Non-Disclosure Agreements and Externalities from Silence*

Jason Sockin
IZA

Aaron Sojourner
W. E. Upjohn Institute

Evan Starr
University of Maryland

January 18, 2024

Abstract

Do employment restrictions which prohibit workers from disclosing misconduct at work, which we refer to as 'broad non-disclosure agreements' (NDAs), distort labor markets? We develop a framework in which the legal risk from violating a broad NDA reduces worker willingness to share negative information about their employers, making it more difficult for high-quality employers to differentiate themselves to jobseekers. Changes in the content of Glassdoor reviews following the passage of three state laws that prohibited employers from using NDAs to conceal unlawful workplace conduct support this idea. By curtailing the flow of negative information, broad NDAs impose negative externalities on jobseekers who value such information and on competing employers who are less able to stand out.

JEL: M55, K31, J58
Keywords: Imperfect Information, Non-Disclosure Agreements, Firm Reputation, Externalities.

*We would like to thank Pierre Azoulay, Alan Benson, Laura Boudreau, Zoe Cullen, David Hoffman, Matthew Knepper, Ioana Marinescu, Kyle Rozema, and Anna Stansbury for helpful comments and feedback. We would also like to thank seminar participants at RAND, Duke University, the University of Southern California, LMU Munich, IZA Workshop on Labor Market Institutions, 4th IDSC of IZA/CAIS Workshop Matching Workers and Jobs Online, 14th Annual People and Organizations Conference, Columbia University Non-Market Effects of Market Power, Georgetown University Strategy and the Business Environment Conference, LERA 73rd Annual Meeting, 2022 Strategy Science Conference, 2024 ASSA Meeting, 4th Annual Washington Area Labor Economics Symposium, 15th Conference on Empirical Legal Studies, the NYU Empirical Contracts Conference, the Society of Labor Economists, the Society for Institutional & Organizational Economics, the Academy of Management, the Strategic Management Society, Yildiz Technical University, the W. E. Upjohn Institute, and the University of Maryland. All mistakes are our own.

"More and more employers fight back with what you signed on the dotted line: The contractual agreements you sign when you onboard as an employee are being used ... to silence negative feedback from employees." – Torres (2019)

1 Introduction

Policy, media, and social interest in employment terms that restrict workers from sharing information about misconduct at work has risen in recent years. We refer to these employment terms as 'broad' non-disclosure agreements (NDAs), although they are also referred to as hush contracts, gagging clauses, concealment contracts, confidentiality agreements, and non-disparagement agreements. Whereas NDAs are often viewed as restricting workers from sharing trade secrets, *broad* NDAs also restrict workers from sharing information beyond trade secrets that may adversely affect the company. Prominent cases of employers using broad NDAs to hide management misconduct—including at Theranos (Rogal, 2020) and the Weinstein Company (McShane, 2021)—exemplify a larger pattern of firms strategically using broad NDAs as a condition of employment to chill voluntary sharing of negative reports about the firm and threatening such volunteers with litigation (Grensing-Pophal, 2019; Meyer, 2019; Torres, 2019).[1] Almost 90% of U.S. employers report using NDAs, almost 60% of U.S. workers report being bound by one (Balasubramanian et al., 2021), and studies of actual NDAs suggest that most of them are written very broadly (Hrdy and Seaman, 2023).

A fundamental concern about broad NDAs is that they may perpetuate misconduct at work by silencing workers with negative information to share (Griffith, 2021). Legal scholars have articulated this concern for decades (Short, 1998; Bast, 1999; Philip, 2002; Lobel, 2016). The #MeToo movement underscored it vividly (Carlson, 2019). Such silencing could impose negative externalities on those who value the suppressed information, including jobseekers who would act on it in their job searches, competing employers who struggle to credibly

[1] For example, Meyer (2019) describes a recent case where a law firm required all newly hired employees to sign agreements not to 'criticize, ridicule, disparage,' the firm or its employees, and later sued employees who posted negative reviews on Glassdoor. Apple also recently amended its 'concealment clauses' following pressure from investors to allow workers to speak freely about harms within the workplace (Rella, 2022). For more examples of broad NDAs, see Section 2.

differentiate themselves (Bryan et al., 2022; Benson et al., 2020; Royle, 2023), and customers who prefer to support employers that promote their values. In the wake of the #MeToo movement, policymakers in several countries and U.S. states proposed—and some passed—new laws prohibiting firms from using broad NDAs to conceal unlawful conduct (Johnson et al., 2019; Topping, 2021), including on December 7, 2022 when President Biden signed the Speak Out Act (Habeshian, 2022), which prohibits firms from using NDAs signed as a condition of employment to conceal claims of illegal sexual harassment or assault.[2]

Despite decades of concern among legal scholars, growing media and social interest, and recent policy decisions, no prior research has quantitatively examined how broad NDAs affect labor markets. Several obstacles prevented such work. Representative data on NDA prevalence only became available recently. Further, because broad NDAs may both improve firm quality by protecting trade secrets (their traditional role) and artificially prop-up firm reputation by censoring negative information, distinguishing between these channels requires an unexpected change to what NDAs can cover, while maintaining trade secret protections. As Hoffman and Lampmann (2019) note, "...to know what hush contracts do ... the gold standard test would be to find a legal regime that switched from enforcement to nonenforcement .... To date, such a natural experiment has been unavailable."

This paper takes a first step toward filling this gap. To set the stage, we develop a simple framework in which broad NDAs limit negative information flows by enabling an employer to threaten litigation against workers who share negative information, which in turn bolsters the reputations of firms offering lower-quality jobs relative to higher-quality ones. Our empirical work examines these predictions by harnessing the first nationally-representative data on NDA prevalence (Balasubramanian et al., 2021) alongside recent state policy changes that

[2]The bill does not narrow NDAs for other illegal actions (violations of workers' right to occupational safety and health, wage and hour standards, organizing rights, freedom from discrimination in hiring, pay, or promotion, etc.) nor for any lawful behavior. On October 28, 2022, the Department of Defense also instituted a new policy that it would not award any contracts "to an entity that requires its employees to sign internal confidentiality agreements or statements that would prohibit or otherwise restrict its employees from lawfully reporting waste, fraud, or abuse related to the performance of a DoD contract." See https://www.federalregister.gov/documents/2022/10/28/2022-23281/defense-federal-acquisition-regulation-supplement-prohibition-on-award-to-contractors-that-require.

'narrowed' NDAs, meaning that the laws began prohibiting employers from using NDAs to conceal unlawful activity but continued allowing NDAs to protect trade secrets. The paper's primary triple-differences design tests if and to what extent an NDA-narrowing policy affects outcomes differently in states that narrowed NDAs relative to states that did not, between industries where NDAs are more prevalent and those where they are less. The main analysis considers how narrowing NDAs affects workers' disclosure of negative experiences at work, public measures of firm reputation, and whether high-job-quality firms become better able to differentiate themselves to jobseekers.

This study provides the first credibly causal evidence substantiating concerns about the negative externalities broad NDAs create by silencing workers. Narrowing NDAs increased the flow and value of negative information about employers, measured by the content of workers' reviews of employers written on Glassdoor. Average firm ratings on Glassdoor fell by 5%, and not all employers' ratings fell equally; rather, dispersion of firm ratings within local labor markets increased after NDAs were narrowed, consistent with a theoretical framework in which broad NDAs artificially bolster the reputation of lower job-quality employers. Evidence on changes in reviewing workers' choices to conceal aspects of their identity suggests reduced retaliation risk is a key mechanism. These results are robust to different control groups, within-individual designs, double-difference designs, other measures of NDA use, alternative ways of handling standard errors, and several alternative explanations. We also document the role of news media in mediating these results. Last, we replicate these results in two other contexts: in a case study of Theranos in which the public revelation of fraud reduced retaliation risk from broad NDAs and in sexual harassment complaints to the U.S. Equal Employment Opportunity Commission (EEOC).

From a policy perspective, these results are consequential because the employer and employee who sign an NDA are unlikely to internalize its costs on jobseekers and competing employers. This is because: (i) NDAs are often intentionally required as a condition of employment before a new hire sees information on hard-to-observe aspects of job quality

(Grensing-Pophal, 2019; Meyer, 2019; Eaton, 2021), (ii) affected parties, including future jobseekers and competiting employers, are not involved in the bilateral contracting process, and (iii) it is hard for outside parties to price information of which they are not aware. Thus, broad NDAs are likely oversupplied and negative information on employers undersupplied, such that policies which enable workers to speak freely—and that limit the ability of firms to restrict disparaging speech generally (as opposed to only unlawful activity)—could improve welfare.[3] In documenting such externalities, we add to other evidence that private contracting can induce broad externalities, e.g., in the context of noncompete agreements (Starr et al., 2019; Johnson et al., 2020).

This work builds on the exit, voice, and loyalty framework of Hirschman (1970).[4] Broad NDAs force post-exit "loyalty" by limiting worker voice even after exit, which in turn imposes costs on 'anyone who would care to listen' (Hirschman (1970), p. 4). A more recent literature emphasizes alternative ways to incentivize individuals to reveal negative information when doing so might invite retaliation (Sockin and Sojourner, 2023)—such as paying whisteblowers to come forward (Dey et al., 2021), increasing workers' outside options (Dahl and Knepper, 2021), or 'hard garbling' survey designs (Boudreau et al., 2022). In contrast, this study suggests that firms use restrictive employment terms to suppress negative information and that regulating such restrictions can increase workers' willingness to supply more-accurate, more-negative information than they otherwise would.

These findings also build on research about asymmetric information in labor markets vis-

[3]Implicit in this assessment of welfare is that new information is supplied truthfully. If the additional negative information is disingenuous, then externalities may be exacerbated as jobseekers struggle even more to assess job quality offered by prospective employers. We do not believe this to be the case for three reasons. One, defamation laws (e.g., libel, slander) already give firms the opportunity to punish workers for lying and these policies did not change when states narrowed NDAs. Two, the legislation either implicitly or explicitly permits "truthful statements or disclosures." Three, if workers were more likely to supply false information, they would have an incentive to also increasingly conceal their job titles to limit their employer's ability to identify and sue them. Narrowing NDAs would have increased volunteers' rate of job title concealment. However, the opposite occurred, suggesting that the new information is not more disingenuous.

[4]This literature typically considers collective voice (Mowbray et al., 2015) or workers expressing their dissatisfaction with management (Morrison, 2014). See also Harju et al. (2021) for a study of how voice affects separations and workplace quality.

à-vis firm reputation,[5] reputation management,[6] and labor market sorting.[7] Respectively, these literatures highlight the importance of reputation as a meaningful asset for firms, the strategies firms undertake to maintain their reputations (e.g., advertising or corporate social responsibility), the tendency for reputations, especially in the digital economy, to become inflated (Filippas et al., 2020; Nosko and Tadelis, 2015), and the importance of non-wage amenities for labor market sorting. This paper is unique in highlighting how broad NDAs create labor market imperfections (Manning, 2013; Schubert et al., 2022) which function as a reputation-preserving device for firms, thereby inhibiting high-road employers from standing out and hindering workers from sorting ex ante on firm quality.

2 Institutional Background

NDAs are one in a class of employment provisions known as restrictive covenants, which restrict what workers can do during and after an employment relationship. For example, Lobel (2021) notes that "a typical employment contract ...regularly includes multiple restrictive clauses—a boilerplate bundle that restricts postemployment competition through noncompete, non-disclosure, non-solicit, non-poaching, non-dealing, innovation assignment, and non-disparagement clauses." These restrictions can be agreed to at the outset of employment, while employed, or in severance arrangements or settlement agreements.

Evidence on the use of NDAs agreed to while employed was largely unavailable until recently. In the first detailed empirical analysis of NDAs—using data on 33,000 workers and 1,800 U.S. firms from 2017 (before the laws we study here were enacted)—Balasubramanian

[5]In particular, while a large body of literature focuses on asymmetric information related to worker quality, a growing literature considers similar issues related to firm reputation (Bryan et al., 2022). For other examples, see Diamond (1989); Tadelis (1999); Carmichael (1984); Shapiro (1983); McDevitt (2011); Cabral and Hortaçsu (2010); Liu and Shankar (2015); Turban and Cable (2003); Lange et al. (2011); Benson et al. (2020); Filippas et al. (2020); Gadgil and Sockin (2020).

[6]See, for example, Melo and Garrido-Morgado (2012); Lii and Lee (2012); Barrage et al. (2020); Lloyd-Smith and An (2019); Akey et al. (2021).

[7]See, for example, Sorkin (2018); Sullivan and To (2014); Maestas et al. (2023); Sockin (2022), and see Batut et al. (2021); Folke and Rickne (2022); Adams-Prassl et al. (2022) as it relates to sorting related to sexual harassment.

et al. (2021) find that NDAs are the most common restrictive covenant, covering 57%[8] of the workforce and in use by 88% of firms for at least some workers.[9] They also show that NDAs are the baseline restrictive covenant: if a worker has (or the firm uses) a noncompetition or nonsolicitation agreement, there is at least a 95% chance that the firm also uses an NDA. Finally, they suggest that NDAs are on average associated with 6% higher earnings, but that this positive relationship is mitigated when the worker has agreed to other restrictions including noncompete and nonsolicitation agreements.

The classic purpose of NDAs is to prohibit the use or disclosure of trade secrets. Indeed, professional, scientific, and technical services has the highest use of NDAs at 70% of workers (Balasubramanian et al., 2021). But legal researchers and practitioner organizations emphasize that the limits on what workers can share has broadened markedly. For example, National Women's Law Center (2020) notes that "since the 1980s, companies have broadened the use of nondisclosure agreements to prohibit workers from speaking up about a range of workplace conditions, including harassment, discrimination, and other violations of worker rights." Lobel (2018) similarly notes that NDAs are "often broadly worded to protect against speaking up against corporate culture or saying anything that would portray the company and its executives in a negative light." Flanagan and Gerstein (2019) emphasize that NDAs often "purport to protect information that is otherwise public, discoverable, or would not otherwise seem to be particularly confidential." A recent review of NDAs collected from trade secret litigation similarly finds that many "encompass publicly available or generally known information" (Hrdy and Seaman, 2023). Broad NDAs can also outright prohibit workers from *truthfully* describing harms at work—including to coworkers—either in separate non-disparagement agreements or via language within an NDA itself.

There are several clear examples of these direct prohibitions on disparagement as well as

[8]Note that a prior study, which was not designed as effectively to capture NDA use, estimated that approximately one-third of U.S. workers were bound by NDAs (Starr et al., 2021).

[9]One limitation of the data from Balasubramanian et al. (2021) is that they do not directly measure non-disparagement agreements. Rather, their surveys refer to restrictions related to sharing the employer's confidential information. There is no representative data on the use of non-disparagement agreements.

descriptions of 'confidential information' that are so broad as to include unlawful misconduct at work. For example, employees at Juvly Aesthetics, a beauty salon, were required to agree to the following as a condition of employment:[10]

> "... disparaging statements concerning individuals within management, other employees, or The Company will be considered insubordination, both in writing, in person, on social media, *on review sites* and on all other venues and are prohibited." (emphasis added)

Similarly, the Stange Law firm required all newly hired support staff and attorneys to sign an employment agreement which stipulated that ,

> "[D]uring and after Employee's employment or association with Law Firm ends, for any reason, Employee will not in any way criticize, ridicule, disparage, libel, or slander Law Firm, its owners, its partners, or any Law Firm employees, either orally or in writing" (Meyer, 2019).

In addition to separate agreements not to disparage the company, NDAs can simply broaden 'confidential information' to prohibit workers from sharing about many aspects of their work environment (Lobel, 2016). One prominent example is the NDA language in a Weinstein Company employment contract.[11] This contract broadly barred employees from disclosing trade secrets and confidential information, where the latter included,

> "...any confidential, private, and/or non-public information obtained by Employee during Employee's employment with the Company concerning the *personal, social, or business activities* of the Company, the Co-Chairmen, or the executives, principals, officers, directors, agents, employees of, or contracting parties ...with, the Company." (emphasis added)

The Weinstein NDA may seem like a potential outlier, given the firm's prominence in the #MeToo movement, but defining confidential information broadly is easy and seemingly standard practice (Hrdy and Seaman, 2023); indeed lawyers recommend firms do so in onboarding material (Grensing-Pophal, 2019). Moreover, standard employee NDA templates

[10] See Consolidated Complaint in Juvly Aesthetics, Case No. 09-CA-300239, available at https://foiaonline.gov/foiaonline/action/public/submissionDetails?trackingNumber=NLRB-2023-002116&type=Request.

[11] The authors obtained this contract confidentially.

(free at https://nondisclosureagreement.com/employee.html) include the following language in their definition of 'confidential information'

> "(e) any other information not generally known to the public which, if misused or disclosed, could reasonably be expected to adversely affect Company's business."

At the time of this writing, this NDA template has been downloaded 272,819 times. As a real life example of this language, Schwans Home Services, Inc. included language within their NDA which prohibited their workers from sharing confidential or proprietary information that would hurt the employer:[12]

> "Employee shall neither directly nor indirectly (i) disclose to any person not in the employ of Employer any Confidential or Proprietary Information, or (ii) use any such information to the Employee's benefit, the benefit of any third party or [e]mployer, or *to the detriment of Employer*" (emphasis added).

How many NDAs contain these broad restrictions? Studying 446 contracts drawn from employee-involved litigation, Hrdy and Seaman (2023) find that 96% of NDAs protect confidential information that extends beyond trade secrets, and that nearly 40% do not exclude any type of information from the NDA. Notably, 60% of the NDAs exclude information that is publicly known or became public after the agreement entered into effect. Further, 93% had no duration limit and less than 1% had a geographic limit, such that most of these NDAs are designed to stay in effect for perpetuity. Though a selected sample, these estimates, along with the standardized template, suggest that many, if not most, NDAs are quite broad.

To what extent are broad confidentiality restrictions in NDAs enforceable in court? Traditionally, courts have applied a reasonableness test towards enforcing NDAs and similar restrictions, balancing the firm's need for protection against the harms to the individuals and public (Bast, 1999; Lobel, 2016; Short, 1998). Notably, among similar restrictions, legal scholars and practitioners suggest that courts have been more willing to enforce NDAs (see Appendix Figure E1). This willingness derives from the presumption that NDAs reflect both

[12] See *Schwans Home Serv., Inc., 364 N.L.R.B. No. 20 2016* (p. 6).

parties' freedom to contract,[13] as well as the idea that NDAs "restrain trade, if at all, to a much less extent than do covenants not to compete" (Decker, 1993).

What is less clear, however, and of significant debate among legal scholars, is whether courts should enforce restrictions on confidentiality that go well beyond trade secrets.[14] For example, Short (1998) describes two cases in which employees with detailed knowledge of harm stemming from lack of safety of the firm's products (tobacco and vehicles) were prohibited from sharing such information due to broad NDAs signed as part of a settlement. The problem, as Bast (1999) notes, is that an NDA "may satisfy the employer's needs, but not the needs of the employee nor the needs of society." In theory, courts can deem an NDA unenforceable ex post based on the public's interest, or because it violates the National Labor Relations Act (Hoffman and Lampmann, 2019). Regardless of the enforceability of their NDAs, workers may still have federal protections as whistleblowers (Hrdy and Seaman, 2023), and can still report claims to the EEOC,[15] although "the bar for proving sexual harassment under Title VII is extremely high" (Ence, 2019).[16]

Despite the potential for courts to rule that an NDA is unenforceable (on public policy or other grounds), or that individuals may still make claims to the EEOC or be whistleblowers, as a practical matter NDAs may nevertheless be effective in silencing workers. For example, Flanagan and Gerstein (2019) emphasize that it is not clear that workers are aware of the protections that they might have. Research suggests, for example, that workers tend to believe agreements they sign are enforceable (Prescott and Starr, 2021), and that even unenforceable contracts can be costly to get out of (Sullivan, 2009), such that the actual enforceability of the provision may matter much less than its existence (Starr et al., 2020).

Even unenforceable NDAs likely chill reporting behavior. In a 2016 report,[17] the EEOC emphasized that "6% to 13% of individuals who experience harassment file a formal com-

[13] As the Restatement (Second) of Contracts chapter 8 notes "In general, parties may contract as they wish, and courts will enforce their agreements without passing on their substance."

[14] See, e.g., the discussions in Hoffman and Lampmann (2019); Short (1998); Lobel (2016); Rogal (2020).

[15] See *EEOC v. Astra, 94 F.3d 738, 744–45 (1st Cir. 1996).*

[16] See also Hafiz (2017); Broden & Mickelsen (2018); Hemel (2017).

[17] See https://perma.cc/W6L5-CNVV.

plaint" because they fear disbelief, blame, or retaliation that could damage their career or reputation (Dahl and Knepper, 2021). These fears appear well-placed. Among those who 'vocally resisted mistreatment' in the workplace, 75% reported experiencing retaliation (Cortina and Magley, 2003). Flanagan and Gerstein (2019) summarize the challenges workers face as follows: "Violating a rule in a coercive contract is typically grounds for termination, yet it may be the only way for an employee to talk to others about conditions in the workplace or find out if the underlying contractual rule is legal." Workers can count on only very imperfect enforcement of their legal rights.

A confusing part of this debate is that much of the caselaw related to NDAs derives from settlement agreements (Short, 1998; Bast, 1999), although broad NDAs are likely more frequently deployed in an employment contract as a condition of employment (Lobel, 2021). For example, even in the Harvey Weinstein case, in which prominent settlement agreements precluded the victims of sexual harassment from speaking out (Perkins, 2022), the employees of the Weinstein Company could also not talk about the harms they witnessed at work without violating the NDA they agreed to as a new hire (Gibson, 2017). In a 2017 statement, the employees begged management to let them out of their NDAs so they could speak out about the harms they observed at work.[18] Similarly, a Theranos whistleblower who later declared that "fraud is not a trade secret" (Carreyrou, 2018) initially feared "Theranos' wrath if he violated the nondisclosure and confidentiality agreements he'd signed *when he was hired*" (Primeaux, 2019) (emphasis added). From a public policy standpoint, settlement agreements can themselves be a concern due to potential coercion, but broad NDAs signed as a condition of employment—before any harm has occurred—are more objectionable. New recruits are unaware of the potential harm they will experience or witness and are unlikely to be adequately compensated for their silence (Balasubramanian et al., 2021).

Law related to NDA enforcement remained largely stable until 2019 (Hoffman and Lampmann, 2019), when the #MeToo movement put a spotlight on how firms use NDAs to conceal

[18] See https://www.newyorker.com/news/news-desk/statement-from-members-of-the-weinstein-company-staff.

and perpetuate misconduct (Carlson, 2019; Silver-Greenberg and Kitroeff, 2020; Griffith, 2021). Many countries subsequently began to reconsider their enforcement policy regarding NDAs, including England, Wales, Canada, Ireland, Australia, and the United States (Topping, 2021; Harris, 2019; Perkins, 2022). In the United States, while many states passed measures protecting workers from sexual harassment in the workplace or regarding post-harassment settlement agreements, initially only California, Illinois, and New Jersey made it unlawful for firms to use NDAs (or other similar documents) in the employment context to conceal any unlawful activity (Johnson et al., 2019). Eventually, other states and localities would do the same,[19] and in December 2022, President Biden signed the Speak Out Act, prohibiting firms from using non-disparagement and non-disclosure agreements signed as a condition of employment to conceal sexual harassment at work.[20] Although the Speak Out Act is only limited to sexual harassment and not all unlawful conduct, President Biden's decision to sign the legislation bolstered the NDA policy debate, further raising the importance of understanding how broad NDAs impact labor market dynamics. To provide empirical evidence on such dynamics, we exploit the adoption of the broadest and earliest laws related to NDAs signed as a condition of employment. We review these laws below.

Beginning on January 1, 2019, California SB 1300 made it an "unlawful employment practice, in exchange for a raise or bonus, or as a condition of employment or continued employment" to require an employee to sign "a nondisparagement agreement or other document that purports to deny the employee the right to disclose information about unlawful acts in the workplace, including, but not limited to, sexual harassment."[21] While the bill does not explicitly reference non-disclosure agreements, they are included in the reference to 'other document[s]' and the bill is often discussed as narrowing NDAs (Johnson et al., 2019; Groff, 2018). The bill also prohibits retaliation against employees who do speak out in ways the new law allows. Importantly, the bill was designed to apply retroactively to all prior

[19] See https://www.jdsupra.com/legalnews/washington-state-s-new-law-on-ndas-and-8936165/. See also https://rb.gy/4lq53 for similar policy in Louisville.

[20] See https://www.congress.gov/bill/117th-congress/senate-bill/4524/text for the bill.

[21] See hhttps://rb.gy/wi8bz for the full bill.

NDAs, not just NDAs agreed to after the bill became law (Akopyan, 2019).

Similarly, Illinois Senate Bill 75, effective January 1, 2020, noted that "Any agreement, clause, covenant, or waiver that is a unilateral condition of employment or continued employment and has the purpose or effect of preventing an employee or prospective employee from making truthful statements or disclosures about alleged unlawful employment practices is against public policy, void to the extent it prevents such statements or disclosures, and severable from an otherwise valid and enforceable contract under this Act."[22] The bill also protects workers against retaliation, but it is not retroactive.

Lastly, beginning on March 18, 2019, New Jersey Senate Bill 121 stipulated that a "provision in any employment contract" that would prohibit current or former employees from revealing "the details relating to a claim of discrimination, retaliation, or harassment" is "against public policy and unenforceable" (Hughes and Nacchio, 2019).[23] The bill also protects workers against any retaliation and makes the party attempting to enforce a contract against public policy liable for attorneys' fees. The bill does not apply retroactively.

Comparing these three, the California law is the broadest in applicability because it specifically highlights non-disparagement agreements (Legittino, 2019), covers all unlawful acts in the workplace, and covers NDAs agreed to prior to the law's implementation (Akopyan, 2019). The Illinois law is less broad in that it only applies to 'unilateral' restrictions in new contracts after the effective date, carving out agreements for which there is some 'bargained-for-consideration.' The New Jersey law is the least broad because it only covers claims of discrimination, retaliation, or harassment, does not address other unlawful behavior, and is not retroactive. Although it is the least broad of the three, the New Jersey law is still much broader than all of the other laws passed during this time which focused almost exclusively on sexual harassment (Usenheimer et al., 2019).[24]

[22]See https://www.ilga.gov/legislation/publicacts/101/101-0221.htm for the full bill.

[23]See https://www.njleg.state.nj.us/2018/Bills/S0500/121_R2.PDF for the full bill.

[24]For more details on these three laws and the weaker laws that passed, see Johnson et al. (2019). Given some of the heterogeneity of these laws, in robustness checks, we examine each state individually and also compare the results with those states that passed laws that covered predominantly sexual harassment.

It is worth emphasizing two further points in relation to these laws. First, although they reduce the legal risk workers face for speaking out about illegal conduct at work and include protections against retaliation, workers may still face retaliation in some form. Second, none of these laws changed the ability of firms to use NDAs to protect trade secrets or confidential information; they only limited the ability of firms to use NDAs to hide unlawful conduct.

Finally, it is important to note that in addition to these laws and the intense media focus on NDAs in the wake of the #MeToo movement, several new organizations were founded—often by prominent survivors of sexual harassment during the #MeToo movement—to educate individuals about their rights under these laws and push for reform. These include Lionness (Griffith, 2021), which offers free consultations with lawyers about violating NDAs, as well as Lift Our Voices, Can't Buy My Silence, and Time's Up.[25] The American Civil Liberties Union ran an article entitled 'Is a Nondisclosure Agreement Silencing You From Sharing Your 'Me Too' Story? 4 Reasons It Might Be Illegal' (Roth, 2018). Taken together, even if broad NDAs were unenforceable before these laws took effect, these laws provided some clarity regarding their unenforceability, and media outlets and organizations sought to better educate individuals about their rights regarding NDAs.

3 Theoretical Framework

To provide some guidance as to how broad NDAs might affect information flows, firm reputation, and differentiation in labor markets, we build a model similar to those in Dahl and Knepper (2021) and Sockin and Sojourner (2023). Let the probability distribution of firm quality be $g(\mu)$. Suppose a worker at quality-μ firm has an experience $\sigma(\mu)$ and, when the

[25]See https://www.liftourvoices.org/, https://www.cantbuymysilence.com/, and https://timesupnow.org/newsroom/times-up-statement-on-employee-non-disclosure-agreements/.

worker is asked to rate their firm, they give rating s to maximize utility, given by:

$$U(s) = \underbrace{-\frac{1}{2}(s-\sigma(\mu))^2}_{Dishonesty\,Penalty} \; - \; \underbrace{\alpha c(s)}_{Retaliation\,Cost}$$

The first part of the utility function reflects a dishonesty cost. The worker suffers a quadratic penalty when their rating differs from their true experience. The second part reflects retaliation costs from a given rating. By convention we assume that the greater is s, the more positive is the rating of the firm (e.g., a better star rating on Glassdoor), such that higher ratings have lower retaliation costs ($c'(s) < 0$). In other words, firms retaliate when people give bad reviews. We also assume that retaliation costs are convex ($c''(s) > 0$), such that a marginally lower rating is associated with much higher retaliation cost when the rating is low than when it is high. The term $\alpha > 0$ reflects the firm's ability to retaliate by suing over a broad NDA. In turn, the narrowing of NDAs can be intepreted as reducing α.[26]

Differentiating $U(s)$ with respect to s and solving for the optimal rating s^* gives:

$$s^*(\sigma(\mu)) = \sigma(\mu) - \alpha c'(s^*(\sigma(\mu))) \tag{1}$$

Define the optimal amount of over-rating (e.g., the difference between the worker's rating and their true experience) as $w^*(\sigma(\mu)) \equiv s^*(\sigma(\mu)) - \sigma(\mu) = -\alpha c'(s^*(\sigma(\mu)))$. Then since $c'(s) < 0$ the optimal rating is to over-rate. Since the extent of over-rating depends only on the marginal retaliation cost, it follows directly that broad NDAs will increase over-rating

[26] Note that this setup is designed to mimic the experience of workers on Glassdoor, the main data source. Workers likely do not go to Glassdoor with the intention of supplying a review of their employer. Rather, they go to look at information about pay and ratings provided by others about their own or others' employers. After allowing the worker to look at a few webpages, Glassdoor requires the worker to share information about their employer before they can continue—they refer to this as their 'give-to-get' model (Marinescu et al., 2018). As long as workers want to see more information, they will provide a report and exercise discretion in what to report about their employer. It is in this sense that this setup mirrors the empirical setting. The retaliation risk is as described by Grensing-Pophal (2019), where the firm can then sue the worker for violating an NDA given their public rating. This differs from reporting a formal complaint to a federal agency like the EEOC, where there is a chance to win a financial settlement and the main choice is whether or not to report, not what to report.

by increasing the marginal retaliation cost of a bad rating: $\frac{\partial w^*(\sigma(\mu))}{\partial \alpha} = -c'(s^*(\sigma)) > 0$. This comparative static yields the model's first prediction: Broader NDAs increase over-rating about workers' experiences at their firms.

Next, consider how the extent of over-rating changes with a worker's experience:

$$\frac{\partial w^*(\sigma)}{\partial \sigma} = -\alpha c''(s^*(\sigma))\frac{\partial s^*(\sigma)}{\partial \sigma} < 0 \tag{2}$$

Given cost function convexity ($c''(s) > 0$), workers over-rate more when they have bad experiences than when they have good experiences.[27] That is, retaliation costs cause workers to rate a bad experience more similarly to a good experience. Thus, convex retaliation costs compress the difference in ratings across good and bad experiences. An increase in NDA broadness will naturally moderate this compression ($\frac{\partial^2 w^*(\sigma)}{\partial \sigma \partial \alpha} < 0$), in turn reducing the difference in optimal worker ratings between poor experiences and better ones.

To aggregate to a 'public reputation,' define the public reputation of firm with quality μ as the average rating of the workers at firms of that quality, $\bar{s}(\sigma(\mu))$. Consider the simplest case where each worker's experience perfectly matches the firm's quality: $\sigma(\mu) = \mu$. Then every worker's rating is identical, such that $\bar{s}(\mu) = s^*(\mu)$. Thus we can think about the distribution of public reputations of firms as $f(\bar{s}(\mu))$ and how it contrasts with the true quality distribution $g(\mu)$. These lead to the core propositions we seek to test in this paper: Based on Equation 1, it follows directly that NDA-induced retaliation costs increase the mean of the public reputation distribution above the mean of the quality distribution. And based on Equation 2, it follows that NDA-induced retaliation costs cause the variance of the public reputation distribution to be compressed relative to the true quality distribution.

[27]The term $s^*(\sigma)$ is also increasing in σ. Using the implicit function theorem to differentiate Equation 1 yields $\frac{\partial s^*(\sigma)}{\partial \sigma} = \frac{1}{1+\alpha c''(s^*(\sigma))}$. Because $c''(s) > 0$, then $\frac{\partial s^*(\sigma)}{\partial \sigma} > 0$.

3.1 Model Extensions

Several additions to the model are possible. For example, individuals may have different experiences with the firm. That is, we could assume a worker's experience equals the firm's true quality plus some worker-specific, mean zero noise term, $\sigma_i(\mu) = \mu + \epsilon_i$, where subscript i denotes worker i. We might also allow workers to learn from some of their co-workers experiences. Allowing for such within-firm dispersion makes the math more complicated, but does not change the fundamental workings of the model—rather it generates an additional prediction that, even within a firm, broad NDAs can compress the distribution of ratings by causing workers to self-censor when they have a bad experience.

Furthermore, the possibility of individualized experiences combined with only a random subset of workers rating their firm raises the possibility that public rankings scramble true quality. If the number of ratings at all firms is large, then the average rating within each firm stabilizes and the rank ordering of firm quality will match the rank ordering of firm reputation. If the number of ratings is small or moderate, however, then unique individual experiences can scramble public reputations to some degree.[28]

One may also consider expanding the model to allow a workers' report to be multidimensional. For example, throughout the empirical analysis, we consider several aspects of Glassdoor reviews that workers make decisions over. These include free-response descriptions about the workplace, providing identifying information such as one's job title, and an overall rating. To the extent that each of these dimensions entails some retaliation risk, the above results would all apply to extensions of this type. Similarly, the main results go through if retaliation risk is a product of these characteristics, which seems likely (e.g., the firm may be even more likely to retaliate when both the rating and content of the review is negative).

Additionally, one may endogenize the use of NDAs by assuming that NDAs are more likely to be used by low quality firms as tools to censor the firm's true reputation. Implicitly,

[28]Note that prior research suggests online reputation systems influence jobseeker behavior more for smaller and newer employers (Benson et al., 2020; Sockin and Sojourner, 2023), as in consumer markets (Luca, 2011).

the setup allows for this possibility given the assumption of a convex retaliation cost function. However, one would have to add a 'first stage' to the model where the worker would consider the utility from agreeing to an NDA versus not——including any downstream dishonesty costs from over-reporting the worker's experience. Effectively, this would add a compensating differential to the worker's utility function, though it would not affect downstream reporting behavior for workers with NDAs. We do not endogenize broad NDAs in general because (a) we are focused on externalities arising from information sharing and variation in public firm reputations, which arise regardless of whether individuals are paid for their silence, and (b) because prior research suggests that few workers actually review or negotiate their NDA.[29] Note, however, that if low quality firms are the only users or enforcers of NDAs, then the model suggests it will help low-road firms pool their reputations with high quality firms.

4 Data

To analyze how broad NDAs influence the information workers share about their employers and consequent employer reputation, we use several unique datasets. We use employer reviews submitted by workers on Glassdoor from January 2015 through June 2021. Although it is not random who submits reviews on Glassdoor, this dataset is perhaps the most relevant for examining the effect of broad NDAs on the flow of labor market information, both because it consists of reviews written by those who would be bound by NDAs—current and former employees—and because jobseekers commonly go there to learn about employers (as opposed to, e.g., Twitter). Royle (2023), for example, highlights that bad Glassdoor reviews are the number one 'red flag' for job seekers. Glassdoor's data collection process works through a 'give-to-get' policy: Individuals can initially access a few reviews until Glassdoor requires them to review their employer before they can continue looking at other reviews. Thus, the data come from individuals who have enough desire to continue looking at other information

[29]By comparison, Starr et al. (2021) find that only 10% of workers negotiate over their noncompete and 86% of workers simply read and sign a noncompete when presented with it. Since NDAs are less restrictive than noncompetes, these percentages are likely upper bounds for the same relationships with NDAs.

on Glassdoor—not necessarily people who came to Glassdoor to review their employer.

To measure information flows, we analyze the free-response text from the 'Pros' and 'Cons' fields of employer reviews.[30] To measure employer reputation, we leverage the worker's overall employer rating on a Likert scale of 1–5 stars, with more stars signaling more satisfaction. Each review also permits 1–5 stars ratings for five sub-categories (career opportunities, compensation and benefits, culture and values, senior management, and work-life balance), three (dis)approval options (CEO's performance, positive business outlook for the firm over the next six months, would refer a friend to the firm), and an option to provide advice to management. Volunteers are asked to provide their job titles and locations, but can leave these characteristics blank.[31] Summary statistics are provided in Appendix Table F1.

Although Glassdoor reviews are intended to be anonymous, workers may still worry about retaliation risk when deciding whether to volunteer a review (Sockin and Sojourner, 2023). Appendix A provides several pieces of evidence suggesting that volunteers face some identity disclosure risks. Lawyers encourage firms to use broad NDAs as a way to discourage workers from leaving negative reviews specifically on Glassdoor, by reminding workers about the restrictions they signed and the consequences of NDA violations as a means to get workers to remove the offending review (Grensing-Pophal, 2019; Meyer, 2019). While Glassdoor takes reviewer anonymity seriously and fights employers' legal attempts to force the company to unmask reviewers, many workers may be reluctant to consider sharing negative information. Being bound to an NDA may compound this reluctance. Such workers may simply refrain from spending the time to investigate the institutional details of an information sharing site like Glassdoor. Workers who investigate the risks they would take on by volunteering a review will find articles like those in the appendix that make such risks salient.

[30] Appendix Figure E2 provides a sample, blank review form. Note that the language asks reviewers to "not post ...trade secrets/confidential information".

[31] The option to conceal one's job title was not always available to every potential reviewer. Figure E4 displays trends in job-title concealment rates and shows a structural decline in the rate after December 2018, reflecting changes to the Glassdoor review submission form on some user platforms. This coincides with the enactment of California SB 1300 and takes place a few months prior to the enactment of New Jersey SB 121. Fortunately, similar shifts are observed among both low- and high-NDA-use industries across both treatment and control states, minimizing concern that this could threaten identification.

The other primary dataset we leverage is 2017 data from Payscale.com, from which we measure variation in the coverage of NDAs.[32] We aggregate the worker-level NDA data to calculate the average rate of NDA use by industry that we then merge with Glassdoor reviews. We use the industry level as opposed to the occupation or industry-occupation level, because, as we discuss later, volunteers may leave their occupation blank as a form of identity concealment to protect against retaliation risk (Sockin and Sojourner, 2023), while industry follows from the always-observed firm. Nevertheless, alternative measures using occupation reveal similar results, as do double-difference models that do not rely on variation in NDA intensity, which we document later in robustness checks. Harmonizing the Glassdoor and Payscale industry classifications results in fifteen industries. Table 1 lists these industries along with their respective shares of workers who report being bound by an NDA.

Table 1: NDA Intensity by Industry in 2017

Industry	NDA incidence
Accommodation and Food Services	44.0%
Agriculture	51.0%
Arts and Entertainment	50.1%
Construction	41.3%
Finance and Insurance	70.1%
Health Care and Social Assistance	55.8%
Information	65.3%
Manufacturing	57.2%
Mining	59.6%
Other Services	48.3%
Professional, Scientific and Technical Services	70.4%
Real Estate	51.9%
Retail Trade	50.5%
Transportation and Warehousing	50.7%
Utilities	66.0%

Notes: The table provides the incidence of NDAs by industry per the unweighted, individual-level Payscale data, which cover 33,000 workers. See Balasubramanian et al. (2021) for more details.

While Balasubramanian et al. (2021) suggest that the use of NDAs is driven primarily

[32] The data, developed in partnership with and discussed initially by Balasubramanian et al. (2021), derive from individual intake data collected by Payscale.com. In particular, individuals who visit the website can fill out information about themselves to receive an estimate of their earnings potential. In 2017, Payscale.com added a question on NDAs to their intake survey. Individuals were incentivized to provide accurate information because the validity of their earnings prediction depended on it. The survey question underlying this data asks workers if they had agreed with their current employer not to share confidential information.

by the value of trade secrets, it remains unknown how ex ante NDAs relate to potential wrongdoing at work. Based on Payscale measures of NDA incidence by industry, Glassdoor information flows and reputation, and a measure of workplace abuse by industry from the 2015 RAND American Working Conditions Survey, Figure E3 suggests that industries that use NDAs more often tend to look like better places to work. This correlation could arise because NDAs help make workplaces better, help conceal abuse, and/or are just associated with other desirable characteristics. Below we describe our main empirical approach, which examines how these correlations change after California, Illinois, and New Jersey 'narrowed NDAs', as described above.

5 Empirical Approach

Based on our theoretical model, the ideal experiment to estimate the causal effect of broad NDAs on how employees rate employers would be to take employees who have signed broad NDAs and then exogenously reduce the legal retaliation risk from the NDA (i.e., reduce α). Our empirical approach mimics this ideal experiment by leveraging variation in (a) how likely a worker is to have signed an NDA and (b) whether the worker's state narrows NDAs. We can further mimic this ideal experiment by pursuing within-worker analyses, which hold fixed idiosyncratic aspects of each worker (e.g., individual experiences, risk aversion, etc.) and only shock the external retaliatory threat via a change to the NDA regime. While this within-worker approach closely mimics the ideal experiment, it also excludes one-time reviewers, who comprise 90% of the Glassdoor sample and might be nudged into reviewing by the narrowing of NDAs. We thus pursue both within-worker and pooled models.

Double-difference and a triple-difference designs allow us to leverage variation in which types of workers are more or less likely to have NDAs and variation in which states narrowed NDAs. For example, we look within states before and after they narrowed NDAs and compare workers with a high likelihood of having an NDA to workers with a low likelihood. This within-state model nets out the effects of other state policies that might have been passed around the same time, to the extent that they have similar effects on workers who vary in their NDA use. This within-state differencing seems important given that some states did pass other laws to improve workplaces by reducing sexual harassment[33]—which otherwise push against finding that narrowing NDAs induces the disclosure of more negative information. That said, differential trends between jobs with low- and high-NDA usage within the state will bias these within-state estimates. Accordingly, we consider jobs with the same NDA intensity in states that did not narrow NDAs to difference out such differential trends. The resulting triple differences model is our main specification, though we also report all double difference models in Appendix G.

Accordingly, the main triple-difference specification is:

$$Y_{ikst} = \beta \times NarrowedNDAs_{st} \times nda_{\iota(k)} + \lambda_{st} + \lambda_{\iota(k)t} + \lambda_{ks} + \epsilon_{ikst} \tag{3}$$

where Y_{ikst} is an outcome for worker i employed at firm k in state s in calendar year–month t, $NarrowedNDAs_{st}$ is an indicator for whether one of the three NDA-narrowing laws was in effect in state s in year–month t, $nda_{\iota(k)}$ is the intensity with which NDAs are used in firm k's industry $\iota(k)$, λ_{st} are state-year-month fixed effects, $\lambda_{\iota(k)t}$ industry-year-month fixed effects, and λ_{ks} firm-state fixed effects. We also report models with individual fixed effects, as noted above. We two-way cluster the standard errors separately by state and industry, the two levels at which the key independent variables are assigned (Abadie et al., 2022). Results are robust to alternative methods for handling standard errors, including bootstrapping and randomization inference, where we repeatedly randomly choose three states and pretend they passed NDA-narrowing laws to construct a distribution of placebo effects against which we can compare our estimates.

In terms of control states, in a double-differences specification, we would want to find

[33] These policies are describe in Johnson et al. (2019) and include limiting forced arbitration and requiring new training. We use states that adopted these policies as a different control set in robustness checks.)

states whose trends would reflect the counterfactual trends the three treated states would have followed had they not narrowed NDAs. However, in a triple-differences specification, the identifying assumption is *not* parallel trends, but rather 'parallel biases' (Olden and Møen, 2022), which are differenced out by that third difference. Given the uncertainty in picking control states, we consider several alternatives. As a baseline, we include all untreated states as controls. Since some states also adopted legislation related to sexual harrassment in the wake of the #MeToo movement (Johnson et al., 2019)—but not related to NDAs—these states form another natural control set that addresses selection into *some* policy adoption (which we refer to as 'weak legislation states'). In Appendix B, we show our results are not sensitive to several other choices of control states. And while the identifying assumption is not parallel trends, we nevertheless present tests from Rambachan and Roth (2023) to document the robustness of our results to violations of parallel trends assumptions.

In the triple-differences specification, the coefficient of interest β is identified by comparing how: (i) the baseline relationship between industry NDA use and information flows or ratings (e.g., panels (a) and (b) of Figure E3) change, (ii) within the same firm-state after the enactment dates of laws narrowing NDAs, and (iii) in states that passed such legislation compared with states that did not. If NDAs have a suppressing effect on negative information flows, then β should be positive for the outcomes related to the disclosure of negative information and negative for outcomes related to firm reputation. The parameter $\frac{\beta}{100}$ describes the average effect of the legal change for a one percentage point increase in NDA intensity, such that β multiplied by the average level of NDA incidence (0.6) describes the average effect of prohibiting firms from using NDAs to conceal wrongdoing.[34]

6 Results

[34]This design nets out any effects of other laws that were passed simultaneously which affect all industries equally. The idea is that if initially NDAs cover 60% of the workers in an industry, then 0% of the NDAs can prohibit workers from sharing information related to unlawful conduct after NDAs are narrowed.

6.1 Narrowing NDAs and Negative Information Flows

Motivated by the theory, in this section we consider how narrowing NDAs affects the content individuals share about their employers and themselves in Glassdoor reviews.

6.1.1 Increase in Negative Information Flows in Glassdoor Reviews

We use the free-response text in Glassdoor reviews, where individuals can communicate the positive and negative aspects of the workplace, to measure information flows about employers. In Table 2, we examine several dependent variables reflecting the content of these fields. As a broad measure of negative information flows, we find that narrowing NDAs results in a 4.5% ($e^{0.074\times0.6}-1$) increase in the length of the cons field, while the cons' share of the review's total text increases by 1.7% ($\frac{0.16\times0.6}{0.570}$). As a placebo test, we find no evidence that the pros section increased in length. Thus, as an overall measure of negative information, narrowing NDAs results in individuals spending more and a greater share of effort elaborating on the downsides of working at their firms. The estimates are even larger when we consider only workers who submit multiple reviews and include worker fixed effects (Table F2). Figure 1 examines dynamic triple-differences event-study specifications for the log length of the cons section under both the benchmark (panel a), within-worker (panel b), and using weak legislation states (panel c) specifications. Across the three specifications, we observe fairly stable estimates in the pre-period, and once NDAs are narrowed, the average length of the cons section rises and remains elevated over the next 2.5 years.

Because these laws were passed in the wake of the #MeToo movement, and all of the laws deal with illegal behavior and sexual harassment in some way, in columns 4 and 5 of Table 2, we examine whether individual reviews are more likely to use language related to harassment and illegal behavior, respectively.[35] We then implement the same triple-

[35] For 'harassment,' we create an indicator equal to one if any of the following terms are mentioned in the cons field of the review: abus, assault, bully, bullied, harass, hostil, humiliat, innuendo, intimidat, mobb, sexual, stalk, threaten, victim, and violen. Similarly, we create an indicator for 'illegal' behavior which captures whether the cons section mentions the terms illegal, not legal, be legal, been legal, is legal, and was legal.Note, we chose these terms to avoid capturing legal as a noun, as 'legal' often describes the law offices

Table 2: Narrowing NDAs and Outcomes Related to Review Text

	Cons share of review text	Log length of Pros section	Log length of Cons section	References harassment in text	References illegality in text	Offers mgmt. advice
	(1)	(2)	(3)	(4)	(5)	(6)
Narrowed NDAs x NDA intensity	0.016**	-0.016	0.074***	0.712***	0.270*	2.855***
	(0.008)	(0.057)	(0.018)	(0.169)	(0.134)	(0.863)
Dependent variable mean	0.523	4.431	4.552	1.948	0.170	56.213
Observations	3645332	3645332	3645332	3644449	3644449	3645332
Adjusted R^2	0.11	0.25	0.15	0.02	0.03	0.12

Notes: The table conveys how the content of worker reviews changed following the narrowing of NDAs. The dependent variable in each regression is listed as the header of each column. For alluding to harassment or illegality in the review text, as well as offering management advice, estimates are reported in percentage points. Regressions include firm-state, industry-year-month, and state-year-month fixed effects. Standard errors are two-way clustered by industry and state. Significance levels: * 10%, ** 5%, *** 1%.

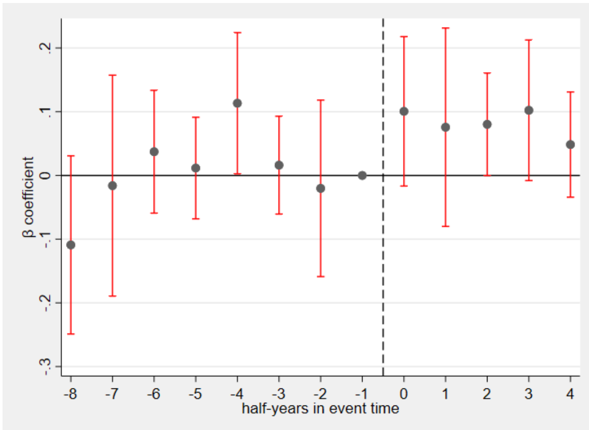
differences estimation to see if reviews discuss harassment or illegal conduct more frequently after these laws passed. Indeed, we observe that reviews indicating harassment increase by 22% ($\frac{0.71 \times 0.6}{1.95}$) on average after NDAs are narrowed, while reviews indicating illegal behavior increased 95% ($\frac{0.270 \times 0.6}{0.170}$). The outsized effect in percentage terms for illegal behavior reflects a low baseline rate, with only 0.17% of reviews mentioning one of these illegal terms, whereas 1.95% of reviews mention a term related to harassment. Finally, workers are 3% more likely to write advice for management (column 6).

6.2 Firm Reputation Falls on Average

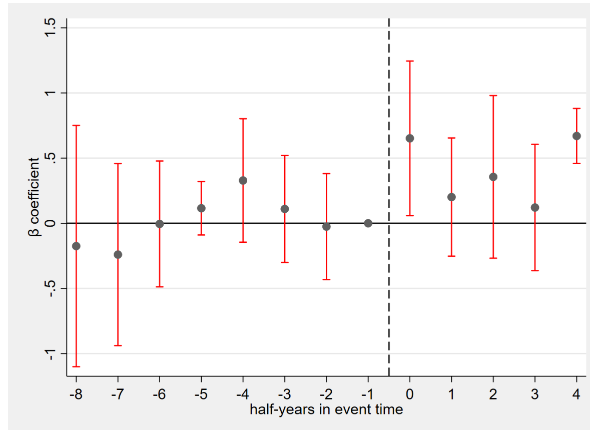
The prior analyses provide evidence that narrowing NDAs increases the flow of negative information about employers. One might expect that jobseekers who consume this information would use it when making labor market decisions. Negative Glassdoor reviews are a key red flag for jobseekers (Royle, 2023). But, if this negative information is difficult for jobseekers to find (e.g., buried within thousands of reviews), then it seems unlikely to matter for labor market decisions. Accordingly, we analyze whether this increased negative information flow leads to changes in firm reputation, which prior work has shown to both drive labor supply

within the firm. Accordingly, the search terms attempt to capture legal when used as an adjective.

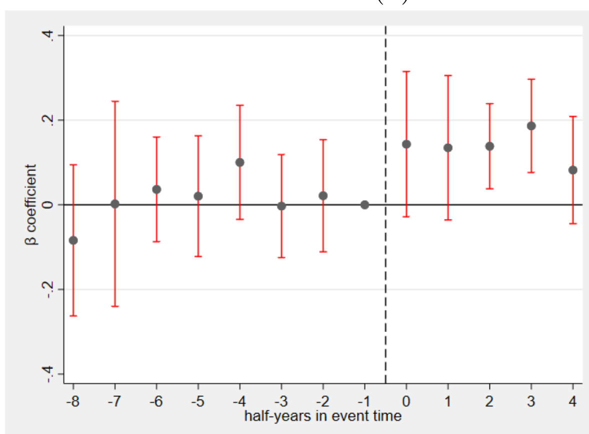
Figure 1: Narrowing NDAs and Negative Provision on Glassdoor, Dynamic Responses



(a) Benchmark



(b) Benchmark with worker fixed effects



(c) Weak legislation states

Notes: The dependent variable is the log length of the 'cons' section. The sample period is 2015–2021 and point estimates are relative to the calendar half-year before the legislation goes into effect. Regression includes firm-state, industry-year-month, and state-year-month fixed effects. Standard errors are two-way clustered by industry and state in panels (a) and (c) and by industry cross state in panel (b). Panels (a) and (b) include all untreated states as controls, where as the control states in Panel (c) are the weak legislation states. Red vertical bars indicate 95% confidence intervals around each point estimate.

(Benson et al., 2020; Sockin and Sojourner, 2023; Bryan et al., 2022) and change employment practices (Dube and Zhu, 2021).

Ex ante, it is not obvious whether narrowing NDAs should affect employer reputation, even if it spurs more negative information flows.[36] For example, employees with bad experience at work can rate their firms as one star (out of five) on Glassdoor, provide no other identifying information about themselves or their experiences, and face little retaliation risk,

[36]In extreme cases, e.g., the Harvey Weinstein scandal, the media likely tars the firm's reputation by itself.

both before and after NDAs are narrowed. In this case, while review text may become more negative after NDAs are narrowed, Glassdoor ratings of firms may not change. However, if retaliation risk is higher when both the content and the rating are negative,[37] then reducing retaliation risk by narrowing NDAs will reduce over-reporting on ratings too. Alternatively, if narrowing NDAs makes individuals more likely to speak up about unlawful conduct at work, then coworkers may become informed about their employer's bad behavior, potentially yielding more-negative reviews and ratings than would have otherwise been provided. Similarly, if narrowing NDAs changes the composition or increases the number of reviewers, as suggested above, then narrowing NDAs could feasibly result in lower average ratings, resulting from more one- and two-star ratings and fewer five-star ratings.

In Table 3, we report on the main triple-difference model to estimate how narrowing NDAs affects the distribution of ratings (i.e., the likelihood of 1, 2, 3, 4, or 5 stars) along with the average rating. Columns 1–5 show that there is a 16% ($\frac{0.039\times0.6}{0.143}$) increase in the share of one-star ratings, a 14% ($\frac{0.026\times0.6}{0.113}$) increase in two-star share, and a 7% ($\frac{0.022\times0.6}{0.189}$) increase in three-star share. Increases in the probability of having one rating category must be offset by decreases in the probability of another, hence a 14.5% ($\frac{0.078\times0.6}{0.323}$) decrease in the share of five-star ratings. In turn, average ratings fall 4.7% ($\frac{0.270\times0.6}{3.479}$) after NDAs are narrowed (Column 6). The estimate is similar for new reviews (column 7) and is approximately twice as large when worker fixed effects are incorporated (column 8). Consistent with the idea that individuals are generally more informed about their firm's wrongdoing, or that dissatisfied employees are now more likely to review, the decline in reported sentiment is observed among alternative measures of worker satisfaction (Appendix Table F3).

Figure 2 reports the dynamic responses of workers' overall ratings under the benchmark (panel a), within-worker (panel b), and weak legislation states (panel c) specifications, with the half-year before the laws are passed serving as the reference period. Notwithstanding

[37] In the model, we could allow a report to have two dimensions, s_1 and s_2, and the cost of retaliation may reflect both dimensions, $c(s_1, s_2)$, with the cross-partial derivative being negative. In practice, this might mean that an employer would be more likely to sue a worker for violating an NDA when they share negative information and give a low rating, relative to just a low rating or just negative information.

Table 3: Narrowing NDAs and the Distribution of Review Ratings

	Distribution of star ratings					Average star rating		
	One star	Two stars	Three stars	Four stars	Five stars	Full sample	New reviewers	Worker FE
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Narrowed NDAs x NDA intensity	0.039***	0.026***	0.022**	-0.010	-0.078**	-0.270***	-0.252***	-0.582***
	(0.013)	(0.002)	(0.010)	(0.006)	(0.027)	(0.079)	(0.044)	(0.220)
Dependent variable mean	0.143	0.113	0.189	0.232	0.323	3.479	3.507	3.233
Observations	3645332	3645332	3645332	3645332	3645332	3645332	3266773	290188
Adjusted R^2	0.09	0.02	0.04	0.03	0.16	0.15	0.15	0.49

Notes: The table shows the triple-differences estimates relaying how the distribution of newly submitted employer reviews shifted following the narrowing of NDAs. The dependent variable in each regression is a dummy for the specific star rating. Regressions include firm-state, industry-year-month, and state-year-month fixed effects. Standard errors are two-way clustered by industry and state in the first seven columns and industry-cross state in the last one. Significance levels: * 10%, ** 5%, *** 1%.

somewhat higher trends four years before the policies take effect, trends appear parallel in the three years before the policies take effect, followed by an immediate and persistent decline in average ratings throughout the post-legislation period of 2.5 years.

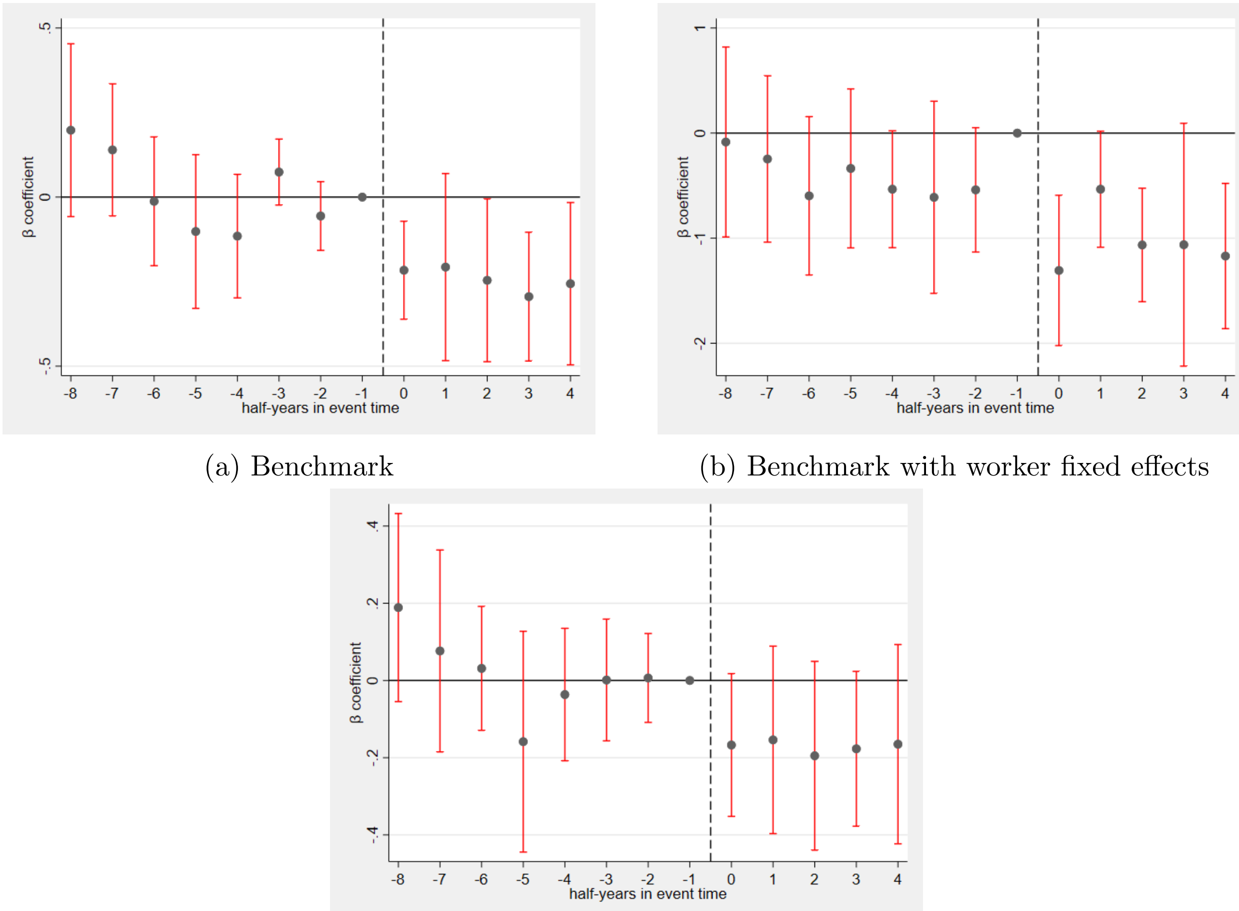
6.2.1 Extensive Margin and Heterogeneity Results

To assess the potential for any extensive margin effects, in this section we examine how narrowing NDAs affects the composition of reviewers, the volume of reviews, and heterogeneity in star ratings by characteristics of the firm or worker. Table 4 displays the triple-differences results reflecting how narrowing NDAs influence the composition of reviewers. While reviewers are similar along many characteristics, after NDAs are narrowed, reviewers are 5% ($\frac{0.037\times 0.6}{0.461}$) more likely to be female and 5% more likely to be long-tenured—both of which seem ex ante more likely to have negative information to share.

To see whether narrowing NDAs increases the volume of reviews, we estimate the triple-difference models using labor market (state-industry) by half-year review count as the outcome. Point estimates are consistent with an increase in the number of reviews, but the estimates are imprecise (Appendix Table F4).

We also examine several dimensions of heterogeneity to see how changes in the composition of reviewers might drive these reputation effects. For example, one might expect that

Figure 2: Narrowing NDAs and Employee Overall Ratings, Dynamic Responses



(a) Benchmark

(b) Benchmark with worker fixed effects

(c) Weak legislation states

Notes: The dependent variable is employee overall star rating. The sample period is 2015–2021 and point estimates are relative to the calendar half-year before the legislation goes into effect. Regression includes firm-state, industry-year-month, and state-year-month fixed effects. Standard errors are two-way clustered by industry and state in panel (a) and (c) and by industry cross state in panel (b). Panels (a) and (b) include all untreated states as controls, where as the control states in Panel (c) are the weak legislation states. Red vertical bars indicate 95% confidence intervals around each point estimate.

women would be more likely to rate their firm lower following the narrowing of NDAs, that current employees would be more affected than former employees, or that the effects would be greater for workers at small employers since smaller employers may be better able to identify those who violate an NDA. Appendix D details these ideas further. While many of the estimates are in line with our expectations, the estimates are surprisingly consistent across each sub-sample, with the only statistically significant difference coming from short-tenure workers rating their firms lower after NDAs are narrowed. We thus conclude that the compositional shifts observed among reviewers are likely not driving the results.

6.3 Dispersion in Firm Ratings Rises

This section tests the prediction from our theoretical framework that narrowing NDAs could increase ratings dispersion. To test this idea, we aggregate to the state-industry-half-year level and measure the inter-quartile range of ratings. We then deploy our triple-difference identification strategy, weighting each observation by the average number of firms represented in each state-industry over time. This ensures that states and industries with few firms are not driving our results. We then decompose the overall results into within-firm and cross-firm dispersion. The results are presented in Table 5.

Consistent with the model, once NDAs are narrowed, dispersion in overall ratings within a local labor market rises, with the inter-quartile range increasing 22% ($\frac{0.831\times 0.6}{2.24}$). How does this overall increase relate to dispersion within and across firms? Turning first to within-firm dispersion, if narrowing NDAs results in some individuals within the firm—but not others—becoming more informed about their employer's wrongdoing, then this could cause dispersion in ratings within the firm to increase. Alternatively, if everybody in the firm becomes informed and now shares the same experience, dispersion in ratings within the firm

Table 4: Narrowing NDAs and Composition of Reviewers

	Indicator for worker type				
	Current employee	Female	Short tenure	Ages 18–30	Manager position
	(1)	(2)	(3)	(4)	(5)
Narrowed NDAs x NDA intensity	-0.020	0.037**	-0.048*	-0.030	0.011
	(0.017)	(0.014)	(0.023)	(0.036)	(0.013)
Dependent variable mean	0.542	0.461	0.605	0.488	0.190
Observations	3645332	1761369	2884577	751233	3146120
Adjusted R^2	0.08	0.13	0.11	0.14	0.10

Notes: The table shows the triple-differences estimates relaying how the distribution of respondents' characteristics changed following the narrowing of NDAs. The dependent variable is an indicator variable for each worker type. Regressions include firm-state, industry-year-month, and state-year-month fixed effects. Standard errors are two-way clustered by industry and state. Significance levels: * 10%, ** 5%, *** 1%.

may fall. Column 2 of Table 5 shows that the narrowing of NDAs increases the inter-quartile range of within-firm ratings by 7% ($\frac{0.191 \times 0.6}{1.54}$).

Table 5: Narrowing NDAs and Interquartile Range of Ratings Within and Between Firms

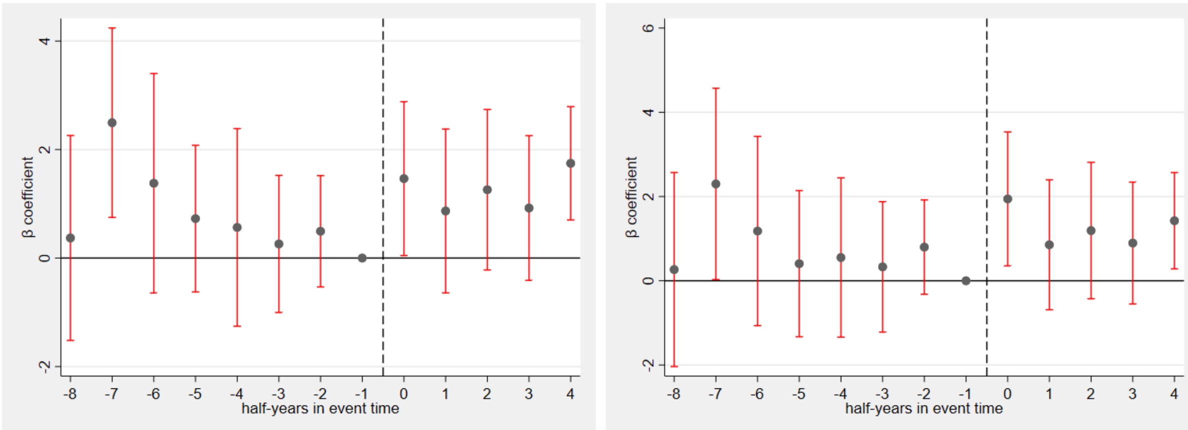
	Overall	Within firms	Across firms
	(1)	(2)	(3)
Narrowed NDAs x NDA intensity	0.831**	0.191***	0.463***
	(0.329)	(0.060)	(0.141)
Dependent variable mean	2.24	1.54	2.03
Mean firms per industry-state-half	145	160	149
Observations	9675	8759	9401
Adjusted R^2	0.54	0.20	0.53

Notes: The table implements triple-difference models for estimating the causal effect narrowing NDAs has on the interquartile range of overall ratings within and between firms. Each observation represents an industry-state-half year and is weighted by the average number of firms within each industry-state each half-year. Regressions include industry-state, industry-year-half, and state-year-half fixed effects. Standard errors are two-way clustered by industry and state. Significance levels: * 10%, ** 5%, *** 1%.

While within-firm dispersion may be indicative of different experiences within the firm, it is likely less important to job seekers than cross-firm dispersion in ratings. The theoretical framework suggests that broad NDAs especially inflate reputations of low-road employers, making it more difficult for high-road employers to distinguish themselves to jobseekers and for jobseekers to in turn recognize them. If, in contrast, narrowing NDAs causes all firms to receive more-negative reviews in equal measure, then while each firm's average rating would decline, neither dispersion in ratings across firms nor the ability of jobseekers to distinguish between firm would change. Column 3 of Table 5 shows that narrowing NDAs increases the interquartile range of ratings across firms by 14% ($\frac{0.463 \times 0.6}{2.03}$).

Figure 3 displays the dynamic effects for the across-firm measure of ratings dispersion. While there is some noise 3.5 years before the policies are passed, dispersion in firms' average ratings rises soon after the policy shocks and remains elevated throughout the post-period. In Appendix Table F6, we consider an alternative measure of dispersion, the standard deviation of ratings, and again conclude that ratings dispersion rises across firms.

Figure 3: Narrowing NDAs and Cross-Firm Dispersion in Firm Ratings, Dynamic Responses



(a) Bechchmark Rating IQR

(b) Weak Legislation States

Notes: The dependent variable is, within each industry-state-half-year, the inter-quartile range of firms' average ratings. The sample period is 2015–2021 and point estimates are relative to the calendar half-year before the legislation goes into effect. Each observation represents an industry-state-half year and is weighted by the average number of observed firms. Panel (a) includes all states, and Panel (b) includes as control states the weak legislation states. Regressions include industry-state, industry-year-half, and state-year-half fixed effects. Standard errors are two-way clustered by industry and state. Red vertical bars indicate 95% confidence intervals around each point estimate.

6.4 Retaliation Risk as a Mechanism

The core mechanism by which NDAs are likely to cause workers to under-share negative information is that they fear legal retaliation for violating their contract. In this section we shed light on this mechanism. Workers who decide to share negative information may try to conceal aspects of their identity, such as their job title, to mitigate retaliation risk (Grothaus, 2020; McClure, 2022). In general, reviews with more negative information tend to have higher rates of identity concealment (Sockin and Sojourner, 2023). However, because narrowing NDAs should reduce the legal risk workers face for sharing negative information, after NDAs are narrowed we expect workers volunteering reviews will be less likely to conceal their job title, especially when providing negative reviews (since those are the ones likely to lead to retaliation). If instead a spurious correlation is responsible for the negative reviews observed above, we would expect identity concealment to rise.

To examine potential identity concealment, in Table 6, we use as a dependent variable an indicator for whether the individual withheld their job title in their review. We examine

Table 6: Effects of Narrowing NDAs on Job Title Concealment, by Review Content

Sample:	Full	Longer cons than pros	Longer pros than cons
	(1)	(2)	(3)
Narrowed NDAs x NDA intensity	-0.044**	-0.058***	-0.034
	(0.016)	(0.009)	(0.023)
Dependent variable mean	0.140	0.145	0.133
Observations	3645332	1898057	1633324
Adjusted R^2	0.12	0.12	0.13

Notes: The table implements triple-difference models for estimating the causal effect narrowing NDAs on rate at which employees conceal their job title when leaving a review with the specific characteristic noted in the column header. Regressions include firm-state, industry-year-month, and state-year-month fixed effects. Standard errors are two-way clustered by state and industry. Significance levels: * 10%, ** 5%, *** 1%.

overall behavior and then split the sample based on whether the review is more positive or more negative. Our expectation is limited effects in more positive reviews, since narrowing NDAs reduces the legal and retaliatory threat relevant to sharing negative—not positive—information. On average, we find that job title concealment falls 19% ($\frac{-0.044 \times 0.6}{0.14}$). Consistent with narrowing NDAs reducing legal risk, concealment rates fall especially for reviews where the cons text is longer than the pros, i.e., when employees discuss more-negative information.[38] Appendix Figure E5 shows the dynamic event-study, confirming that the reduction in concealment occurs just after NDAs were narrowed and persists thereafter. In contrast, we detect no statistically significant change in concealment among relatively positive reviews.

This concealment analysis is important for a second reason: Volunteers' identity concealment decisions have externality costs. Consumers of reviews find reviews less valuable when the volunteer conceals aspects of their identity (Sockin and Sojourner, 2023). Without a volunteer identifying information, a given negative review is less credible in general and, for users, it is more difficult to judge the relevance to their own situation. As a result, because narrowing NDAs makes workers more likely to share negative information and reduces concealment behavior, narrowing NDAs increases both the flow of negative information and the value of supplied information to jobseekers.

[38]The effect is even more pronounced for the full sample when we incorporate worker fixed effects (see Appendix Table F2 for the triple-difference estimate and Appendix Figure E6 for the event study).

6.5 Sensitivity Analyses

This section probes robustness of the results. Due to the extensive nature of these analyses, we summarize them here and point interested readers to Appendix B for a detailed discussion.

We first document robustness to various control groups (Appendix B.1), including: (i) states that neighbor the three treated states, and (ii) states that have high coverage in the Glassdoor data. Results are also consistent using double-differences designs—either within-state across industries or within-industries across states—building up to the triple-differences specification (see Appendix G for the double-difference models). While the identifying assumption of triple-difference designs is parallel biases and not parallel trends, we also examine robustness of the results to violations of parallel trends using the method outlined by Rambachan and Roth (2023). Some of our results are robust to large violations of parallel trends, while others are more sensitive. The results are also robust to models with individual fixed effects. We also show that the results are robust to considering only multi-state employers that have an establishment in both a treated state and a control state. These models are stringent because they show that the effects are driven by within-firm behavior in treated states versus untreated states, suggesting that the effects of narrowing NDAs in one state does not fully spillover to the company's establishments in control states. The results are also robust to imputing the firm's location for reviews in which such information is unavailable (Appendix B.6).

Appendix B.2 shows that the results are also robust to both stacked designs (given the slightly staggered policy adoption) and equally weighting states and industries. It also shows the results hold when considering alternative NDA measures—including by occupation and industry-occupation—bolstering the interpretation of these effects as pertaining to the narrowing of NDAs. The results are also robust to various ways of handling standard errors (Appendix B.4), including the wild cluster bootstrap and a randomization inference procedure whereby we compare our estimate for overall ratings with that obtained from randomly assigning three states to be treated.

We further examine the extent to which specific states or industries drive the results (Appendix B.3). The effects are strongest in California, which seems plausible because its law was both retroactive and the broadest. The industry-specific results suggest that the highest NDA use industries (i.e., finance and insurance and professional and technical services) drive the results. The results are robust to dropping professional, technical, and scientific services—both overall and only in California—suggesting that changes in the Silicon Valley technology sector are not solely driving the results.

We also examine several alternative potential explanations for the results. Correlated changes in minimum wages do not change results (Appendix B.5), and neither do various measures of the severity of the COVID-19 pandemic (Appendix B.8). While the start of the #MeToo movement, corresponding with the public revelation of Harvey Weinstein's sexual abuse scandal, led to directionally similar changes, these effects appear short-lived and do not explain the results observed after NDAs are narrowed (Appendix B.7).

We also rule out the possibility that firm review-planting behavior might drive any observed findings. That is, a potential response of firms to newly negative reviews on Glassdoor is to engage in review planting campaigns—known broadly as 'sock-puppetry.' While Glassdoor tries to prevent reviews from being planted, some employers do intentionally engage in positive review planting campaigns (Fuller and Winkler, 2020), competitor degradation, or both (Mayzlin et al., 2014).[39] If an employer plants more positive ratings of themselves, then the estimated effect of NDA-narrowing laws on the average firm rating would be biased toward zero—working against finding a negative effect on average ratings. The fact that a negative effect is observed suggests that self-promotion does not overturn the direction of the ratings results. Alternatively, a negatively-shocked employer might have strengthened incentives to plant negative reviews for their competitors, which would amplify the negative effects we observe on average ratings—but would be inconsistent with the increase in

[39] According its responses to frequently asked questions regarding content submission, Glassdoor takes a number of steps to prevent planted reviews. For one, Glassdoor requires an email address to ensure the respondent is "a real person" and respondents must "verify their account via email before any of their posts are shared." Further, Glassdoor makes a "commitment to review every post before it appears on the site."

across-firm ratings dispersion.[40]

To address this potential firm response empirically, we use a proxy for detecting planted reviews following the methodology of Sockin and Sojourner (2023). The key idea is that employer-planted reviews will be more likely to occur as discontinuous spikes in the arrival rate of new reviews for a firm as employers engage in company-wide promotions or accumulative reviews prior to the announcement of awards recognizing employer quality on Glassdoor (Fuller and Winkler, 2020)—thereby breaking the prevailing trend.[41] Appendix Table F17 reports estimates using the main specification but with an indicator for such spikes in reviews, at various growth rate thresholds for considering a review to be suspect, as the outcome. These laws have little effect on this proxy for the share of planted reviews. If anything, the effect is negative. In Appendix Table F18, reviews flagged as planted according to various thresholds are excluded and the effects on average ratings persist.

Another concern is that workers are potentially unaware of these laws. To examine this idea, following Rees-Jones and Rozema (2023), we obtain the number of articles from state-specific news media that mention NDAs, as well as the respective word length of each article, in each state for each calendar month by scraping `newslibrary.com`.[42] Dynamic event study plots, presented in Appendix Figure E9, suggest that NDAs became increasingly common in the news before these laws passed, and discussion of NDAs fell afterwards. Incorporating this media variation into our main models by controlling for an interaction of the lagged measure of NDA-based news coverage with the industry-level measure for NDA intensity illuminates how the Glassdoor results for negative information disclosure change

[40] Competitor-degrading is likely a more expensive, more difficult strategy than self-promotion. Each employer has multiple labor-market competitors, not all of whom are necessarily known to the employer. For an employer with C competitors to achieve the same gain in average review difference with their competitors, it would need to induce roughly C times as many negative reviews for competitors as positive reviews for themselves. Self-promotion is also almost certainly an easier task to accomplish. It requires only identifying one's own positively-disposed employees and encouraging them to post sincerely. In contrast, competitor-degradation requires either persuading employees to lie or finding competitors' negatively-disposed employees and encouraging them to post sincerely.

[41] For a recent example of how this works in practice, see `https://www.newsweek.com/ceos-message-telling-employees-manipulate-glassdoor-reviews-backfires-1704968`.

[42] In particular, we scraped articles including the term NDA, Non-disclosure, Nondisclosure, Non-disparagement, Nondisparagement, or Confidentiality in the title.

when accounting for changes in media attention. The resulting triple-differences estimates are presented in Appendix Table F19, and the coefficients on the news-related covariates are available in Appendix Table F5. Accounting for two years of lagged news coverage attenuates the estimated magnitude of the NDA-narrowing effect on the cons share of the review text by 44%, rendering it no longer statistically significant. The NDA-narrowing effect on the log length of the cons section moderates as well but by a more modest 5%.

6.5.1 Replication in Another Context

While Glassdoor offers a natural setting in which to study shifts in workers' provision of information about employers, since it is precisely the place where workers go to discover such information, we illustrate the robustness of our findings by examining sexual harassment complaints filed with the EEOC. Relative to Glassdoor, reporting to the EEOC is higher stakes, and has the possibility of winning damages in a legal case. The pivotal choice is also different, in that on Glassdoor the question is *what* to report, whereas in the EEOC the question is *whether* to report. Nevertheless, as reported in Appendix C.1, we similarly find that narrowing NDAs is associated with a rise in sexual harassment reporting.

6.5.2 Theranos Case Study as an Alternative Shock to Retaliation Risk

The core mechanism in this study is that legal retaliation risk associated with broad NDAs causes workers to reduce their willingness to supply negative information about their employer. While the main analysis exploits reductions in retaliation risk coming from state laws—but this is not the only way to reduce retaliation risk from broad NDAs. In particular, NDAs cannot legally cover public information, such that the public revelation of misconduct can also reduce retaliation risk from broad NDAs. To bolster this mechanism, we turn to a case study of Theranos. The benefit of a case study is that we know, based on documentary evidence, that Theranos used NDAs to cover up illegal misconduct and withhold information from employees. The public revelation of the Theranos fraud in Carreyrou

(2015) simultaneously informed some Theranos workers about the fraud perpetrated at their employer and reduced the retaliation risk to employees from discussing Theranos' misconduct publicly. Thus, in Appendix C.2, we look at Glassdoor ratings before and after the Carreyrou (2015) article, at Theranos relative to other companies in the biotech and pharmaceutical industry. Although the reduction in NDA-related retaliation risk stems from the public revelation of information (and not state laws), we find concordant results for this case study as in our main analysis—that Theranos' NDAs were propping up its public reputation, i.e., Glassdoor rating, relative to other employers in the industry.

7 Discussion

This section reviews limitations of the analysis and describes areas of valuable future research. We conclude with a summary of the main results and their implications for policy.

7.1 Limitations and Future Research

A first limitation of the analysis comes from the inability to observe "true" wrongdoing, a challenge common in studying misbehavior. As a result, it is difficult to distinguish between two interpretations of the results: That narrowing NDAs (i) reveals existing wrongdoing, or (ii) encourages more wrongdoing. We think (ii) is unlikely for several reasons. First, negative information on employers is durable and can be shared at any point in time, even if the experiences occurred years ago. Indeed, the broadest law of the three we study is California's and it applies retroactively to existing NDAs. Second, if firms cannot use NDAs to prohibit workers from speaking out, then they have less incentive to allow wrongdoing in the first place. Accordingly, the likely interpretation of the results is that policies narrowing NDAs encourage workers to report negative experiences, not that they create more of them.

A second limitation of the study is that the results stem almost exclusively from data on Glassdoor. This website is important precisely for studying employer reputation since jobseekers use it for this purpose, but it is an open question whether the results generalize to other information sharing platforms. We expect this would depend upon credibility and anonymity on the platform. Glassdoor respondents have some anonymity and credibility (e.g., a hidden but verified e-mail address or social media account)—which can be accompanied by voluntary provision of identifying characteristics in reviews. Consider two alternative platforms on different ends of this spectrum, LinkedIn and Reddit.[43] On LinkedIn, users have little to no anonymity and, consequently, more credibility. Profiles contain names, display portraits, and host a user's posts. On Reddit, users are completely anonymous and do not need to verify their employment (though they can reveal it). The effects from narrowing NDAs should be more muted for both platforms. Workers may face retaliation or future employment disadvantages from posting negative information about a company on LinkedIn, whereas the ex ante anonymity likely makes the posting of negative information already more common, albeit less credible, on Reddit.

A few other limitations bear noting. First, while we have studied one potential firm response to negative information revelation—review planting behavior—there are surely other responses that merit further investigation. This could involve engaging in other types of concealment behavior, hiring only trusted associates, or certain types of workers.

Second, with our data we are unable to study how these negative information flows affect subsequent labor market sorting behavior. Given our findings and the research showing that non-wage amenities account for a substantial portion of the value workers gain from a job (Sullivan and To, 2014; Sorkin, 2018; Maestas et al., 2023; Sockin, 2022; Bidwell et al., 2015), we expect this prior literature to have under-estimated the extent to which amenities affect sorting behavior, since the extensive use of broad NDAs has likely distorted the perceptions that prospective jobseekers have about workplaces' amenities. Ideally, one could model the firm-level labor flows in response to the narrowing of NDAs, incorporating potential firm responses as noted above. Without such firm-level data, we leave this for future work.

[43] Though Reddit is a broader information-sharing platform that is primarily not used for disseminating labor market information, there are popular threads specific to workplace experiences, e.g., Antiwork.

Lastly, information frictions like those related to NDAs are just one of a variety of frictions in the labor market. Other frictions like thin labor markets (Schubert et al., 2022), search and matching frictions (Manning, 2013), or related labor market imperfections that give firms monopsony power (Sokolova and Sorensen, 2021), may interact with broad NDAs in important ways. We hope future researchers will explore the potential interplay that may occur between restrictive covenants that suppress worker voice and labor market competition.

7.2 Concluding Remarks

This study is motivated by the recent policy and social interest regarding broad NDAs, the decades-old legal concerns that firms use them to prop-up their reputations by hiding negative information, and the concomitant negative externalities on workers and competing firms (Bast, 1999; Hoffman and Lampmann, 2019). Its main contribution is providing the first quantitative analysis of these externalities with evidence on how broad NDAs suppress the flow and value of negative information about employers, inflate firm reputation, and compress the reputation distribution, ultimately making it more difficult for high-road employers to differentiate themselves from low-road employers.

Our findings have important implications for policy regarding contracts that restrict worker voice. Whereas non-disclosure clauses that protect trade secrets may have theoretical economic justification, non-disparagement clauses do not provide this same benefit. Unless they are narrowly tailored, they generally prohibit employees from sharing *truthful* negative information about their employers. Since the parties to non-disparagement clauses likely will not internalize the costs of silence imposed on others, and because non-disparagement clauses are designed only to suppress the flow of information on employers, the negative externalities we identify here are likely to apply even more strongly. Policymakers have recently expressed similar concerns: A February 2023 National Labor Relations Board decision in *McLaren Macomb* determined that a broad non-disparagement agreement signed as part of a settlement violates the National Labor Relations Act,[44] though the decision has no effect on supervisors or management (Burke et al., 2023). The decision appears to apply also to non-disparagement agreements signed as a condition of employment.[45]

While the study's focus is on ex ante NDAs—e.g., signed as a condition of employment—policymakers have also been concerned about NDAs signed in severance and settlement agreements, after any harm has been experienced. For example, in January 2022 California's 'Silenced No More Act' prohibited firms from using NDAs signed to as part of a settlement to conceal harassment, discrimination, or retaliation.[46] While results suggest that ex post NDAs also likely create negative externalities by silencing workers, such policies may come with important trade-offs for the individuals who experienced the wrongdoing in the first place. Individuals who have experienced harm at work may not wish to share their negative experiences, and may prefer to receive a compensating differential for their silence. Given that these ex post NDA payments are typically private and endogenous to the expected or actual wrongdoing, they are necessarily difficult to study. Nevertheless, we hope that future research will engage not only with the potential externalities that NDAs create, but also with the distinction between ex ante and ex post NDAs, how much directly-harmed workers value their freedom to speak out, and how much they are compensated for giving it up.

---

[44] The "Non-Disclosure" portion of the agreement at issue read "At all times hereafter, the Employee promises and agrees not to disclose information, knowledge or materials of a confidential, privileged, or proprietary nature of which the Employee has or had knowledge of, or involvement with, by reason of the Employee's employment. At all times hereafter, the Employee agrees not to make statements to Employer's employees or to the general public which could disparage or harm the image of Employer, its parent and affiliated entities and their officers, directors, employees, agents and representatives."

[45] See the follow-up memo from the General Counsel at https://www.nlrb.gov/news-outreach/news-story/nlrb-general-counsel-issues-memo-with-guidance-to-regions-on-severance. In the memo, the general counsel answers the question "How does this decision affect other employer communications with employees, such as pre-employment or offer letters?" as follows: "Based on extant Board law, overly broad provisions in any employer communication to employees that tend to interfere with, restrain or coerce employees' exercise of Section 7 rights would be unlawful if not narrowly tailored to address a special circumstance justifying the impingement on workers' rights."

[46] See https://silencednomore.org/the-silenced-no-more-act.

References

A. Abadie, S. Athey, G. W. Imbens, and J. M. Wooldridge. When Should You Adjust Standard Errors for Clustering? The Quarterly Journal of Economics, 138(1):1–35, 10 2022.

A. Adams-Prassl, K. Huttunen, E. Nix, and N. Zhang. Violence against women at work. 2022.

P. Akey, S. Lewellen, I. Liskovich, and C. Schiller. Hacking Corporate Reputations. 2021. Working paper.

H. Akopyan. California enacts legislation to combat discrimination and harassment. https://www.defenselitigationinsider.com/2019/02/06/california-enacts-legislation-to-combat-discrimination-and-harassment/, 2019.

A. C. Baker, D. F. Larcker, and C. C. Wang. How much should we trust staggered difference-in-differences estimates? Journal of Financial Economics, 144(2):370–395, 2022.

N. Balasubramanian, E. Starr, and S. Yamaguchi. Bundling postemployment restrictive covenants: When, why, and how it matters. Working paper, 2021.

L. Barrage, E. Chyn, and J. Hastings. Advertising and environmental stewardship: Evidence from the bp oil spill. American Economic Journal: Economic Policy, 12(1):33–61, February 2020.

C. M. Bast. At what price silence: are confidentiality agreements enforceable. Wm. Mitchell L. Rev., 25:627, 1999.

C. Batut, C. Coly, and S. Schneider-Strawczynski. It's a man's world: culture of abuse,# metoo and worker flows. Working Paper, 2021.

R. Beck. President biden?s proposed ban of (most) noncompetes: Protection strategies and steps to take now. https://www.faircompetitionlaw.com/2020/12/02/president-bidens-proposed-ban-of-most-noncompetes-protection-strategy-and-steps-to-take-now/, 2020.

A. Benson, A. Sojourner, and A. Umyarov. Can reputation discipline the gig economy?: Experimental evidence from an online labor market. Management Science, 66:1802–1825, 2020.

M. Bidwell, S. Won, R. Barbulescu, and E. Mollick. I used to work at goldman sachs! how firms benefit from organizational status in the market for human capital. Strategic Management Journal, 36(8):1164–1173, 2015.

J. Boitnott. 3 reasons to think again before bashing your company on glassdoor. https://www.entrepreneur.com/article/247010, 2015.

L. Boudreau, S. Chassang, A. González-Torres, and R. Heath. Monitoring harassment in organizations. Working Paper, 2022.

L. Broden & Mickelsen. Can you report a sexual assault if you signed a nda? https://www.brodenmickelsen.com/blog/can-you-report-a-sexual-assault-if-you-signed-a-nda/, 2018.

K. A. Bryan, M. Hoffman, and A. Sariri. Information frictions and employee sorting between startups. Working Paper 30449, National Bureau of Economic Research, September 2022.

C. R. Burke, A. Reganata, and A. Telegen. Nlrb targets confidentiality and non-disparagement provisions. https://www.tradesecretslaw.com/2023/02/articles/restrictive-covenants/nlrb-targets-confidentiality-and-non-disparagement-provisions/, 2023.

L. Cabral and A. Hortaçsu. The dynamics of seller reputation: Evidence from ebay. The Journal of Industrial Economics, 58(1):54–78, 2010.

B. Callaway and P. H. Sant'Anna. Difference-in-differences with multiple time periods. Journal of Econometrics, 2020.

G. Carlson. Gretchen carlson: Fox news, i want my voice back. https://www.nytimes.com/2019/12/12/opinion/gretchen-carlson-bombshell-movie.html, 2019.

H. L. Carmichael. Reputations in the labor market. American Economic Review, 74(4):713–725, 1984.

J. Carreyrou. Hot startup theranos has struggled with its blood-test technology. https://www.wsj.com/articles/theranos-has-struggled-with-blood-tests-1444881901, 2015.

J. Carreyrou. Theranos whistleblower shook the company and his family. https://www.wsj.com/articles/theranos-whistleblower-shook-the-companyand-his-family-1479335963, 2018.

D. Cengiz, A. Dube, A. Lindner, and B. Zipperer. The effect of minimum wages on low-wage jobs. The Quarterly Journal of Economics, 134(3):1405–1454, 2019.

L. M. Cortina and V. J. Magley. Raising voice, risking retaliation: Events following interpersonal mistreatment in the workplace. Journal of Occupational Health Psychology, 8(4):247, 2003.

J. F. Coyle. A short history of the choice-of-law clause. U. Colo. L. Rev., 91:1147, 2020.

G. B. Dahl and M. M. Knepper. Why is workplace sexual harassment underreported? the value of outside options amid the threat of retaliation. Technical report, National Bureau of Economic Research, 2021.

K. H. Decker. Covenants Not to Compete. Aspen Publishers Online, 1993.

A. Dey, J. Heese, and G. Pérez-Cavazos. Cash-for-information whistleblower programs: Effects on whistleblowing and consequences for whistleblowers. Journal of Accounting Research, 59(5): 1689–1740, 2021.

D. W. Diamond. Reputation acquisition in debt markets. Journal of Political Economy, 97(4): 828–862, 1989.

S. Dube and C. Zhu. The disciplinary effect of social media: Evidence from firms' responses to glassdoor reviews. Journal of Accounting Research, 2021.

D. Eaton. New restrictions on employee confidentiality and non-disparagement pacts. https://www.sandiegouniontribune.com/business/story/2021-10-25/new-restrictions-on-employee-confidentiality-and-non-disparagement-pacts, 2021.

J. Ence. I like you when you are silent: The future of ndas and mandatory arbitration in the era of# metoo. J. Disp. Resol., page 165, 2019.

A. Filippas, J. J. Horton, and J. Golden. Reputation inflation. 2020.

J. Flanagan and T. Gerstein. 'sign on the dotted line': How coercive employment contracts are bringing back the lochner era and what we can do about it. USFL Rev., 54:441, 2019.

O. Folke and J. Rickne. Sexual harassment and gender inequality in the labor market. The Quarterly Journal of Economics, 137(4):2163–2212, 2022.

A. Fuller and R. Winkler. How companies secretly boost their glassdoor ratings. https://www.wsj.com/articles/companies-manipulate-glassdoor-by-inflating-rankings-and-pressuring-employees-11548171977, 2020.

S. Gadgil and J. Sockin. Caught in the act: How corporate scandals hurt employees. (3774639), June 2020.

C. Gibson. How ndas kept the lid on harassment scandals ? and why that might be changing. https://www.washingtonpost.com/lifestyle/style/how-ndas-kept-the-lid-on-harassment-scandals--and-why-that-might-be-changing/2017/10/25/62af1e30-b99d-11e7-a908-a3470754bbb9_story.html, 2017.

A. Goodman-Bacon. Difference-in-differences with variation in treatment timing. Journal of Econometrics, 225(2):254–277, 2021.

L. Grensing-Pophal. How to protect your employer brand on glassdoor. Society for Human Resource Management, October 2019.

E. Griffith. How the world learns about bosses behaving badly. https://www.nytimes.com/2021/06/05/business/nondisclosure-agreements-lioness.html, 2021.

S. Groff. California restricts confidentiality provisions concerning information related to sexual harassment. https://www.jacksonlewis.com/publication/california-restricts-confidentiality-provisions-concerning-information-related-sexual-harassment, 2018.

M. Grothaus. Every employee's worst nightmare, getting outed on glassdoor, could become a reality. Wired, February 2020.

S. Habeshian. Biden signs bill to curb use of sexual harassment ndas. https://www.theemployerhandbook.com/can-an-company-force-current-and-former-employees-not-to-bash-it-on-glassdoor-is-that-legal/, 2022.

H. Hafiz. How legal agreements can silence victims of workplace sexual assault. https://www.theatlantic.com/business/archive/2017/10/legal-agreements-sexual-assault-ndas/543252/, 2017.

D. S. Hamermesh. The changing distribution of job satisfaction. Working Paper 7332, National Bureau of Economic Research, September 1999.

J. Harju, S. Jäger, and B. Schoefer. Voice at work. Technical report, National Bureau of Economic Research, 2021.

E. Harris. Despite #metoo glare, efforts to ban secret settlements stop short. https://www.nytimes.com/2019/06/14/arts/metoo-movement-nda.html, 2019.

D. Hemel. How nondisclosure agreements protect sexual predators. https://www.vox.com/the-big-idea/2017/10/9/16447118/confidentiality-agreement-weinstein-sexual-harassment-nda, 2017.

A. O. Hirschman. Exit, voice, and loyalty: Responses to decline in firms, organizations, and states, volume 25. Harvard university press, 1970.

D. A. Hoffman and E. Lampmann. Hushing contracts. Wash. UL Rev., 97:165, 2019.

C. A. Hrdy and C. B. Seaman. Beyond trade secrecy: Confidentiality agreements that act like noncompetes. Yale Law Journal, 133, 2023.

P. Hughes and M. Nacchio. New jersey bars nondisclosure agreements in all discrimination settlements, purports to bar arbitration agreements. https://www.natlawreview.com/article/new-jersey-bars-nondisclosure-agreements-all-discrimination-settlements-purports-to, 2019.

A. Johnson, K. Menefee, and R. Sekaran. Progress in Advancing Me Too Workplace Reforms in 20statesby2020. National Women's Law Center, 2019.

M. S. Johnson, K. Lavetti, and M. Lipsitz. The labor market effects of legal restrictions on worker mobility. Available at SSRN 3455381, 2020.

M. Karabarbounis and S. Pinto. What can we learn from online wage postings? evidence from glassdoor. Economic Quarterly, 104:173–189, 01 2019.

D. Lange, P. M. Lee, and Y. Dai. Organizational reputation: A review. Journal of Management, 37(1):153–184, 2011.

Y. Lee, S. Ng, T. J. Shevlin, and A. Venkat. The Effects of Tax Avoidance News on Employee Perceptions of Managers and Firms: Evidence from Glassdoor.com Ratings. The Accounting Review, 2021.

T. Legittino. California employment law changes you need to know. https://fkks.com/news/california-employment-law-changes-you-need-to-know, 2019.

Y.-s. Lii and M. Lee. Doing right leads to doing well: When the type of csr and reputation interact to affect consumer evaluations of the firm. Journal of Business Ethics, 105:69–81, 06 2012. doi: 10.1007/s10551-011-0948-0.

Y. Liu and V. Shankar. The dynamic impact of product-harm crises on brand preference and advertising effectiveness: An empirical analysis of the automobile industry. Management Science, 61(10):2514–2535, 2015.

P. Lloyd-Smith and H. An. Are corporate social responsibility and advertising complements or substitutes in producing firm reputation? Applied Economics, 51(21):2275–2288, 2019.

O. Lobel. Enforceability tbd: From status to contract intellectual property law. BUL Rev., 96:869, 2016.

O. Lobel. Ndas are out of control. heres what needs to change. Harvard Business Review, 2018.

O. Lobel. Boilerplate collusion: Clause aggregation, antitrust law & contract governance. Minn. L. Rev., 106:877, 2021.

M. Luca. Reviews, Reputation, and Revenue: The Case of Yelp.com. Harvard Business School Working Papers 12-016, Harvard Business School, Sept. 2011.

N. Maestas, K. J. Mullen, D. Powell, T. von Wachter, and J. B. Wenger. The value of working conditions in the united states and the implications for the structure of wages. American Economic Review, 113(7):2007–47, July 2023.

A. Manning. Monopsony in motion: Imperfect competition in labor markets. Princeton University Press, 2013.

I. Marinescu, N. Klein, A. Chamberlain, and M. Smart. Incentives can reduce bias in online employer reviews. Academy of Management Proceedings, 2018(1):11189, 2018.

D. Mayzlin, Y. Dover, and J. Chevalier. Promotional reviews: An empirical investigation of online review manipulation. American Economic Review, 104(8):2421–55, August 2014.

T. McClure. Glassdoor ordered to reveal identity of negative reviewers to new zealand toymaker. https://amp.theguardian.com/world/2022/jul/19/glassdoor-ordered-to-reveal-identity-of-negative-reviewers-to-new-zealand-toymaker, 2022.

R. C. McDevitt. Names and reputations: An empirical analysis. American Economic Journal: Microeconomics, 3(3):193–209, 2011.

J. McShane. She broke her nda with harvey weinstein in 2017. here?s how she wants to change the system for others. https://www.thelily.com/she-broke-her-nda-with-harvey-weinstein-in-2017-heres-how-she-wants-to-change-the-system-for-others/, 2021.

T. Melo and A. Garrido-Morgado. Corporate reputation: A combination of social responsibility and industry. Corporate Social Responsibility and Environmental Management, 19(1):11–31, 2012.

E. Meyer. Can a company require new employees not to bash it on glassdoor? is that legal? https://www.theemployerhandbook.com/can-an-company-force-current-and-former-employees-not-to-bash-it-on-glassdoor-is-that-legal/, 2019.

E. W. Morrison. Employee voice and silence. Annu. Rev. Organ. Psychol. Organ. Behav., 1(1): 173–197, 2014.

P. K. Mowbray, A. Wilkinson, and H. H. Tse. An integrative review of employee voice: Identifying a common conceptualization and research agenda. International Journal of Management Reviews, 17(3):382–400, 2015.

National Women's Law Center. Limiting nondisclosure and nondisparagement agreements that silence workers: Policy recommendations. https://nwlc.org/wp-content/uploads/2020/04/NDA-Factsheet-4.27.pdf, 2020.

C. Nosko and S. Tadelis. The limits of reputation in platform markets: An empirical analysis and field experiment. Working Paper 20830, National Bureau of Economic Research, January 2015.

A. Olden and J. Møen. The triple difference estimator. The Econometrics Journal, 25(3):531–553, 2022.

Z. Perkins. An nda from harvey weinstein cost me my career ? at last, banning them feels within reach. https://www.theguardian.com/commentisfree/2022/dec/15/nda-harvey-weinstein-confidentiality-clause-abuse, 2022.

R. M. Philip. Silence at our expense: Balancing safety and secrecy in non-disclosure agreements. Seton Hall. L. Rev., 33:845, 2002.

J. Prescott and E. Starr. Subjective beliefs about contract enforceability. Working Paper, 2021.

E. Primeaux. Whistleblower helped dismantle biotech juggernaut theranos in his 'zero-strategy' defense. https://www.fraud-magazine.com/cover-article.aspx?id=4295006794, 2019.

A. Rambachan and J. Roth. A More Credible Approach to Parallel Trends. The Review of Economic Studies, 90(5):2555–2591, 02 2023.

A. Rees-Jones and K. Rozema. Price isn?t everything: Behavioral response around changes in sin taxes. National Tax Journal, 76(1):000–000, 2023.

E. Rella. Apple makes historic move by removing employee gag orders for talking about harassment, discrimination apple has been facing mounting pressure from shareholders and activists to remove "concealment clauses" for employees. https://www.entrepreneur.com/business-news/apple-removes-gag-orders-for-workers-talking-about/441048, 2022.

L. Rogal. Secrets, lies, and lessons from the theranos scandal. Hastings LJ, 72:1663, 2020.

E. Roth. Is a nondisclosure agreement silencing you from sharing your ?me too? story? 4 reasons it might be illegal. https://www.aclu.org/news/womens-rights/nondisclosure-agreement-silencing-you-sharing-your-me-too , 2018.

J. Roth, P. H. Sant'Anna, A. Bilinski, and J. Poe. What's trending in difference-in-differences? a synthesis of the recent econometrics literature. Journal of Econometrics, 235(2):2218–2244, 2023.

O. R. Royle. The top 15 red flags for job seekers?and it?s bad news for leaders who say ?work hard, play hard? or ?we are like family? https://fortune.com/2023/04/04/top-15-red-flags-job-seekers-bad-news-leaders-say-work-hard-play-hard-we-are-family/, 2023.

S. Sanga. Choice of law: an empirical analysis. Journal of Empirical Legal Studies, 11(4):894–928, 2014.

G. Schubert, A. Stansbury, and B. Taska. Employer concentration and outside options. Available at SSRN 3599454, 2022.

C. Shapiro. Premiums for high quality products as returns to reputations. The Quarterly Journal of Economics, 98(4):659–679, 1983.

J. L. Short. Killing the messenger: The use of nondisclosure agreements to silence whistleblowers. U. Pitt. L. Rev., 60:1207, 1998.

J. Silver-Greenberg and N. Kitroeff. How bloomberg buys the silence of unhappy employees. www.nytimes.com/2020/03/02/business/michael-bloomberg-nda.html, 2020.

J. Sockin. Show me the amenity: Are higher-paying firms better all around? CESifo Working Paper 9842, 2022.

J. Sockin and M. Sockin. A pay scale of their own: Gender differences in variable pay. (3512598), December 2019.

J. Sockin and A. Sojourner. What's the inside scoop? challenges in the supply and demand for information on employers. Journal of Labor Economics, 2023.

A. Sokolova and T. Sorensen. Monopsony in labor markets: A meta-analysis. ILR Review, 74(1): 27–55, 2021.

I. Sorkin. Ranking firms using revealed preference. The Quarterly Journal of Economics, 133(3): 1331–1393, 2018.

E. Starr, J. Frake, and R. Agarwal. Mobility constraint externalities. Organization Science, 30(5): 961–980, 2019.

E. Starr, J. Prescott, and N. Bishara. The behavioral effects of (unenforceable) contracts. The Journal of Law, Economics, and Organization, 36(3):633–687, 2020.

E. P. Starr, J. Prescott, and N. D. Bishara. Noncompete agreements in the us labor force. The Journal of Law and Economics, 64(1):53–84, 2021.

C. A. Sullivan. The puzzling persistence of unenforceable contract terms. Ohio St. LJ, 70:1127, 2009.

P. Sullivan and T. To. Search and nonwage job characteristics. Journal of Human Resources, 49 (2):472–507, 2014.

S. Tadelis. What's in a name? reputation as a tradeable asset. American Economic Review, 89(3): 548–563, 1999.

A. Topping. Harvey weinstein pa says abusers still have the legal power to silence victims. https://www.theguardian.com/world/2021/oct/10/harvey-weinstein-pa-says-abusers-still-have-the-legal-power-to-silence-victims, 2021.

M. Torres. What to know before you post negative comments about your company. https://www.huffpost.com/entry/writing-bad-review-employer-anonymity_l_5d9cca6de4b087efdba3bfe7, 2019.

D. B. Turban and D. M. Cable. Firm reputation and applicant pool characteristics. Journal of Organizational Behavior, 24(6):733–751, 2003.

G. Usenheimer, N. Merola, and A. Dana. #metoo inspires legislative changes across the united states. https://www.lexology.com/library/detail.aspx?g=4470934c-7a2f-4ae9-86af-e670cd28799d, 2019.

Y. Zhou and C. Makridis. Financial Misconduct, Reputation Damage and Changes in Employee Satisfaction. Working paper, 2021.

Online Appendix

Non-Disclosure Agreements and Externalities from Silence

A Is Glassdoor Anonymous?

While it might seem like posting on Glassdoor is completely anonymous, firms can often directly discern which employees have written which reviews, or Glassdoor is sometimes required to reveal the identities of individuals who are posting,[47] though they resist doing so. Below we highlight several instances of identity revelation on Glassdoor, as well as how firms identify such individuals, and the risks that firms can sue workers for violating NDAs.

- *"...I'm a Glassdoor admin for my company and on the HR team. I'd say 90% of the time we know who left the posting because people tend to write them immediately after they have been put on a PIP, Have been fired, or quit. It's unusual for other employees to write good or bad reviews daily for a company. Timing is a give away. People also tend to write like they talk and give insights into the department they just left, or say things inadvertently give up their identity. Be careful and thoughtful in your reviews."*[48]
- *"I recently left a job after a few years of intense work and long hours. After leaving, I left a critical but fair Glassdoor review. Since then, I have received multiple emails from the CEO of my former employer asking me to remove the review and including (what I perceive to be) threats of the industry being "small" and mentioning my new manager by name."*[49]
- *"Looking closely at a review can give you a clue about the employee: If a review were left by someone who worked in your company in the late 80s, you wouldn't have as much luck figuring out who the employee is. However, most reviews are left by recently fired employees or those who resigned recently. If your organisation has a reputation management team in place, reading between the lines of a review can help reveal which employee left a review. Usually, disgruntled employees will leave possible clues about their roles and job positions as they rant in a negative review. Put your detective hat on and you might be able to resolve the issue."* [50]

[47] See https://amp.theguardian.com/world/2022/jul/19/glassdoor-ordered-to-reveal-identity-of-negative-reviewers-to-new-zealand-toymaker.

[48] See https://www.quora.com/Is-there-a-way-for-a-company-to-identify-an-anonymous-review-on-Glassdoor.

[49] See https://www.askamanager.org/2021/05/my-old-employer-is-trying-to-intimidate-me-into-removing-a-glassdoor-review.html.

[50] See https://cloutly.com/blog/are-glassdoor-reviews-anonymous/.

- In an article from Entrepreneur.com titled '3 Reasons to Think Again Before Bashing Your Company on Glassdoor', Boitnott (2015) highlights the following issues.[51]

  - *"Your comments could land you on the unemployment line, potentially causing irreparable damage to your reputation or career."*
  - *"People may recognize you ... As clever as you think your efforts are at disguising your identity, your co-workers are cleverer than you think. Without you even realizing it, the wheels could be turning in the background. Your co-workers and supervisors may be whispering about the post, analyzing the wording used in it, all while you're going about your day, completely unaware.*
  - *"Even if you think you disguised all clues that you were the person behind the anonymous post, to be truly accurate about your workplace experience you'd have to give some details about your experience. You may not even realize you're one of only a few people who didn't get a raise last year, for instance, and complaining about your salary rut gives away your identity. If you're part of a small startup team (I'd say anywhere between five and 50 people, or thereabouts), you probably want to avoid venting through Glassdoor altogether."*
  - *"Once your boss has identified it's you, what will happen? "Free speech" doesn't necessarily protect you in a situation like this, especially if you're in a right to work state, where your employer can fire you without cause. If you remain on board, you may find yourself feeling shunned or shut out of group gatherings. Take these potential outcomes into consideration before you post."*
  - *"Technology can track you ... If you're posting to Glassdoor on work-issued devices, including smartphones or tablets, keep in mind that your server administrator may be able to track your activity. Glassdoor may protect your identity but they can't do anything in an instance like this, since your company owns the equipment you're using. This definitive proof that you're the one who posted the review could serve as documentation in your boss's decision to discipline or terminate you.*

- Torres (2019) wrote an article entitled "What To Know Before You Post Negative Comments About Your Company" discusses whether reviews are 'legally actionable.'[52]

  - *"Anonymity online is not guaranteed: The Supreme Court has repeatedly decided that we have a right to anonymous free speech as protected by the First Amendment. But that doesn't guarantee your employer can't find out who you are when posting anonymously. 'There's nothing that absolutely protects someone's anonymity,' Mackey said."*

[51] See https://www.entrepreneur.com/article/247010.

[52] See https://www.huffpost.com/entry/writing-bad-review-employer-anonymity_l_5d9cca6de4b087efdba3bfe7.

– *"More and more employers fight back with what you signed on the dotted line: The contractual agreements you sign when you onboard as an employee are being used as an additional tool to silence negative feedback from employees. 'A lot of employers are now moving to making claims around either violations of nondisclosure agreements that employees signed or trade secret claims,' Mackey said."*

– *"Take the former Theranos employees who sought to come forward about the blood-testing company's fraudulent practices. Theranos alleged that one of its anonymous whistleblowers was disclosing 'trade secrets' and threatened to sue if she didn't stop. In that case, the threats did not stop the truth about Theranos from eventually coming out, but it demonstrates the legal tactics bad employers can use to keep you from sharing the truth to the world."*

– *"In these cases, Mackey said that a lot of the purposes for why employers may want to unmask these individuals is not because they're after people telling something they shouldn't be telling about the confidentiality of the business, 'they're doing it to punish someone, to intimidate them, to harass them and to make them silent, when they would otherwise speak out."'*

B Summary of Robustness Checks

In this section we assess the sensitivity of results to alternative control groups, alternative measures of NDA intensity, alternative ways of handling standard errors, alternative methods for handling staggered adoption, alternative weighting schemes, the potential for firms to plant reviews, industry heterogeneity, state heterogeneity, the issue of missing locations in reviews, and whether the results are driven by the #MeToo movement or the advent of COVID-19.

B.1 Different Control Groups and Double-Difference Designs

In Table F8, we consider how the main results change when we consider three alternative sets of control states. First, we consider regional neighbors only, under the assumption that neighbors may operate similarly or face similar exposure to economic shocks to the treated states. Second, we focus only on states that are the most represented in the Glassdoor data. Third, we consider the fact that several other states passed restrictions related to the #MeToo movement around this time period (see Johnson et al. (2019) for details).

In addition, and although triple-difference designs rely on a parallel biases assumption (and not parallel trends), we nevertheless study how sensitive our estimates are to a violation of parallel trends using the relative magnitude approach developed in Rambachan and Roth (2023). In this approach we allow the post-trends to be violated as a multiple of the largest period-to-period change in the pre-trends (denoted $\bar{M}$). In this regard, we follow the recommendation outlined in Roth et al. (2023) to report the "breakdown" value of $\bar{M}$. In Figure E10, for information disclosure and ratings, we show the 90% confidence interval for the first treatment period (the default in Rambachan and Roth (2023)) and the full post-treatment period (the right column) under a range of values for $\bar{M}$, for both our benchmark specification and the specification with individual fixed effects. The effect on overall ratings is the most robust, especially when worker fixed effects are included. In that specification, the full post-period results remain statistically significant until $\bar{M} = 0.85$. Our results for the log length of the cons section are less resilient, primarily because the violation of parallel trends in the post period are based on one particularly noisy pre-period two years before the laws passed.

A related issue is that the triple difference analyses build in variation from control groups across states and low-NDA incidence industries. In Appendix G, we revisit the main models, building from all the double-differences to the triple difference specification. In general, the results are robust to using different control groups based on the within-state cross-industry and within-industry cross-state comparisons. The results are also robust to including individual fixed effects (Table F2).

One may also be concerned about within-firm, cross-state spillovers as potentially contaminating the control groups. That is, wrongdoing within a firm in a treated state may spillover to establishments of the firm in untreated states. That said, our effects if anything only strengthen when we restrict the control set to employers with establishments in both treated and control states (Table F12). If such behavior is happening, it should attenuate estimates, making them conservative. Moreover, ratings results are robust to focusing only on single-state firms (Table F22).

B.2 Stacked Designs, Weighting, and Other NDA Measures

A recent literature highlights several concerns with staggered adoption in two-way fixed effects models (Goodman-Bacon, 2021; Callaway and Sant'Anna, 2020; Baker et al., 2022). The results are unlikely to be affected by these issues because the policies were adopted within a year of each other and because we exploit within-state across-industry variation in the incidence of NDAs. Nevertheless, we employ a stacked regression to address the fact that some treated states serve as control states (Cengiz et al., 2019). To do this, we create a dataset with just one treated state and all control states. We then append another dataset with a different treated state and all control states. We repeat this for the third treated state, such that the data are stacked but that within each dataset, there is no variation coming from other treated states. We then implement the triple-differences specification with fixed effects for each dataset. The results reaffirm the main findings (Table F7: column 1).

In column 2 of Table F7, we address the fact that certain states and industries are more prevalent in the data. For instance, California represents 15% of the sample, where as Illinois and New Jersey reflect 5.0% and 2.5%, respectively. To redistribute weight towards smaller states and industries, we consider an alternative specification, where industry-state pairings have equal weight, meaning each review in industry n and state s has weight $1/\sum 1_{ns}$. Running the baseline triple-differences regression giving each industry-state equal weight does not change the takeaway results from the Glassdoor analyses.

We also consider alternative measures of NDA incidence. Using the same survey from Payscale.com, we calculate the share of workers that are covered by NDAs within occupations and within industry-occupation pairs, where occupations reflect *onet50* occupation categories. To obtain occupations for Glassdoor reviews, we use a mapping from job title to occupation that was constructed based on Glassdoor's textual analysis machine learning algorithm. We then re-estimate a triple-differences model using a continuous measure of NDA intensity across occupations (column 3) and industry-occupation pairs (column 4). The results remain robust when utilizing these measures instead. Importantly though, these estimates include only those reviews with revealed job titles, which is an endogenous outcome itself. Therefore, we do not use either of these alternatives as the preferred specification.[53]

B.3 Estimates by Industry and State

We also consider the extent to which the results are driven by certain industries. Rather than allowing the treatment to vary with the NDA intensity between industries, we consider an alternative approach in which we estimate the coefficient of a standalone post-legislation indicator separately for each industry. Figure E8 shows the results. Interestingly, industries with low-NDA incidence appear to have more positive ratings following the passage of this legislation. This may be due to the fact that California, Illinois, and New Jersey passed numerous other laws alongside the NDA provisions that sought to improve the workplace in light of the #MeToo movement, in addition to other policies such as raising the minimum wage. However, as in the main results, industries with high-NDA incidence, such as professional, scientific and technical services as well as finance and insurance, have more negative reviews (and reduced likelihood of concealment). We also see a negative slope in both graphs, and it is this additional difference that nets out the effects of any non-NDA-interactive state policies implemented around the same time.

Next, we evaluate the extent to which reputation results are driven by either of the three state policies, which do differ from each other. We re-estimate the triple-differences specification, but isolating the effect for each policy separately by excluding the two other treated states entirely. The results, detailed in Table F9, reveal that California and Illinois have robust negative effects on overall ratings, while the estimate for New Jersey is smaller in absolute value and nosier such that we fail to reject a null effect. That California's effect is the largest is not necessarily surprising, since California's policy was the broadest, including

[53]This necessary sample restriction is non-trivial, as the sample sizes are cut by more than 40% when variation between occupations is incorporated.

explicitly non-disparagement agreements, and the only retroactive policy. It is not obvious why we observe different effects in Illinois versus New Jersey.

One may also be concerned the results are driven by a singular industry shock in one of the treated states. For instance, the professional, scientific, and technical services industry exhibits the highest rate of NDA use and employer reviews from California exhibit the strongest response to the narrowing of NDAs. Are high-skilled service-sector workers in California, which comprise more than one-quarter of the sample among treated states, driving the results? Reassuringly, they are not. Excluding reviews from this industry in California (Table F10) or from the sample entirely (Table F11) does not alter our findings.

B.4 Alternative Approaches to Standard Errors

First, we iterate over four different approaches for clustering standard errors, including only state, industry cross state, firm and state, and using the wild cluster bootstrap. Across all specifications, the results hold (Table F13).

We also implement a randomized inference approach for the ratings results to gauge how often one would generate a result as negative as the observed triple-differences estimate from randomly allocating states between treatment and control groups. There are three possible treatments that can be assigned: January 2019, March 2019, and January 2020. We randomly assign three states to these treatments pulling from a uniform distribution, assign the rest to the control group, and record the triple-differences estimate for the main reputation result. After repeating this exercise five-hundred times, 5% of the simulation's coefficients fall below the true estimate (Figure E7).

B.5 Robustness to Changes in the Minimum Wage

Next, we consider the possibility that the decline in overall ratings is driven by differences in wage growth after the laws. This poses an identification threat because higher NDA intensity is strongly associated with greater pay—correlations of 0.62–0.69 across industries and occupations—and the three states that narrowed NDAs also increased their minimum wages around the same time.[54] If faster wage growth results in greater job satisfaction (Hamermesh, 1999), then our triple-differences estimates may reflect a wage effect rather than the narrowing of NDAs. To address this concern, we turn to Glassdoor pay data.[55] Specifically, we calculate the average log earnings among full-time workers by calendar half-year within a given labor market, i.e., the pairing of a state and job type (industry, occupation, industry-occupation, firm, or firm-occupation), and incorporate this measure into the triple-differences model (Table F14). The robustly negative estimates from narrowing NDAs change little, suggesting the observed effects appear unrelated to wage changes.

B.6 Addressing Reviews with Missing Locations

Our analysis is restricted to reviews for which a location is available, in order to assign reviews to treated or control states. However, leaving the location of the review blank is not uncommon, representing nearly 41% of reviews. To attempt to incorporate these reviews into the analysis, we implement an imputation procedure by which reviews are assigned to their highest likelihood state. Although the location for these reviews is missing, the firm is not. The intuition behind the imputation process is to use all of the firms' reviews for which the location is not missing to estimate a latent distribution of the firms' reviews across states. We then assign every review for firm k with a blank location to the state s with the highest probability of origination, $p_{k,s} = \sum 1_{ks} / \sum 1_k$. If concealing location is not random and instead a strategic decision when revealing negative information, failing to incorporate these reviews may bias results. At the same time though, incorporating these

[54]According to data from the Federal Reserve Economic Database (FRED), California raised the state minimum wage from \$11.00 in 2018 to \$12.00 in 2019, New Jersey raised the state minimum wage from \$8.60 in 2018 to \$10.00 in 2019, and Illinois raised the state minimum wage from \$8.25 in 2019 to \$10.00 in 2020.

[55]Other works have found Glassdoor pay data to be representative within industries and occupations. Karabarbounis and Pinto (2019) show the data broadly match first and second moments by industry and region using the Quarterly Census for Employment and Wages and the Panel Study of Income Dynamics, while Sockin and Sockin (2019) find correlations of about 0.9 and 0.8 for the first and second moments, respectively, between industry-occupation pairs using the American Community Survey.

no-location reviews injects measurement error that will likely bias estimates toward zero.[56] We re-estimate the baseline specification for overall ratings, incorporating reviews for which there is a reasonably-high probability the review is from state s, iteratively lowering the threshold $\bar{p}$ for inclusion into the sample, i.e. $p_{k,s} \geq \bar{p}$. Using a lower $\bar{p}$ introduces more measurement error into the state coding. While incorporating these reviews attenuates the magnitude of the effect on overall ratings, estimates remain robustly negative (Table F15).

B.7 Could #MeToo Drive The Results?

We also generate evidence against the possibility that the reputation results can be attributed simply to a broader willingness to come forward with compromising information about low-road firms after the #MeToo movement (in addition to using other states that passed legislation related to #MeToo, per Table F8). On October 5th, 2017, it was first reported by the New York Times that Harvey Weinstein had sexually harassed employees for decades, and this revelation ultimately led to accusations against a number of CEOs including those for Wynn Resorts, Guess?, and CBS.[57] If the #MeToo movement led to a large and persistent shift in workers' willingness to disclose negative information about firms that spread throughout industries prone to using NDAs, particularly in California, then the triple-differences estimate could reflect this change rather than the new legislation. Table F16 shows that there is a decline in overall rating among high-NDA industries in the treated states within three months following the revelation of the Weinstein scandal. However, the effect fades over the subsequent months. We also find little evidence in support of a longer run effect that spills over into the treatment period. In retrospect, this may not be surprising since if it were the #MeToo movement that were driving the effect then we should observe stronger pre-trends pushing in the same direction as the results before NDAs were narrowed.

[56]The same will be true if individuals select a different state than the one they work in.

[57]See Weinstein, Wynn Resorts, Guess?, and CBS for the initial news coverage.

B.8 Robustness to the COVID-19 Pandemic

A final concern is that the results are driven by the COVID-19 pandemic. We address this possibility by leveraging data on monthly deaths and cases per capita by state. The benchmark triple difference equation is re-run but adding controls for state-month cases and deaths per capita interacted with NDA intensity (see Table F20). The information flow, concealment, and reputation estimates are robust to including these controls.

C Replication in other contexts

In this section we analyze the extent to reports of sexual harrassment to the EEOC change after NDAs are narrowed. We also examine a case study of Theranos to test a different form retaliation risk—the public disclosure of information.

C.1 Sexual Harrassment Complaints to the EEOC

To bolster the Glassdoor results in a different setting with a very different reporting structure, we examine how narrowing NDAs affects sexual harassment complaints to the EEOC. Broad NDAs may not legally prevent workers from making high-stakes claims to the EEOC (Ence, 2019), but the potential chilling effects of an NDA itself and a lack of knowledge related to worker rights can still deter individuals from making such claims (Hafiz, 2017; Broden & Mickelsen, 2018). If narrowing NDAs and the concomitant media and educational efforts gave individuals more clarity that they could speak out about unlawful conduct without fear of losing a lawsuit, then we might expect workers who have experienced sexual harassment to be more likely to file a claim.

The EEOC data are available at the gender-state-year (industry is not available), precluding a triple-difference analysis. Given that the outcomes are counts, we estimate a double-difference, Poisson fixed effects specification, including fixed effects for state and year. We weight each cell by employment, such that small cells do not exert an outsize influence on the results. The standard errors are clustered by state. Dynamic event study plots for filings by gender are presented in Figure E12, while the overall estimates are presented in Table F21. Figure E12 shows that EEOC sexual harassment complaints rose 40% overall, 46% for men, and 27% women. The overall effect for women is lower in part because sexual harassment cases for women was somewhat elevated in the years before these laws were enacted, reflecting perhaps that these laws were a response to rising cases related to the #MeToo movement. Overall, even though workers with broad NDAs could legally report cases to the EEOC, these results provide evidence that broad NDAs nevertheless deterred workers from raising claims of misconduct to official agencies.

C.2 Public Disclosure as a Reduction to NDA Retaliation Risk

The main mechanism in the theory is that NDAs are associated with retaliation risk. While our main analysis implies that laws that narrow NDAs reduce this retaliation risk, in this section we exploit a case-study approach that leverages a different way NDA-related retaliation risk can fall: The public disclosure of private information.

In particular, we examine how public disclosure of fraud at Theranos influences information flows, employer ratings, and within-industry dispersion in ratings. This example is informative because Theranos (i) engaged in illegal activity—defrauding investors with a faulty product—and (ii) vigorously used and enforced NDAs as a way to conceal that information, including from other co-workers within the company. Whisteblower Tyler Schultz feared that sharing about the fraud perpetrated at Theranos might violate an NDA he signed when hired (Primeaux, 2019). Similarly, whistleblower Erika Cheung had to sign an NDA even before she interviewed with Theranos, and when she left Theranos she was warned "against posting anything about Theranos on online forums" (Rogal, 2020), as was typical employee exit protocol. They ultimately pushed past these fears and shared negative information with reporter John Carreyrou, who revealed the fraud in a Wall Street Journal article published on October 16, 2015 (Carreyrou, 2015). The exposé also informed other employees at Theranos about the fraud. One Glassdoor review from July 2016 noted, "It was not fun to find out company news in the headlines [and] not in the office."[58]

Since NDAs cannot legally cover public information, the Carreyrou (2015) article reduced the retaliation risk that workers faced from speaking out about this information that may have been previously covered by their NDA. Indeed, public information is expressly carved out in 60% of the NDAs reviewed by Hrdy and Seaman (2023).

Figure E11 reveals how employees' reviews and ratings of Theranos changed after the Carreyrou article, relative to other companies in the biotechnology and pharmaceutical sector. Before the Carreyrou article, Glassdoor reviews of Theranos were more positive than the industry average (panel a) and Theranos had a higher overall rating than its industry's average (panel b). After the public revelation of fraud, however, Theranos employees wrote more negative reviews, such that one-to-two years after the initial Carreyrou article, 59% of the review text for Theranos' reviews was in the 'cons' section, compared with 54% among the rest of the industry. Similarly, overall ratings for Theranos fell from an average above 3.3 before the article to below 2.5 afterwards, well below industry peers whose ratings instead grew over this time period (panel b). These results are consistent with broad NDAs bolstering Theranos's reputation by keeping unlawful conduct from both the public and other employees. After the public revelation (which also reduced retaliation risk), however, employees were more likely to describe the company negatively and give it a low rating.

These results are illustrative but not definitive of the effect of broad NDAs. For one, the shock is an information revelation shock, not a sudden choice by Theranos to narrow their NDAs. Rather, Theranos likely kept using NDAs even in the fallout from the revelation of the fraud, even if employees could now talk about this public information.[59] Moreover, the revelation of the fraud and the public pressure facing Theranos may have also directly contributed to a decline in quality of the work environment, as found in prior work (Zhou

[58] See https://www.refinery29.com/en-us/2019/03/227149/elizabeth-holmes-theranos-glassdoor-reviews-employees.

[59] We were not able to access a Theranos NDA to ascertain the precise language of the contract.

and Makridis, 2021; Lee et al., 2021), though the magnitude of this decline is much larger relative to other scandals (Gadgil and Sockin, 2020).

D Heterogeneity in Ratings

Our heterogeneity analyses are driven by both theoretical and practical concerns. From a theoretical perspective, NDAs threaten legal and financial costs on workers if they speak out and share negative information. Accordingly, narrowing NDAs is likely to have the strongest effects on ratings from workers who have more negative information to share and who experience larger reductions in the probability of negative consequences of violating an NDA. Two hypotheses follow. First, while both current and former employees would face legal risk from violating an NDA, current employees face more substantial potential retaliation by the firm, given that a current worker is still employed there. Therefore, the laws' new protections against firm retaliation for violating an NDA may have greater bite for current employees. Second, given that 84.1% of the sexual harassment charges filed with the EEOC in 2018 are from women, they likely have more negative information to share.

Table F22 shows the results of the main triple-differences specification, but allowing for heterogeneous estimates within various sample partitions. Columns 1 and 2 split the sample by whether the individual is a current or former employee and finds that while both provide more negative reviews after NDAs are narrowed, current employees increase provision of negative information more than former employees, though the difference is not precise. Columns 3 and 4 split the sample into workers who had or currently have a short tenure (at most two years) with the firm or a long tenure (more than two years). Shorter-tenure employees drive the observed negative effects. Columns 5 and 6 split the sample by gender. Contrary to our expectation, men and women similarly provide more-negative reviews. Many respondents do not report their gender, perhaps to purposefully obscure their identity, which adds noise to these cuts of the data. Nevertheless, the results do not suggest that narrowing

NDAs causes women to share more negative information than men.

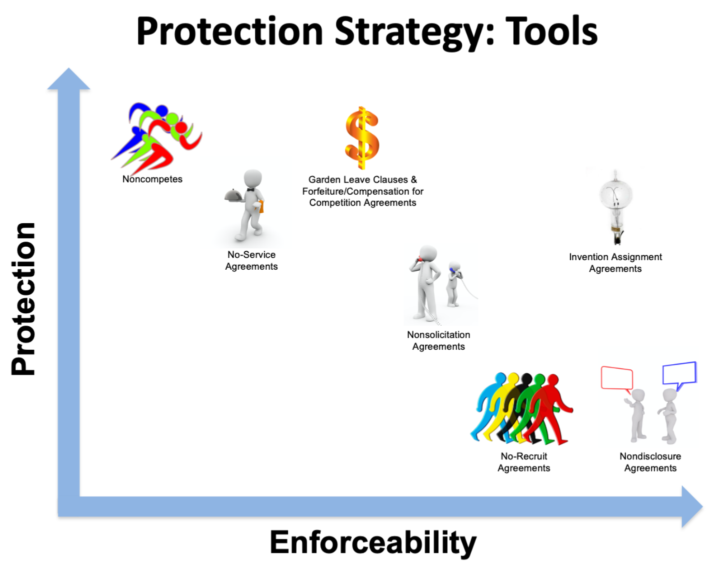
We also examine three dimensions of firm heterogeneity. First, one way that firms can potentially avoid the policies that weakened NDAs is by using choice of law and forum provisions that stipulate that in the event of contract breach a different state's law be applied (Sanga, 2014; Coyle, 2020). While we do not know what choice of law/forum provision each firm has in its employment contract, the empirical specification is designed on the assumption that single-state employers have chosen the laws of the state in which they operate, while multi-state employes could choose myriad state laws. We divide firms based on whether their employer reviews on Glassdoor stem from a single state, or come from multiple states. Columns 7 and 8 reports the results, splitting the sample by whether the firm is a single-state or multi-state employer. Consistent with expectation, we find that the negative reviews stemming from narrowing NDAs are driven more by firms operating in a single state, though the effects are not statistically distinguishable from that observed among multiple-state firms.

Second, we allow for different effects for small and large firms, as workers employed by smaller firms likely face greater risk of retaliation from being less able to blend in among a pool of coworkers. We define a small (large) firm as an employer whose size is below (above) the sample median. Columns 9 and 10 reveal that while the effect is negative for both, it is larger among small firms (though not statistically distinguishable), consistent with workers at smaller firms feeling less burdened by retaliatory risk following the passage of these laws.

Third, an important question is whether the decline is driven by bad-reputation firms receiving worse reviews, or good-reputation firms whose reputations were inflated by the use of NDAs. It is possible that the returns to reputation are non-linear, such that firms with 'good' reputations have stronger incentives to keep negative information from coming out. We classify firm-state pairs by whether the average rating among reviews submitted in 2018—the year prior to enactment of either of the three laws—was above or below average, and re-estimate the triple-differences specification on each sub-sample. The former we label as "high rated" firms and the latter "low rated" ones. In columns 11 and 12, we find that both low-rated and high-rated firms receive more-negative reviews after NDAs are narrowed, and the difference in the declines is not statistically significant.

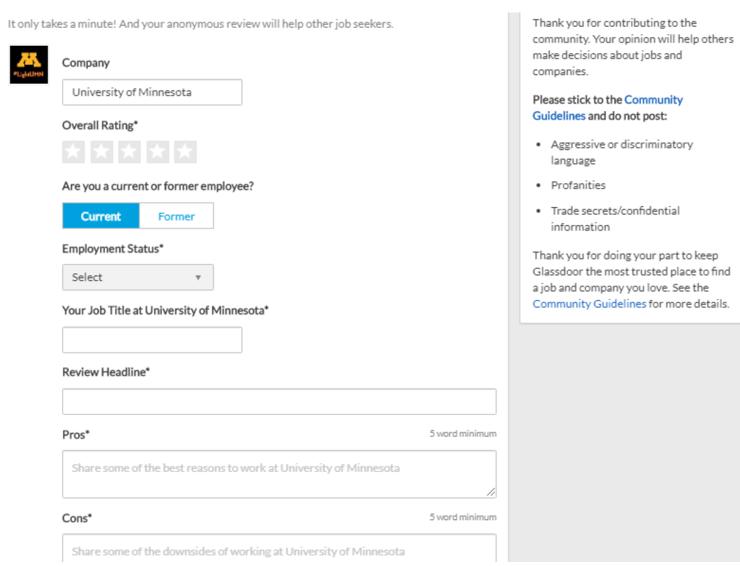
E Additional Figures

Figure E1: Enforceability of NDAs and Related Restrictions



Notes: Figure available in Beck (2020).

Figure E2: Example of Blank Employer Review Form


It only takes a minute! And your anonymous review will help other job seekers.

Company

University of Minnesota

Overall Rating*

Are you a current or former employee?

Current Former

Employment Status*

Select

Your Job Title at University of Minnesota*

Review Headline*

Pros* 5 word minimum

Share some of the best reasons to work at University of Minnesota

Cons* 5 word minimum

Share some of the downsides of working at University of Minnesota

Thank you for contributing to the community. Your opinion will help others make decisions about jobs and companies.

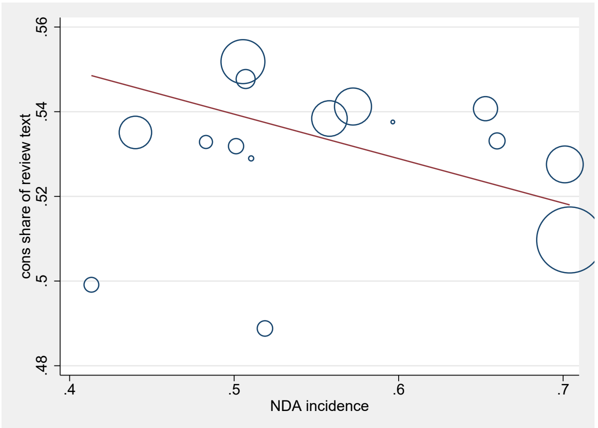
Please stick to the Community Guidelines and do not post:

- Aggressive or discriminatory language
- Profanities
- Trade secrets/confidential information

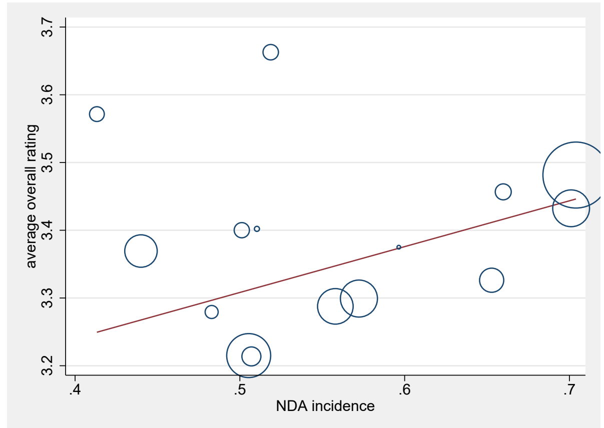
Thank you for doing your part to keep Glassdoor the most trusted place to find a job and company you love. See the Community Guidelines for more details.

Notes: This figure is a screenshot of submitting an employer review for the University of Minnesota.

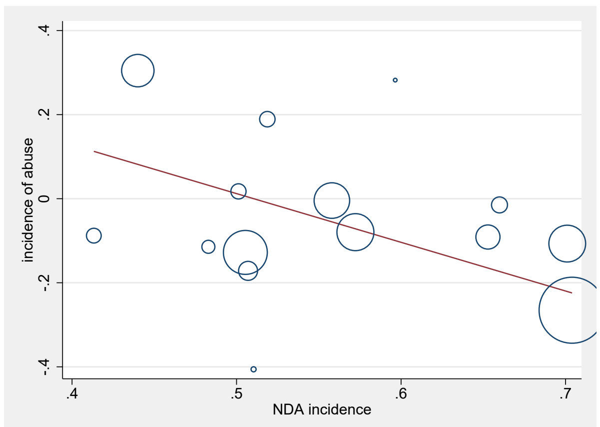
Figure E3: Correlation of Glassdoor and RAND American Working Conditions Survey Measures with NDA Incidence across Industries



(a) Cons share of review text



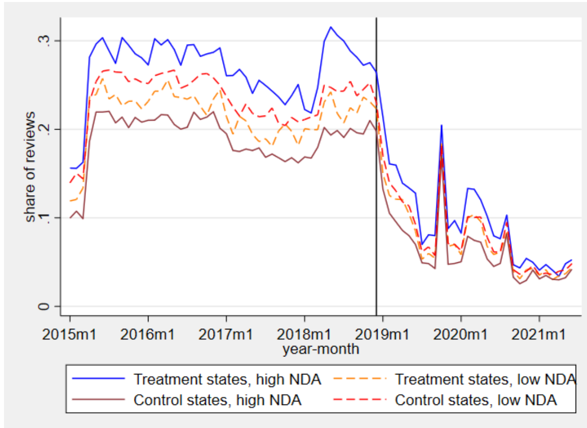
(b) Overall rating



(c) Workplace Abuse from the RAND 2015 American Working Conditions Survey

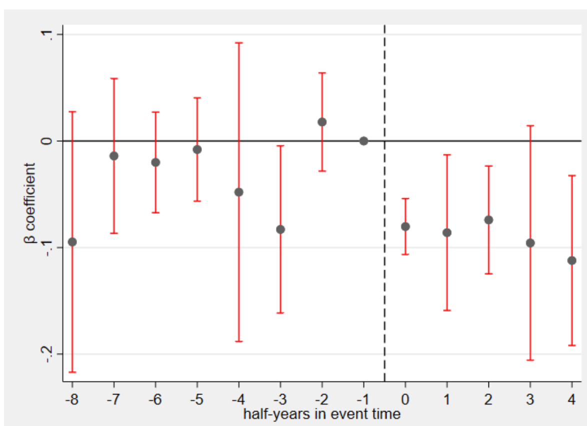
Notes: These figures display scatterplots between industries of the average Cons share of review text and the average overall star rating from Glassdoor, as well as a standardized z-score measure of abuse or harassment from the 2015 American Working Conditions Survey. Each dot indicates a different industry, and dots are weighted according to their sample size of Glassdoor reviews. The standardized measure from the American Working Conditions Survey reflects the aggregation of responses to q70a (verbal abuse), q70b (unwanted sexual attention), q70c (threats), q70d (humiliating behavior), q71a (physical violence), q71b (bullying/harassment), and q71c (sexual harassment).

Figure E4: Structural Shift in Rates of Job Title Concealment Among 1–3 Stars Reviews



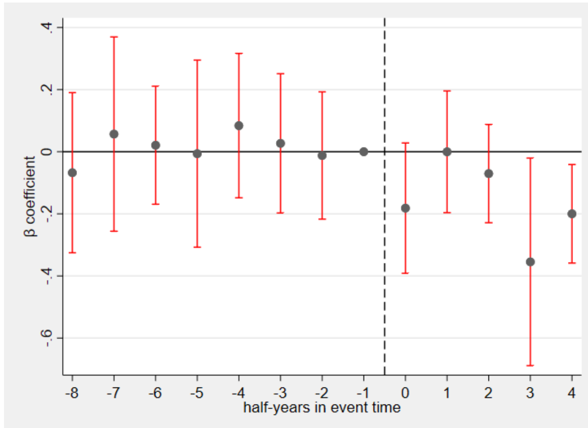
Notes: This figure plots the average rate of job title concealment among new reviews submitted each calendar month over the full sample period. Sample of reviews is partitioned into four subsets: treatment states (California, Illinois, and New Jersey) and control states (all others), as well as high NDA industries (above-average NDA intensity) and low NDA industries (below-average NDA intensity). Vertical line indicates the end of 2018, when there is a structural break in rates of job title concealment, likely associated with structural changes to the review submission platform. This structural shift occurs simultaneously with the enactment of California's legislation and a few months prior to the enactment of New Jersey's legislation. However, the change affected both treatment and control states and industries similarly.

Figure E5: Narrowing NDAs and Job Title Concealment for Reviews Where Length of 'Cons' Section is Longer than the 'Pros' Section, Dynamic Response



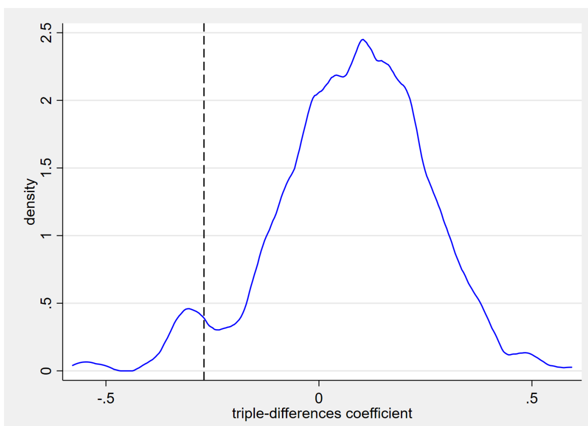
Notes: The dependent variable is an indicator for whether the reviewer concealed their job title, conditional on the 'cons' section being longer than the 'pros' section. The sample period is 2015–2021 and point estimates are relative to the calendar half-year before the legislation goes into effect. Regressions include firm-state, industry-year-month, and state-year-month fixed effects. Standard errors are two-way clustered by industry and state. Red vertical bars indicate 95% confidence intervals around each point estimate.

Figure E6: Narrowing NDAs and Job Title Concealment with Worker Fixed Effects, Dynamic Response



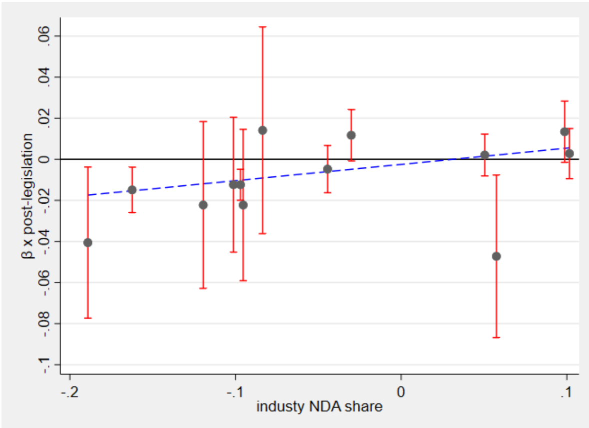
Notes: The dependent variable is an indicator for whether the reviewer concealed their job title. The sample period is 2015–2021 and point estimates are relative to the calendar half-year before the legislation goes into effect. Regressions include worker, firm-state, industry-year-month, and state-year-month fixed effects. Standard errors are clustered by industry cross state. Red vertical bars indicate 95% confidence intervals around each point estimate.

Figure E7: Distribution of Triple-Differences Estimates Under Randomized Inference

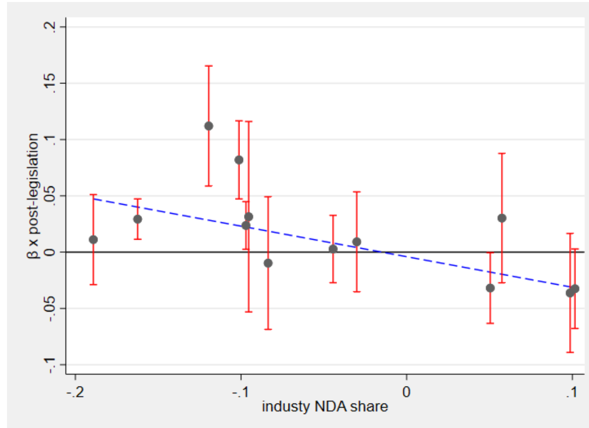


Notes: The figure illustrates the density of triple-differences coefficients for overall rating when states are randomly assigned between treatment and control groups. There are three possible treatment periods: January 2019 (corresponding to California), March 2019 (corresponding to New Jersey), and January 2020 (corresponding to Illinois). We draw randomly from a uniform distribution to assign one of the fifty states or the District of Columbia to January 2019, a second of the fifty-one to March 2019, and a third to January 2020. The remaining forty-eight are assigned to the control group. We re-draw if the same state is assigned to two treatments. We then record the estimate from estimating a triple-differences specification. We repeat this procedure 500 times and plot the distribution. The dashed line indicates the main triple-differences estimate under the true treatment and control assignment.

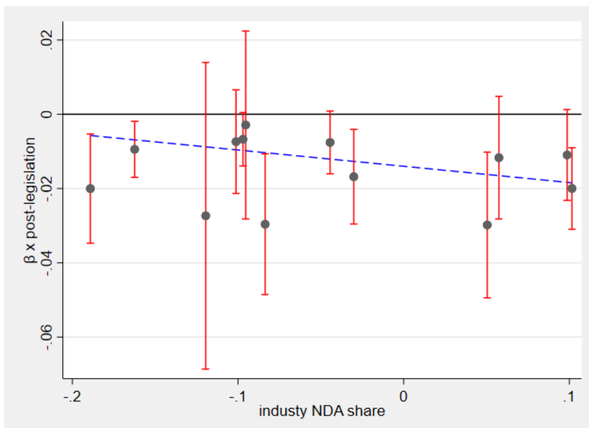
Figure E8: Industry-Specific Difference-in-Differences Estimates



(a) Log length cons section



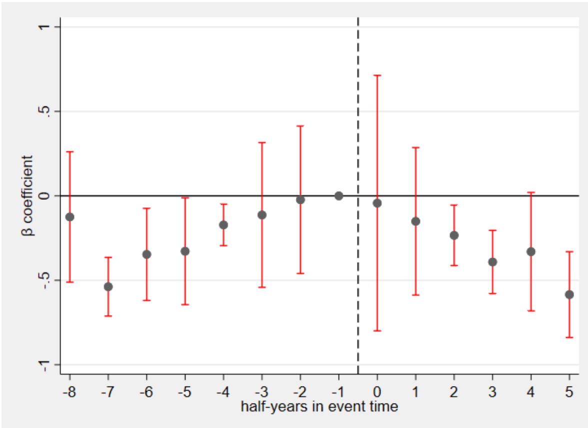
(b) Overall rating



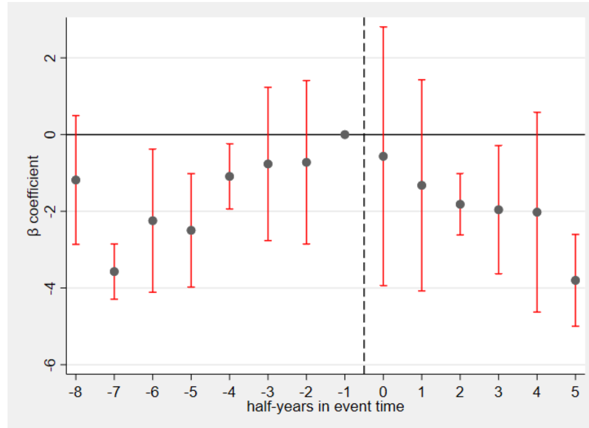
(c) Conceals job title | longer cons than pros

Notes: These figures display difference-in-differences estimates from comparing CA-IL-NJ to all other states before and after NDAs are narrowed, separately for each industry, where the x-axis reflects demeaned NDA intensity. Regressions include firm-state and year-month fixed effects. Vertical red bars indicate a 95% confidence interval around each point estimate. Standard errors are clustered by state. Dashed blue line reflects linear lines of best fit through the point estimates, with industries weighted by their respective sample sizes. From left-to-right, the industries are: Construction, Accommodation and Food Services, Other Services, Arts and Entertainment, Retail Trade, Transportation and Warehousing, Real Estate, Health Care and Social Assistance, Manufacturing, Information, Utilities, Finance and Insurance, and Professional, Scientific and Technical Services. Excluded are two industries—Agriculture and Mining—for which standard errors are particularly large due to thin samples.

Figure E9: Articles and Respective Word Counts Mentioning NDAs, Dynamic Responses



(a) Log number of articles

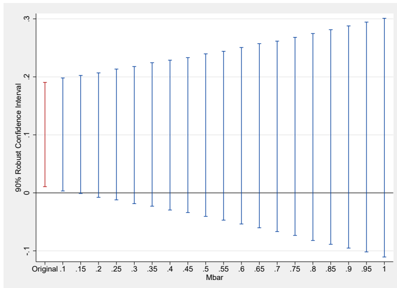


(b) Log number of words

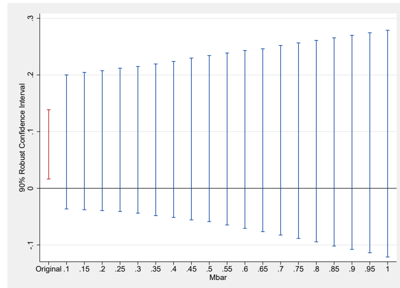
Notes: The figures present dynamic difference–in–difference models estimating how the extent of news coverage pertaining to NDAs evolved after California, Illinois, and New Jersey passed laws narrowing NDAs compared with all other states. Data on the number of news articles discussing NDAs and the length of such articles are obtained at the state-month level following Rees-Jones and Rozema (2023). Red vertical bars indicate 95% confidence intervals around each point estimate.

Figure E10: Rambachan and Roth (2023) Test for First and Full Post-Treatment Periods on the Provision of Negative Information

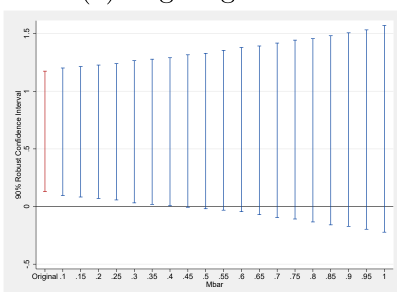
First Post-Period | Full Post-Period



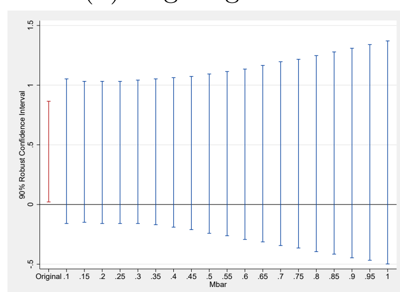
(a) Log length cons



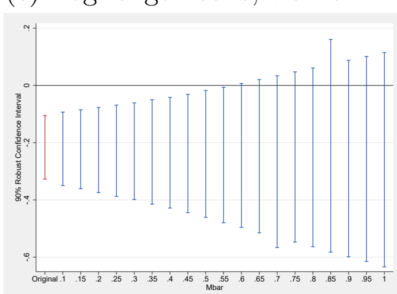
(b) Log length cons



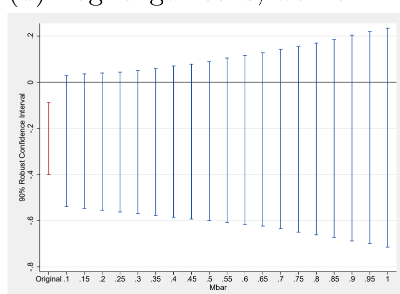
(c) Log length cons, worker FE



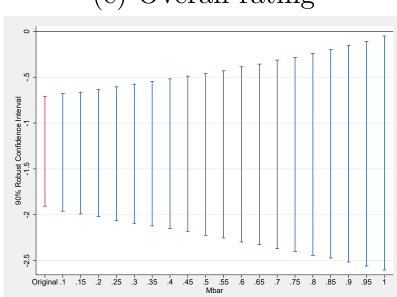
(d) Log length cons, worker FE



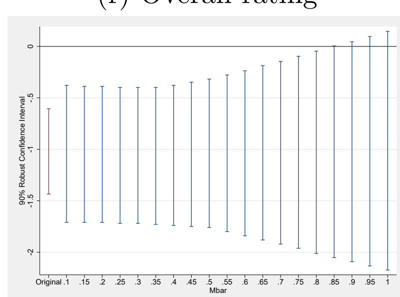
(e) Overall rating



(f) Overall rating



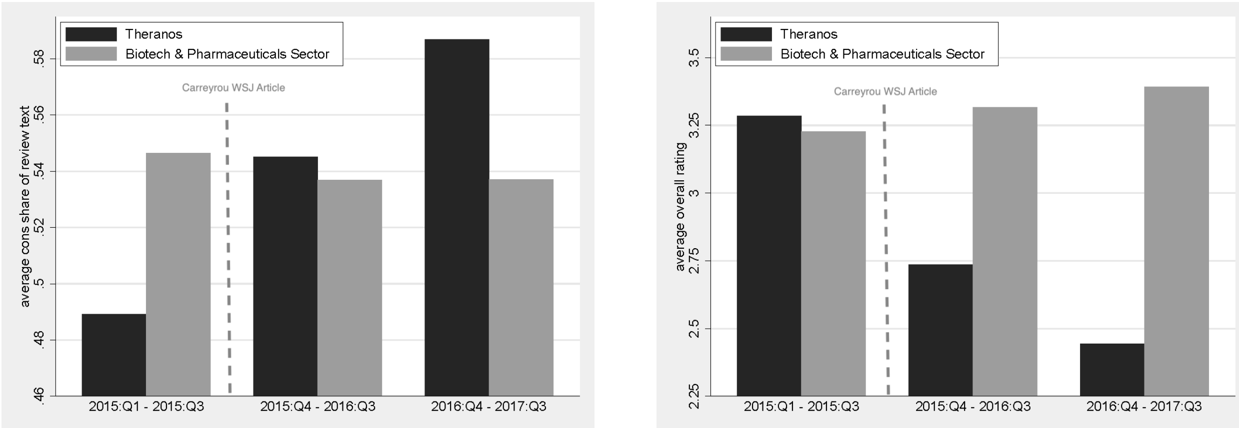
(g) Overall rating, worker FE



(h) Overall rating, worker FE

Notes: The figures above reflect the robust confidence intervals from implementing the approach of Rambachan and Roth (2023) using the *honestdid* package in Stata for the triple-differences event studies of each main outcome. Post-trends can be violated by no more than a multiple ($\bar{M}$) of the largest period-to-period change in the pre-trends. Vertical blue bars represent 90% confidence intervals.

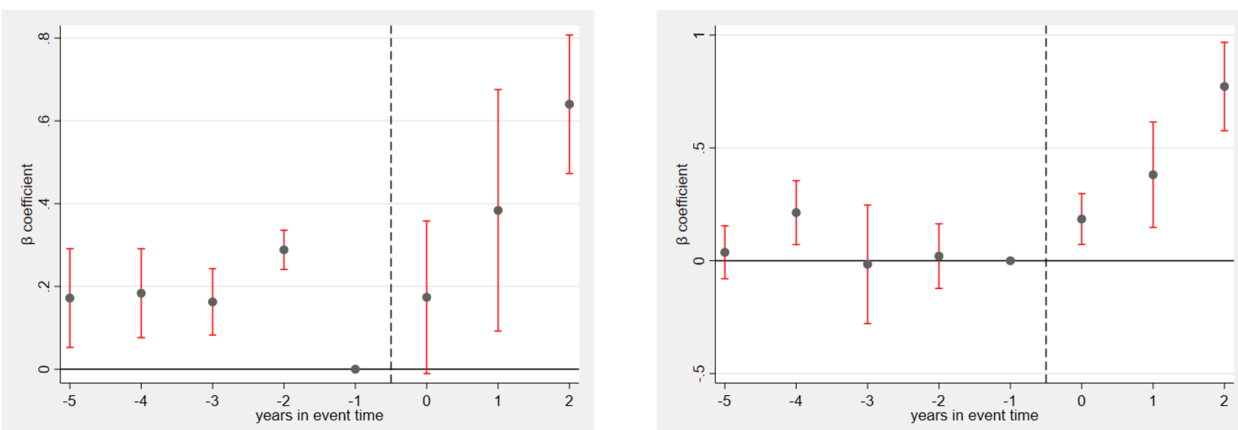
Figure E11: Employee Sentiment Before and After Wall Street Journal Report on Theranos



(a) Cons share of review text (b) Overall rating

Notes: Figure displays the average share of review text attributable to the cons section (panel a) and average overall rating (panel b) among reviews from Theranos (dark gray, left-side bars) and all other employers in the Biotechnology and Pharmaceuticals sector (light gray, right-side bars). Each bar reflects the sample average during the time periods along the x-axis. Wall Street Journal report on Theranos was published on October 16, 2015 such that left-most bar in the graph refers to the time period before the information revelation, while the other two are in the post-revelation phase.

Figure E12: Narrowing NDAs and Complaints to EEOC, Dynamic Responses



(a) Female EEOC filings (b) Male EEOC filings

Notes: The dependent variable is the count of charges filed alleging sexual harassment by gender at the state-year level. The model is estimated via a fixed effects poisson model with fixed effects for state and year. The sample period is 2015–2021 and point estimates are relative to the calendar year before legislation goes into effect. Standard errors are clustered by state. Red vertical bars indicate 95% confidence intervals around each point estimate.

F Additional Tables

Table F1: Summary Statistics for Glassdoor Dependent Variables

Dependent variable	Reviews (millions)	Mean	Median	Standard deviation	5th percentile	95th percentile
	(1)	(2)	(3)	(4)	(5)	(6)
Overall rating	3.88	3.49	4.00	1.41	1.00	5.00
Career opportunities	3.44	3.29	3.00	1.46	1.00	5.00
Compensation and benefits	3.44	3.37	4.00	1.36	1.00	5.00
Culture and values	3.43	3.44	4.00	1.52	1.00	5.00
Senior leadership	3.40	3.15	3.00	1.55	1.00	5.00
Work-life balance	3.44	3.38	4.00	1.44	1.00	5.00
Conceal job title | longer cons than pros	2.07	0.14	0.00	0.35	0.00	1.00
Would refer a friend to firm	3.30	0.63	1.00	0.48	0.00	1.00
Positive business outlook	3.17	0.54	1.00	0.50	0.00	1.00
Approve of CEO	2.66	0.52	1.00	0.50	0.00	1.00
References harassment in text	3.88	0.02	0.00	0.14	0.00	0.00
References illegality in text	3.88	0.00	0.00	0.04	0.00	0.00
Offers management advice	3.88	0.56	1.00	0.50	0.00	1.00
Log length of review text	3.88	5.28	5.15	0.87	4.14	6.88
Log length of pros section	3.88	4.42	4.23	0.87	3.30	6.09
Log length of cons section	3.88	4.54	4.36	1.00	3.26	6.44
Pros share of review text	3.88	0.48	0.48	0.20	0.13	0.82

Notes: Table provides summary statistics (mean, median, standard deviation, fifth percentile, and ninety-fifth percentile) for each of the review-level dependent variables.

Table F2: Effect of Narrowing NDAs on Review Outcomes with Worker Fixed Effects

	Cons share of review text	Log length cons section	Overall rating	Conceal job title
	(1)	(2)	(3)	(4)
Narrowed NDAs x NDA intensity	0.058**	0.344**	-0.582***	-0.147**
	(0.027)	(0.149)	(0.220)	(0.061)
Observations	290188	290188	290188	290188
Adjusted R^2	0.33	0.41	0.49	0.22

Notes: The table conveys how other outcomes related to newly submitted employer reviews changed, according to the triple-differences specification, following the narrowing of NDAs. The dependent variable in each regression is listed as the header of each column. Regressions include worker, firm-state, industry-year-month, and state-year-month fixed effects. Standard errors are clustered by industry cross state. Significance levels: * 10%, ** 5%, *** 1%.

Table F3: Narrowing NDAs and Alternative Outcomes for Employee Sentiment

	Star ratings					Indicators		
	Culture and values	Senior mgmt.	Career opp.	Comp. and benefits	Work-life balance	Would refer a friend to firm	Positive business outlook	Approve of CEO
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Narrowed NDAs x NDA intensity	-0.200***	-0.292***	-0.201**	-0.228***	-0.184***	-0.098***	-0.046*	-0.044**
	(0.059)	(0.071)	(0.077)	(0.064)	(0.033)	(0.013)	(0.024)	(0.015)
Dependent variable mean	3.435	3.137	3.285	3.364	3.370	0.629	0.535	0.521
Observations	3212937	3187633	3226611	3223887	3228388	3092463	2968011	2514093
Adjusted R^2	0.16	0.16	0.15	0.19	0.15	0.11	0.13	0.15

Notes: The table conveys how other outcomes related to newly submitted employer reviews changed, according to the triple-differences specification, following the narrowing of NDAs. The dependent variable in each regression is listed as the header of each column. Regressions include firm-state, industry-year-month, and state-year-month fixed effects. Standard errors are two-way clustered by industry and state. Significance levels: * 10%, ** 5%, *** 1%.

Table F4: Narrowing NDAs and Arrival of New Reviews

	Full US Within High NDA Ind.		Within CA-IL-NJ High vs. Low NDA Ind.		Triple Difference	
	(1)	(2)	(3)	(4)	(5)	(6)
CA-IL-NJ	875					
	(799)					
Narrowed NDAs	1031	829*			619	
	(645)	(430)			(355)	
Narrowed NDAs x NDA intensity			4368	4474	2792	2770
			(4242)	(4437)	(1854)	(2151)
Dependent variable mean	702	702	1461	1461	425	425
Observations	2652	2652	585	585	9113	9113
Adjusted R^2	0.39	0.43	0.59	0.60	0.87	0.91
Half-Year FE	✓	✓	✓			
Industry FE	✓		✓	✓		
State FE		✓	✓	✓		
Industry-Half-Year FE					✓	✓
State-Industry FE					✓	✓
State-Half-Year FE				✓		✓

Notes: The dependent variable is the number of reviews submitted within an industry-state-half year. Regressions in first two columns are clustered by state, next two columns clustered by industry, and the final two columns clustered by industry and state. Significance levels: * 10%, ** 5%, *** 1%.

Table F5: Narrowing NDAs and Incorporating NDA News Coverage, Coefficients

	Cons share of review text		Log length cons section	
	(1)	(2)	(3)	(4)
Narrowed NDAs x NDA intensity	0.016**	0.009	0.074***	0.070***
	(0.008)	(0.007)	(0.018)	(0.020)
Log(once-lagged state-halfyear NDA articles) x NDA intensity		0.004		0.005
		(0.002)		(0.008)
Log(twice-lagged state-halfyear NDA articles) x NDA intensity		0.006		0.005
		(0.006)		(0.021)
Log(thrice-lagged state-halfyear NDA articles) x NDA intensity		0.012***		0.019
		(0.003)		(0.018)
Log(fourth-lagged state-halfyear NDA articles) x NDA intensity		0.002		0.003
		(0.005)		(0.025)
Log(once-lagged state-halfyear NDA words) x NDA intensity		-0.000		0.000
		(0.001)		(0.003)
Log(twice-lagged state-halfyear NDA words) x NDA intensity		-0.002		-0.003
		(0.002)		(0.006)
Log(thrice-lagged state-halfyear NDA words) x NDA intensity		-0.003***		-0.005
		(0.001)		(0.004)
Log(fourth-lagged state-halfyear NDA words) x NDA intensity		0.000		-0.001
		(0.001)		(0.005)

Notes: The table implements the benchmark triple-differences model estimating the causal effect narrowing NDAs had on the cons share of the review text and the log length of the Cons section when relative news coverage regarding NDAs across states over time is incorporated. Regressions include firm-state, industry-year-month, and state-year-month fixed effects. Standard errors are two-way clustered by industry and state. Significance levels: * 10%, ** 5%, *** 1%.

Table F6: Narrowing NDAs and Standard Deviation of Ratings Within and Across Firms

	Overall	Within firms	Across firms
	(1)	(2)	(3)
Narrowed NDAs x NDA intensity	0.047*	0.122**	0.061**
	(0.027)	(0.053)	(0.031)
Dependent variable mean	1.39	1.04	1.30
Mean firms per industry-state-half	149	160	149
Observations	9443	8759	9401
Adjusted R^2	0.68	0.19	0.62

Notes: The table implements triple-difference models for estimating the causal effect narrowing NDAs has on the standard deviation of overall ratings within and between firms. Each observation represents an industry-state-half year. Regressions include firm-state, industry-year-half, and state-year-half fixed effects. Regressions are clustered by industry-state. Each industry-state is weighted by the average number of firms in the sample each half-year. Significance levels: * 10%, ** 5%, *** 1%.

Table F7: Stacked Designs, Weighting, and Alternative Measures of NDA Intensity

Dependent variable	Stacked (1)	Equally weighted (2)	Alternative NDA intensity: Occupation (3)	Alternative NDA intensity: Industry x occupation (4)
Log length cons section	0.075***	0.266**	0.037**	0.071***
	(0.018)	(0.095)	(0.018)	(0.017)
N	9890031	3645332	2801470	2157826
Overall rating	-0.274***	-0.372**	-0.144**	-0.179***
	(0.080)	(0.171)	(0.058)	(0.061)
N	9890031	3645332	2801470	2157826
Conceals job title | longer cons than pros	-0.060***	-0.054	–	–
	(0.012)	(0.033)	–	–
N	5293792	1898057	–	–

Notes: The table conveys the triple-difference estimates when alternative regression specifications or alternative measures of NDA intensity are used in lieu of the benchmark model with industry-level NDA intensity. The "stacked" model replicates the control sample thrice–for each treatment state–incorporating firm-state, state-year-month, industry-year-month, and treatment-state sample fixed effects. Standard errors are two-way clustered by industry and state. The "equally weighted" model weights each review by $1/N_{\iota(k)st}$ such that each industry-state-year receives equal weight, where the sample is restricted to industry-state pairings that receive on average at least 100 reviews annually. Standard errors are two-way clustered by industry and state. The "occupation" model use occupation level NDA intensity from Payscale data in lieu of industry level incorporating firm-state, occupation-state, occupation-year-month, and state-year-month fixed effects and restricting the sample to occupations with at least 20 Payscale observations. Standard errors are two-way clustered by occupation and state. The "industry x occupation" model use industry-occupation level NDA intensity from Payscale data in lieu of industry level incorporating firm-state, occupation-state, occupation-industry, occupation-year-month, industry-year-month, and state-year-month fixed effects and restricting the sample to industry-occupation pairs with at least 20 Payscale observations. Standard errors are two-way clustered by industry-occupation and state. Significance levels: * 10%, ** 5%, *** 1%.

Table F8: Alternative Choice of Control States

Dependent variable	All states (1)	Neighbor states (2)	High coverage states (3)	Weaker legislation states (4)
Log length cons section	0.074***	0.058*	0.066**	0.119***
	(0.018)	(0.029)	(0.023)	(0.026)
N	3645332	1628960	2515222	1487320
Overall rating	-0.270***	-0.265**	-0.235**	-0.190*
	(0.079)	(0.102)	(0.082)	(0.094)
N	3645332	1628960	2515222	1487320
1(Conceals job title | longer cons than pros)	-0.058***	-0.054**	-0.042***	-0.041
	(0.009)	(0.022)	(0.011)	(0.025)
N	1898057	838157	1300171	756281
IQR across-firm ratings	0.463***	0.411**	0.578***	0.597***
	(0.141)	(0.183)	(0.189)	(0.173)
N	9401	2667	2519	1880

Notes: The table illustrates the robustness of our triple-differences estimates to the choice of control sample. "All states" includes the other 47 states and Washington DC. "Neighbor states" reflects the 11 states (Arizona, Delaware, Indiana, Iowa, Kentucky, Missouri, Nevada, New York, Oregon, Pennsylvania, and Wisconsin) that share a contiguous border with California, Illinois, or New Jersey. "High coverage states" refer to the 10 states (other than California, Illinois and New Jersey) that represent at least 2.5% of the review sample (Florida, Georgia, Massachusetts, New York, North Carolina, Ohio, Pennsylvania, Texas, Virginia, and Washington). "Weaker legislation states" refers to the 7 states that implemented more narrowly-focused legislation regarding the use of non-disclosures around the same time (Maryland, New York, Oregon, Tennessee, Virginia, Vermont, and Washington). These laws were weaker than those studied here because they typically covered only NDAs in the context of sexual harassment. See Johnson et al. (2019) for details. Standard errors are two-way clustered by industry and state. Significance levels: * 10%, ** 5%, *** 1%.

Table F9: Isolating the Treatment Effect of Narrowing NDAs on Overall Rating, by State

	Lone Treatment State					
	California		Illinois		New Jersey	
	(1)	(2)	(3)	(4)	(5)	(6)
Narrowed NDAs x NDA intensity	-0.369***	-0.705***	-0.138***	-0.228	0.045	-0.467
	(0.031)	(0.249)	(0.033)	(0.307)	(0.040)	(0.385)
Observations	3371488	261223	3010818	222615	2920304	212698
Adjusted R^2	0.15	0.49	0.15	0.49	0.15	0.49
Worker FE		✓		✓		✓

Notes: The table illustrates the heterogeneity by treatment state underlying the triple-differences estimate for overall rating. For each state listed in the column sub-headers, we exclude the other two treatment states entirely and estimate the main triple-difference with that lone treatment state. For columns (1) and (2), Illinois and New Jersey are dropped; for columns (3) and (4), California and New Jersey; for columns (5) and (6), California and Illinois. Regressions include firm-state, industry-year-month, and state-year-month fixed effects. Standard errors are two-way clustered by industry and state for odd columns and by industry cross state for even columns. Significance levels: * 10%, ** 5%, *** 1%.

Table F10: Narrowing NDAs and Review Outcomes, Excluding the Professional, Scientific, and Technical Services Industry in California

	Log length cons section		Conceal job title if longer cons than pros		Overall rating	
	Full sample	When excluded	Full sample	When excluded	Full sample	When excluded
	(1)	(2)	(3)	(4)	(5)	(6)
Narrowed NDAs x NDA intensity	0.074***	0.103***	-0.058***	-0.030*	-0.270***	-0.227***
	(0.025)	(0.031)	(0.017)	(0.018)	(0.047)	(0.064)
Observations	3645332	3424266	1898057	1798555	3645332	3424266

Notes: The table implements the benchmark triple-differences model estimating the causal effect narrowing NDAs had on the average log length of the 'cons' section of reviews, the average overall star rating, and the likelihood of concealing one's job title when supplying a negative review when reviews from employers in California's professional, scientific, and technical services industry are excluded. Regressions include firm-state, industry-year-month, and state-year-month fixed effects. Standard errors are clustered by industry and state. Significance levels: * 10%, ** 5%, *** 1%.

Table F11: Effects of Narrowing NDAs on Review Outcomes, Excluding the Professional, Scientific, and Technical Services Industry in All States

	Log length cons section		Conceal job title if longer cons than pros		Overall rating	
	Full sample	When excluded	Full sample	When excluded	Full sample	When excluded
	(1)	(2)	(3)	(4)	(5)	(6)
Narrowed NDAs x NDA intensity	0.074***	0.109***	-0.058***	-0.042**	-0.270***	-0.275***
	(0.025)	(0.034)	(0.017)	(0.021)	(0.047)	(0.070)
Observations	3645332	2483748	1898057	1348176	3645332	2483748

Notes: The table implements the benchmark triple-differences model estimating the causal effect narrowing NDAs had on the average log length of the 'cons' section of reviews, the average overall star rating, and the likelihood of concealing one's job title when supplying a negative review when reviews from employers in the professional, scientific, and technical services industry are excluded. Regressions include firm-state, industry-year-month, and state-year-month fixed effects. Standard errors are clustered by industry and state. Significance levels: * 10%, ** 5%, *** 1%.

Table F12: Effects of Narrowing NDAs on Review Outcomes, Only Firms with an Establishment in a Treated and a Control State

	Cons share of review text	Log length cons section	Overall rating	Conceal job title if longer cons than pros
	(1)	(2)	(3)	(4)
Narrowed NDAs x NDA intensity	0.030***	0.106***	-0.374***	-0.069***
	(0.007)	(0.024)	(0.050)	(0.011)
Observations	2932329	2932329	2932329	1578722
Adjusted R^2	0.10	0.14	0.14	0.12

Notes: The table implements the benchmark triple-differences model estimating the causal effect narrowing NDAs had on the average log length of the 'cons' section of reviews, the average overall star rating, and the likelihood of concealing one's job title when supplying a negative review when the sample is restricted to only reviews from employers that have establishments in one of the three treated states and a control state. Regressions include firm-state, industry-year-month, and state-year-month fixed effects. Standard errors are clustered by industry and state. Significance levels: * 10%, ** 5%, *** 1%.

Table F13: Alternative Choice of Standard Error Clustering

	Clustering of standard errors				
Dependent variable	Industry and state	State	Industry cross state	Firm and state	Wild cluster bootstrap
	(1)	(2)	(3)	(4)	(5)
Log length cons section	0.074***	0.074***	0.074***	0.074***	0.077**
	(0.018)	(0.026)	(0.025)	(0.019)	[0.045]
Overall rating	-0.270***	-0.270***	-0.270***	-0.270***	-0.273*
	(0.079)	(0.086)	(0.047)	(0.083)	[0.082]
Conceals job title | longer cons than pros	-0.058***	-0.058***	-0.058***	-0.058***	-0.057**
	(0.009)	(0.015)	(0.017)	(0.012)	[0.022]
IQR across-firm ratings	0.463***	0.463***	0.463**	–	0.463*
	(0.141)	(0.132)	(0.188)	–	[0.098]

Notes: The table re-estimates each of the main results under different clustering methods for the standard errors. The baseline clustering method is two-way clustering by industry and state. For ease of implementing wild cluster bootstrapping due to large sample sizes for the log length of cons section, overall rating, and job title concealment, we relax the time-related fixed effects to industry-year and state-year. Standard errors are reported in parentheses. For wild cluster bootstrapping, we conduct 5,000 replications and report p-values in brackets. Significance levels: * 10%, ** 5%, *** 1%.

Table F14: Effect on Average Rating Accounting for Differences in Wage Growth

	Level of aggregation for mean earnings				
	Industry x State	Occupation x State	Industry x Occupation x State	Firm x State	Firm x Occupation x State
	(1)	(2)	(3)	(4)	(5)
Narrowed NDAs x NDA intensity	-0.301***	-0.292***	-0.301***	-0.283***	-0.254***
	(0.073)	(0.063)	(0.072)	(0.069)	(0.046)
Mean log earnings	-0.036	0.147***	0.157***	0.169***	0.193***
	(0.040)	(0.026)	(0.026)	(0.013)	(0.029)
Observations	2949937	2040339	1989399	2423585	1164850

Notes: The table conveys the triple-difference estimates when average pay within a labor market is incorporated into the model. Mean log earnings is calculated using Glassdoor pay data by calendar half-year at the level of aggregation detailed in the header of each column. Sample excludes reviews from the first half of 2021 because our Glassdoor pay dataset extends only as far as 2020. The dependent variable in each regression is employee overall star rating. NDA intensity reflects the benchmark industry level. Samples for Columns 2, 3, and 5 are necessarily restricted to reviews for which job title is available. Standard errors are two-way clustered by industry and state. Significance levels: * 10%, ** 5%, *** 1%.

Table F15: Imputation of Missing Location

	Overall rating				
	(1)	(2)	(3)	(4)	(5)
Narrowed NDAs x NDA intensity	-0.270***	-0.236***	-0.205***	-0.206***	-0.213***
	(0.079)	(0.074)	(0.069)	(0.069)	(0.071)
Threshold for including missing location	none	100%	75%	50%	25%
Dependent variable mean	3.479	3.471	3.469	3.468	3.462
Observations	3645332	3967305	4473555	4917232	5498879
Adjusted R^2	0.15	0.15	0.15	0.15	0.15

Notes: The table explores whether the exclusion of reviews without the location left blank is driving the reduction in average ratings following the passage of these laws. The dependent variable in each regression is employee overall star rating. Regressions include firm-state, industry-year-month, state-year-month, a location left blank dummy–year–month fixed effects. Standard errors are two-way clustered by industry and state. Significance levels: * 10%, ** 5%, *** 1%.

Table F16: Effect on Overall Rating by NDA Intensity Following Weinstein Scandal

	CA only		CA, IL, NJ		Triple difference
	3 mths	12 mths	3 mths	12 mths	
	(1)	(2)	(3)	(4)	(5)
After Weinstein Scandal x NDA intensity	-0.254**	-0.004	-0.279**	-0.081	-0.062
	(0.091)	(0.051)	(0.057)	(0.060)	(0.080)
Narrowed NDAs x NDA intensity					-0.270***
					(0.079)
Observations	85972	133650	131991	203440	3645332
Adjusted R^2	0.20	0.19	0.20	0.18	0.15
Firm FE	✓	✓	✓	✓	
State-Year-Month FE	✓	✓	✓	✓	✓
Industry-Year-Month FE					✓
Firm-State FE					✓

Notes: The table reflects how overall ratings evolve in California, Illinois, and New Jersey following the public revelation of the Harvey Weinstein scandal on October 5th, 2017. The pre-period for each specification is the twelve calendar months preceding this date. Short-term and long-term effects following this event are estimated using post-periods of three and twelve calendar months, respectively. Standard errors for the former two columns are clustered by industry and the latter three columns by industry and state. Significance levels: * 10%, ** 5%, *** 1%.

Table F17: Narrowing NDAs and the Extent of Review Planting

	Growth threshold for identifying sock puppetry		
	100%	50%	25%
	(1)	(2)	(3)
Narrowed NDAs x NDA intensity	-0.000	-0.004	-0.044**
	(0.015)	(0.011)	(0.017)
Dependent variable mean	0.073	0.210	0.356
Observations	3645332	3645332	3645332
Adjusted R^2	0.43	0.39	0.37

Notes: The table explores whether the presence of reviews identified as possible sock puppetry changes following the passage of these laws. For each firm in each year-month, we calculate the log change in reviews relative to the three months prior (g_{kt}^{B}) and the three months after (g_{kt}^{A}). The x% percent cutoff refers to firm-year-months in which $g_{kt}^{B} \geq x\%$ and $g_{kt}^{A} \geq x\%$. The dependent variable for each specification reflects an indicator variable for satisfying this criteria. Regressions include firm-state, industry-year-month, and state-year-month fixed effects. Standard errors are two-way clustered by state and industry. Significance levels: * 10%, ** 5%, *** 1%.

Table F18: Accounting for Possible Review Planting in Submission of Reviews

	Overall rating			
	(1)	(2)	(3)	(4)
Narrowed NDAs x NDA intensity	-0.270***	-0.278***	-0.259***	-0.249***
	(0.079)	(0.075)	(0.071)	(0.060)
Growth threshold for excluding reviews	none	100%	50%	25%
Dependent variable mean	3.479	3.432	3.419	3.420
Observations	3645332	3365791	2841821	2308570
Adjusted R^2	0.15	0.14	0.13	0.13

Notes: The table explores whether the presence of reviews identified as possible sock puppetry are driving the reduction in average ratings following the passage of these laws. For each firm in each year-month, we calculate the log change in reviews relative to the three months prior (g_{kt}^B) and the three months after (g_{kt}^A). The x% percent cutoff refers to firm-year-months in which $g_{kt}^B \geq x\%$ and $g_{kt}^A \geq x\%$. The sample for each specification *excludes* reviews from all firm-year-months satisfying this criteria. The dependent variable in each regression is employee overall star rating. Regressions include firm-state, industry-year-month, and state-year-month fixed effects. Standard errors are two-way clustered by state and industry. Significance levels: * 10%, ** 5%, *** 1%.

Table F19: Narrowing NDAs and Overall Ratings, Incorporating NDA News Coverage

	Cons share of review text		Log length cons section	
	(1)	(2)	(3)	(4)
Narrowed NDAs x NDA intensity	0.016**	0.009	0.074***	0.070***
	(0.008)	(0.007)	(0.018)	(0.020)
With two years lagged NDA news x NDA intensity		✓		✓

Notes: The table implements the benchmark triple-differences model estimating the causal effect narrowing NDAs had on the log length of the Cons section, the likelihood of concealing one's job title when supplying a negative review, and the average overall star rating when relative news coverage regarding NDAs across states over time is incorporated. Regressions include industry-state, industry-year-month, and state-year-month fixed effects. Standard errors are two-way clustered by industry and state. Significance levels: * 10%, ** 5%, *** 1%.

Table F20: Narrowing NDAs and Review Outcomes, Incorporating COVID-19 Outcomes

	Log length cons section		Conceal job title if longer cons than pros		Overall rating	
	(1)	(2)	(3)	(4)	(5)	(6)
Narrowed NDAs x NDA intensity	0.074***	0.088***	-0.058***	-0.049***	-0.270***	-0.267***
	(0.018)	(0.013)	(0.009)	(0.010)	(0.079)	(0.087)
Log(COVID-19 cases per capita) x NDA intensity		-0.029		0.009		-0.001
		(0.026)		(0.007)		(0.024)
Log(COVID-19 deaths per capita) x NDA intensity		0.015		-0.026**		-0.002
		(0.020)		(0.010)		(0.035)

Notes: The table implements the benchmark triple-differences model estimating the causal effect narrowing NDAs had on the average log length of the 'cons' section of reviews, the average overall star rating, and the likelihood of concealing one's job title when supplying a negative review when (the logarithm of) COVID-19 cases and deaths per capita over time are contemporaneously incorporated. COVID-19 cases and deaths are reported at the state-month level. State populations reflect 2020 levels and are obtained from the Federal Reserve Economic Database (FRED). Regressions include firm-state, industry-year-month, and state-year-month fixed effects. Standard errors are two-way clustered by industry and state. Significance levels: * 10%, ** 5%, *** 1%.

Table F21: Narrowing NDAs and EEOC Filings

	EEOC complaints		
	Total	Females	Males
	(1)	(2)	(3)
Narrowed NDAs	0.284**	0.189	0.307***
	(0.127)	(0.118)	(0.102)
Dependent variable mean	194.58	139.79	33.09
Observations	1.07e+09	4.93e+08	5.66e+08
Pseudo-R^2	0.95	0.93	0.81

Notes: Notes: The dependent variables are the number of EEOC charges filed alleging sexual harassment. The EEOC data are at the state-year level and include fixed effects for state and year. Fully saturated poisson fixed effects models are estimated. State-year-gender cells in each regression are weighted by total employment. The sample period is 2015–2021. Standard errors are clustered by state. Significance levels: * 10%, ** 5%, *** 1%.

Table F22: Heterogeneity in Narrowing NDAs and Employees' Overall Ratings of Firms

	Worker characteristic						Firm characteristic					
	Current	Former	Short tenure	Long tenure	Male	Female	Operates one state	Operates many states	Small	Large	Low rated	High rated
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Narrowed NDAs x NDA intensity	-0.255***	-0.194***	-0.272***	-0.201***	-0.289**	-0.229*	-0.409**	-0.254***	-0.285**	-0.205**	-0.232***	-0.301**
	(0.073)	(0.037)	(0.060)	(0.061)	(0.103)	(0.119)	(0.160)	(0.068)	(0.122)	(0.076)	(0.030)	(0.137)
Dependent variable mean	3.87	3.02	3.47	3.46	3.45	3.33	3.60	3.46	3.54	3.42	3.18	3.75
Observations	1919640	1616427	1705661	1084757	914036	773746	394951	3250258	1716326	1904742	1478239	1539896
Adjusted R^2	0.19	0.16	0.17	0.17	0.16	0.15	0.20	0.14	0.18	0.11	0.11	0.13
P-value for test of difference		0.512		0.047		0.415		0.167		0.583		0.559

Notes: The table investigates whether the effect on average overall ratings following the passage of these three laws under the triple-differences specification differs among various subsets of reviews, partitioned according to either worker or firm characteristics. Short (long) tenure employees refer to workers with at most (least) two years of work experience at the firm. The number of states the firm operates in is determined by calculating the number of unique states from which there is an employee review in the Glassdoor data. Volunteers are not asked to reveal their gender when submitting an employer review, but for a subset of respondents, gender is obtained through other aspects of the platform, such as a user profile. Low (high) rated firms reflect employers for which their average overall ratings in 2018 are above (below) average. Regressions include firm-state, industry-year-month, and state-year-month fixed effects. Standard errors are two-way clustered by industry and state. Significance levels: * 10%, ** 5%, *** 1%.

G Decomposing Triple Differences

Table G1: Narrowing NDAs and Cons Share of Review Text

	Full US Within High NDA Ind.		Within CA-IL-NJ High vs. Low NDA Ind.		Triple Difference		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
CA-IL-NJ	-0.013**						
	(0.005)						
Narrowed NDAs	0.003	0.005**			0.003**		
	(0.003)	(0.002)			(0.001)		
Narrowed NDAs x NDA intensity			0.038**	0.037**	0.015***	0.016***	0.016**
			(0.016)	(0.015)	(0.001)	(0.004)	(0.008)
Dependent variable mean	0.509	0.509	0.514	0.514	0.521	0.521	0.523
Observations	1862470	1862470	816693	816693	3879305	3879305	3645332
Adjusted R^2	0.01	0.01	0.12	0.12	0.02	0.02	0.11
Year-Month FE	✓	✓	✓				
Industry FE	✓	✓					
State FE		✓					
Industry-Year-Month FE					✓	✓	✓
State-Industry FE					✓	✓	
State-Year-Month FE				✓		✓	✓
Firm-State FE			✓	✓			✓

Notes: The table implements difference–in–differences and triple-difference models for estimating the causal effect narrowing NDAs had on the share of review text written within the cons section of each employer review. High- and low-NDA usage refers to industries for whom the share of workers covered by NDAs is above- or below-average, respectively. The first two columns are clustered by state, the next two by industry, and the final three two-way clustered by industry and state. Significance levels: * 10%, ** 5%, *** 1%.

Table G2: Narrowing NDAs and Log Length of Cons Section

	Full US Within High NDA Ind.		Within CA-IL-NJ High vs. Low NDA Ind.		Triple Difference		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
CA-IL-NJ	-0.001						
	(0.023)						
Narrowed NDAs	0.006	0.009			-0.000		
	(0.008)	(0.006)			(0.003)		
Narrowed NDAs x NDA intensity			0.180***	0.181***	0.106***	0.100***	0.074***
			(0.042)	(0.043)	(0.009)	(0.021)	(0.018)
Dependent variable mean	4.594	4.594	4.578	4.578	4.537	4.537	4.552
Observations	1862470	1862470	816693	816693	3879305	3879305	3645332
Adjusted R^2	0.09	0.08	0.15	0.15	0.09	0.09	0.15
Year-Month FE	✓	✓	✓				
Industry FE	✓	✓					
State FE		✓					
Industry-Year-Month FE					✓	✓	✓
State-Industry FE					✓	✓	
State-Year-Month FE				✓		✓	✓
Firm-State FE			✓	✓			✓

Notes: The table implements difference–in–differences and triple-difference models for estimating the causal effect narrowing NDAs had on the length of the cons section of employer reviews. High- and low-NDA usage refers to industries for whom the share of workers covered by NDAs is above- or below-average, respectively. Regressions in first two columns are clustered by state, next two columns clustered by industry, and the final three columns two-way clustered by industry and state. Significance levels: * 10%, ** 5%, *** 1%.

Table G3: Narrowing NDAs and Job Title Concealment Among Negative Reviews

	Full US Within High NDA Ind.		Within CA-IL-NJ High vs. Low NDA Ind.		Triple Difference		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
CA-IL-NJ	0.027***						
	(0.008)						
Narrowed NDAs	-0.018***	-0.019***			-0.016***		
	(0.005)	(0.005)			(0.002)		
Narrowed NDAs x NDA intensity			-0.271***	-0.271***	-0.030***	-0.047***	-0.058***
			(0.057)	(0.056)	(0.001)	(0.010)	(0.009)
Dependent variable mean	0.165	0.165	0.164	0.164	0.144	0.144	0.145
Observations	941093	941093	411187	411187	2067139	2067139	1898057
Adjusted R^2	0.07	0.07	0.14	0.14	0.07	0.07	0.12
Year-Month FE	✓	✓	✓				
Industry FE	✓	✓					
State FE		✓					
Industry-Year-Month FE					✓	✓	✓
State-Year-Month FE				✓		✓	✓
Firm-State FE			✓	✓			✓

Notes: The table implements difference–in–differences and triple-difference models for estimating the causal effect narrowing NDAs had on the rate at which employees conceal their job title when leaving a review for which the cons section was longer than the pros section. High- and low-NDA usage refers to industries for whom the share of workers covered by NDAs is above- or below-average, respectively. The first two columns are clustered by state, the next two by industry, and the final three two-way clustered by industry and state. Significance levels: * 10%, ** 5%, *** 1%.

Table G4: Narrowing NDAs and Overall Ratings of Firms

	Full US Within High NDA Ind.		Within CA-IL-NJ High vs. Low NDA Ind.		Triple Difference		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
CA-IL-NJ	0.076*						
	(0.041)						
Narrowed NDAs	-0.034*	-0.045***			-0.014		
	(0.017)	(0.013)			(0.009)		
Narrowed NDAs x NDA intensity			-0.232	-0.224*	-0.315***	-0.303***	-0.270***
			(0.132)	(0.122)	(0.003)	(0.052)	(0.079)
Dependent variable mean	3.584	3.584	3.526	3.526	3.489	3.489	3.479
Observations	1862470	1862470	816693	816693	3879305	3879305	3645332
Adjusted R^2	0.03	0.02	0.15	0.15	0.02	0.03	0.15
Year-Month FE	✓	✓	✓				
Industry FE	✓	✓					
State FE		✓					
Industry-Year-Month FE					✓	✓	✓
State-Industry FE					✓	✓	
State-Year-Month FE				✓		✓	✓
Firm-State FE			✓	✓			✓

Notes: The table implements difference–in–differences and triple-differences models estimating the causal effect narrowing NDAs had on the average overall rating of new employer reviews. High- and low-NDA use refers to industries for which the share of workers covered by NDAs is above- or below-average, respectively. The first two columns are clustered by state, the next two by industry, and the final three two-way clustered by industry and state. Significance levels: * 10%, ** 5%, *** 1%.

Table G5: Narrowing NDAs and Interquartile Range of Within-Firm Ratings

	Full US Within High NDA Ind.			Within CA-IL-NJ High vs. Low NDA Ind.		Triple Difference		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
CA-IL-NJ	0.021							
	(0.025)							
Narrowed NDAs	0.003	0.010	0.024			-0.005		
	(0.029)	(0.026)	(0.015)			(0.013)		
Narrowed NDAs x NDA intensity				0.063	-0.448*	0.309**	0.327**	0.191**
				(0.287)	(0.215)	(0.156)	(0.156)	(0.084)
Dependent variable mean	1.560	1.560	1.539	1.594	1.576	1.579	1.581	1.576
Observations	1521	1521	2575	468	565	4836	4810	8759
Adjusted R^2	0.04	0.13	0.16	0.05	0.12	0.19	0.21	0.21
Half-Year FE	✓	✓	✓	✓	✓			
State FE		✓	✓	✓	✓			
Industry-Half-Year FE						✓	✓	✓
State-Industry FE						✓	✓	✓
State-Half-Year FE							✓	✓
State-Industry 50+ firms on average	✓	✓		✓		✓	✓	
Weighted by average firm count			✓		✓			✓

Notes: The table implements difference–in–differences and triple-difference models for estimating the causal effect narrowing NDAs had on the mean of the interquartile ranges of firms' ratings within an industry-state-half year. High- and low-NDA usage refers to industries for whom the share of workers covered by NDAs is above- or below-average, respectively. The first three columns are clustered by state, the next two by industry, and the final three two-way clustered by industry and state. Significance levels: * 10%, ** 5%, *** 1%.

Table G6: Narrowing NDAs and Interquartile Range of Across-Firm Ratings

	Full US Within High NDA Ind.			Within CA-IL-NJ High vs. Low NDA Ind.		Triple Difference		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
CA-IL-NJ	-0.098***							
	(0.034)							
Narrowed NDAs	0.180***	0.192***	0.092***			0.081**		
	(0.046)	(0.051)	(0.024)			(0.033)		
Narrowed NDAs x NDA intensity				0.905*	0.853	0.998***	0.971***	0.463**
				(0.465)	(0.502)	(0.352)	(0.373)	(0.188)
Dependent variable mean	2.137	2.137	2.194	2.081	2.082	2.045	2.043	2.081
Observations	1521	1521	2648	468	581	4836	4810	9401
Adjusted R^2	0.18	0.25	0.24	0.18	0.19	0.47	0.47	0.53
Half-Year FE	✓	✓	✓	✓	✓			
State FE		✓	✓	✓	✓			
Industry-Half-Year FE						✓	✓	✓
State-Industry FE						✓	✓	✓
State-Half-Year FE							✓	✓
State-Industry 50+ firms on average	✓	✓		✓		✓	✓	
Weighted by average firm count			✓		✓			✓

Notes: The table implements difference–in–differences and triple-difference models for estimating the causal effect narrowing NDAs had on the interquartile range of the mean firm rating within an industry-state-half year. High- and low-NDA usage refers to industries for whom the share of workers covered by NDAs is above- or below-average, respectively. The first three columns are clustered by state, the next two by industry, and the final three two-way clustered by industry and state. Significance levels: * 10%, ** 5%, *** 1%.