

Author Accepted Version

Please cite as: Burt, Callie H. Forthcoming. The Not-So-Curious Case of the Bathroom Safety Study: Ideological Bias and the Erosion of Scholarly Scrutiny. *Theory & Society*.

The Not-So-Curious Case of the Bathroom Safety Study: Ideological Bias and the Erosion of Scholarly Scrutiny

Abstract

Scholars have long warned that ideological homogeneity in social science can compromise the error-correcting mechanisms of scientific inquiry, but how this process operates in practice, and how fragile claims become stabilized as settled knowledge, remains insufficiently understood. In this article, we examine how field-level ideological commitments embed contestable assumptions upstream of empirical testing, shaping what can be asked, measured, and concluded. We conduct an Ideological Epistemic Analysis (IEA) of a widely cited and highly publicized study treated as definitive evidence that gender-identity-prioritizing public-accommodations policies pose no public-safety risk. We show that a priori ideological commitments, operating through category reconstitution—especially the displacement of sex as an analytically relevant category—structure the study’s framing, constrain which mechanisms can be examined, and shape interpretation before data are examined. We further document how a methodologically flawed design produced findings that were nonetheless converted into settled science through canonization bias, a process by which ideologically congenial but fragile evidence is elevated to authoritative status through citation, media uptake, and policy discourse. Rather than reflecting a curious anomaly or isolated failure, the case illuminates a broader epistemic pattern in which ideological homogeneity weakens organized skepticism and facilitates the canonization of fragile claims. We also suggest avenues for mitigating ideological epistemic distortions and promoting more robust social science.

Keywords: ideological homogeneity; ideological bias; gender identity; transgender research; metascience; canonization bias; knowledge production

THE NOT-SO-CURIOUS CASE OF THE BATHROOM SAFETY STUDY: IDEOLOGICAL BIAS AND THE EROSION OF SCHOLARLY SCRUTINY

Concerns have mounted that ideological homogeneity has undermined the rigor and reliability of social scientific knowledge (e.g., Duarte et al., 2015; Honeycutt & Jussim, 2020; Martin, 2016; Rubin, 2025; Savolainen, 2024; Wright & DeLisi, 2015). Recent years have seen a growing body of scholarship warning that when scholars largely share the same ideological commitments, this uniformity can foster ideological epistemic distortions, as systematic biases in knowledge that favor these shared beliefs and values (Burt, forthcoming; Clark & Winegard, 2020; Duarte et al., 2015; Honeycutt & Jussim, 2020; Iceland & Silver, 2024; Martin, 2016).

In some fields characterized by extreme viewpoint homogeneity, ideological framing legitimizes one worldview while rendering dissent epistemically suspect, even morally illegitimate (e.g., recoding disagreement as ignorance or bigotry). In these situations, the mechanisms of scholarly scrutiny, such as peer review, become less effective at filtering bias (e.g., Duarte et al., 2015; Rauch, 2021; Savolainen, 2024), particularly when findings are ideologically congenial, such that findings that support the prevailing view are subject to relatively lax scrutiny (Clark & Winegard, 2020; Duarte et al., 2015; Savolainen, 2024). As replication and adversarial analyses become less common and dissenting perspectives fade from view, results that reinforce the dominant ideology may be accepted with increasingly minimal scrutiny (Duarte et al., 2015). At that point, science almost loses its “scientificity,” that is, its grounding in critical interrogation and evidence-based adjudication and instead can become advocacy conducted under the banner of science.

Recognizing the threats posed by ideological homogeneity to the integrity and credibility of science, scholars in several social science disciplines have documented its prevalence (Langbert, 2018; Magness & Waugh, 2022), the risk of consequent epistemic distortions (Al Gharbi, 2018; Iceland & Silver, 2024; Martin, 2016; Wright & DeLisi, 2015), and the mechanisms by which shared ideological commitments can shape, and sometimes warp, scientific inquiry (Duarte et al., 2015; Honeycutt & Jussim, 2020; Rubin, 2025). Building on this work, Burt (forthcoming) introduced ideological epistemic analysis (IEA) as an analytic framework for diagnosing whether, where, and how ideological commitments produce epistemic distortions. This initial work was offered both as a general contribution, namely, articulating the IEA framework, and a substantive one, specifically examining how shared commitments to gender-identity ideology may bias transgender research. This application revealed how ideological commitments could shape transgender research at multiple stages of knowledge production, including framing, measurement, design, and interpretation, producing skewed findings congenial to gender-identity ideology (in this case, Democrats’ “much more favorable attitudes toward and support for trans inclusive policies” being due to their “superior transgender knowledge compared to Republicans”; Zell & Burnett, 2024, p.1), but not necessarily consistent with reality.

Here, we extend the initial IEA application in content and substance, focusing on a single, widely publicized and highly influential study that has been canonized as

evidence that gender-identity–prioritizing¹ policies (**GIPPs**) for public accommodations pose “no safety risk”. Specifically, we conduct an IEA case study of Hasenbush, Flores, and Herman’s (2019) article, “Gender Identity Nondiscrimination Laws in Public Accommodations,” which asserts “that fears of increased safety and privacy violations as a result of nondiscrimination laws are not empirically grounded” (abstract). Whereas Burt’s initial IEA case study examined ideologically inflected but otherwise competent research, the current focus is on an ideologically inflected study whose methodological flaws are so severe that its findings should not be treated as credible evidence in scientific or policy discussions. Despite this, as of this writing, the article has been cited over 173 times (Google Scholar, accessed December 2025), and attracted extraordinary attention in scholarly discourse, legal advocacy, and mainstream media as providing definitive empirical evidence that GIPPs do not undermine safety.² Illustrating its influence, Altmetric, which tracks the online attention articles get from news outlets, social media, blogs, and the like, ranks the article in the 99th percentile (36th of the nearly 350k tracked articles of a similar age) across all journals.³

After examining the study and identifying its fatal flaws, as we do here, one might be tempted to describe its publication and influence as a curious case—a puzzling lapse in scholarly standards that happens and is corrected, perhaps in a short methodological response. However, we submit that the study’s publication and canonization is neither puzzling nor an isolated lapse, but is instead indicative of a field-level problem. The study’s publication and uncritical canonization as authoritative evidence despite severe methodological flaws are precisely what we should expect in a field characterized by ideological homogeneity and epistemic closure.⁴ Studies often get things wrong. What is unusual and revealing in this case (and requires explanation) is why this one got things so wrong in the specific ways it did: why peer review failed to catch errors that are, on close inspection, not subtle; why subsequent scholarship canonized rather than

¹ Following Burt (forthcoming), we refer to laws and institutional policies that allow individuals access to formerly or typically sex-separated spaces based on self-declared gender identity (regardless of sex) as gender-identity-prioritizing policies (GIPPs). These are, in effect, gender-identity-choice policies, as they retain the two (previously) sex-based facility categories, but prioritize gender identity over sex when the two are in conflict, shifting the criterion for access from sex to self-declaration. Under GIPPs, an assertion grants access: once an individual states that their gender identity corresponds to a given facility (e.g., the women’s locker room) there is no lawful ground for exclusion. We use the term *gender-identity-prioritizing* rather than simply gender-identity policies because such laws do not require that individuals use the facility matching a declared gender identity or create new facilities corresponding to the many gender identities people may claim, and not all people claim a gender identity at all. Thus, the laws give priority to gender identity when gender-identity-based access is claimed.

² For illustration, one NBC News article discussing the study and interviewing its authors, entitled “No link between trans-inclusive policies and bathroom safety,” described the study as “the first of its kind to rigorously test the relationship between nondiscrimination laws in public accommodations and reports of crime in public restrooms” (Moreau, 2018). As we will discuss, this characterization as “rigorous” is not warranted by the study.

³ Altmetric counted 338 news outlets referencing this article, 10 Wikipedia references, and nearly 3000 tweets. For the Altmetric data see: <https://archive.ph/VhdYV> (archived page as of October 2025).

⁴ To be clear, we use epistemic closure in the social-science sense, following Burt (forthcoming), as a condition in which a research community’s evidentiary and interpretive space becomes restricted such that challenges to core assumptions are foreclosed prior to empirical adjudication.

corrected it; and what these patterns reveal about the epistemic conditions of the field. Rather than being questions limited to one study's methodology, these are questions about how ideological homogeneity inflects the knowledge production process and impedes the error-correcting mechanisms of science. A standalone methodological critique, however thorough, cannot address these questions. An IEA is a diagnostic framework that can, by tracing the mechanisms and field-level conditions under which ideologically congenial but methodologically flawed work not only passes peer review but acquires the status of settled science. We use the HFH article as a case study to illuminate the field-level conditions that made its trajectory possible, conditions of considerably broader significance than any single study's methodology.

Thus, we add to Burt's initial IEA application, which traced more nuanced ideological inflections within otherwise competent work, to reveal a more acute form of distortion: how severely compromised studies can shape both scholarly consensus and public policy under the mantle of scientific authority. We also refine aspects of the IEA approach and clarify mechanisms through which shared ideological commitments can distort the knowledge base, highlighting the role of category reconstitution, defined as ideology-driven boundary-shifting of the conceptual categories, and canonization bias. In doing so, we elaborate how language choices, conceptual and interpretive slippage, and the stacking of weak evidence can transform modest or flawed findings into entrenched 'facts', which can be amplified in media coverage and political debate. In this way, tentative or weak findings can become established knowledge not by virtue of rigor, replication, and robustness, but through ideologically selective interpretation and amplification (e.g., Honeycutt & Jussim, 2020).

In addition, we briefly highlight institutional biasing effects prevalent in some ideologically inflected fields. Some research organizations function as what we call advocacy-oriented institutes, which are mission-driven research centers explicitly committed to advancing a sociopolitical goal. However, in fields characterized by ideological homogeneity and epistemic closure, such a policy-advocacy ecosystem creates epistemic conditions that make ideological epistemic distortions more likely to occur and less likely to be detected or corrected.

We focus on the field of transgender research. We do so not only because evidence suggests the field has been severely warped by ideological commitments, but also because scholarship on transgender issues is frequently cited in legislative debates, court cases, and medical decisions. As such, the stakes are especially high. This work affects people's lives, and the public is paying attention. When methodological flaws are overlooked because they yield ideologically congenial findings, the consequences extend beyond the integrity of the knowledge to public trust in science.

In short, this paper both refines the IEA model, by expanding the framework to further elucidate how ideological epistemic distortions are produced and sustained and expands our understanding of ideological biases in transgender research. In what follows, we review the IEA framework and the contours of gender-identity ideology; elaborate the analytical framework to focus on category reconstitution and canonization biases; conduct an in-depth analysis of a widely publicized and highly cited study to illustrate mechanisms of bias; and identify claims that, we submit, have been prematurely elevated to "facts," before concluding by encouraging more pluralistic, theoretical, and reliable research.

IDEOLOGICAL EPISTEMIC ANALYSIS

IEAs are a distinct subtype of scientific critique characterized by a focus on ideological biases in the knowledge-production process. Rather than focusing on theoretical or methodological shortcomings in isolation, or critiquing ostensibly flawed articles as such, IEAs examine how shared ideological priors, institutional incentives, and classificatory assumptions shape the research process, including what is asked, what counts as an effect, how results are interpreted, and which findings are amplified or canonized. The goal is not to spotlight particular scholars or impute motives, but to identify and illuminate field-level mechanisms through which shared ideological commitments produce epistemic distortions. As an analytic framework rather than a substantive theory, IEA does not compete with existing sociology of knowledge traditions, which explain why ideological distortion of knowledge is possible in general, but provides a structured diagnostic procedure for locating specific manifestations in particular fields.

Central to IEA is a non-pejorative conception of ideology as “shared descriptive, explanatory, and normative beliefs—i.e., beliefs about the way the world is, why, and how the world should be” (Burt, forthcoming). In this usage, central to ideologies are ideological pseudofacts, defined as normative, partial, uncertain or simply false beliefs, which are treated as established truths (i.e., justified true beliefs) by adherents. Pseudofacts are typically intertwined with sacred values, which are core normative commitments regarded as morally mandatory and insulated from empirical challenge (Baron & Spranca, 1997; Wright & DeLisi, 2015). For a fuller account of ideology's defining characteristics and their epistemic consequences, see Burt (forthcoming).

IEAs differ from conventional critiques by focusing on where and how shared ideological commitments shape the knowledge-production process. IEAs interrogate scholarship at multiple stages where ideological biases may induce epistemic distortions: research questions asked and unasked, framing of explanations and confounders; measurement of key constructs, interpretation, and canonization, whereby lax scrutiny combined with repetition transforms tentative or ideologically congenial findings into accepted knowledge (Honeycutt & Jussim, 2020; Savolainen, 2024). By illuminating these influences, IEAs clarify blind spots, expose signals of ideological distortions, indicate where theoretical, conceptual, and methodological refinement is needed, and function as heterodox perspectives in homogenous fields.

Here, we expand the IEA framework in several ways, including by elaborating the mechanisms of category reconstitution and canonization bias and foregrounding the role of advocacy-oriented institutional environments as structural conditions that systematically weaken error-correcting mechanisms. From an IEA perspective, such environments are not merely contextual backgrounds but can serve as part of the causal explanation for ideological epistemic distortions. Identifying these environments is therefore useful for diagnosing how ideological commitments may become embedded upstream of empirical testing and downstream in interpretation and canonization.

IEAs necessarily involve delineating the hegemonic ideology's orthodoxy and lexicon. A fuller account of gender-identity ideology is provided in Burt (forthcoming; for even more detailed treatment, see Ekman, 2023; Jones, 2023; Lawford-Smith, 2022; Stock,

2021; Sullivan & Todd, 2023). Here we summarize the key assumptions most directly relevant to the case study that follows.

GENDER IDENTITY IDEOLOGY

Rooted in postmodern traditions that emphasize the discursive construction of reality, gender-identity ideology (**GII**) is a distinct ideological framework that has come to dominate social science research on sex, gender, and transgender issues (see, e.g., Lawford-Smith, 2022; Stock, 2021; Sullivan & Todd, 2023). Under GII, gender identity replaces sex as the socially and institutionally relevant category, and sex is redescribed as “assigned at birth” and destabilized as a material biological category. GII fulfills the defining criteria of an ideology: it consists of interrelated beliefs shared within a group in opposition to competing beliefs, is sociopolitically action-oriented, and is anchored in ideological pseudofacts and sacred values that together constitute its orthodoxy (see Burt, forthcoming).

Ideological pseudofacts vary in how they function epistemically, and distinguishing their types clarifies the specific distortions they produce in downstream research. First, *premature certainty pseudofacts* are empirically unsettled propositions that are treated as established facts. Second, *normativity masquerading pseudofacts* are political-moral values framed as empirical claims. Third, *conceptual-as-empirical pseudofacts* are conceptual stipulations recast as empirical observations, either when theoretically constructed categories are misrepresented as natural kinds or when empirically grounded natural kinds are redescribed as socially constructed without evidentiary justification. In practice, pseudofacts can reflect a combination of subtypes.

The key ideological pseudofacts of GII exemplify these subtypes. *Sex constructionism*—that sex is social construction imposed on an underlying continuum rather than a biological classification—operates primarily as a conceptual-as-empirical pseudofact. Here, sex, a biological category, is reconceived as a social assignment or continuum, and this reconception is presented as if it were an empirical discovery rather than a normative redefinition.⁵ *Gender-identity essentialism*—that all humans have an innate, sex-

⁵ Consistent with our point, many scholars appear to understand the claim that “sex is non-binary” as an empirical discovery rooted in recent advances in biology and medicine rather than a conceptual redefinition (e.g., Ainsworth, 2023; Fuentes, 2025). If contemporary claims that sex is non-binary were grounded in new empirical discoveries, they would be expected to track the timing of those discoveries. They do not. The empirical phenomena most often invoked in support of a non-binary conception of sex (e.g., “intersex” conditions or differences of sexual development (DSDs), chromosomal mosaics, and phenotypic variation) are not new discoveries. Such conditions have been clinically recognized and medically described since at least the mid-twentieth century, without producing a corresponding reconceptualization of sex as non-binary. In biology, sex is defined by reproductive role, which is organized around the production of large or small gametes, and is therefore binary as a classificatory system, notwithstanding continuous variation in sex-linked traits or rare, atypical developmental pathways (e.g., Coyne & Maroja, 2023; Goymann et al., 2023; Janicke et al., 2016; Lehtonen et al., 2016; Marinov, 2020; Rehman, 2023; Stock, 2021; Wright, 2025). DSD conditions do not constitute additional sexes or a sex spectrum, but instead represent expected deviations within a binary reproductive framework, as biological systems routinely exhibit phenotypic variation without altering categorical structure. Biological categories are defined by organizing functions, not by the absence of developmental exceptions. Accordingly, contemporary claims that sex is non-binary are better understood as instances of the category reconstitution we describe, rather than as conclusions compelled by recent empirical discoveries in biology or medicine.

independent gender identity—is primarily a premature certainty pseudofact, which is presented as a settled universal human psychological property, despite the fact the evidence is nonexistent, indirect, or interpretively loaded.⁶ *Gender-identity precedence*—that one’s status as a man/boy or woman/girl (or neither) is determined by gender identity not sex—is a primarily a conceptual-as-empirical pseudofact, with a secondary normativity masquerading component, insofar as the normative position (“gender identity *should* determine status”) is framed as a descriptive truth. Finally, a core sacred value underwriting GII’s political aims—that replacing sex-based policies with gender-identity-based policies is ‘necessary for the well-being of transgender people’—begins as normativity-masquerading and is progressively transformed, through ideologically aligned scholarship including the case study examined here, into a premature-certainty pseudofact treated as an established empirical premise.

These orthodoxies and political aims are reinforced and naturalized through a distinctive lexicon, including terms such as “assigned sex at birth,” “trans-inclusive,” “trans-exclusionary,” and “cisgender,” which encode ideological assumptions and render alternatives suspect or unintelligible (Burt, forthcoming). Altogether, these elements shape the epistemic terrain of transgender research by channeling how questions are framed; narrowing the hypotheses deemed legitimate; and recasting critical scrutiny as prejudiced and ignorant. In this way, GII functions not merely as a set of contested claims but as a normative regime that governs scholarly inquiry and evaluation (Burt, forthcoming; Sullivan & Todd, 2023). As the case study later shows, these ideological distortions are compounded by a broader absence of scientific theory, a point we return to in the conclusion.

Category Reconstitution

A central biasing mechanism operating through these linguistic choices is what we call *category reconstitution*, defined as the reshaping of conceptual boundaries so that categories conform to ideological commitments.⁷ Category reconstitution is an ontological recategorization; that is, a change in what the category refers to or what the phenomenon categorized is taken to be in reality. The epistemic distortion arises because the downstream implementations inherit the ontological stipulation and thus occlude alternative framings. Category reconstitution is present not whenever scholars adopt contested terminology, of course, but when (a) a normative claim is embedded in classification, *and* (b) that classificatory choice renders a competing frame (such as sex-based accounts of privacy or safety), conceptually unintelligible or inexpressible.

In GII, the most significant example of category reconstitution is the substitution of gender identity for sex as the relevant analytic and policy category. This conceptual-as-empirical substitution is presented as an empirical correction rather than an

⁶ To be clear, our claim is not that ‘gender identity’ is itself a pseudofact, as it is a concept; the pseudofact is the (empirically uncertain) claim that gender identity is a universal, innate, sex-independent property of all humans.

⁷ Category reconstitution encompasses several related processes, including concept creep (i.e., the moral-driven expansion of harm-related categories to include milder or more ambiguous phenomena; see Haslam, 2016), as well as its inverse forms of contraction, substitution, or erasure, whereby categories that become ideologically inconvenient are displaced or eliminated. We focus on this latter case.

ideological/normative redefinition, effectively rendering sex analytically obsolete, which amounts in practice to sex erasure or denialism (Burt, 2023a; Jones, 2023). Once normalized, research can no longer register sex-based effects or conflicts, since the category of sex has been conceptually displaced by gender identity (or even erased as a relevant category). Category reconstitution is thus a process by which ideological commitments produce epistemic distortions in ways that often escape scrutiny, in part, because the reconstituted terms restrict what can be articulated or even conceptualized.⁸

For example, under GII, “woman” (or “man”) is reconstituted into a gender identity class. Consequently, sex-based policies for female spaces that exclude transwomen are recast as ‘discrimination against a subset of women’ rather than as the maintenance of female-only space (Burt, forthcoming). GII’s lexicon thus does not merely describe reality but enacts the reconstituted ontology it presupposes, transforming a contested ideological stance into an unquestionable empirical and moral truth. We can see these processes at work in the slogan “Trans women are women,” which operates as both a conceptual-as-empirical pseudofact (purporting to be a factual statement about the world, when, in fact, it is a contested conceptual/normative claim) and a sacred value, functioning as a moral-ideological axiom. To dispute this claim is often treated as intrinsically harmful or bigoted rather than a conceptual or normative disagreement (e.g., Jenkins, 2016, p. 396). The epistemic consequences include the foreclosure of sex-based framings and the redefinition of dissent as bigotry or ignorance. Once the ontological recategorization is adopted as a foundational assumption about reality, category reconstitution can unfold through the following four structured transformations:

1. *Ontological transformation*: “woman” and “man” are redefined as gender-identity classes; “sex” is redescribed as “assigned at birth,” reframing sex as an administratively conferred category rather than a material constraint and thereby diminishing its explanatory relevance.
2. *Analytic transformation*: “impact” and “harm” are operationalized with respect to gender identity categories, rendering sex-based effects analytically invisible.
3. *Policy transformation*: sex-separated spaces are redescribed as “exclusionary” or “discriminatory”. For example, female-only spaces are cast as “excluding

⁸ After writing and reflecting on this section, we realized (perhaps belatedly, and certainly ironically) that our critique parallels Foucault’s (1972) in important ways. We argue that GII enacts the kind of discursive regulatory regime he described. However, whereas Foucault adopted an anti-realist stance, contending that reality is discursively constructed, we take an ontologically realist position. We argue that GII can distort what is known and what can be said about reality (e.g., sex), but that reality itself exists independent of discourse.

As a quick backdrop, Foucault (1972) argued that power and knowledge are co-constitutive. That is, power produces the forms of knowledge through which it operates, and knowledge in turn sustains power. His aim was diagnostic rather than prescriptive, serving as a critique of how power naturalizes itself through regimes of truth; his was not a call to harness discourse as a means of transformation. Butler (1990), and subsequent gender theorists, transformed Foucault’s descriptive insight into a political instrument. By embracing Foucault’s premise that discourse constitutes reality, they effectively sought to redefine reality (sex) through discourse. Thus, what began as a critique of power’s productive capacity by Foucault thus became, in Butlerian practice, a project of epistemic and linguistic control, one that arguably instantiates a new discursive regime while operating under an emancipatory rationale.

transgender women,” because transwomen are presupposed as members of the ontological class “women.”

4. *Normative transformation*: disagreement over policy design, namely opposition to GII-consistent policy, is reframed as discrimination or bigotry, and skepticism is treated as epistemically and morally suspect.

In sum, category reconstitution converts contested ideological claims into foundational classificatory assumptions, foreclosing rival framings by definition rather than by empirical evaluation and generating epistemic distortions that persist downstream while remaining largely invisible.

Advocacy-Oriented Knowledge Environments

The risk of ideological epistemic distortions in the field of transgender research is further amplified by the institutional milieu. Much of this research occurs in or is shaped by what we call *advocacy-oriented knowledge environments*. These are contextual preconditions or epistemic ecologies in which ideological aims become constitutive of the research mission, shaping not only what is studied but how.

Much transgender research (and LGBTQ scholarship more broadly, especially that closely tied to policy debates) occurs in or is funded by such advocacy-driven institutes, think tanks, or philanthropic programs explicitly dedicated to advance LGBTQ equality. The most prominent example is the Williams Institute at UCLA Law School, an explicitly advocacy-oriented research center founded with the stated mission “to advance LGBTQ+ law and public policy through rigorous research.”⁹ The Institute’s mission statement replaces a scientific telos with a moral-political telos (advancing equality law). To be clear, this orientation *does not* invalidate the scholarship produced there, but it undoubtedly shapes epistemic incentives. Research agendas, interpretive frameworks, and dissemination practices tend to emphasize evidence consistent with the moral legitimacy and policy efficacy of putatively ‘pro-LGBTQ policies,’ including GIPPs. The Hasenbush et al. (2019) study examined here was conducted by scholars affiliated with the Williams Institute and reflects the incentives and interpretive frames typical of this environment.

This dynamic is not unique to transgender or LGBTQ research but has parallels in pharmaceutical research, where financial conflicts of interest require mandatory disclosure because structural incentives can bias study design, interpretation, and publication even when individual researchers act in good faith. Advocacy-oriented knowledge environments pose a similar epistemic concern. When the funders, institutional homes, and professional rewards are organized around advancing a moral or policy goal, ideological commitments function as a kind of conflict of interest, systematically privileging congenial evidence and discouraging skeptical interpretation.

With this account of GII and advocacy-oriented knowledge environments in place, we turn to examine how these orthodoxies and lexical choices have shaped the production and interpretation of empirical research, focusing on a single study that illustrates these dynamics in practice. We organize our critical analyses into sections corresponding to research framing and question, methods, interpretations, and canonization biases,

⁹ <https://williamsinstitute.law.ucla.edu/legacy-campaign/> (accessed December 2025).

noting that there is overlap in these discussions given their intertwinement in the research process.

**CASE STUDY: HASENBUSH, FLORES, AND HERMAN (2019)
GENDER IDENTITY NONDISCRIMINATION LAWS IN PUBLIC ACCOMMODATIONS**

In their 2019 study, Hasenbush, Flores, and Herman (HFH) aim to empirically evaluate a salient concern raised by non-supporters of GIPPs for public accommodations (e.g., restrooms, locker rooms, changing rooms, shelters): that gender self-identification policies undermine safety. HFH write that the primary argument opponents levy against these policies is that GIPPs create “a loophole for sexual predators to access women’s public restrooms and locker rooms, thus decreasing women’s and girls’ safety and privacy in such spaces” (p.70). Addressing this empirical gap is the aim of their study, which is described in their abstract and which we reproduce here:

Legislation, regulations, litigation, and ballot propositions affecting public restroom access for transgender people increased drastically in the last three years. Opponents of gender identity inclusive public accommodations nondiscrimination laws often cite fear of safety and privacy violations in public restrooms if such laws are passed, while proponents argue that such laws are needed to protect transgender people and concerns regarding safety and privacy violations are unfounded. No empirical evidence has been gathered to test such laws’ effects. This study presents findings from matched pairs analyses of localities in Massachusetts with and without gender identity inclusive public accommodation nondiscrimination ordinances. Data come from public record requests of criminal incident reports related to assault, sex crimes, and voyeurism in public restrooms, locker rooms, and dressing rooms to measure safety and privacy violations in these spaces. This study finds that the passage of such laws is not related to the number or frequency of criminal incidents in these spaces. Additionally, the study finds that reports of privacy and safety violations in public restrooms, locker rooms, and changing rooms are exceedingly rare. *This study provides evidence that fears of increased safety and privacy violations as a result of nondiscrimination laws are not empirically grounded* (p.70, emphasis added).

The article presents itself and has been touted as a neutral, even rigorous, evidence-based empirical analysis; however, it embeds GII commitments that profoundly shape the study’s framing, questions, and interpretations. Even more significantly, the study suffers from methodological flaws that are so severe that its findings should not be treated as credible evidence in scientific or policy discussions. We discuss each in turn.

Ideological Research Framing and Research Question Formulation

Consistent with the category-reconstitution process outlined above, and like most research in this field (e.g., Morgenroth et al., 2024a; Morgenroth et al., 2024b; Ward et al., 2025; Zell & Burnett, 2024; see Burt, forthcoming, for more examples), HFH’s article adopts GII lexicon, embedding contested ontological assumptions upstream of empirical analysis. For illustration, the article, including in the title, refers to GIPPs as “gender identity nondiscrimination laws.” To be sure, we recognize that Massachusetts officially designates these ordinances using this terminology, and that adopting a law’s legal name alone is certainly not evidence of ideological embedding. The problem is that the article does more than reproduce a statutory label. The framing repeatedly presupposes or relies on the discrimination framing, characterizing sex-based policies as “proscribing transgender inclusion,” treating opposition as affectively driven and empirically weak, and presenting the protective function of GIPPs as a background premise rather than a hypothesis requiring evidence. This substantive endorsement of

the ontological claim encoded in the label—that sex-based policies are discriminatory against transgender people—is misguided and a salient manifestation of category reconstitution.¹⁰

Adopting a working definition of discrimination as injurious differential treatment on the basis of membership in a category, we can see that sex-based policies do not discriminate against transgender people because they do not differentiate on the basis of transgender status. The operative classificatory criterion is sex, and that applies uniformly; gender identity or transgender status is irrelevant. The discrimination framing becomes coherent only once sex is displaced and gender-identity precedence is endorsed (i.e., gender identity is treated as the relevant basis of category membership, such that transwomen are treated as a subset of women for purposes of access). At that point, exclusion from female spaces is recast as discrimination against a subset of women rather than as the maintenance of female-only space. That is the ontological transformation at issue, and HFH’s article’s framing relies on it.

Similarly, as is also common, HFH’s article describe GIPPs as “inclusive” policies and contrast them with policies said to exclude transgender people. The effect of this framing is a kind of linguistic sleight of hand: it treats access based on gender identity as inclusion and sex-based access rules as exclusion, even though transgender people are not excluded from sex-based public accommodations as such; rather, like everyone else, they are classified on the basis of sex (for more detail on this point, see Burt, forthcoming).

Another minor, but noteworthy, example of ideological language is the article’s use of the term “sex-segregated” in several places over the more neutral “sex-separated”. Segregated carries obvious ideological-normative connotations. The term “segregation” is a heavily loaded term in the U.S. context, overwhelmingly associated with exclusion, injustice, and historical racial oppression. Using the term “sex-segregated spaces” implicitly analogizes designated single-sex spaces to morally condemned forms of racial exclusion, suggesting that such arrangements are inherently problematic or discriminatory.

Notably, HFH’s article is framed as an empirical response to what the authors call a “drastic” increase in legislation requiring “sex-segregated restrooms and changing facilities to be limited to use based on the sex on a person’s birth certificate” (p.70), with a particular focus on North Carolina’s much-maligned “bathroom bill” (i.e., Public Facilities Privacy & Security Act, or H.B. 2). HFH’s description of H.B.2 epitomizes the conceptual and linguistic distortions introduced by GII’s reconstituted categories:

¹⁰ Differential treatment alone is not discrimination on this definition; otherwise, no program or policy-based categorization could exist without being discriminatory, and the term would lose its analytic utility. We understand differential treatment as discriminatory when (a) the differentiated dimension (in this case, transgender status) is the one that the policy targets, and (b) the policy is unfavorable or injurious in its treatment of that category. What (b) implies can be discussed at much greater length, but in this case (a) does sufficient work. Under sex-based policies, transgender people are not excluded as transgender people; rather, like everyone else, they are classified by sex. A non-trans female who wishes to use the men’s restroom because it is empty and the women’s has a long line is also not allowed to use the other-sex facility because of her sex. To be clear, such policies may be experienced as exclusionary and may be argued to be unjust. But under this definition, that does not make them discriminatory. Conversely, the fact that a policy is not discriminatory in this sense does not make it desirable, justifiable, or appropriate.

While this law would have no legal impact on restroom use among cisgender (i.e., nontransgender) people, it meant that transgender people who had transitioned from their sex assigned at birth to a different gender would be required to use the restroom of their sex assigned at birth unless they had legally changed their birth certificate. The new law also opened up the possibility of increased harassment and policing, both social and actual, of gender nonconforming people in public restrooms, whether they were transgender or not (p.70).

This paragraph obscures the operative legal change in several ways. First the claim that law had “no legal impact on restroom use among cisgender people” reflects the category reconstitution that erases sex as the operative legal category. Deploying GII terminology, “sex assigned at birth,” implies that sex is not a biological reality but instead a social construction created from an arbitrary social assignment. This lexical choice serves to accept and encode gender identity as the proper classificatory axis, *thereby obscuring what actually changes*: the eligibility criteria for access (from sex to gender identity). As such, HFH delimit what counts as “impact”. By treating sex-based provisions as impacting only transgender users, HFH ignore composition effects. Shifting eligibility from sex to gender identity alters who qualifies and thus changes the provision *for everyone*. This category reconstitution facilitates what we refer to as the “non-impact” impact logic underpinning much transgender research and GIPP advocacy, rooted in the neglect of sex. What this framing conceals is the actual “impact” of GIPPs: the elimination of sex-separated provisions, especially female-only spaces.

To justify their framing (and the non-impact argument), HFH, like others in this genre, draw extensively from literature that obscures the distinctions between the rights claims of LGB individuals and those of transgender individuals. However, these situations are not analogous (Burt, 2023b). The [dis]analogy often made between GIPPs and same-sex marriage illustrates the flaw in this reasoning. Expanding marriage eligibility to same-sex couples did not change what marriage is for existing participants. Whatever one thinks about same-sex marriage, it is a genuine case of a “non-impact impact”. GIPPs are different. Changing the eligibility criterion from sex to gender identity materially alters the provision itself by changing the composition of the group, and, therefore, the conditions the category was designed to secure (i.e., a single-sex space). In short, the shift from sex-based to identity-based classification, anchored in the gender-identity-precedence pseudofact with a side of sex constructionism,¹¹ defines sex-based composition effects out of “impact.”

Returning back to the above quote, one can see in the second sentence that the authors shift from ideological description to speculative moral narrative, continuing the pattern of ideological framing. The “possibility” of increased harassment is not a new condition introduced by H.B. 2 (or GIPPs) but a constant feature of social life. Furthermore, the

¹¹ We note that our critique does not depend on establishing that HFH personally endorse the metaphysical claim that sex is socially constructed or not materially real. As discussed, ideological commitments operate through shared classificatory assumptions, lexical conventions, and institutional incentives that shape research regardless of individual scholar’s explicit beliefs or intentions. Whether HFH consciously embrace sex constructionism is not the relevant question. What matters analytically is that their framework’s operational assumptions (i.e., treating sex as an administratively conferred label, defining impact exclusively in terms of gender identity categories, and rendering sex-based composition effects invisible) produce the same epistemic distortions that explicit endorsement of GII would produce.

sentence illustrates the asymmetry of skepticism typical of ideologically saturated research. HFH's study purports to empirically test whether GIPPs increase safety and privacy violations. That is, the study ostensibly examines whether safety concerns about male access to female spaces are empirically justified, with the implication (and conclusion) that in the absence of confirmatory evidence from aggregate crime rates, such fears or claims of safety concerns are unwarranted. Yet they assert *without evidence* that H.B. 2 "opened up the possibility of increased harassment and policing ... of gender nonconforming people in public restrooms". In other words, the (inconsistent with GII) harm they set out to test is treated with skepticism, while the harm they do not test is asserted as real.

We can see this same dynamic in their pro-GIPP framing and their research approach. HFH write: "The primary argument levied against the passage of public accommodations nondiscrimination policies *that protect transgender people* is that the policy creates a loophole for [male] sexual predators to access women's public restrooms and locker rooms, thus decreasing women's and girls' safety and privacy in such spaces" (emphasis added p.70). In this early statement, the authors assert that GIPPs protect transgender people, without presenting any evidence that they do. This assumption of untested efficacy (possibly a premature certainty pseudofact) mirrors the broader pattern of one-sided evidentiary scrutiny in ideologically inflected scholarship, where only dissenting claims are required to clear an evidentiary bar. Hypothetical harms and benefits consistent with GII's narrative are asserted as real without robust scientific evidence. This selective application of evidentiary standards, as the suspension of skepticism for ideologically favored claims, is a key mechanism influencing epistemic distortions. Perhaps obviously, the bracketed male insertion is ours. Notably, this quote omits the sex of the predators, referring only to "sexual predators." This omission is noteworthy. The entire mechanism motivating safety concerns is sex-based (i.e., males accessing female spaces under GIPPs), and leaving the sex of the predator unstated naturalizes the erasure of sex as the operative analytical category. That HFH's article can describe the central safety concern without naming sex illustrates how category reconstitution operates. Once sex has been rendered analytically invisible, even the mechanism the study purports to test is obscured.

As noted, HFH's study is framed around H.B. 2 and the surge of 'exclusionary', 'anti-trans' legislation, rooted in political opportunism,¹² e.g., "forms of tactical innovation the Religious right have taken" (p.72). Whatever one thinks about it, this sex-based legislative activity is clearly a response to a preceding, rapid and ongoing wave of policy changes that prioritize gender identity over sex in regulating access to (previously or currently) sex-separated spaces. By omitting this initiating policy shift, the article depicts a unilateral upsurge of 'anti-trans' sentiment. This asymmetric framing¹³ facilitates a

¹² This is, of course, not a denial that political opportunism has played a role in this debate (of course it has). This is instead a critique of one-sided framing and characterization of opposition (or even questioning) of GIPPs as reflecting "anti-LGBT" sentiment and misinformation, rather than a more complicated policy debate involving trade-offs and significant changes.

¹³ The broader political dynamic is, of course, not unique to this issue. We have seen such reactive, conservative legislation in recent decades, for example, in laws restricting marriage to opposite-sex couples. Laws, such as the Defense of Marriage Act, were enacted in explicit response to court rulings and local policies actively shifting away from the status quo and toward same-sex marriage rights, backed by a strong LGB movement. Our point is not to equate these policy issues or movements—indeed, we disanalogize them here—but to note the same sequencing: a culturally

neglect of the fact bills such as H.B. 2 were designed to codify the longstanding implicit or explicit sex-based norms for public accommodations grounded in building codes, administrative regulations, and common-law assumptions about privacy and safety, which were operative in every U.S. jurisdiction not that long ago. HFH's framing of H.B. 2 as a legal innovation "limiting transgender people's restroom access" neglects the fact that sex-based access was previously the standing arrangement in nearly all localities. In North Carolina, the initial "change" came with the city of Charlotte's change to gender-identity-based access, not the reverse.

The article's literature review reflects the same ideological tilt. GIPP opponents' concerns are characterized as affective or irrational involving "gender panic" or "disgust"; LGB (non-impact) precedents are uncritically generalized to ("impact") transgender policies; and principled feminist/legal critiques of self-ID in sex-separated spaces are noticeably absent. The net effect is an ideological-normative framing: that GIPPs are morally just, even necessary "antidiscrimination" laws. As an example, HFH assert: "The arguments used to justify anti-LGBT policies tend to be emotionally stirring, though often lack empirical validity," (p.72), implying that the arguments used to justify GIPPs are different—i.e., reasoned, logical, and empirically warranted. However, this distinction is asserted rather than shown. No analysis is provided to establish that pro-GIPP arguments are empirically grounded in a way their critics' arguments are not.

Against this GII backdrop, HFH introduce their research question as follows:

The asserted motivations for *proscribing transgender inclusion in public accommodations* are the perceived negative externalities of increased harassment and victimization in public spaces such as bathrooms and locker rooms. The current project examines the policy motivation portion. In essence we ask: 'Are inclusive transgender public accommodations laws associated with these negative externalities?' If not, then public policy may be seeking a policy solution to solve a problem that does not exist or that would exist even if policies have changed (pp.72-73; emphasis added).¹⁴

One would think, given their framing, that HFH's study actually examines whether GIPPs create "a loophole for sexual predators to access women's public restrooms and locker rooms." That is, the study assesses whether male predators exploited gender identity-based access provisions to prey on females. But it does not. Instead, the study examine changes in aggregate crime rates, "inclusive of criminal incidents in public restrooms, public locker rooms, and public changing rooms"; data incapable of detecting the underlying concern.

significant policy change backed by a strong social movement is proposed or adopted, followed by counter-mobilization backed by a competing political movement to preserve the status quo. In the case of GIPPs, legislative efforts that HFH characterize as "anti-trans" emerged *after, and in reaction to*, the adoption or proposal of GIPPs.

¹⁴ Notably, the second part of this question "or that would exist even if policies have changed" is incoherent given their framing. The "primary argument levied against" GIPPs, as they themselves note, is that such policies "create[] a loophole for [male] sexual predators to access women's public restrooms and locker rooms" (p.70). The stated mechanism of risk is that GIPPs permit entry on the basis of self-declaration, thereby creating conditions for male access and potential predation. Given this, the phenomenon cannot "exist even if policies have changed," because the claim *is about policy-enabled access*.

This alone is not invalidating but combined with the assumed and asserted function of GIPPs, it is surely close to it. Specifically, the study's research question and associated hypotheses (and later interpretations) almost completely ignore the motivation for and asserted function of GIPPs: protecting transgender people. Although briefly noted in the discussion¹⁵ of their "hypotheses," (a list of possible outcomes: no change, increase, decrease in crime rates), the prevailing assumption that GIPPs protect and increase the safety of transgender people is largely ignored. According to GII and HFH's logic, GIPPs should *decrease* crime rates by reducing crimes against transgender people. That is, if GIPPs do not undermine the safety of females and do increase the safety of transgender people, the net result is a *decrease* in crime rates.

However, this obvious inferential implication the guiding logic is almost entirely neglected. This allows the authors to depict a null finding (no change in crime) and the presumed "lack of negative externalities," as support for GIPPs, even in the absence of evidence of the alleged benefit. What is more, using their logic—that GIPPs reduce crimes against transgender people—a null finding could be consistent with increased crimes against females in formerly female-only spaces. Inasmuch as GIPPs reduce crimes against transgender people, as their arguments imply, a finding of no change could mask, by offsetting, increased offenses by male predators against females in formerly female-only spaces.

Thus, we again see the selective application of empirical standards. A null result, where no detectable change occurs, is improperly posited as 'myth busting' of GIPP opponents, while not recognized as equally 'myth busting' of the protective value of GIPPs. These asymmetries are (or this incoherence is) masked by the ideological framing, where opposition to GIPPs is framed as discriminatory and sex has been effectively erased.

Indeed, the authors use the very language of myth busting in their framing. HFH write that their null hypothesis "is motivated both by previous scholarship dispelling myths in other areas of LGBT rights as well as the insufficiency of single policies to lead to vast changes in victimization rates." In effect, the study's framing preloads the evidentiary interpretation of the failure to reject the null hypothesis, reframing it as a confirmatory finding that supports the ideological claim that opposition to GIPPs is myth-driven. As discussed later in the interpretation section, failure to reject the null hypothesis (i.e., no difference or no change) does not constitute generalizable evidence in favor of no difference or no change. (As we will see, this problem is compounded by the design's near-certain inability to detect an effect at all.)

Finally, the ideological framing also obscures privacy concerns and how GIPPs change the legal contexts of some behaviors. Although "privacy" appears in the study's title and is presumably central to the intended contribution, the outcome measure—police records of criminal incidents—does not incorporate privacy issues that do not rise to

¹⁵ Immediately prior to presenting their hypotheses, the authors mention that "counter to arguments against such laws, [GIPPs may] reduce victimization rates [against] sexual and gender minorities, which could make a discernible reduction in overall crime rates" (p.73). However, this idea that GIPPs might reduce crime is not simply an idea in passing, it is a central background assumption and motivation for GIPPs. The authors assert that GIPPs protect transgender people. Thus, on this view, GIPPs reduce crimes against transgender people.

criminality. Privacy is inherently subjective and context-dependent and concerns expectations of bodily exposure, the composition of the space, and who is present, not just whether a criminal act occurred. A male body present in a space where females undress, shower, or use the bathroom is reasonably viewed by some as a privacy violation, regardless of the criminal status of the behavior. Moreover, the vast majority of women who feel their privacy has been violated by an unwanted male presence in a changing room, shower, or bathroom in a GIPP locality will likely not call the police. Their outcome simply does not assess privacy in any rigorous way.¹⁶

No less important, the study's focus on "criminal incidents" across GIPP and non-GIPP localities overlooks the fact that GIPPs *alter the legal classification or treatment of certain conduct* such that privacy and safety violations may increase *without affecting crime rate changes*. Some behaviors considered crimes under sex-based policies could become either unenforceable or effectively redefined as noncriminal under GIPPs because male persons self-identifying into women's spaces are treated as if they are female. This blind spot is not incidental but follows directly from the erasure of sex as a legally operative category under GIPPs.

For example, under a sex-based policy, a male who exposes his genitals in a women's locker room, shower, or restroom could be charged with indecent exposure or public lewdness, even if he claimed no sexual intent. The crime is defined in part by the context (a male's nudity in a female-only space), and female discomfort or distress is recognized as a harm. Under GIPPs, however, if that male says he identifies as a woman, authorities may treat their nudity as legitimate for the space, removing the sex-based element from the offense. What was once a prosecutable act becomes reclassified as permissible, even if criminal motives (e.g., intentional exposure for personal sexual pleasure) and harm (e.g., women present feel violated or intimidated) remain.

Thus, a key shift, which is not adequately considered by HFH's study (or any work in this area, to our knowledge), is the change in the baseline legality of male presence in a women's space. Once male presence is deemed lawful by policy, behaviors that once relied upon "being male in a women's space" as an element of the offense or a marker of intent lose that legal foundation or presumptive criminality.

In sum, HFH's study is not merely affected by isolated terminological choices or background assumptions but instead is fundamentally structured by a GII framing. By ignoring sex as a legally and materially operative category, the framing delimits what counts as policy "impact," constrains which harms are recognizable, and facilitates the interpretation of null findings as confirmatory. This framing shapes the research question, the choice of outcome measures, and the standards of skepticism applied to competing claims. Against this backdrop, the study's methodological limitations are not incidental technical defects, but compound and amplify the distortions introduced at the framing stage.

Methodological Flaws: Opaque Reporting, Tiny Sample, and Near-Zero Events

Methodological errors are not uncommon in social research, and peer review is designed to detect and correct them before publication. When the process is working, flaws are caught in peer review before a study is accepted into the literature or shortly thereafter.

¹⁶ We are grateful to a reviewer for drawing our attention to the privacy issue.

In this case, however, those mechanisms appear to have failed. HFH's study was published and canonized with methodological flaws that are pervasive, fundamental, and incompatible with the standards of rigorous empirical analysis. The study's methods are neither transparent nor replicable, and the evidentiary claims far exceed what its data can support. Nevertheless, the article has been repeatedly cited in scholarship, media reporting, and policy debates as rigorous empirical evidence that GIPPs pose no safety risk. The following analysis details the scope and seriousness of these methodological failings, which together reveal not just weak research design but an erosion of scholarly scrutiny.

To be sure, we acknowledge that assessing the public-safety effects of GIPPs is empirically challenging, given low base rates, municipal units of analysis, the infeasibility of experimental designs, among other factors. However, the problem with the analysis is not just that it confronts a hard question with imperfect tools, but that its design lacks the minimum features required for informative inference while its non-findings are nonetheless interpreted and later canonized as dispositive evidence.

For their analyses, HFH examine police-recorded criminal incidents gleaned from open records requests in public restrooms, locker rooms, and changing rooms in a small set of Massachusetts municipalities around Boston before and after the adoption of local GIPPs. The study aims to compare municipalities that adopted such ordinances with matched municipalities that did not, using pre-post comparisons over a one- to two-year period and focusing on offenses such as sexual assault, voyeurism, and indecent exposure. Based on these comparisons, the study's goal is to assess whether the passage of GIPPs is associated with relative increases in safety or privacy violations in these spaces, compared to matched controls.

Turning to limitations, although not a methodological weakness, per se, the lack of methodological transparency is an oversight in the study that colors all aspects of the analysis. Key aspects of the study design, matching procedures, data inclusion criteria, and analytic decisions are insufficiently described, making the analyses difficult to follow and impossible to replicate. Importantly, the analyses are not just unclear but are inconsistent or misleading in places. For example, the study is framed as a rigorous quasi-experimental design, with terms such as "matched comparisons", "genetic matching", and "difference-in-differences" to describe the approach. This description suggests a level of methodological sophistication appropriate for causal inference in observational research. However, close inspection reveals that the actual analyses fall short of these standards.

For example, although the article uses the "difference-in-differences" (DiD) label and references work describing and performing the standard regression-based DiD design, the study does not actually implement it. In a standard DiD analysis, researchers estimate whether the change over time in an outcome (here: crime rates) in the treatment group (GIPP locales) differs from the change over time in a comparable control group (non-GIPP localities) in a regression model that includes group, time, and crucially, a group-by-time-interaction to isolate the effect of the policy (Donald & Lang, 2007; Goodman-Bacon, 2021). The key output is this interaction term, which is the estimate of the average treatment effect—the extent to which the outcome changed in the treated group more (or less) than in the control group. Notably, DiD models rely on the crucial parallel trends assumption: that, absent the policy change, treated and control units

would have followed similar trends over time (Dimick & Ryan, 2014; Donald & Lang, 2007). None of these components are present in HFH’s analysis. There is no regression model estimating group-by-time interaction, the analysis does not test for or present evidence regarding the parallel trends assumption, and the presented evidence raises concern about potential violations.¹⁷ Therefore, article’s description of the approach as a “difference-in-difference[s] method” is misleading, as the analysis lacks the basic elements that define the method as widely implemented,¹⁸ and the description creates a false impression of methodological credibility that the design does not support.

Perhaps the most notable example of the lack of transparency—and the most significant limitation of the study—is that nowhere does the article clearly report the (small) study N (the number of jurisdictions included in the primary analyses). Table 2 misleadingly implies a broad comparative analysis across seven to nine GIPP jurisdictions. However, close inspection of Table 2 and the footnotes suggests that the effective N for the primary analyses is just *three* Massachusetts municipalities: Medford, Melrose, and Newton.¹⁹ These treated localities are not explicitly identified as the focal treated localities, and the sample size is not stated outright. Instead, the sample size must be inferred by the reader. Omitting a basic design feature of this kind is not defensible in an empirical publication. That such a basic design feature must be reconstructed by inference exemplifies the broader problem: the study’s reporting obscures foundational limitations through omission and ambiguity.

Ultimately, a sample of this size offers no meaningful statistical power. In such a design, any inferential statistical testing—however carefully executed—is largely exercise without substance. With only three treated units, there is no capacity to detect generalizable effects and no reliable basis for estimating uncertainty. The use of Fisher’s exact tests, manually calculated difference scores, and standardized difference estimates creates the appearance of rigor, which is lacking. Inferential statistics are meant to quantify uncertainty, test hypotheses, or generalize beyond observed data. With three cases, inferential statistics can credibly do none of those things. In effect, what is

¹⁷ The data presented by HFH strongly suggest that the parallel trends assumption is violated. Prior to the policy change, the matched (untreated) localities reported a baseline incident rate of 2.5 per 100,000, while the GIPP (treated) localities reported zero incidents and thus a zero rate. This is not a trivial imbalance. When treated units begin with consistently lower outcome levels than controls, and these disparities persist or widen post-treatment (as they do here), it becomes impossible to attribute differences in trends to the policy rather than to pre-existing, systematic differences between localities. Despite claiming to have matched on pre-policy crime indices, the analyses did not match on the actual outcome of interest: public accommodation crime incidents. The failure to achieve pre-treatment equivalence directly undermines the validity of the comparisons. This is a flaw that renders any causal interpretation unreliable.

¹⁸ The authors might argue that they approximated a treatment-by-time interaction informally by calculating the change in incidents from pre- to post-policy in each treated locality, doing the same for its matched control locality, subtracting the control locality’s change from the treated locality’s change (i.e., a crude difference-in-differences), and then standardizing that result per 100,000 people in the treated city (as reported in the “Difference per 100,000 (clear matched)” column in Table 3, which is not described clearly). However, this approach falls short of what constitutes a “DiD analysis”. Simply calculating a difference-in-differences by hand, without modeling uncertainty or evaluating comparability over time, is an analysis, but it is not a DiD analysis.

¹⁹ All three of these are located in the inner-ring suburbs of Boston, with direct Massachusetts Bay Transportation Authority (MBTA) access to downtown, and as the authors noted, Boston has had a GIPP since 2002.

presented as a quasi-experimental analysis is, in reality, a descriptive account of three cases dressed in the trappings of formal statistical inference. This approach, and the evidentiary claims built on it, conveys a level of inferential rigor the design cannot support.

Adding the base-rate problem—the rarity of the criminal events analyzed—to the tiny sample size further weakens the design. The events HFH examine (e.g., sexual assaults, voyeurism, and indecent exposure) in public restrooms and locker rooms are, thankfully, exceedingly rare, with counts often at or near zero across treated and comparison localities and observed periods (see HFH, Table 3). This matters because their substantive claim is about policy-induced changes in already rare events, which creates several linked problems. First, when the outcome is rare regardless of policy, then most locality-years will show zero or near-zero incidents whether the policy matters or not. In such settings, zero or near-zero events is largely what we would expect under multiple different underlying realities (no effect, a small increase, moderate increase that does not show up in a given year, or heterogeneous effects). Given this, the rarity makes the data weakly diagnostic. Second, HFH’s argument is not that these events are rare—although they do make that point—but that their analysis can adjudicate whether fears are empirically grounded. However, given that incidents are rare to begin with, the study has very little ability to detect the effects people are actually worried about unless the effects are large and obvious. (And many expressing concern have stated that the fear is not a large increase in these crimes, but rather that some incidents that would otherwise have been prevented will now occur along with broader privacy and uncertainty accompanying GIPPs, see Burt, 2020, 2022). The rarity interacts with the small sample size. With only a handful of treated localities, any true increase would have to manifest in a very conspicuous way to move the aggregate counts. Otherwise, the modal observation remains zero or near-zero. Thus, the issue is not merely “low power” in the textbook sense, but that the design occurs in a setting where the expected observed outcome is potentially the same under different competing hypotheses. Finally, because the incidents are rare, any conclusions become highly sensitive to slight differences in police reporting or recording, matching quality, classification rules, and the like. Thus, recognizing these events are exceedingly rare, as HFH do, does not support the inference that “therefore the policy is harmless”. The consequence is that even under the best conditions, observing significant differences will be unlikely even differences exist.

Furthermore, these are not the best conditions. Although the article uses language implying methodological rigor and sophistication for the matched pairs analyses, the implementation is neither transparent nor analytically defensible. The article indicates that each treated locality is matched and compared with non-treated (i.e., non-GIPP) localities for each of the GIPP localities examined. However, if we look at Table 2, we see that that Medford is compared to Beverly and Watertown; Melrose is compared to Beverly and Beverly (perhaps the duplication is a typo); and Newton is compared to Brookline (noted as having ‘unclear enforceability’) and Arlington. This repeated use of Beverly is a problem for the analysis, as it violates the assumption of independent comparison.²⁰ Any idiosyncrasies in Beverly (crime reporting practices, demographics,

²⁰ In matched-pair designs, the independence assumption generally requires that each treated-control pair constitute an independent unit of comparison. Reusing the same control locality across multiple treated localities violates this assumption by inducing statistical dependence across

behavior, law enforcement) now affect two (of three) treatment comparisons, and one twice. Worse still, Brookline—classified as a non-GIPP locality—appears to *have* a GIPP but without clear enforceability, which makes it wholly inappropriate as a non-GIPP locality. The matched localities simply do not appear robust.²¹

Although the description suggests a sophisticated matching strategy, insufficient detail is provided to instill confidence in the rigor of the comparisons. The matching process is described only in general terms, without key diagnostics that would allow readers to assess their validity, including balance statistics, covariate comparison tables for treated versus matched localities, full matching diagnostics, distributions of propensity scores, explanations of how matched localities were combined in the analysis, and discussions of unmatched cases or the sensitivity of results to different matching specifications. Without this information, it is impossible to evaluate whether the matches were truly comparable, or whether the results might be an artifact of poor matching quality. This lack of clarity about how the genetic matching outputs were used (or possibly overridden, see confusing discussion on p. 74), makes it impossible to rule out post hoc or outcome-informed selection. But again, with three treated jurisdictions and near-zero base rates, the best this could be even with excellent matches is a rough descriptive study of GIPP effects in a handful of Massachusetts localities around Boston where enforcement is unclear.

Finally, beyond problems with how the outcome was conceived and measured, there are equally serious concerns about the validity of the treatment itself. Perhaps most fatal to the study's causal logic is the fact that even municipal officials (e.g., police departments, town clerks, and human rights staff) were often unclear on the content, enforceability, or even existence of their own local GIPPs (see p.74). If those tasked with implementing the policies are uncertain about their scope or substance, it is unlikely that residents, visitors, or potential offenders would have meaningful awareness of them. This problem is intensified in a geographically dense and highly mobile region of the Boston area around Massachusetts, where individuals routinely cross municipal boundaries for work, errands, schooling, and social activities. The idea that someone contemplating misconduct would adjust their behavior based on mostly invisible jurisdictional lines—and a poorly publicized local policy of ambiguous legal significance—seems highly unrealistic. The short time span (1-2 years) covered contributes to this uncertainty.

In sum, HFH's bathroom safety study suffers from profound methodological deficiencies at nearly every level: ambiguous treatment (and definitions), uncertain and apparently poorly matched comparisons, indefensible statistic inference from a sample

comparisons and inflating the influence of a single control unit, which compromises the reliability of standard errors and weakens the validity of inferential claims.

²¹ The article also includes an appendix titled “Placebo Matched Pairs Analysis” that re-conducts the analysis using a second matching procedure and adds “limited GIPANDOs” as a kind of placebo comparison. This appendix, however, does little to resolve the core concerns raised here. It still rests on the small sample with low base rates, omits the information needed to evaluate the exercise rigorously, including a clear statement of the effective sample, transparent identification of all localities used in each comparison, and the balance diagnostics necessary to assess whether the matches were credible. At most, then, the appendix offers a thin robustness gesture rather than a persuasive demonstration of analytic rigor, as it remains constrained by the same weaknesses that pervade the study.

of three treated units, inconsistent and under-specified outcome data, and a lack of transparency that obscures rather than clarifies core analytic decisions. Even if some elements reflect ordinary lapses in methodological rigor, the cumulative weight of these problems renders the study incapable of supporting the causal claims it advances, discussed next.

Findings and Interpretations

HFH's study's empirical results are limited. Their "DiD model" fails to reject the null hypothesis of no difference in treated and untreated (GIPP and non-GIPP) localities in crime rates after the passage of GIPPs. They thus do not observe a relative increase (or statistically significant difference) in the recorded incidents analyzed after passage of GIPPs, compared with their matched jurisdictions. In the null-hypothesis statistical testing framework HFH employ, this result indicates insufficient statistical evidence to detect differences; this finding does not constitute affirmative evidence of no differences because the design is too weakly informative and the uncertainty too great. In other words, failure to reject the null is properly interpreted as a non-finding that leaves the substantive question unresolved rather than as confirmatory evidence of equivalence or safety.

This is not, however, how the study findings are interpreted. Instead, a weak local non-finding is treated as confirmation that the laws are unrelated to safety and privacy violations in these spaces, and further advanced as dispositive evidence that GIPPs pose no threat to public safety in these domains. In doing so, an uncertain statistical result is elevated into a prematurely settled empirical conclusion that the study is not designed or statistically powered to support.

Notably, the authors do acknowledge that their null findings "may be driven by the small number of localities included in the analysis," and, to address that concern, they conduct a post-hoc power analysis "to assess whether [their] finding of no difference was a result of [their] small sample size" (p.77). These results are interpreted as showing that the findings would likely remain non-significant even in a larger sample. However, post-hoc power analyses are widely recognized as uninformative because they do not provide meaningful information beyond the observed result. In this case, the problem is especially severe. With an effective sample of three treated units, near-zero outcome counts, nonrandom case selection, and contestable matches, the exercise cannot distinguish between a genuine null effect in these settings and a design that is simply too limited and weakly informative to detect a relative difference. Rather than resolving the study's indeterminacy, the power analysis simply reflects it.

While we acknowledge the inherent challenges of studying this issue, the cumulative deficiencies documented here mean that the study's findings can support only one very modest claim: that there was no relative increase in police reports of crimes in public accommodations in three Boston-area jurisdictions with unclear or inconsistently enforced policies compared to that observed in their matched localities. And even that claim is qualified by the fact that behaviors formerly chargeable or recordable as crimes may no longer generate comparable criminal incident reports under GIPPs, which further weakens any inference from these data to actual safety, much less to privacy effects.

Category Reconstitution in HFH

Although embedded in the above discussion, we briefly summarize the category reconstitution pipeline discussed above as it manifests in HFH's study to trace how upstream ontological commitments structure downstream analytic, interpretive, and policy reasoning. As discussed, ideological commitments need not be explicitly declared to be analytically operative. Category reconstitution functions through often invisible lexical choices, classificatory assumptions, and framing decisions that enact ideologically inflected ontological commitments. In fact, the absence of explicit declaration is diagnostic of the degree to which the reconstituted categories have become naturalized as background assumptions rather than contested claims. What follows traces how these implicit operations structure HFH's analysis at each stage. Examining these transformations clarifies how the study's conceptual framing constrains what can register as an effect and shapes how evidence can be interpreted.

Ontological Transformation

The study adopts the reconstituted ontology of GII, in which sex is no longer treated as a meaningful or policy-relevant category. Within this framework, gender identity functions as the operative basis of category membership, while sex is reduced to a retrospective administrative label. This ontological commitment is never stated explicitly, but instead is embedded in the framing, analysis, and interpretation in ways that shape which classifications are treated as real and meaningful, and which are rendered analytically obsolete or unintelligible.

Analytic Transformation

Once this ontology is in place, HFH operationalize "impact" and "harm" in ways that exclude sex-based composition effects from consideration. In particular, the GIPP shift from single-sex to mixed-sex spaces is effectively ignored. Within this analytic frame, the central mechanism motivating safety concerns—male access to female spaces—*does not register as an effect* to be considered or measured.

Policy Transformation

HFH then redescribe sex-based public-accommodations provisions as "proscribing transgender inclusion" and frame GIPPs as "inclusive nondiscrimination laws." These descriptors imply, for example, that transwomen are members of the category "women," such that sex-based policies are recoded as exclusion of a subset of women rather than as the maintenance of female-only spaces. The policy dispute is thereby shifted from questions of justification, scope, and tradeoffs to a moralized question of inclusion versus discrimination.

Normative Transformation

Finally, dissenting positions grounded in sex-based rationales are framed as affectively driven and epistemically weak, for example, "emotionally stirring" arguments that "often lack empirical validity" (p.72), while the pro-GIPP position is assumed to be protective and morally justified. Within this frame, skepticism toward GIPPs is no longer legible as a principled disagreement over classification or risk, but is recast as prejudice, moral panic, or irrational fear.

Altogether, these transformations clarify why HFH's design and inferences have the structure they do. The study presents itself as testing whether GIPPs create "a loophole for [male] sexual predators to access women's public restrooms and locker rooms" (p.

70); however, the analytic framework, which is shaped upstream by category reconstitution, renders the sex-based mechanism (male access) at the heart of that claim effectively invisible. This supports the study's reliance on aggregate incident counts and treatment of null findings as dispositive reassurance.

This synthesis is also key to explaining the study's unusual public and scholarly uptake. When category reconstitution has already converted a policy disagreement into a moralized dispute about "inclusion" versus "exclusion," a weak empirical design can nonetheless appear to resolve the controversy, because it is seen as vindicating the taken-for-granted frames. This is the kind of upstream foreclosure that sets the stage for the downstream canonization processes examined next.

In short, HFH's framework presupposes that sex-based policies are discriminatory against transgender people. Once that presupposition is built into the conceptual foundation, the study's empirical task shifts from weighing competing risks to demonstrating the presupposition empirically. Within this frame, searching for evidence of unintended harms can be morally suspect, and null findings are more easily interpreted as dispositive and politically actionable

Canonization Bias

Canonization bias refers to the process by which empirically fragile or methodologically flawed findings are elevated to the status of authoritative or settled knowledge not through cumulative evidentiary support but through their ideological congeniality (Burt, forthcoming; Honeycutt & Jussim, 2020; Jussim et al., 2019). Once canonized, such findings are no longer treated as provisional contributions to an ongoing empirical debate but instead function as background assumptions that structure subsequent research, interpretation, and policy discourse. Not simply overcitation, canonization bias involves the suspension of ordinary standards of methodological scrutiny and organized skepticism combined with the foreclosure of further inquiry. Within an ideologically homogeneous research milieu, the suspension of organized skepticism facilitates not only the uptake of such flawed results, but their premature stabilization as settled knowledge, discouraging reanalysis, marginalizing dissent, and narrowing the scope of legitimate scholarly debate (Honeycutt & Jussim, 2020; Savolainen, 2024).

Here, canonization bias is observable when an empirically fragile study is cited as dispositive evidence resolving an empirical question, rather than as one contribution among others subject to uncertainty, qualification, or ongoing debate. HFH's article fails to meet basic standards of methodological rigor and transparency, and in our view should not have passed peer review in its current form. Nevertheless, the article has been widely cited and treated as authoritative evidence that safety or privacy concerns regarding GIPPs are empirically unfounded and therefore irrational or indicative of prejudice (e.g., Atwood et al., 2024; Morgenroth et al., 2024a).

Specifically, HFH's study is cited as resolving the empirical question of whether GIPPs affect public safety in relevant domains. For example, Turnbull-Dugarte and McMillan (2023) cite HFH's article as showing "no empirical link" between GIPPs and changes in criminal incident reports, framing safety concerns as a persuasive strategy aimed at "impressionable citizens" rather than as an open empirical question. Similarly, Atwood et al. (2024) describe the claim that predatory males might exploit trans-inclusive policies to access women's spaces as having been "debunked as a myth" and as "empirically

unsubstantiated,” citing HFH’s article (p.171; see also, Morgenroth et al., 2024a; Valcore et al., 2021; Zanghellini, 2020). Lamble (2024) likewise characterizes safety concerns as “clearly misdirected,” asserting that “there is no credible empirical evidence that trans rights pose a threat to women,” with HFH offered as supporting authority (p.510).

The claim that there is “no credible empirical evidence that GIPPs pose a threat to women” warrants direct scrutiny. First, documented incidents, including media, police reports, and court records, including cases such as the 2021 Wi Spa incident in Los Angeles, in which a male lewdly exposed their genitalia in the women’s section of the spa and invoked transgender identity as justification, constitute empirical evidence. In this incident, authorities initially declined to act on the grounds that the individual’s presence in the women’s section was lawful under the applicable GIPP, before it emerged the individual was a convicted, registered sex offender with other recent exposure incidents, and charges were ultimately filed (Queally & Chabria, 2021; McGahan, 2021; Stuart, 2021). Ignoring such occurrences while treating an aggregate null finding from three municipalities as definitive scientific evidence is an instance of the asymmetric evidentiary skepticism we discuss here.

Second, the claim cannot be evaluated in isolation from the broader evidence base. Existing research indicates that males offend at substantially higher rates than females for violent and sexual crimes, and that offending tracks sex not gender identity, such that transwomen offend at rates comparable to non-trans males rather than females (Dhejne et al., 2011). That too is empirical evidence grounding the concern that male access to female spaces carries risk. A null finding from an underpowered study does not rebut that prior evidence base but adds null results to a literature where available evidence already supports the concern. A pharmaceutical analogy is useful here. An underpowered, flawed trial that fails to detect a plausible side effect does not establish “no credible evidence” of that side effect, particularly when prior evidence lends credence to its existence. Characterizing HFH’s null finding as establishing that safety concerns are empirically unfounded is thus an inferential error.

In short, across the subsequent HFH citations, the absence of detected effects in a single, methodologically constrained study is progressively upgraded from “no empirical link,” to “debunked myth,” to “no credible empirical evidence,” and treated as affirmative refutation rather than as a non-finding. Importantly, this interpretive overreach originates in the authors’ own public-facing presentation of the study. A study press-release published by the Williams Institute,²² which includes quotes from the first and second authors, states that GIPPs “do not jeopardize safety in public bathrooms” and that the study “provides evidence” that alleged safety incidents are “rare and unrelated to the laws.” In the press release, the second author asserts: “This study should provide some assurance that these types of public accommodations laws provide necessary protections for transgender people and maintain safety and privacy for everyone.” These claims go well beyond what the study’s analytic design can support. In this way, the study results are no longer presented as provisional or contingent, but as settled background knowledge foreclosing further empirical inquiry and reframing concern about or objection to GIPPs as irrational or bigoted.

²² <https://williamsinstitute.law.ucla.edu/press/trans-pub-acc-laws-press-release/>

This overinterpretation is also observed in media and policy discourse, where HFH's study is invoked to assert that public safety concerns regarding GIPPs are not merely unproven but *refuted* by science. For example, *The Boston Globe* reported that a "first-of-its-kind study ... refutes the premise that the state's transgender antidiscrimination law threatens public safety," citing HFH as authoritative support (Ebbert, 2018). Similarly, an *NBC News* article on HFH's study reported "there is no evidence that letting transgender people use public facilities that align with their gender identity increases safety risks," and quoted advocates asserting that the study "once and for all puts this issue to rest" (Moreau, 2018).

FROM CASE-STUDY TO FIELD-LEVEL EPISTEMIC DIAGNOSIS

The preceding analysis of canonization bias suggests that HFH's (2019) study came to function as authoritative evidence despite being almost methodologically non-informative. The question is whether this trajectory reflects an anomalous failure of scholarly judgment or instead reveals a broader epistemic pattern characteristic of ideologically homogeneous research domains.

Readers may rightly point out that a single flawed study, even one that is highly cited, does not, on its own, warrant a diagnosis of field-level epistemic distortion. Ordinary scientific practice tolerates isolated errors, weak designs, and interpretive excess, which are typically corrected through replication, critique, and cumulative evaluation. To be sure, some features of HFH's study are plausibly idiosyncratic. These include the unusually small number of treated jurisdictions, the opaque reporting of sample size and matching procedures, and the specific inconsistencies in the implementation of quasi-experimental language. Any one of these shortcomings, taken alone, could reflect ordinary lapses in peer review or author judgment rather than a systematic problem. Here, however, the epistemic significance lies in their co-occurrence. In HFH's study, upstream ontological commitments structure the analytic frame, that frame renders key mechanisms invisible, and indeterminate results are then interpreted as reassuring.

Moreover, less idiosyncratic and more epistemically diagnostic are other features that align with recurrent patterns observed in the field of transgender research and help explain the study's influence. First, as discussed, HFH's uncritical adoption of the dominant gender-identity lexicon ("gender identity nondiscrimination," "inclusive policies," "sex assigned at birth") serves as an ontological commitment that structures the entire analysis. These terms embed contested normative assumptions upstream of empirical testing and enact category reconstitution by displacing sex as an analytically relevant category. As a result, sex-based mechanisms, which are central to the policy controversy at issue, are rendered conceptually unintelligible rather than empirically evaluated. This pattern is not unique to HFH's study but is widely observable in transgender research, where alternative framings are often excluded by definition rather than examined (see discussion in Burt, forthcoming).

Second, the study exhibits pronounced asymmetry in evidentiary skepticism. Claims consistent with gender-identity ideology, such as the assumption that GIPPs protect transgender people or that sex-based policies increase harassment of gender-nonconforming individuals, are treated as background premises rather than hypotheses requiring empirical support. By contrast, sex-based safety concerns are framed as speculative, emotive, or myth-driven and subjected to empirical testing under standards

the study's design cannot credibly satisfy. This selective suspension of skepticism is a well-documented feature of ideologically inflected inquiry, in which congenial claims are presumed while uncongenial claims are problematized (Clark & Winegard, 2020; Jussim, 2017).

Third, HFH's study exemplifies a null-to-refutation interpretive slide, as the systematic upgrading of atheoretical non-findings into affirmative evidence that an effect does not exist. As discussed above, the study's design is severely underpowered, characterized by near-zero base rates, ambiguous treatment definitions, and minimal analytic leverage. Under ordinary inferential norms, such results leave the substantive question unresolved. Yet HFH's null findings are repeatedly interpreted, both by the authors and by subsequent citers, as evidence that safety concerns are "not empirically grounded," without appropriate caveats. This transformation of indeterminacy into reassurance is not a statistical misunderstanding in isolation, but a predictable outcome when null results align with a dominant ideological frame.

Ultimately the greatest significance of this case in revealing field-level distortions is the way in which its non-findings were transformed into settled knowledge and subsequently insulated from correction. Crucially, it is this combination of upstream category reconstitution, asymmetric skepticism, and interpretive overreach that makes canonization possible. The study's authority does not derive from evidentiary strength, methodological rigor, or theoretical insight, but from its alignment with a preexisting ideological consensus that treats the moral legitimacy of GIPPs as settled. In this way, even methodologically fragile work can function as dispositive evidence, because it appears to empirically vindicate assumptions already treated as true (and 'good').

To be sure, this field-level diagnosis is open to refutation. Several counterfactual patterns would undermine the claim that HFH's influence reflects field-level ideological biases rather than ordinary scholarly error. These could include (a) routine critical engagement with the study's methodological limitations in the literature citing it; (b) subsequent empirical work explicitly attempting to replicate or improve upon its design; (c) symmetrical skepticism applied to claims about both the risks and benefits of GIPPs; or (d) explicit acknowledgment, in scholarly and public discourse, that the empirical question remains unsettled due to limited or weak evidence. The absence of these corrective responses, combined with the study's rapid elevation to myth-busting authority, supports the inference that HFH's study's trajectory is heavily influenced by canonization bias operating within an ideologically homogeneous epistemic environment.

Seen in this light, HFH is not merely a flawed study that escaped correction, but a revealing case of how shared ideological commitments erode standards of scrutiny, permitting epistemically fragile work to stabilize as settled knowledge while foreclosing further inquiry. The bathroom safety study is thus best understood not as a curious lapse in scholarly standards, but as a predictable outcome of the epistemic conditions generated by ideological homogeneity and its attendant epistemic discontent.

CONCLUSION

This article conducts an ideological epistemic analysis (IEA) of a widely cited and highly publicized study that has been treated as dispositive evidence that GIPPs pose “no safety risk.” In doing so, it refines the IEA framework by elaborating two linked mechanisms, category reconstitution and canonization bias, and demonstrates how an almost non-informative design can nonetheless be transformed into authoritative “evidence” when it aligns with an entrenched ideological frame. The central problem, on this account, is not merely that values shape research agendas (they always do), but that in ideologically homogeneous domains the normal error-correcting mechanisms of science—organized skepticism, adversarial engagement, and careful calibration of evidentiary claims to design strength—can weaken precisely when they are most needed (Burt, forthcoming; Duarte et al., 2015; Rauch, 2021; Savolainen, 2024; Wright & DeLisi, 2015).

Importantly, our goal in this particular IEA case study, like IEAs in general, is not motive attribution or scholar indictment. The point is that predictable distortions arise when a field’s ideological priors become sufficiently uniform that dissenting framings are treated as epistemically suspect or morally illegitimate, and this can occur even when individual scholars act in good faith. Under those conditions, ideologically inflected assumptions can foreclose downstream empirical testing. In the present case, category reconstitution functions as an ontological and analytic constraint: by redescribing sex as “assigned at birth,” shifting membership in sex categories onto gender identity, and treating sex-based composition effects as conceptually irrelevant, the focal empirical mechanism in the public controversy (male access to female spaces) becomes difficult to register as an “effect” to be measured at all. Once the admissible space of effects is narrowed in this way, null findings from underpowered and opaque analyses can be readily upgraded into “myth-busting” conclusions. Canonization bias then stabilizes these conclusions as background premises. In this case, a non-finding is repeatedly cited as refutation, equivocal evidence is laundered into certainty, and the bounds of legitimate debate shrink. The epistemic and social stakes of this dynamic are high in any domain, but especially in transgender research, where scholarship is routinely invoked in court filings, legislative debate, and institutional policy, and where public-facing claims of settled science can be amplified far beyond what the evidence warrants.

The remedy implied by this analysis is not a demand that scholars purge values, which is impossible, but a demand to rebuild and institutionalize the conditions of reliable error correction in contested, high-stakes research domains (e.g., Duarte et al., 2015; Jussim et al., 2019). Several straightforward practices would directly target several epistemic failures identified here: (1) basic design transparency as a non-negotiable norm (including clear reporting of sample sizes, comparisons, time windows, inclusion criteria, etc.) (Ioannidis, 2005; Munafò et al., 2014); (2) interpretive discipline that matches claims to designs (Jussim et al., 2016); (3) routine independent reanalysis and adversarial replication (see Isch et al., 2025; Kahneman, 2003), ideally incorporating researchers with divergent priors, to prevent epistemic closure from functioning as a substitute for robustness (Duarte et al., 2015; Smith, 2025; Wright & DeLisi, 2015); (4) citation hygiene norms that discourage “debunked myth” language absent cumulative, methodologically credible evidence, and that require explicit qualification when citing fragile or non-informative designs (Jussim et al., 2019); and (5) conflict-of-interest analogs for advocacy-oriented knowledge environments, emphasizing disclosure and

structural separation between moral-political dissemination goals and evidentiary claims.

However, perhaps the most epistemically consequential deficiency in much contemporary transgender research is the absence of scientific theory, defined as articulated explanatory frameworks that specify mechanisms, generate testable hypotheses, and delineate scope conditions (Brint, 2025; Cohen, 1989; Turner, 1986). If present, theory is often superficial (Iceland & Silver, 2024) or replaced with normativity in ideologically homogeneous, advocacy-oriented scholarly domains (Rubin, 2025). That is, upstream of weak, flawed methods or biased interpretation is the lack of an articulated explanatory framework capable of specifying mechanisms, generating competing hypotheses, and identifying what empirical findings would count as disconfirming. Where theory is absent or thin, ideology is easily mistaken for explanation (Rubin, 2025), such that normative commitments function as implicit premises rather than as claims requiring empirical justification, which allows internally incoherent claims to persist and empirically fragile findings to be interpreted as confirmatory regardless of design strength or informativeness.

Under conditions of ideological homogeneity, explanatory development further languishes. Favored narratives increasingly crowd out alternatives, anomalies are discounted, and empirical uncertainty is converted into apparent consensus (Iceland & Silver, 2024). In this context, null or non-informative findings can be rhetorically upgraded into authoritative conclusions and subsequently canonized as settled knowledge (Clark & Winegard, 2020; Honeycutt & Jussim, 2020). Addressing this deficit requires treating theory not as optional commentary (or moral orientation) but as an essential epistemic constraint, which disciplines interpretation, forces engagement with counterevidence, and restores the error-correcting capacities of scientific inquiry, especially in contested fields with sociopolitical resonance. Taken together, the absence or underdevelopment of theory, the narrowing of admissible explanations, and the failure of responsible inference create conditions under which epistemically fragile work can nonetheless stabilize as authoritative, dispositive knowledge.

In conclusion, IEAs are useful for bringing these upstream conditions into view, demonstrating how shared moral-ideological commitments, when unconstrained by theory and buffered from adversarial scrutiny, can allow weak evidence to acquire the status of settled science or strong authority. Repairing these conditions is both a tractable scientific problem and a prerequisite for sustaining the credibility of social science in domains where empirical claims are rapidly translated into law, policy, and public understanding. When research appears to track ideological commitments more reliably than evidence, public trust erodes. Rebuilding theoretical discipline, pluralistic scrutiny, and responsible inference is not only valuable for knowledge production but is central to preserving social science as a genuinely truth-seeking enterprise rather than ideological advocacy conducted under scientific branding.

Declarations:

Ethical Approval: Not applicable (No animal or human subjects)

Conflict of Interest: The authors declare they have no conflict of interest.

Funding: No funding was received for this project.

Data Availability Statement: No original datasets were generated or analyzed for this study. All materials examined consist of publicly available publications, documents, and citations.

REFERENCES

- Ainsworth, C. (2015). Sex redefined: The idea of two sexes is simplistic. *Nature*, *518*, 288–291.
- Al Gharbi, M. (2018). Race and the race for the White House: On social research in the age of Trump. *The American Sociologist*, *49*, 496–519.
- Atwood, S., Morgenroth, T., & Olson, K. R. (2024). Gender essentialism and benevolent sexism in anti-trans rhetoric. *Social Issues and Policy Review*, *18*(1), 171–193.
- Baron, J., & Spranca, M. (1997). Protected values. *Organizational behavior and human decision processes*, *70*(1), 1–16.
- Brint, S. (2025). High-leverage sociological concepts and the progress of theory, part one. *Theory and Society*, *54*(5), 705–730.
- Burt, C. H. (2020). Scrutinizing the U.S. Equality Act 2019: A feminist examination of definitional changes and sociolegal ramifications. *Feminist Criminology*, *15*(4), 363–409.
- Burt, C. H. (2022). Discounting females, denying sex, and disregarding dangers from self-ID—a reply and a defense of open debate. *Journal of Controversial Ideas*, *2*(1). doi:10.35995/jci02010015
- Burt, C. H. (2023a). Feminist lesbians as anti-trans villains: A comment on Worthen and elaboration. *Sexuality & Culture*, *27*(1), 161–190.
- Burt, C. H. (2023b). Sex, gender, and the US Equality Act: Confusion, conflict, and consequences. In A. Sullivan & S. Todd (Eds.), *Sex and Gender Identity: A Reader* (pp. 139–155). London: Routledge.
- Burt, C. H. (forthcoming). Ideological homogeneity and its epistemic discontents: A case study of research on transgender issues. *Theory and Society*.
- Butler, J. (1990). *Gender trouble: Feminism and the subversion of identity*. New York: Routledge.
- Clark, C. J., & Winegard, B. M. (2020). Tribalism in war and peace: The nature and evolution of ideological epistemology and its significance for modern social science. *Psychological Inquiry*, *31*(1), 1–22.
- Cohen, B. P. (1989). *Developing sociological knowledge: Theory and method* (Second ed.). Chicago: Nelson-Hall.
- Coyne, J., & Maroja, L. S. (2023, July/August). The ideological subversion of biology. *Skeptical Inquirer*, *47*, 4. Retrieved from <https://skepticalinquirer.org/2023/06/the-ideological-subversion-of-biology/>
- Dhejne, C., Lichtenstein, P., Boman, M., Johansson, A. L., Långström, N., & Landén, M. (2011). Long-term follow-up of transsexual persons undergoing sex reassignment surgery: Cohort study in Sweden. *PLOS ONE*, *6*(2), e16885.
- Dimick, J. B., & Ryan, A. M. (2014). Methods for evaluating changes in health care policy: The difference-in-differences approach. *Jama*, *312*(22), 2401–2402.
- Donald, S. G., & Lang, K. (2007). Inference with difference-in-differences and other panel data. *The Review of Economics and Statistics*, *89*(2), 221–233.
- Duarte, J. L., Crawford, J. T., Stern, C., Haidt, J., Jussim, L., & Tetlock, P. E. (2015). Political diversity will improve social psychological science. *Behavioral and Brain Sciences*, *38*, e130.
- Ebbert, S. (2018, September 12). Study finds no link between transgender rights law and bathroom crimes. *The Boston Globe*. Retrieved from <https://www.bostonglobe.com/metro/2018/09/11/study-finds-link-between->

- transgender-rights-law-and-bathroom-crimes/1YWqSptLXOSiobmbH0RBMM/story.html
- Ekman, K. E. (2023). *On the meaning of sex: Thoughts about the new definition of woman* (K. Maki, Trans.). Australia: Spinifex.
- Foucault, M. (1972). *The archaeology of knowledge* (A. M. Sheridan Smith, Trans.). London: Tavistock.
- Fuentes, A. (2025). *Sex is a spectrum: The biological limits of the binary*. Princeton University Press.
- Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics*, 225(2), 254–277.
- Goymann, W., Brumm, H., & Kappeler, P. M. (2023). Biological sex is binary, even though there is a rainbow of sex roles: Denying biological sex is anthropocentric and promotes species chauvinism. *Bioessays*, 45(2), 2200173.
- Hasenbush, A., Flores, A. R., & Herman, J. L. (2019). Gender identity nondiscrimination laws in public accommodations: A review of evidence regarding safety and privacy in public restrooms, locker rooms, and changing rooms. *Sexuality Research and Social Policy*, 16(1), 70–83.
- Haslam, N. (2016). Concept creep: Psychology's expanding concepts of harm and pathology. *Psychological Inquiry*, 27(1), 1–17.
- Honeycutt, N., & Jussim, L. (2020). A model of political bias in social science research. *Psychological Inquiry*, 31(1), 73–85.
- Iceland, J., & Silver, E. (2024). When all you have is a hammer: how social justice distorts what we know about racial disparities. *Theory and Society*, 53(5), 1073–1092.
- Ioannidis, J. P. (2005). Why most published research findings are false. *PLoS medicine*, 2(8), e124.
- Isch, C., Tetlock, P. E., & Clark, C. J. (2025). Reflections on adversarial collaboration from the adversaries: was it worth it? *Theory and Society*, 54, 929–946.
- Janicke, T., Häderer, I. K., Lajeunesse, M. J., & Anthes, N. (2016). Darwinian sex roles confirmed across the animal kingdom. *Science Advances*, 2(2), e1500983.
- Jenkins, K. (2016). Amelioration and inclusion: Gender identity and the concept of woman. *Ethics*, 126(2), 394–421.
- Jones, J. C. (2023). The history of sex: Sex denial and gender identity ideology. In A. Sullivan & S. Todd (Eds.), *Sex and Gender Identity: A Reader* (pp. 69–85): Routledge.
- Jussim, L. (2017). Accuracy, bias, self-fulfilling prophecies, and scientific self-correction. *Behavioral and brain sciences*, 40, 65.
doi:<https://doi.org/10.1017/S0140525X16000339>
- Jussim, L., Crawford, J. T., Anglin, S. M., Stevens, S. T., & Duarte, J. L. (2016). Interpretations and methods: Towards a more effectively self-correcting social psychology. *Journal of Experimental Social Psychology*, 66, 116–133.
- Jussim, L., Krosnick, J. A., Stevens, S. T., & Anglin, S. M. (2019). A social psychological model of scientific practices: Explaining research practices and outlining the potential for successful reforms. *Psychologica Belgica*, 59(1), 353–372.
- Kahneman, D. (2003). Experiences of collaborative research. *American Psychologist*, 58(9), 723–730.
- Lamble, S. (2024). Confronting complex alliances: Situating Britain's gender critical politics within the wider transnational anti-gender movement. *Journal of Lesbian Studies*, 28(3), 504–517.
- Langbert, M. (2018). Homogenous: The political affiliations of elite liberal arts college faculty. *Academic Questions*, 31(2).
- Lawford-Smith, H. (2022). *Gender-Critical Feminism*. Oxford University Press.

- Lehtonen, J., Kokko, H., & Parker, G. A. (2016). What do isogamous organisms teach us about sex and the two sexes? *Philosophical Transactions of the Royal Society B: Biological Sciences*, 371(1706).
- Magness, P. W., & Waugh, D. (2022). The hyperpoliticization of higher ed. *The Independent Review*, 27(3), 359–370.
- Marinov, G. K. (2020). In humans, sex is binary and immutable. *Academic Questions*, 33, 152–177.
- Martin, C. C. (2016). How ideology has hindered sociological insight. *The American Sociologist*, 47, 115–130.
- McGahan, J. (2021, December 2). What really happened at Wi Spa. *Los Angeles Magazine*. Retrieved from <https://lamag.com/lgbtq/how-wi-spa-sparked-bloody-battle-over-trans-rights/>
- Moreau, J. (2018, September 19). No link between trans-inclusive policies and bathroom safety, study finds. *NBC News*. Retrieved from <https://www.nbcnews.com/feature/nbc-out/no-link-between-trans-inclusive-policies-bathroom-safety-study-finds-n911106>
- Morgenroth, T., Axt, J. R., & Westgate, E. C. (2024a). What underlies the opposition to trans-inclusive policies? The role of concerns about male violence versus attitudes toward trans people. *Personality and Social Psychology Bulletin*, 50(4), 533–549.
- Morgenroth, T., Means, K. K., Mueller, A. S., & Sass, K. D. (2024b). The Contradictory Nature of Anti-Transgender Rhetoric. *Social Psychological and Personality Science*, 15(7), 848–862.
- Munafò, M., Noble, S., Browne, W. J., Brunner, D., Button, K., Ferreira, J., . . . Lindquist, M. (2014). Scientific rigor and the art of motorcycle maintenance. *Nature Biotechnology*, 32(9), 871–873.
- Queally, J., & Chabria, A. (2021, September 2). Indecent exposure charges filed against trans woman over L.A. spa incident. *Los Angeles Times*. Retrieved from <https://www.latimes.com/california/story/2021-09-02/indecent-exposure-charges-filed-trans-woman-spa>
- Rauch, J. (2021). *The constitution of knowledge: A defense of truth*. Brookings Institution Press.
- Rehman, R. (2023). “Intersex” does not violate the sex binary. *The Linacre Quarterly*, 90(2), 145–154.
- Rubin, A. T. (2025). Normativity is not a replacement for theory. *Theory and Society*, 54, 795–850.
- Savolainen, J. (2024). Unequal treatment under the law: Race, crime & retractions. *Current Psychology*, 43(17), 16002–16014.
- Smith, J. (2025). From heterodoxy to pluralism: The need for a conservative paradigm in social science. *Theory and Society*, 54, 1119–1133.
- Stock, K. (2021). *Material girls: Why reality matters for feminism*. Hachette UK.
- Stuart, G. (2021, July 30). Five women have filed police reports alleging indecent exposure at Wi Spa. *Los Angeles Magazine*. Retrieved from <https://web.archive.org/web/20210801181435/https://www.lamag.com/citythinkblog/wi-spa-police-reports/>
- Sullivan, A., & Todd, S. (Eds.). (2023). *Sex and Gender: A Contemporary Reader*. Taylor & Francis.
- Turnbull-Dugarte, S. J., & McMillan, F. (2023). “Protect the women!” Trans-exclusionary feminist issue framing and support for transgender rights. *Policy Studies Journal*, 51(3), 629–666.
- Turner, J. H. (1986). *The structure of sociological theory* (4th ed.). Chicago: The Dorsey Press.
- Valcore, J., Fradella, H. F., Guadalupe-Diaz, X., Ball, M. J., Dwyer, A., DeJong, C., . . . Worthen, M. G. (2021). Building an intersectional and trans-inclusive

- criminology: Responding to the emergence of “gender critical” perspectives in feminist criminology. *Critical Criminology*, 29(4), 687–706.
- Ward, A., Schenk-Day, A., Woods, M., & Haynie, D. L. (2025). Moral panic and the legislative attack on transgender rights: state-level pathways to anti-trans Lawmaking in the US. *Social Problems*, spaf076.
<https://doi.org/10.1093/socpro/spaf076>
- Wright, C. (2025). Why there are exactly two sexes. *Archives of Sexual Behavior*, 54, 3941–3945.
- Wright, J., & DeLisi, M. (2015). *Conservative criminology: A call to restore balance to the social sciences*: Routledge.
- Zanghellini, A. (2020). Philosophical problems with the gender-critical feminist argument against trans inclusion. *Sage Open*, 10(2), 2158244020927029.
- Zell, E., & Burnett, M. I. (2024). Transgender knowledge mediates U.S. political differences in prejudice and support for trans-inclusive policies. *Sex Roles*, 90, 1879–1890