The four-fifths rule is not disparate impact

A woeful tale of epistemic trespassing in algorithmic fairness

Elizabeth Anne Watkins
elizabeth.watkins@intel.com
Intel Labs Intelligent Systems Research
New York, New York, USA

Jiahao Chen
cjiahao@gmail.com
Responsible AI LLC
New York, New York, USA

ABSTRACT

Computer scientists are trained in the art of creating abstractions that simplify and generalize. However, a premature abstraction that omits crucial contextual details creates the risk of epistemic trespassing, by falsely asserting its relevance into other contexts. We study how the field of responsible AI has created an imperfect synecdoche by abstracting the four-fifths rule (a.k.a. the 4⁄5 rule or 80% rule), a single part of disparate impact discrimination law, into the disparate impact metric. This metric incorrectly introduces a new deontic nuance and new potentials for ethical harms that were absent in the original 4⁄5 rule. We also survey how the field has amplified the potential for harm in codifying the 4⁄5 rule into popular AI fairness software toolkits. The harmful erasure of legal nuances is a wake-up call for computer scientists to self-critically re-evaluate the abstractions they create and use, particularly in the interdisciplinary field of AI ethics.

CCS CONCEPTS

• **Social and professional topics → Computing / technology policy**; **Governmental regulations**.

KEYWORDS

disparate impact, AI ethics, discrimination law, metrics, fairness, bias, optimization, employment, civil rights

ACM Reference Format:
Elizabeth Anne Watkins and Jiahao Chen. 2024. The four-fifths rule is not disparate impact: A woeful tale of epistemic trespassing in algorithmic fairness. In *The 2024 ACM Conference on Fairness, Accountability, and Transparency (FAccT '24), June 03–06, 2024, Rio de Janeiro, Brazil.* ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3630106.3658938

1 INTRODUCTION

Premature abstraction and epistemic trespassing. The field of computer science is oriented around two epistemic motivations: first, to simplify complex problems into mathematical abstractions, and second, to generalize by reusing these same abstractions across other domains [32, 85]. The creation and application of abstractions are integral to defining computer languages and symbolic logic in artificial intelligence [1, 10, 11, 74, 81]. Abstractions discard irrelevant details, which not only reduce cognitive load, but also enable generalizations through use. However, abstractions sometimes result in ontological conflicts, particularly when the details removed in a first formulation, especially those removed out of ignorance as to their salience and those necessary to establish a more general context, are regarded by others to be integral to defining the core concept in the context from which the abstraction is constructed. These premature abstracts, malformed through ontological errors, cause downstream epistemic errors when reused beyond their original scope, resulting in "research debt" [56]. Well-intentioned computer scientists who lack the critical perspective on the initial context may attempt to apply the reified abstraction as a concept and resource in its own right, feeling like they are simply practicing the aphorism that "all models are wrong, but some are useful" [9]. Nevertheless, such "premature abstraction" [23]—using an abstraction without a critical perspective on the original context of its creation—is problematic behavior. By "not staying in their lane", computer scientists can create semantic confusion when reborrowing the premature abstraction back into the original context. Rather than providing genuine contributions to the problem at hand, they become "epistemic trespassers", i.e., "thinkers who have competence or expertise to make good judgments in one field, but move to another field where they lack competence—and pass judgment nevertheless" [4].

Our contributions. In this paper, we argue that epistemic trespassing has formed around the terms "disparate impact" and "four-fifths rule", which poses significant epistemic and deontic risks in real-world, regulated decision-making contexts. By "four-fifths rule", also known as the eighty-percent rule, we refer to the guidelines widely implemented in employment and civil rights contexts to determine whether a hiring or selection process has a disparate impact on protected groups. According to this guideline, a selection process might have "disparate impact" if the selection rate for a particular or protected group is found to be equal to or less than four-fifths of the dominant group. In Section 2, we introduce the various legal concepts that share the name "disparate impact": "disparate impact" ($\mathrm{DI}^{\mathrm{law}}$), a body of U.S. discrimination *law*; "disparate impact" ($\mathrm{DI}^{\mathrm{finding}}$), a legal *finding* by a court or regulator as to whether $\mathrm{DI}^{\mathrm{law}}$ has been violated; and "disparate impact analysis" ($\mathrm{DI}^{\mathrm{analysis}}$), the chain of legal reasoning that argues for a $\mathrm{DI}^{\mathrm{finding}}$. In Section 3, we detail how these concepts have co-opted into "disparate impact" ($\mathrm{DI}^{\mathrm{metric}}$), the *metric* introduced into the algorithmic fairness literature as an imperfect synecdoche of the "four-fifths rule"; we quote the definition of $\mathrm{DI}^{\mathrm{metric}}$ in its entirety in Definition 1.1. In Section 4, we describe the spread of $\mathrm{DI}^{\mathrm{metric}}$ in algorithmic fairness toolkits. In Section 5, we discuss the societal

FAccT '24, June 03–06, 2024, Rio de Janeiro, Brazil

ACM ISBN 979-8-4007-0450-5/24/06
https://doi.org/10.1145/3630106.3658938

ramifications that result from the semantic clashes between the reborrowed abstraction $\mathrm{DI}^{\mathrm{metric}}$ when used in situations where the original legal terms of art apply. To facilitate our discussion, we provide the relevant regulatory paragraph which defines the 4/5 rule in the Equal Employment Opportunity Commission (EEOC)'s Uniform Guidelines on Employee Selection Procedures (UGESP) of 1978 in Appendix A [49–52, 77, 79], an abridgement of which is stated in Table 1.

Definition 1.1 (Disparate impact metric ("80% rule", $\mathrm{DI}^{\mathrm{metric}}$) [22]). Given data set $D = (X, Y, C)$, with protected attribute X (e.g., race, sex, religion, etc.), remaining attributes Y, and binary class to be predicted C (e.g., "will hire"), we will say that D has disparate impact if

$$\frac{\Pr(C = \mathrm{YES}|X = 0)}{\Pr(C = \mathrm{YES}|X = 1)} \leq \tau = 0.8 \tag{1}$$

for positive outcome class YES and majority protected attribute 1 where $\Pr(C = c|X = x)$ denotes the conditional probability (evaluated over D) that the class outcome is $c \in C$ given protected attribute $x \in X$. Note that under this definition disparate impact is determined based on the given data set and decision outcomes.

Related work. The algorithmic fairness literature is sprinkled with various degrees of awareness of the epistemic trespassing problem around "disparate impact". Feldman et al. [22] state that "The terminology of 'right' and 'wrong', 'positive' and 'negative' that is used in classification is an awkward fit when dealing with majority and minority classes, and selection decisions." We revisit this phenomenon of deontic polarization in Section 3. Other papers focus expressly on issues around the de-/re-contextualization inherent in creating and applying abstractions. Bao et al. [5] comment that "Decontextualization of the data creates further problems when algorithmic fairness papers imply that their results have consequences for how [responsible AIs] work (or should work)." Selbst et al. [67] describe "the portability trap" and others that risk creating social harms through overgeneralizations. Martin Jr et al. [38] calls for greater community participation for creating better models and abstractions. Jacobs and Wallach [29] describes risks of abstracting concepts which are challenging to measure, such as gender and teacher effectiveness. The choices of mathematical formalisms around population and data are critical to effectivel achieving fairness goals Mitchell et al. [46], and yet may hide harmful "methodological blindspots" with which the discipline at large must contend [14]. Xiang and Raji [86] draws on theories of disparate impact in their discussion of how machine learning practitioners often misunderstand the legal concepts they attempt to operationalize. To our knowledge, however, we are the first to provide the full synthesis of the extent of epistemic trespassing that has happened around the terms "disparate impact" and "four-fifths rule", which is particularly problematic when reborrowed into the contexts of regulated decision-making not just because of the semantic clash with $\mathrm{DI}^{\mathrm{law}}$, but because of the ubiquity of $\mathrm{DI}^{\mathrm{metric}}$.

2 THE LEGAL CONCEPTS OF DISPARATE IMPACT

In this section, we present the key elements needed by U.S. courts and regulators, which together form $\mathrm{DI}^{\mathrm{analysis}}$ and give rise to a $\mathrm{DI}^{\mathrm{finding}}$ under $\mathrm{DI}^{\mathrm{law}}$. An epistemic trespasser may (falsely) presume that this is simply a matter of applying the 4/5 rule and computing $\mathrm{DI}^{\mathrm{metric}}$ (1) to establish $\mathrm{DI}^{\mathrm{finding}}$. They may even turn to one of the toolkits in Section 4 to perform this computation. This chain of reasoning is an example of a fallacious synecdoche, where $\mathrm{DI}^{\mathrm{metric}}$ vainly stands in for the entire process of $\mathrm{DI}^{\mathrm{analysis}}$ and entire body of $\mathrm{DI}^{\mathrm{law}}$. On the contrary, a proper $\mathrm{DI}^{\mathrm{finding}}$ under $\mathrm{DI}^{\mathrm{law}}$ requires $\mathrm{DI}^{\mathrm{analysis}}$ under $\mathrm{DI}^{\mathrm{law}}$ as shown in Figure 1. $\mathrm{DI}^{\mathrm{analysis}}$ is a complex iterative and multistage test, undertaken with reference to the facts of the specific case.

2.1 Establishing a *prima facie* case of disparate impact

The starting point for a disparate impact assessment is finding statistical evidence of a pattern of unintentional discrimination, which affects a protected class, before turning to mitigation and defences. The resulting *prima facie* case can be established using an appropriate test statistic which compares a relevant population to the specific population that is alleged to have suffered disparate impact along protected class lines, combined with a causal link [76]. For example, to assess if women suffer disparate impact in the hiring of firefighters drawing from all NYC to service Brooklyn, a compliance team could use a χ^2-test (an appropriate statistical test) to compare the $\frac{\text{women in brooklyn}}{\text{men in brooklyn}}$ or $\frac{\text{women in nyc}}{\text{men in nyc}}$ (relevant population), with $\frac{\text{women firefighters servicing brooklyn}}{\text{men firefighters servicing brooklyn}}$ (population in question)[1]. If the test statistic shows a statistically significant difference, this forms evidence to be presented in court or to a regulator.

Selecting a relevant comparison population. The example above highlights an ambiguity in defining the relevant population that forms the basis for comparison when computing a test statistic. Should the reference population be the population of New York City (as the source of applicants), just the borough of Brooklyn (the service area), or something else? Some cases failed to establish $\mathrm{DI}^{\mathrm{finding}}$ because they chose too broad a reference population [3]. On the other hand, the use of general population statistics is not always inadmissible [36]. The set of relevant populations that courts will accept can turn on the legislative history as well as the facts of the case. In one recent case [80], a dispute about the appropriate comparison population drew on analogies to a range of cases in fair housing, but ultimately turned on the differences between legislative intents when writing housing and employment regulations. Ultimately, the choice of relevant comparison population is complex, contingent, and contextual, and cannot be easily abstracted away.

Selecting an appropriate test statistic. Once a reference population has been established, the reference population and the population under review need to be compared. In modern times, this comparison is a statistical one, but the tests that are indicated differ based on the facts of the case. Commonly used test statistics include χ^2 and Fisher's exact tests, each of which is considered reliable, but can occasionally disagree [73]. When conflicts between valid

[1] We acknowledge the existence of genders that fall outside of the gender binary. The law typically compares against each other class individually (one vs one), rather than comparing a class against all other classes simultaneously (one vs rest)

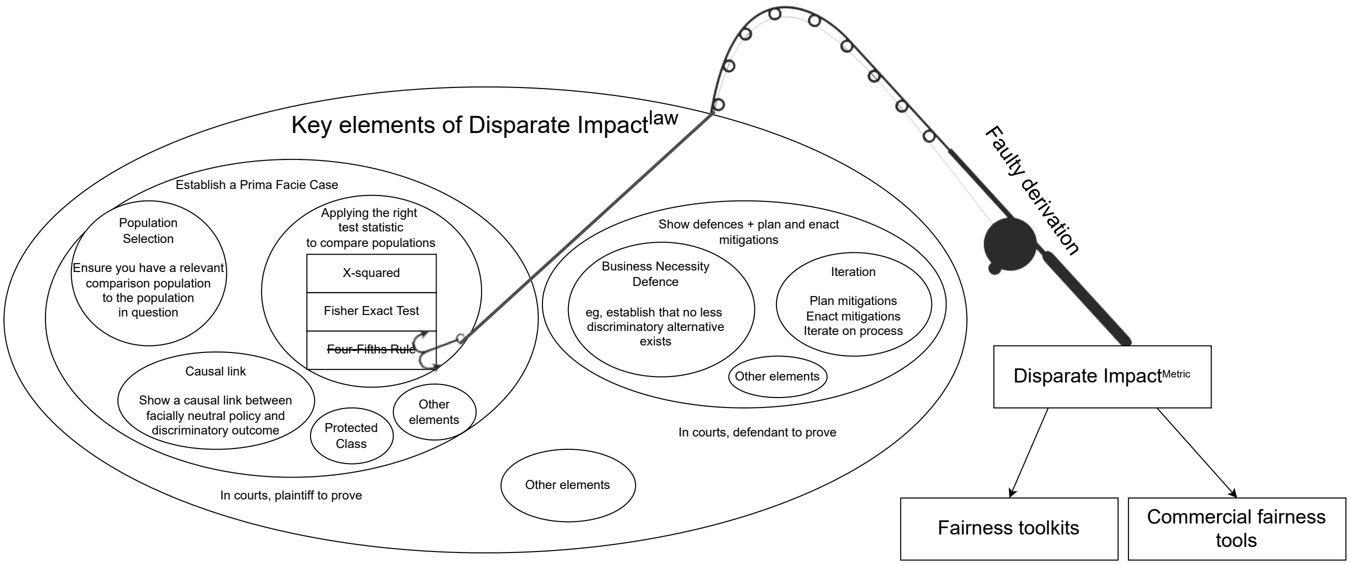
Key elements of Disparate Impact^{law}
Establish a Prima Facie Case
Population Selection
Ensure you have a relevant comparison population to the population in question
Applying the right test statistic to compare populations
X-squared
Fisher Exact Test
~~Four-Fifths Rule~~
Causal link
Show a causal link between facially neutral policy and discriminatory outcome
Protected Class
Other elements
In courts, plaintiff to prove
Other elements
Show defences + plan and enact mitigations
Business Necessity Defence
eg, establish that no less discriminatory alternative exists
Iteration
Plan mitigations
Enact mitigations
Iterate on process
Other elements
In courts, defendant to prove
Faulty derivation
Disparate Impact^{Metric}
Fairness toolkits
Commercial fairness tools

Figure 1: Premature abstraction of the legal term "disparate impact" (DI^{law}), showing synecdoche of the 4/5 rule in the "disparate impact" metric ($\text{DI}^{\text{metric}}$) [22].

statistical tests arise, the court needs to make a call based on the facts of the case, as "'[S]tatistics [...] come in infinite variety [...] their usefulness depends on all of the surrounding facts and circumstances.'"[17] As above, the choice of testing is complex, contingent, and contextual, and cannot be easily abstracted away.

The 4/5 rule is not an appropriate test statistic. In contrast to the tests mentioned above, the 4/5 rule from which $\text{DI}^{\text{metric}}$ is faultily derived is considered less favourably. *It is neither necessary nor sufficient that* (1) *constitutes* $\text{DI}^{\text{finding}}$ *in courts at all* - it is only used in employment contexts by resource-constrained regulators out of court [34, 84]. Courts will simply place greater weight on significance testing than the 4/5 rule, for reasons similar to those which inspired the exceptions in original regulation (Appendix A) - principally, the greater consistency of statistical significance [31].

2.2 Demonstrating a business necessity defence, or arriving at one through mitigation

A prima facie case is not the only step in $\text{DI}^{\text{analysis}}$ and does not automatically lead to a $\text{DI}^{\text{finding}}$. If significant discrepancies are found, legal and compliance teams will look to justify the practice causing the discrepancy using business necessity justifications. Here, the context matters. In employment cases, it is enough to show a "nexus between its hiring requirement and the employment goals" [42]. In a fair lending or machine learning context, regulators may ask for evidence that the model chosen is the least discriminatory of all models which provide sufficient value (generally, profit) [16]. In a disability context, compliance teams may show that reasonable accommodations cannot rectify the alleged disparate impact[58]. If mitigations are unavailable or simply too burdensome, the alleged discriminatory practice need not result in $\text{DI}^{\text{finding}}$ as the above cases show. However, the discovery of mitigations and an assessment of their burden are complex matters, contingent on the facts of the case, and reliant on context.

Iteration. If workable mitigations are found, they must be documented and carried out so that compliance teams can establish a business necessity defence in the future to a regulator or court. For example, if a less discriminatory alternative model is found in the process of demonstrating a business necessity defence, a bank concerned with fair lending is bound to use the less discriminatory alternative [21]. However, the less discriminatory alternative should be reassessed from the beginning, leading to an iterative process which ought to end in a process or model that can be defended in a disparate impact claim, either because no discrimination remains or because the business necessity defence can be made out.

2.3 Brief regulatory history of the 4/5 rule

The 4/5 rule was historically developed and used as an early screening test to decide if further regulatory scrutiny was needed for compliance with Title VII employment discrimination laws. The earliest mention of the 4/5 rule can be traced to regulatory guidance from California in 1972 [19, 20, 72]. We have been unable to find any official, written justification of *why* the precise value of $\tau = 4/5$ was chosen; however, there is anecdotal evidence from the meeting of the authors of [72] that "The 80% Test was born out of two compromises: (1) a desire expressed by those writing and having input into the Guidelines to include a statistical test as the primary step but knowing from an administrative point of view a statistical test was not possible for the FEPC consultants who had to work the enforcement of the Guidelines, and (2) a way to split the middle between two camps, the 70% camp and the 90% camp. A way was found to use both. In the way the 80% Test was defined by TACT, if there was no violation of the 80% Test, then there would be no reason to apply statistical significance tests." [7, Ch. 1, pp. 29–30]. In other words, the 4/5 rule was meant to be a practical rule of thumb that was borne out of a split-the-difference compromise as to the precise value of τ, as a means to reduce the need for rigorous statistical testing, which was considered scarce at the time.

The rule-of-thumb intent of the 4/5 rule has been codified into the federal UGESP of 1978 through a convoluted process of harmonization and clarification [2]. The earliest federal document we have been able to find that mentions the 4/5 rule is a 1974 memo [53] (republished in [41]) which states the 4/5 rule as a trigger for further regulatory scrutiny by the Office of Federal Contract Compliance (OFCC, one of the agencies who co-issued the UGESP; now the Office of Federal Contract Compliance Programs, OFCCP). An interview with Eleanor Holmes Norton in 1978, who was at the time the Chair of the EEOC, highlighted the need to triage the large volume of EEOC investigations, which was at the time estimated to be up to 80,000 per year, and the need to prioritize cases that had potential for the the largest dollar amounts of redress [34]. This context is alluded to in the 93 questions and answers that followed the UGESP [49, 50] (Appendix A.2): "This "4/5ths" or "80%" rule of thumb is not intended as a legal definition, but is a practical means of keeping the attention of the enforcement agencies on serious discrepancies in rates of hiring, promotion and other selection decisions." [49, Q.11].

Today, DI$^{\text{law}}$ has grown beyond the Title VII rights to employment nondiscrimination into other civil rights. However, the use of the 4/5 rule itself has ***not*** been exported concomitantly to these other areas. For example, the Department of Justice (DOJ)'s *Title VI Legal Manual* [78, §7], which contains a detailed discussion of how to prove DI$^{\text{finding}}$ in the context of Title VI civil rights, explicitly disclaims that "not every type of disparity lends itself to the use of the four-fifths rule, even with respect to employment decisions", and furthermore details how individual agencies like the Department of Education (DOE) and the Environmental Protection Agency (EPA) consider different statistical evidence as relevant to Title VI DI$^{\text{finding}}$. Similarly, Regulations B and C, which codify fair lending regulations, do not mention the 4/5 rule in the context of DI$^{\text{finding}}$ in fair lending law [13].

Furthermore, the contemporary importance placed by regulators of the 4/5 rule in its original Title VII context has arguably diminished in importance relative to other statistical tests. The Office of Federal Contract Compliance Programs (OFCCP), whose precursor issued the 1974 memo referenced above, no longer publishes any reference to the 4/5 rule in its current compliance manual [54]. In a recent case [15], a federal court also decided that technical satisfaction of the 4/5 rule was insufficient to defend against an allegation of DI$^{\text{finding}}$, and that other tests of statistical significance sufficed to establish a prima facie case of DI$^{\text{finding}}$ as outlined in Section 2.1. The importance placed on statistical significance testing over the 4/5 rule in the context of DI$^{\text{analysis}}$ is understandable, given the greater knowledge of statistical testing and greater availability of statistical computing resources today, both of which were rare when the UGESP was published in 1978.

The history of the 4/5 rule reinforces our message that DI$^{\text{law}}$ is complex, contingent, and contextual. The 4/5 rule, while part of the regulations around Title VII DI$^{\text{law}}$, was never intended to be a legal definition of DI$^{\text{finding}}$, and that its use in DI$^{\text{analysis}}$ does ***not*** generalize to other contexts of DI$^{\text{law}}$.

2.4 Summary

The legal approaches to disparate impact analysis and mitigation are complex, expensive, and necessary to avoid eight-digit regulatory fines, court judgments carrying similar cost, and reputational damage. Both compliance teams and plaintiffs in court need to make subtle yet consequential choices about reference populations, statistical tests, defences, mitigation strategies, and other considerations, with reference to the particular regulatory scheme and facts of the case. While computer scientists can help with tasks like establishing statistical evidence, there is simply no substitute for legal expertise to establish DI$^{\text{finding}}$ through an appropriate contextual DI$^{\text{analysis}}$. In the large, DI$^{\text{metric}}$ is irrelevant for DI$^{\text{finding}}$. Computer scientists risk epistemic trespassing in overreaching for the limited places where quantitative computations are called for, and by arguing for the synecdoche of DI$^{\text{metric}}$ in place of DI$^{\text{analysis}}$, and for ignoring the precise legal contexts of statistical evidence used in DI$^{\text{law}}$.

3 CRITICAL ANALYSIS OF THE GENERALIZATION OF DISPARATE IMPACT

Having now reviewed the original legal contexts of disparate impact, we present in this section a critical "derivation" of DI$^{\text{metric}}$ [22] from the regulation stating the 4/5 rule (Appendix A). While Definition 1.1 claims to generalize the 4/5 rule [22]; we present in Table 1 a sequence of logical transformations (introduced in Definition 3.1), showing that several premature abstractions and *ad hoc* redefinitions are necessary in this "derivation", which is therefore erroneous. The flawed generalization means that Definition 1.1 *no longer correctly describes the original regulatory use of the 4/5 rule.* To state this claim more precisely, we now introduce some formal logical definitions for the notions of premature abstraction and epistemic trespassing that we have previously introduced.

Definition 3.1. Let Γ be some context in which the statement x is true, written $\Gamma \vdash x$; y be a statement that is more general than x, written $x < y$, by virtue of omission of details; and Γ' be a more general context than Γ, written $\Gamma \prec \Gamma'$. Furthermore, assume that the generality relations $<$ and $\prec$ are transitive. Then, an **inductive generalization** (I) is the logical inference rule

$$\frac{\Gamma \vdash x\Gamma \prec \Gamma' \; x < y}{\Gamma' \vdash y}(\text{I}).$$

A **decontextualization** (D) is an inductive generalization (I) where $\Gamma \prec \Gamma'$ is axiomatically presumed to be true and $x = y$ identically, i.e., only the context is asserted to be generalized and not the statement. An **abstraction** (A) is an inductive generalization (I) where $\Gamma = \Gamma'$ identically and $x < y$ is axiomatically presumed to be true, i.e., only the statement is asserted to be generalized and not the context. A **premature abstraction** (P) is an inductive generalization (I) where both $\Gamma \prec \Gamma'$ and $x < y$ are both axiomatically presumed to be true, i.e., both the statement and context are asserted to be generalized. A **recontextualization** (R) is the logical inference rule

$$\frac{\Gamma' \vdash x \quad \Gamma \prec \Gamma'}{\Gamma \vdash x}\text{(R)}.$$

The terms de-/re-contextualization have been previously used to describe the processes of socio-technical change [30, 70]. The reciprocal relationship turns out to be a specific instance of deontic semantics, which shows up as meaning latent in the values of binary random variables. We now define the following concept:

Definition 3.2. A **deontically-polarized binary (DPB) variable** is a random variable V taking either a positive value or a negative value.

The deontic meaning assigned to a binary variable is relevant when computing metrics of algorithmic bias. First, Definition 1.1 explicitly builds upon the notion of equality of outcomes, which compares base rates for the positive outcome $C = \text{YES}$ only (conditioned on the protected attribute). Replacing $C = \text{YES}$ with $C = \text{NO}$ in Definition 1.1 creates a different metric, $\Pr(C = \text{NO}|X = 0)/\Pr(C = \text{NO}|X = 1)$, which in general will not be the same value as the original ratio $\Pr(C = \text{YES}|X = 0)/\Pr(C = \text{YES}|X = 1)$. Second, the deontic value of being in the majority group $X = 1$ literally defines the denominator of the ratio: the metric is not symmetric with respect to interchanging $X = 1$ and $X = 0$ classes. Third, the deontic polarizations of positive/negative outcomes ($C \in \{\text{YES}, \text{NO}\}$) and majority/minority groups ($X \in \{0, 1\}$) take on moral dimensions of good/bad and inclusion/exclusion that need to be considered when qualitatively assessing ethical harms.

While we do not define precisely the "more general" relationships $<$ or $\prec$, the definitions above nevertheless suffice for the critique summarized in Table 1, which distinguishes between abstractions that are correct generalizations (of the form $x < y$) and those that are not (denoted $x <^{(*)} y$), as well as correct and incorrect logical transformations (the latter are denoted by a suffixed *). For example, the text "A ~~selection~~ positive outcome rate for any race, sex, or ethnic group which is less than four-fifths of the rate for the group with the highest rate defines disparate impact." describes the change necessary to turn the preceding statement x_2 (using the word "selection") into the current statement x_3 (using the phrase "positive outcome"), and similarly for mutating one context Γ_i into another Γ_{i+1}.

The logical flow of Table 1 can be summarized as a process of **epistemic trespassing**, being a fallacious premature abstraction (P*) based on faulty inductive premises $\Gamma_1 \prec^{(*)} \Gamma_4$ and $x_1 <^{(*)} x_5$,[2] followed by a recontextualization (R) into the context of the data set D:

$$\frac{D \prec \Gamma_5 \qquad \dfrac{\Gamma_1 \vdash x_1 \quad \Gamma_1 \prec^{(*)} \Gamma_4 \quad x_1 <^{(*)} x_5}{\Gamma_4 \vdash x_5}\text{(P}^*\text{)}}{D \vdash x_5}\text{(R)}.$$

While the second step is logically valid, the first step involves problematic assertions which invalidate the premises upon which the premature abstraction was defined.

The individual steps reveal the precise logical faults worth detailing, as are the concomitant implicit, yet necessary, widenings of context to enable abstracting away of now-irrelevant details. $\Gamma_1 \prec \Gamma_2$ enables the notion of (enforcement) agency, to be discarded in $x_1 <^{(*)} x_2$, which now claims the 4/5 rule as an operational definition, and glosses over all the legal requirements described in Section 2. The generalization $x_2 < x_3$, while abstracting away the binary decision, needs to retain the deontic polarity presumed in that it is a good thing for people to be employed, which must be preserved in the widening $\Gamma_2 \prec \Gamma_3$ even without the employment context. Feldman et al. [22] have acknowledged such deontic polarity as "awkward". The generalization $x_3 <^{(*)} x_4$ abstracts away "protected attributes", a reference to the legal notion of protected class in DI^{law}. However, the "binary" modifier collapses nuance in the comparisons to be measured and introduces an ethical harm of overly broad categorization, making the widening of the corresponding context $\Gamma_3 \prec^{(*)} \Gamma_4$ problematic. For example, rather than considering each racial group separately relative to some reference racial group, the nuance is flattened into a simple pairwise comparison of out-group performance relative to in-group performance, and lays bare the deontic subtext that belies the comparison. This problem is further exacerbated in the transformation $x_4 <^{(*)} x_5$, which redefines the reference group as the majority group $X = 1$. The only way to view this redefinition as an abstraction is to assume that *in all reference populations, the majority group and most advantaged group are identical.* This change in semantics alters the description of model minorities that are not the majority group ($X = 0$) but are nevertheless the group more likely to have the better outcomes, $\Pr(C = \text{YES}|X = 0) > \Pr(C = \text{YES}|X = 1)$, and *decouples the deontic polarity of the outcome C from the deontic polarity of the in-group membership X*, requiring now the management of two separate sets of deontic semantics. All these semantic changes culminate in the recontextualization going from Γ_4 to an empirical data set D, which requires categorical assignments into the DPB variables X and C in order to be defined. Furthermore, the rows of D explicitly define the reference population to be assessed, raising practical issues around representativeness and sampling bias that must be considered.

3.1 Possibilities for removing deontic polarization

The analysis above demonstrate the composition of multiple abstractions that were necessary to arrive at Definition 1.1. It is also clear that other abstractions of the 4/5 rule are also possible, being analogous to (1.1) as a codification of $D \vdash x_5$, albeit corresponding to statements other than x_5. Thus, it is possible to ameliorate one of the most problematic aspects of $\text{DI}^{\text{metric}}$ by redefining the test to remove the deontic aspect of the protected attribute X. For example, $\Gamma_4 \vdash x_4$ could have been codified for a specific data set, $D \vdash x_4$, as the symmetrized ratio

$$\min\left(\frac{\Pr(C = \text{YES}|X = 0)}{\Pr(C = \text{YES}|X = 1)}, \frac{\Pr(C = \text{YES}|X = 1)}{\Pr(C = \text{YES}|X = 0)}\right) \le \tau = 0.8, \qquad (2)$$

or equivalently,

$$\tau \le \frac{\Pr(C = \text{YES}|X = 0)}{\Pr(C = \text{YES}|X = 1)} \le \frac{1}{1 - \tau}. \qquad (3)$$

This redefinition removes deontic polarization by symmetrization: it no longer matters which group $X = 1$ or $X = 0$ serves as the basis for comparison and goes into the denominator. As we will see later

[2] These statements follow from the transitive relations $\Gamma_1 \prec \Gamma_2 \prec \Gamma_3 \prec^{(*)} \Gamma_4$ and $x_1 <^{(*)} x_2 < x_3 <^{(*)} x_4 <^{(*)} x_5$.

Formal notation	Scope	Text	Comments
$\Gamma_1 \vdash x_1$	Certain federal agencies and employment decisions	A selection rate for any race, sex, or ethnic group which is less than four-fifths of the rate for the group with the highest rate will generally be regarded by the Federal enforcement agencies as evidence of disparate impact.	Abridged from Appendix A
$\Gamma_2 \vdash x_2$	Certain ~~federal agencies and~~ employment decisions	A selection rate for any race, sex, or ethnic group which is less than four-fifths of the rate for the group with the highest rate ~~will generally be regarded by the Federal enforcement agencies as evidence of~~ defines disparate impact.	(P*); $\Gamma_1 \prec \Gamma_2$ discards agency; $x_1 <^{(*)} x_2$ ignores Section 2
$\Gamma_3 \vdash x_3$	~~Certain employment decisions~~ any DPB decision involving race, sex or ethnic groups	A ~~selection~~ positive outcome rate for any race, sex, or ethnic group which is less than four-fifths of the rate for the group with the highest rate defines disparate impact.	(P)
$\Gamma_4 \vdash x_4$	Any DPB decision involving ~~race, sex or ethnic groups~~ groups defined by any DPB protected attribute	A positive outcome rate for any ~~race, sex, or ethnic group~~ binary protected attribute which is less than four-fifths of the rate for the group with the highest rate defines disparate impact.	(P*); $x_3 <^{(*)} x_4$ and $\Gamma_3 \prec^{(*)} \Gamma_4$ introduce harms of categorization
$\Gamma_4 \vdash x_5$	Any DPB decision involving groups defined by any DPB protected attribute	A positive outcome rate for any binary protected attribute which is less than four-fifths of the rate for the ~~group with the highest rate~~ majority group defines disparate impact.	(A*); $x_4 <^{(*)} x_5$ redefines relevant population
$D \vdash x_5$	Data $D = (X, Y, C)$ on some ~~any~~ DPB decision C and some ~~any~~ DPB protected attribute X	A positive outcome rate for any binary protected attribute which is less than four-fifths of the rate for the majority group defines disparate impact.	(R) yields Definition 1.1

Table 1: Necessary abstractions to derive Definition 1.1 from the original U.S. federal regulatory guidance on disparate impact. DPB is short for deontically-polarized binary variable (Definition 3.2). * denotes logically problematic steps. See Section 3 for details.

in Section 4, some practical implementations of $\text{DI}^{\text{metric}}$ do indeed define the 4/5 rule in terms of (3) instead of (1). Nevertheless, the *legal* significance of the symmetrized ratio is unclear, and raises further thorny issues of the legality of "reverse discrimination" [26–28].

An alternative to removing the deontic polarization of X is to attempt a different way to abstract away the "race, sex or ethnic group" of Γ_3 and x_3, which assigns to X the deontic meaning of C rather than giving X its own, separate, deontic semantics, and preserves the ability to reason about non-binary X with multiple categories. For example, consider $\Gamma_{4'} \vdash x_{4'}$:

Any DPB decision involving a categorical protected attribute	A positive outcome rate for any categorical protected attribute which is less than four-fifths of the rate for the group with the highest rate defines disparate impact.

which could have been codified $D \vdash x_{4'}$ into the ratios

$$\rho^{(4')}(x) = \frac{\Pr(C = \text{YES}|X = x)}{\max_{x'} \Pr(C = \text{YES}|X = x')} \le \tau, \text{where } x \neq x', \qquad (4)$$

where the denominator encodes the notion of "group with the highest rate" and not "majority group". A single metric could have been constructed from summary statistics of these ratios; one plausible metric, $R^{(4')}$, is simply to consider the worst case:

$$R^{(4')} = \min_x \rho^{(4')}(x). \qquad (5)$$

In the special case of a binary X (with no deontic polarization needed), (5) reduces to (3); it is therefore accurate to characterize (5) as the correct generalization of (3) to categorical X.

While the above shows that it is possible to remove the deontic polarization necessary in X, the preceding discussion also shows how it is impossible to completely remove deontic polarization from any redefinition of $\text{DI}^{\text{metric}}$, for two reasons. First, the deontic polarization of C is necessary for correctly computing $\text{DI}^{\text{metric}}$. Consider the confusion matrix $\begin{array}{|c|cc|}\hline & C=1 & C=0 \\ \hline X=1 & P_1 & N_1 \\ X=0 & P_0 & N_0 \\ \hline\end{array}$, where P_x is the number of people receiving the positive outcome $C = 1$ that belong to $X = x$, and N_x being the corresponding negative count. A simple computation of $\text{DI}^{\text{metric}}$ yields the ratio $(1 + N_1/P_1)/(1 + N_0/P_0)$. If $C = 0$ were the positive outcome, $\text{DI}^{\text{metric}}$ would instead take the reciprocal value $(1 + N_0/P_0)/(1 + N_1/P_1)$, which is in general different. Such concerns have been acknowledged in the early literature [68, 69], which have noted the possibility for apparently contradictory statistical evidence when measuring differences in selection rates vs. differences in non-selection rates. Second, the assumption of universal positive polarity in $C = 1$ neglects more complex nuances; in the original context of employment, the holistic consideration of the underemployment of women [83], youths [12], and racial minorities [48]; exploitative labor conditions that affect vulnerable workers like those in lower-income countries [71], children [60], and trafficked slaves [57, 82]; and other concerns around people with disabilities [44, 66], social class [62], immigration [55], labor organizing [63], freelancing [75], and corporate social responsibility [18], are all necessary for determining the deontic value of an employment selection. Similar deontic assumptions must be confronted in other contexts, such as granting bail to those who cannot afford it in the context of criminal justice [5].

These possibilities for ameliorating a single problematic aspect of $\mathrm{DI}^{\mathrm{metric}}$, while instructive for understanding how to improve the quantitative definition, nevertheless do not redress all the various stages of premature abstraction that enable the epistemic trespassing of the 4/5 rule. Rather, this discussion lays bare how far the meaning of $\mathrm{DI}^{\mathrm{metric}}$ has deviated from the original meaning of the 4/5 rule in $\mathrm{DI}^{\mathrm{law}}$. Therefore, we argue that giving the name "disparate impact" to $\mathrm{DI}^{\mathrm{metric}}$ is a problematic practice, and exhort the algorithmic fairness community to stop this usage of "disparate impact", to reduce the inevitable epistemic trespassing in conflating $\mathrm{DI}^{\mathrm{metric}}$ with $\mathrm{DI}^{\mathrm{finding}}$. Worryingly, such epistemic trespassing is already manifest in AI fairness toolkits, as we will now discuss.

4 SPREADING THE 4/5 RULE IN FAIRNESS TOOLKITS

Interest in fairness and disparate impact within the computer science discipline has grown greatly since 2015. Perhaps in response to this growing demand for applicable fairness heuristics which can be implemented into statistical models, a new field has emerged of "AI ethics" and "AI fairness" toolkits. Such toolkits are usually open-source code, but commercial offerings do exist. These technical packages operationalize guidelines for "fair" decision-making into tests which end-users can build into their own model-development processes to assess their own models' treatment of disparate groups, or use as-is.

A number of papers have critiqued the presumptions and organizational imperatives of toolkits, in particular how these toolkits prioritize the decision-making of privileged technologists [47], how they frame the work of AI ethics as an individual rather than systematic endeavor [43] and may fail to address practitioner needs [33, 61]. Rather, we focus on toolkits as constructions that collect instruments, processes, and actions in prescriptive ways that make a deliberate representation of expertise [40]. Fairness toolkits perpetuate the epistemic trespassing we have detailed above in Section 3, which lends undue weight to $\mathrm{DI}^{\mathrm{metric}}$ by giving it the same name as $\mathrm{DI}^{\mathrm{law}}$. These toolkits are clearly not built to handle the full complexities of $\mathrm{DI}^{\mathrm{law}}$ as sketched in Figure 1, and since few, if any, users of these toolkits will be aware of the nuances of disparate impact that we differentiate in this paper, offering $\mathrm{DI}^{\mathrm{metric}}$ under a name like "disparate impact" ought to provoke concern about unintended legal claims that are unwarranted from simply computing $\mathrm{DI}^{\mathrm{metric}}$. In particular, we cannot pretend that the act of simply computing $\mathrm{DI}^{\mathrm{metric}}$ constitutes the entirety of "disparate impact analysis", without risking significant confusion with the $\mathrm{DI}^{\mathrm{analysis}}$ that must be conducted in regulatory review or in a court of law. Overall, the inadequacy of toolkits to assure legal protections, combined with their widespread popularity, makes our argument both compelling and urgent.

4.1 Fairness toolkits are popular

In this section, we briefly overview some AI fairness toolkits that present functionality for computing bias metrics, and highlight any references made to "disparate impact" or the 4/5 rule in their documentation. Table 2 shows some crude statistics that indicate the relative popularity of each toolkit. Altogether, these toolkits have been downloaded at least 600,000 times, which ought to raise concerns about the scale of unintended and improper legal claims that may be perpetrated across all sorts of use cases.

Microsoft Fairlearn **(390,000+ downloads)** [8] is the only fairness toolkit we surveyed here which does not use "Disparate Impact" in its naming of $\mathrm{DI}^{\mathrm{metric}}$, and also does not suggest any thresholds (in particular those that align with the 4/5 rule). Furthermore, Fairlearn's documentation acknowledges risks inherent in "portability traps" and the like. We commend the authors of Fairlearn for their care in avoiding epistemic trespassing.[3]

Aequitas **(80,000+ downloads)** [64] relies heavily on the 0.8–1.25 thresholds which characterize (3), and in fact exhibits additional epistemic trespassing by applying these thresholds to metrics other than $\mathrm{DI}^{\mathrm{metric}}$. For instance, the main example for their Bias Report states that "any disparity measure between 0.8 and 1.25 will be deemed fair. (This is inline with the 80 percent rule for determining disparate impact)." The corresponding report claims that meeting the 4/5 rule will ensure a "pass" grade for the audit: "If disparity for a group is within 80 percent and 125 percent of the value of the reference group on a group metric (e.g. False Positive Rate), this audit will pass."

pymetrics Audit-AI **(20,000+ downloads)** has a README explicitly cites EEOC and the 4/5 rule. They then provide a sample model problem describing a ratio of a "lowest-passing" population to the "highest-passing" population, describing a "ratio [that] is greater than .80 (4/5ths), the legal requirement enforced by the EEOC, the model would pass the check for practical significance." While the author takes care to denote that the EEOC guidelines originate in the hiring space, they explicitly generalize the rule to all domains (including outside employment) without warning users that different rules may apply.

IBM AI Fairness 360 (AIF360, **175,000+ downloads)** [6] depicts the 4/5 threshold in their GUI tutorial. In their notebook tutorial on a medical expenditure data set, they note that "$1 - \min(DI, 1/DI) < 0.2$ is typically desired for classifier predictions to be fair", which is equivalent to (3).

Salesforce's Einstein is a proprietary tool including bias safeguarding, which depicts the four-fifths threshold for $\mathrm{DI}^{\mathrm{metric}}$ in their demo under the name "Disparate Impact". Their customer story indicates that Einstein is used in a finance context, where particularly onerous anti-discrimination law applies.

Fairplay AI's Mortgage Fairness Monitor is a proprietary tool which measures mortgage fairness by county. The tool uses $\mathrm{DI}^{\mathrm{metric}}$, termed Adverse Impact Ratio. The thresholds used are <80%, between 80% and 90%, and over 90%. Their target market is finance,

[3] We found an example where a data scientist could not find suggested thresholds in Fairlearn, and so looked to the thresholds in AI Fairness 360, found 80% thresholds and ended up using the 4/5 rule anyway [59].

Name	GitHub stars	PyPI downloads	Paper citations[1]	Name of DI metric or similar	Suggests an 80% threshold
Aequitas	458	83750	106	Impact parity[2]	Yes[3]
AIF360	1635	178736	340	Disparate impact ratio	Yes
Fairlearn	1190	391898	58	Selection rate ratio	No
Audit-AI	273	21159	N/A	4/5 test	Yes
Salesforce Einstein	N/A	N/A	N/A	Disparate impact	Yes
Fairplay Mortgage Fairness Monitor	N/A	N/A	N/A	Adverse impact ratio	Yes
H2O.ai	N/A	N/A	N/A	Adverse impact ratio	Yes

Table 2: Statistics of popularity for several major AI fairness toolkits as of 2022-02-16. Notes: 1. Citation counts taken from Google Scholar. 2. Documentation also refers to $\mathrm{DI}^{\mathrm{metric}}$ as "proportional parity" or "minimizing disparate impact". 3. Documentation recommends the 80% threshold not just for $\mathrm{DI}^{\mathrm{metric}}$, but for multiple similar metrics.

where particularly onerous anti-discrimination law applies, yet contains no mention of the 4/5 rule as being relevant for fair lending $\mathrm{DI}^{\mathrm{finding}}$.

H2O.ai offers a responsible ML workflow paper [25], which acknowledges that "it is not clear that the use of this [80%] threshold is directly relevant to testing fairness for measures other than the AIR." A blog post which describes "Disparate Impact Analysis" or DIA, states that "The regulatory agencies will generally regard a selection rate for any group which is less than four-fifths (4/5) or eighty percent of the rate for the group with the highest selection rate as constituting evidence of adverse impact" immediately following the sentence of describing "discrimination in hiring, housing, etc., or in general any public policy decisions", which can be read as epistemic trespassing in claiming the relevance of the 4/5 rule in all public policy decisions. The same post claims that "Disparate Impact Analysis is one of the tools that is broadly applicable to a wide variety of use cases under the regulatory compliance umbrella, especially around intentional discrimination." The "Disparate Impact Analysis" workflow is not the same as $\mathrm{DI}^{\mathrm{analysis}}$ - for instance, intent is not always relevant to establishing $\mathrm{DI}^{\mathrm{analysis}}$ under $\mathrm{DI}^{\mathrm{law}}$. Other tutorials also explicitly reference the same 0.8–1.25 range of (3) to "be flagged as disparate."

5 RISKS FROM EPISTEMIC TRESPASSING OF THE 4/5 RULE

Since most computer scientists are not lawyers, the obvious risk of the epistemic trespassing of the 4/5 rule is that presenting the 4/5 rule as the entirety of $\mathrm{DI}^{\mathrm{analysis}}$ carries obvious legal risks for users of software developed by otherwise well-intentioned computer scientists. Disregarding the necessary legal context for $\mathrm{DI}^{\mathrm{finding}}$ runs the risk of treating $\mathrm{DI}^{\mathrm{metric}}$ as an instance of Maslow's hammer: "if the only tool you have is a hammer, to treat everything as if it were a nail" [39]. The reality, as discussed above in Section 2, is that computing $\mathrm{DI}^{\mathrm{metric}}$ and demonstrating (1) is ***neither necessary nor sufficient*** to establish $\mathrm{DI}^{\mathrm{finding}}$. Alternative formulations and tests like (3) or (5) do not address this disconnect.

The conflation of the 4/5 rule with $\mathrm{DI}^{\mathrm{law}}$, $\mathrm{DI}^{\mathrm{finding}}$, and $\mathrm{DI}^{\mathrm{metric}}$, also risks instantiating Goodhart's Law: "When a measure becomes a target, it ceases to be a good measure," or, more accurately described as when "optimization causes a collapse of the statistical relationship between a goal which the optimizer intends and the proxy used for that goal"[37]. The instantiation of Goodhart's law is particularly visible within toolkits that place undue emphasis on the 4/5 rule, which encourages epistemic trespassing, and thus creates a self-fulfilling prophecy.

Furthermore, the epistemic trespassing of the 4/5 rule leads to simple but wrong answers of how to actually remediate fairness shortcomings, as algorithmic fairness toolkits and literature pursue satisfaction of $\mathrm{DI}^{\mathrm{metric}}$. The conflation of $\mathrm{DI}^{\mathrm{metrics}}$ with $\mathrm{DI}^{\mathrm{finding}}$ creates a slippery slope of presuming that if (1) constitutes unfairness, then fairness resolution is simply a matter of transforming the data D so that (1) is now falsified. Such a misconception underlies the presentation of the disparate impact remover technique [22] that is again available in many of the toolkits of Section 4. In reality, the proper remediation of $\mathrm{DI}^{\mathrm{finding}}$ involves following regulatory guidelines which are specific to each regulatory and legal context (see [54, §7] and [78, §VIII] for two specific examples), or following the steps of redress determined in court cases. Such remedial actions require a causal analysis in $\mathrm{DI}^{\mathrm{analysis}}$ to understanding why $\mathrm{DI}^{\mathrm{finding}}$ was determined, in order for such remediation to actually be effective.

While we do not yet have data around the effect that toolkits which emphasize the 4/5 rule will have on communities, we do identify that such toolkits can have effect of avoiding the scrutiny and oversight processes designed to protect those very communities both by using the 4/5 rule as the measure of compliance and by downstream debiasing tasks that emphasize $\mathrm{DI}^{\mathrm{metric}}$. Because the 4/5 rule is substantially less onerous than requirements under $\mathrm{DI}^{\mathrm{law}}$, and more easily gamed, passing $\tau \geq 4/5$ off as a certification of compliance presents a potential moral hazard.

A moral hazard is a concept drawn from economics, in which one party in a contract, transaction, or agreement with another party (i.e., a firm and a government agency) provides misleading information about their participation in that agreement, either because they are either incentivized in that direction, because they are protected from the risks that may be produced, or both. We fear that the 4/5 rule, when used as compliance certification, may either incentivize or facilitate firms' engaging in moral hazard, either intentionally or unintentionally gamifying their metrics, not to make the (admittedly cost- and time-intensive) investment to ensure good faith reduction of discrimination.

Finally, though $\mathrm{DI}^{\mathrm{metric}}$ is not $\mathrm{DI}^{\mathrm{law}}$, we note that we see relevance spillover is occurring not just in other U.S. regulatory contexts, but even into non-U.S. jurisdictions [65], especially because of the emphasis that the 4/5 rule is given in toolkits! We want to reiterate that the simple definition of $\mathrm{DI}^{\mathrm{metric}}$, and its corresponding simple resolution through debiasing, should not stand in for $\mathrm{DI}^{\mathrm{analysis}}$ and do not spur the iterative actions (such as identifying less discriminatory models) required by law to protect our communities.

6 LIMITATIONS

Several components of our analysis are subject to rightful critique. For one, the authors of the foundational paper [22] referred to in this work, themselves delimited their abstraction. They explicitly bounded their selection of the four-fifths values to the purpose of "*notational convenience only*" (emphasis theirs). Furthermore, the relatonship between that paper and the proliferation of the four-fifths metric in fairness toolkits is abstract, and causation is a challenge to assess.

7 CONCLUSION AND OUTLOOK

On this 50th anniversity of the 4/5 rule, we document the epistemic trespassing around the 4/5 rule and the phrase "disparate impact" as testament to the deceptive complexity of operationalizating the seemingly obvious and self-evident constitutional right to nondiscrimination in practice. While nobody wants computer systems to discriminate, reaching for a single $\mathrm{DI}^{\mathrm{metric}}$ to encompass the entire body of $\mathrm{DI}^{\mathrm{law}}$ is overly reductive and trivializes important aspects of establishing $\mathrm{DI}^{\mathrm{finding}}$ through $\mathrm{DI}^{\mathrm{analysis}}$. The epistemic trespassing inherent in conflating all of these notions of disparate impact does a disservice to real-world decision-making systems that must operate in regulatory contexts where $\mathrm{DI}^{\mathrm{law}}$ applies, and is unfortunately manifest in multiple, popular software toolkits. The very real potential for causing harm through well-intentioned misuse of these toolkits requires computer scientists to be more self-critical in their zeal for abstraction, and to be willing to revise initial abstractions when ontological errors in their formation are later elucidated. The self-awareness of the limitations of computational thinking via abstractions is essential for working across disciplinary boundaries, particularly with lawyers, who primarily reason by analogy to specific cases and appeals to authority. Such self-criticism will be essential for incrementally improving upon the practice of *ethical* decision-making, around which awareness on processes like checklists, model cards, and datasheets is emerging [24, 35, 45], and for enabling future research toward more effective debiasing.

ACKNOWLEDGMENTS

We thank Aleksander Eskilson for helpful discussions, Rumman Chowdhury for feedback on an earlier draft, and Michael McKenna for the immense contributions.

REFERENCES

[1] Harold Abelson, Gerald Jay Sussman, and Julie Sussman. 1996. *Structure and Intepretation of Computer Programs* (2 ed.). MIT Press, Cambridge, MA. https://mitpress.mit.edu/sites/default/files/sicp/index.html

[2] Ad Hoc Group on Uniform Selection Guidelines. 1981. *A professional and legal analysis of the uniform guidelines on employee selection procedures.* American Society for Personnel Administration, Berea, Ohio.

[3] Alexander v. Edgewood Mgmt. Corp. 2019. Civil Case No. 15-1140 (D.D.C. Jun. 25).

[4] Nathan Ballantyne. 2019. Epistemic trespassing. *Mind* 128, 510 (2019), 367–395. https://doi.org/10.1093/mind/fzx042

[5] Michelle Bao, Angela Zhou, Samantha A Zottola, Brian Brubach, Sarah Desmarais, Aaron Seth Horowitz, Kristian Lum, and Suresh Venkatasubramanian. 2021. It's COMPASlicated: The Messy Relationship between RAI Datasets and Algorithmic Fairness Benchmarks. In *Thirty-fifth Conference on Neural Information Processing Systems Datasets and Benchmarks Track (Round 1) (NeurIPS '21).* OpenReview, OpenReview.net, 18 pages. https://openreview.net/forum?id=qeM58whnpXM

[6] R. K. E. Bellamy, K. Dey, M. Hind, S. C. Hoffman, S. Houde, K. Kannan, P. Lohia, J. Martino, S. Mehta, A. Mojsilović, S. Nagar, K. Natesan Ramamurthy, J. Richards, D. Saha, P. Sattigeri, M. Singh, K. R. Varshney, and Y. Zhang. 2019. AI Fairness 360: An extensible toolkit for detecting and mitigating algorithmic bias. *IBM Journal of Research and Development* 63, 4/5 (2019), 4:1–15. https://doi.org/10.1147/JRD.2019.2942287

[7] Dan Biddle. 2006. *Adverse Impact and Test Validation: A Practitioner's Guide to Valid and Defensible Employment Testing* (2 ed.). Gower, Aldershot. https://doi.org/10.4324/9781315263298

[8] Sarah Bird, Miro Dudík, Richard Edgar, Brandon Horn, Roman Lutz, Vanessa Milan, Mehrnoosh Sameki, Hanna Wallach, and Kathleen Walker. 2020. *Fairlearn: A toolkit for assessing and improving fairness in AI.* Technical Report MSR-TR-2020-32. Microsoft. https://www.microsoft.com/en-us/research/publication/fairlearn-a-toolkit-for-assessing-and-improving-fairness-in-ai/

[9] George E. P. Box. 1976. Science and Statistics. *J. Amer. Statist. Assoc.* 71, 356 (1976), 791–799. https://doi.org/10.1080/01621459.1976.10480949

[10] Felice Cardone. 2021. Games, Full Abstraction and Full Completeness. In *The Stanford Encyclopedia of Philosophy* (Spring 2021 ed.), Edward N. Zalta (Ed.). Metaphysics Research Lab, Stanford University, Stanford, CA.

[11] Giuseppe Castagna. 1997. *Object-Oriented Programming: A Unified Foundation.* Birkhäuser, Boston. https://doi.org/10.1007/978-1-4612-4138-6

[12] Brendan Churchill and Chabel Khan. 2021. Youth underemployment: A review of research on young people and the problems of less(er) employment in an era of mass education. *Sociology Compass* 15, 10 (2021), e12921. https://doi.org/10.1111/soc4.12921

[13] Consumer Financial Protection Bureau. 2020. Supplement I to Part 1002 - Official Interpretations. Comment for 1002.6 - Rules Concerning Evaluation of Applications. Comment 6(a)-2. Effects test. https://www.consumerfinance.gov/rules-policy/regulations/1002/interp-6/

[14] Samuel Deng and Achille Varzi. 2019. Methodological blind spots in machine learning fairness: Lessons from the philosophy of science and computer science. In *Workshop on Human-Centric Machine Learning at the 33rd Conference on Neural Information Processing Systems (NeurIPS 2019).* arXiv, Vancouver, Canada, 5 pages. arXiv:1910.14210

[15] EEOC v. Schuster Co. 2021. No. 13-CV-4063, 2021 U.S. Dist. LEXIS 79815 (N.D. Iowa Apr. 13, 2021).

[16] Elston v. Talladega County Bd. of Educ. 1993. (997 F.2d 1394, 84 Ed. Law Rep. 122). , 1394 pages.

[17] Hazelwood School District et al. v. United States. 1977. (433 U.S. 299).

[18] Michael Etter, Christian Fieseler, and Glen Whelan. 2019. Sharing Economy, Sharing Responsibility? Corporate Social Responsibility in the Digital Age. *Journal of Business Ethics* 159, 4 (2019), 935–942.

[19] Fair Employment Practice Commission. 1974. Report: July 1, 1971 – June 30, 1972. https://digitalcommons.law.ggu.edu/caldocs_agencies/49

[20] Fair Employment Practice Commission. 1975. Report: July 1, 1972 – June 30, 1974. https://digitalcommons.law.ggu.edu/caldocs_agencies/51

[21] FDIC. 2019. Policy statement on discrimination in lending. https://www.fdic.gov/regulations/laws/rules/5000-3860.html

[22] Michael Feldman, Sorelle A. Friedler, John Moeller, Carlos Scheidegger, and Suresh Venkatasubramanian. 2015. Certifying and Removing Disparate Impact. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.* ACM, New York, NY, USA, 259–268. https://doi.org/10.1145/2783258.2783311 arXiv:1412.3756

[23] W. C. Fletcher. 1940. Premature Abstraction. *The Mathematical Gazette* 24, 259 (1940), 73–85. http://www.jstor.org/stable/3606739

[24] Timnit Gebru, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Daumé Iii, and Kate Crawford. 2021. Datasheets for datasets. *Commun. ACM* 64, 12 (2021), 86–92.

[25] Navdeep Gill, Patrick Hall, Kim Montgomery, and Nicholas Schmidt. 2020. A Responsible Machine Learning Workflow with Focus on Interpretable Models, Post-hoc Explanation, and Discrimination Testing. *Information* 11, 3 (29 Feb. 2020), 137. https://doi.org/10.3390/info11030137

[26] Alan H. Goldman. 2015. *Justice and Reverse Discrimination.* Princeton University Press, Princeton, NJ. https://doi.org/10.1515/9781400868605

[27] K. Greenawalt. 1983. *Discrimination and Reverse Discrimination.* Knopf, New York. https://books.google.com/books?id=mTAQAQAAMAAJ

[28] Barry R Gross. 1977. *Reverse discrimination.* Prometheus Books, Buffalo, NY. https://archive.org/details/reversediscrimin00gros

[29] Abigail Z. Jacobs and Hanna Wallach. 2021. Measurement and Fairness. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (Virtual Event, Canada) *(FAccT '21)*. Association for Computing Machinery, New York, NY, USA, 375–385. https://doi.org/10.1145/3442188.3445901

[30] Monique Janneck. 2010. Challenges of Software Recontextualization: Lessons Learned. In *CHI '10 Extended Abstracts on Human Factors in Computing Systems* (Atlanta, Georgia, USA) *(CHI EA '10)*. Association for Computing Machinery, New York, NY, USA, 4613–4628. https://doi.org/10.1145/1753846.1754202

[31] Jones v. City of Bos. 2014. 752 F.3d 38, 51 (1st Cir. 2014).

[32] Jeff Kramer. 2007. Is Abstraction the Key to Computing? *Commun. ACM* 50, 4 (April 2007), 36–42. https://doi.org/10.1145/1232743.1232745

[33] Michelle Seng Ah Lee and Jat Singh. 2021. The Landscape and Gaps in Open Source Fairness Toolkits. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 699, 13 pages. https://doi.org/10.1145/3411764.3445261

[34] Paul London. 1978. A Conversation with Eleanor Holmes Norton. *Employee Relations Law Journal* 3 (1978), 314–326.

[35] Michael A. Madaio, Luke Stark, Jennifer Wortman Vaughan, and Hanna Wallach. 2020. *Co-Designing Checklists to Understand Organizational Challenges and Opportunities around Fairness in AI.* Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3313831.3376445

[36] Mandala v. NTT Data, Inc. 2020. 975 F.3d 202 (2d Cir. 2020). https://casetext.com/case/mandala-v-ntt-data-inc-1 "a number of courts have denied motions to dismiss disparate impact claims using general population statistics to challenge [criminal history checks]".

[37] David Manheim and Scott Garrabrant. 2018. Categorizing variants of Goodhart's Law. arXiv:1803.04585

[38] Donald Martin Jr, Vinodkumar Prabhakaran, Jill Kuhlberg, Andrew Smart, and William S Isaac. 2020. Extending the machine learning abstraction boundary: A Complex systems approach to incorporate societal context. arXiv:2006.09663

[39] Abraham Harold Maslow. 1966. *The Psychology of Science: A Reconnaissance.* John Dewey Society Lectureship Series, Vol. 8. Harper & Row, New York, 15.

[40] Shannon Mattern. 2021. Unboxing the Toolkit. https://tool-shed.org/unboxing-the-toolkit

[41] Kevin S. McGuiness. 1976. *No. SC-2, Government memoranda on affirmative action programs: a study of compliance agency documents affecting non-construction federal contractors.* Vol. 1. Equal Employment Advisory Council, Washington, DC.

[42] Meditz v. City of Newark. 2011. (658 F. 3d 364).

[43] Nick Merrill, Michael Madaio, and Richmond Wong. 2022. Seeking Like a Toolkit: How Toolkits Envision the Work of AI Ethics. arXiv:2202.08792

[44] Boris Miethlich and Anett G. Oldenburg. 2019. Social Inclusion Drives Business Sales: A Literature Review on the Case of the Employment of Persons With Disabilities. In *33nd International Business Information Management Association Conference (IBIMA), Education Excellence and Innovation Management through Vision 2020, Granada, Spain, 10-11.04.2019.* IBIMA Publishing, King of Prussia, PA; King of Prussia, PA, 6253–6267. https://doi.org/10.33543/16002.62536267

[45] Margaret Mitchell, Simone Wu, Andrew Zaldivar, Parker Barnes, Lucy Vasserman, Ben Hutchinson, Elena Spitzer, Inioluwa Deborah Raji, and Timnit Gebru. 2019. Model Cards for Model Reporting. In *Proceedings of the Conference on Fairness, Accountability, and Transparency* (Atlanta, GA, USA) *(FAT* '19)*. Association for Computing Machinery, New York, NY, USA, 220–229. https://doi.org/10.1145/3287560.3287596

[46] Shira Mitchell, Eric Potash, Solon Barocas, Alexander D'Amour, and Kristian Lum. 2021. Algorithmic fairness: Choices, assumptions, and definitions. *Annual Review of Statistics and Its Application* 8 (2021), 141–163.

[47] Gina Neff. 2020. *From Bad Users and Failed Uses to Responsible Technologies: A Call to Expand the AI Ethics Toolkit.* Association for Computing Machinery, New York, NY, USA, 5–6. https://doi.org/10.1145/3375627.3377141

[48] Jennifer L. Nelson and Steven P. Vallas. 2021. Race and inequality at work: An occupational perspective. *Sociology Compass* 15, 10 (2021), e12926. https://doi.org/10.1111/soc4.12926

[49] Eleanor Holmes Norton, Alan K. Campbell, Drew S. Days, Welden Rougeau, and Kent A. Peterson. 1979. Adoption of Questions and Answers To Clarify and Provide a Common Interpretation of the Uniform Guidelines on Employee Selection Procedures. *Federal Register* 44, 43 (2 March 1979), 11996–12009. https://www.loc.gov/item/fr044043

[50] Eleanor Holmes Norton, Alan K. Campbell, Drew S. Days, Weldon J. Rougeau, and Kent A. Peterson. 1980. Adoption of Additional Questions and Answers To Clarify and Provide a Common Interpretation of the Uniform Guidelines on Employee Selection Procedures. *Federal Register* 45, 87 (2 May 1980), 29530–1. https://www.loc.gov/item/fr045087

[51] Eleanor Holmes Norton, Drew S. Days, and Jule M. Sugarman. 1977. Uniform Guidelines on Employee Selection Procedures: Notice of Proposed Rulemaking. *Federal Register* 42, 251 (30 Dec. 1977), 65542–65552. https://www.loc.gov/item/fr042251

[52] Eleanor Holmes Norton, Richard J. Devine, and Drew S. Days. 1978. Proposed Uniform Guidelines on Employee Selection Procedures: Issues of Particular Interest for Public Hearing and Meeting. *Federal Register* 43, 55 (21 March 1978), 11812–3. https://www.loc.gov/item/fr043055

[53] Office of Federal Contract Compliance. 1974. Memorandum No. 8: Testing and Selection Order Guidance.

[54] Office of Federal Contract Compliance Programs, U.S. Department of Labor. 2021. Federal Contract Compliance Manual. https://www.dol.gov/agencies/ofccp/manual/fccm

[55] Liesbet Okkerse. 2008. How to measure labour market effects of immigration: a review. *Journal of Economic Surveys* 22, 1 (2008), 1–30. https://doi.org/10.1111/j.1467-6419.2007.00533.x

[56] Chris Olah and Shan Carter. 2017. Research Debt. https://doi.org/10.23915/distill.00005

[57] Orlando Patterson and Xiaolin Zhuo. 2018. Modern Trafficking, Slavery, and Other Forms of Servitude. *Annual Review of Sociology* 44, 1 (2018), 407–439. https://doi.org/10.1146/annurev-soc-073117-041147

[58] Payan v. Los Angeles Community College Dist. 2021. (11 F. 4th 729). , 729 pages.

[59] prog.world. 2022. Fair modeling with Fairlearn. https://prog.world/fair-modeling-with-fairlearn/ accessed 2022-02-16.

[60] Amir Radfar, Seyed Ahmad Ahmadi Asgharzadeh, Fernando Quesada, and Irina Filip. 2018. Challenges and perspectives of child labor. *Industrial psychiatry journal* 27, 1 (Jan-Jun 2018), 17–20.

[61] Brianna Richardson, Jean Garcia-Gathright, Samuel F. Way, Jennifer Thom, and Henriette Cramer. 2021. Towards Fairness in Practice: A Practitioner-Oriented Rubric for Evaluating Fair ML Toolkits. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 236, 13 pages. https://doi.org/10.1145/3411764.3445604

[62] John E. Roemer. 1982. *A General Theory of Exploitation and Class.* Harvard University Press, Cambridge, MA. https://doi.org/10.4159/harvard.9780674435865

[63] Benjamin I. Sachs. 2007-2008. Employment Law as Labor Law. *Cardozo Law Review* 29 (2007-2008), 2685–2748. https://heinonline.org/HOL/P?h=hein.journals/cdozo29&i=2707

[64] Pedro Saleiro, Benedict Kuester, Loren Hinkson, Jesse London, Abby Stevens, Ari Anisfeld, Kit T. Rodolfa, and Rayid Ghani. 2019. Aequitas: A Bias and Fairness Audit Toolkit. arXiv:1811.05577 [cs.LG]

[65] Javier Sánchez-Monedero, Lina Dencik, and Lilian Edwards. 2020. What Does It Mean to 'solve' the Problem of Discrimination in Hiring? Social, Technical and Legal Perspectives from the UK on Automated Hiring Systems. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (Barcelona, Spain) *(FAT* '20)*. Association for Computing Machinery, New York, NY, USA, 458–468. https://doi.org/10.1145/3351095.3372849

[66] Angela L. Sauer, Andra Parks, and Patricia C. Heyn. 2010. Assistive technology effects on the employment outcomes for people with cognitive disabilities: A systematic review. *Disability and Rehabilitation: Assistive Technology* 5, 6 (2010), 377–391. https://doi.org/10.3109/17483101003746360

[67] Andrew D. Selbst, Danah Boyd, Sorelle A. Friedler, Suresh Venkatasubramanian, and Janet Vertesi. 2019. Fairness and Abstraction in Sociotechnical Systems. In *Proceedings of the Conference on Fairness, Accountability, and Transparency* (Atlanta, GA, USA) *(FAT* '19)*. Association for Computing Machinery, New York, NY, USA, 59–68. https://doi.org/10.1145/3287560.3287598

[68] Elaine W. Shoben. 1977. Probing the Discriminatory Effects of Employee Selection Procedures with Disparate Impact Analysis Under Title VII. *Texas Law Review* 56, 1 (Dec. 1977), 1–45. https://scholars.law.unlv.edu/facpub/575

[69] Elaine W. Shoben. 1978. Differential Pass-Fail Rates in Employment Testing: Statistical Proof under Title VII. *Harvard Law Review* 91, 4 (1978), 793–813. http://www.jstor.org/stable/1340356

[70] Edouard J. Simon, Monique Janneck, and Dorina Gumm. 2006. Understanding Socio-Technical Change: Towards a Multidisciplinary Approach. In *Social Informatics: An Information Society for all? In Remembrance of Rob Kling*, Jacques Berleur, Markku I. Nurminen, and John Impagliazzo (Eds.). Springer US, Boston, MA, 469–479.

[71] Jeremy Snyder. 2010. Exploitation and Sweatshop Labor: Perspectives and Issues. *Business Ethics Quarterly* 20, 2 (2010), 187–213. http://www.jstor.org/stable/25702393

[72] State of California Fair Employment Practice Commission. 1972. Guidelines on employee selection procedures.

[73] Stevenson v. City & County of San Francisco. 2016. C-11-4950 MMC (N.D. Cal. Jan. 5, 2016). "The primary reason for the experts' divergent opinions is that the experts employed different testing methods. Dr. Haan used the 'Fisher's Exact' test, the results of which, he states, show no significant statistical disparity (see Rolnick Decl. Ex. 18 at 14), while Dr. Gutman used the 'Chi Square' test, the results of which, he states, do show a significant statistical disparity (see Randle

Decl. Ex. 4 at 14)".

[74] Christopher Strachey. 2000. Fundamental Concepts in Programming Languages. *Higher-Order and Symbolic Computation* 13, 1/2 (2000), 11–49. https://doi.org/10.1023/A:1010000313106

[75] Will Sutherland, Mohammad Hossein Jarrahi, Michael Dunn, and Sarah Beth Nelson. 2020. Work Precarity and Gig Literacies in Online Freelancing. *Work, Employment and Society* 34, 3 (2020), 457–475. https://doi.org/10.1177/0950017019886511

[76] Texas Dept. of Housing and Community Affairs v. Inclusive Communities Project, Inc. 2015. (576 U.S. 519). 'A robust causality requirement ensures that "[r]acial imbalance [...] does not, without more, establish a prima facie case of disparate impact" and thus protects defendants from being held liable for racial disparities they did not create.'.

[77] Harold R. Tyler, Michael H. Moskov, Ethel Bent Walsh, Robert E. Hampton, Arthur E. Flemming, Richard Albrecht, Eleanor Holmes Norton, Alan K. Campell, Ray Marshall, and Griffin B. Bell. 1971. Adoption of Employee Selection Procedures. Supplementary Information: An overview of the 1978 Uniform Guidelines on Employee Selection Procedures. *Federal Register* 43, 166 (2 Oct. 1971), 38290–38295. https://www.loc.gov/item/fr043166

[78] U.S. Department of Justice, Civil Rights Division, Federal Coordination and Compliance Section. 2021. Title VI legal manual. https://www.justice.gov/crt/fcs/T6Manualhttps://www.justice.gov/crt/book/file/1364106/download

[79] U.S. Government Publishing Office. 1978. 41B C.F.R. 60 Pt. 60-3(D). Adverse impact and the "four-fifths rule.". https://www.ecfr.gov/current/title-41/subtitle-B/chapter-60/part-60-3

[80] Villafana v. Cnty. of San Diego. 2020. 57 Cal.App.5th 1012, 271 Cal. Rptr. 3d 639 (Cal. Ct. App. 2020). https://casetext.com/case/villafana-v-cnty-of-san-diego

[81] Philip Wadler. 2015. Propositions as Types. *Commun. ACM* 58, 12 (nov 2015), 75–84. https://doi.org/10.1145/2699407

[82] Ronald Weitzer. 2015. Human Trafficking and Contemporary Slavery. *Annual Review of Sociology* 41, 1 (2015), 223–242. https://doi.org/10.1146/annurev-soc-073014-112506

[83] Johanna Weststar. 2011. *A Review of Women's Experiences of Three Dimensions of Underemployment.* Springer New York, New York, NY, 105–125. https://doi.org/10.1007/978-1-4419-9413-4_6

[84] Tina Williams. 2020. Nondiscrimination Obligations of Federal Contractors and Subcontractors: Procedures To Resolve Potential Employment Discrimination. *Federal Register* 85, 218 (10 Nov. 2020), 71553–75. https://www.govinfo.gov/content/pkg/FR-2020-11-10/pdf/2020-24858.pdfhttps://www.federalregister.gov/documents/2020/11/10/2020-24858/rin-1250-aa10

[85] Jeannette M. Wing. 2006. Computational Thinking. *Commun. ACM* 49, 3 (March 2006), 33–35. https://doi.org/10.1145/1118178.1118215

[86] Alice Xiang and Inioluwa Deborah Raji. 2019. On the legal compatibility of fairness definitions. In *Workshop on Human-Centric Machine Learning at the 33rd Conference on Neural Information Processing Systems (NeurIPS '19).* arXiv, Vancouver, Canada, 6 pages. arXiv:1912.00761

A THE 4/5 RULE IN THE UGESP

For ease of reference, we quote verbatim the entire paragraph from the U.S. Code of Federal Regulations that describes the 4/5 rule. This paragraph forms part of the Uniform Guidelines on Employee Selection Procedures (29 CFR §1607). We ignore the minor legal nuance that distinguishes disparate impact ($\mathrm{DI}^{\mathrm{finding}}$) from adverse impact, and treat them synonymously.

> **29 CFR §1607.4(D) Adverse impact and the "four-fifths rule"**. . A selection rate for any race, sex, or ethnic group which is less than four-fifths (4/5) (or eighty percent) of the rate for the group with the highest rate will generally be regarded by the Federal enforcement agencies as evidence of adverse impact, while a greater than four-fifths rate will generally not be regarded by Federal enforcement agencies as evidence of adverse impact. Smaller differences in selection rate may nevertheless constitute adverse impact, where they are significant in both statistical and practical terms or where a user's actions have discouraged applicants disproportionately on grounds of race, sex, or ethnic group. Greater differences in selection rate may not constitute adverse impact where the differences are based on small numbers and are not statistically significant, or where special recruiting or other programs cause the pool of minority or female candidates to be atypical of the normal pool of applicants from that group. Where the user's evidence concerning the impact of a selection procedure indicates adverse impact but is based upon numbers which are too small to be reliable, evidence concerning the impact of the procedure over a longer period of time and/or evidence concerning the impact which the selection procedure had when used in the same manner in similar circumstances elsewhere may be considered in determining adverse impact. Where the user has not maintained data on adverse impact as required by the documentation section of applicable guidelines, the Federal enforcement agencies may draw an inference of adverse impact of the selection process from the failure of the user to maintain such data, if the user has an underutilization of a group in the job category, as compared to the group's representation in the relevant labor market or, in the case of jobs filled from within, the applicable work force.

A.1 Legal scope

The legal scope of this paragraph is defined in an earlier section, which we also quote verbatim for ease of reference and to illustrate the full complexity of the legal scope in which the 4/5 rule is defined.

> **29 CFR §1607.2 Scope**.
> **A. Application of guidelines**. These guidelines will be applied by the Equal Employment Opportunity Commission in the enforcement of title VII of the Civil Rights Act of 1964, as amended by the Equal Employment Opportunity Act of 1972 (hereinafter "title VII"); by the Department of Labor, and the contract compliance agencies until the transfer of authority contemplated by the President's Reorganization Plan No. 1 of 1978, in the administration and enforcement of Executive Order 11246, as amended by Executive Order 11375 (hereinafter "Executive Order 11246"); by the Civil Service Commission and other Federal agencies subject to section 717 of title VII; by the Civil Service Commission in exercising its responsibilities toward State and local governments under section 208(b)(1) of the Intergovernmental-Personnel Act; by the Department of Justice in exercising its responsibilities under Federal law; by the Office of Revenue Sharing of the Department of the Treasury under the State and Local Fiscal Assistance Act of 1972, as amended; and by any other Federal agency which adopts them.
> **B. Employment decisions**. These guidelines apply to tests and other selection procedures which are used as a basis for any employment decision. Employment decisions include but are not limited to hiring, promotion, demotion, membership (for example, in a labor organization), referral, retention, and licensing and

> certification, to the extent that licensing and certification may be covered by Federal equal employment opportunity law. Other selection decisions, such as selection for training or transfer, may also be considered employment decisions if they lead to any of the decisions listed above.
>
> **C. Selection procedures**. These guidelines apply only to selection procedures which are used as a basis for making employment decisions. For example, the use of recruiting procedures designed to attract members of a particular race, sex, or ethnic group, which were previously denied employment opportunities or which are currently underutilized, may be necessary to bring an employer into compliance with Federal law, and is frequently an essential element of any effective affirmative action program; but recruitment practices are not considered by these guidelines to be selection procedures. Similarly, these guidelines do not pertain to the question of the lawfulness of a seniority system within the meaning of section 703(h), Executive Order 11246 or other provisions of Federal law or regulation, except to the extent that such systems utilize selection procedures to determine qualifications or abilities to perform the job. Nothing in these guidelines is intended or should be interpreted as discouraging the use of a selection procedure for the purpose of determining qualifications or for the purpose of selection on the basis of relative qualifications, if the selection procedure had been validated in accord with these guidelines for each such purpose for which it is to be used.
>
> **D. Limitations**. These guidelines apply only to persons subject to title VII, Executive Order 11246, or other equal employment opportunity requirements of Federal law. These guidelines do not apply to responsibilities under the Age Discrimination in Employment Act of 1967, as amended, not to discriminate on the basis of age, or under sections 501, 503, and 504 of the Rehabilitation Act of 1973, not to discriminate on the basis of disability.
>
> **E. Indian preference not affected**. These guidelines do not restrict any obligation imposed or right granted by Federal law to users to extend a preference in employment to Indians living on or near an Indian reservation in connection with employment opportunities on or near an Indian reservation.

A.2 Clarifying questions and answers

Finally, we include two relevant question and answer pairs that were published as follow-up to the initial publication of the UGESP [49].

> 11. Q: What is a substantially different rate of selection?
>
> A: The agencies have adopted a rule of thumb under which they will generally consider a selection rate for any race, sex, or ethnic group which is less than four-fifths (4/5ths) or eighty percent (80%) of the selection rate for the group with the highest selection rate as a substantially different rate of selection. See Section 4D. This "4/5ths" or "80%" rule of thumb is not intended as a legal definition, but is a practical means of keeping the attention of the enforcement agencies on serious discrepancies in rates of hiring, promotion and other selection decisions.
>
> For example, if the hiring rate for whites other than Hispanics is 60%, for American Indians 45%, for Hispanics 48%, and for Blacks 51%, and each of these groups constitutes more than 2% of the labor force in the relevant labor area (see Question 16), a comparison should be made of the selection rate for each group with that of the highest group (whites). These comparisons show the following impact ratios: American Indians 45/60 or 75%; Hispanics 48/60 or 80%; and Blacks 51/60 or 85%. Applying the 4/5ths or 80% rule of thumb, on the basis of the above information alone, adverse impact is indicated for American Indians but not for Hispanics or Blacks.
>
> 12. Q: How is adverse impact determined?
>
> A: Adverse impact is determined by a four step process.
>
> (1) calculate the rate of selection for each group (divide the number of persons selected from a group by the number of applicants from that group).
>
> (2) observe which group has the highest selection rate.
>
> (3) calculate the impact ratios, by comparing the selection ratefor each group with that of the highest group (divide the selection rate for a group by the selection rate for the highest group).
>
> (4) observe whether the selection rate for any group is substantially less (i.e., usually less than 4/5ths or 80%) than the selection rate for the highest group. If it is adverse impact is indicated in most circumstances. See Section 4D.
>
> For example:

Applicants	Hired	Selection Rate Percent Hired
80 White	48	48/80 or 60%
40 Black	12	12/40 or 30%

> A comparison of the black selection rate (30%) with the white selection rate (60%) shows that the black rate is 30/60, or one-half (or 50%) of the white rate. Since the one-half (50%) is less than 4/5ths (80%) adverse impact is usually indicated.
>
> The determination of adverse impact is not purely arithmetic however; and other factors may be relevant. See, Section 4D.