

Abusive Supervision: A Systematic Review and Fundamental Rethink

Thomas Fischer*
Geneva School of Economics and Management
University of Geneva
thomas.fischer@unige.ch

Amy Wei Tian
Curtin Business School
Curtin University
amy.tian@curtin.edu.au

Allan Lee
University of Exeter Business School
University of Exeter
Allan.Lee@exeter.ac.uk

David J. Hughes
Alliance Manchester Business School
The University of Manchester
david.hughes-4@manchester.ac.uk

*Corresponding author

Abusive Supervision: A Systematic Review and Fundamental Rethink

Within organizations, leaders are often powerful individuals who wield influence over many aspects of employees' working lives (e.g., Schyns & Schilling, 2013); and as the popular phrase goes: "with great power there must also come great responsibility." In other words, leaders should wield their power and influence carefully, refraining from destructive behaviors. Many leaders do so and are sources of great support for their employees. Others, however, are not. Holding positions of power can be corruptive (Bendahan, Zehnder, Pralong, & Antonakis, 2015), meaning that some leaders exploit and mistreat followers (e.g., Krasikova, Green, & LeBreton, 2013; Schmid, Pircher Verdorfer, & Peus, 2017), behaving more like proverbial villains than superheroes (Giurge, van Dijke, Zheng, & De Cremer, 2019). We focus on this dark or destructive side of leadership in this article.

The most studied form of destructive leadership is "abusive supervision", defined by Tepper (2000; p. 178) as a subjective evaluation resting on "subordinates' perceptions of the extent to which supervisors engage in the sustained display of hostile verbal and nonverbal behaviors, excluding physical contact". It is hard to think of many topics that could be more important to the wellbeing of workers and the effectiveness of organizations than leaders who are, or at least appear to be, abusive. Meta-analyses have reported that evaluations of abusive supervision are positively associated with myriad undesirable outcomes (e.g., Mackey, Frieder, Brees, & Martinko, 2017; Zhang & Liao, 2015), leading Tepper, Simon, and Park (2017; p. 145) to succinctly conclude that "the preponderance of work to date suggest[s] that abusive supervision undermines individual, unit, and organizational functioning".

Because abusive supervision is associated with so many consequential and deleterious outcomes, it is important that the research base is robust enough to provide nuanced and reliable findings that can be used to generate nuanced and reliable policy recommendations. For example, the research base should be able to provide us with evidence relating to questions such as: How much, and how frequently does leader 'wrongdoing' result in

evaluations of abuse? Should we invest time and money in selecting out abusive leaders and/or training leaders to be non-abusive? Should we focus less on leaders and more on increasing employee resilience or changing organizational culture? Research on abusive supervision touches upon these issues, but several conceptual and methodological concerns limit our ability to precisely assess the nature, prevalence, causes, effects, and buffers of abusive supervision. Thus, in order to provide a robustness check of existing research, we conducted a systematic review of abusive supervision research since Tepper's (2000) seminal work. In addition, we offer pragmatic guidance for future research.

As a commonly studied topic, several previous reviews exist which have tended to summarize findings (often using meta-analysis) about antecedents (e.g., Zhang & Bednall, 2016) or consequences of abusive supervision (e.g., Park et al., 2019; Zhang & Liao, 2015; Zhang, Liu, Xu, Yang, & Bednall, 2019), reviewed abusive supervision in specific contexts (e.g., Yu, Xu, Li, & Kong, 2020; Zhang & Liu, 2018), synthesized theoretical frameworks or identified 'gaps' within the literature (e.g., Mackey et al., 2017; Martinko, Harvey, Brees, & Mackey, 2013; Tepper et al., 2017). In contrast, our review combines systematic and narrative review techniques to provide a comprehensive overview of abusive supervision research and identify the major challenges that need to be overcome in order to advance the field.

The article contains six sections. First, we describe the article search strategy. Second, we review the way in which abusive supervision is conceptualized, highlight the challenges this poses, and provide guidance on how clarifying the conceptualization can help to progress research. Third, we examine the degree to which the dominant measurement tool (i.e., 15-item survey of Tepper, 2000) reflects common conceptualizations of abusive supervision and provides clear guidance on how we can improve assessment of abusive supervision. Fourth, we use systematic review techniques to provide a comprehensive catalogue of the correlates (i.e., antecedents, outcomes, mediators, and moderators) of abusive supervision and consider

how typical study designs can be improved in light of the ethical and pragmatic challenges facing abusive supervision research. Fifth, we report novel and simulated data on another unique challenge in this field, namely, that abusive supervision is a low base rate phenomenon, which raises numerous methodological concerns. Sixth, we provide a brief discussion and re-cap our recommendations for overcoming the challenges identified and for rethinking abusive supervision research.

We hope that the observations and recommendations made will help to reorient the field such that future findings will be robust enough to generate meaningful policy implications. Specifically, we see our review as a starting point for rethinking research on abusive supervision through careful consideration of the three interrelated questions that drive this article: *Are we conceptualizing abusive supervision correctly? Are we measuring abusive supervision in an optimal manner? Are current empirical studies well-suited to provide the evidence required to advance knowledge?* These questions are crucial, because if the primary studies are flawed, meta-analytic studies and narrative summaries will also be flawed (e.g., Antonakis, 2017) and might produce misleading conclusions and policy recommendations.

Before proceeding, we wish to clarify our use of terminology. Research examining abusive supervision variously uses leader, manager, supervisor, and boss to refer to the suggested perpetrator of abuse, and follower, employee, subordinate, and team member to refer to the suggested victim of abuse. All of these terms have their strengths and limitations. Still, to avoid perpetuating conceptual confusion by using different sets of terminology, we use the widely adopted nomenclature of leader and follower.

Search Strategy, Study Inclusion, and Coding Criteria

Search strategy

We used four search approaches to identify relevant empirical studies. First, we reviewed the reference lists of 11 qualitative and quantitative review articles of abusive

supervision (Mackey et al., 2017; Martinko et al., 2013; Park et al., 2019; Peng, Mitchell, & Schaubroeck, 2019; Tepper, 2007; Tepper et al., 2017; Yu et al., 2020; Zhang & Bednall, 2016; Zhang & Liu, 2018; Zhang & Liao, 2015; Zhang et al., 2019). Second, we searched for journal articles published in English in the scientific databases ProQuest and PsychInfo, up until October 2020, using specific keywords linked to abusive supervision, such as “abusive supervision”, “abusive supervisor”, “abusive leader”, “abusive manager”, and “abusive leadership”. Third, we searched Google Scholar to obtain studies that cited either Tepper’s (2000) seminal study, or Mitchell and Ambrose (2007). We chose Tepper (2000) because his 15-item measure of abusive supervision led to the inception of the field of abusive supervision research and is until today widely used in research, and we chose Mitchell and Ambrose (2007) as its shortened two-dimensional measure of abusive supervision (passive-aggressive and active-aggressive) has also been often adopted in the abusive supervision literature. Finally, we searched online for in press articles at journals that have published abusive supervision articles in the past (e.g., *Journal of Applied Psychology*, *Journal of Business Ethics*, *The Leadership Quarterly*, *Journal of Organizational Behavior*).

Selection criteria

We applied three criteria to select primary studies. First, abusive supervision is based on the definition proposed by Tepper (2000), and the measure of abusive supervision is based on Tepper (2000), either using all 15 items or a short version of Tepper’s scale (e.g., Mitchell & Ambrose, 2007). Second, we included only published journal articles because this ensured that data were subjected to a rigorous peer review process. Third, we included only empirical articles which contain a quantitative measure of abusive supervision.

Our keyword search in electronic databases returned 9,955 results, and a Google scholar search for studies citing Tepper (2000) or Mitchell and Ambrose (2007) returned 5,070 results. Duplicate articles were removed before the remaining titles, abstracts, and

method sections were screened for inclusion. Among the studies, there are 28 existing qualitative and quantitative reviews for the coding process (11 review articles on abusive supervision and 17 review articles on various destructive leadership styles). A total of 380 studies (with 490 independent samples) met the inclusion criteria and were used in our review. This number is notable as it reflects the steeply increasing volume of articles included in recent reviews of abusive supervision (e.g. Mackey et al., 2017, 112 studies with 140 independent samples; Martinko et al., 2013, 82 studies; Park et al., 2019, 79 studies; Zhang & Bednall, 2016, 74 studies; Zhang & Liao, 2015, 96 studies with 119 samples; Zhang & Liu, 2018, 74 studies; Yu et al., 2020, 36 studies with a focus in hospitality and tourism). For a summary of the journals that most frequently publish abusive supervision research please see Appendix Table A1.

Coding procedure

The initial coding scheme was developed based on the extant abusive supervision literature. Using this initial coding scheme, two co-authors independently coded 10 randomly selected studies. The coding was discussed between all co-authors, ensuing discrepancies and problems were resolved, and a refined coding scheme obtained. Based on this refined scheme, one of the two co-authors proceeded to code all the remaining studies. Subsequently, for assessing the average inter-coder percentage of agreement, the same two co-authors coded 15% randomly selected studies using the refined coding scheme. Cohen's Kappa estimated for this procedure was 96%. Any discrepancies were resolved through discussion.

Conceptualizing Abusive Supervision

How scholars think about (i.e., theorize and conceptualize) abusive supervision is a fundamental issue that shapes all other aspects of the field (e.g., measurement and study design). Tepper's (2000) initial definition, which remains the most popular, describes abusive

supervision as “subordinates' perceptions of the extent to which supervisors engage in the sustained display of hostile verbal and nonverbal behaviors, excluding physical contact” (p. 178). A definition is an “exact statement or description of the nature, scope, or meaning of something” (Oxford English Dictionary), and any construct definition should clearly articulate its boundaries such that it can be distinguished from related concepts, including closely related antecedents and outcomes (MacKenzie, 2003). In this sense, Tepper’s (2000) original definition is strong; stating clearly that abusive supervision does not refer to leader behaviors but to followers subjective evaluations of these behaviors, because, “[t]he same individual could view a supervisor's behavior as abusive in one context and as nonabusive in another context, and two subordinates could differ in their evaluations of the same supervisor's behavior” (p. 178).

Nevertheless, scholars have often treated abusive supervision scores as markers of leader behavior. For example, abusive supervision has been described as “nonphysical hostility perpetrated by managers against their subordinates” (Tepper, Henle, Lambert, Giacalone, & Duffy, 2008; p. 721), and “nonphysical hostility perpetrated by employees’ immediate supervisors” (Tepper, Moss, & Duffy, 2011; p. 279). Thus, there is confusion. Ratings of abusive supervision are variously considered to represent a behavioral or evaluative construct. Perhaps the label itself is a source of misunderstanding. Supervising, refers to being in charge, and thus, to leader behaviors rather than follower evaluations.

The difference between leader behaviors and follower evaluations might seem an overly critical, even minor, concern. However, although behaviors and evaluations of behaviors correlate, behaviors and evaluations are distinct concepts with distinct antecedents and consequences (see Banks, Fischer, Gooty, & Stock, 2020). What we refer to as ‘evaluations’ are sometimes called ‘perceptions’. Whereas perception describes the process of ‘becoming aware of’ an event, evaluation describes the process of ‘making judgements’ about

an event (c.f. Ross & Nisbett, 2011). Evaluation is therefore the more appropriate terminology because abusive supervision research requires followers to make subjective assessments of leader behaviors, rather than to simply acknowledge their existence. Thus, evaluations can differ from behaviors for more reasons than biased perceptions only. Consequently, conflating evaluations of leader behaviors with the behaviors themselves impedes both theoretical and empirical precision. Similar critiques have been levied previously and retorted. For example, Tepper et al. (2017; p. 126) argued:

“From the construct’s inception, those contributing rigorous scholarship to the domain (i.e., the kind of research that has found a home in the finest outlets in organizational behavior and industrial and organizational psychology) have assumed that subordinate reports of abusive supervision reflect a subjective evaluation and have never claimed that these reports capture supervisors’ behavior in an objective sense.”

Based on the current systematic review, it is our estimation that this conclusion does not hold, and research still conflates behaviors and evaluations. Even articles published after Tepper et al.’s (2017) review, including some by Tepper himself, do not meet Tepper et al.’s (2017) standards of rigorous scholarship. See the following examples:

- Yu, Duffy, and Tepper (2018; p. 2299) in the *Academy of Management Journal*:
“The leadership literature has identified a variety of behaviors that supervisory leaders employ [...]. The abusive supervision behavioral domain consists of hostile acts that include yelling at direct reports, and derogating, blaming, and ostracizing them”.
- Fiset, Robinson, and Saffie-Robertson (2019; p. 756) in the *European Journal of Work and Organizational Psychology*: “...abusive supervision, defined as

sustained displays of “hostile, verbal, and nonverbal behaviors, excluding physical contact” (Tepper, 2000, p. 178)”.

- Richard, Boncoeur, Chen, and Ford (2020; p. 549) in the *Journal of Business Ethics*: “Abusive supervision lacks an ethical component and is defined as the extent to which an immediate supervisor engages in consistent hostile verbal and non-verbal actions towards a subordinate”.

Our review reveals that a sizeable, and possibly growing, proportion of the literature draws on Tepper’s (2000) definition but conflates evaluations of abusive supervision and leader behaviors, which risks numerous erroneous conclusions. In principle, deviating from an originally proposed definition is not a problem. Newly proposed constructs often evolve over time and empirical work reflecting such evolutions can inform subsequent theoretical refinements (Hughes & Evans, 2018; Shaffer, DeGeest, & Li, 2016). Thus, the fact that the domain of abusive supervision is now conceptualized in various ways is not necessarily problematic, particularly because understanding leaders’ abusive behaviors and followers’ evaluations of such behaviors are both important. However, conflating leader behaviors with follower evaluations, both of which will have different antecedents and outcomes, serves only to obfuscate our understanding of these important phenomena. Not least because confused conceptualizations produce imprecise measurement and empirical estimates, which we discuss next.

Measuring Abusive Supervision

Our review focused on articles that used Tepper’s (2000) 15-item measure of abusive supervision or derivatives (e.g., Mitchell & Ambrose, 2007). For a detailed review of the specific questionnaires used see Appendix Table A2. Most studies collected a single assessment of abusive supervision via employee ratings of their leader (N = 452) or leader self-ratings (N = 19).

To examine the quality of measurement within the field, we conducted an item-level review of Tepper's (2000) abusive supervision scale. Unfortunately, our review revealed that regardless of who completes the questionnaire, important conceptual and psychometric limitations undermine the accuracy and precision of the measurement (Hughes, 2018). It is important to take an item-level perspective because all measurement in a survey study occurs between a respondent reading an item and responding to it. Everything after that point is a test or reflection of some theoretical or statistical model (Hughes, 2018). Table 1 summarizes the findings of our item-level review, and we now examine the shortcomings in measuring abusive supervision in more detail.

Insert Table 1 about here

First, measures clearly reflect the conceptual ambiguities discussed above: the conflation of leader behavior and follower evaluation. The first item of Tepper's (2000) scale, "Ridicules me" (p. 189), serves as illustration. The item requires a subjective evaluation to judge if a concrete behavior can be classified as ridiculing; merely perceiving and acknowledging the existence of single behaviors is not enough. For instance, when a leader makes a joke about a person, some team members might regard the joke as loosening the atmosphere whereas others regard the joke as derogatory or ridiculing. Hence, in the spirit of Tepper (2000), "[t]he same individual could view a supervisor's behavior [here: a leader's joke] as abusive in one context and as nonabusive in another context, and two subordinates could differ in their evaluations of the same supervisor's behavior [here: joking]" (p. 178). Thus, "Ridicules me" requires a subjective follower evaluation. Still, the item rests on behavioral language focusing on the leader as the actor and not on the follower's evaluation of the leader's acts. Such a conflation of leader behaviors and follower evaluations characterizes also other items of Tepper's scale. For instance, "Invades my privacy" and "Is

rude to me” (p. 190) depend not only on leader behaviors but on context- and person-specific evaluation standards for privacy and ruthlessness too.

Second, and relatedly, some items are implicitly double-barreled. Contrary to the definition of abusive supervision, all items adopt behavioral language that suggests the scale is assessing a supervisor behavior. At the same time, however, some of these items are not purely descriptive but also evaluative. For example, the item “Invades my privacy” refers to the behavior of a leader. However, whether a behavior is evaluated as invading the own privacy or, for instance, displaying concern for personal matters, is ultimately in the ‘eye of the beholder’. Thus, the implicitly double-barreled formulation makes the item conflated; and other items of Tepper’s scale share the same problem (see Table 1).

Third, some items are explicitly double-barreled. For example, the item “Blames me to save himself/herself embarrassment” assesses the act of blaming and the supposed motive of self-preservation underlying such an act. In such instances, it would be unclear which aspect an item response reflects—that is, whether the response refers to the act or the supposed underlying motive—which is another source of inaccurate measurement (Borsboom, Mellenbergh, & van Heerden, 2004; Hughes, 2018).

Fourth, the content of some items is ambiguous, because it is unclear if these items are to be understood literally or figuratively. For example, when responding to the item “Tells me I’m incompetent”, it is unclear whether the item refers only to instances when the leader literally used these words, or if it would be sufficient if the follower ‘felt’ criticized by the leader, even if the leader used different language. Such ambiguous items make it necessary for the follower to infer the supposed meaning of the items, and it is likely that different followers come to different inferences. Hence, de facto followers respond to different items and as such to different constructs—another source of mismeasurement.

Fifth, some items refer to supervisor behaviors that are negative but not necessarily abusive. For example, “Breaks promises he/she makes” is certainly unreliable but not abusive if the act is infrequent, explained, and justified. Like double-barreled items, those items that assess a related but distinct construct capture ‘construct irrelevant’ variance, which contaminates measurement and prevents proper causal inferences drawn from using the scales (Borsboom et al., 2004). Problematically, however, double-barreled, figurative, and construct-unrelated items lead to mixing the measurement of abusive supervisor behaviors with other concepts. Such confusion, then, can have substantial detrimental effects upon the accuracy of the evidence-base and the resulting policy recommendations of abusive supervision research. Taken together, avoiding the five types of mismeasurement is a major challenge for moving forward research on abusive supervision.

Empirical Research on Abusive Supervision

Table 2 contains summary statistics regarding the study design characteristics of empirical research on abusive supervision. Although the adequacy and appropriateness of any study design must be evaluated in regard to the study goals and context, summary statistics can provide us with an overview of abusive supervision research. In brief, 424 utilized a sample of working adults, 52 used student samples and 5 used a mix of students and employees. Most studies (N = 446) used questionnaires to measure the amount of naturally occurring abusive supervision and its correlates, with only 44 utilizing experimental methods. Those that used survey methods most frequently employed a cross-sectional design (N = 247). However, a non-trivial number of studies measured variables at different points of time (i.e., used time lagged designs; N = 176). In addition, 23 studies used a longitudinal design, of which 13 utilized diary designs examining within-person variation over time (ranging from 5-21 days; e.g., Bormann, 2017; Vogel & Mitchell, 2017)

[Insert Table 2 about here](#)

Thus, the typical abusive supervision study is survey-based using convenience or snowball samples of participants drawn from a range of organizations. In total, 91% of studies utilized such a design and many sought to examine causal hypotheses (e.g., employees' evaluations of abusive supervision cause employee depression), which is problematic because such designs are susceptible to endogeneity. The concept of endogeneity refers to several well-known, but rarely addressed, model misspecifications that often bias parameter estimates (Antonakis, Bendahan, Jacquart, & Lalive, 2010, 2014). As Hughes, Lee, Tian, Newman, and Legood (2018; p. 558) describe it:

“...endogeneity refers to an instance when a predictor variable (whether classed as predictor, mediator, or moderator) is correlated with the error term of the outcome variable (see Antonakis et al., 2010, 2014 for details). In other words, an endogenous predictor is related to the measured outcome variable in two or more ways, usually in the way theorized (e.g., as a meaningful cause), but also in some unanticipated way(s) (e.g., common method bias, reciprocal effects, relationship with a common cause).”

The consequence of endogeneity bias can be substantial and make it difficult to draw causal conclusions because it is impossible to know whether and to what degree the estimate of interest represents the theorized relationship (i.e., the causal effect) or the unanticipated relationship (i.e., endogeneity bias). It is important to keep in mind that, because most studies use designs are heavily influenced by endogeneity, the variables we discuss in the remainder of this section, whether proposed as antecedents, mediators, moderators, or outcomes, are best considered to be correlates. This is because regardless of theoretical rationale, endogeneity biases prevent us making firm conclusions about the direction and magnitude of the effects observed.

Concerns of endogeneity have been frequently discussed over the last few years and numerous articles already provide clear discussions in technical (e.g., Antonakis et al., 2010)

and non-technical terms (e.g., Martin, Hughes, Epitropaki, & Thomas, 2020). These articles highlight the importance of using multiple study designs to build a cumulative body of evidence robust to different types of endogeneity bias, argue for the forgotten importance of experimental designs, and discuss statistical solutions such as the use of instrumental variables. Many, if not all, of the suggestions within these articles apply to abusive supervision research, too. Hence, we do not reiterate these points. However, we note very strongly that future research must do everything possible to address concerns of endogeneity. In addition, we provide specific examples of how this might be achieved following the systematic review of the correlates of abusive supervision.

Outcomes of abusive supervision

Numerous outcomes have been examined in relation to abusive supervision (see Appendix Figure A1 for additional details of every outcome explored and the range of correlations reported). Because the number of outcome variables is so large, we have grouped them into broad categories to facilitate comprehension and discussion (see Figure 1). For example, many studies explore the link between abusive supervision and follower behaviors. The outcomes broadly cover desirable (e.g., organizational citizenship behavior, task performance, and creativity/innovation) and undesirable behaviors (e.g., employee deviance, counterproductive work behavior and hostility/aggression). Another common outcome category relates to follower wellbeing, such as burnout, anxiety, depression, stress/distress, and negative emotional responses such as anger or frustration.

Insert Figure 1 about here

When looking at the patterns of these associations, the picture is rather uniform: abusive supervision is ‘bad news’. That is to say, researchers report that employees’ evaluations of abusive supervision are negatively correlated with desirable outcomes such as follower task performance (Peng, Schaubroeck, & Li, 2014), organizational citizenship

behavior (Zellars, Tepper, & Duffy, 2002), job attitudes (Aryee, Chen, Sun, & Debrah, 2007), and well-being (Lin, Wang, & Chen, 2013). Employees' evaluations of abusive supervision are also positively correlated with undesirable outcomes, such as deviant behavior (e.g., Mitchell & Ambrose, 2007), turnover intentions (Haar, de Fluiter, & Brougham, 2016), and counterproductive work behavior (Wei & Si, 2013). Such findings have been well documented in previous meta-analyses and reviews.

However, the range of associations between abusive supervision and many outcomes is surprisingly large (see also Appendix Figure A1). For instance, the correlation between abusive supervision and employee deviance ranges from $-.15$ (Sungu, Hu, & Weng, 2020) to $.74$ (Tröster & Van Quaquebeke, 2020). Similarly, the correlation between abusive supervision and counterproductive work behavior ranges from $-.26$ (Ahmad, Athar, Azam, Hamstra, & Hanif, 2019) to $.71$ (Goswami, Park, & Beehr, 2020). The substantial variability in the magnitude of associations between abusive supervision and outcomes could be a result of differences in study design or the use of designs that are susceptible to endogeneity biases (Hughes et al., 2018). Alternatively, the substantial variability could indicate that the effects of abusive supervision are contingent on certain boundary conditions or moderators.

Moderators of the consequences of abusive supervision

Figure 2 shows that many moderating variables have been proposed and tested, which makes a concise summary impossible, especially because most moderators have been tested only once, hence did not yet prove robust to repeated testing. Interestingly, of the moderating variables that have been studied multiple times, many have shown inconsistent effects.

Insert Figure 2 about here

For example, some studies demonstrate that power distance orientation—a cultural value indicating the extent to which individuals accept the unequal distribution of power in organizations (Carl, Gupta, & Javidan, 2004; Hofstede, 1980)—weakens the negative effects

of abusive supervision (Lin et al., 2013; Low, Sambasivan, & Ho, 2019; Peltokorpi & Ramaswami, 2019), whereas others find that the effects of abusive supervision are exacerbated by higher levels of power distance (Lian, Ferris, & Brown, 2012; Richard et al., 2020). Similarly, meaning at work has been found to both attenuate (Pradhan & Jena, 2017) and exacerbate (Harris, Kacmar, & Zivnuska, 2007) the negative relationship between abusive supervision and employee outcomes. The moderating role of social support within the workplace also offers contradictory findings. For example, research suggests that abusive supervision coupled with supervisory support (e.g., in the form of LMX relationship quality or perceived supervisor support) exacerbates the negative effects of abusive supervision (Hobman, Restubog, Bordia, & Tang, 2009; Lian et al., 2012). However, perceived support from co-workers sometimes has been found to either attenuate (Hobman et al., 2009; Liao & Liu, 2015; Pradhan & Jena, 2017) or accentuate (Caesens, Nguyen, & Stinglhamber, 2019; Wu & Hu, 2009) the negative effects of abusive supervision.

Mediators of abusive supervision

Because leadership is typically viewed as a process whereby leader behaviors (e.g., abusive supervision) influence distal outcomes (e.g., employee performance) through more proximate mediating variables (e.g., follower motivation: Fischer, Dietz, & Antonakis, 2017), many studies within our review examined mediating mechanisms. Recent reviews of the leadership literature (Hughes et al., 2018; Inceoglu, Thomas, Chu, Plans, & Gerbasi, 2018) have identified five categories of mediators of leader behavior: i) motivational, ii) relational, iii) cognitive/social-cognitive, iv) affective, or v) identification-based. We draw upon these frameworks here, and as evident in Figure 1, the myriad of proposed mediators largely fit into one of these categories (Appendix Figure A2 provides additional details of the most measured variables within each mediator category).

Many of the mediator models tested are posted as tests of causal theories of workplace behavior, including theories of social learning, displaced aggression, social exchange, ego depletion, and affective events (see review by Tepper et al., 2017). Unfortunately, three limitations reduce the ability of these studies to inform theoretical advance. First, the problem of endogeneity which we discussed earlier. Second, because studies typically examine a single mediator, they cannot provide comparative tests of different mediators from different categories or theoretical perspectives (Fischer et al., 2017). As highlighted in previous leadership reviews (e.g., Hughes et al., 2018), there is much conceptual and empirical overlap between many of the mediators examined (e.g., LMX and interpersonal justice), and as such, it is likely that the literature suffers from construct proliferation and redundancy (Shaffer et al., 2016). Relatedly, few studies have examined conceptually dissimilar mediators concurrently (i.e., those from different mediator categories), meaning that we cannot say, for example, whether relational or affective mechanisms are more powerful, and whether their effects are additive or not (see Fischer et al., 2017). Third, the same mediating mechanisms can be explained using various theories, some of which are very different. For example, followers' negative emotional reactions to abusive supervision have been taken as indicative of support for affective event theory (Avey, Wu, & Holley, 2015), conservation of resources (Seo & Chung, 2019), the multi-motive model of reactions to interpersonal threat (Simon, Hurst, Kelley, & Judge, 2015), the appraisal theory of discrete emotions (Peng, Schaubroeck, Chong, & Li, 2019), the emotional process theory of abusive supervision (Tröster & Van Quaquebeke, 2020), the social functional view of emotions (Yu & Duffy, 2020), and an approach–avoidance framework (Ferris, Yan, Lim, Chen, & Fatimah, 2016). If the same evidence can be used to support multiple distinct theories, then the theories are not specified clearly enough and our empirical tests of them are weak to allow myriad interpretations.

Looking forward, we should aim to generate specific theories that posit high-risk and falsifiable propositions that we can test rigorously (Meehl, 1990). Future research is needed that a) explores the mechanisms of abusive supervision in a more systematic way, and b) uses a study design that is optimal for accurately assessing the indirect relationship between abusive supervision and outcomes. Thus, although the one-mediator-per-article model and diversity of theories used make synthesis difficult, our comprehensive but parsimonious taxonomy of mediators can help to guide future research. For example, using this taxonomy researchers can easily identify mediating variables that operate through the same or different broad mechanisms and posit competing tests (Fischer et al., 2017; Hughes et al., 2018).

Antecedents of abusive supervision

Unlike many other leadership variables, the potential antecedents of abusive supervision have received substantial and growing attention (N = 186), and scholars start acknowledging the multidimensional nature of employee evaluations of abusive supervision (Wang, Van Iddekinge, Zhang, & Bishoff, 2019). Put simply, whenever an employee completes an abusive supervision scale, the responses reflect at least three meaningful sources of variance: a) leader characteristics and behaviors, b) employee characteristics and behaviors, or c) features of the environment (see also Appendix Figures A3).

Leader characteristics and behaviors. This source of variation reflects the fact that certain leaders are more likely to behave in an abusive manner than others and thus are rated by followers as higher in abusive supervision. Commonly studied leader characteristics include the HEXACO personality dimensions (Breevaart & de Vries, 2017), dark personality traits (i.e., Machiavellianism, psychopathy, narcissism; e.g., Kiazad, Restubog, Zagenczyk, Kiewitz, & Tang, 2010; Lyons, Moorman, & Mercado, 2019), attachment styles (e.g., Robertson, Dionisi, & Barling, 2018), stress/distress (e.g., Li, Wang, Yang, & Liu, 2016), and negative emotions (e.g., Gabler, Nagy, & Hill, 2014).

Follower characteristics. Like leader characteristics, follower characteristics also contribute to evaluations of abusive supervision. The logic being that followers will differ in the degree to which they (i) regard specific behaviors as abusive and/or (ii) elicit abusive behavior from their supervisors. For example, prominent leadership theories suggest that leaders and followers have unique relationships that are influenced, in part, by follower characteristics. For instance, Henle and Gross (2014) found that employees' self-ratings of emotional stability and conscientiousness negatively predicted self-reports of abusive supervision. Drawing on victim precipitation theory, they argued that the expressions of low levels of conscientiousness and emotional stability are often irritating, create tension and conflict, or violate social norms, which is likely to elicit abusive leader behavior (obviously, such an empirical finding is no moral justification of abuse). More generally and going beyond abusive supervision, Guntner, Klonek, Lehmann-Willenbrock, and Kauffeld (2020) show how the behavior of followers influences the behavior of leaders.

Follower personality might also shape the way that employees assess their environment, meaning they might have different thresholds for what constitutes abuse, or indeed be more or less aware of abuse within their environment. Such personality-driven variance in evaluations of abusive supervision are, to some degree, independent of leader behavior (Wang et al., 2019). For instance, scholars argue that some employees may be predisposed to attribute negative events to external factors and will therefore be more likely to perceive their supervisors as abusive (e.g., Zhang & Bednall, 2016). In line with this attribution argument, employees high in negative affectivity or neuroticism may selectively recall more negative events than employees low on these traits, which may partly explain the correlations between evaluations of abusive supervision and employee negative affectivity and neuroticism (e.g., Wang et al., 2019; Zhang & Bednall, 2016).

Features of the environment. Several studies have examined situational characteristics as predictors of abusive supervision (see Appendix Figure A4). For example, some authors have argued that leaders may engage in abusive supervision as a form of displaced aggression when they experience difficulties inside or outside of the workplace. That is, the life situation outside of work might be predictive of abusive supervision behavior. For instance, studies have found that a leader's evaluation of abuse from their supervisor is related to evaluations of abusive supervision by their own subordinates (e.g., Mawritz, Mayer, Hoobler, Wayne, & Marinova, 2012).

Experimental research on abusive supervision

Most studies reviewed used survey-based study designs without applying instrumental variable estimation to account for endogeneity (Antonakis et al., 2010). Hence, these studies are ill-equipped to provide stable estimates of causal antecedents and effects of abusive supervision (see e.g., Fischer et al., 2017; Hughes et al., 2018). Promisingly, however, 35 articles with 44 independent samples used experimental designs, which are the best method for overcoming endogeneity bias (Lonati, Quiroga, Zehnder, & Antonakis, 2018; Podsakoff & Podsakoff, 2019). By randomly selecting participants from the population and randomly assigning them to different experimental groups, researchers can reduce several threats to causal inference and be more confident that observed differences in outcomes are due to the manipulation. However, even when complete randomization is not possible, experimental designs can still be valuable (Martin et al., 2020). Table 3 provides an overview of experimental articles and groups them according to whether participants were 'leaders' or 'followers' and contains an overview of the experimental design, participant sample, method of manipulation, participant recruitment strategy, and where the experiment took place.

[Insert Table 3 about here](#)

Although experiments are the best method for establishing causality, our review of existing experiments suggests that experimentation in the case of abusive supervision provides a number of unique challenges. Below we focus on three particularly important issues that future research must address.

Manipulate the experience of abusive supervision. In order to conduct a true experimental study we must manipulate the independent variable. In this literature, that would often require manipulating abusive supervision, such that some participants are willfully subject to abusive behavior. This is obviously a huge problem from an ethical perspective. Nevertheless, existing research has identified a number of experimental manipulations that can be used as inspiration for future research (see Table 3).

Vignette-style scenarios have proved particularly popular. For example, Tröster and Van Quaquebeke (2020) randomly assigned participants to conditions using a 2 (abusive supervision: low vs. high) by 2 (LMX: high vs. low) between-subjects factorial design. Participants read vignettes depicting their supervisor as either abusive or non-abusive, and the leader-follower relationship as either good or bad. Following the scenarios, participants were asked to rate whether they thought that they had done something to threaten their relationship with the leader (self-blame) and if they would feel guilty in such a situation. Perhaps a more ‘realistic’ adaptation of such manipulations involves asking participants to recall past experiences. When assessing the effects of abusive supervision on followers, we can ask them to recall a specific instance when a supervisor has acted abusively, neutrally, or positively before assessing some cognitive or affective outcome (e.g., retaliation on a digital voodoo doll, Liang et al., 2018). When assessing the triggers of abusive supervision, we can ask them to recall some aspect of a past project (e.g., task difficulty; Collins & Jackson, 2015) or employee performance (e.g., poor vs. strong performance) before asking them to carry out a realistic activity (e.g., writing an e-mail to the imagined employee; Ju et al., 2019).

The benefit of recall or vignette-type designs is that they profit from the causal precision of experimental designs without breaching ethical standards of research. However, studies reliant on memory recall, hypothetical choices, and self-reported or symbolic outcomes provide limited information about the magnitude of the effects of abusive supervision in ‘real life’. In addition, hypothetical studies are prone to demand effects and social desirability, especially if combined with manipulation checks (Lonati et al., 2018). Furthermore, a few studies use priming (see Table 3), which also entails limitations for causal interpretability (c.f. Sturm & Antonakis, 2015).

As we noted earlier, for ensuring causal rigor it would be most effective to directly manipulate abusive behaviors. Only 4 out of 35 articles do so, and all of them involve confederates (Mitchell & Ambrose, 2012; Qin, Huang, Johnson, Hu, & Ju, 2018; Tu, Bono, Shum, & LaMontagne, 2018; Yu & Duffy, 2020), which violates the journal’s ethical standards for experimentation. Thus, although existing experiments have made some progress in testing the antecedents and effects of abusive supervision in a causal manner, stronger and ethically defensible experimental designs are required. We acknowledge that designing robust and ethical experiments of abusive supervision is difficult and that we must be creative in our study designs. Nevertheless, below are some potentially fruitful suggestions.

First, scholars can consider a method informally referred to as ‘inverting the treatment’ (see Eden, 2020). Inverting the treatment in an abusive supervision study might, for example, entail an experimental condition which actively removes abuse from participants but allows abusive supervision to naturally occur within the control group. For example, imagine time pressure was demonstrated to be a causal antecedent of abusive supervision and we had participant leaders of participant work teams working under time pressure. In condition one, leaders would be trained how to act compassionately towards followers under time pressure, thus reducing abusive supervision. In condition two, leaders would receive no

training. Condition two would, assuming time pressure is a reliable causal antecedent, see more leaders behave abusively, but this would not have been directly introduced by the researcher. We could then compare the effects of condition one (actively removing abuse) with condition two (leaders who ‘naturally’ behave abusively).

Second, scholars can make use of natural experiments, whereby variation in abusive supervision occurs naturally. Typical examples in the literature are changes in law or regulation, as well as crises like pandemics (Sieweke & Santoni, 2020). More pertinent to the study of abusive supervision, however, could be for instance the designation of a new leader. Some organizations and industries (e.g., academia, politics, unions, consulting etc.) design leader turnover into their operating models, by having fixed-term appointments, for example. Such organizations or industries provide opportunities for natural experiments (i.e., exogenous source of leader turnover; and some of the leaders will behave abusively more often than others). Combining the natural experiment with a regression discontinuity design, to control for the lack of random assignment, could be a powerful route to studying abusive supervision in the future (Antonakis et al., 2010).

Although we see ‘inverting the treatment’ and natural experiments as fruitful research designs, implementing these designs might be challenging because abusive supervision is a low base rate phenomenon (i.e., in its current conceptualization, abusive supervision occurs relatively infrequently or at relatively low levels), a challenge that we discuss later. Nevertheless, these two designs are promising options that used in conjunction with studies that use other forms of manipulation can help yield a stronger evidence base.

Use active control conditions. Of the 37 experiments that manipulated abusive supervision (typically using vignettes), 23 used a passive control group in which the leader was non-abusive (e.g., Tröster & Van Quaquebeke, 2020). A further 6 experiments compared the effects of an abusive leader to positive leadership styles: supportive/positive (Burton &

Hoobler, 2006; Burton, Hoobler, & Kernan, 2011), transformational (Hurst, Simon, Jung, & Pirouz, 2019), positive (Burton & Barber, 2019; Lopes, Kamau, & Jaspal, 2019), and jovial leadership (Liang et al., 2016). Using passive control groups (i.e., non-abuse) or positive control groups (e.g., positive leadership) creates an issue of demand effects and unfair comparisons (Lonati et al., 2018; Martin et al., 2020), because it is not clear whether the effects stem from abusive supervision, positive leadership, or a mixture of the two. One way to deal with this issue is to provide a manipulation of different levels of abusive supervision. Schyns, Felfe, and Schilling (2018), for instance, examine four different conditions: constructive leadership, laissez-faire leadership, mild abuse, and strong abuse. Although the study has the limitations of hypothetical vignettes (see previous discussion), the article serves as an example of best practice in manipulating multiple levels of abuse.

Assess substantive consequences and incentivize. 21 of the 29 experimental samples exploring follower responses to abusive supervision assessed self-reported attitudes, emotions or behavioral intentions. In addition, in studies that explored antecedents of abusive supervision, 6 of 10 actually assessed intentions to engage in abusive supervision, rather than enacting abusive supervision. Because hypothetical situations and self-reported outcomes have no real-world consequences, they are particularly prone to demand effects, cheap talk, and socially desirable responses (Antonakis, 2017). However, some studies have assessed more realistic outcomes. For instance, Greenbaum, Hill, Mawritz, and Quade (2017) exposed participants to an abusive vs. non-abusive leader, and asked participants to complete self-graded competitive anagram tasks. Participants were rewarded depending upon their self-reported performance levels. Following completion of the task, the true scores of the anagram tasks were calculated and the discrepancy between the true score and self-reported score (i.e., artificially inflated self-report scores) was taken as a real-world marker of unethical behavior. Another example of a slightly more realistic outcome than self-report can be found in Liang et

al. (2018) who had participants recall a specific act of abusive supervision and enact revenge on an electronic ‘voodoo doll’ which represented their leader. Other examples of substantive consequences we found were in studies by Fiset et al. (2019); Mitchell and Ambrose (2012); Shao, Li, and Mawritz (2018); and Yu and Duffy (2020).

To summarize, non-experimental study designs of abusive supervision might suffer from endogeneity. However, while experiments appear to be gaining momentum in this area, used experimental designs suffer from various limitations, such as being merely hypothetical vignette studies. Stronger experimental designs that are more ecologically valid are still required if we wish to make causal claims about real-world outcomes. We offered exemplary studies from the literature base and some novel suggestions (e.g., ‘inverting the treatment and natural experiments). At this point, we want to acknowledge that conducting abusive supervision studies that are causally identified, ecologically ‘valid’, and ethically sound represents a very difficult challenge. In this regard, we agree with Martin et al. (2020) that demanding all abusive supervision studies to “...reach the highest standards of experimental design is likely to prove counterproductive. Indeed, we believe that all well-designed research that takes into account the ability of the design to determine causality, and tempers conclusions accordingly, can make significant contributions to the ‘cumulative body of research’ (Shaver, 2020)”. Nevertheless, we need many more studies that take causality seriously and do everything possible to limit the effects of endogeneity.

Low Base Rates of Reported Abusive Supervision: A Special Challenge for Empirical Research

A large-scale U.S. survey found that 13% of employees experienced psychological aggression at least once a week (Schat, Frone, & Kelloway, 2006), and abusive supervision studies typically find “a low average mean ($\mu = 1.78, \sigma = 0.46$) [of abusive supervision]” (Mackey et al., 2017; p. 13) meaning that “exposure to abusive supervision is rare” (Tepper et

al., 2017; p. 125). Our systematic review revealed a similar pattern. In nearly all reviewed samples, the across-item average of abusive supervision, as measured by Tepper's scale (2000) or one of its adaptations, indicated that abusive behaviors never or rarely occurred.

From a humanitarian point of view, it is relieving that abusive supervision is rare. From a scientific point of view, however, the relative absence of reported abusive supervision poses a statistical challenge. When calculating correlations, regressions, SEMs, or most other models, we are concerned with variance and co-variance, and usually, we use variance in one variable (e.g., abusive supervision) to explain variance in another (e.g., follower wellbeing). Obviously, such models rest on the assumption that there is substantive variance in abusive supervision, and when this assumption is violated, the models will often return inaccurate parameter estimates (Angrist & Pischke, 2014; Hahn, 1977). Thus, the fact that evaluations of abusive supervision are rare raises a critical question: Is it proper to extrapolate the relationship between abusive supervision and negative outcomes observed at zero levels or low levels of abuse to intermediary and high levels of abuse? That is, given that the empirical data is largely restricted in its range to observations of no or rare abuse, is it proper to draw conclusions about the effects of intermediary or high levels of abuse?

Additional data on the base rate of abusive supervision

Before addressing the above question, we gathered and analyzed additional data to assess whether abusive supervision really has a low base rate. Via ProlificAcademic¹, we collected a sample of 1,530 working adults from a variety of industries, who each completed Tepper's (2000) 15-item measure of abusive supervision. Such data allowed us to check the

¹ Prolific (<http://www.prolific.ac>) is an online crowdsourcing data acquisition platform explicitly designed for participant recruitment by the scientific community. The platform allows researchers to recruit participants using pre-screening requirements. For our study screening criteria were being a fluent in the English language, being employed (either full or part time), having a direct supervisor. As the study was conducted during the Covid pandemic, we also stipulated that participants were still working. 26 participants were excluded for failing attention checks or not finishing the survey.

relative frequency of reported abusive supervision for each individual item and not only for the scale overall. This is important, because each item reflects the frequency with which abusive leader behaviors are evaluated as such.

The results are striking and consistent with past research (Mackey et al., 2017; Schat, Frone, & Kelloway, 2006; Tepper et al., 2017). Across all 15 items, the average level of reported abusive supervision across respondents was closer to “never” than “occasional” (see Figure 3). Even the most common abusive supervision behaviors (e.g., the 15th item, “My supervisor lies to me”) are still rare. In addition, those items with comparatively higher scores (e.g., “lies to me”) do not measure abuse directly, but arguably assess other forms of negative behaviors (compare Figure 3 and Table 1). Moreover, occasional abuse is observed for only about 10% of the cases; and even when combined, instances of moderately frequent or very frequent abuse are limited to less than 10% of the sample. Therefore, we can conclude that abusive supervision is a low base rate phenomenon, and research data mostly captures variation between no abuse and rare abuse.

Insert Figure 3 about here

Demonstrating the risk to extrapolate from low base rate phenomena

Tepper et al. (2017) say that the “preponderance of work to date suggests that abusive supervision undermines individual, unit, and organizational functioning” (p. 145). Such a claim, however, is based on research with zero or low levels of abusive supervision; previous work simply shows that low levels of abuse coincide with low levels of negative and high levels of positive outcomes. Although it is a plausible speculation that high levels of abusive supervision lead to low levels of positive outcomes, such a claim would be speculative. Further, because data with high levels of abusive supervision has rarely been examined we do not know whether and to what degree high levels of abusive supervision affect followers.

To illustrate the problem of low base rates, we simulated 1,530 observations, which is identical in size to our ProlificAcademic sample. For each observation we generated values of abusive supervision and an exemplary outcome variable, follower job performance. We simulated abusive supervision with a skewed distribution to reflect that abusive supervision is a low base rate phenomenon (for details, see Stata Code A1). Thereby, we calibrated the distribution such that the generated sample had a mean and standard deviation that was similar to previous samples ($\mu = 1.84$, $\sigma = .45$, compared with $\mu = 1.78$, $\sigma = 0.46$ for the studies of Mackey et al.'s meta-analysis, and $\mu = 1.49$, $\sigma = .61$ for our own data set).

In addition, we specified follower job performance by a constant term, a random (error) term, and a negative causal impact of abusive supervision at low levels of abuse. Deliberately, however, we did not specify a causal impact of intermediate to high levels of abusive supervision on follower job performance. The reason is that the very point of the simulated dataset is to demonstrate the risk of improperly extrapolating causal relationships from low to high levels of abusive supervision. Indeed, our data showed that an otherwise properly specified regression model predicted a negative causal impact of abusive supervision on follower job performance—which is correct for low, but not intermediate or high levels of abusive supervision. Figure 4 illustrates this point (see also Stata Code A1 in the appendix). Interested readers can re-run these analyses and/or modify certain parameters to examine variations of the model.

Insert Figure 4 about here

Therefore, a linear extrapolation of the effects observed at low levels of abuse to high levels of abuse is not empirically supported but rests on a set of assumptions. For instance, typical regression models implicitly assume that each case of abuse leads to a uniform decrease of positive outcomes. However, there are different reasons why this assumption might not hold (see Fischer et al., 2017). For instance, high levels of abuse might lead to the

development of psychological coping mechanisms among victims, which could weaken the effect of repeated abuse. Alternatively, high levels of abusive supervision might lead to much stronger negative effects, because at lower levels followers are able to recover more quickly (c.f. Ross & Nisbett, 2011). Likewise, there might be a dynamic link between the negative outcomes and the level of abuse, such that negative outcomes like low performance lead to even higher levels of abuse. Such bi-directionality would violate assumptions of linear extrapolation too (Fischer et al., 2017). Any of the listed alternatives are possible, and that we do not know which is most likely, presents a glaring challenge for abusive supervision research.

For qualifying our arguments: The simulated data does not affirm or prove a causal directionality or effect size between abusive supervision and follower performance, but is a stylized example. As such, the simulated data simply highlights that—as commonly done—linearly extrapolating the effect of abusive supervision beyond the range for which there is a considerable amount of observations in the data is not empirically grounded. Extrapolations beyond this variance range entail the implicit assumption that there is a stable abusive supervision-performance link, and when this assumption does not hold, causal inference is flawed (Fischer et al., 2017). If scholars wish to study the effects of high levels of abusive supervision, then they need to collect data which contains a broad range of such observations. Ethically defensible experimental manipulations that actively create variation in levels of abusive supervision, like those we discussed earlier, offer one solution to this challenge.

Discussion and Recommendations for Rethinking Abusive Supervision Research

We reviewed 490 independent samples measuring abusive supervision, and most of them show that the construct is associated with a range of deleterious outcomes for employees, teams, and organizations. However, our review also identified a number of challenges facing the conceptualization, measurement, and empirical study of abusive

supervision. Throughout our review, we provided achievable solutions to address these challenges. Now, we finish by summarizing and synthesizing the proposed solutions.

Rethinking conceptualization and measurement

Our review uncovered that the current conceptualization and measurement of abusive supervision conflates leader behaviors with followers' evaluations of those behaviors. Future research must address this conflation in two ways: (i) develop an unambiguously behavioral and (ii) an unambiguously evaluative construct.

First, scholars need to develop clear behavioral measures of abusive supervision. As evidenced by hundreds of articles that use Tepper's (2000) abusive supervision construct as a behavioral predictor variable for follower outcomes, many scholars seek to study the impact of abusive supervisors on followers. However, there is increasing doubt that questionnaire measures really measure leadership behaviors and not, for instance, (dis)liking or affect towards the leader (e.g., Martinko, Mackey, et al., 2018; Yammarino, Cheong, Kim, & Tsai, 2020). Thus, we concur with a recent call for articles to assess abusive supervision with non-questionnaire-based techniques to increase objectivity in measurement (Fischer, Hambrick, Sajons, & van Quaquebeke, 2020). There are multiple possible ways of doing so. One way is (field) experimentation. Since the 1930s, scholars like Lewin, Lippitt, and White (1939) experimentally manipulated leadership, and Eden (2020) has recently taken stock of and given advice for field experimentation in leadership research. Another option is the use of behavioral coding schemes. Bales (1950) developed a behavioral coding scheme for leadership, and more recently, Jacquart and Antonakis (2015) used coding of leaders' speeches. Hence, we call for scholars to develop a coding scheme for abusive supervision.

Second, scholars need to develop a clearly evaluative construct of abuse too, because as Tepper (2000) outlines, negative consequences of abusive supervision likely rest on whether or not followers evaluate the leader as abusive. Hence, whereas we lament the

conflation of behaviors and evaluations, we do not suggest that subjective evaluations should be ignored. To the contrary, we call for research exploring subjective evaluations in their own right, because evaluations are more proximal antecedents of follower-level outcomes than leader behaviors (Banks et al., 2020). We believe that such an evaluative construct can be developed by revising Tepper's (2000) conflated construct. We sketch a revision of the definition and measurement of the current abusive supervision construct toward a truly evaluative construct that we suggest calling "*experience of supervisor abuse*".

Tepper's (2000; p. 178) definition clearly states that abusive supervision reflects subordinate evaluation of supervisors' behavior as *hostile verbal and nonverbal* (but not physically so), over a *sustained* period. We question the necessity to rule out non-physical and non-sustained cases of abuse as not abusive. After all, physical transgressions of supervisors might be perceived as abusive, just like cases of occasional and non-sustained abuse too. Hence, instead of ruling out that physical and non-sustained transgressions are not abusive, the revised construct should treat this question as an empirical, not a conceptual one; future empirical research can address this question. Thus, we suggest a revised and simplified definition of the experience of supervisor abuse: *The extent to which supervisor behavior is evaluated as abusive*.

Subsequently, in contrast to the current measurement items, which are neither unambiguously behavioral nor unambiguously evaluative, we suggest refining existing formulations for developing a truly evaluative questionnaire. For instance, the first item in Tepper's (2000) scale is "Ridicules me", whereas "I feel ridiculed" would better fit Tepper and colleagues' (2000; 2017) ambition to study follower evaluations; and it better fits our revised definition, too. Taken together, we suggest that future research should revise the items of the original scale and future research should psychometrically validate such a revised scale (Hughes, 2018). Then, scholars can examine the impact of leader behaviors on followers'

evaluations of abusive supervision, and subsequently, the impact of these evaluations on other relevant variables. Parsing and precisely assessing leader behaviors and followers' evaluations can be the starting point for studying the causal links between these conceptually distinct phenomena and for building a more rigorous knowledge base of abusive supervision.

Rethinking study designs

Study design in general. Our review shows that most abusive supervision studies sought to understand causal effects, yet used study designs that are poorly suited to the task. To address this concern, we make three suggestions for future research. First, whenever possible, scholars should seek including exogenous sources of variation in their study. Experimental manipulations are the obvious, but not the only way for doing so (see Eden, 2020; Sieweke & Santoni, 2020). Second, however, if such exogenous variation cannot be identified, it is necessary to statistically correct for the endogenous nature of supervisory behaviors. A recent review article provides good guidance for doing so (Hill, Johnson, Greco, O'Boyle, & Walter, 2020). Third, studies should examine abusive supervision as a processual and hence longitudinal phenomenon. Tepper (2000) defines abusive supervision as a sustained phenomenon. Fischer et al. (2017) outline how improper modelling of temporal unfolding can lead to severe empirical distortions, and they offer guidelines for modeling leadership processes correctly.

Low base rates. Thankfully, employee evaluations suggest that abusive supervision is a rare phenomenon, if we assume that this observation does not primarily reflect measurement error (e.g., underreporting of abuse, misleading questionnaire items). Nevertheless, low base rates pose statistical challenges if we want to study the effects of intermediate and high levels of abuse. One solution is experiments, which can ensure statistically meaningful levels of abusive supervision are observed. As we have discussed previously, ethically defensible

methods of manipulating exposure to abusive supervision have been identified and we encourage scholars to follow such examples.

Another answer to the challenge of low base rates is studying abusive supervision in contexts where abuse is more likely to happen. Arguably, for instance, competitive organizational cultures (e.g., sport, investment banking) and industries in which employees have low bargaining power (e.g., arts, cleaning) might be settings where abusive supervision is relatively more prevalent. There, scholars might naturally find the full variance range of abuse. If, however, a study cannot ensure such a full variance range for whatever reasons, it is still feasible to study abusive supervision. However, doing so requires intellectual honesty and transparency about the work. If one is examining the differential impact of no abuse to low or lower intermediate levels of abuse, then scholars should declare that. The problem is not studying infrequent cases of abuse but overstating and incorrectly extrapolating findings.

Limitations

Before concluding, we acknowledge that, as all research, our review has limitations too. Notably, we focused on published articles to highlight how abusive supervision is studied in the literature and to avoid the double inclusion of data (e.g., unpublished dissertation later published in a peer-reviewed journal). While focusing only on published studies is common (e.g., Harms, Credé, Tynan, Leon, & Jeung, 2017), we note that we may have potentially missed studies that are relevant to our review and might have missed important trends that stem from publication bias (Siddaway, Wood, & Hedges, 2019; Steel, Beugelsdijk, & Aguinis, 2021). In addition, we only included articles that were written in English, which could have produced ethnocentric biases (Steel et al., 2021). Furthermore, we used two independent coders at two different coding stages to code a set of randomly selected studies, and we reach an inter-coder agreement of above 95%. Whereas such an approach

substantively reduces subjectivity, it does not fully eliminate the risk of measurement error. Human coding is always prone of biases.

Conclusion

We reported a systematic and critical review of abusive supervision research. Alongside providing a comprehensive catalogue of the different correlates (antecedents, outcomes, mediators, moderators) of abusive supervision that serves as a useful map of empirical research in the field, we identified challenges facing the conceptualization, measurement, and empirical study of abusive supervision. Most importantly, however, we provided achievable recommendations for addressing these challenges, which if enacted will serve to improve the veracity of abusive supervision research. Thus, although our review is critical of the typical abusive supervision study, we believe it is constructive and can contribute to improving and rethinking the study of one of the most disconcerting and therefore most important topics of organizational research: abuse of followers by leaders. Only through rigorous research can we achieve more definitive information regarding how frequent abusive supervision is, how to best select and train leaders to reduce abusive supervision, and how to help followers be resilient when faced with abusive supervision.

Table 1. Analysis of abusive supervision items

	Negative leader behavior	Implicitly double-barreled	Explicitly double-barreled	Negative, but not necessarily abusive	Figurative
1. Ridicules me.	X	X			
2. Tells me my thoughts or feelings are stupid.	X				X
3. Gives me the silent treatment.	X	X			
4. Puts me down in front of others.	X	X			
5. Invades my privacy.	X	X			
6. Reminds me of my past mistakes and failures.	X			X	
7. Doesn't give me credit for jobs requiring a lot of effort.	X			X	
8. Blames me to save himself/herself embarrassment.	X		X		
9. Breaks promises he/she makes.	X			X	
10. Expresses anger at me when he/she is mad for another reason.	X		X		
11. Makes negative comments about me to others.	X	X			
12. Is rude to me.	X	X			
13. Does not allow me to interact with my coworkers.	X			X	
14. Tells me I'm incompetent.	X				X
15. Lies to me.	X			X	

Table 2. Summary statistics for several study design characteristics

Study characteristics	Frequency	%
Study designs		
<i>Cross-sectional self-rated</i>	140	28.57
<i>Cross-sectional multiple source</i>	107	21.84
<i>Time-separated self-rated</i>	105	21.43
<i>Time-separated multiple source</i>	71	14.49
<i>Longitudinal (except diary) self-rated</i>	7	1.43
<i>Longitudinal (except diary) multiple source</i>	3	.61
<i>Diary Study self-rated</i>	11	2.24
<i>Diary Study multiple source</i>	2	.41
Randomized control trial experiment	43	8.78
Field experiment	1	.20
Level of analysis		
Individual - between	467	95.31
Individual - within	7	1.43
Multilevel/team level	16	3.27
Most frequently sampled countries		
United States of America	155	31.63
China	120	24.49
Philippines	20	4.08
Pakistan	18	3.67
South Korea	17	3.47

Note. Percentages are based on total number of samples (N = 490).

Table 3. Overview of experimental studies of abusive supervision

Authors	Experimental Design	Participants	Manipulation Method	Recruitment	Setting
Participants as <i>Followers</i>					
Brees, Martinko, and Harvey (2016)	Pretest – Test	Working adults (N=756)	Videoed vignette scenario	Convenience sample	Laboratory
Burton and Hoobler (2006)	Randomized control: 2 (abusive vs. supportive leader)	Students (N=196)	Videoed vignette scenario	Convenience sample	Laboratory
Burton and Barber (2019)	Randomized control: 2 (No video vs. mindfulness video) x 2 (positive supervisor vs. abusive supervisor)	Students (N=263)	Written vignette scenario	Convenience sample	Laboratory
Burton et al. (2011)	Randomized control: 2 (abusive vs. supportive leader)	Students (N= 134)	Written vignette scenario	Convenience sample	Laboratory
Caesens et al. (2019)	Randomized control (abusive vs non-abusive leader)	Working adults (N= 212)	Written vignette scenario	Online sample (Prolific)	Laboratory
Farh and Chen (2014)	Randomized control: 2 (individual-level abusive supervision: high vs. low) x 2 (team-level abusive supervision: high vs. low)	Students (N=276)	Written vignette scenario	Convenience sample	Laboratory
Fiset et al. (2019)	Study 1: Randomized control, 2 (abusive supervision: high vs. neutral) x 2 (vision: high vs. low)	Working adults (N=205)	Written vignette scenario	Online sample (MTurk)	Laboratory
	Study 2: Randomized control 3 (abusive supervision: injury initiation versus performance promotion vs. neutral) x 2 (vision: high vs. low)	Working adults (N=294)	Written vignette scenario	Online sample (MTurk)	Laboratory
Greenbaum et al. (2017)	Randomized control (abusive vs non abusive)	Undergraduate students (N=151)	Priming Study (ruminate and write)	Convenience sample	Laboratory

Hurst et al. (2019)	Randomized control (abusive vs transformational leadership)	Working adults (N=433)	about a past abusive authority figure) Written vignette scenario	Online sample	Laboratory
Liang et al. (2018)	Study 1: Randomized control: 3 (abusive supervision/no retaliation vs. abusive supervision/retaliation vs. control) Study 2: Randomized control: 2 (supervisory treatment: abusive supervision vs. neutral interaction) ×2 (retaliation: no retaliation vs. retaliation)	Working adults (N=195) Students (N=150)	Recall exercise about an abusive incident – no trigger Recall exercise about an abusive v – no trigger	Online sample (MTurk) Convenience sample	Laboratory Laboratory
Lopes et al. (2019)	Randomized control: 2 (abusive vs positive leadership condition)	Employed adults (N=100)	Videoed vignette scenario	Recruited via adverts at organizations	Laboratory
Martinko, Randolph-Seng, et al. (2018)	Randomized control: 3 (leader performance: low vs. average vs. high) × 3 (follower performance: low vs. average vs. high).	Students and working adults (N=282)	Videoed and written vignette scenarios	Convenience sample	Laboratory
Mitchell and Ambrose (2012)	Randomized control: 2 (aggression: high vs. low) x 2 (fear of retaliation: high vs. low)	Students (N=242)	Feedback received from teacher	Convenience sample	Field
Park, Carter, DeFrank, and Deng (2018)	Randomized control: 2 (abusive supervision: high vs. low) by 2 (gender similarity: different vs. same)	Working adults (N=222)	Written vignette scenario	Convenience sample	Laboratory
Rice, Taylor, and Forrester (2020)	Randomized control: 2 (abusive supervision: low vs. high) x 2 (leader political skill: low vs. high)	Study 1: students (N=53) Study 2: students (N=31)	Written vignette scenario Written vignette scenario	Convenience sample Convenience sample	Laboratory Laboratory

Schmid, Pircher Verdorfer, and Peus (2018)	Randomized control: 3 (abusive supervision vs. exploitative leadership vs. organization-directed destructive leader behaviors)	Working adults (N=297)	Written vignette scenario	Convenience sample	Laboratory
Schyns et al. (2018)	Randomized control: 4 (constructive vs. laissez-faire vs. mild abusive vs. strong abusive supervision)	Pilot: working adults (N=207)	Written vignette scenario	Online sample (Qualtrics)	Laboratory
		Study 1: working adults (N=310)	Written vignette scenario	Online sample (Qualtrics)	Laboratory
		Study 2: working adults (N=234)	Written vignette scenario	Online sample (Respondi)	Laboratory
Shao et al. (2018)	Randomized control: 2 (peer abusive supervision: high vs. low)	Undergraduate students (N=84)	Written vignette scenario	Convenience sample	Laboratory
Smallfield, Hoobler, and Kluemper (2020)	Randomized control: 2 (team helping behavior: high vs. low) x 2 (leader performance feedback: positive vs. negative)	MBA students (N=167)	Videoed and written vignette scenarios	Convenience sample	Laboratory
Tillman, Gonzalez, Crawford, and Lawrence (2018)	Randomized control (abusive supervision vs. nonabusive supervision)	Students (N=427)	Written vignette scenario	Convenience sample	Laboratory
Tröster and Van Quaquebeke (2020)	Randomized control: 2 (abusive supervision: low vs. high) x 2 (LMX: high vs. low)	Working adults (N=200)	Written vignette scenario	Online sample (MTurk)	Online
Wang and Jiang (2015)	Randomized control (abusive vs nonabusive supervision)	Working adults (N=196)	Written vignette scenario	Convenience sample	Laboratory
Xu et al. (2020)	Randomized control: 2 (abusive vs. non abusive) x 2 (rivalry vs. non-rivalry)	Students (N=156)	Written vignette scenario	Convenience sample	Laboratory
Yu and Duffy (2020)	Randomized control: 2 (abuse vs. nonabuse) x 3 attribution (injury initiation vs. performance promotion vs. nonmanipulated attribution).	University students and staff (N=156)	Confederate actor (supervisor)	Convenience sample	Laboratory

Participants as Leaders					
Camps, Stouten, Euwema, and De Cremer (2020)	Randomized control: 2 (self-doubt: low vs high) x 2 (interpersonal justice: low vs high)	Students (N=312)	Priming (self-doubt) and written vignette scenario (interpersonal justice)	Convenience sample	Laboratory
Collins and Jackson (2015)	Randomized control (easy vs hard maths test)	Organizational leaders (N=161)	Task difficulty manipulation	Convenience sample attending a leadership program	Field
Ju et al. (2019)	Randomized control (abusive vs. nonabusive supervisory behavior)	Study 2a: supervisors (N=102)	Recall exercise - trigger of writing an abusive paragraph	Online sample (Sojump.com)	Online
		Study 2b: supervisor (N=194)	Write email to subordinates (abusive supervision manipulation)	Online sample (MTurk)	Online
		Study 3: supervisor (N=282)	Recall exercise – no trigger	Online sample (MTurk)	Online
Karagonlar and Neves (2020)	Randomized control: 2 (Supervisor's concern for the organization: high vs low) x 2 (Subordinate's concern for the organization: high vs low)	Students (N=78)	Written vignette scenario	Convenience sample	Laboratory
Liang et al. (2016)	Randomized control (hostility vs joviality)	Full-time supervisors (N = 101)	Priming Study (recall an interaction for each conditions)	Online Sample (MTurk)	Laboratory
Qin et al. (2018)	Study 1A: Randomized control (abusive vs. non abusive supervision condition)	Students (N=64)	Chat group communication between leader and	Convenience sample	Laboratory , WeChat (instant

	Study 1B: Randomized control (abusive vs. nonabusive supervision condition)	Supervisors (N=100)	subordinates (confederates)	Online text messages from leader and subordinates (does not really exist)	Online sample (MTurk)	communication tool)
Shum, Gatling, and Tu (2020)	Randomized control: 2 (abusive vs. non abusive)	Hospitality supervisor (N=285)		Written vignette scenario	Online sample (MTurk)	Online
Tu et al. (2018)	Randomized control: 2 (leadership: abusive vs. nonabusive) x 2 (team performance: high vs. low)	Undergraduate students (N=93 teams)		Prerecorded script-based interactions (3 videos and 2 audio-tapes) between leader (actor) and team.	Convenience sample	Laboratory
Taylor, Griffith, Vadera, Folger, and Letwin (2019)	Study 1: Randomized control: 2 (abusive supervisor: high vs. low) x 2 (supervisor relational disidentification: high vs. low).	Students (N=125) and Working adults (N=163)		Written Vignette Scenario; Write email to subordinates	Convenience sample (students), Online sample (MTurk)	Online
	Study 2: Randomized control (abusive behavior: high vs. low)	Upper-level management course students (N=462)		Written vignette scenario;	Convenience sample	Online
Walter, Lam, Van Der Vegt, Huang, and Miao (2015)	Randomized control: 2 (perceived subordinate performance: lower vs. higher) 2 (outcome dependence: lower vs. higher) x 2 (liking: lower vs. higher)	Students (N=157)		Written vignette scenario	Convenience sample	Laboratory
Watkins, Fehr, and He (2019)	Randomized control (instrumentality beliefs vs. control)	Working adults (N=81)		Written vignette scenario	Online sample (MTurk)	Laboratory

Figure 1. An overview of the antecedents, mediators, and outcomes associated with abusive supervision

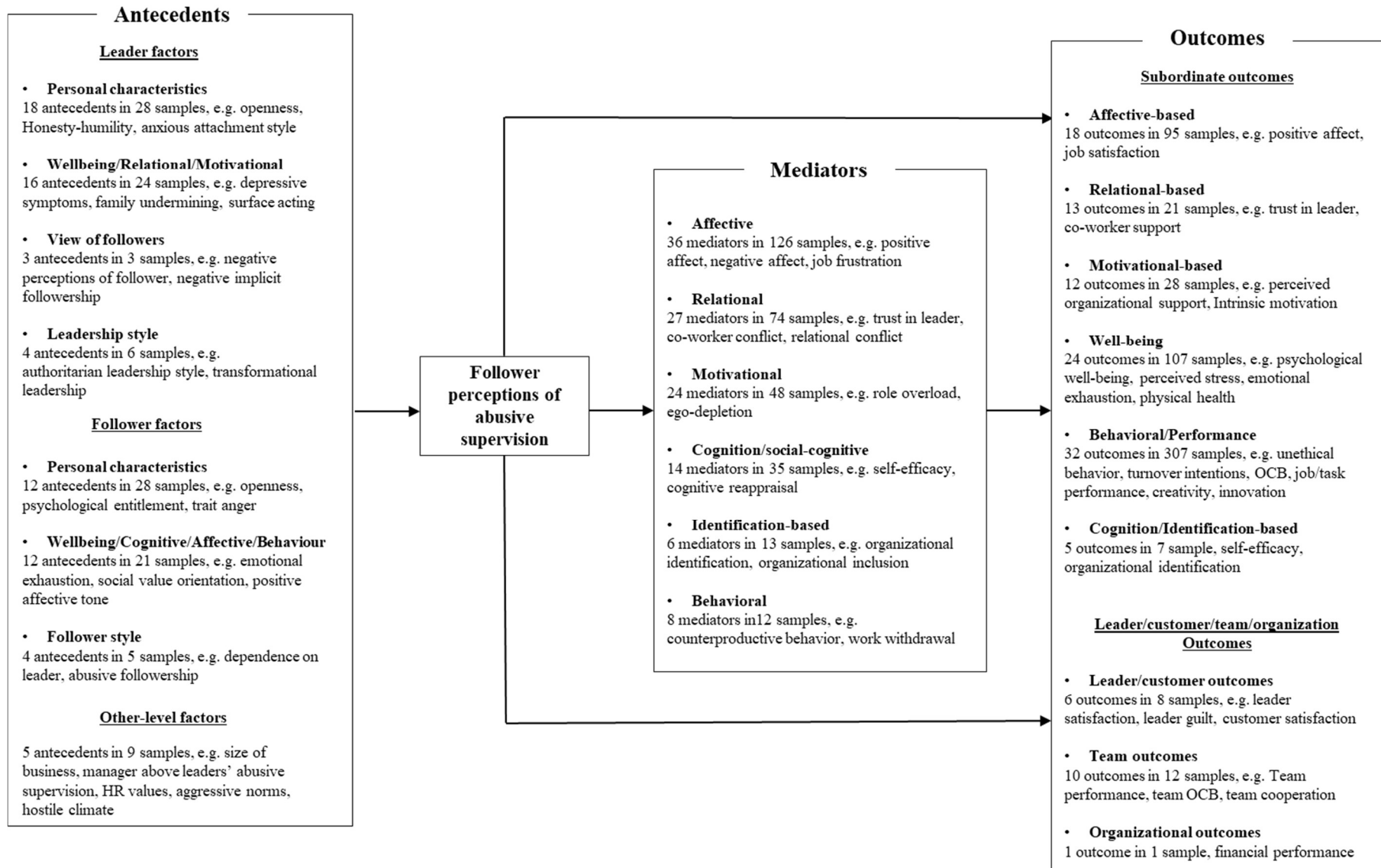


Figure 2 A summary of moderators associated with abusive supervision as the antecedent

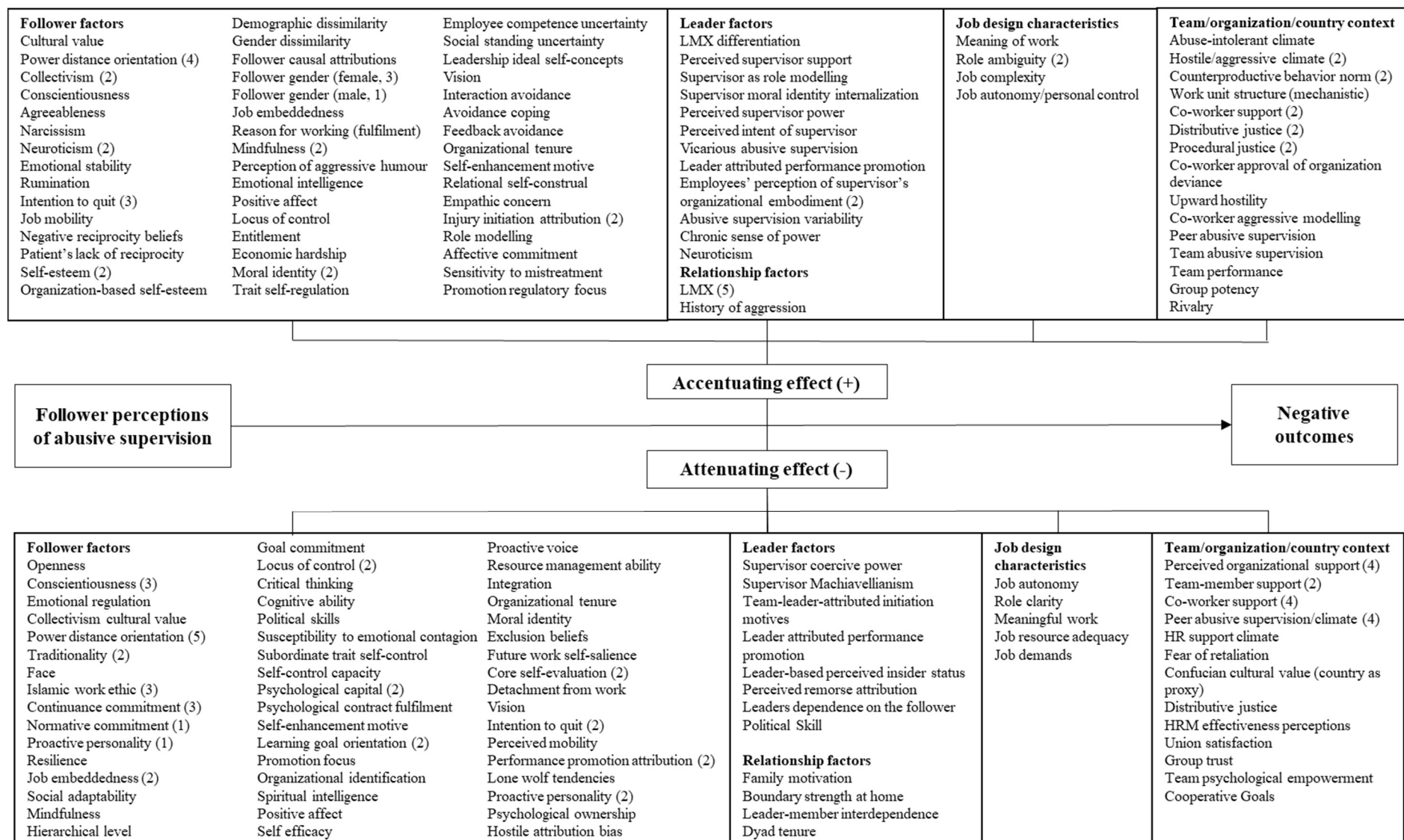


Figure 3. The base rate of abusive supervision

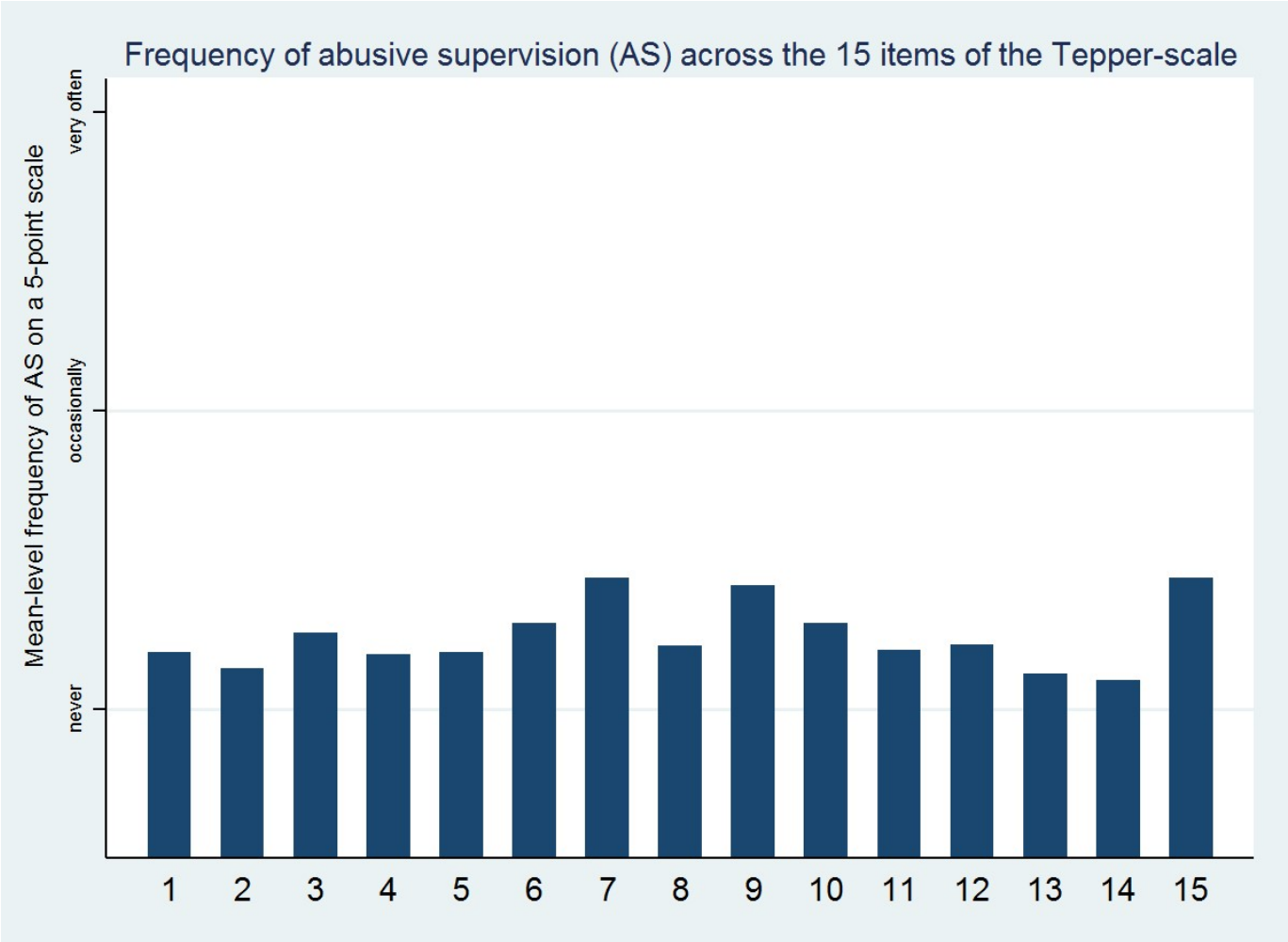
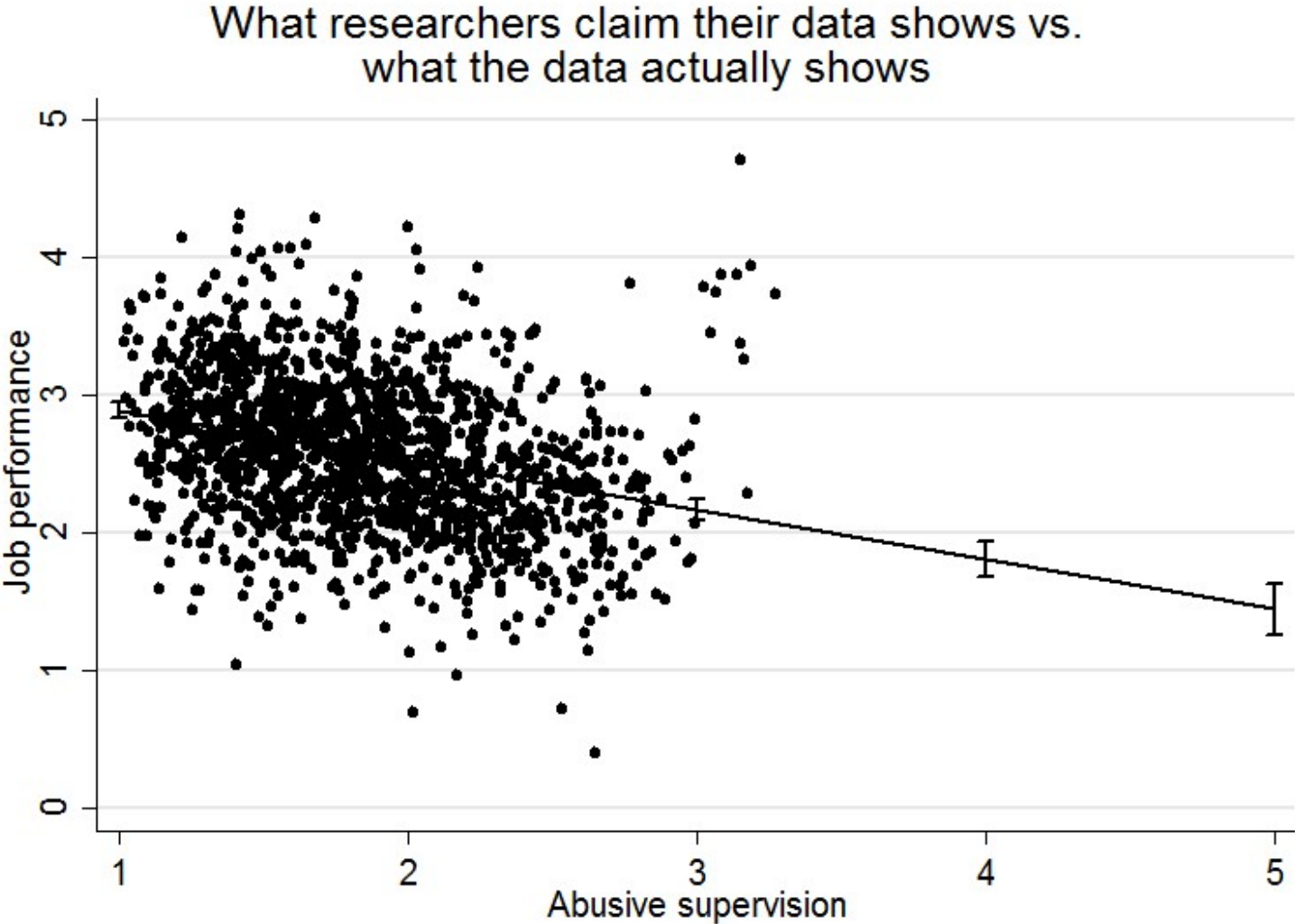


Figure 4. Illustrating the undue extrapolation of causal relationships from low to high base rates



APPENDIX

Table A1: Frequency of articles by journal with five or more samples

Journal title	Frequency	%
<i>Journal of Applied Psychology</i>	67	13.67
<i>Journal of Business Ethics</i>	30	6.12
<i>The Leadership Quarterly</i>	28	5.71
<i>Academy of Management Journal</i>	26	5.31
<i>Journal of Organizational Behavior</i>	25	5.10
<i>Journal of Business and Psychology</i>	17	3.47
<i>Journal of Leadership & Organizational Studies</i>	16	3.27
<i>Frontiers in Psychology</i>	13	2.65
<i>Journal of Management</i>	13	2.65
<i>Journal of Managerial Psychology</i>	12	2.45
<i>Personnel Review</i>	11	2.24
<i>Organizational Behavior and Human Decision Process</i>	10	2.04
<i>European Journal of Work and Organizational Psychology</i>	9	1.84
<i>Personnel Psychology</i>	8	1.63
<i>Journal of Occupational and Organizational Psychology</i>	7	1.43
<i>The International Journal of Human Resource Management</i>	6	1.22
<i>Human Relations</i>	5	1.02
<i>International Journal of Contemporary Hospitality Management</i>	5	1.02
<i>Journal of Social Psychology</i>	5	1.02
<i>Leadership and Organization Development Journal</i>	5	1.02
<i>Personality and Individual Differences</i>	5	1.02
<i>Social Behavior and Personality</i>	5	1.02
Total	328	66.94

Note. Percentages are based on total number of samples (N = 490).

Table A2: Measurement of abusive supervision studies

Scale	No of items	Frequency	%
Tepper (2000) original scale	15	280	58.33
Shortened scale commonly used			
- Tepper (2000) – shortened	5	13	2.71
- Tepper (2000) – shortened	6	18	3.75
- Tepper (2000) – shortened scale with less than 10 studies	(range between 1 to 14)	45	9.38
- Mitchell & Ambrose (2007) – short	5	107	22.29
- Mitchell & Ambrose (2007) – long	10	4	.83
- Aryee et al (2007)	10	13	2.71
Scale anchor			
Frequency	/	249	54.37
Agreement	/	200	43.67
Others (e.g., likelihood)		9	1.96

Note. Percentages for *Scale* are based on the total number of samples that provided scale information (N = 480). Percentages for *Scale anchor* are based on the total number of samples that provided scale anchor information (N = 458).

Figure A1. A detailed summary of outcomes of abusive supervision

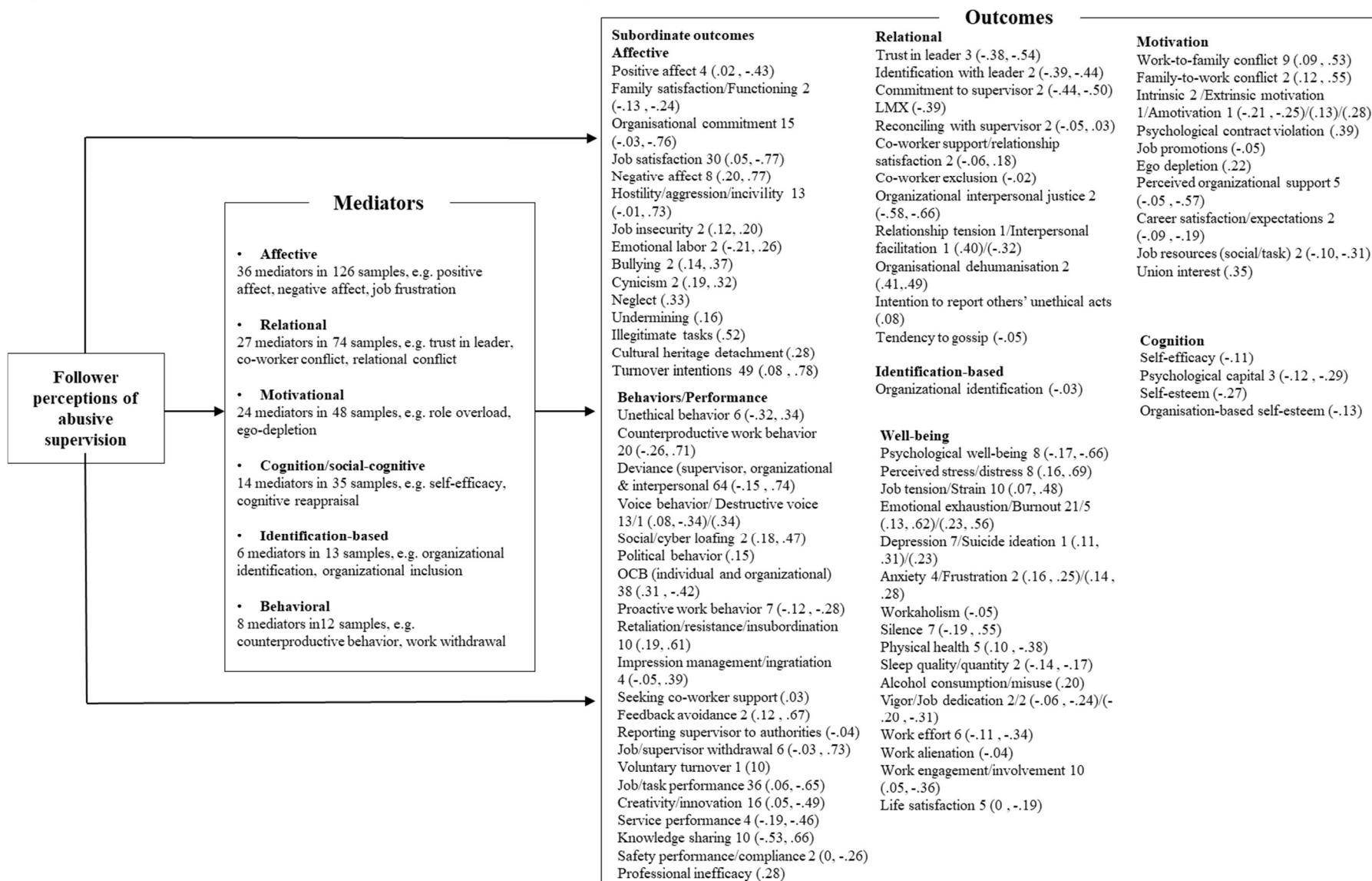


Figure A2. A detailed summary of mediators of abusive supervision

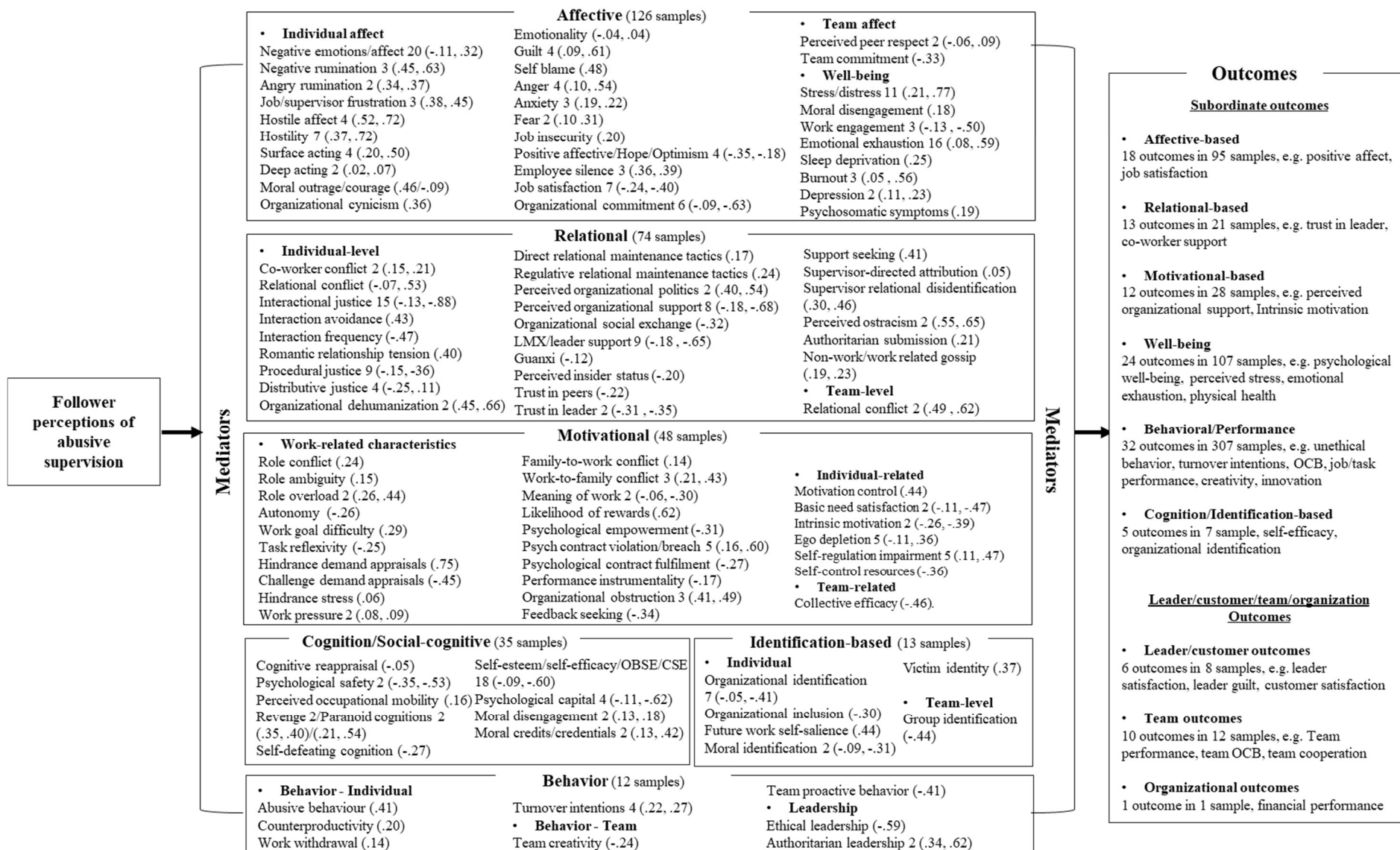


Figure A3. A detailed summary of antecedents of abusive supervision

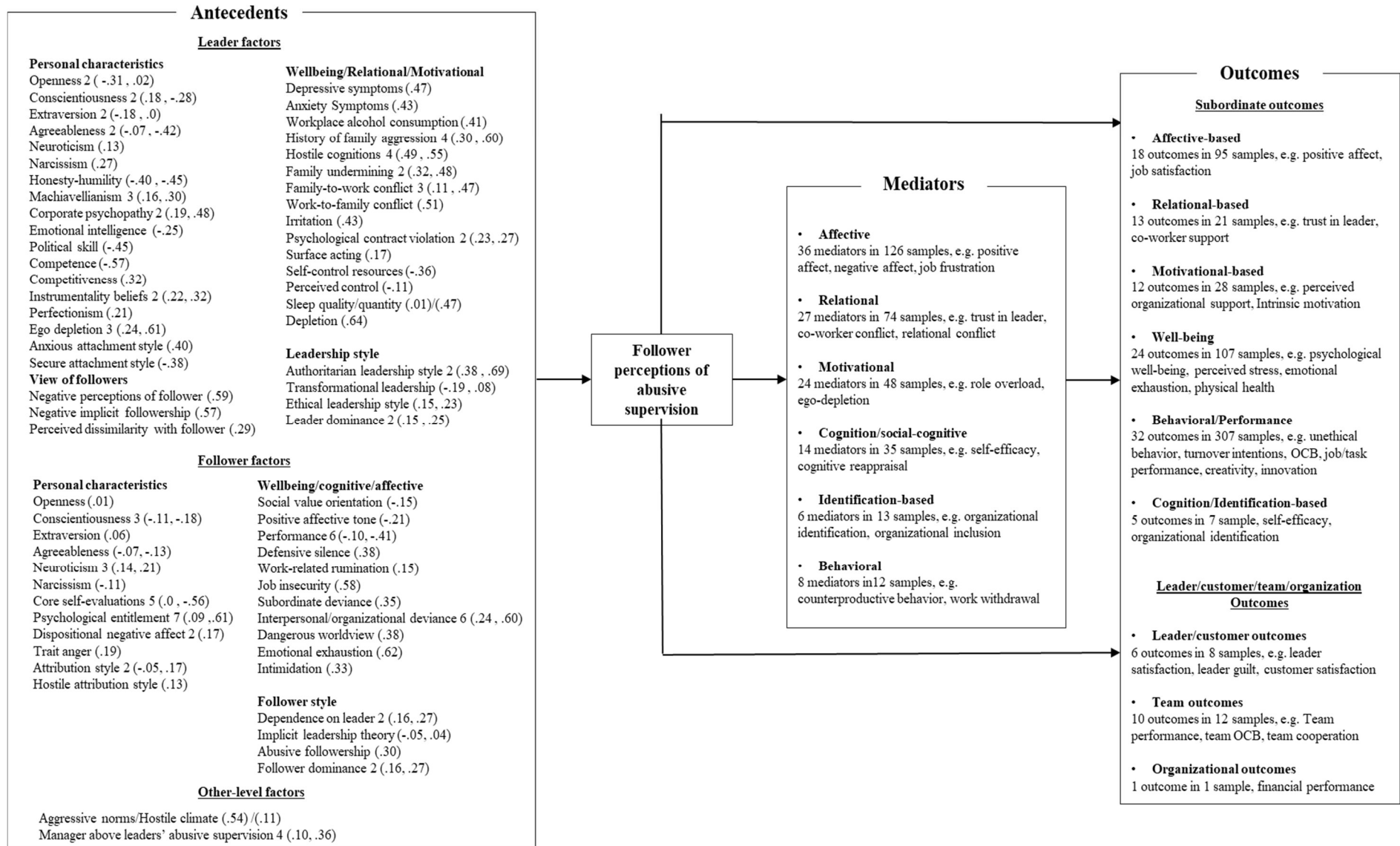
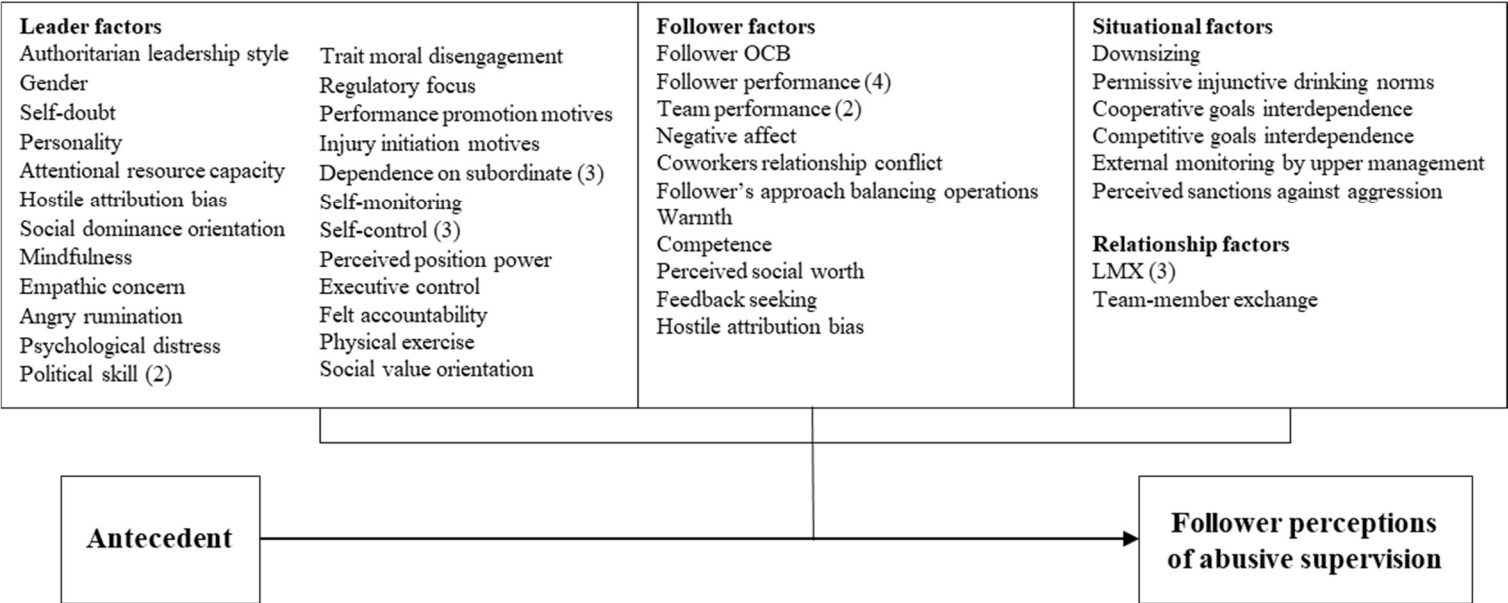


Figure A4 .A detailed summary of moderators associated with abusive supervision as the outcome



Stata Code A1: Stata code underlying the stylized illustration

```
clear
set seed 123
set obs 1530

//generating the abusive supervision variable
gen as = 1+2.5*rbeta(2, 4)

//generating the performance variable (with different levels for high/low AS)
gen y = 3.5 - .5*as + .5*rnormal()
replace y = 3.5 + .5*rnormal() if as>3

//creating the regression
reg y as

//plotting the regression- predictions of the abusive supervision-performance link
margins, at(as = (1(1)5))
marginsplot, ylabel(1(1)5) xlabel(1(1)5) plot1opts(lcolor(black) mlabcolor(black) ///
msymbol(none)) ciopts(lcolor(black)) graphregion(fcolor(white)lcolor(white)) ///
scheme(s2mono) ///
title("What researchers claim their data shows vs." "what the data actually shows") ///
xtitle("Abusive supervision") legend(off) ///
ytitle("Job performance") name(four, replace) ///
addplot(scatter y as, msymbol(circle) msize(small) mcolor(black%10) ///
mlcolor(black%10))
```

References

- Ahmad, J., Athar, M. R., Azam, R. I., Hamstra, M. R., & Hanif, M. (2019). A resource perspective on abusive supervision and extra-role behaviors: The role of subordinates' psychological capital. *Journal of Leadership & Organizational Studies*, 26(1), 73-86.
- Angrist, J. D., & Pischke, J.-S. (2014). *Mastering 'metrics: The path from cause to effect*: Princeton University Press.
- Antonakis, J. (2017). On doing better science: From thrill of discovery to policy implications. *The Leadership Quarterly*, 28(1), 5-21.
- Antonakis, J., Bendahan, S., Jacquart, P., & Lalive, R. (2010). On making causal claims: A review and recommendations. *The Leadership Quarterly*, 21(6), 1086-1120.
- Antonakis, J., Bendahan, S., Jacquart, P., & Lalive, R. (2014). Causality and endogeneity: Problems and solutions. In *The Oxford Handbook of Leadership and Organizations* (pp. 93): Oxford University Press New York, NY.
- Aryee, S., Chen, Z. X., Sun, L.-Y., & Debrah, Y. A. (2007). Antecedents and outcomes of abusive supervision: Test of a trickle-down model. *Journal of Applied Psychology*, 92(1), 191.
- Avey, J. B., Wu, K., & Holley, E. (2015). The influence of abusive supervision and job embeddedness on citizenship and deviance. *Journal of Business Ethics*, 129(3), 721-731.
- Bales, R. F. (1950). A set of categories for the analysis of small group interaction. *American Sociological Review*, 15(2), 257-263.
- Banks, G., Fischer, T., Gooty, J., & Stock, G. (2020). Ethical leadership: Mapping the terrain for concept cleanup and a future research agenda. *The Leadership Quarterly*.
- Bendahan, S., Zehnder, C., Pralong, F. P., & Antonakis, J. (2015). Leader corruption depends on power and testosterone. *The Leadership Quarterly*, 26(2), 101-122.
- Bormann, K. C. (2017). Linking daily ethical leadership to followers' daily behaviour: The roles of daily work engagement and previous abusive supervision. *European Journal of Work and Organizational Psychology*, 26(4), 590-600.
- Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2004). The concept of validity. *Psychological Review*, 111(4), 1061-1071.
- Brees, J., Martinko, M., & Harvey, P. (2016). Abusive supervision: Subordinate personality or supervisor behavior? *Journal of Managerial Psychology*.
- Breevaart, K., & de Vries, R. E. (2017). Supervisor's HEXACO personality traits and subordinate perceptions of abusive supervision. *The Leadership Quarterly*, 28(5), 691-700.
- Burton, J. P., & Barber, L. K. (2019). The role of mindfulness in response to abusive supervision. *Journal of Managerial Psychology*.
- Burton, J. P., & Hoobler, J. M. (2006). Subordinate self-esteem and abusive supervision. *Journal of Managerial Issues*, 340-355.
- Burton, J. P., Hoobler, J. M., & Kernan, M. C. (2011). When research setting is important: The influence of subordinate self-esteem on reactions to abusive supervision. *Organization Management Journal*, 8(3), 139-150.
- Caesens, G., Nguyen, N., & Stinglhamber, F. (2019). Abusive supervision and organizational dehumanization. *Journal of Business and Psychology*, 34(5), 709-728.
- Camps, J., Stouten, J., Euwema, M., & De Cremer, D. (2020). Abusive supervision as a response to follower hostility: a moderated mediation model. *Journal of Business Ethics*, 164(3), 495-514.
- Carl, D., Gupta, V., & Javidan, M. (2004). Power distance. *Culture, leadership, and organizations: The GLOBE study of*, 62(2004), 513-563.

- Collins, M. D., & Jackson, C. J. (2015). A process model of self-regulation and leadership: How attentional resource capacity and negative emotions influence constructive and destructive leadership. *The Leadership Quarterly*, 26(3), 386-401.
- Eden, D. (2020). The science of leadership: A journey from survey research to field experimentation. *The Leadership Quarterly*.
- Farh, C. I., & Chen, Z. (2014). Beyond the individual victim: Multilevel consequences of abusive supervision in teams. *Journal of Applied Psychology*, 99(6), 1074.
- Ferris, D. L., Yan, M., Lim, V. K., Chen, Y., & Fatimah, S. (2016). An approach–avoidance framework of workplace aggression. *Academy of Management Journal*, 59(5), 1777-1800.
- Fischer, T., Dietz, J., & Antonakis, J. (2017). Leadership process models: A review and synthesis. *Journal of Management*, 43(6), 1726-1753.
- Fischer, T., Hambrick, D., Sajons, G., & van Quaquebeke, N. (2020). Beyond the ritualized use of questionnaires: Toward a science of actual behaviors and psychological states. *The Leadership Quarterly*, 31(4), 101449.
- Fiset, J., Robinson, M. A., & Saffie-Robertson, M. C. (2019). Masking wrongs through brilliance: The moderating effect of vision on the relationship between abusive supervision and employee outcomes. *European Journal of Work and Organizational Psychology*, 28(6), 756-768.
- Gabler, C. B., Nagy, K. R., & Hill, R. P. (2014). Causes and consequences of abusive supervision in sales management: A tale of two perspectives. *Psychology & Marketing*, 31(4), 278-293.
- Giurge, L. M., van Dijke, M., Zheng, M. X., & De Cremer, D. (2019). Does power corrupt the mind? The influence of power on moral reasoning and self-interested behavior. *The Leadership Quarterly*.
- Goswami, A., Park, H. I., & Beehr, T. A. (2020). Does the congruence between leaders' implicit followership theories and their perceptions of actual Followers matter? *Journal of Business and Psychology*, 35(4), 519-538.
- Greenbaum, R. L., Hill, A., Mawritz, M. B., & Quade, M. J. (2017). Employee Machiavellianism to unethical behavior: The role of abusive supervision as a trait activator. *Journal of Management*, 43(2), 585-609.
- Güntner, A. V., Klonek, F. E., Lehmann-Willenbrock, N., & Kauffeld, S. (2020). Follower behavior renders leader behavior endogenous: The simultaneity problem, estimation challenges, and solutions. *The Leadership Quarterly*, 31(6), 101441.
- Haar, J. M., de Fluiter, A., & Brougham, D. (2016). Abusive supervision and turnover intentions: The mediating role of perceived organisational support. *Journal of Management and Organization*, 22(2), 139.
- Hahn, G. J. (1977). The hazards of extrapolation in regression analysis. *Journal of Quality Technology*, 9(4), 159-165.
- Harms, P., Credé, M., Tynan, M., Leon, M., & Jeung, W. (2017). Leadership and stress: A meta-analytic review. *The Leadership Quarterly*, 28(1), 178-194.
- Harris, K. J., Kacmar, K. M., & Zivnuska, S. (2007). An investigation of abusive supervision as a predictor of performance and the meaning of work as a moderator of the relationship. *The Leadership Quarterly*, 18(3), 252-263.
- Henle, C. A., & Gross, M. A. (2014). What have I done to deserve this? Effects of employee personality and emotion on abusive supervision. *Journal of Business Ethics*, 122(3), 461-474.
- Hill, A., Johnson, S., Greco, L., O'Boyle, E., & Walter, S. (2020). Endogeneity: A review and agenda for the methodology-practice divide affecting micro and macro research. *Journal of Management*.

- Hobman, E. V., Restubog, S. L. D., Bordia, P., & Tang, R. L. (2009). Abusive supervision in advising relationships: Investigating the role of social support. *Applied Psychology, 58*(2), 233-256.
- Hofstede, G. (1980). Motivation, leadership, and organization: Do American theories apply abroad? *Organizational Dynamics, 9*(1), 42-63.
- Hughes, D. J. (2018). Psychometric validity: Establishing the accuracy and appropriateness of psychometric measures. *The Wiley handbook of psychometric testing: A multidisciplinary approach to survey, scale and test development*.
- Hughes, D. J., & Evans, T. R. (2018). Putting ‘emotional intelligences’ in their place: Introducing the integrated model of affect-related individual differences. *Frontiers in Psychology, 9*, 2155.
- Hughes, D. J., Lee, A., Tian, A. W., Newman, A., & Legood, A. (2018). Leadership, creativity, and innovation: A critical review and practical recommendations. *The Leadership Quarterly, 29*(5), 549-569.
- Hurst, C., Simon, L., Jung, Y., & Pirouz, D. (2019). Are “bad” employees happier under bad bosses? Differing effects of abusive supervision on low and high primary psychopathy employees. *Journal of Business Ethics, 158*(4), 1149-1164.
- Inceoglu, I., Thomas, G., Chu, C., Plans, D., & Gerbasi, A. (2018). Leadership behavior and employee well-being: An integrated review and a future research agenda. *The Leadership Quarterly, 29*(1), 179-202.
- Jacquart, P., & Antonakis, J. (2015). When does charisma matter for top-level leaders? Effect of attributional ambiguity. *Academy of Management Journal, 58*(4), 1051-1074.
- Ju, D., Huang, M., Liu, D., Qin, X., Hu, Q., & Chen, C. (2019). Supervisory consequences of abusive supervision: An investigation of sense of power, managerial self-efficacy, and task-oriented leadership behavior. *Organizational Behavior and Human Decision Processes, 154*, 80-95.
- Karagonlar, G., & Neves, P. (2020). No more Mr. Nice Guy: social value orientation and abusive supervision. *Journal of Managerial Psychology*.
- Kiazad, K., Restubog, S. L. D., Zagenczyk, T. J., Kiewitz, C., & Tang, R. L. (2010). In pursuit of power: The role of authoritarian leadership in the relationship between supervisors’ Machiavellianism and subordinates’ perceptions of abusive supervisory behavior. *Journal of Research in Personality, 44*(4), 512-519.
- Krasikova, D. V., Green, S. G., & LeBreton, J. M. (2013). Destructive leadership: A theoretical review, integration, and future research agenda. *Journal of Management, 39*(5), 1308-1338.
- Lewin, K., Lippitt, R., & White, R. K. (1939). Patterns of aggressive behavior in experimentally created “social climates”. *The Journal of Social Psychology, 10*(2), 269-299.
- Li, Y., Wang, Z., Yang, L.-Q., & Liu, S. (2016). The crossover of psychological distress from leaders to subordinates in teams: The role of abusive supervision, psychological capital, and team performance. *Journal of Occupational Health Psychology, 21*(2), 142.
- Lian, H., Ferris, D. L., & Brown, D. J. (2012). Does power distance exacerbate or mitigate the effects of abusive supervision? It depends on the outcome. *Journal of Applied Psychology, 97*(1), 107.
- Liang, L. H., Brown, D. J., Lian, H., Hanig, S., Ferris, D. L., & Keeping, L. M. (2018). Righting a wrong: Retaliation on a voodoo doll symbolizing an abusive supervisor restores justice. *The Leadership Quarterly, 29*(4), 443-456.

- Liang, L. H., Lian, H., Brown, D. J., Ferris, D. L., Hanig, S., & Keeping, L. M. (2016). Why are abusive supervisors abusive? A dual-system self-control model. *Academy of Management Journal*, 59(4), 1385-1406.
- Liao, Z., & Liu, Y. (2015). Abusive supervision and psychological capital: a mediated moderation model of team member support and supervisor-student exchange. *Frontiers of Business Research in China*, 9(4), 576.
- Lin, W., Wang, L., & Chen, S. (2013). Abusive supervision and employee well-being: The moderating effect of power distance orientation. *Applied Psychology*, 62(2), 308-329.
- 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.
- Lopes, B. C., Kamau, C., & Jaspal, R. (2019). Coping with perceived abusive supervision: The role of paranoia. *Journal of Leadership & Organizational Studies*, 26(2), 237-255.
- Low, Y. M., Sambasivan, M., & Ho, J. A. (2019). Impact of abusive supervision on counterproductive work behaviors of nurses. *Asia Pacific Journal of Human Resources*.
- Lyons, B. D., Moorman, R. H., & Mercado, B. K. (2019). Normalizing mistreatment? Investigating Dark Triad, LMX, and abuse. *Leadership & Organization Development Journal*.
- MacKenzie, S. B. (2003). The dangers of poor construct conceptualization. *Journal of the Academy of Marketing Science*, 31(3), 323-326.
- Mackey, J. D., Frieder, R. E., Brees, J. R., & Martinko, M. J. (2017). Abusive supervision: A meta-analysis and empirical review. *Journal of Management*, 43(6), 1940-1965.
- Martin, R., Hughes, D. J., Epitropaki, O., & Thomas, G. (2020). In pursuit of causality in leadership training research: A review and pragmatic recommendations. *The Leadership Quarterly*.
- Martinko, M., Harvey, P., Brees, J., & Mackey, J. (2013). A review of abusive supervision research. *Journal of Organizational Behavior*, 34(S1), S120-S137.
- Martinko, M., Mackey, J., Moss, S., Harvey, P., McAllister, C., & Brees, J. (2018). An exploration of the role of subordinate affect in leader evaluations. *Journal of Applied Psychology*.
- Martinko, M., Randolph-Seng, B., Shen, W., Brees, J., Mahoney, K., & Kessler, S. (2018). An examination of the influence of implicit theories, attribution styles, and performance cues on questionnaire measures of leadership. *Journal of Leadership & Organizational Studies*, 25(1), 116-133.
- Mawritz, M. B., Mayer, D. M., Hoobler, J. M., Wayne, S. J., & Marinova, S. V. (2012). A trickle-down model of abusive supervision. *Personnel Psychology*, 65(2), 325-357.
- Meehl, P. E. (1990). Appraising and amending theories: The strategy of Lakatosian defense and two principles that warrant it. *Psychological Inquiry*, 1(2), 108-141.
- Mitchell, M. S., & Ambrose, M. L. (2007). Abusive supervision and workplace deviance and the moderating effects of negative reciprocity beliefs. *Journal of Applied Psychology*, 92(4), 1159.
- Mitchell, M. S., & Ambrose, M. L. (2012). Employees' behavioral reactions to supervisor aggression: An examination of individual and situational factors. *Journal of Applied Psychology*, 97(6), 1148.
- Park, H., Hoobler, J. M., Wu, J., Liden, R. C., Hu, J., & Wilson, M. S. (2019). Abusive supervision and employee deviance: A multifoci justice perspective. *Journal of Business Ethics*, 158(4), 1113-1131.

- Park, J. H., Carter, M. Z., DeFrank, R. S., & Deng, Q. (2018). Abusive supervision, psychological distress, and silence: The effects of gender dissimilarity between supervisors and subordinates. *Journal of Business Ethics, 153*(3), 775-792.
- Peltokorpi, V., & Ramaswami, A. (2019). Abusive supervision and subordinates' physical and mental health: the effects of job satisfaction and power distance orientation. *The International Journal of Human Resource Management, 1-27*.
- Peng, A., Mitchell, R., & Schaubroeck, J. M. (2019). Abusive supervision. In *Oxford Research Encyclopedia of Business and Management*.
- Peng, A., Schaubroeck, J., Chong, S., & Li, Y. (2019). Discrete emotions linking abusive supervision to employee intention and behavior. *Personnel Psychology, 72*(3), 393-419.
- Peng, A., Schaubroeck, J., & Li, Y. (2014). Social exchange implications of own and coworkers' experiences of supervisory abuse. *Academy of Management Journal, 57*(5), 1385-1405.
- Podsakoff, P. M., & Podsakoff, N. P. (2019). Experimental designs in management and leadership research: Strengths, limitations, and recommendations for improving publishability. *The Leadership Quarterly, 30*(1), 11-33.
- Pradhan, R. K., & Jena, L. K. (2017). Employee performance at workplace: Conceptual model and empirical validation. *Business Perspectives and Research, 5*(1), 69-85.
- Qin, X., Huang, M., Johnson, R. E., Hu, Q., & Ju, D. (2018). The short-lived benefits of abusive supervisory behavior for actors: An investigation of recovery and work engagement. *Academy of Management Journal, 61*(5), 1951-1975.
- Rice, D. B., Taylor, R., & Forrester, J. K. (2020). The unwelcoming experience of abusive supervision and the impact of leader characteristics: turning employees into poor organizational citizens and future quitters. *European Journal of Work and Organizational Psychology, 1-18*.
- Richard, O. C., Boncoeur, O. D., Chen, H., & Ford, D. L. (2020). Supervisor abuse effects on subordinate turnover intentions and subsequent interpersonal aggression: The role of power-distance orientation and perceived human resource support climate. *Journal of Business Ethics, 164*(3), 549-563.
- Robertson, J. L., Dionisi, A. M., & Barling, J. (2018). Linking attachment theory to abusive supervision. *Journal of Managerial Psychology*.
- Ross, L., & Nisbett, R. E. (2011). *The person and the situation: Perspectives of social psychology*: Pinter & Martin Publishers.
- Schat, A. C., Frone, M. R., & Kelloway, E. K. (2006). Prevalence of Workplace Aggression in the US Workforce: Findings From a National Study.
- Schmid, E. A., Pircher Verdorfer, A., & Peus, C. (2017). Shedding light on leaders' self-interest: Theory and measurement of exploitative leadership. *Journal of Management, 0149206317707810*.
- Schmid, E. A., Pircher Verdorfer, A., & Peus, C. V. (2018). Different shades—different effects? Consequences of different types of destructive leadership. *Frontiers in Psychology, 9*, 1289.
- Schyns, B., Felfe, J., & Schilling, J. (2018). Is It Me or You?—How Reactions to Abusive Supervision Are Shaped by Leader Behavior and Follower Perceptions. *Frontiers in Psychology, 9*, 1309.
- Schyns, B., & Schilling, J. (2013). How bad are the effects of bad leaders? A meta-analysis of destructive leadership and its outcomes. *The Leadership Quarterly, 24*(1), 138-158.
- Seo, Y., & Chung, S. W. (2019). Abusive supervision, psychological capital, and turnover intention: Evidence from factory workers in China. *Relations Industrielles / Industrial Relations, 74*(2), 377-404.

- Shaffer, J. A., DeGeest, D., & Li, A. (2016). Tackling the problem of construct proliferation: A guide to assessing the discriminant validity of conceptually related constructs. *Organizational Research Methods, 19*(1), 80-110.
- Shao, P., Li, A., & Mawritz, M. (2018). Self-protective reactions to peer abusive supervision: The moderating role of prevention focus and the mediating role of performance instrumentality. *Journal of Organizational Behavior, 39*(1), 12-25.
- Shaver, J. M. (2020). Causal identification through a cumulative body of research in the study of strategy and organizations. *Journal of Management, 46*(7), 1244-1256.
- Shum, C., Gatling, A., & Tu, M.-H. (2020). When do abusive leaders experience guilt? *International Journal of Contemporary Hospitality Management*.
- Siddaway, A. P., Wood, A. M., & Hedges, L. V. (2019). How to do a systematic review: A best practice guide for conducting and reporting narrative reviews, meta-analyses, and meta-syntheses. *Annual Review of Psychology, 70*, 747-770.
- Sieweke, J., & Santoni, S. (2020). Natural experiments in leadership research: An introduction, review, and guidelines. *The Leadership Quarterly, 31*(1), 101338.
- Simon, L. S., Hurst, C., Kelley, K., & Judge, T. A. (2015). Understanding cycles of abuse: A multimotive approach. *Journal of Applied Psychology, 100*(6), 1798.
- Smallfield, J., Hoobler, J. M., & Kluepfer, D. H. (2020). How team helping influences abusive and empowering leadership: The roles of team affective tone and performance. *Journal of Organizational Behavior, 41*(8), 757-781.
- Steel, P., Beugelsdijk, S., & Aguinis, H. (2021). The anatomy of an award-winning meta-analysis: Recommendations for authors, reviewers, and readers of meta-analytic reviews. *Journal of International Business Studies, 52*(1), 23-44.
- Sturm, R. E., & Antonakis, J. (2015). Interpersonal power: A review, critique, and research agenda. *Journal of Management, 41*(1), 136-163.
- Sungu, L. J., Hu, E., & Weng, Q. (2020). Goal Commitment Buffers the Negative Effects of Perceived Abusive Supervision. *The Journal of Psychology, 154*(4), 273-291.
- Taylor, S. G., Griffith, M. D., Vadera, A. K., Folger, R., & Letwin, C. R. (2019). Breaking the cycle of abusive supervision: How disidentification and moral identity help the trickle-down change course. *Journal of Applied Psychology, 104*(1), 164.
- Tepper, B. J. (2000). Consequences of abusive supervision. *Academy of Management Journal, 43*(2), 178-190.
- Tepper, B. J. (2007). Abusive supervision in work organizations: Review, synthesis, and research agenda. *Journal of Management, 33*(3), 261-289.
- Tepper, B. J., Henle, C. A., Lambert, L. S., Giacalone, R. A., & Duffy, M. K. (2008). Abusive supervision and subordinates' organization deviance. *Journal of Applied Psychology, 93*(4), 721.
- Tepper, B. J., Moss, S. E., & Duffy, M. K. (2011). Predictors of abusive supervision: Supervisor perceptions of deep-level dissimilarity, relationship conflict, and subordinate performance. *Academy of Management Journal, 54*(2), 279-294.
- Tepper, B. J., Simon, L., & Park, H. M. (2017). Abusive supervision. *Annual Review of Organizational Psychology and Organizational Behavior, 4*, 123-152.
- Tillman, C. J., Gonzalez, K., Crawford, W. S., & Lawrence, E. R. (2018). Affective responses to abuse in the workplace: The role of hope and affective commitment. *International Journal of Selection and Assessment, 26*(1), 57-65.
- Tröster, C., & Van Quaquebeke, N. (2020). When victims help their abusive supervisors: The role of LMX, self-blame, and guilt. *Academy of Management Journal*.
- Tu, M.-H., Bono, J. E., Shum, C., & LaMontagne, L. (2018). Breaking the cycle: The effects of role model performance and ideal leadership self-concepts on abusive supervision spillover. *Journal of Applied Psychology*.

- Vogel, R. M., & Mitchell, M. S. (2017). The motivational effects of diminished self-esteem for employees who experience abusive supervision. *Journal of Management*, 43(7), 2218-2251.
- Walter, F., Lam, C. K., Van Der Vegt, G. S., Huang, X., & Miao, Q. (2015). Abusive supervision and subordinate performance: Instrumentality considerations in the emergence and consequences of abusive supervision. *Journal of Applied Psychology*, 100(4), 1056.
- Wang, G., Van Iddekinge, C. H., Zhang, L., & Bishoff, J. (2019). Meta-analytic and primary investigations of the role of followers in ratings of leadership behavior in organizations. *Journal of Applied Psychology*, 104(1), 70.
- Wang, R., & Jiang, J. (2015). How abusive supervisors influence employees' voice and silence: The effects of interactional justice and organizational attribution. *The Journal of Social Psychology*, 155(3), 204-220.
- Watkins, T., Fehr, R., & He, W. (2019). Whatever it takes: Leaders' perceptions of abusive supervision instrumentality. *The Leadership Quarterly*, 30(2), 260-272.
- Wei, F., & Si, S. (2013). Tit for tat? Abusive supervision and counterproductive work behaviors: The moderating effects of locus of control and perceived mobility. *Asia Pacific Journal of Management*, 30(1), 281-296.
- Wu, T.-Y., & Hu, C. (2009). Abusive supervision and employee emotional exhaustion: Dispositional antecedents and boundaries. *Group & Organization Management*, 34(2), 143-169.
- Xu, E., Huang, X., Jia, R., Xu, J., Liu, W., Graham, L., & Snape, E. (2020). The "evil pleasure": Abusive supervision and third-party observers' malicious reactions toward victims. *Organization Science*.
- Yammarino, F. J., Cheong, M., Kim, J., & Tsai, C.-Y. (2020). Is leadership more than "I like my boss"? In *Research in Personnel and Human Resources Management*: Emerald Publishing Limited.
- Yu, L., & Duffy, M. K. (2020). The whiplash effect: The (moderating) role of attributed motives in emotional and behavioral reactions to abusive supervision. *Journal of Applied Psychology*.
- Yu, L., Duffy, M. K., & Tepper, B. J. (2018). Consequences of downward envy: A model of self-esteem threat, abusive supervision, and supervisory leader self-improvement. *Academy of Management Journal*, 61(6), 2296-2318.
- Yu, Y., Xu, S., Li, G., & Kong, H. (2020). A systematic review of research on abusive supervision in hospitality and tourism. *International Journal of Contemporary Hospitality Management*.
- Zellars, K. L., Tepper, B. J., & Duffy, M. K. (2002). Abusive supervision and subordinates' organizational citizenship behavior. *Journal of Applied Psychology*, 87(6), 1068.
- Zhang, & Bednall, T. C. (2016). Antecedents of abusive supervision: A meta-analytic review. *Journal of Business Ethics*, 139(3), 455-471.
- Zhang, J., & Liu, J. (2018). Is abusive supervision an absolute devil? Literature review and research agenda. *Asia Pacific Journal of Management*, 35(3), 719-744.
- Zhang, Y., & Liao, Z. (2015). Consequences of abusive supervision: A meta-analytic review. *Asia Pacific Journal of Management*, 32(4), 959-987.
- Zhang, Y., Liu, X., Xu, S., Yang, L.-Q., & Bednall, T. C. (2019). Why abusive supervision impacts employee OCB and CWB: A meta-analytic review of competing mediating mechanisms. *Journal of Management*, 45(6), 2474-2497.