.g., commitment and sexuality measures) are not even invariant across gender (Sakaluk, Fisher, and Kilshaw 2021).

A robust, generalizable literature requires diversity in terms of not just the participant samples, but also the researchers themselves. When White Western researchers develop psychological theories and methodologies without incorporating diverse perspectives, they often assume that their resulting scholarship is objective and neutral, when it is in fact steeped in unacknowledged cultural biases (Lewis Jr 2021). For example, much of the relationship communication literature is built on the idea that open, direct communication strategies are superior to indirect communication strategies. Yet, this is likely a culturally specific principle—based on individualistic values—that may not generalize well to cultures outside of North America (Williamson 2024a). Most problematically, North American models of "effective communication" have frequently been exported to other cultures with the implicit assumption that they are universal. For example, marriage and family therapists in East Asia are frequently trained in practices originally developed in the United States (Tseng et al. 2020). Such practices need to be adapted to local cultural contexts. For example, one research team recently developed a set of core competencies for family therapists in Taiwan (the Taiwan Family Therapist Core Competencies; Chao and Lou 2018). Overall, findings are likely to be more generalizable when diverse voices are incorporated at every stage of the research process.

5.2 | Improving the Diversity and Generalizability of Relationship Research

An important initial step toward improving the generalizability of relationship research is for researchers to more accurately situate their results within the context of their samples (McGorray et al. 2023; Williamson et al. 2022). In addition to providing detailed participant demographics, researchers can include a Constraints on Generality statement that specifies which populations their findings could or could not reasonably be expected to generalize to (Simons, Shoda, and Lindsay 2017). More ambitiously, researchers can utilize new tools and paradigms to greatly diversify both their samples and their collaborative teams (e.g., Williamson 2024b), such as Many Labs (e.g., Klein et al. 2014) and the Psychological Science Accelerator (Moshontz et al. 2018). For example, one recent collaboration partnered with the Psychological Science Accelerator to examine ideal partner preferences across 43 countries and 22 languages (N = 10,358; Eastwick et al. 2024). This project provided both a large-N replication of prior (semi-contentious) findings while simultaneously documenting the degree to which these findings varied across different cultural contexts.

Finally, there is much that could be done to improve the generalizability of relationship science's methodologies. In the context of self-report measures, this work could include translating popular measures into different languages (Benet-Martinez 2007), and probing them for measurement invariance (e.g., Ogan and Kanter 2024), including at the dyadic level (Sakaluk, Fisher, and Kilshaw 2021). In some cases, new or adapted measures may be required to capture culturally specific phenomena that are missed by existing instruments. Meanwhile, tools for behavioral observation (e.g., couple observation

paradigms and coding schemes) could be modified to be more appropriate across different cultural contexts (Friedlander, Lee, and Escudero 2019; Williamson 2024a).

6 | Conclusions

The discipline of psychology is in the midst of a methodological renaissance, whereby researchers are critically re-examining best practices for every stage of the research process. We posit that, like other fields, relationship science has a great deal to gain from this movement. In addition to continuing to encourage the adoption of open science practices (e.g., preregistration, power analyses, and data sharing), there are many exciting new approaches and tools the field could adopt to help improve the credibility of its findings. The field would strongly benefit from large-scale validity assessments of its self-report measures (e.g., CORE Lab 2024; Kim et al. 2024), as well as its experimental manipulations (Chester and Lasko 2021), observational coding schemes (e.g., Williamson 2024a), and other forms of assessment. Stronger efforts should be made to account for method biases (e.g., Joel et al. 2024; Orth 2013), such as with the use of structural equation modeling (Sakaluk et al. 2024; Wang and Eastwick 2020). Large, multi-lab collaborations (e.g., with the Psychological Science Accelerator, Moshontz et al. 2018) may hold particular promise for helping to diversify the field's samples, measures, and researchers. Overall, measurement challenges seem poised to take center stage in relationship science in future years, particularly as the field seeks to become more globally representative.

Acknowledgments

This work was supported by a Canada Social Sciences and Humanities Research Council (SSHRC) Insight grant awarded to Samantha Joel. The authors would also like to thank Dr. John Sakaluk for supplying R code to assist with the article scraping process.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data that support the findings of this study are openly available in OSF at https://osf.io/ukcz6/.

Endnotes

[1] Although measurement invariance is treated as a dichotomous test in this paper (which 14 of 15 tested measures failed), measurement invariance is really a spectrum, such that some measures possess more or less invariance than others (see critique by Wetzel and Roberts 2020).

[2] We have avoided using the popularized WEIRD acronym (Western, Educated, Industrialized, Religious, Democratic), because its dimensions have been criticized for being atheoretical and reductionist (e.g., Rochat 2010; Sakaluk and Daniel 2022; Syed 2021). The frequent dichotomization of "WEIRD" versus "non-WEIRD" countries artificially lumps certain categories together (e.g., not all educated, industrialized countries are Western). Meanwhile, the acronym entirely ignores the role of race.

References

Ackerman, R. A., and D. A. Kenny. 2016. "APIMPowerR: An Interactive Tool for Actor-Partner Interdependence Model Power Analysis [Computer Software]." https://robert-a-ackerman.shinyapps.io/APIMPowerRdis/.

Andersen, J. P., and C. Zou. 2015. "Exclusion of Sexual Minority Couples From Research." *Health Science Journal* 9, no. 6: 1–9.

Aron, A., E. Melinat, E. N. Aron, R. D. Vallone, and R. J. Bator. 1997. "The Experimental Generation of Interpersonal Closeness: A Procedure and Some Preliminary Findings." *Personality and Social Psychology Bulletin* 23, no. 4: 363–377. https://doi.org/10.1177/0146167297234003.

Baker, Z. G., E. A. Gentzis, E. M. Watlington, et al. 2020. "Reflections on a Registered Report Replicating a Body of Dyadic Cross-Sectional Research." *Personal Relationships* 27, no. 4: 907–938. https://doi.org/10.1111/pere.12343.

Baumgartner, H., B. Weijters, and R. Pieters. 2021. "The Biasing Effect of Common Method Variance: Some Clarifications." *Journal of the Academy of Marketing Science* 49, no. 2: 221–235. https://doi.org/10.1007/s11747-020-00766-8.

Beer, J., P. Eastwick, and J. X. Goh. 2023. "Hits and Misses in the Last Decade of Open Science: Researchers From Different Subfields and Career Stages Offer Personal Reflections and Suggestions." *Social Psychological Bulletin* 18: 1–23. https://doi.org/10.32872/spb.9681.

Bem, D. J. 2011. "Feeling the Future: Experimental Evidence for Anomalous Retroactive Influences on Cognition and Affect." *Journal of Personality and Social Psychology* 100, no. 3: 407–425. https://doi.org/10.1037/a0021524.

Benet-Martinez, V. 2007. "Cross-Cultural Personality Research: Conceptual and Methodological Issues." In *Handbook of Research Methods in Personality Psychology*, edited by R. W. Robins, C. R. Fraley, and R. F. Krueger, 170–189. New York: Guilford Press.

Berscheid, E., and H. T. Reis. 1998. "Attraction and Close Relationships." In *The Handbook of Social Psychology*, edited by D. T. Gilbert, S. T. Fiske, and G. Lindzey, 4th ed., 193–281. New York: Oxford University Press.

Bolger, N., and J. P. Laurenceau. 2013. *Intensive Longitudinal Methods: An Introduction to Diary and Experience Sampling Research*. New York: Guilford Press.

Brady, L. M., S. A. Fryberg, and Y. Shoda. 2018. "Expanding the Interpretive Power of Psychological Science by Attending to Culture." *Proceedings of the National Academy of Sciences* 115, no. 45: 11406–11413. https://doi.org/10.1073/pnas.1803526115.

Bühler, J. L., and U. Orth. 2024. "How Relationship Satisfaction Changes Within and Across Romantic Relationships: Evidence From a Large Longitudinal Study." *Journal of Personality and Social Psychology* 126, no. 5: 930–945. https://doi.org/10.1037/pspp0000492.

Button, K. S., J. P. Ioannidis, C. Mokrysz, et al. 2013. "Power Failure: Why Small Sample Size Undermines the Reliability of Neuroscience." *Nature Reviews Neuroscience* 14, no. 5: 365–376. https://doi.org/10.1038/nrn3475.

Chao, W., and Y. C. Lou. 2018. "Construction of Core Competencies for Family Therapists in Taiwan." *Journal of Family Therapy* 40, no. 2: 265–286. https://doi.org/10.1111/1467-6427.12204.

Cheon, B. K., I. Melani, and Y. Y. Hong. 2020. "How USA-Centric Is Psychology? An Archival Study of Implicit Assumptions of Generalizability of Findings to Human Nature Based on Origins of Study Samples." *Social Psychological and Personality Science* 11, no. 7: 928–937. https://doi.org/10.1177/1948550620927269.

Chester, D. S., and E. N. Lasko. 2021. "Construct Validation of Experimental Manipulations in Social Psychology: Current Practices and Recommendations for the Future." *Perspectives on Psychological Science* 16, no. 2: 377–395. https://doi.org/10.1177/1745691620950684.

Cheung, I., L. Campbell, E. P. LeBel, et al. 2016. "Registered Replication Report: Study 1 From Finkel, Rusbult, Kumashiro, & Hannon (2002)." *Perspectives on Psychological Science* 11, no. 5: 750–764. https://doi.org/10.1177/1745691616664694.

Chopik, W. J., C. R. Chartier, L. Campbell, and M. Brent Donnellan. 2020. "Relationship Science and the Credibility Revolution: An Introduction to the First Part of the Special Issue." *Personal Relationships* 27, no. 1: 132–137. https://doi.org/10.1111/pere.12312.

Cook, T. D., and D. T. Campbell. 1979. *Quasi-experimentation: Design and Analysis Issues for Field Settings*. Chicago: Rand McNally.

Cordova, J. V., C. J. E. Fleming, M. I. Morrill, et al. 2014. "The Marriage Checkup: A Randomized Controlled Trial of Annual Relationship Health Checkups." *Journal of Consulting and Clinical Psychology* 82, no. 4: 592–604. https://doi.org/10.1037/a0037097.

CORE Lab. 2024. "A Novel, Network-Based Approach to Assessing Romantic Relationship Quality." *Perspectives on Psychological Science*: 17456916231215248. https://doi.org/10.1177/17456916231215248.

Corneille, O., and B. Gawronski. 2024. "Self-reports Are Better Measurement Instruments Than Implicit Measures." *Nature Reviews Psychology* 3: 1–12. https://doi.org/10.1038/s44159-024-00376-z.

Dang, J., K. M. King, and M. Inzlicht. 2020. "Why Are Self-Report and Behavioral Measures Weakly Correlated?" *Trends in Cognitive Sciences* 24, no. 4: 267–269. https://doi.org/10.1016/j.tics.2020.01.007.

da Silva Frost, A., and P. W. Eastwick. 2024. "Experimental Tests of the Role of Ideal Partner Preferences in Relationships." https://doi.org/10.31219/osf.io/azjgs.

da Silva Frost, A., and A. Ledgerwood. 2020. "Calibrate Your Confidence in Research Findings: A Tutorial on Improving Research Methods and Practices." *Journal of Pacific Rim Psychology* 14: e14. https://doi.org/10.1017/prp.2020.7.

Delatorre, M. Z., and A. Wagner. 2020. "Marital Quality Assessment: Reviewing the Concept, Instruments, and Methods." *Marriage & Family Review* 56, no. 3: 193–216. https://doi.org/10.1080/01494929.2020.1712300.

Ding, C. G., C. F. Chen, and T. D. Jane. 2023. "Improving the Performance of the Unmeasured Latent Method Construct Technique in Common Method Variance Detection and Correction." *Journal of Organizational Behavior* 44, no. 3: 519–542. https://doi.org/10.1002/job.2673.

Eastwick, P. W., J. Sparks, E. J. Finkel, et al. 2024. "A Worldwide Test of the Predictive Validity of Ideal Partner Preference-Matching." *Journal of Personality and Social Psychology* 128, no. 1: 123–146. https://doi.org/10.1037/pspp0000524.

Elson, M., I. Hussey, T. Alsalti, and R. B. Arslan. 2023. "Psychological Measures Aren't Toothbrushes." *Communications Psychology* 1: 25. https://doi.org/10.1038/s44271-023-00026-9.

Eronen, M. I. 2020. "Causal Discovery and the Problem of Psychological Interventions." *New Ideas in Psychology* 59: 100785. https://doi.org/10.1016/j.newideapsych.2020.100785.

Fabrigar, L. R., D. T. Wegener, and R. E. Petty. 2020. "A Validity-Based Framework for Understanding Replication in Psychology." *Personality and Social Psychology Review* 24, no. 4: 316–344. https://doi.org/10.1177/1088868320931366.

Fincham, F. D., P. C. Garnier, S. Gano-Phillips, and L. N. Osborne. 1995. "Preinteraction Expectations, Marital Satisfaction, and Accessibility: A New Look at Sentiment Override." *Journal of Family Psychology* 9, no. 1: 3–14. https://doi.org/10.1037//0893-3200.9.1.3.

Finkel, E. J., P. W. Eastwick, and H. T. Reis. 2015. "Best Research Practices in Psychology: Illustrating Epistemological and Pragmatic Considerations With the Case of Relationship Science." *Journal of Personality and Social Psychology* 108, no. 2: 275–297. https://doi.org/10.1037/pspi0000007.

Finkel, E. J., J. A. Simpson, and P. W. Eastwick. 2017. "The Psychology of Close Relationships: Fourteen Core Principles." *Annual Review of Psychology* 68, no. 1: 383–411. https://doi.org/10.1146/annurev-psych-010416-044038.

Flake, J. K., and E. I. Fried. 2020. "Measurement Schmeasurement: Questionable Measurement Practices and How to Avoid Them." *Advances in Methods and Practices in Psychological Science* 3, no. 4: 456–465. https://doi.org/10.1177/2515245920952393.

Flake, J. K., J. Pek, and E. Hehman. 2017. "Construct Validation in Social and Personality Research: Current Practice and Recommendations." *Social Psychological and Personality Science* 8, no. 4: 370–378. https://doi.org/10.1177/1948550617693063.

Floyd, K., and N. T. Woo. 2020. "Loneliness and Social Monitoring: A Conceptual Replication of Knowles et al." *Personal Relationships* 27, no. 1: 209–223. https://doi.org/10.1111/pere.12304.

Frey, R., A. Pedroni, R. Mata, J. Rieskamp, and R. Hertwig. 2017. "Risk Preference Shares the Psychometric Structure of Major Psychological Traits." *Science Advances* 3, no. 10: e1701381. https://doi.org/10.1126/sciadv.1701381.

Friedlander, M. L., M. Lee, and V. Escudero. 2019. "What We Do and Do Not Know About the Nature and Analysis of Couple Interaction." *Couple and Family Psychology: Research and Practice* 8, no. 1: 24–44. https://doi.org/10.1037/cfp0000114.

Gawronski, B., M. Morrison, C. E. Phills, and S. Galdi. 2017. "Temporal Stability of Implicit and Explicit Measures: A Longitudinal Analysis." *Personality and Social Psychology Bulletin* 43, no. 3: 300–312. https://doi.org/10.1177/0146167216684131.

Gerlach, T. M., R. C. Arslan, T. Schultze, S. K. Reinhard, and L. Penke. 2019. "Predictive Validity and Adjustment of Ideal Partner Preferences Across the Transition into Romantic Relationships." *Journal of Personality and Social Psychology* 116, no. 2: 313–330. https://doi.org/10.1037/pspp0000170.

Greenwald, A. G., and C. K. Lai. 2020. "Implicit Social Cognition." *Annual Review of Psychology* 71, no. 1: 419–445. https://doi.org/10.1146/annurev-psych-010419-050837.

Gurven, M. D. 2018. "Broadening Horizons: Sample Diversity and Socioecological Theory Are Essential to the Future of Psychological Science." *Proceedings of the National Academy of Sciences* 115, no. 45: 11420–11427. https://doi.org/10.1073/pnas.1720433115.

Hardwicke, T. E., R. T. Thibault, J. E. Kosie, J. D. Wallach, M. C. Kidwell, and J. P. Ioannidis. 2022. "Estimating the Prevalence of Transparency and Reproducibility-Related Research Practices in Psychology (2014–2017)." *Perspectives on Psychological Science* 17, no. 1: 239–251. https://doi.org/10.1177/1745691620979806.

Henrich, J., S. J. Heine, and A. Norenzayan. 2010. "The Weirdest People in the World?" *Behavioral and Brain Sciences* 33, no. 2-3: 61–83. https://doi.org/10.1017/s0140525x0999152x.

Hinnekens, C., L. Stas, F. Gistelinck, and L. L. Verhofstadt. 2020. "'I Think You Understand me.' Studying the Associations Between Actual, Assumed, and Perceived Understanding Within Couples." *European Journal of Social Psychology* 50, no. 1: 46–60. https://doi.org/10.1002/ejsp.2614.

Hussey, I., and S. Hughes. 2020. "Hidden Invalidity Among 15 Commonly Used Measures in Social and Personality Psychology." *Advances in Methods and Practices in Psychological Science* 3, no. 2: 166–184. https://doi.org/10.1177/2515245919882903.

Jakobsen, M., and R. Jensen. 2015. "Common Method Bias in Public Management Studies." *International Public Management Journal* 18, no. 1: 3–30. https://doi.org/10.1080/10967494.2014.997906.

Jeong, S., and Y. Lee. 2019. "Consequences of Not Conducting Measurement Invariance Tests in Cross-Cultural Studies: A Review of Current

Research Practices and Recommendations." *Advances in Developing Human Resources* 21, no. 4: 466–483. https://doi.org/10.1177/15234223198 70726.

Joel, S., P. W. Eastwick, C. J. Allison, et al. 2020. "Machine Learning Uncovers the Most Robust Self-Report Predictors of Relationship Quality Across 43 Longitudinal Couples Studies." *Proceedings of the National Academy of Sciences* 117, no. 32: 19061–19071. https://doi.org/10.1073/pnas.1917036117.

Joel, S., P. W. Eastwick, and E. J. Finkel. 2018. "Open Sharing of Data on Close Relationships and Other Sensitive Social Psychological Topics: Challenges, Tools, and Future Directions." *Advances in Methods and Practices in Psychological Science* 1: 86–94. https://doi.org/10.1177/2515245917744281.

Joel, S., J. K. Sakaluk, J. Kim, D. S. Khera, and S. C. E. Stanton. 2024. "Pseudo Effects: How Method Biases Can Produce Spurious Findings About Close Relationships." https://doi.org/10.31219/osf.io/gsakt.

Kane, H. S., C. McCall, N. L. Collins, and J. Blascovich. 2012. "Mere Presence Is Not Enough: Responsive Support in a Virtual World." *Journal of Experimental Social Psychology* 48, no. 1: 37–44. https://doi.org/10.1016/j.jesp.2011.07.001.

Karney, B. R., M. A. Kreitz, and K. E. Sweeney. 2004. "Obstacles to Ethnic Diversity in Marital Research: On the Failure of Good Intentions." *Journal of Social and Personal Relationships* 21, no. 4: 509–526. https://doi.org/10.1177/0265407504044845.

Kelley, T. L. 1927. *Interpretation of Educational Measurements*. Oxford, England: World Book Co.

Kenny, D. A. 2019. "Enhancing Validity in Psychological Research." *American Psychologist* 74, no. 9: 1018–1028. https://doi.org/10.1037/amp0000531.

Kenny, D. A., D. A. Kashy, and W. L. Cook. 2006. *Dyadic Data Analysis*. New York: Guilford Press.

Kim, J. J., S. Joel, A. M. Gonzales, et al. 2024. "Construct Redundancy in Relationship Science: A Single General Factor Captures Most of the Variance Across 34 Common Relationship Measures." *Manuscript Under Review*.

Klein, R. A., K. A. Ratliff, M. Vianello, et al. 2014. "Investigating Variation in Replicability: A 'Many Labs' Replication Project." *Social Psychology* 45, no. 3: 142–152. https://doi.org/10.1027/1864-9335/a000178.

Landy, J. F., M. (L.) Jia, I. L. Ding, et al. 2020. "Crowdsourcing Hypothesis Tests: Making Transparent How Design Choices Shape Research Results." *Psychological Bulletin* 146, no. 5: 451–479. https://doi.org/10.1037/bul0000220.

Lane, S. P., and E. P. Hennes. 2018. "Power Struggles: Estimating Sample Size for Multilevel Relationships Research." *Journal of Social and Personal Relationships* 35, no. 1: 7–31. https://doi.org/10.1177/0265407517710342.

Ledgerwood, A., K. M. Lawson, M. W. Kraus, et al. 2024. "Disrupting Racism and Global Exclusion in Academic Publishing: Recommendations and Resources for Authors, Reviewers, and Editors." *Collabra: Psychology* 10, no. 1: 121394. https://doi.org/10.1525/collabra.121394.

Lewis, N. A. Jr. 2021. "What Counts as Good Science? How the Battle for Methodological Legitimacy Affects Public Psychology." *American Psychologist* 76, no. 8: 1323–1333. https://doi.org/10.1037/amp0000870.

Lilienfeld, S. O., and A. N. Strother. 2020. "Psychological Measurement and the Replication Crisis: Four Sacred Cows." *Canadian Psychology/Psychologie Canadienne* 61, no. 4: 281–288. https://doi.org/10.1037/cap0000236.

Malhotra, N. K., T. K. Schaller, and A. Patil. 2017. "Common Method Variance in Advertising Research: When to Be Concerned and How to Control for it." *Journal of Advertising* 46, no. 1: 193–212. https://doi.org/10.1080/00913367.2016.1252287.

Maul, A. 2017. "Rethinking Traditional Methods of Survey Validation." *Measurement: Interdisciplinary research and perspectives* 15, no. 2: 51–69. https://doi.org/10.1080/15366367.2017.1348108.

McGorray, E. L., L. F. Emery, A. Garr-Schultz, and E. J. Finkel. 2023. "'Mostly White, Heterosexual Couples': Examining Demographic Diversity and Reporting Practices in Relationship Science Research Samples." *Journal of Personality and Social Psychology* 125, no. 2: 316–344. https://doi.org/10.1037/pspi0000417.

McNulty, J. K., and B. R. Karney. 2002. "Expectancy Confirmation in Appraisals of Marital Interactions." *Personality and Social Psychology Bulletin* 28, no. 6: 764–775. https://doi.org/10.1177/0146167202289006.

Moshontz, H., L. Campbell, C. R. Ebersole, et al. 2018. "The Psychological Science Accelerator: Advancing Psychology Through a Distributed Collaborative Network." *Advances in Methods and Practices in Psychological Science* 1, no. 4: 501–515. https://doi.org/10.1177/2515245918797607.

Nederhof, A. J. 1985. "Methods of Coping With Social Desirability Bias: A Review." *European Journal of Social Psychology* 15, no. 3: 263–280. https://doi.org/10.1002/ejsp.2420150303.

Nisbett, R. E., and T. D. Wilson. 1977. "Telling More Than We Can Know: Verbal Reports on Mental Processes." *Psychological Review* 84, no. 3: 231–259. https://doi.org/10.1037/0033-295x.84.3.231.

Nosek, B. A., T. E. Hardwicke, H. Moshontz, et al. 2022. "Replicability, Robustness, and Reproducibility in Psychological Science." *Annual Review of Psychology* 73, no. 1: 719–748. https://doi.org/10.1146/annurev-psych-020821-114157.

Ogan, M. A., and J. B. Kanter. 2024. "Uncovering Measurement Bias: Differential Item Functioning by Income in the Couples Satisfaction Index." *Journal of Family Psychology* 38, no. 6: 966–976. https://doi.org/10.1037/fam0001247.

Olson, J. G., S. I. Rick, D. A. Small, and E. J. Finkel. 2023. "Common Cents: Bank Account Structure and Couples' Relationship Dynamics." *Journal of Consumer Research* 50, no. 4: 704–721. https://doi.org/10.1093/jcr/ucad020.

Orth, U. 2013. "How Large Are Actor and Partner Effects of Personality on Relationship Satisfaction? The Importance of Controlling for Shared Method Variance." *Personality and Social Psychology Bulletin* 39, no. 10: 1359–1372. https://doi.org/10.1177/0146167213492429.

Podsakoff, P. M., S. B. MacKenzie, J. Y. Lee, and N. P. Podsakoff. 2003. "Common Method Biases in Behavioral Research: A Critical Review of the Literature and Recommended Remedies." *Journal of Applied Psychology* 88, no. 5: 879–903. https://doi.org/10.1037/0021-9010.88.5.879.

Podsakoff, P. M., S. B. MacKenzie, and N. P. Podsakoff. 2012. "Sources of Method Bias in Social Science Research and Recommendations on How to Control it." *Annual Review of Psychology* 63, no. 1: 539–569. https://doi.org/10.1146/annurev-psych-120710-100452.

Podsakoff, P. M., N. P. Podsakoff, L. J. Williams, C. Huang, and J. Yang. 2024. "Common Method Bias: It's Bad, It's Complex, It's Widespread, and It's Not Easy to Fix." *Annual Review of Organizational Psychology and Organizational Behavior* 11, no. 1: 17–61. https://doi.org/10.1146/annurev-orgpsych-110721-040030.

Pollitt, A. M., K. L. Blair, and P. J. Lannutti. 2023. "A Review of Two Decades of LGBTQ-Inclusive Research in JSPR and PR." *Personal Relationships* 30, no. 1: 144–173. https://doi.org/10.1111/pere.12432.

Rad, M. S., A. J. Martingano, and J. Ginges. 2018. "Toward a Psychology of Homo Sapiens: Making Psychological Science More Representative of the Human Population." *Proceedings of the National Academy of Sciences* 115, no. 45: 11401–11405. https://doi.org/10.1073/pnas.1721165115.

Randall, A. K., and M. A. Curran. 2023. "Editorial Synthesis for 2 in 2023 a Collaboration Between IARR's Two Journals: Recognizing the Need for Greater Inclusivity in Relationship Science." *Personal Relationships* 30, no. 1: 4–18. https://doi.org/10.1111/pere.12477.

Reis, H. T., S. M. Smith, C. L. Carmichael, et al. 2010. "Are You Happy for Me? How Sharing Positive Events With Others Provides Personal and Interpersonal Benefits." *Journal of Personality and Social Psychology* 99, no. 2: 311–329. https://doi.org/10.1037/a0018344.

Remedios, J. D. 2022. "Psychology Must Grapple With Whiteness." *Nature Reviews Psychology* 1, no. 3: 125–126. https://doi.org/10.1038/s44159-022-00024-4.

Rhemtulla, M., D. Borsboom, and R. van Bork. 2017. "How to Measure Nothing." *Measurement: Interdisciplinary Research and Perspectives* 15, no. 2: 95–97. https://doi.org/10.1080/15366367.2017.1369785.

Rochat, P. 2010. "What Is Really Wrong With A Priori Claims of Universality? Sampling, Validity, Process Level, and the Irresistible Drive to Reduce." *Behavioral and Brain Sciences* 33, no. 2-3: 107–108. https://doi.org/10.1017/s0140525x10000233.

Rohrer, J. M. 2018. "Thinking Clearly About Correlations and Causation: Graphical Causal Models for Observational Data." *Advances in Methods and Practices in Psychological Science* 1, no. 1: 27–42. https://doi.org/10.1177/2515245917745629.

Rohrer, J. M., and K. Murayama. 2023. "These Are Not the Effects You Are Looking for: Causality and the Within-/Between-Persons Distinction in Longitudinal Data Analysis." *Advances in Methods and Practices in Psychological Science* 6, no. 1: 25152459221140842. https://doi.org/10.1177/25152459221140842.

Sabik, N. J., J. L. Matsick, K. McCormick-Huhn, and E. R. Cole. 2021. "Bringing an Intersectional Lens to 'Open' Science: An Analysis of Representation in the Reproducibility Project." *Psychology of Women Quarterly* 45, no. 4: 475–492. https://doi.org/10.1177/03616843211035678.

Sakaluk, J. K., & Camanto, O. J. (2024). dySEM: A R Package for Structural Equation Models of Latent Variables With Dyadic Data.

Sakaluk, J. K., and A. Daniel. 2022. "How EIRD is Sex Research?: A Commentary and Reanalysis of Klein et al.(2021)." *Journal of Sex Research* 59, no. 7: 818–825. https://doi.org/10.1080/00224499.2022.2087854.

Sakaluk, J. K., A. N. Fisher, and R. E. Kilshaw. 2021. "Dyadic Measurement Invariance and its Importance for Replicability in Romantic Relationship Science." *Personal Relationships* 28, no. 1: 190–226. https://doi.org/10.1111/pere.12341.

Sakaluk, J. K., S. Joel, C. Quinn-Nilas, et al. 2024, January 3. "A Renewal of Dyadic Structural Equation Modelling With Latent Variables: Clarifications, Methodological Advantages and Trends, Reflections, and New Directions." https://doi.org/10.31234/osf.io/yp485.

Schimmack, U. 2019. "The Implicit Association Test: A Method in Search of a Construct." *Perspectives on Psychological Science* 16, no. 2: 396–414. https://doi.org/10.1177/1745691619863798.

Schimmack, U. 2021. "The Validation Crisis in Psychology." *Meta-Psychology* 5. https://doi.org/10.15626/mp.2019.1645.

Sharkey, J. A., J. S. Feather, and S. Goedeke. 2022. "The Current State of Relationship Science: A Cross-Disciplines Review of Key Themes, Theories, Researchers and Journals." *Journal of Social and Personal Relationships* 39, no. 4: 864–885. https://doi.org/10.1177/02654075211047638.

Simmons, J. P., L. D. Nelson, and U. Simonsohn. 2011. "False-positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant." *Psychological Science* 22, no. 11: 1359–1366. https://doi.org/10.1177/0956797611417632.

Simons, D. J., Y. Shoda, and D. S. Lindsay. 2017. "Constraints on Generality (COG): A Proposed Addition to All Empirical Papers." *Perspectives on Psychological Science* 12, no. 6: 1123–1128. https://doi.org/10.1177/1745691617708630.

Syed, M. 2021. "WEIRD Times: Three Reasons to Stop Using a Silly Acronym." https://getsyeducated.blogspot.com/2021/06/weird-times three-reasons-to-stop-using.html.

Syed, M., and U. K. Kathawalla. 2022. "Cultural Psychology, Diversity, and Representation in Open Science." *Cultural methods in psychology: Describing and transforming cultures*: 427–454. https://doi.org/10.1093/oso/9780190095949.003.0015.

Thalmayer, A. G., C. Toscanelli, and J. J. Arnett. 2021. "The Neglected 95% Revisited: Is American Psychology Becoming Less American?" *American Psychologist* 76, no. 1: 116–129. https://doi.org/10.1037/amp0000622.

Tseng, C., A. K. Wittenborn, A. J. Blow, W. Chao, and T. Liu. 2020. "The Development of Marriage and Family Therapy in East Asia (China, Taiwan, Japan, South Korea and Hong Kong): Past, Present and Future." *Journal of Family Therapy* 42, no. 4: 477–498. https://doi.org/10.1111/1467-6427.12285.

Vazire, S., S. R. Schiavone, and J. G. Bottesini. 2022. "Credibility beyond Replicability: Improving the Four Validities in Psychological Science." *Current Directions in Psychological Science* 31, no. 2: 162–168. https://doi.org/10.1177/09637214211067779.

Waldinger, R. J., and M. S. Schulz. 2006. "Linking Hearts and Minds in Couple Interactions: Intentions, Attributions, and Overriding Sentiments." *Journal of Family Psychology* 20, no. 3: 494–504. https://doi.org/10.1037/0893-3200.20.3.494.

Wang, Y. A., and P. W. Eastwick. 2020. "Solutions to the Problems of Incremental Validity Testing in Relationship Science." *Personal Relationships* 27, no. 1: 156–175. https://doi.org/10.1111/pere.12309.

Weiss, R. L. 1980. "Strategic Behavioral Marital Therapy: Toward a Model for Assessment and Intervention." In *Advances in Family Intervention, Assessment and Theory*, edited by J. P. Vincent, Vol. 1, 229–271. Greenwich: JAI Press.

Westfall, J., and T. Yarkoni. 2016. "Statistically Controlling for Confounding Constructs Is Harder Than You Think." *PLoS One* 11, no. 3: e0152719. https://doi.org/10.1371/journal.pone.0152719.

Wetzel, E., and B. W. Roberts. 2020. "Commentary on Hussey and Hughes (2020): Hidden Invalidity Among 15 Commonly Used Measures in Social and Personality Psychology." *Advances in Methods and Practices in Psychological Science* 3, no. 4: 505–508. https://doi.org/10.1177/2515245920957618.

Williams, L. J., and A. K. McGonagle. 2016. "Four Research Designs and a Comprehensive Analysis Strategy for Investigating Common Method Variance With Self-Report Measures Using Latent Variables." *Journal of Business and Psychology* 31, no. 3: 339–359. https://doi.org/10.1007/s10869-015-9422-9.

Williams, L. J., and E. H. O'Boyle. 2015. "Ideal, Nonideal, and No-Marker Variables: The Confirmatory Factor Analysis (CFA) Marker Technique Works When it Matters." *Journal of Applied Psychology* 100, no. 5: 1579–1602. https://doi.org/10.1037/a0038855.

Williamson, H. C. 2024a, July. "Models of Couple Communication Center White American Modes of Interaction." In *Methodological Challenges in Relationship Science [Symposium]*, edited by J. J. Kim (chair), Boston, U.S. A: International Association for Relationship Research.

Williamson, H. C. 2024b. "Innovative Strategies for Increasing Sample Size and Diversity in Family Science Research." *Journal of Marriage and Family* 86, no. 5: 1393–1409. https://doi.org/10.1111/jomf.12961.

Williamson, H. C., J. X. Bornstein, V. Cantu, O. Ciftci, K. A. Farnish, and M. T. Schouweiler. 2022. "How Diverse Are the Samples Used to Study Intimate Relationships? A Systematic Review." *Journal of Social and Personal Relationships* 39, no. 4: 1087–1109. https://doi.org/10.1177/02654075211053849.

Wysocki, A. C., K. M. Lawson, and M. Rhemtulla. 2022. "Statistical Control Requires Causal Justification." *Advances in Methods and Practices in Psychological Science* 5, no. 2: 25152459221095823. https://doi.org/10.1177/25152459221095823.

Yarkoni, T. 2022. "The Generalizability Crisis." *Behavioral and Brain Sciences* 45: e1. https://doi.org/10.1017/s0140525x20001685.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.