UC Merced
Proceedings of the Annual Meeting of the Cognitive Science Society

Title
The Hair Club for Boys: How children and adults judge disparate impact rules

Permalink
https://escholarship.org/uc/item/3c22d3gn

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 46(0)

Authors
Popat, Aarthi
Amemiya, Jamie
Heyman, Gail D.
et al.

Publication Date
2024

Copyright Information
This work is made available under the terms of a Creative Commons Attribution License, available at https://creativecommons.org/licenses/by/4.0/

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

The Hair Club for Boys: How children and adults judge disparate impact rules

Aarthi K. Popat (aarthi.popat@yale.edu)
Department of Psychology, 100 College Street
New Haven, CT 06510

Jamie Amemiya (amemiya@oxy.edu)
Department of Psychology, 1600 Campus Road
Los Angeles, California 90041

Gail D. Heyman (gheyman@ucsd.edu)
Department of Psychology, 2941 Muir Ln
La Jolla, CA 92093

Caren M. Walker (cmwalker@ucsd.edu)
Department of Psychology, 2941 Muir Ln
La Jolla, CA 92093

Abstract

Disparate impact rules are formally neutral but indirectly discriminate against protected groups (i.e., by targeting a characteristic that is more prevalent in a given group). Because these rules are not obviously malicious, they have been widely enacted to circumvent policies against explicit discrimination. In a series of four experiments, we show that adults and children are sensitive to the moral implications of disparate impact rules. However, we also find that they are more accepting of these rules when strong justification is provided, compared to rules with no justification. Crucially, demographic differences also impact people's judgments of disparate impact rules and their creators. We find that conservatives and those from groups not directly affected by the rule tend to be more accepting of it. By studying people's reasoning about disparate impact rules, this work aims to identify the mechanisms by which these rules may evade detection. Finally, we discuss how these insights may inform the development of interventions that highlight the problematic effects of indirectly discriminatory policies.

Keywords: exclusion; disparate impact; explanation; gender

Introduction

The Trump Administration imposed a travel ban in 2017 prohibiting people from seven majority-Muslim countries and two non-Muslim countries from entering the United States. As justification, the administration cited national security concerns and diplomacy—it did not mention religion (Gerstein & Lin, 2018). This policy, often referred to as the "Muslim ban", is one example of a disparate impact rule: It is formally neutral, but indirectly discriminates against a protected group by focusing on a characteristic (i.e., nationality) that is associated with that group (e.g., people of Muslim faith). Policies like this face public and legal pushback (Seicshnaydre, 2016; Ellis, 2017)—particularly when they implicate historically marginalized groups—but they persist in local and national governments (Ceballos et al., 2021; Serwer, 2019). The present research examines how children and adults evaluate disparate impact rules, and what factors influence their evaluations.

Humans begin to engage in moral reasoning early in life. Even infants and young toddlers are surprised by unfair resource allocation (Geraci & Surian, 2011; Sloane et al., 2012; Sommerville et al., 2013), and by the age of three, children start to endorse equality as a norm (LoBue et al., 2011; McAuliffe et al., 2017). This penchant for fairness evolves as children grow: Four-year-olds reject unequal dividends out of spite for advantaged peers (McAuliffe et al., 2014), and six-year-olds prefer fair distributors to generous ones (Shaw et al., 2012). Older children can make more nuanced judgments about fairness. In particular, they can understand and reject instances of *discrimination*. Preschoolers can understand that discrimination is unacceptable (Theimer et al., 2001; Killen et al., 2002), and by age seven, children negatively evaluate gender- and race-based social exclusion (see e.g., Killen & Stangor, 2001a). However, it is only at age eight that children begin to appropriately identify more *implicit* discriminatory behavior with respect to race and gender (Scott et al., 2003; Brown & Bigler, 2004).

Disparate impact rules are a form of implicit discrimination because they do not focus directly on group membership, targeting associated characteristics instead. These rules can therefore act as loopholes to evade allegations of overt discrimination. Existing work shows that adults evaluate loophole behavior (e.g., deviant actions that are consistent with a possible—but unintended—interpretation of a rule) less positively than compliant behavior but more positively than overt defiance (Bridgers et al., 2021). Children as young as eight years old can reliably generate loophole behavior (e.g., when told to stop jumping on the couch, switching to the other couch; Murthy et al., 2023). Given that the capacity to recognize loophole behavior as well as more implicit forms of gender- and race-based discrimination develops in middle childhood, it is possible that children's moral judgments of disparate impact rules will differ with age.

Because they are indirect, disparate impact rules raise competing hypotheses for their true intention: Either the rule has inherent instrumental value, or it is actually intended to exclude a social group. Reasoners must weigh both hypotheses when developing moral judgments about a disparate impact rule. It is likely that school-aged children can flexibly consider and evaluate both possibilities, given that even very young children can consider multiple hypotheses in causal reasoning contexts (Cesana-Arlotti et al., 2022; Alderete & Xu, 2023; Goddu et al., 2021). We propose that two factors might shape how children and adults weigh one hypothesis over the other: (1) whether compelling justification for the rule is provided, and (2) whether they observe exceptions that refute the "exclusion" hypothesis.

First, justifications are known to shape moral reasoning about discrimination in both children and adults. For exam-

In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society.* ©2024 The Author(s). This work is licensed under a Creative Commons Attribution 4.0 International License (CC BY).

ple, children—like adults—typically accept inequality more readily if it is explained by merit-based differences (Starmans et al., 2017). Justification can even lead older children to condone exclusion more readily (Killen & Stangor, 2001b). However, with age, children accept such justification selectively: Eight-year-old children accept compelling explanations, and reject idiosyncratic explanations, for inequality (Schmidt et al., 2016).

We also consider the role of exceptions in licensing hypotheses about the true intention of a disparate impact rule. Learners may favor the "instrumental value" hypothesis if they observe exceptions that provide evidence against intentional exclusion (akin to tokenism; Lee, 2020). In the case of the "Muslim ban", some reasoners may point to the fact that two non-majority Muslim countries were also implicated, and thus believe the argument that the rule was intended to increase national security. Prior research shows that even young children's hypotheses are sensitive to the presence of exceptions: For example, they treat uniform samples differently from those that include one unique item (Hochmann et al., 2017; Lapidow et al., 2022; Lapidow et al., 2021).

Finally, we consider that individual social factors may also impact people's evaluations of disparate impact rules, even in context of justification and exceptions. For example, children are known to be more tolerant of discrimination when it does not impact their own in-group—e.g., five-year-old girls are more likely to identify discriminatory behavior against other girls (vs. against boys; Brown & Bigler, 2004). Moreover, conservative (vs. liberal) adults are more tolerant of overtly discriminatory speech targeting race and gender (Chong et al., 2022). We therefore explore whether those not affected by a disparate impact rule and more conservative adults are less skeptical of such rules. The current research examines how these factors, along with justification and exceptions, impact judgments of disparate impact rules.

Experiment 1

Exp. 1 was an exploratory study conducted with adults to determine whether hearing justification, observing exceptional cases, holding the same identity as the targeted group, and/or political affiliation increase skepticism about a gender-based disparate impact rule in the context of a novel developmental paradigm. We chose a rule that implicates gender because gender is a highly salient social category in children's lives (Rhodes & Mandalaywala, 2017).

Method

Participants. We recruited 184 undergraduates at a large public West Coast university (122 women, 58 men, 3 non-binary, 1 did not report; 83 Asian/Pacific Islander, 42 Latine/x, 26 Multiracial, 20 White, 6 Middle Eastern or North African, 3 Black, 1 Native American or Alaskan Native, 3 did not report). Undergraduates participated in exchange for course credit. 13 identified as extremely liberal, 65 as liberal, 29 as slightly liberal, 27 as moderate, 2 as slightly conservative, 6 as conservative, 7 as extremely conservative, and 35 did not identify with a political ideology. Sixteen additional participants were excluded for failing our attention check.

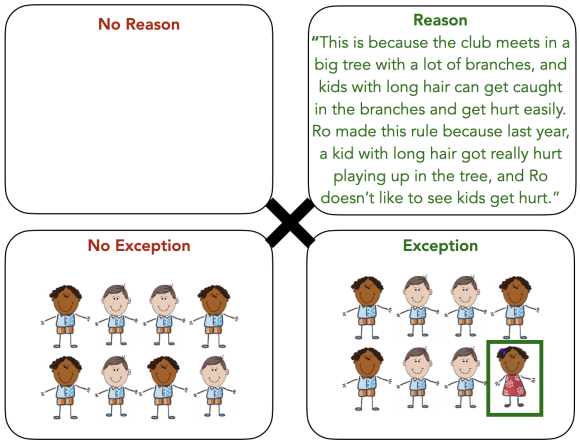

Figure 1: Participants in Exps. 1 and 2 were assigned to one of four conditions crossing reason and exception.

Materials. In a Qualtrics survey, participants read a vignette illustrated by colorful pictures.

Procedure. Participants first learned about Sunny School, a fictional school on a faraway planet. Then, they learned about the students at Sunny School: boys, girls, and a special alien student named Ro. We told participants that boys had to wear blue shirts and girls had to wear pink dresses to ensure that they could distinguish girls from boys in this paradigm. Participants heard that there is an exclusive club at Sunny School, and Ro gets to decide who is and isn't allowed to be in the club. Ro makes a rule that indirectly excludes gender-stereotypical girls (Hill & Tenenbaum, 2022): Only children with short hair are allowed to be in the club. Our 2x2 paradigm crossed justification with the presence of an exception (i.e., inclusion of a short-haired girl in the club; see Figure 1). After learning about Ro's rule, participants either heard a compelling justification for the rule (reason condition) or no justification (no-reason condition). Then, participants saw a picture of the members of the club. In the no-exception condition, participants saw eight boys in the club. In the exception condition, participants saw seven boys and one short-haired girl (exception) in the club. Participants were therefore assigned to one of four conditions: reason/exception, reason/no-exception, no-reason/exception, no-reason/no-exception (see Figure 1).

Dependent measures. Participants responded to three main questions assessing their evaluation of the rule. First, participants judged the rule's morality ("Is it okay or not okay that this rule exists?"; "How [okay/not okay] is it?" with response options "a little [not] okay", "[not] okay", or "really [not] okay"). Responses were coded from 1 = really not okay to 6 = really okay. Second, participants heard a conversation where two other alien characters briefly discussed the "exclusion" and "instrumental value" hypotheses for Ro's rule,

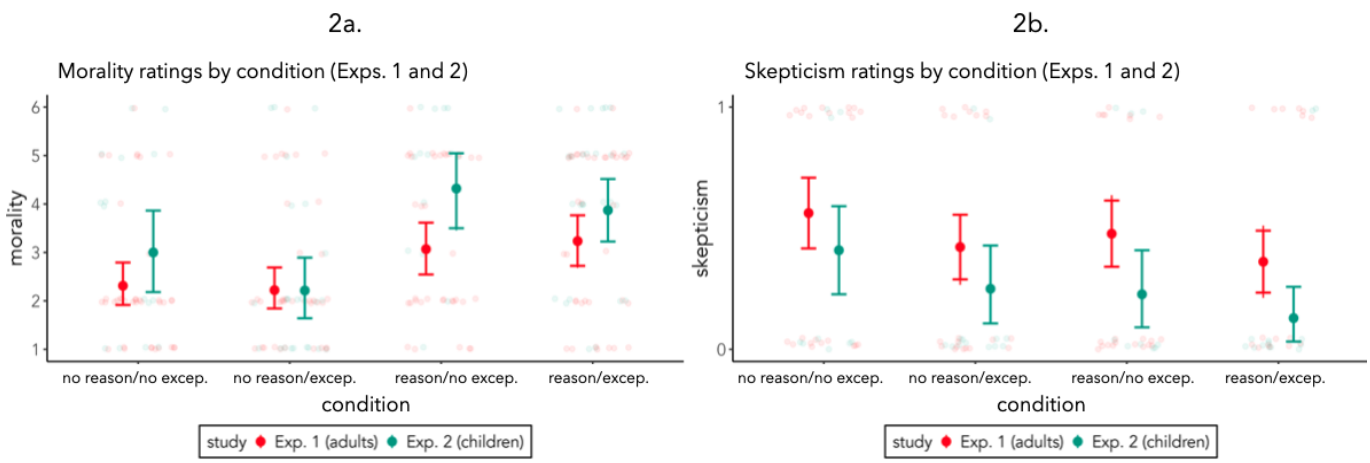

Figure 2: Regardless of the presence of exception, justification increased positive morality judgments for children and adults in Exps. 1 and 2. However, manipulations did not impact skepticism about the true intention of the rule.

and then evaluated the true intention of the rule (i.e., "Do you think the rule is about boys or short hair?"). Responses were coded as 1 = skeptical (inferred the rule is actually about boys) or 0 = accepted rule (believed the short hair rule at face value). Third, participants made judgments about the creator's gender bias ("Which four friends do you think Ro wants to play with after school?"; participants could select between short-haired girls (2), long-haired girls (3), short-haired boys (3), and long-haired boys (2). Responses were coded to tally the number of boys selected by each participant; more "boy" selections would constitute an inference of gender bias. Participants also reported a number of other demographic factors, including conservatism on a seven-point scale from "Very Liberal" to "Very Conservative".

Results and Discussion

Morality. Overall, participants negatively judged the disparate impact rule, $t(183) = -6.21, p < .001$. As shown in Figure 2, condition predicted morality judgments, $F(3,180) = 4.31, p = .006$, but there was no interaction between the reason and exception study factors. This is because of the strength of the reason manipulation. Posthoc paired contrasts indicated that participants who heard a reason rated the rule as more moral than those who did not, $B = 0.88, p < .001, 95\%CI[0.39, 1.37]$. However, there was no such main effect of the exception manipulation, $B = 0.04, p = .88, 95\%CI[-0.45, 0.53]$.

True intention of rule. There was no effect of condition on participants' inferred intention of the rule, $\chi^2(3) = 4.17, p = .24$. Overall, skepticism did not differ from chance.

Creator's bias. There was an overall effect of condition on beliefs that the rule creator is biased, $F(3,180) = 4.47, p = .005$, as well as a significant interaction between the reason and exception study factors, $B = -0.40, p = .048, 95\%CI[-0.80, -0.003]$. That is, observing an exception reduced participants' inference that the creator was biased, but only if the creator stated a reason; condition difference between reason/exception vs. reason/no-exception: $B = -0.25, p < .001, 95\%CI[-0.39, -0.11]$. We did not find a main effect of the reason manipulation ($p > .05$).

Demographic factors. Female participants (vs. non-female identifying participants) rated the rule as less moral, $B = -0.69, p = .01, 95\%CI[-1.21, -0.16]$, but did not differ on rule intention or creator bias measures. More politically conservative participants rated the rule as more moral, $B = 0.27, p = .03, 95\%CI[0.03, 0.51]$, and were less skeptical of the rule, $B = -0.32, p = .04, 95\%CI[-0.64, -0.03]$, but conservatism did not relate to the bias measure ($p > .05$).

In Exp. 1, the combination of justification and exception mitigated perceptions of the rule-maker's gender bias. However, justification on its *own* moderated moral criticism of the disparate impact rule. It follows that different mechanisms may target different inferences related to disparate impact rules—judgments of the rule itself and judgments of the rule's creator.

Critically, across conditions, members of the group marginalized by the rule (here, women) were more sensitive to the rule's implications than those who did not belong to this group: They negatively evaluated the rule regardless of justification or exceptions. In Experiment 2, we explore whether we find these same patterns among children. We also examine whether these judgments vary with age.

Experiment 2

Exp. 2 was an exploratory replication of Exp. 1, conducted with children.

Method

Participants. We recruited 103 5- to 12-year-old children (M_{age} = 7.48 years, SD_{age} = 2.01; 47 girls, 56 boys) on ChildrenHelpingScience.com (Lookit; Scott & Schulz, 2017).

Two participants were excluded due to missing demographic information.

Materials and Procedure. Exp. 2 was identical to Exp. 1, except that the stimuli were presented to participants in an asynchronous storybook format via images and audio recordings.

Results and Discussion

Morality. Children's moral judgments did not differ from chance. As shown in Figure 2, we found an overall effect of condition on morality judgments, $F(3,99) = 6.41, p < .001$. We did not find evidence of an interaction between the reason and exception study factors. Paired contrasts indicated a main effect of the reason manipulation, such that children who heard a reason rated the rule as more moral children than those who did not, $B = 1.49, p < .001, 95\%CI[0.75, 2.23]$. We did not find a main effect of the exception manipulation, $B = -0.62, p = .10, 95\%CI[-1.36, 0.13]$.

True intention of rule. Overall, children were not very skeptical of the true intention of the disparate impact rule, $t(102) = -6.06, p < .001$. We did not find an overall effect of condition on children's inferred intention of the rule, $\chi^2(3) = 5.53, p = .14$.

Creator's bias. Children across conditions inferred some level of gender bias on Ro's part, $t(102) = 3.20, p = .002$. However, we did not find an overall effect of condition on children's inferences about Ro's gender bias, $F(3,99) = 1.18, p = .32$.

Demographic factors. Girls did not differ from boys in their morality ratings or skepticism, but girls were less likely to infer that Ro was biased towards boys—presumably because they reasoned that Ro should prefer girls (i.e., reflecting participants' own in-group bias), $B = -0.61, p < .001, 95\%CI[-0.96, -0.27]$. Older participants rated the rule as less moral, $B = -0.32, p < .001, 95\%CI[-0.51, -0.13]$, but age did not relate to other dependent measures.

Here, like with adults, justification impacted children's moral judgments about the disparate impact rule more than the presence of an exception did. However, it remains unclear how strong a reason needs to be to have a mitigating effect. Given that adults and even young children can systematically calibrate the quality of different explanations (Rips, 2002; Mills et al., 2017; Liquin & Lombrozo, 2022), prefer rich explanations to empty (circular) ones (Baum et al., 2008; Corriveau & Kurkul, 2014), and prefer circular explanations to no explanation (Mercier et al., 2014), it is possible that a circular reason would be less impactful than a compelling reason in raising moral judgments of the disparate impact policy and its creator. In Experiment 3, we examine different types of justification to assess its causal impact in increasing acceptance of disparate impact rules.

Experiment 3

Exp. 3 (pre-registered) was a modified version of Exp. 1, uniquely focusing on the role of different forms of justification in mitigating skepticism about disparate impact rules. Specifically, we examined the impact of a compelling justification (a strong-reason) and a circular explanation (a minimal-reason) on adults' judgments about a disparate impact rule. We hypothesized that the compelling reason would be more effective than both the circular explanation and no explanation (no-reason) in mitigating skepticism.

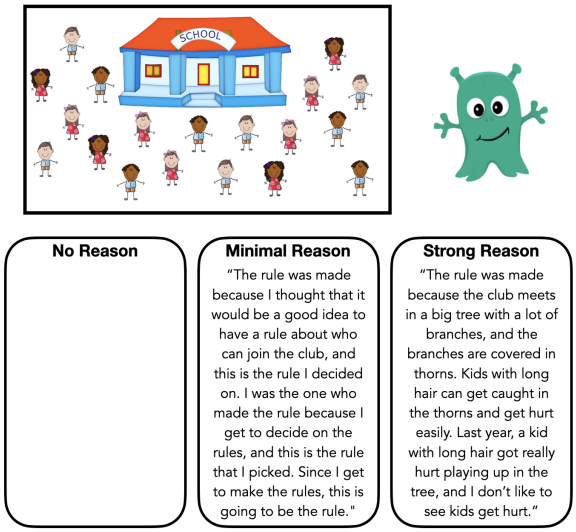

Figure 3: Participants in Exp. 2 heard no reason, a minimal reason, or a strong reason after hearing Ro's rule.

Method

Participants. We recruited 150 adult participants on Prolific. One participant was excluded due to attention check failure, resulting in a final sample of 149 participants (74 women, 73 men, 1 non-binary, 1 did not report; 94 White, 20 Black, 15 Asian/Pacific Islander, 13 Multiracial, 6 Latine/x, 1 Native American or Alaskan Native, 8 did not report). 23 identified as extremely liberal, 45 as liberal, 18 as slightly liberal, 31 as moderate, 13 as slightly conservative, 14 as conservative, 3 as extremely conservative, and 2 did not report/did not identify with a political ideology.

Materials & Procedure. Exp. 3 had a similar design to Exp. 1: participants heard about Sunny School, met the students and Ro, and learned about Ro's "short hair" club. However, Exp. 3 also had crucial differences. First, instead of seeing a visual representation of the club members, participants referenced a picture of *all* the students at Sunny School. This was done to avoid drawing explicit attention to the disproportionate representation of boys in the club; we wanted to see if participants would make this connection themselves. Additionally, Exp. 3 focused exclusively on the role of justification: Participants were either in the strong-reason, minimal-reason, or no-reason condition (see Figure 3). In the strong-reason condition, participants heard a similar compelling reason to the one they heard in Exp. 1. In the minimal-reason condition, participants heard a circular explanation (a lengthy version of "The rule was made because I make the rules!" that

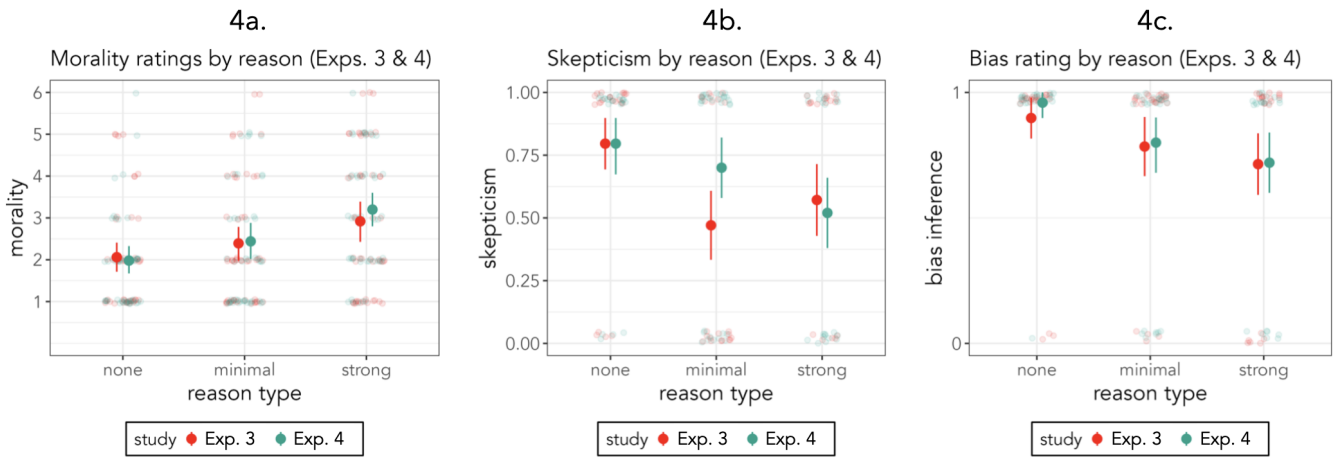

Figure 4: Participants in Exp. 3 ("short-hair" rule) and Exp. 4 ("long hair" rule) were made more accepting and less skeptical of the disparate impact rule and its creator when they heard compelling justification.

was matched in length to the strong reason). In the no-reason condition, participants heard no reason for the rule.

As in Exps. 1 and 2, participants answered questions about the rule's morality, its true intent, and Ro's potential gender bias. However, we made minor changes for clarity and consistency with the Exp. 3 framing. In particular, we adapted the bias measure such that participants were instructed to select one of three *groups* of children (instead of four individual children): One group with four boys in it, one with four girls, and one with two boys and two girls. Finally, because we did not manipulate exception here, all boy characters had short hair and all girl characters had long hair.

Results and Discussion

Figure 4 shows the dependent measures in Exp. 3 by condition.

Morality. Participants across conditions judged the disparate impact rule negatively, $t(148) = -8.40, p < .001$. However, we found an overall effect of condition on positive morality judgments, $F(2,146) = 4.20, p = .02$. Paired contrasts indicated that a strong reason increased moral judgments relative to the no-reason condition, $B = 0.86, p = .01, 95\%CI[0.27, 1.45]$. There were no other condition differences.

True intention of rule. Overall, participants were skeptical of the rule, $t(148) = 2.76, p < .001$. We found a main effect of condition on participants' inferred intention of the rule, $\chi^2(3) = 11.60, p = .003$. Paired contrasts indicated that a strong-reason reduced skepticism of the rule relative to the no-reason condition, $B = -1.07, p = .02, 95\%CI[-1.97, -0.18]$, as well as the minimal-reason condition relative to the no-reason condition, $B = -1.48, p = .001, 95\%CI[-2.37, -0.59]$. The reason conditions did not differ from one another.

Creator's bias. We found a trending effect of condition for judgments of bias, $\chi^2(3) = 5.24, p = .07$. Paired contrasts indicated that a strong-reason reduced judgments of bias relative to the no-reason condition, $B = -1.26, p = .03, 95\%CI[-2.37, -0.15]$. There were no other condition differences.

Demographic factors. We did not find any gender differences in ratings. However, as in Study 1, conservative-leaning participants rated the rule as more moral, $B = 0.21, p = .01, 95\%CI[0.07, 0.36]$, were less skeptical of the rule, $B = -0.44, p < .001, 95\%CI[-0.68, -0.23]$, and were less likely to infer the creator was biased, $B = -0.05, p = .02, 95\%CI[-0.09, -0.01]$.

Politically polarized judgments of this gender-based disparate impact rule raised a new question: are conservative participants less sensitive to group-based exclusion in general, or is their tolerance of discrimination specific to historically marginalized groups (here, girls)? To test this, in Experiment 4, we assess people's judgments of a disparate impact rule that targets boys.

Experiment 4

Exp. 4 (pre-registered) was a replication of Exp. 3 using a disparate impact rule that targeted boys to test whether judgments of disparate impact rules are specific to rules that target historically marginalized groups (i.e., girls).

Method

Participants. We recruited 150 adult participants on Prolific. One participant was excluded due to attention check failure, yielding a final sample of 149 participants (75 women, 74 men; 83 White, 21 Asian, 18 Black, 13 Multiracial, 12 Hispanic/Latine/x, 1 Native American or Alaskan Native, 1 Middle Eastern or North African). 17 participants identified as extremely liberal, 46 as liberal, 22 as slightly liberal, 26 as moderate, 12 as slightly conservative, 14 as conservative, 3

as extremely conservative, and 9 did not report.

Exp. 4 was identical to Exp. 3, with one crucial difference: Ro makes a rule about long hair instead ("Only kids with long hair are allowed to be in the club"), indirectly excluding boys. Hypotheses for Exp. 4 were also identical to Exp. 3: we predicted that the strong reason would increase judgments of the rule as morally acceptable, decrease skepticism, and reduce inferences about the rule-maker's gender bias more than the minimal reason or no reason. We also planned to explore the effect of political conservatism on participant judgments.

Results

Morality. Like in Exp. 3, participants negatively judged the morality of the rule across conditions, $t(148) = -7.58, p < .001$. We found an overall effect of condition on morality judgments, $F(2,146) = 8.78, p < .001$. Paired contrasts indicated that a strong-reason increased moral judgments relative to the no-reason condition, $B = 1.22, p < .001, 95\%CI[0.64, 1.80]$. In addition, the strong-reason increased moral judgments relative to the minimal-reason condition, $B = 0.76, p = .01, 95\%CI[0.18, 1.34]$. The minimal-reason condition did not differ from the no-reason condition.

True intention of rule. Participants were also skeptical of the rule across conditions, $t(148) = 4.43, p < .001$. We found an overall effect of condition on participants' inferred intention of the rule, $\chi^2(2) = 8.82, p = .01$. Paired contrasts indicated that a strong-reason (vs. no-reason) reduced skepticism of the rule, $B = -1.28, p = .005, 95\%CI[-2.17, -0.39]$. There were no other significant condition differences.

Creator's bias. We found an overall effect of condition on judgments of bias, $\chi^2(2) = 10.17, p = .006$. Paired contrasts indicated that a strong-reason (vs. no-reason) reduced judgments of bias, $B = -1.28, p = .005, 95\%CI[-2.17, -0.39]$. There were no other significant condition differences.

Demographic factors. Female participants were more skeptical of the rule, $B = 1.83, p = .02, 95\%CI[0.13, 1.54]$, but did not differ from non-female identifying participants on morality or bias judgments. Unlike in Exp. 3, political conservatism did not predict any measures for Exp. 4 (ps $> .05$).

Given that Exp. 4 participants who witnessed the exclusion of boys did not judge the rule differently from Exp. 3 participants who witnessed the exclusion of girls, people may be sensitive to the implications of disparate impact rules even when they do not target protected groups. However, it seems that the influence of political ideology is specific to judgments of disparate impact rules that target *marginalized* groups.

General Discussion

Creators of disparate impact rules harness indirect forms of discrimination to evade detection. Across four experiments, we found that children and adults largely reject disparate impact rules, are skeptical of their true intent, and infer that the rule-maker is biased against the group the rule targets. However, when given a reason for the rule, people of all ages judge it to be more acceptable. Crucially, we find evidence that, for some measures, these justifications do not even have to be compelling to be effective. In some cases, this can be deeply problematic: Disparate impact rules may persist in society even without a plausible justification.

These results present a novel addition to existing literature on the power of justification in causal and social reasoning contexts. While prior work has shown that providing some rationale can amend reasoners' moral judgments and willingness to accept inequality (Kanngiesser et al., 2021; Starmans et al., 2017), here we show that this is also effective in an ambiguous context. Elucidating the power of justification here is particularly important because of potential implications for the legal rights of marginalized groups.

Perhaps unsurprisingly, we find that social factors also impact judgments of disparate impact rules. In particular, when the rule discriminates against a historically marginalized group, adults who do *not* belong to that group are more tolerant of the rule. This tolerance is consistent with prior work showing that both children and adults are more tolerant of discrimination against their out-group (Brown & Bigler, 2004; Chong et al., 2022). Further, conservative adults tend to be more accepting of disparate impact rules than liberals are. Finally, we find that children's tolerance for disparate impact rules decreases with age—even in the context of justification.

Showing participants a single exception to the rule in Exps. 1 and 2 did not change their moral judgments or skepticism about the disparate impact rule. However, the presence of an exception did moderate participants' inferences about the rule creator's gender bias. Thus, while exceptions may not directly impact judgments of disparate impact rules (at least in this context), they may affect judgments of rule-makers. A possible explanation for this effect is that exceptions provide evidence that—although the rule itself may be unfair—the rule-maker still technically adheres to the rule and is therefore principled. This perception may provide more legitimacy to the justification they provide. Alternatively, exceptions in this case may have called attention to the gender disparity within the club, thus making people more skeptical about the rule's true intent. Future work will further explore the impact of exceptions (e.g., tokenism) on judgments of indirect discrimination.

Future work will also examine how people judge disparate impact rules when compared to blatantly discriminatory or entirely benign rules. Additionally, because participant identity and political ideology impact judgments of disparate impact rules in this novel scenario, future work will evaluate adults' judgments of real-world disparate impact rules (e.g., voter identification laws). Taken together, these findings have important implications for studying the powerful effects of justification in sociopolitical contexts. Ultimately, uncovering how disparate impact rules persist in a society that ostensibly rejects discrimination will inform efforts to eliminate them.

Acknowledgments

We thank members of the Early Learning and Cognition Lab for input and feedback. In particular, we thank research assistants Jordan Cannone and Kathleen Nguyen for data processing assistance. This work was supported by a Jacobs Foundation Research Fellowship to C.M.W.

References

Alderete, S., & Xu, F. (2023). Three-year-old children's reasoning about possibilities. *Cognition*, *237*, 105472.

Baum, L. A., Danovitch, J. H., & Keil, F. C. (2008, June). ChildrenÄôs sensitivity to circular explanations. *Journal of Experimental Child Psychology*, *100*(2), 146–155. Retrieved 2024-01-21, from https://www.sciencedirect.com/science/article/pii/S0022096507001415 doi: 10.1016/j.jecp.2007.10.007

Bridgers, S., Schulz, L., & Ullman, T. D. (2021). Loopholes, a window into value alignment and the learning of meaning. In *Proceedings of the annual meeting of the cognitive science society* (Vol. 43).

Brown, C., & Bigler, R. (2004, September). Children's Perceptions of Gender Discrimination. *Developmental psychology*, *40*, 714–26. doi: 10.1037/0012-1649.40.5.714

Ceballos, C. I., Engstrom, D. F., & Ho, D. E. (2021). Disparate limbo: how administrative law erased antidiscrimination. *Yale LJ*, *131*, 370.

Cesana-Arlotti, N., Varga, B., & Téglás, E. (2022). The pupillometry of the possible: an investigation of infants' representation of alternative possibilities. *Philosophical Transactions of the Royal Society B*, *377*(1866), 20210343.

Chong, D., Citrin, J., & Levy, M. (2022, October). The Realignment of Political Tolerance in the United States. *Perspectives on Politics*, 1–22. Retrieved 2024-01-21, from https://www.cambridge.org/core/journals/perspectives-on-politics/article/realignment-of-political-tolerance-in-the-united-states/7813ABE313EC0008AB3E282292A318D7 (Publisher: Cambridge University Press) doi: 10.1017/S1537592722002079

Corriveau, K. H., & Kurkul, K. E. (2014). "why does rain fall?": Children prefer to learn from an informant who uses noncircular explanations. *Child development*, *85*(5), 1827–1835.

Ellis, R. (2017, January). *Protesters mass at airports to decry Trump's immigration policies | CNN Politics.* Retrieved 2024-01-10, from https://www.cnn.com/2017/01/28/politics/us-immigration-protests/index.html

Geraci, A., & Surian, L. (2011). The developmental roots of fairness: Infants' reactions to equal and unequal distributions of resources. *Developmental science*, *14*(5), 1012–1020.

Gerstein, J., & Lin, J. C. F. (2018, June 26). *Why these 7 countries are listed on trump's travel ban.* https://www.politico.com/interactives/2018/trump-travel-ban-supreme-court-decision-countries-map/. (Accessed: 2024-02-01)

Goddu, M. K., Sullivan, J. N., & Walker, C. M. (2021). Toddlers learn and flexibly apply multiple possibilities. *Child Development*, *92*(6), 2244–2251. Retrieved 2024-02-01, from https://onlinelibrary.wiley.com/doi/abs/10.1111/cdev.13668 (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/cdev.13668) doi: 10.1111/cdev.13668

Hill, D. B., & Tenenbaum, H. R. (2022). Children's physical gender schemas: Acquisition and features. *Sex Roles*, *87*(7), 406–418.

Hochmann, J.-R., Tuerk, A. S., Sanborn, S., Zhu, R., Long, R., Dempster, M., & Carey, S. (2017, December). Children's representation of abstract relations in relational/array match-to-sample tasks. *Cognitive Psychology*, *99*, 17–43. doi: 10.1016/j.cogpsych.2017.11.001

Kanngiesser, P., Mammen, M., & Tomasello, M. (2021). Young children's understanding of justifications for breaking a promise. *Cognitive Development*, *60*, 101127.

Killen, M., Lee-Kim, J., McGlothlin, H., Stangor, C., & Helwig, C. C. (2002). How children and adolescents evaluate gender and racial exclusion. *Monographs of the society for research in child development*, i–129.

Killen, M., & Stangor, C. (2001a). Children's social reasoning about inclusion and exclusion in gender and race peer group contexts. *Child development*, *72*(1), 174–186.

Killen, M., & Stangor, C. (2001b). Children's Social Reasoning about Inclusion and Exclusion in Gender and Race Peer Group Contexts. *Child Development*, *72*(1), 174–186. Retrieved 2024-01-11, from https://onlinelibrary.wiley.com/doi/abs/10.1111/1467-8624.00272 (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/1467-8624.00272) doi: 10.1111/1467-8624.00272

Lapidow, E., Goddu, M. K., & Walker, C. M. (2022). Reasoning from samples to populations: Children use variability information to predict novel outcomes. In *Proceedings of the annual meeting of the cognitive science society* (Vol. 44).

Lapidow, E., Tandon, T., Goddu, M., & Walker, C. M. (2021). A tale of three platforms: Investigating preschoolers' second-order inferences using in-person, zoom, and lookit methodologies. *Frontiers in Psychology*, *12*, 731404.

Lee, A. R. (2020). All you have gotten is tokenism. *Prejudice, stigma, privilege, and oppression: A behavioral health*

handbook, 387–399.

Liquin, E. G., & Lombrozo, T. (2022). Motivated to learn: An account of explanatory satisfaction. *Cognitive Psychology*, *132*, 101453.

LoBue, V., Nishida, T., Chiong, C., DeLoache, J. S., & Haidt, J. (2011). When getting something good is bad: Even three-year-olds react to inequality. *Social Development*, *20*(1), 154–170.

McAuliffe, K., Blake, P. R., Steinbeis, N., & Warneken, F. (2017, February). The developmental foundations of human fairness. *Nature Human Behaviour*, *1*(2), 1–9. Retrieved 2024-01-10, from https://www.nature.com/articles/s41562-016-0042 (Number: 2 Publisher: Nature Publishing Group) doi: 10.1038/s41562-016-0042

McAuliffe, K., Blake, P. R., & Warneken, F. (2014, December). Children reject inequity out of spite. *Biology Letters*, *10*(12), 20140743. Retrieved 2024-01-10, from https://royalsocietypublishing.org/doi/10.1098/rsbl.2014.0743 (Publisher: Royal Society) doi: 10.1098/rsbl.2014.0743

Mercier, H., Bernard, S., & Clément, F. (2014, September). Early sensitivity to arguments: How preschoolers weight circular arguments. *Journal of Experimental Child Psychology*, *125*, 102–109. Retrieved 2024-01-21, from https://www.sciencedirect.com/science/article/pii/S0022096513002488 doi: 10.1016/j.jecp.2013.11.011

Mills, C. M., Danovitch, J. H., Rowles, S. P., & Campbell, I. L. (2017). Children's success at detecting circular explanations and their interest in future learning. *Psychonomic bulletin & review*, *24*, 1465–1477.

Murthy, S. K., Bridgers, S., Parece, K. M., Glassman, E. L., & Ullman, T. (2023). Comparing the Evaluation and Production of Loophole Behavior in Children and Large Language Models.

Rhodes, M., & Mandalaywala, T. M. (2017). The development and developmental consequences of social essentialism. *Wiley Interdisciplinary Reviews: Cognitive Science*, *8*(4), e1437.

Rips, L. J. (2002). Circular reasoning. *Cognitive science*, *26*(6), 767–795.

Schmidt, M. F., Svetlova, M., Johe, J., & Tomasello, M. (2016). Children's developing understanding of legitimate reasons for allocating resources unequally. *Cognitive Development*, *37*, 42–52.

Scott, K. E., Henkel, M. A., Moens, O. M., Devine, P. G., & Shutts, K. (2023). ChildrenÄôs evaluations of and reactions to racial discrimination. *Developmental Psychology*, *59*(7), 1190–1202. (Place: US Publisher: American Psychological Association) doi: 10.1037/dev0001542

Seicshnaydre, S. E. (2016, August). *Disparate Impact and the Limits of Local Discretion after <I>Inclusive Communities</I>* [SSRN Scholarly Paper]. Rochester, NY. Retrieved 2024-01-10, from https://papers.ssrn.com/abstract=2823813 doi: 10.2139/ssrn.2823813

Serwer, A. (2019, January). *Trump Is Making It Easier to Get Away With Discrimination.* Retrieved 2024-01-10, from https://www.theatlantic.com/ideas/archive/2019/01/disparate-impact/579466/ (Section: Ideas)

Shaw, A., DeScioli, P., & Olson, K. R. (2012, November). Fairness versus favoritism in children. *Evolution and Human Behavior*, *33*(6), 736–745. Retrieved 2024-01-10, from https://www.sciencedirect.com/science/article/pii/S1090513812000608 doi: 10.1016/j.evolhumbehav.2012.06.001

Sloane, S., Baillargeon, R., & Premack, D. (2012). Do infants have a sense of fairness? *Psychological science*, *23*(2), 196–204.

Sommerville, J. A., Schmidt, M. F., Yun, J.-e., & Burns, M. (2013). The development of fairness expectations and prosocial behavior in the second year of life. *Infancy*, *18*(1), 40–66.

Starmans, C., Sheskin, M., & Bloom, P. (2017). Why people prefer unequal societies. *Nature Human Behaviour*, *1*(4), 1–7.

Theimer, C., Killen, M., & Stangor, C. (2001, January). Young ChildrenÄôs Evaluations of Exclusion in Gender-Stereotypic Peer Contexts. *Developmental psychology*, *37*, 18–27. doi: 10.1037/0012-1649.37.1.18