Article

Organizational Research Methods
2026, Vol. 29(2) 177–201

DOI: 10.1177/10944281241300952
journals.sagepub.com/home/orm
Sage

Manipulation in Organizational Research: On Executing and Interpreting Designs from Treatments to Primes

Kira F. Schabram[1], Christopher G. Myers[2], and Ashley E. Hardin[3]

Abstract

While other applied sciences systematically distinguish between manipulation designs, organizational research does not. Herein, we disentangle distinct applications that differ in how the manipulation is deployed, analyzed, and interpreted in support of hypotheses. First, we define two archetypes: *treatments*, experimental designs that expose participants to different levels/types of a manipulation of theoretical interest, and *primes*, manipulations that are not of theoretical interest but generate variance in a state that is. We position these and creative derivations (e.g., interventions and invariant prompts) as specialized tools in our methodological kit. Second, we review 450 manipulations published in leading organizational journals to identify each type's prevalence and application in our field. From this we derive our guiding thesis that while treatments offer unique advantages (foremost establishing causality), they are not always possible, nor the best fit for a research question; in these cases, a non-causal but accurate test of theory, such as a prime design, may prove superior to a causal but inaccurate test. We conclude by outlining best practices for selection, execution, and evaluation by researchers, reviewers, and readers.

Keywords

experimental design, quasi-experimental design, manipulation, treatment, prime, invariant prompt, intervention

Manipulations represent a dominant paradigm for studying behavior in organizations. In their classic application—commonly termed a "treatment" (Antonakis et al., 2010; Campbell & Stanley, 1967)—researchers randomly expose participants to different types or levels of an objectively demonstrable

[1]Foster School of Business, University of Washington, Seattle, WA, USA
[2]Carey Business School, Johns Hopkins University, Baltimore, MD, USA
[3]Olin Business School, Washington University in St. Louis, St Louis, MO, USA

Corresponding author:
Kira Schabram, Foster School of Business, University of Washington, Box 353226, 4295 E. Stevens Way East, Seattle, WA 98105, USA.
Email: schabram@uw.edu

stimulus in order to study its effect (Austin et al., 2002). This method and terminology were imported from long-standing traditions in the natural sciences (Steffens, 2007), medicine (Gaw, 2009), and economics (Brue & Grant, 2013) and now represents an entrenched research paradigm in the fields of organizational behavior, industrial–organizational psychology, entrepreneurship, and management. The value of treatments lies in their functionality as tools for isolating a phenomenon of interest, carefully controlling its expression, and providing a causal test of its effects on key outcomes (Shadish et al., 2002).

Though treatments are often what organizational scholars envision when considering experimental or quasi-experimental research, manipulations lend themselves to other functions. In psychological and organizational research, for instance, scholars frequently employ manipulations with the intent of priming variance in cognitions, emotions, or other psychological states of interest, particularly those that are rare, deviant, or difficult to capture (Forster & Liberman, 2013). This distinction between more treatment-like and more prime-like manipulations is critical. Though similar manipulations could be used in each paradigm, they differ in their underlying purpose—and importantly their statistical analysis and interpretation—when deployed in pursuit of particular research questions. Unfortunately, because there is little guidance on this distinction in our field to date, there has been misapplication, leading to systematic errors, which ultimately undermine the reliability and replicability of our science.

To illuminate this point: imagine a team of researchers considering the question of whether personnel decisions made by an algorithm (vs. by a human) would result in lower perceptions of fairness (an example adapted from Newman et al., 2020). This hypothesis could be tested by randomly assigning participants to one of two conditions in which either an algorithm or a human makes a layoff decision, followed by a measure of perceived fairness. Analysis would involve a statistical comparison of fairness means across the two conditions. Now consider a different research team interested in examining the link between perceptions of fairness and organizational commitment. While they could pursue their question via a variety of methods (e.g., a field survey of individuals' perceived fairness and commitment in their organizations), for both pragmatic and theoretical reasons (explored further below), the most suitable approach might involve priming variation in the perception of fairness among participants via a manipulated stimulus and then measuring commitment. Having read Newman et al. (2020), these researchers could generate variance in fairness perceptions by randomly assigning participants to the same two conditions (i.e., algorithmic vs. human decision). Importantly, to have fidelity to the intent of their study, their primary analysis would need to test the association between the measured state (perceptions of fairness) and organizational commitment, not the mean comparison approach of conditions employed by the former set of scholars.

On the surface, these two data collection efforts would differ only by their inclusion of an assessment of organizational commitment. However, the purpose of the manipulation and how it should be treated in the study's analysis are entirely different (for respective best practices see Lonati et al., 2018; Sajons, 2020). In the first example, using the manipulation as a treatment, the researchers are conceptually interested in the effects of the manipulated stimulus itself (i.e., the layoff decision-maker being an algorithm or a human). In the second, using the manipulation as a prime, the researchers are not interested in the decision-maker at all. They care about this stimulus (algorithmic vs. human decision) only insofar as it induces variance in their state of interest (i.e., perceptions of fairness) among participants, which can then be used to understand the impact of this state on outcomes of interest.

These two types of manipulations reflect distinct approaches to conducting research and each serves valuable functions as part of a methodological "toolkit" for organizational researchers. Both types have the potential to probe theoretical propositions of relevance and importance to organizational research, but only to the extent that they are designed, analyzed, and interpreted in ways that align with the nature of the question at hand. Unfortunately, as a field, due to an early and persistent emphasis on the benefits of treatment designs, we tend to approach all manipulations as if they are treatments (or should be). Without guidance for the relevant uses of primes and treatments, the risk of misapplication and systematic error, which undermine the reliability and replicability of our science, is high. Such concerns have been addressed in other applied sciences via a tradition of explicitly distinguishing between different types of manipulations (see Oehlert, 2000; O'Keefe, 2003; Welsh et al., 2013). The aim of this paper is to present a framework of manipulation types that does the same for organizational scholarship.

To do so, we begin by defining the two main types of manipulation: treatments and primes, as well as additional approaches derived from each of these broader archetypes (i.e., interventions and invariant prompts). We outline the types of research questions to which each is best suited, describe the appropriate statistical approach to their interpretation, and identify exemplars of each in extant literature. In this section, we also carefully discuss trade-offs of each, including considerations of their respective risks and shortcomings. Next, we present a methodological content analysis of all studies which employed a manipulation published in five leading management and organizational science journals over the course of a year ($n = 450$ manipulations employed in 326 studies across 98 papers). This allows us to document variation in how manipulations are currently designed, employed, and analyzed. We demonstrate that though scholars use all manipulation types, there are no codified practices around how and when to select or execute each, which can lead to misapplication. We highlight a particular concern that the majority of primes are analyzed and interpreted as if they are treatments, leading to a mismatch between theory and method, and potentially inaccurate conclusions that undermine scientific progress. In doing so, we unpack two paradoxical pressures we see as contributing to the covert and/or misapplication of primes in our field: a lack of codified, accepted community practices for the implementation of manipulations, and normative pressures to adhere to one standard of empirical quality in manipulation designs that does not suit the wide variety of research questions commonly explored in applied psychological and organizational scholarship. To address these issues, we conclude our manuscript with best practices for researchers utilizing manipulations, as well as for reviewers and editors evaluating their adequacy and appropriateness.

Echoing other recent guidelines (see Eden, 2017), we stress that ours is not a comprehensive review of methods. Numerous other publications offer explicit guidelines on experimental methods (e.g., Lonati et al., 2018), quasi-experimental methods (e.g., Grant & Wall, 2009), or both (e.g., Shadish et al., 2002). Throughout, we direct readers to these invaluable resources for best and cutting-edge empirical practices. Our core aim is to conceptually and philosophically disentangle the ways that manipulations are already used (and could be used) in our field; not to advocate for one over another, but rather to facilitate a transparent discussion of their application and trade-offs. If our effort is successful, readers will have greater confidence in designing manipulation studies, justifying and transparently describing their choices, as well as in evaluating others' usage of manipulation (cf. Eden, 2017). Our hope is that such guidance for a priori planning, contextualizing designs, and transparency in communication, can be a useful step toward codifying practices in the field.

Distinguishing Types of Manipulations

Around the turn of the millennium, some scholars expressed genuine worry that manipulation designs were being relegated to a "lower status in the pecking order of organizational research methods" (Dobbins et al., 1988; Highhouse, 2009, p. 2). In one review, laboratory experiments had dropped to less than 5% of publications in three top management journals (Scandura & Williams, 2000). Today, however, the pendulum has decisively swung back (e.g., Minson et al., 2023). Manipulations (and experimental designs more broadly) are prevalent not only across micro-organizational research, but becoming more common in other management domains such as entrepreneurship (Hsu et al., 2024), where they are applied to a broad range of research questions and literatures. We posit that there are two distinct types of manipulations (treatments and primes), which anchor the variation observed across manipulations in our field. Below, we define each type, highlight their relative advantages and disadvantages, and reference exemplary applications. In outlining this typology, we stress that neither treatments nor primes are inherently superior or universally preferable in organizational research: the principal criterion for choosing one over the other is alignment between the manipulation and the theoretical focus of the study.

Treatments

Treatments are manipulations in which researchers intentionally expose participants to different types or levels of an objectively demonstrable stimuli (i.e., different conditions) that are of central theoretical interest (see top model in Figure 1). Treatments are principally concerned with the direct effect of this differential exposure on an outcome, particularly examining mean differences in that outcome by condition. This logic of manipulation underpins classic experimental methods in other applied fields, such as the randomized-control trial (RCT) of novel medical treatments, where exposure to a particular stimulus (i.e., receiving doses of a new medication, vs. not receiving any medication or receiving a placebo) is the phenomenon of interest, and key outcomes are compared across the individuals randomly assigned to each condition. As noted in our introduction, the treatment design reflects an experimental tradition imported from pioneering work in the natural and hard sciences (Brue & Grant, 2013; Gaw, 2009; Steffens, 2007).

Applications of treatments to explore the impact of workplace stimuli abound. Recent published examples of treatments in organizational behavior research include examinations of the effect of exposure to a happy vs. angry interaction partner on information search behaviors (Rees et al., 2020), the impact of displaying family photos (vs. photos of nature, strangers, or the self) on unethical behavior (Hardin et al., 2020), or the effect of communication medium (telephone vs. face-to-face) on interpretations of surface acting (Brodsky, 2020). Entrepreneurship scholars have deployed treatments to demonstrate that when entrepreneurs practice—as opposed to listen to a TED talk about—loving-kindness meditation they make more sustainable decisions (Engel et al.,

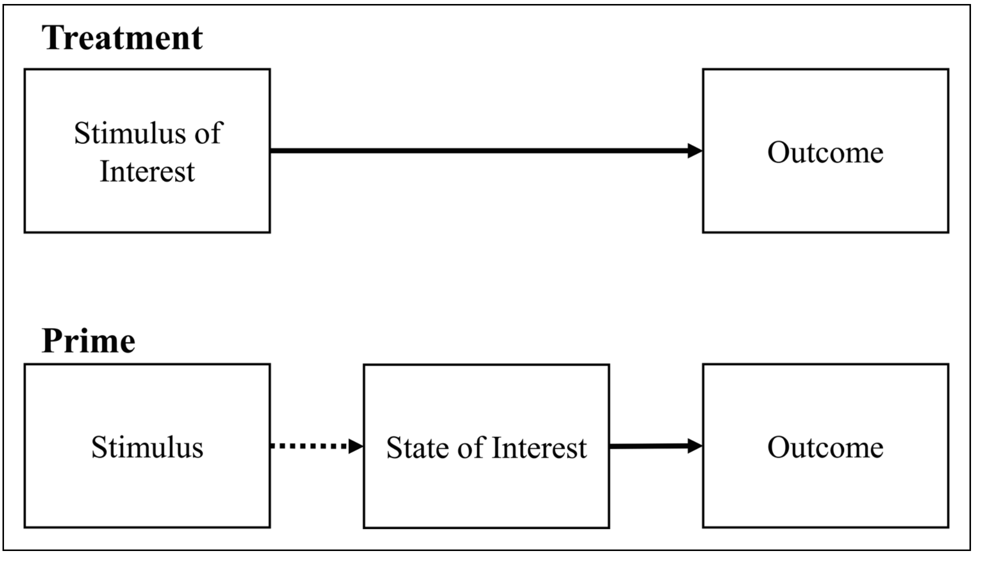

Figure 1. Conceptual models of manipulation types.

2020), and that hypothetical new venture investors differ in their likelihood to fund female entrepreneurs across different global crises (financial, COVID-19, or no crisis; Yu et al., 2024). Furthermore, treatments also form the basis of recent large-scale field experiments that varied founder information (e.g., gender, experience of failure vs. success) to probe job applicant (Abraham & Burbano, 2022) and firm hiring (Botelho & Chang, 2023) decisions. In all these examples, the research question and methodological design lend themselves to a comparison of means across the different levels of the independent variable (e.g., happy vs. angry partner, photo of a family member vs. landscape, financial crisis vs. pandemic, company founder gender) through comparative methods (e.g., t-tests, ANOVA) or regression methods using the condition as a categorical predictor variable.[1]

The primary advantage of treatments lies in their singular ability to support causal claims, the key goal of traditional experimental design (Shadish et al., 2002). Put more colorfully by Pinker (2012, p. 123), such randomized, controlled experiments are the best means to untangle "the social science rat's nest of confounded variables." For this reason, a treatment is the first (and often only) type of manipulation encountered in methodological training and the gold standard to which most scholars refer when they reference experimental convention (Ellsworth & Gonzalez, 2007; Festinger, 1953).

Treatments, however, like all research methods, have limitations. Treatments are "not a panacea" (Lonati et al., 2018, p. 20) for bad methodological practice including internal and statistical validity threats (e.g., inappropriate comparisons, demand effects, incorrect inference) as well as threats to external and ecological validity (see Lonati et al., 2018 for further discussion of these threats and their amelioration). Treatments are also less suitable for another crucial scientific goal: providing meaningful evidence of the explanation for detected variance (Bacharach, 1989; Kerlinger, 1973). This shortcoming is often addressed by incorporating treatments into multi-study examinations of a phenomenon, paired with other designs (e.g., field, archival, or qualitative studies) including other manipulation designs that probe underlying mechanisms.

To sum, when scholars adhere to best practices that leverage their elegant simplicity and ability to establish causality, treatments represent the classic ideal of manipulation (see Table 1). Such treatments are objective in nature, arguably behavioral whenever possible (Banks et al., 2023), can be judged on their face to have occurred (e.g., communication happened by telephone or face-to-face; Brodsky, 2020), and closely match the condition stimuli to the concepts they represent (Bacharach, 1989). Treatments are most powerfully suited to hypotheses regarding a comparison of phenomena across different factors or efforts to confirm causal links between an independent and dependent variable.

Primes

In prime designs, the manipulation functions as a tool to generate variance in a particular state of interest (attitudes, emotions, behaviors, etc.) in order to understand the impact of that state on an outcome (Hauser et al., 2018; Sajons, 2020), as shown in the bottom model of Figure 1. Returning to our opening example, for the second research team, the value of the manipulation (exposure to human vs. algorithmic decisions) lies in its ability to generate variance in participants' subjective perceptions of fairness, which creates a fuller representation of the range of the independent variable. In this research design, the condition (human vs. algorithm) is useful only insofar as it creates the necessary variation in the state of interest (perceptions of fairness) to assess its relationship with the outcome of interest (workplace commitment). The researchers' theory and hypothesis have nothing to do with algorithmic decision-making: many other activities that generated variance in perceptions of fairness (e.g., compensating some participants equitably and others inequitably for their performance in the experiment; see Sajons, 2020) would be equally valid as a prime in the study design. In the analysis, the independent variable would be the measured state, which would be analyzed using methods that identify association (e.g., correlation, regression, etc.) between the state and the outcome of interest.[2]

Table 1. Typology of Manipulation Designs.

	Treatments (and Related Derivations)	Primes (and Related Derivations)
Purpose	The goal of a *treatment* design is isolating a phenomenon of interest, carefully controlling its expression, and providing a causal test of its effects on key outcomes.	The goal of a *prime* design is generating variance in cognitions, emotions, or other psychological states of interest, particularly those that are rare, deviant, or difficult to capture.
Design	Expose participants to manipulated types or levels of an objectively demonstrable stimulus in order to study the effect of the stimulus conditions, as the independent variable, on an outcome of interest (which may then itself be associated with other distal outcomes, e.g., in an *intervention*).	Expose participants to different types or levels of a stimulus (or to a single stimulus, in the case of an *invariant prompt*) to induce variation in participants' emotion, cognition, behavior, or other state, which is measured and used as the independent variable to test its association with an outcome of interest.
Analysis and Interpretation	Comparison of means across the different types or levels of the manipulated stimulus through comparative methods (e.g., t-tests, ANOVA) or regression methods using the condition as a categorical predictor variable. Significant differences across stimulus conditions supports claims (including, given effective randomization, causal claims) for the effects of the stimulus on the outcome.	Analyzing the relationship between the measured state variable and the outcome variable using methods that identify association (e.g., correlation, regression, etc.). Significant association between the state variable and the outcome variable (given effective steps taken to address endogeneity and control for confounding factors) supports claims for the relationship between the state and the outcome.
Risks and Tradeoffs	Particularly vulnerable to issues concerning internal and external validity of the manipulation. Less suitable to examining theoretical mechanisms or processes.	Particularly vulnerable to issues associated with subjective measures, self-reports, and demand effects. Subject to endogeneity concerns and unsuited to examining causal effects.

Recent publications demonstrate how prime manipulations can inform theory when carefully designed and analyzed. Chua and Jin (2020) randomly assigned participants to intercultural (versus same-culture) dyads to prime opportunities for conflict; displayed conflict, their independent variable, was measured by coding video recordings of the interactions in the context of creative collaboration. Klein and colleagues (2020) assigned undergraduates to different referent audiences (family member, close friend, or graduate student) to prime student perceptions of their assigned audience's relative status. Analysis revealed these relative status perceptions, as reported by the subjects, were positively related to goal commitment and downstream performance. In entrepreneurship, Frederiks and colleagues (2019) sought to examine the degree to which potential entrepreneurs could recognize business opportunities associated with a technical innovation based on different future-oriented cognitive processing. To do so, participants completed different behavioral tasks (e.g., anagrams, subjective probability tasks), which created variance in the usage of different future-oriented cognitive processes (Janiszewski & Wyer, 2014). In all these examples, what is of theoretical interest is the primed state (e.g., perceived relative status, future-oriented cognitive processing), *not* the mechanics of the manipulation (e.g., evaluation by friends/family/graduate students, solving anagrams). In other words, what is of conceptual and empirical consequence is how the induced state relates to the outcome variable, not the impact of the manipulation activity itself.

The most obvious limitation of primes is that their analysis relies on a measured state (self-reported by the participant or coded by observers), rather than the distinction of which manipulated stimulus participants were exposed to. Notably, when one models a measured state as the independent variable, the assumption of random assignment to stimulus conditions is absent. Though the measured state variable would obviously be influenced by the stimulus, it may also include variance due to external forces or other factors endogenous to the model (such as participants' attitudes, norms, or past experiences; Sajons, 2020). In short, primes, at best, fulfill two of the three criteria for causal effects: the state generally does precede the outcome and the two can be reliably correlated, but other causes for their relationship cannot be automatically ruled out (Antonakis et al., 2010). Researchers must therefore take care to address issues of endogeneity in analyzing and interpreting the results of these designs.

Stated more directly, insufficiently addressed endogeneity in a prime design could potentially constitute a "fatal flaw" in a study (just as it would in all other non-treatment designs) and a threat to our field's mandate to accurately inform policy (Antonakis, 2017). On the continuum from causal to correlational designs, primes fall alongside other quasi-experimental approaches (Campbell & Stanley, 1967; Grant & Wall, 2009) that seek to change a key independent variable of interest, but that relax the strict criteria of a randomized-control trial. Unlike treatments, researchers must therefore take extreme care to avoid making any causal inferences in their analysis of primes (Aguinis & Vandenberg, 2014). Scholars should address issues of endogeneity, common method variance, and alternative explanations in their analytical approach, for example via the inclusion of control variables, instrumental variable techniques, or other modeling approaches that address endogeneity (Antonakis et al., 2010; Sajons, 2020). Indeed, as highlighted in recent research methods literature, such issues can be productively addressed through (1) basing the indirect effect of the prime on the outcome on estimates from an instrumental variable regression, given endogeneity issues stemming from measurement errors or omitted variables (Bastardoz et al., 2023; Sajons, 2020) and (2) estimating the effect of the measured state on the outcome via instrumental variable regression (Maydeu-Olivares et al., 2020).[3]

In addition, whereas treatments aim to be as objective as possible (i.e., capturing the concept of interest in a clear, unequivocal manipulation), most primes intentionally examine the subjective experience of participants. The primed state of interest can certainly be a behavior (like that of displayed conflict in our earlier example; Chua & Jin, 2020) but is more often a cognition, attitude, or emotion. Accordingly, primes are necessarily subject to issues commonly associated with self-report measures (Banks et al., 2023) as well as to demand effects, defined as changes in participant conduct due to cues about what is appropriate in the setting (Zizzo, 2010, p. 75). While this risk cannot be eliminated, it can be mitigated via established practices such as accounting for social desirability, avoiding unfair comparison in condition design, incorporating experimenter blinding protocols or using blind coders to measure the state (e.g., coding videos of participants post-manipulation), and including the most unobtrusive manipulations possible, particularly those occurring outside of artificial settings (Eden, 2017; for other recommendations, see also Banks et al., 2023; Khademi et al., 2021; Lonati et al., 2018; Zizzo, 2010).

In exchange for such tradeoffs, primes provide researchers with several key theoretical and pragmatic advantages. First, our field aspires to use the most accurate measure available to represent the focal construct (Grant & Wall, 2009; Locke, 1986). Conceptual fit constitutes not only the substantive goal of scientific endeavors but also the means to that end, superordinate to other considerations like statistical analysis (Fiedler et al., 2021). If a researcher's conceptual model focuses on the relationship between a state and an outcome, the most appropriate variable to analyze would be the one that is the closest fit to the state construct (the measured state), rather than the activity that influenced the state (the manipulated stimulus). In other words, reading a scenario about having power is not the same as feeling powerful, just as recalling a time one was hungry would not be the same as feeling hunger. A prime approach avoids creating this false equivalence between the stimulus and the state of interest.

Second, primes are well-suited for harnessing and testing the effects of more nuanced variance in an underlying variable of interest, a function central to the science of organizations (Kerlinger, 1973). Analyzing a measure of the state allows researchers to assess the full range of the state on an outcome, rather than representing this range as merely a few data points (the discrete conditions). Primes are not unique in this way, as a fuller range of a state of interest could also be captured by other associational research designs, such as field surveys. However, inducing the state via a prime design can be quite useful when states that are of theoretical interest may be hard to capture from a sampling or timing perspective (i.e., when the full range of the state might not naturally occur in a particular sample or at a particular point in time, or might occur among unique or vulnerable populations; see Restubog et al., 2023). We see this reflected in recent applications in our field inducing particular moods (Umphress et al., 2020), attitudes (Foulk et al., 2020; Qin et al., 2020), and cognitions (Frederiks et al., 2019). Schabram and Heng (2022, p. 461) explicitly acknowledge that a prime design allowed them to test the effects of compassion on burnout among members of a convenient and theoretically important population (business school students) who “tend to offer less self- or other-compassion.”

To sum, primes are ideally suited to maximize conceptual fit and capture nuanced variance in an underlying state of interest (see Table 1). They lend themselves to examining unusual or rare phenomena—what Cortina and colleagues (2017, p. 274) refer to as “study[ing] the exceptional and not just the average.” In return for such advantages, however, this design requires scholars to avoid causal claims, anticipate issues of endogeneity, and take care in measuring and interpreting the primed states in a way that contends with the possibility of social desirability, the influence of obtrusive manipulations, and demand effects (akin to issues previously raised about manipulation checks; see for instance Ejelöv & Luke, 2020).

Other Manipulation Types: Derivations of Treatments and Primes

Thus far, we have focused on treatments and primes as two archetypes of manipulations used in organizational research. However, these two approaches serve only as broad anchors of the myriad of ways researchers can employ manipulations in their empirical strategies. Across the field, we observe two particular derivations of treatments and primes (i.e., sub-types within the broader typology of manipulations) that warrant specific mention.

Invariant Prompts. In certain studies, scholars might employ a variation of a prime in which all participants are exposed to a *single* manipulation. We term such one-condition manipulations an invariant prompt design. As in a more typical prime design, the intention is to create variance in a state of interest. However, those employing invariant prompts have reason to believe that a single prompt will create sufficient variance in the intended state due to a variety of individual or contextual factors that might alter individuals’ reaction or response to the manipulation. For instance, in our opening example, researchers might be confident that simple exposure to algorithmic decisions (i.e., all participants read about the algorithm making personnel decisions and none read about a human) might suffice to prompt varying levels of participants’ perceptions of fairness.

Published examples of invariant prompts often involve critical incident recall designs. Priesemuth and Bigelow (2020) asked participants to recall abusive supervision to create variance in their independent variable—enacted abusive supervision—and examine the downstream consequences on social worth and job performance. Lyons et al. (2020) asked participants to recall an episode in the last six months in which a coworker has come out to them to measure heterosexual identity threat and its impact on participants’ response. At a more macro level, DeCelles et al. (2020) recruited social activists, asked them to think about the social issues most important to them in order to prompt (and measure) anger and its impact on collective action intentions. Invariant prompts are not constrained to hypothetical activities or recall tasks however, and can make use of clever, single-condition behavioral tasks. For instance, Yeomans and colleagues (2020) had government executives discuss just one hot-button issue with a disagreeing partner to prime "conversational receptiveness" and examine its impact on interpersonal engagement.

Invariant prompts present many of the same tradeoffs previously outlined for primes, most notably the inability to claim causality. Each of the above examples reminds us that invariant prompts, like primes more generally, are particularly useful for studying behaviors that may occur infrequently (e.g., a coworker coming out, hot-button topics), that participants may be reluctant to self-report in surveys (e.g., abusive supervision, guilt, anger, heterosexual identity threat), or that may be challenging to ethically manipulate in a lab setting (e.g., abusive supervision). In this way, invariant prompts harness a key benefit of quasi-experimentation noted by Grant and Wall (2009), namely not having to sort a portion of the participants into conditions that might be deemed unethical because they involve harm, raise feelings of inequity or paternalism, or may be culturally inappropriate (see also Schein, 2015). A single-condition approach to manipulation, when appropriate, offers additional unique advantages. For instance, while certain demand effects remain—such as the influence of social desirability and experimenter effects—using a single condition does eliminate concerns about asymmetric demands between treatment and control groups (Lonati et al., 2018).

Interventions. In detailing the types of manipulations above, we have focused on studies examining the impact of a stimulus or state (the independent variable of interest) on one particular outcome. However, many studies in our field have multiple constructs theoretically impacted by the independent variable, often in sequence as mediating paths, designed to capture the intervening mechanism and provide explanation for why an effect is occurring. We term these types of manipulations interventions and identify them as a variation of the classic treatment design. In this design, a treatment is the start of a causal chain, where scholars are interested in the specific effects of the manipulated stimulus (treatment) on an immediate outcome, while simultaneously hypothesizing the downstream effects of this immediate outcome on other variables of interest (e.g., examining the immediate outcome as a mediating mechanism for the effects of a manipulated variable on some distal outcome of interest). Returning once more to our opening example, this approach might be deployed if researchers were theoretically interested in both the causal impact of algorithmic vs. human decision-making on fairness perceptions *and* the association of these fairness perceptions with organizational commitment—or stated differently, exploring the indirect effect of algorithmic decision-making on organizational commitment with fairness perceptions as the explanatory mediator of this effect.

As indicated by their name, in our literature, this variation of a treatment design lends itself particularly well to the test of workplace interventions, such as the impact of a work–family enrichment training (treatment) through perceptions work-to-family enrichment (immediate state outcome) on the distal outcome of job satisfaction (Heskiau & McCarthy, 2020) or the effect of an online networking treatment, through networking self-efficacy, on reemployment. While interventions can function as part of a multi-study portfolio, we note a trend among organizational research that such designs, when conducted in the field, are often ambitious enough to be published as stand-alone studies. For instance, in a field experiment in rural Ghana, Slade Shantz and colleagues (2020) assigned 40 new cooperatives to a flat or hierarchical control structure (treatment) to examine how collective psychological ownership (as an immediate outcome of the intervention) subsequently impacted conflict.

A Content Analysis of Manipulations in Organizational Research

Having defined archetypes and derivations of manipulations and identified published exemplars of each, we sought to systematically establish their respective usage in our field. To do so, we conducted a targeted methodological content analysis (for examples see Antonakis et al., 2010; Casper et al., 2007; Eby, 2022; Grant & Wall, 2009; Scandura & Williams, 2000), selecting five top, "big tent" journals spanning micro-, meso-, and macro-perspectives in our diverse field: *Academy of Management Journal, Administrative Sciences Quarterly, Journal of Applied Psychology, Personnel Psychology,* and *Organizational Behavior and Human Decision Processes.* We sourced all empirical articles (277) published in a one-year period (2020) and coded all containing at least one manipulation study. In total, we identified 98 articles containing 450 manipulations across 326 studies.

We randomly divided these articles among the authors and coded each for three pieces of information: (1) the type of manipulation employed, (2) the specific mechanics of the manipulation, and (3) the (mis)match between manipulation design and analysis. Not surprisingly, given the lack of discussion on the topic, manipulations were rarely labelled by the terminology proposed herein (or any other consistent nomenclature) and coding could fall into gray areas (e.g., presenting the analysis of a prime manipulation, both correctly using regression and incorrectly as mean comparison). Therefore, this process was iterative and interspersed with regular discussion of our coding and consultation of how manipulation types are conceptually distinguished in related fields (see Oehlert, 2000; O'Keefe, 2003; Welsh et al., 2013).

The Current State of Manipulations

Of the manipulations coded, a clear majority (77%) functioned as treatments. The use of treatments as a manipulation tool dominates our contemporary field. But, it does not escape notice that almost one-fifth of studies (18%) used manipulations that we coded as primes, while the remaining 5% were split into the two sub-types of treatments and primes described earlier: invariant prompts (3%) and interventions (2%). These ratios varied considerably across the journals coded (see Table 2).

As reflected in the number of manipulations (450) compared to the number of studies (326) in our review, studies commonly incorporated multiple, concurrent manipulations (e.g., a 2×2 design), often of the independent variable in conjunction with a mechanism. For instance, Cowen and Montgomery (2020) employed a treatment-by-treatment design in which CEO gender (female vs. male) and CEO response to the organizational failure (sympathy vs. unqualified apology response) were manipulated via a simulated news story to examine consumer response. Watkins and Umphress (2020) used a treatment-by-prime design in which individuals exercised versus read about exercise (a treatment) while feelings of injustice were primed via the experience of whether a confederate answered a staged phone call in their presence (a prime because the researchers were interested in the effect of injustice, not of phone etiquette). In our analysis, 110 studies (34%) used multiple manipulation

Table 2. 2020 Literature Review: Stimuli Type by Journal.

Journal	Treatment (%)	Prime (%)	Invariant Prompt (%)	Intervention (%)	Total Stimuli
Academy of Management Journal	79	14	4	4	28
Administrative Science Quarterly	70	10	20	0	10
Journal of Applied Psychology	65	27	7	0	55
Organizational Behavior and Human Decision Processes	80	15	2	3	344
Personnel Psychology	46	46	0	8	13
Overall (%)	77	18	3	2	450 (100%)

Note: Based on 450 stimuli used across 326 studies in 98 articles.

designs. This often involved a 2 × 2 design with treatments, primes, or one of each (or even higher levels of complexity such as 2 × 2 × 2, 2 × 3, etc.). We identified 69 treatment-by-treatment designs, 20 treatment-by-prime designs, 13 treatment-by-treatment-by-treatment designs, six prime-by-prime designs, one prime-by-treatment-by-treatment design, and one invariant-prompt-by-prime design.

Our review allowed us to clarify that different design choices—including the underlying mechanics of the manipulation—tend to be associated with different manipulation types in a way that could be informative to future scholars. Let us highlight a few key contrasts to make this point (for the entire list, see Table 3). For treatments, vignette designs were the most common (43%), followed by the approach of giving different information across conditions (21%). For example, Roulin and Krings (2020) had participants take on the role of applying for a job at a fictitious organization, wherein the manipulation was embedded in the description of the organization's culture. Such designs were less common for prime manipulations, though they were used at times (14% and 6%, respectively); Umphress and colleagues (2020) had participants read various stories (e.g., someone in a long-term relationship discovering that her partner was cheating, a person peacefully walking on a beach) to prime variance in participant mood (negative and positive). We did not observe any invariant prompts or interventions using vignettes or scenarios.

In contrast, the most common approach to prime variance in a state was via "recall a time" type tasks (37%). Such was the case for Koopman and colleagues (2020), who asked participants to recall and write about a negative or positive event they had experienced in order to generate variance in daily affect. This approach was also used frequently in invariant prompt designs (62%, several of which were noted as critical incident designs, a specific type of recall). Treatments (2%) and interventions (0%) generally did not employ "recall a time" designs.

The vast majority of intervention stimuli were behavioral in nature (82%), such as Slade Shantz et al.'s (2020) manipulation of organizational hierarchy type in a field experiment where co-ops were designed to have a hierarchical or flat structure through training programs. In contrast to their prevalence in interventions, behavioral designs were deployed at similar rates for as primes (14%), treatments (15%), and invariant prompts (15%). For example, as previously noted, Schabram and Heng (2022) instructed participants to engage in different daily acts of compassion as a prime to create variance in the experience of compassion.

Misalignment in Prime Use and Analysis

Through our content analysis, we observe the prevalence of various types of manipulations in the field. Our content analysis also allowed us to consider the fit between research question, data

Table 3. 2020 Literature Review: Manipulation Stimuli Across Types.

Stimuli Type	Treatment (%)	Prime (%)	Invariant Prompt (%)	Intervention (%)	Total Stimuli
"Recall a Time"	2	37	62	0	44
Vignette	43	14	0	0	160
Behavioral Task	15	14	15	82	74
Reading Task	11	11	8	0	48
Writing Task	1	11	8	9	13
Different Information	21	6	0	9	79
Different Instructions	2	4	8	0	11
Exposure (in-person)	4	1	0	0	15
Video Vignettes	1	1	0	0	6

Note: Based on 450 stimuli used across 326 studies in 98 articles.

collection, and analysis. We unearthed extensive misalignment between the use of prime manipulations (as categorized by their conceptual or theoretical intent) and their analysis. Specifically, though 99.5% of studies intended as treatments were analyzed as such (i.e., via mean comparisons or similar approaches to compare outcomes across manipulated conditions), only a small minority of studies intended as primes (5.8%) were analyzed as such (i.e., via associational methods using the measured state as the independent variable). Instead, most prime analyses employed the manipulated condition as the independent variable (i.e., a treatment approach). Below, we discuss the threats arising from this mismatch in primes' conceptualization and analysis, as well as offer some potential reasons underlying the mismatch.

Empirical Threats of Analyzing Primes as Treatments

As noted earlier, conceptual fit and accurate measurement are core goals of any scientific endeavor (Fiedler et al., 2021; Grant & Wall, 2009). To the extent that a study's theory and hypotheses emphasize the effects of a particular state on key outcomes, using a manipulated stimulus as a proxy for that state, particularly when a more direct measure of the state is available (e.g., in a "manipulation check" measure), provides a sub-standard empirical estimation of the effects of interest. Researchers may find themselves—in the pursuit of harnessing the causal interpretability of a treatment design—unfortunately drawing conclusions (however causal) that do not accurately reflect their hypothesis.

Scholars might point to the presence of significant manipulation checks in some of these identified mismatch studies as addressing the concerns of using conditions as stand-ins for the measured state. Indeed, we acknowledge that there is an appealing logic to demonstrating that a manipulation resulted in a significant, expected difference in the primed state to assess manipulation validity (Fiedler et al., 2021), but then using the manipulated conditions as the independent variable in order to capitalize on the causal explanatory ability that comes from random assignment to conditions in treatment designs (Fayant et al., 2017). However, as we illustrate in Figure 2, conducting a manipulation check of the condition on the primed state and testing the effect of the condition on the outcome variable gives statistical evidence for every relationship between the constructs *except* the one of interest: the effect of the primed state on the outcome. Studies analyzed in this way cannot be accurately interpreted as providing support (or not) for a hypothesis regarding the effect of the state on the outcome. A scholar could only claim that the manipulation itself (e.g., observing algorithmic decision-making) has an effect on the outcome of interest (e.g., organizational commitment), but we cannot say that this effect is due to the intervening state (e.g., perceptions of fairness) as we cannot rule out alternative mechanisms: observing algorithmic decision-making might cue some

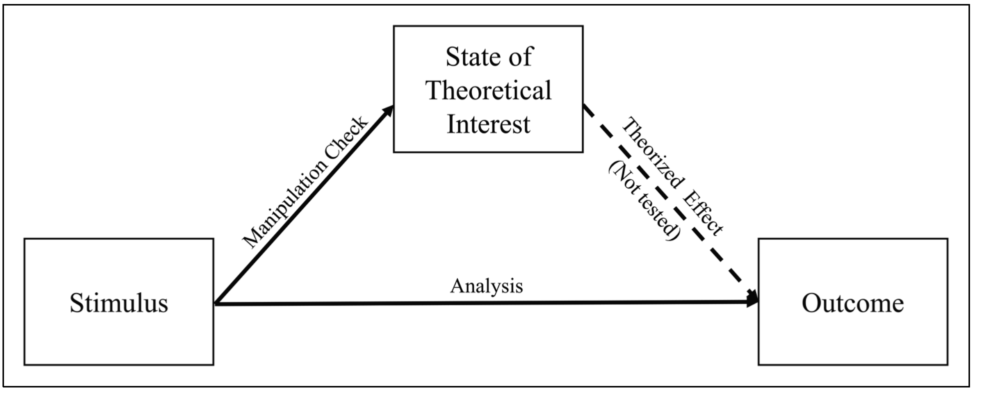

Figure 2. Faulty prime analysis.

other state or feeling that alters commitment, such as anxiety about job loss from being replaced by an automated algorithm. Ironically, using the categorical condition as the predictor because it seems the most rigorous approach (i.e., because it provides a causal test) means using subpar proxy measures, something for which our literature has been criticized (Boyd et al., 2013; Gruijters, 2022). A direct measure of that state (i.e., a self-report scale or other method for assessing of the state) would be preferable to using the stimulus as a more distal proxy for the state.

Analyzing prime designs with the state measure rectifies this empirical concern. Indeed, a number of studies (including examples in our content analysis) do focus their analyses on the state–outcome relationship in prime designed studies; however, these analyses often come with caveats or as a supplement to a more traditional treatment-style analysis. As an example, Study 2 in Anderson and Galinsky's (2006) oft-cited study of priming power (via a recall task) and risk-taking behavior includes both analytical approaches: an ANOVA to establish mean differences in risk-taking across power conditions *and* a correlation between the degree to which participants expressed a sense of power in their written responses with risk-taking behavior. Only by including the latter are the authors able to show a direct test of the relationship between the state and outcome, the closest conceptual fit to the hypothesized effect (though admittedly a rudimentary one in this case, given the reliance on a simple correlation as a supplemental analysis, without steps taken to address any of the concerns noted earlier regarding endogeneity and demand effects).

In describing primes earlier, we also noted an additional empirical advantage in their analysis of a measured state variable: expanding the range of values considered when examining a phenomenon of interest. A treatment-type analysis (i.e., mean comparisons across conditions) nullifies this benefit. Using the categorical condition variable as the predictor, by definition, empirically constricts the variance in the primed state when analyzing its effects (Hunter & Schmidt, 1990; Irwin & McClelland, 2003; Royston et al., 2006), which can lead to limited or misleading claims, particularly when the mean values across conditions (though statistically distinct from one another) reflect only a small portion of the practical range of a state. In other words, showing that on average, participants across conditions differ from one another in a particular state via a manipulation check is not the same as reflecting objectively high and low levels of that state.

This is not an idle concern: mean differences of states across conditions that are statistically, but not practically, distinct are readily found in published papers in our field. Returning to Anderson and Galinsky's (2006; Study 4) power mind-set manipulation, the authors found significant differences in coder-rated power across conditions (7-point scale; high power, $M = 3.77$, $SD = 1.02$, low power $M = 1.03$, $SD = 0.71$). However, claiming these conditions reflect comparisons across high and low levels of power seems somewhat misleading, as the comparison is really between two groups that have very low vs. slightly below midpoint feelings of power. In our review we identified other instances of this compression, such as Yam and colleagues' (2020; Study 1) examination of the effects of inducing perceptions of robot anthropomorphism (7-point scale; anthropomorphism condition, $M = 2.88$, $SD = 1.19$, control condition, $M = 2.51$, $SD = 1.16$), which demonstrate similar challenges. Both studies report significant results for the manipulation check (i.e., the state means differed significantly across conditions), but the means for both conditions fell below the midpoint of the scale, suggesting that claims about "high" levels of each state may be unwarranted.[4]

One might argue that this is actually an empirical benefit of treatments over primes, because using the categorical condition variable constitutes a more conservative test of the phenomenon than a state measure which falls on a continuum. However, even if there is a significant effect of the conditions on an outcome (or significant effects for both the condition and the measured state), testing and reporting the effect of the condition presents incorrect estimates of the true effect of interest (i.e., different coefficients than what one would observe with the measured variable). This impacts not only the results of the particular study but has downstream implications for future research seeking to meta-analyze results, derive expected effect sizes for subsequent studies, or replicate prior work. The observed effects of the measured state allow for more accurate conceptual integration and comparison across studies, as the meaning of higher and lower values of the state are more aligned (e.g., comparing correlations between measured power and risk propensity across multiple studies vs. comparing results of high- or low-power conditions across studies where each condition has a vastly different mean value). Such an argument also assumes one is confident that the conditions have no other plausible mechanism through which they might influence the outcome, and that the expected form of the relationship is well-known, as collapsing a range of data into a limited set of condition mean values masks empirically important nuance in the form of the effect, such as a curvilinear relationship (for an example see Quinn et al., 2021).

Moreover, unnecessary restriction of variance is the best-case scenario: what about situations in which the results of treatment (outcome means by condition) and prime (correlational) analyses differ? This would be particularly likely when a manipulation creates variance in a state, but in an unexpected direction. In the case of our algorithmic decision-making example, this could be the case when a particular participant, perhaps having witnessed biased hiring decisions in their past work experience, believes that an algorithm will be fairer than a human. In that scenario, the algorithmic decision manipulation would result in higher perceptions of fairness, rather than the intended lower perceptions. This individual, by virtue of being an outlier, would have an outsized and problematic effect in analyses of fairness on organizational commitment that used the manipulated decision-making condition as the independent variable. Scholars would be forced to decide on how to handle the outlier, one of the most enduring and inconsistent methodological challenges in our field (Aguinis et al., 2013). But this would not be an issue when using the participant's actual measured sense of fairness. In a prime, any variance is a boon (as it can be captured appropriately in a correlational analysis), preventing a possible misinterpretation of results arising from the suppression of a significant prime-outcome relationship when analyzed as a treatment (i.e., a Type II error).

Normative Pressures Underlying Mismatched Analyses

Our content analysis revealed that primes are being published in our field's top journals, demonstrating a level of acceptance for this approach. The prevalence of mismatched analyses, however, suggests that primes' acceptance is murkier than it initially appears. Indeed, we contend that empirical efforts to fit prime "pegs" into treatment "holes" may be, at least in part, due to normative pressure. Given the status of treatments as the archetypical experiment (Campbell & Stanley, 1967) and the emphasis placed on causality as *the* benefit of experimental designs (Shadish et al., 2002), researchers in our field may feel explicit or implicit pressure to analyze prime studies as if they were treatments, and may believe that analyzing a prime design using the measured state represents a failed experiment (Hauser et al., 2018). Such pressure is often evident in comments from reviewers and editors (as well as published guidelines in the field; see, for example, Antonakis, 2017; Lonati et al., 2018) who encourage authors to conform to treatment-style analyses.

We also gleaned direct evidence of this pressure among the minority of primes in our review that were analyzed as such. In the absence of guidance for publishing these kinds of manipulations, prime studies often included a substantial section justifying their approach to analysis (see for example Frederiks et al., 2019, pp. 333–334). A common, but cumbersome, solution was to call on prior publications for evidence of the validity of the approach: Koopman et al.'s methods (2020) referenced and elaborated on Klein et al.'s (2020) prime-type design, while Schabram and Heng (2022) cited Quinn et al. (2021) who in turn, drew upon similar efforts in communications research to support their approach (O'Keefe, 2003). These citation chains demonstrate the recognition and desire for authors to analyze their work in ways that best align with their (prime) design, while simultaneously making clear the lack of well-established guidance to ground the validity of the approach. When scholars lack explicit guidance and instead do their best based on successful prior publications, their practices may naturally deviate as they follow each idiosyncratic exemplar.

Moreover, our content analysis only captured published studies and we have no idea how many prime-type studies from this time period were judged as "unpublishable" because of design-analysis mismatch (e.g., non-significant results for a prime when analyzed as a treatment). Though there has been extensive debate in our field about "significosis" (Antonakis, 2017, p. 5)—from what should constitute cut-offs of significance (Lance et al., 2006) to whether they should exist at all (Halsey et al., 2015; Kennedy-Shaffer, 2019)—significance continues to drive which projects are published (Abelson, 1997; Fiedler et al., 2021; Lance et al., 2006) and researchers in this position might feel that their study "failed." Such scholars could be forced to contemplate one of several sub-optional options including abandoning the project, thereby contributing to our field's file drawer problem (Franco et al., 2014; Rosenthal, 1979), or exercising concerning practices including re-running the study until it "works" (exacerbating the issue of "capitalizing on chance"; Baumeister & Leary, 1997, p. 313). However, if we recognize the legitimacy of prime designs, when they fit the theory, and judge studies that hypothesize and test the correlational finding between the measured state and outcome as appropriate, their authors might not feel this pressure to contemplate these suboptimal paths.

Best Practices and Considerations for Employing Manipulations

From our review, it is clear that our field deploys multiple types of manipulations. In response to the current lack of explicit differentiation or accompanying guidance, we offer below a list of best practices for the use of manipulation designs in organizational research, drawing upon the groundwork laid by methodological experts in our field, as well as adjacent disciplines.

A Priori Planning

Foremost, we advocate for scholars to carefully reflect on the nature of their manipulations and to *a priori* consider the design and analysis implied by their theory and hypotheses. As our opening example highlighted, essentially identical data collections can be used to examine decidedly different hypothesized effects. The crucial distinction lies in how each team makes use of the manipulation and state measure as a function of their hypotheses. We argue that decisions need to be made in the design phase rather than after data collection, as this *a priori* planning of design and analysis is aligned with rigorous scientific practices, akin to conducting a power analysis before deciding on a sample size.

Such an early decision also allows scholars to aim for the highest standards for their respective manipulation, be that an emphasis on objective and/or behavioral operationalizations for treatments, or taking care to avoid demand effects when using subjective and/or self-evaluations for primes. Without singling out any particular study for criticism, in our review, treatments most often used hypothetical vignettes while primes favored recall-a-time manipulations (see Table 3); the former have been labeled as best to be avoided in true experiments because they fuel non-consequential decision making (see commandment IV; Lonati et al., 2018), while the latter have been criticized for "the assumption that everyone has relevant memories" (Khademi et al., 2021, p. 3) and because they risk demand effects via unfair comparison when treated participants receive stronger cues than those in a neutral control (see commandment III; Lonati et al., 2018). Following McGrath's (1981) advice that the research process is not so much a set of problems to be solved by making the "right choice," but instead a set of interlocking dilemmas to be contended with, we hope that our typology and content analysis can offer those planning a manipulation design inspiration for which designs are most appropriate, rather than just most common, for a given research question.[5]

Crucially, we caution researchers who designed a treatment against thoughtlessly analyzing it as a prime when a comparison of the conditions results in non-significant outcomes. Conversely, those designing a prime should have confidence in their design and analyze it using the measured state as the independent variable, rather than as a treatment as was common in our literature review). *A priori* planning should give scholars enough guidance to limit themselves to the most appropriate method without feeling compelled to hedge or change analysis strategies. We do note one compelling exception, as guidelines for field experiments (Eden, 2017; see also King et al., 2013) allow for "quasification" in these studies when evolving circumstances undo randomization and thereby render a treatment impossible.

Contextualizing and Complementing Manipulation Designs

In addition to designing a specific manipulation, scholars should also carefully consider each individual manipulation study in the context of the larger manuscript, particularly given the trend toward multi-study papers. The field of organizational research has come a long way since Spencer and colleagues (2005, p. 847) wrote that "though using multiple methods to test a theoretical account would be ideal, we feel that in most situations requiring such multiple methods would be setting such a high standard that progress in the field might well be impeded." Over the past two decades, multi-study papers, triangulation, and a plurality of data analytic approaches have proliferated (Cortina et al., 2017; Wellman et al., 2023). Papers that were part of our content analysis included between one and eleven manipulation studies ($M = 3.29$, $SD = 1.98$) and scholars should think carefully about when and why to deploy different designs.

Spencer and colleagues' (2005) point may still be apt when it comes to interventions, which often incorporate more elaborate and ambitious field applications and are quite robust in isolation. In contrast, the other designs lend themselves as distinct complements in a multi-study package. Treatments, for instance, may pair well with inductive efforts in multimethod papers following either an "explore-and-test" logic, wherein qualitative insights are supplemented with deductive tests, or a "test-and-explore" approach wherein a deductive test establishes a causal link that is then pursued via inductive inquiry (Wellman et al., 2023).

Primes and invariant prompts also lend themselves well to pairing with another data source as part of a multi-method paper, such as field survey to enhance ecological validity, or a treatment-design to establish causality. For instance, they might be suited to a twist on full-cycle micro organizational behavior approaches (Chatman & Flynn, 2005), in which the phenomenon is first identified in the field (Cialdini, 1995) and then constructively replicated via controlled designs. A multi-study paper could first establish the phenomenon via prime and then replicate it via conceptually complementary treatments (Ilgen, 1985). They would also function well as part of a generalization and extension approach (Tsang & Kwan, 1999), wherein researchers triangulate their predictions across multiple imprecise but complementary replications, leveraging significant effects that hold across different approaches.

Transparency in Communication

Organizational behavior research reflects a collective body of knowledge that is inherently diverse and expansive (Agarwal & Hoetker, 2007; Ferris et al., 2008; Kouchaki, 2020). Specialization means that our field is frequently split into insular communities (Alvesson & Gabriel, 2013) with their own writing conventions (Alvesson & Sandberg, 2013). Establishing nomenclature is not only a foundational practice in any epistemological endeavor (Ohl, 2018) but crucial to overcome division in our "big tent" field (for influential exemplars see Chan, 1998; Kozlowski & Klein, 2000). Naming and defining distinct categories of manipulation can create a bridging language around how diverse topics are being studied.

Herein, we took the parsimonious approach of importing labels from other fields whenever possible.[6] For instance, the term treatment is widely used in other disciplines and often synonymous with experimental manipulation (e.g., "Experimental manipulation … which are also called treatment variables or factors," Allen, 2017, p. 274). By adopting a common language around manipulation types in our own field, we believe that authors, editors, and reviewers can be more confident in their shared understanding when producing and evaluating research utilizing manipulations. Moreover, a shared lexicon to "facilitate scientific communication" (Chan, 1998, p. 234) eases meaningful replication and extension, supporting the field's increased emphasis on rigorous theory-pruning (Leavitt et al., 2010) and the broader efforts of the open-science movement (Kouchaki, 2020; Open Science Collaboration, 2012, 2015). One could argue that such labeling should be welcome by even the harshest critics of any of the manipulation designs as it allows for a more streamlined discussion of their limitations. Furthermore, our labels integrate our scholarship with other domains of applied science (such as medicine or economics). Simple and consistent language around the processes by which we draw conclusions might even help bridge the oft-acknowledged divide between academia and those who stand to benefit from our work outside the ivory tower (Byington & Felps, 2017).

Our final recommendation when it comes to transparency in communication is for scholars to share—whether in their manuscript or an online repository—as much of their data and design as possible. For treatments and interventions that incorporated a measured state, for the purpose of a manipulation check, this would mean sharing that measured state even if it is not part of the central analysis. For primes and invariant prompts, despite the manipulation technically taking a backseat to the primed state in the test of hypotheses, scholars should carefully describe their manipulation as well as report descriptive statistics for each condition as supplemental data (i.e., data not used to assess support for a hypothesis). This may be paradoxical advice, given that we have stressed that the effect of the stimulus is irrelevant to the theoretical purpose of the study. However, we argue that this is an important practice in the interest of open science, as it demonstrates the impact of the stimuli used for readers to better understand the impact of the design. Even if the study's research question does not pertain to the direct effect of the manipulation, this data may be of interest to other scholars such as those conducting a meta-analysis or those interested in adapting the design to study topics related to the manipulated conditions (as in our algorithmic decision-making example).

Conclusion

Those concerned with the state of methodological practice have noted that "many of the problems facing the field are not likely be solved without radical shifts in its philosophy" (Ferris et al., 2012, p. 94). While this may be true in general, we see the mindful design and transparent discussion of manipulation designs as more low-hanging fruit.

Though well-intentioned, adherence to a singular stringent convention that suits one manipulation type (i.e., treatments), has not prevented others from emerging, but has created confusion and promoted misapplication and potentially incorrect conclusions. By carefully distinguishing treatments and primes we seek to offer organizational scholars a consistent framework for articulating and planning their studies, particularly when they differ from the "conventional" treatment archetype. Acknowledging different types of manipulations not only helps clarify how primes might be used, but also provides greater clarity about what treatments are, by disentangling what they are not. Both treatments and primes (and their derivations) should be recognized as specialized, valuable tools in the methodological kit of organizational researchers. Our intention is not to advocate for any one design. Each manipulation type has its advantages and limitations—as there is no perfect method—and should be deployed when it provides the best test of a particular hypothesis. Theory-method fit is always the objective.

We reiterate that the typology we present here is primarily conceptual in focus. We have not articulated specific statistical techniques for analyzing the different types of manipulations (beyond general orientations towards mean-comparison vs. associational approaches), but have instead directed readers to others who better detail their execution (e.g., on the use of manipulation checks, Hauser et al., 2018; on techniques for addressing endogeneity in prime manipulations, Antonakis et al., 2010; Sajons, 2020). We hope our introduction of terminology and guidelines provides researchers greater confidence when deploying manipulations and readers with a standard for evaluating the legitimacy of their claims.

Acknowledgments

All authors contributed equally to this manuscript. Our work was seeded when we found ourselves—as authors and reviewers—engaged in discussions about the best path forward for prime-designed manipulations during the review process for several distinct research projects. We sincerely thank those earlier review teams, as well as associate editor Justin DeSimone and the two anonymous reviewers of this paper, for engaging with us in a fruitful conversation to disentangle manipulation types. We thank Beth Campbell and participants at the Wharton OB conference for helpful feedback on earlier versions of this work. We thank Young Won Rhee for early research assistance.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

ORCID iDs

Kira F. Schabram https://orcid.org/0000-0002-1879-5492
Christopher G. Myers https://orcid.org/0000-0001-7788-8595
Ashley E. Hardin https://orcid.org/0000-0001-6466-3952

Notes

1. Analysis of treatments often also includes an assessment of whether the manipulation "worked" via a manipulation check. The uses and drawbacks of these sorts of checks in treatment designs have been discussed extensively elsewhere in the literature (Ejelöv & Luke, 2020; Hauser et al., 2018), so we do not delve into them here. We simply note that the elegant simplicity of a treatment lies in its reliance on random assignment to control for any varying factors or confounds in the design (Grant & Wall, 2009), including varying attention to engagement with the manipulation by participants in each condition. Treatments are also generally more objective in nature (Cortina et al., 2017) and can be judged on their face to have occurred or not. Taken together, this suggests that manipulation checks in these designs are at best unnecessary, and at worst can drive demand effects (Lonati et al., 2018; Zizzo, 2010) and threaten the validity of the causal claim that is the hallmark of treatments (Hauser et al., 2018).
2. In this way, a variable that might be deemed a "manipulation check" measure in a traditional treatment design is the focal independent variable for analysis in a prime design. Prior work regarding the various uses of "manipulation check" measures has noted the benefit of being able to use these measures as a form of secondary "internal analysis" examining the correlation between a state and outcome of interest when a treatment "fails" (Hauser et al., 2018, p. 3). However, we argue that this measured state-outcome association is actually

of greatest interest in a prime design (as it is closest to the theorized relationship), and should be the primary test of the hypothesis, rather than cast as a consolation to a "failed" treatment. Indeed, the treatment's "failure" may be an artifact of forcing a prime-type design into a treatment-type analysis, as discussed further below.

3. A complete discussion of methods to address endogeneity in correlational designs is beyond the scope of this paper, and we refer readers to the rich, well-established literature on these approaches (e.g., Antonakis et al., 2010; Bastardoz et al., 2023; Sajons, 2020). However, we stress again that the particular choice of method and approach to address these issues must be theoretically driven and should not be a knee-jerk reaction to assuage perceived institutional norms and pressures. For example, researchers might choose to include the manipulated stimulus conditions as variables in the model, but must remember that simply including the stimuli in the model does not automatically improve the ability to claim causality in the state-outcome relationship (Sajons, 2020)—any effect of the state on an outcome is still correlational (see also Hauser et al., 2018 for a discussion in the context of manipulation checks).
4. We note here a trend in published papers toward describing conditions as being "lower" or "higher" (emphasis added) representations of a state (rather than "low" or "high"). While semantically more accurate, these adjustments do not fully resolve the underlying conceptual misfit highlighted here.
5. One additional dilemma that bears mention in these design decisions is that of cost. All research is expensive in one way or another (see Buhrmeister et al., 2011; Eden, 2017). Thoughtful a priori design can help make sure resources are used judiciously in a "take only what you need" paradigm. For instance, adding multiple conditions when the nature of the study's hypothesis would suit an invariant prompt unnecessarily amplifies costs, contributing to growing rifts between haves and have-nots in research (Leavitt et al., 2021) and potentially contributing to underpayment of participants, an increasingly concerning critique of the behavioral sciences (Felstiner, 2011; Samuel, 2018). Similarly, running many incremental variations of the same design instead of considering a complementary package of designs can be wasteful at best and, at worst, risk contamination of subject pools (see Kraut et al., 2004). We stress of course, that any logistical considerations are secondary—when a research question does not lend itself to a particular design, theory must always trump efficiency.
6. The term prime may face a higher hurdle to adoption in our particular domain, due to extensive, contemporary scrutiny of social/behavioral "priming" practices, defined as subtle interventions intended to affect people's behavior (Chivers, 2019; Sherman & Rivers, 2020). We stress that primes are not synonymous with social priming.

References

Abelson, R. P. (1997). On the surprising longevity of flogged horses: Why there is a case for the significance test. *Psychological Science*, *8*(1), 12–15. https://doi.org/10.1111/j.1467-9280.1997.tb00536.x

Abraham, M., & Burbano, V. (2022). Congruence between leadership gender and organizational claims affects the gender composition of the applicant pool: Field experimental evidence. *Organization Science*, *33*(1), 393–413. https://doi.org/10.1287/orsc.2021.1442

Agarwal, R., & Hoetker, G. (2007). A Faustian bargain? The growth of management and its relationship with related disciplines. *Academy of Management Journal*, *50*(6), 1304–1322. https://doi.org/10.5465/amj.2007.28165901

Aguinis, H., Gottfredson, R. K., & Joo, H. (2013). Best practice recommendations for defining, identifying, and handling outliers. *Organizational Research Methods*, *16*(2), 270–301. https://doi.org/10.1177/1094428112470848

Aguinis, H., & Vandenberg, R. J. (2014). An ounce of prevention is worth a pound of cure: Improving research quality before data collection. *Annual Review of Organizational Psychology and Organizational Behavior*, *1*(1), 569–595. https://doi.org/10.1146/annurev-orgpsych-031413-091231

Allen, M. (2017). Experimental manipulation. In M. Allen (Ed.), *The SAGE encyclopedia of communication research methods* (pp. 474–479). Sage.

Alvesson, M., & Gabriel, Y. (2013). Beyond formulaic research: In praise of greater diversity in organizational research and publications. *Academy of Management Learning & Education*, *12*(2), 245–263. https://doi.org/10.5465/amle.2012.0327

Alvesson, M., & Sandberg, J. (2013). Has management studies lost its way? Ideas for more imaginative and innovative research. *Journal of Management Studies*, *50*(1), 128–152. https://doi.org/10.1111/j.1467-6486.2012.01070.x

Anderson, C., & Galinsky, A. D. (2006). Power, optimism, and risk-taking. *European Journal of Social Psychology*, *36*(4), 511–536. https://doi.org/10.1002/ejsp.324

Antonakis, J. (2017). On doing better science: From thrill of discovery to policy implications. *The Leadership Quarterly*, *28*(1), 5–21. https://doi.org/10.1016/j.leaqua.2017.01.006

Antonakis, J., Bendahan, S., Jacquart, P., & Lalive, R. (2010). On making causal claims: A review and recommendations. *The Leadership Quarterly*, *21*(6), 1086–1120. https://doi.org/10.1016/j.leaqua.2010.10.010

Austin, J. T., Scherbaum, C. A., & Mahlman, R. A. (2002). History of research methods in industrial and organizational psychology: Measurement, design, analysis. In S. G. Rogelberg (Ed.), *Handbook of research methods in industrial and organizational psychology* (pp. 3–33). Blackwell Publishing.

Bacharach, S. B. (1989). Organizational theories: Some criteria for evaluation. *The Academy of Management Review*, *14*(4), 496–515. https://doi.org/10.2307/258555

Banks, G. C., Woznyj, H. M., & Mansfield, C. A. (2023). Where is "behavior" in organizational behavior? A call for a revolution in leadership research and beyond. *The Leadership Quarterly*, *34*(6), 101581. https://doi.org/10.1016/j.leaqua.2021.101581

Bastardoz, N., Matthews, M. J., Sajons, G. B., Ransom, T., Kelemen, T. K., & Matthews, S. H. (2023). Instrumental variables estimation: Assumptions, pitfalls, and guidelines. *The Leadership Quarterly*, *34*(1), 101673. https://doi.org/10.1016/j.leaqua.2022.101673

Baumeister, R. F., & Leary, M. R. (1997). Writing narrative literature reviews. *Review of General Psychology*, *1*(3), 311–320. https://doi.org/10.1037/1089-2680.1.3.311

Botelho, T. L., & Chang, M. (2023). The evaluation of founder failure and success by hiring firms: A field experiment. *Organization Science*, *34*(1), 484–508. https://doi.org/10.1287/orsc.2022.1592

Boyd, B. K., Bergh, D. D., Ireland, R. D., & Ketchen, D. J. (2013). Constructs in strategic management. *Organizational Research Methods*, *16*(1), 3–14. https://doi.org/10.1177/1094428112471298

Brodsky, A. (2020). Virtual surface acting in workplace interactions: Choosing the best technology to fit the task. *Journal of Applied Psychology*, *106*(5), 714–733. https://doi.org/10.1037/apl0000805

Brue, S. L., & Grant, R. R. (2013). *The evolution of economic thought*. South-Western.

Buhrmeister, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science*, *6*(1), 3–5. https://doi.org/10.1177/1745691610393980

Byington, E. K., & Felps, W. (2017). Solutions to the credibility crisis in management science. *Academy of Management Learning & Education*, *16*(1), 142–162. https://doi.org/10.5465/amle.2015.0035

Campbell, D. T., & Stanley, J. C. (1967). *Experimental and quasi-experimental designs for research* (2015th ed.). Ravenio Books.

Casper, W. J., Eby, L. T., Bordeaux, C., Lockwood, A., & Lambert, D. (2007). A review of research methods in IO/OB work-family research. *Journal of Applied Psychology*, *92*(1), 28–43. https://doi.org/10.1037/0021-9010.92.1.28

Chan, D. (1998). Functional relations among constructs in the same content domain at different levels of analysis: A typology of composition models. *Journal of Applied Psychology*, *83*(2), 234–246. https://doi.org/10.1037/0021-9010.83.2.234

Chatman, J. A., & Flynn, F. J. (2005). Full-cycle micro-organizational behavior research. *Organization Science*, *16*(4), 434–447. https://doi.org/10.1287/orsc.1050.0136

Chivers, T. (2019). What's next for psychology's embattled field of social priming. *Nature*, *576*(7786), 200–202. https://doi.org/10.1038/d41586-019-03755-2

Chua, R., & Jin, M. (2020). Across the great divides: Gender dynamics influence how intercultural conflict helps or hurts creative collaboration. *Academy of Management Journal*, *63*(3), 903–934. https://doi.org/10.5465/amj.2016.1319

Cialdini, R. B. (1995). A full-cycle approach to social psychology. In G. G. Brannigan, & M. R. Merrens (Eds.), *The social psychologists: Research adventures* (pp. 53–72). McGraw-Hill.

Cortina, J. M., Aguinis, H., & DeShon, R. P. (2017). Twilight of dawn or of evening? A century of research methods. *Journal of Applied Psychology*, *102*, 274–290. https://doi.org/10.1037/apl0000163

Cowen, A. P., & Montgomery, N. V. (2020). To be or not to be sorry? How CEO gender impacts the effectiveness of organizational apologies. *Journal of Applied Psychology*, *105*(2), 196–208. https://doi.org/10.1037/apl0000430

DeCelles, K. A., Sonenshein, S., & King, B. G. (2020). Examining anger's immobilizing effect on institutional insiders' action intentions in social movements. *Administrative Science Quarterly*, *65*(4), 847–886. https://doi.org/10.1177/0001839219879646

Dobbins, G. H., Lane, I. M., & Steiner, D. D. (1988). A note on the role of laboratory methodologies in applied behavioural research: Don't throw out the baby with the bath water. *Journal of Organizational Behavior*, *9*(3), 281–286. https://doi.org/10.1002/job.4030090308

Eby, L. T. (2022). Reflections on the Journal of Applied Psychology in times of change. *Journal of Applied Psychology*, *107*(1), 1–8. https://doi.org/10.1037/apl0001000

Eden, D. (2017). Field experiments in organizations. *Annual Review of Organizational Psychology and Organizational Behavior*, *4*(1), 91–122. https://doi.org/10.1146/annurev-orgpsych-041015-062400

Ejelöv, E., & Luke, T. J. (2020). Rarely safe to assume: Evaluating the use and interpretation of manipulation checks in experimental social psychology. *Journal of Experimental Social Psychology*, *87*, 103937. https://doi.org/10.1016/j.jesp.2019.103937

Ellsworth, P. C., & Gonzalez, R. (2007). Questions and comparisons: Methods of research in social psychology. In M. A. Hogg, & J. Cooper (Eds.), *The Sage handbook of social psychology* (pp. 24–42). Sage.

Engel, Y., Ramesh, A., & Steiner, N. (2020). Powered by compassion: The effect of loving-kindness meditation on entrepreneurs' sustainable decision-making. *Journal of Business Venturing*, *35*(6), 105986. https://doi.org/10.1016/j.jbusvent.2019.105986

Fayant, M.-P., Sigall, H., Lemonnier, A., Retsin, E., & Alexopoulos, T. (2017). On the limitations of manipulation checks: An obstacle toward cumulative science. *International Review of Social Psychology*, *30*(1), 125–130. https://doi.org/10.5334/irsp.102

Felstiner, A. (2011). Working the crowd: Employment and labor law in the crowdsourcing industry. *Berkeley Journal of Employment and Labor Law*, *32*(1), 143–203.

Ferris, G. R., Hochwarter, W. A., & Buckley, M. R. (2012). Theory in the organizational sciences: How will we know it when we see it? *Organizational Psychology Review*, *2*(1), 94–106. https://doi.org/10.1177/2041386611423696

Ferris, G. R., Ketchen, D. J., & Buckley, M. R. (2008). Making a life in the organizational sciences: No one ever said it was going to be easy. *Journal of Organizational Behavior*, *29*(6), 741–753. https://doi.org/10.1002/job.533

Festinger, L. E. (1953). Laboratory experiments. In Festinger L., & Katz D. (Eds.), *Research methods in the behavioral sciences* (pp. 136–172). Holt, Rinehart, & Winston.

Fiedler, K., McCaughey, L., & Prager, J. (2021). Quo vadis, methodology? The key role of manipulation checks for validity control and quality of science. *Perspectives on Psychological Science*, *16*(4), 816–826. https://doi.org/10.1177/1745691620970602

Forster, J., & Liberman, N. (2013). Knowledge activation. In Kruglanski A. W., & Higgins E. T. (Eds.), *Social Psychology, Second Edition: Handbook of Basic Principles* (pp. 201–231). Guilford Publications.

Foulk, T. A., Pater, I. E. D., Schaerer, M., Plessis, C. D., Lee, R., & Erez, A. (2020). It's lonely at the bottom (too): The effects of experienced powerlessness on social closeness and disengagement. *Personnel Psychology*, *73*(2), 363–394. https://doi.org/10.1111/peps.12358

Franco, A., Malhotra, N., & Simonovits, G. (2014). Publication bias in the social sciences: Unlocking the file drawer. *Science*, *345*(6203), 1502–1505. https://doi.org/10.1126/science.1255484

Frederiks, A. J., Englis, B. G., Ehrenhard, M. L., & Groen, A. J. (2019). Entrepreneurial cognition and the quality of new venture ideas: An experimental approach to comparing future-oriented cognitive processes. *Journal of Business Venturing*, *34*(2), 327–347. https://doi.org/10.1016/j.jbusvent.2018.05.007

Gaw, A. (2009). *Trial by fire: Lessons from the history of clinical trials*. SA Press.

Grant, A. M., & Wall, T. D. (2009). The neglected science and art of quasi-experimentation: Why-to, when-to, and how-to advice for organizational researchers. *Organizational Research Methods*, *12*(4), 653–686. https://doi.org/10.1177/1094428108320737

Gruijters, S. L. (2022). Making inferential leaps: Manipulation checks and the road towards strong inference. *Journal of Experimental Social Psychology*, *98*, 104251. https://doi.org/10.1016/j.jesp.2021.104251

Halsey, L. G., Curran-Everett, D., Vowler, S. L., & Drummond, G. B. (2015). The fickle P value generates irreproducible results. *Nature Methods*, *12*(3), 179–185. https://doi.org/10.1038/nmeth.3288

Hardin, A. E., Bauman, C. W., & Mayer, D. M. (2020). Show me the… family: How photos of meaningful relationships reduce unethical behavior at work. *Organizational Behavior and Human Decision Processes*, *161*, 93–108. https://doi.org/10.1016/j.obhdp.2020.04.007

Hauser, D. J., Ellsworth, P. C., & Gonzalez, R. (2018). Are manipulation checks necessary? *Frontiers in Psychology*, *9*, 998. https://doi.org/10.3389/fpsyg.2018.00998

Heskiau, R., & McCarthy, J. M. (2020). A work–family enrichment intervention: Transferring resources across life domains. *Journal of Applied Psychology*, *106*(10), 1573–1585. https://doi.org/10.1037/apl0000833

Highhouse, S. (2009). Designing experiments that generalize. *Organizational Research Methods*, *12*(3), 554–566. https://doi.org/10.1177/1094428107300396

Hsu, D. K., Mitchell, J. R., & Cao, X. (2024). Examining psychological mediators in entrepreneurship: Experimental designs, remedies, and recommendations. *Entrepreneurship Theory and Practice*, *48*(1), 418–445. https://doi.org/10.1177/10422587231152824

Hunter, J. E., & Schmidt, F. L. (1990). Dichotomization of continuous variables: The implications for meta-analysis. *Journal of Applied Psychology*, *75*(3), 334–349. https://doi.org/10.1037/0021-9010.75.3.334

Ilgen, D. R. (1985). *Laboratory research: A question of when, not if*. Michigan State University, Department of Psychology.

Irwin, J. R., & McClelland, G. H. (2003). Negative consequences of dichotomizing continuous predictor variables. *Journal of Marketing Research*, *40*(3), 366–371. https://doi.org/10.1509/jmkr.40.3.366.19237

Janiszewski, C., & Wyer, R. S. (2014). Content and process priming: A review. *Journal of Consumer Psychology*, *24*(1), 96–118. https://doi.org/10.1016/j.jcps.2013.05.006

Kennedy-Shaffer, L. (2019). Before p< 0.05 to beyond p< 0.05: Using history to contextualize p-values and significance testing. *The American Statistician*, *73*(1), 82–90. https://doi.org/10.1080/00031305.2018.1537891

Kerlinger, F. N. (1973). *Foundations of behavioral research*. Holt Rinehalt and Winston.

Khademi, M., Schmid Mast, M., Zehnder, C., & De Saint Priest, O. (2021). The problem of demand effects in power studies: Moving beyond power priming. *The Leadership Quarterly*, *32*(4), 101496. https://doi.org/10.1016/j.leaqua.2021.101496

King, E. B., Hebl, M. R., Botsford Morgan, W., & Ahmad, A. S. (2013). Field experiments on sensitive organizational topics. *Organizational Research Methods*, *16*(4), 501–521. https://doi.org/10.1177/1094428112462608

Klein, H. J., Lount, R. B., Jr., Park, H. M., & Linford, B. J. (2020). When goals are known: The effects of audience relative status on goal commitment and performance. *Journal of Applied Psychology*, *105*(4), 372–389. https://doi.org/10.1037/apl0000441

Koopman, J., Conway, J. M., Dimotakis, N., Tepper, B. J., Lee, Y. E., Rogelberg, S. G., & Lount, R. B., Jr. (2020). Does CWB repair negative affective states, or generate them? Examining the moderating role of trait empathy. *Journal of Applied Psychology*, *106*(10), 1493–1516. https://doi.org/10.1037/apl0000837

Kouchaki, M. (2020). OBHDP Editorial: Where we are, how we got here, and where we're going. *Organizational Behavior and Human Decision Processes*, *158*, A1–A2. https://doi.org/10.1016/j.obhdp.2020.05.001

Kozlowski, S. W. J., & Klein, K. J. (2000). A multilevel approach to theory and research in organizations: Contextual, temporal, and emergent processes. In K. J. Klein, & S. W. J. Kozlowski (Eds.), *Multilevel theory, research and methods in organizations: Foundations, extensions, and new directions* (pp. 3–90). Jossey-Bass.

Kraut, R., Olson, J., Banaji, M., Bruckman, A., Cohen, J., & Couper, M. (2004). Psychological research online: Report of board of scientific affairs' advisory group on the conduct of research on the internet. *American Psychologist*, *59*(2), 105–117. https://doi.org/10.1037/0003-066X.59.2.105

Lance, C. E., Butts, M. M., & Michels, L. C. (2006). The sources of four commonly reported cutoff criteria: What did they really say? *Organizational Research Methods*, *9*(2), 202–220. https://doi.org/10.1177/1094428105284919

Leavitt, K., Mitchell, T. R., & Peterson, J. (2010). Theory pruning: Strategies to reduce our dense theoretical landscape. *Organizational Research Methods*, *13*(4), 644–667. https://doi.org/10.1177/1094428109345156

Leavitt, K., Schabram, K., Hariharan, P., & Barnes, C. M. (2021). Ghost in the machine: On organizational theory in the age of machine learning. *Academy of Management Review*, *46*(4), 750–777. https://doi.org/10.5465/amr.2019.0247

Locke, E. A. (1986). Ecological validity or abstraction of essential elements. In E. A. Locke (Ed.), *Generalizing from laboratory to field settings* (pp. 3–9). Lexington Books.

Lonati, S., Quiroga, B. F., Zehnder, C., & Antonakis, J. (2018). On doing relevant and rigorous experiments. Review and recommendations. *Journal of Operations Management*, *64*, 19–40. https://doi.org/10.1016/j.jom.2018.10.003

Lyons, B. J., Lynch, J. W., & Johnson, T. D. (2020). Gay and lesbian disclosure and heterosexual identity threat: The role of heterosexual identity commitment in shaping de-stigmatization. *Organizational Behavior and Human Decision Processes*, *160*, 1–18. https://doi.org/10.1016/j.obhdp.2020.03.001

Maydeu-Olivares, A., Shi, D., & Fairchild, A. J. (2020). Estimating causal effects in linear regression models with observational data: The instrumental variables regression model. *Psychological Methods*, *25*(2), 243–258. https://doi.org/10.1037/met0000226

McGrath, J. E. (1981). Dilemmatics: The study of research choices and dilemmas. *American Behavioral Scientist*, *25*(2), 179–210. https://doi.org/10.1177/000276428102500205

Minson, J. A., Bendersky, C., de Dreu, C., Halperin, E., & Schroeder, J. (2023). Experimental studies of conflict: Challenges, solutions, and advice to junior scholars. *Organizational Behavior and Human Decision Processes*, *177*, 104257. https://doi.org/10.1016/j.obhdp.2023.104257

Newman, D. T., Fast, N. J., & Harmon, D. J. (2020). When eliminating bias isn't fair: Algorithmic reductionism and procedural justice in human resource decisions. *Organizational Behavior and Human Decision Processes*, *160*, 149–167. https://doi.org/10.1016/j.obhdp.2020.03.008

Oehlert, G. W. (2000). *A first course in design and analysis of experiments*. W. H. Freeman.

Ohl, M. (2018). *The art of naming*. MIT Press.

O'Keefe, D. J. (2003). Message properties, mediating states, and manipulation checks: Claims, evidence, and data analysis in experimental persuasive message effects research. *Communication Theory*, *13*(3), 251–274. https://doi.org/10.1111/j.1468-2885.2003.tb00292.x

Open Science Collaboration. (2012). An open, large-scale, collaborative effort to estimate the reproducibility of psychological science. *Perspectives on Psychological Science*, *7*(6), 657–660. https://doi.org/10.1177/1745691612462588

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251), 943–944. https://doi.org/10.1126/science.aac4716

Pinker, S. (2012). *The better angels of our nature: Why violence has declined*. Penguin Publishing Group.

Priesemuth, M., & Bigelow, B. (2020). It hurts me too! (or not?): Exploring the negative implications for abusive bosses. *Journal of Applied Psychology*, *105*(4), 410–421. https://doi.org/10.1037/apl0000447

Qin, X., Chen, C., Yam, K. C., Huang, M., & Ju, D. (2020). The double-edged sword of leader humility: Investigating when and why leader humility promotes versus inhibits subordinate deviance. *Journal of Applied Psychology*, *105*(7), 693–712. https://doi.org/10.1037/apl0000456

Quinn, R. W., Myers, C. G., Kopelman, S., & Simmons, S. (2021). How did you do that? Exploring the motivation to learn from others' exceptional success. *Academy of Management Discoveries*, *7*(1), 15–39. https://doi.org/10.5465/amd.2018.0217

Rees, L., Chi, S.-C. S., Friedman, R., & Shih, H.-L. (2020). Anger as a trigger for information search in integrative negotiations. *Journal of Applied Psychology*, *105*(7), 713–731. https://doi.org/10.1037/apl0000458

Restubog, S. L. D., Schilpzand, P., Lyons, B., Midel Deen, C., & He, Y. (2023). The vulnerable workforce: A call for research. *Journal of Management*, *49*(7), 2199–2207. https://doi.org/10.1177/01492063231177446

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin*, *86*(3), 638–641. https://doi.org/10.1037/0033-2909.86.3.638

Roulin, N., & Krings, F. (2020). Faking to fit in: Applicants' response strategies to match organizational culture. *Journal of Applied Psychology*, *105*(2), 130–145. https://doi.org/10.1037/apl0000431

Royston, P., Altman, D. G., & Sauerbrei, W. (2006). Dichotomizing continuous predictors in multiple regression: A bad idea. *Statistics in Medicine*, *25*(1), 127–141. https://doi.org/10.1002/sim.2331

Sajons, G. B. (2020). Estimating the causal effect of measured endogenous variables: A tutorial on experimentally randomized instrumental variables. *The Leadership Quarterly*, *31*(5), 101348. https://doi.org/10.1016/j.leaqua.2019.101348

Samuel, A. (2018, May 15). *Amazon's Mechanical Turk has reinvented research*. JSTOR Daily. https://daily.jstor.org/amazons-mechanical-turk-has-reinvented-research/

Scandura, T. A., & Williams, E. A. (2000). Research methodology in management: Current practices, trends, and implications for future research. *Academy of Management Journal*, *43*(6), 1248–1264. https://doi.org/10.2307/1556348

Schabram, K., & Heng, Y. T. (2022). How other and self-compassion reduce burnout through resource replenishment. *Academy of Management Journal*, *65*(2), 453–478. https://doi.org/10.5465/amj.2019.0493

Schein, E. H. (2015). Organizational psychology then and now: Some observations. *Annual Review of Organizational Psychology and Organizational Behavior*, *2*(1), 1–19. https://doi.org/10.1146/annurev-orgpsych-032414-111449

Shadish, W. S., Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton Mifflin.

Sherman, J. W., & Rivers, A. M. (2020). Social priming: Time to ditch a dubious term? *Nature*, *579*(7797), 29. https://doi.org/10.1038/d41586-020-00616-1

Slade Shantz, A. F., Kistruck, G. M., Pacheco, D. F., & Webb, J. W. (2020). How formal and informal hierarchies shape conflict within cooperatives: A field experiment in Ghana. *Academy of Management Journal*, *63*(2), 503–529. https://doi.org/10.5465/amj.2018.0335

Spencer, S. J., Zanna, M. P., & Fong, G. T. (2005). Establishing a causal chain: Why experiments are often more effective than mediational analyses in examining psychological processes. *Journal of Personality and Social Psychology*, *89*(6), 845–851. https://doi.org/10.1037/0022-3514.89.6.845

Steffens, B. (2007). *Ibn Al-Haytham: First scientist*. Morgan Reynolds Pub.

Tsang, E. W., & Kwan, K.-M. (1999). Replication and theory development in organizational science: A critical realist perspective. *The Academy of Management Review*, *24*(4), 759–780. https://doi.org/10.2307/259353

Umphress, E. E., Gardner, R. G., Stoverink, A. C., & Leavitt, K. (2020). Feeling activated and acting unethically: The influence of activated mood on unethical behavior to benefit a teammate. *Personnel Psychology*, *73*(1), 95–123. https://doi.org/10.1111/peps.12371

Watkins, T., & Umphress, E. E. (2020). Strong body, clear mind: Physical activity diminishes the effects of supervisor interpersonal injustice. *Personnel Psychology*, *73*(4), 641–667. https://doi.org/10.1111/peps.12384

Wellman, N., Tröster, C., Grimes, M., Roberson, Q., Rink, F., & Gruber, M. (2023). Publishing multimethod research in AMJ: A review and best-practice recommendations. *Academy of Management Journal*, *66*(4), 1007–1015. https://doi.org/10.5465/amj.2023.4004

Welsh, B. C., Braga, A. A., Bruinsma, G. J. N., & Bruinsma, G. (2013). *Experimental criminology: Prospects for advancing science and public policy*. Cambridge University Press.

Yam, K. C., Bigman, Y. E., Tang, P. M., Ilies, R., De Cremer, D., Soh, H., & Gray, K. (2020). Robots at work: People prefer—and forgive—service robots with perceived feelings. *Journal of Applied Psychology*, *106*(10), 1557–1572. https://doi.org/10.1037/apl0000834

Yeomans, M., Minson, J., Collins, H., Chen, F., & Gino, F. (2020). Conversational receptiveness: Improving engagement with opposing views. *Organizational Behavior and Human Decision Processes*, *160*, 131–148. https://doi.org/10.1016/j.obhdp.2020.03.011

Yu, W., Fei, J., Peng, G., & Bort, J. (2024). When a crisis hits: An examination of the impact of the global financial crisis and the COVID-19 pandemic on financing for women entrepreneurs. *Journal of Business Venturing*, *39*(2), 106379. https://doi.org/10.1016/j.jbusvent.2024.106379

Zizzo, D. J. (2010). Experimenter demand effects in economic experiments. *Experimental Economics*, *13*(1), 75–98. https://doi.org/10.1007/s10683-009-9230-z

Author Biographies

Kira F. Schabram (schabram@uw.edu) is an assistant professor and Evert McCabe Endowed Fellow at the Foster School of Business. She earned her doctorate from the University of British Columbia. Her research focuses on employees who seek to make the world a better place and factors that make such work more sustainable.

Christopher G. Myers (cmyers@jhu.edu) is an associate professor of management and organization at the Johns Hopkins University Carey Business School. He received his PhD from the University of Michigan Ross School of Business. His research explores processes of individual learning and development in knowledge-intensive work organizations.

Ashley E. Hardin (aehardin@wustl.edu) is an associate professor of organizational behavior at Washington University Olin Business School in St. Louis. She received her PhD from the University of Michigan's Ross School of Business. Her research examines interpersonal processes in organizations and how these processes are influenced by people's nonwork lives.