Abstract: Some scientific studies purport to show differences in, for example, ability or intelligence between members of two social groups, such as race or gender. Given that such science has historically been used to support morally objectionable practices such as discrimination, those interested in preventing these morally objectionable ends often engage in criticizing the science. This paper first examines exactly what good it does to criticize such science, which I call ”inferiorizing science,” and then gives recommendations on which critical strategies are more effective than others. I conclude that methodological criticisms are the most effective, because they shed doubt on whether there is even an actual difference between the groups at all. Another common strategy, claiming that the differences are real but only caused by social or environmental influences, is less effective, because the distinction between social/environmental and biological is problematic.

Did they prove anything scientifically about us before they killed us? No. They killed us first and then tried to get some scientific proof about why we should die.

-Toni Morrison, Song of Solomon

1 Introduction: Science and Oppression

Science and oppression have a mixed history. On the one hand, science has often been used to legitimize oppression. Scientific racism, or the belief that racist ideologies were appropriately based on scientific results, was used to justify American slavery and the persecution of Jews during the Holocaust. Similar “scientific” beliefs have perpetuated the subordination of women and the denial of legal protection to the LGBTQ+ population. Unfortunately not all examples are so archaic. In The Bell Curve, Hernstein and Murray (1994) used correlative data to claim that racial differences in measured intelligence are partly genetic. Baron-Cohen (2005) cited extensive psychological research
of differences between women and men in support of then-Harvard University president Lawrence Summers claim that “there is reasonably strong evidence of taste differences between little girls and little boys that are not easy to attribute to socialization,” a claim meant to justify gender disparities in engineering.

On the other hand, the view that “better science” led to the rejection of racism seems to be pervasive. Contextualizing his analysis of the term “racism,” Blum (2002, p. 4) states, “The rejection of racism was fed...by developments within the sciences that, independently, had begun to throw into question the idea of a hierarchy of discrete human populations or ‘races.’” Similarly, Anderson (2010, p. 50), in The Imperative of Integration, thanks “the discrediting of ‘scientific’ racism” for the rejection of explicit racism in America since the Civil Rights Movement.

This paper examines various strategies that are used to criticize science which purports to show that members of two social groups – e.g., race and gender – are, by some measure, different. First, in section 2 I will explain why it is so important to levy these criticisms: the answer is not so straightforward. Then, in section 3 I will present and categorize common types of criticisms. Finally, in section 4 I will argue that some of these common strategies for criticism are better – i.e., more effective – than others. This will yield a recommendation for those in the business of criticizing this science about how they should do so.

2 Why criticize inferiorizing science?

Why should we engage in criticisms of science which purports to show that members of two social groups are different in some way? The most intuitive answer is that we are worried that such science will be used to support a morally unacceptable conclusion. The argument from the scientific results to that conclusion may go something like this:

1There are other group categorizations besides those based on race and gender, and inferiorizing
science can be used to uphold hierarchies between these groups as well. But I will focus on race and
gender, just given the plethora of examples with respect to these groups.
There is a difference between groups F and G on some measure M.

For any groups P and Q, if there is a difference between P and Q on some measure M, then it is morally permissible to accept a difference between P and Q on some measure N.

Therefore, it is morally permissible to accept a difference between F and G on measure N.

This argument requires some explanation. First, P1 is supposed to be the conclusion of whatever scientific study that shows differences between groups. Such a study would have taken group membership – F and G – as an independent variable, and some measure M as a dependent variable. Evidence that the groups are different with respect to M will support P1. P2 is a conditional about any groups P and Q which, when filled in with F and G, connects P1 as the antecedent with a moral proposition as the consequent. In an attempt to be general, I used as the consequent that it is “morally permissible to accept a difference between groups P and Q on some measure N.” I have used “accept” to capture the thought that if there is a difference that is morally permissible to accept, then we should not automatically feel inclined to do something to change or remove that difference (the opposite of “accept” being “reject”). Note that the accepted difference is on measure N, not M, although I leave open the possibility that N=M. C just follows by modus ponens from P1 and P2, so the argument is valid.

Let me go through an example to make this clear. Say that F and G are women and men, and that M is some measure of their intelligence. Say that we have scientific evidence that there is a difference between women and men on that measure of intelligence, such that P1 has empirical support. (I leave to the side for right now what we might say about such science; that is the subject of section §3.) We could use N=M, in which case the relevant instance of P2 says that if there is such a difference in intelligence between men and women, then it is morally permissible to accept a difference in intelligence (say, on future men and women, or men and women who were not in the study sample). However, we could also use a different measure N, such as lifetime earnings or educational attainment. P2 would then say that if there is a difference between these groups with respect
to intelligence, then we are permitted to accept a difference with respect to earnings or education. Depending on the relationship between M and N, we may be more or less likely to agree with P2. In any case, if we do grant P2, then we have to also grant C, and conclude that the difference on N – intelligence, earnings, or education – is morally permissible to accept.

I will assume that in many such cases the conclusion C is morally unacceptable (for some given F, G, and N). I will not provide independent arguments for this assumption; I take it that my readers will generally agree with me. If the argument given above is valid and we want to deny the conclusion, we must deny (at least) one of the premises.

This is where criticizing the science comes in. If the science that purportedly justified P1 is faulty in some respect, then we may no longer have good reason to think that P1 is true: the groups may not be actually different on measure M. If we can reject P1, we can reject C. I take it this is the most straightforward role for criticizing the science to play in rejecting various socially objectionable conclusions.

The problem is that P2 is much easier to criticize than P1. P2 commits what is known as the naturalistic fallacy. The antecedent of P2 is an “is” statement – it is a factual claim about how the world is. The consequent is an “ought” statement – it is a claim about how the world ought to be (in this case, about how it is permissible for the world to be). Most philosophers agree that statements of the form of P2 are fallacious, even before we fill in the details of F, G, M, and N. Rejecting P2 enables us to reject C, without rejecting P1.

Just to make this point patently clear, I will call science that supports claims of the form of P1 “inferiorizing science.” The idea is that, in showing that two groups are different on some measure M, one is also showing that one group is worse; that group is inferior (with respect to M). However, perhaps counterintuitively, inferiority does not imply anything about how people should be treated, or what we morally should accept about how their lives generally turn out. In fact, it may even reasonably imply that the “inferior” group should be treated in a way that removes the difference; perhaps if women
are less good at math, for example, than men, they should be given more training until that difference disappears.²

So why go to the trouble of criticizing the science, focusing on P₁, when it seems we have good arguments against P₂? My hypothesis is that a variation on P₁ is involved in a more insidious argument, of the following form:

\[
P₁^*: \text{There is a necessary difference between groups F and G on some measure M.}
\]
\[
P₂^*: \text{For any groups P and Q, if there is a necessary difference between P and Q on some measure M, then it is morally permissible to accept a difference between P and Q on some measure N.}
\]
\[
C: \text{Therefore, it is morally permissible to accept a difference between F and G on measure N.}
\]

The only difference between this argument and the first argument is the inclusion of the word “necessary” in P₁ and P₂. The conclusion is the same. This revised argument is more insidious than the first because P₂ is much more difficult to reject than P₂. P₂* is based on an ought-implies-can principle. If the difference is necessary, then there can’t not be a difference, so it won’t be the case that there ought not be a difference, in which case the difference will be permissible (and thus permissible to accept). Although there may be strategies for rejecting P₂*, they are certainly not as straightforward as just identifying P₂ as committing the naturalistic fallacy.

For those of us who wish to reject C, the most viable option now is to reject P₁*. In section 3, I turn to ways in which P₁* is commonly rejected.

3 Common critical strategies

There are two ways of showing that P₁* is false. One can show:

²In a similar vein, [Jordan-Young (2010) p. 5] says, “In an era where diversity is celebrated, the idea of ‘sex in the brain’ no longer equals an endorsement of male superiority, and critics of the idea are increasingly cast as not only antiscience, but antidiversity . . . [W]hile the notion of innately different preferences in men and women was once politically suspect, it is now often suggested that accepting these innate differences will encourage a more rational approach to equality.”

³I don’t intend “necessary” to denote some strong version of the word, like logical or metaphysical necessity. I mean something more like “immutable” or “inflexible.”
1. There is no actual difference between groups F and G on M.

2. The difference between groups F and G on M is actual, but not necessary.

3.1 Showing the difference is not actual

In the context of criticizing inferiorizing science, showing that a difference between groups F and G is not actual will amount to saying that the science that purports to show this difference does not actually do so. (This is slightly different from showing that there is no actual difference; instead, the critics are just saying that we don’t have evidence that there is an actual difference, so we shouldn’t think that there is one.) There are common ways in which critics of inferiorizing science levy their criticisms. I will explain the general idea behind each of these ways, and give some examples. My categories are not necessarily exhaustive; my aim is merely to highlight some patterns.

1. Researcher bias. Critics of inferiorizing science often point out that the scientists who conduct the studies in question are – like many of us would be – somewhat biased about their results before their studies are conducted. These biases can affect every part of the scientific method, from study design to interpretation of results.\footnote{For example, Fine (2010) often discusses the effect of confirmation bias, or “experimenter expectancy effects;” for instance, an experimenter visiting a maternity ward to study newborns will likely know the gender of the baby they are studying, which could predispose them to expect the baby to act in certain ways (p. 114).}

2. Statistical analysis. Sometimes critics of inferiorizing science have noticed that the statistical analysis of the data upon which a study’s conclusion is based has been performed poorly. This can include everything from downright calculation mistakes to questionable choice of statistical method.\footnote{Unfortunately, there are many examples of this. Here are a few: (1) In an influential study, Bateman (1948) claimed that competitiveness in male flies led to reproductive success. However, this was actually only true for part of his sample; the conclusion could only be drawn because data were separated, for no clear reason, into two different subsets for analysis (Fine 2017, p. 35, citing Tang-Martinez 2010). (2) Jordan-Young (2010) points out that there are common so-called Type II statistical errors in studies of brain differences between human males and females, wherein “the chances of finding an association between prenatal hormones and later sex-typed traits is inflated” (p. 88). (3) Gould (1996) gives many examples of times when scientists historically botched the statistical analysis involved in studies of racial differences, coincidentally always to the effect of showing white superiority.}
3. **Measurement tools.** As often happens in the social sciences, measurement tools that are operationalizable by researchers only serve as an imperfect proxy for what they seek to measure. Researchers’ choice of a measurement tool, even if not their employment of it, can be criticized. The point of these criticisms is to say that, while the research may have shown that two groups are different with respect to \( M \) (e.g., IQ or, historically, brain size), \( M \) only imperfectly captures what we intended to measure (e.g., intelligence).\(^6\)

4. **Definitions.** Especially with respect to gender, the definitions researchers use are often criticized. If definitions of the independent or dependent variables change over time or are otherwise contradictory within the literature, results that superficially appear to be consistent will not actually present such a unified front.\(^7\)

5. **Broad methodological critique.** The types of criticisms I am calling “broad methodological critiques” are those that target an entire research program for having faulty methodology. While these criticisms may be legitimate, they usually apply more broadly than just inferiorizing science.\(^8\)

### 3.2 Showing the difference is not necessary

Even if one instance of inferiorizing science is methodologically flawless, meaning that it does support the conclusion that groups \( F \) and \( G \) are actually different on measure \( M \), this does not imply that such a difference is necessary. The latter is needed to support \( P1^* \).

\(^6\)For several examples with respect to race and intelligence measures, see Gould (1996).

\(^7\)There are several examples of this criticism. Jordan-Young (2010, chpt. 6) points out that definitions of female sexuality have changed over time. Thus studies that correlated hormone exposure with feminine versus masculine sexual behavior in one time period (especially, pre-1980s) are actually incompatible with studies conducted in a different time period (post 1980s) that correlate the same independent variable with differently-delineated sexual behavior. Longino (2012) also points out that studies of sexuality and aggression don’t consistently define these two phenomenon; different research methods typically use different proxies for each, which makes studies using these different methods incommensurable. Fausto-Sterling (2000) discusses the various ways in which sex itself has been defined by the medical community over time in order to classify various subsets of intersex individuals. Finally, Richardson (2010) criticizes Carrel and Willard (2005) for comparing the genetic differences between men and women to that between humans and chimpanzees, because doing so implies that the “human” genome is somehow not defined by the genes of both sexes.

\(^8\)For example, Fine (2010) criticizes the use in general of brain scans (p. 134), animal studies (p. 102), and statistical significance (p. 145).
There is even something positive about showing that the difference in question is not necessary without showing that it is not actual. Actual differences can be used to motivate social change. Showing that there is a difference now but that there need not be one could be used to argue the ways in which we should restructure society such that the difference disappears. Social science that shows evidence of discrimination (workforce, housing, etc.), for example, is often cited to motivate policy change, or at least validate the existence of oppressive systems.

One common strategy against inferiorizing science is to claim that whatever difference researchers showed is a socially caused one. The idea is to show that the difference couldn’t be necessary if the difference was caused (exclusively or primarily) by some contingent factor. For example, if women are in fact worse at math than men, but that difference is caused by the sorts of extracurricular activities they are encouraged to join, or the toys they are encouraged to play with as children, then the difference seems to be contingent, rather than necessary. In section 4, I will evaluate the effectiveness of this strategy.

4 Recommendations for criticizing inferiorizing science

There are two ways to criticize inferiorizing science: point out a methodological problem to cast doubt on the claim that there is an actual difference, or show that the difference

---

9An example: Studies of toy preferences in young children (a year old or even younger) have found that there are differences between which toys female and male babies prefer (Fine 2010, p. 204-205, citing Alexander et al. 2009). Fine notes that actually the differences weren’t that large: baby boys still spent about 37% of their time playing with the feminine toy (a doll). Then she remarks,

Still, there are differences, and at first glance these findings seem to toll the bell for the idea that children’s gendered play preferences are purely socially constructed. The reason is that infants at this young age, so far as we know, are not aware of their own sex. They can’t therefore be basing their behavior on reasoning along the lines of I am a girl and girls do not play with trucks . . . Yet do these subtle differences reflect hardwired predispositions that differ between the sexes . . . ? Or do they reflect babies’ sensitivity to their social and physical worlds? (Fine 2010, p. 205).
(while perhaps actual) is not necessary. Which of these strategies is likely to be the most effective?

The methodological critiques are certainly effective. Methodological criticism is not particular to inferiorizing science; all scientific studies need to be methodologically sound. If scientists are unable to confirm a hypothesis – in this case, an inferiorizing hypothesis – then we shouldn’t believe the hypothesis is true. Thus the methodological critiques are straightforwardly effective at showing that we shouldn’t believe there is an actual difference between groups.

It is harder to claim that the second type of criticism, showing that there is not a necessary difference, even if it is actual, is effective. Unfortunately, by my evaluation, critics of inferiorizing science have too often confused the word “necessary” in P1* with the word “innate” or “biological.” Although I could problematize the idea that socially-caused differences are always contingent differences, I instead want to focus on the idea that biologically-caused differences are always necessary differences. This is not the case. Our biology is inextricably linked to our environment, including our social environment. Hormones are a good example of this: prenatal hormone exposure, for instance could be viewed as a part of the prenatal environment, although it is to some extent biologically caused. Environment can also have an effect on hormone levels throughout one’s lifetime. Are the resulting traits or behaviors biological or social? And does this make them necessary or contingent? I suggest the biological/innate versus social/environmental dichotomy is misleading and unhelpful for criticizing inferiorizing science.

That which is biological can be changed, and sometimes via social interventions. Where does this leave the critic of inferiorizing science? Showing that a difference is environmentally caused may help support the thesis that the difference is contingent. But the critic can also say that the difference is contingent while maintaining that the

---

10 Other epigenetic effects, like DNA methylation, are also good examples for showing that it is hard to distinguish biological and environmental causes.

11 For a lengthy critique of the so-called nature versus nurture debate along similar lines, see Pigliucci (2001).
difference is biological.

As a result, merely claiming that the purported difference is socially or environmentally caused is not likely to be an effective strategy, both because necessity and biologically caused are not equivalent and because the dichotomy between environmental and biological causes is a false one. There is another option available to critics of inferiorizing science, though: showing that some environmental intervention can remove the difference. This does have the effect of showing that the difference is not necessary (if it were necessary, it could not be removed). Note that there is a slight distinction between pointing to an environmental cause of the difference and showing that an environmental cause can remove the difference. I submit that the latter is a more effective critical strategy, because it points out circumstances in which the difference is not actual, thereby showing that the difference cannot be necessary.

5 Conclusion

The contributions of this paper are twofold. First, in section 2, I provided a plausible justification for criticizing inferiorizing science, namely, that it is important to reject $P1^*$ in order to reject $C$. Given that critics of inferiorizing science are trying to show that the group differences the science purportedly measured are not necessary, there are two ways to reject $P1^*$: showing that the difference is not actual (in which case it cannot be necessary), or showing that the difference is actual but not necessary. Section 3 gave examples of both of these strategies.

An example: Fine tells us that when large, real amounts of money are at stake, or when tasks are framed as “puzzle solving” instead of “mathematical” tasks, sex no longer predicts willingness to take financial risk, contrary to popular belief that men are prone to riskier behavior, and thus better suited to certain lucrative careers (Fine 2017, p. 158-159, citing Henrich & McElreath 2002 and Carr & Steele 2010). Similarly, when mental rotation tasks are framed as predicting artistic design ability, or as a task that women are generally better at, or when gender is deemphasized, differences between women and men at these tasks is erased (Fine 2010, p. 27-29, citing Voyer et al. 1995 for a meta-analysis, and Sharps et al. 1994, McGlone & Aronson 2006, Hausmann et al. 2009, and Moe 2009 for the altered studies that show no differences). One good explanation for these discrepancies is stereotype threat: when members of any gender are faced with tasks they expect themselves to do well or poorly at, they are more likely to perform accordingly. It’s possible that the original difference is a biological one, which is changed in the presence of different or unusual environmental circumstances (in this case, removal of the stereotype).
The second contribution is to argue that the former of these strategies – casting doubt on the claim that there is an actual difference, and doing so via methodological critiques – has been more effective than the latter. In section 4, I argued that showing that the differences are actual but not necessary cannot be adequately accomplished merely by claiming that the differences are socially or environmentally (rather than biologically) caused. The distinction between biological and social doesn’t line up with the distinction between necessary and contingent, and, furthermore, biologically caused traits are flexible and can often be molded by the environment. Therefore, critics of inferiorizing science need to use other strategies to claim that group differences are actual but not necessary. One suggestion is to demonstrate, with follow up studies, the effectiveness of some intervention at removing the difference. If the difference can be removed, it cannot be necessary.

In summary, my recommendations for critics of inferiorizing science are as follows: (1) Use methodological criticisms whenever possible, thereby removing the evidential weight of studies that claim to show there are group differences. (2) If one is trying to show that a group difference is actual but not necessary, do not merely suggest that the difference is socially or environmentally cased; rather, show that some intervention removes the difference.
References


Cambridge, MA: Harvard University Press.


