JOURNAL OF CONTROVERSIAL IDEAS

Article

Theory vs. Evidence: Unconscious Bias in Medical Decisions

Stewart Justman [1,*]

[1] University of Montana, Missoula, MT 59812, USA

* Corresponding author: stewart.justman@gmail.com

Submitted: 23 January 2024, accepted: 6 January 2025, published: 30 April 2025

Abstract: Racial disparities of medical care have been well documented for decades. That much is clear. Less clear are the cause or causes. By the time of the historic 2003 report by the Institute of Medicine (IOM), investigators had uncovered a pervasive pattern of disparities – often, however, without access to clinical or socioeconomic data which might help explain them. It was in these circumstances that the authors of the IOM report adopted the theory of unconscious bias as the ultimate explanation of observed disparities of care. The theory of a profound bias working outside the holder's awareness and control seemed to many to account for patterns of disparate treatment as nothing else could. However, to this day there exists little good evidence that such bias warps clinical decisions – certainly not enough to bear out the sweeping theory of a psychological mechanism that operates automatically. Impressive in principle but doubtful in practice, the theory of unconscious bias does not account for the evidence but covers its absence, just as it did when it was enshrined in the IOM report in 2003.

Keywords: Unconscious bias; racial disparities; unequal treatment; clinical decisions

How to cite: Justman, S. Theory vs. Evidence: Unconscious Bias in Medical Decisions. *Journal of Controversial Ideas* **2025**, *5*(1), 3; doi:10.63466/jci05010003.

1. An Impasse and Its Solution

In the latter 20th century, concerned investigators documented a pattern of entrenched disparities of medical treatment in the United States. In one notable case, upon reviewing over 1.7 million inpatient discharge abstracts, Harris et al. found, after controlling for age, insurance status, hospital characteristics, and (in part) disease severity, that "blacks were less likely than whites to receive major therapeutic procedures in 37 of 77 (48%)

conditions."[1] Such a pervasive pattern of differential treatment raised questions about the practice of medicine in American hospitals. For some a study like this must have constituted, in and of itself, an indictment of American medicine.

However, as suggestive as its findings are, and as carefully as it was conducted, the Harris study does not support such an indictment. Restricted by the information at their disposal, the authors lacked both clinical and economic data that could have had a bearing on patterns of treatment. Regarding clinical data, they note that their investigation was "limited by the availability of diagnostic information on the hospital discharge abstract." Similarly, although the study accounts for insurance status, "the groups could still differ in benefits and coverage levels, particularly among the privately insured, and in levels of supplemental insurance coverage. In addition, there could be other economic factors, such as ability to pay copayments and deductibles, that may restrict access to procedures for certain groups." If only because of these significant gaps in the data, Harris et al. do not allege racial bias. As they point out, a judgment of bias "may be supported only after all other potential factors have been examined and disproved," and in this case they were unable to eliminate other possible explanations for the observed differences of care.

At once suggestive and inconclusive, the Harris study exemplifies the state of the evidence in the decade preceding the historic report issued by the Institute of Medicine (IOM) in 2003: *Unequal Treatment: Confronting Racial and Ethnic Disparities in Healthcare* (which has been cited over 11,000 times). As noted in a review article appended to the digital edition of the IOM report, studies of treatment patterns conducted in the 1990s typically lacked access to detailed clinical records, and if only for this reason were barred by their own methods from drawing any definite conclusion about the cause or causes underlying their findings;[2] at best they could enumerate possible topics for investigation, some of which (such as deficiencies of communication between doctors and minority patients) later came to life in the literature.

Even studies that did consider clinical status were not necessarily able to eliminate other potential confounders. In the annotated bibliography included in *Unequal Treatment*, most studies of cardiovascular disparities (an emphasis of the literature) do not control for socioeconomic variables. Hence, for example, a study of some 275,000 patients with acute myocardial infarction which finds discrepancies of treatment (albeit no difference in hospital mortality) but acknowledges that the lower use of thrombolytic therapy for black patients may reflect "unmeasured confounders" including "cultural and socioeconomic factors."[3] Hence too a study of the differential care of black and white cardiac patients in a single setting, which concedes that "race may be only a surrogate marker for other socioeconomic factors (such as educational level, employment status, and family support structures) that may affect decisions about care to an equal or greater extent."[4] The literature had reached an impasse.

---

1 Harris et al., 1997: 91.

2 Geiger, 2003: 419. Subsequent page references for Smedley et al.'s *Unequal Treatment*, in which this chapter appears, are given in my text.

3 Taylor et al., 1998: 1022. Twenty years later it was still not possible to conclude by elimination that bias underlies the observed disparity between black and white patients with heart failure admitted to the ICU by a cardiologist. See Breathett et al., 2018.

4 Peterson et al., 1997: 485.

Given the literature's inability to account for confounders, it is misleading to argue, as *Unequal Treatment* does, that "When differences of treatment attributable to insurance, access to care, health status, and other factors are eliminated, racial and ethnic healthcare disparities still remain" (160).[5] At the time of *Unequal Treatment*, bias on the part of practitioners was not, in fact, an inescapable inference, which is why many studies did not propose bias even hypothetically. In these circumstances, the opportune arrival of the theory of unconscious bias allowed critics – prominently including the authors of *Unequal Treatment* – to presume or suggest a bias whose existence had not been established.

2. "Automatic, Unconscious, Ubiquitous"

Even as the medical literature of the 1990s found itself unable to determine the cause or causes of disparities of care, the theory of unconscious bias took shape in social psychology. In a formulation that would be adopted in full in *Unequal Treatment*, Greenwald and Banaji maintained that social behavior often proceeds "in an implicit or unconscious fashion"[6] and, more specifically, that racial stereotyping can and does operate just so. A year after the inconclusive study of 1.7 million discharge forms, Greenwald co-authored a report on the Implicit Association Test (IAT) which has now garnered over 17,000 citations,[7] the test itself having become a staple of the 21st-century literature on racial disparities, including those in medicine. (The test-taker pairs black or white faces and negative or positive attributes, with differences of speed indexing bias.) The IAT is liable to cogent objections,[8] and the authors of *Unequal Treatment* do not invoke it, possibly because it has nothing of evidentiary value to offer them. They do, however, adopt the theory of unconscious bias, and one can understand why. If the disparities literature of the 1990s was beset with evidentiary difficulties and found pervasive but not necessarily predictable patterns of disparate treatment, the theory of unconscious bias cut through such knots and offered a clear vision, one in which doctors were driven by a motive operating with the regularity of a mechanism, principle, or indeed law. If the Harris study is equivocal, unconscious bias is axiomatic.

Influential in importing the concept of unconscious bias into medicine was Michelle van Ryn, whose work is cited at many points in Chapter 4 of *Unequal* Treatment, concerned with the clinical encounter. According to van Ryn and Fu, racial stereotyping reflects "the automatic, unconscious, and ubiquitous nature of fundamental social cognition processes" to which doctors are subject.[9] A categorical claim like this differs in spirit from a meticulously empirical study like an inspection of discharge records. Harris et al. conclude that "[t]he inconsistency of findings across conditions implies that the factors that may be causing the observed racial and gender differences in procedure rates (e.g., unmeasured severity of illness, patient or physician decision-making, communication styles, stereotypes or racism/sexism), have different impacts depending

---

[5] Here and elsewhere, the reference is to the primary text of *Unequal Treatment*, as distinguished from the articles appended to the digital edition.

[6] Greenwald & Banaji, 1995: 4.

[7] Greenwald et al., 1998: 1464–80.

[8] Among these is that it measures the taker's familiarity with a stereotype, not whether or not the taker subscribes to the stereotype. A resource for criticism of the IAT is https://osf.io/74whk.

[9] Van Ryn & Fu, 2003: 252.

on the condition."[10] Thus, where disparities of care occur with puzzling irregularity (according to Harris et al.), unconscious bias works ubiquitously (according to van Ryn and Fu). For Harris et al. bias on the part of doctors represents one possible explanation among many for observed disparities of care, and as we have seen, it is their position that one can infer bias only after eliminating all rival explanations. By contrast, the belief that unconscious bias works like a law of our social nature seems more of a first principle than a conclusion to be reached in a case-by-case manner after methodically working through the evidence. In place of the uncertain conclusions of the Harris study it introduces a note of the absolute. "Automatic," "fundamental," "ubiquitous": these are formidable adjectives.

Unequal Treatment cites the disparities literature extensively, but supplements it with a theory of unconscious bias drawn from social psychology. It is as if the Harris study were purged of its limitations by the ambitious theory of an unconscious mechanism that operates all at once "systematically" (169), "automatically" (169, 173), and "universal[ly]" (170, 172). For the critic of racial disparities this theory has many attractions aside from its impressive language. It purports to be grounded in "normal and pervasive processes associated with social categorization," as *Unequal Treatment* puts it (173). It explains how doctors who sincerely abhor racism can act like racists despite themselves. And because it is phrased with the utmost generality and does not tie itself down to predictions, it can be reconciled with any reported result. Nonetheless, if doctors were driven by a mechanism as regular as unconscious bias is represented to be, one wonders "why differences occur in some conditions but not others" in the Harris study (100) and elsewhere. Moreover, if unconscious bias is directed not only against black patients but minority patients generally (as in the myriad references to "racial and ethnic" disparities in *Unequal Treatment*), why would "Asian people in the aggregate fare the same or better compared to White people for most examined measures" in surveys of health disparities even now?[11]

Even as the IAT attracted multitudes of takers, critics called into question the widely held assumption that unconscious biases like those purportedly measured by it translate into behavior.[12] *Unequal Treatment* makes the same dubious assumption without invoking the IAT. Although committed to the theory that unconscious bias does not just linger harmlessly in the mind of the doctor, the authors offer minimal evidence for the influence of bias over medical practice. At many points in the text a study by van Ryn and Burke is cited as evidence of that influence even though it offers no such thing; titled "The Effect of Patient Race and Socio-economic Status on Physicians' Perceptions of Patients," it includes the following disclaimer: "this study does not provide direct evidence regarding quality of care and it is unclear whether these differences in perceptions [of white and black, and affluent and non-affluent patients] are associated with differences in care or outcome."[13] Actual "direct evidence" of racially biased clinical practice in *Unequal Treatment* is thin. Of five studies cited in conjunction with van Ryn and Burke's in Chapter 4 of *Unequal Treatment*, only two actually consider clinical decisions, of which one yielded findings that await replication, according to the authors of *Unequal Treatment* themselves

---

[10] Harris et al., 1997: 100.

[11] See www.kff.org/key-data-on-health-and-health-care-by-race-and-ethnicity/?entry=executive-summary-introduction. The document is dated June 11, 2024.

[12] Blanton et al., 2009.

[13] Van Ryn & Burke, 2000: 825. For the erroneous claim that this study shows the impact of bias on clinical practice, see *Unequal Treatment,* pp. 166, 173, 200, 201, 426, 430–31.

(164). The study judged of greatest evidentiary value is clearly that by Schulman et al. concerning disparities in cardiovascular care – the only trial of the influence of bias on clinical decisions discussed at length in *Unequal Treatment*. However, once we look into this study and the controversy it evoked, we see that it exemplifies the use of the theory of unconscious bias not to account for the evidence but to cover its absence.

Something of a cause célèbre, the trial reported by Schulman et al. in 1999 in the *New England Journal of Medicine* saw the use of videotaped black and white actors to present identical histories and reports of chest pain to a total of 720 physicians.[14] The wording of the original publication suggests that blacks and women were 40% less likely to be referred for catheterization than whites and men, a disturbing figure picked up by the media. However, in the controversy that ensued, Schwartz et al. showed that in fact black men, white men, and white women were referred at an identical rate of 90%, while black women were referred at a rate 12% lower, for some reason.[15] Schwartz et al. professed themselves bewildered by this result, as well they should have been. As noted by the editors of the journal in a published clarification, the findings of the Schulman study "depended largely on the response to the 70-year-old black actress and, to a lesser extent, on the response to the 55-year-old black actress." The editors conclude that in the study as published "the evidence of racism and sexism was overstated."[16] Thus, the striking gap in referral rates proclaimed by the media reflected not the bias of the doctors but the misleading presentation of statistical findings and the abilities of two actresses. None of this prevented the study's authors from proposing, both in the original report and their response to Schwartz et al., that bias working below the threshold of consciousness dictated the disparity in referral rates – whatever it actually was.

Adopting the theory of a bias that operates autonomously like an internal auto-pilot, Schulman et al. note that

> Our finding that the race and sex of the patient influence the recommendations of physicians independently of other factors may suggest bias on the part of the physicians. However, our study could not assess the form of bias. Bias may represent overt prejudice on the part of physicians or, more likely, could be the result of subconscious perceptions rather than deliberate actions or thoughts. Subconscious bias occurs when a patient's membership in a target group automatically activates a cultural stereotype in the physician's memory regardless of the level of prejudice the physician has. (624–25)

Similarly, in their response to Schwartz et al., the authors comment that "with luck, our study will foster a more honest dialogue between physicians and patients and encourage the medical profession to seek ways to eliminate unconscious bias that may influence physicians' clinical decisions."[17] Thus, not even the finding that black men, white men, and white women were referred at an identical rate, and not even the public disclosure that their results were skewed by bad acting, could shake the authors' strong suspicion that unconscious bias was at work in the behavior of the study's participants and that a

[14] Schulman, Berlin, Harless et al., 1999: 624–25.
[15] Schwartz et al., 1999.
[16] Curfman and Kassirer, 1999.
[17] Schulman, Berlin, Escarce et al., 1999.

sort of conspiracy of silence has prevented doctors in general from recognizing that the same bias works in them.

Let it be said that a bias against black women but not black men – a bias, moreover, so deeply embedded in a cohort of doctors that it operates "automatically" – is ludicrous. For Schulman et al. it appears that unconscious bias represents a wild card that can explain any result, no matter how anomalous or absurd. It explained the results of their study before the critique by Schwartz et al., and it explained them afterward. And this is the trial that bears the greatest weight in the section of *Unequal Treatment* devoted to a discussion of the evidence for bias-driven clinical decisions. With studies like this as their building-blocks, the authors seem to know their evidence is too weak, as yet, to support a categorical indictment of the medical profession. Hence their practice of making tentative or preliminary accusations, as at the end of Chapter 4: "Bias, stereotyping, prejudice, and clinical uncertainty on the part of healthcare providers may contribute to racial and ethnic disparities in healthcare. While indirect evidence from several lines of research supports this statement, a greater understanding of the prevalence and influence of these processes is needed and should be sought through research" (178). With a sort of suggestive finding of biased care issued while gaps in the evidence remain, it appears that a determination of racial or ethnic bias is not a last resort after all. It is as if a verdict of "Guilty, but pending" had been issued in the middle of a trial.

The fact is that *Unequal Treatment* offers virtually no support for its own allegation that biased decisions by doctors are responsible for disparities of treatment. In fact, the authors concede that they use theory to make up for the lack of evidence on this critical point, acknowledging that "Unfortunately, little research has been conducted to elucidate how patient race or ethnicity may influence physician decision-making and how these influences affect the quality of care provided. In the absence of such research, the study committee drew upon a mix of theory and relevant research to understand how clinical uncertainty, biases or stereotypes, and prejudice might operate in the clinical encounter" (9). Later investigators who found no evidence of biased decision-making but sermonized about unconscious bias anyway continued the policy of tendentious speculation introduced in *Unequal Treatment*.

At the time of *Unequal Treatment*, the theory of implicit (that is, unconscious) bias was more impressive than the evidence, and so it remained. A systematic review published in 2017 notes that "the scientific community working in this area agrees that … there is currently a lack of good evidence for a direct negative influence of biases" on care.[18] And yet if unconscious bias exists, it should warp clinical practice precisely because it operates involuntarily. If it doesn't warp practice, what's all the fuss about?[19] A systematic review in 2018 concluded that "There is limited research examining the impact of implicit bias on patient care and outcomes. Most studies have been vignette-based and reveal mixed results, with four demonstrating a relationship between implicit bias and patient care/outcomes and eight finding no association."[20] Of the four supposedly positive studies, one consists of a dissertation that found a correlation of IAT scores and treatment "regardless of patient race"; another correlated IAT scores with "poorer

[18] Fitzgerald & Hurst, 2017: 19: 14.
[19] On the link between implicit bias and behavior, see Gawronski et al., 2022.
[20] Maina et al., 2018: 224.

anticipated therapeutic bonds [sic] with Black versus White patients among counselors"; and another, reported twice (as discussed below), was actually predominantly negative. It appears that to date there has been a total of exactly one successful trial of the influence of unconscious bias on clinical decisions. If unconscious bias operated as automatically and insidiously as its theory prescribes, surely studies of its influence on practice would not return such poor results.[21]

However, it seems the theory of unconscious bias and its malign influence does not stand or fall on the strength of the evidence. As we will see momentarily, two of the three founders of the IAT took part in negative trials of the influence of unconscious bias on clinical practice; in neither case, it seems, were they led to rethink the bias itself. So too, when the authors of *Unequal Treatment* call for "research to replicate [the Schulman] findings and explore how social cognitive processes may operate to … affect the structure, processes, and outcomes of care" (236–37), they fudge a bias that necessarily compromises care (because it is involuntary, and because the stresses of clinical practice make doctors revert to their default settings) into one that may or may not do so. This change having been made, they need not worry about the validity of the theoretically underlying bias if findings of biased practice should fail to materialize. In other words, a bias ambitiously described in their own pages in terms like "systematic" and "automatic" turns out to be compatible with any result, including null ones.

The behavior of unconscious bias in studies is actually strangely erratic. A 2007 study by Green and colleagues, including Banaji (a founder of the IAT), found a correlation between IAT scores and racially differential recommendations of thrombolytic therapy for acute coronary syndrome (in clinical vignettes).[22] With the authors declaring their study "the first evidence of unconscious (implicit) race bias among physicians, its dissociation from conscious (explicit) bias, and its predictive validity," readers at the time might well have felt they were present at the creation of a literature. However, the first successful study of the influence of unconscious bias on clinical practice was also the last.

3. Immunity to Evidence

In 2008 a team of investigators including Greenwald (a co-founder of the IAT) reported a study in which pediatricians were given case vignettes, with two black and two white patients, to gauge the quality of their recommendations for pain control, urinary tract infection (UTI), attention deficit hyperactivity disorder (ADHD), and asthma. On the IATs given afterward they showed some bias, though less than the norm. Despite the theoretically automatic operation of this motive, the authors found "no significant relationship" between measures of implicit bias and treatment, except in the case of UTI, for which black but not white children received the ideal option. The authors conclude,

> Inconsistent with our speculation and with research by Green et al., we did not find a relationship between implicit measures and quality of care. Future research will need to study a nationally representative sample of pediatricians to determine whether pediatricians, generally, hold less implicit race bias than other MDs and others in

[21] On social science that appears more sound than it actually is, see Honeycutt & Jussim, 2022).
[22] Green et al., 2007.

> society and if so why, and whether and under what conditions implicit racial bias may influence quality of care.[23]

Rarely in the literature is the theory of an unconscious bias that contaminates clinical practice conceded to be "speculation."

Seemingly convinced that some sort of positive finding could be squeezed out of their data, Sabin and Greenwald conducted "further analyses" which were published a few years later.[24] Flexible analyses, small studies, and hot fields have been flagged as risk factors for false findings.[25] All three are involved in the retake of the 2008 report.

The Results section of the retake's Abstract mentions only a positive result. The actual results are as follows:

> There were no significant associations between implicit attitudes and stereotypes about race and any of the treatment recommendation options for UTI, ADHD, and asthma. For pain, participants with greater implicit pro-White bias were more likely to agree with prescribing a narcotic medication for postsurgical pain for the White patient but more likely to disagree with prescribing it for the African American patient. (991)

While differences between recommendations of oxycodone by race were small, the implication appears to be that an entrenched myth about the abuse of opioids in black households got in the way of appropriate pain control. Nevertheless, in three of four instances there was no correlation between IAT-detected bias and the treatments recommended in racially parallel case vignettes. This was not what the authors anticipated. "On the basis of previous research," they note, "we expected that physicians' implicit pro-White biases might be related to poorer quality of care for an African American patient than for a White patient" (992). Why they retained this expectation after the negative findings of their own previous research remains unclear.

The relative absence of biased decisions in this report, and their absence in the original trial, are certainly not what we ourselves would expect if unconscious bias, albeit at modest or moderate levels, operated in the automatic manner described in *Unequal Treatment* and elsewhere. Moreover, the latter of the two articles yielded two paradoxical sub-findings: that pro-White bias was associated with poor pain treatment for White children[26] and better treatment of ADHD in children of both races. ("We speculate that there is an unknown variable to account for [the ADHD] finding," write the authors.)[27] Why would a motive working quite outside the party's volition and awareness fail to act in most cases and act paradoxically in some? Sabin and Greenwald are no more bothered by this anomaly than another founder of the IAT (Nosek) was bothered by the non-correlation between implicit bias and treatment recommendations in a later study,[28] or than Schulman and colleagues were bothered by the absurdity of a white bias against black women but not black men.

---

[23] Sabin et al., 2008: 684. The following year, two of these authors along with the third founder of the IAT argued that subconscious bias appears to influence clinical decisions. See Sabin et al., 2009.

[24] Sabin & Greenwald, 2012: 990.

[25] Ioannidis, 2005.

[26] "Physicians who demonstrated stronger pro-White bias on the Race IAT were more likely to agree with prescribing ibuprofen for the White patient (not the best option)." Sabin & Greenwald, 2012: 991.

[27] Sabin & Greenwald, 2012: 992.

[28] Oliver et al., 2014.

Schulman et al. lecture their critics about unconscious bias in the same issue and the same section of the *New England Journal of Medicine* in which the editors point out that they exaggerated "the evidence of racism and sexism" in their study. Sabin and Greenwald, at the end of the second of two articles offering, in sum, virtually no evidence of the influence of unconscious bias on clinical decisions – but several bewildering paradoxes – extol "the science of unconscious bias."[29] Instead of questioning the power of this supposedly automatic motive, they simply carve out open-ended exceptions to its scope of operation, concluding that "Implicit attitudes and stereotypes may not influence care for many chronic and acute pediatric conditions." Evidently, unconscious bias is consistent with a correlation, no correlation, or a perverse correlation with clinical decisions. Eleven years after *Unequal Treatment* a team of investigators including Nosek observed that "The conditions predicting when implicit biases will predict behavior or not are not yet fully understood,"[30] which appears to means that we are not only unable to predict clinical behavior from unconscious bias, we cannot even begin to do so. What exactly is the status of a bias which acts inexplicably and seems compatible with any and every outcome?

The theory of unconscious bias possesses a strange ability to withstand any showing, including several negative trials not discussed here.[31] Indeed, the physician-subjects in a trial that showed no effect of bias on recommendations for total knee replacement reported that their decisions *could* be influenced by bias and that learning about unconscious bias "could improve patient care."[32] The notion of unconscious bias has become so intuitive that it thrives even without confirming evidence, like a sort of anaerobic cultural organism. Given the theory's hardiness, it comes as no surprise that a study built on the concept of unconscious bias finds no connection between bias and treatment but still does not question the existence of the former.

To further the investigation of disparities of care in the Emergency Department, Musey and Kline ran a study in which doctors who saw black and white patients with chest pain estimated their risk for acute coronary syndrome (ACS). "Our hypothesis was based on the assumption that providers do not explicitly set out to treat patients differently based on gender or racial background but rather that there are a number of unconscious biases that continue to drive these disparities."[33] While the study finds that doctors underestimated the risk of ACS in black patients (a pattern consistent with other investigations), it also finds, paradoxically, no significant difference in either treatment or outcome between black and white patients. Grounding their study on the theory of unconscious bias, the authors report that the presumed bias had "no effect on provider behavior" (149). That is, after theorizing that doctors do not intend to treat patients differently based on race but end up doing so because they are under the influence of unconscious bias, Musey and Kline found that the doctors in their study did not end up

---

[29] Sabin went on to editorialize about "the science of implicit bias and evidence of its influence on clinical care" without mentioning two of her own studies that found weak evidence, if any, of the influence of implicit bias on care. See Sabin, 2022.

[30] Oliver et al., 2014: 185.

[31] E.g., Blair et al., 2014; Haider et al., 2011. On strained interpretations of the doctor–patient encounter, see Justman, 2024.

[32] Oliver et al., 2014: 183.

[33] Musey & Kline, 2017: 147.

treating black and white patients differently at all. But if the whole point of unconscious bias is that *it determines your behavior even as you go on imagining that you are quite free of it*, then it is hard to square the behavior shown in this study with the concept of unconscious bias. Indeed, according to theory, it is not so much people's "explicit" or conscious biases that predict their behavior as their "implicit" or unacknowledged ones.[34]

Perhaps in the course of making clinical decisions, doctors in the Musey and Kline study did not act in the automatic way one would respond to a cue on an IAT measured in milliseconds. As if recognizing this, Musey and Kline surmise that the doctors may have made a conscious effort to reduce disparities of care, perhaps for idealistic reasons, perhaps to avoid the "medicolegal consequences" of a missed myocardial infarction (149). Is it possible for someone to overcome his or her own bias? An article on the topic of stereotypes and their activation which has been cited almost as many times as *Unequal Treatment* itself finds that holding a stereotype does not necessarily mean acting on it. Using undergraduate psychology students as subjects, three experiments led Devine to conclude that a stereotype is like a bad but breakable habit.

> Automatic stereotype activation functions in much the same way as a bad habit. Its consequences are spontaneous and undesirable, at least for the low-prejudice person. For those who have integrated egalitarian ideals into their value system, a conflict would exist between these ideals and expressions of racial prejudice.[35]

Analogously, the disparities literature begins with the assumption that most doctors sincerely abhor prejudice. From there, however, it argues that prejudice continues to operate insidiously, beneath the level of consciousness.

The notion that people, including doctors, can stop and think about what they are doing is so pedestrian that it would not be worth citing social psychology to support it, but for two factors. One is that those who invoke the doctrine of unconscious bias to explain disparities of treatment and outcome contend that the conflict between espoused egalitarian ideals and actual bias simply ensures that the bias will remain covert, much as if it took on a life of its own once expelled from consciousness. Secondly, the Devine article happens to be the source given by Schulman et al. for the finding that "Subconscious bias occurs when a patient's membership in a target group automatically activates a cultural stereotype in the physician's memory regardless of the level of prejudice the physician has."[36] This phrasing misses the point of the Devine article, which is that automatic "activation" of a stereotype does *not* condemn the holder to behave accordingly. Habits can be broken. This in itself might explain a striking pattern in the disparities literature, namely, that doctors in general are judged to have the same level of implicit bias as the general population, and yet evidence of its adverse effect on their clinical decisions is elusive.

4. Split Personality

To this day there exists little good evidence that unconscious bias warps clinical practice – certainly not enough to bear out the sweeping theory of a psychological mechanism that

[34] See, e.g., Sabin et al., 2008: 678.
[35] Devine, 1989: 15.
[36] Schulman, Berlin, Harless et al., 1999: 624–25.

operates automatically in white practitioners of medicine. The shortfall of evidence has not deterred allegations of racism. Ever since *Unequal Treatment* proposed its theory of unconscious bias before, not after, rival explanations of disparate care had been sifted and eliminated, critics have been quick to accuse American medicine of a systematic bias against patients of color. Hence, for example, the oft-cited claim that black newborns suffer higher mortality when attended by white doctors. Years after this claim entered the literature and went into circulation, it emerged that the mortality difference approaches zero and loses statistical significance once very low birth weight is factored in.[37]

What is it about the theory of unconscious bias that seems so persuasive to critics of medicine that they embrace it regardless of the state of the evidence? Perhaps the appeal of this theory owes something to cultural clichés such as "Everyone has a dark side" or "The truth lies under the surface."

A half century ago the critic Lionel Trilling, in a discussion of "The Authentic Unconscious," looked into "the firmly entrenched belief that beneath the appearance of every human phenomenon there lies concealed a discrepant actuality."[38] If you seek the truth, look under the surface of propriety and convention, and in particular under the surface of the self. Under the propriety of bourgeois morality lies the turmoil of the unconscious; under the veneer of the medical profession's belief in equality lies unreformed, if also unwitting, bigotry. In the representative words of *Unequal Treatment*, "While it is reasonable to assume that the vast majority of healthcare providers find prejudice morally abhorrent and at odds with their professional values, healthcare providers, like other members of society, may not recognize manifestations of prejudice in their own behavior" (162). In the spirit of *Unequal Treatment*, the literature on unconscious bias views many, possibly most doctors as split personalities, professing respectable principles but impelled by ugly motives secret even from themselves.

Since its enshrinement in *Unequal Treatment*, the language of unconscious bias has become so customary that it is easy to forget how extreme it really is. As critics of the IAT have said, "If the race IAT reliably predicts discriminatory behavior that cannot be consciously controlled, then society should take note."[39] All the more should society beware of doctors who cannot control themselves.

Doctors under the influence of unconscious bias are said to behave in an automatic manner. They are in a dissociated state, unaware that they are acting out secret motives – virtually sleepwalking. According to *Unequal Treatment*, "when individuals do not have the time, capacity, opportunity, or motivation to assess situations fully and deliberately, implicit attitudes automatically shape people's responses to objects, individuals, and groups. These conditions of time pressure, high cognitive demand, and stress are common to many healthcare settings, making these settings 'ripe' for the activation of stereotypes" (173). In effect, because doctors are busy, they become automata. The prospect of doctors actually acting like automata, actually practicing medicine in a dissociated state, actually defaulting to their racist settings, even bringing about the death of black infants, this prospect is so monstrous that it would not be enough simply to criticize or

---

[37] Borjas & VerBruggen, 2024.
[38] Trilling, 1972: 141.
[39] Blanton, Jaccard, Klick et al., "Strong Claims and Weak Evidence: Reassessing the Predictive Value of the IAT": 567.

even reprimand such doctors. They would have to be removed from the practice of medicine immediately.

Funding: This research received no external funding.

Conflict of Interest: The author declares no conflict of interest.

References

Irene Blair, John Steiner, Rebecca Hanratty, et al. "An Investigation of Association Between Clinicians' Ethnic or Racial Bias and Hypertension Treatment, Medication Adherence and Blood Pressure Control". *Journal of General Internal Medicine* 29 (2014): 987–95. [CrossRef] [PubMed]

Hart Blanton, James Jaccard, Jonathan Klick, et al. "Strong Claims and Weak Evidence: Reassessing the Predictive Validity of the IAT". *Journal of Applied Psychology* 94 (2009): 567–82. [CrossRef] [PubMed]

George Borjas and Robert VerBruggen. "Physician-Patient Racial Concordance and Newborn Mortality". *PNAS* 121 (2024): e2409264121. [CrossRef]

Khadijah Breathett, Wehui Liu, Larry Allen, et al. "African Americans Are Less Likely to Receive Care by a Cardiologist During an Intensive Care Unit Admission for Heart Failure". *Journal of American College of Cardiology: Heart Failure* 6 (2018): 413–20. [CrossRef]

Gregory Curfman and Jerome Kassirer. "Letter". *New England Journal of Medicine* 341 (1999): 287.

Patricia Devine. "Stereotypes and Prejudice: Their Automatic and Controlled Components". *Journal of Personality and Social Psychology* 56 (1989): 5–18. [CrossRef]

Chloë Fitzgerald and Samia Hurst. "Implicit Bias in Healthcare Professionals: A Systematic Review". *BMC Medical Ethics* (2017): 18. [CrossRef]

Bertram Gawronski, Alison Ledgerwood, and Paul Eastwick. "Implicit Bias $\neq$ Bias on Implicit Measures". *Psychological Inquiry* 33 (2022): 139–55. [CrossRef]

H. Jack Geiger. "Racial and Ethnic Disparities in Diagnosis and Treatment: A Review of the Evidence and a Consideration of Causes". In Brian Smedley, Adrienne Stith, and Alan Nelson, eds. *Unequal Treatment: Confronting Racial and Ethnic Disparities in Health Care* (Washington, DC: National Academies Press, 2003).

Alexander Green, Dana Carney, Daniel Pallin, et al. "Implicit Bias among Physicians and its Prediction of Thrombolysis Decisions for Black and White Patients". *General Journal of Internal Medicine* 22 (2007): 1231–38. [CrossRef]

Anthony Greenwald and Mahzarin Banaji. "Implicit Social Cognition: Attitudes, Self-Esteem, and Stereotypes". *Psychological Review* 102 (1995): 4–27. [CrossRef]

Anthony Greenwald, Debbie McGhee, and Jordan Schwartz. "Measuring Individual Differences in Social Cognition: The Implicit Association Test". *Journal of Personality and Social Psychology* 74 (1998): 1464–80. [CrossRef] [PubMed]

Adil Haider, Janel Sexton, N. Sriram, et al. "Association of Unconscious Race and Social Class Bias with Vignette-Based Clinical Assessments by Medical Students". *JAMA* 306 (2011): 942–51. [CrossRef]

D. Robert Harris, Roxanne Andrews, and Anne Elixhauser. "Racial and Gender Differences in Use of Procedures for Black and White Hospitalized Adults". *Ethnicity & Disease* 7 (1997): 91–105.

Nathan Honeycutt and Lee Jussim. "Political Bias in the Social Sciences: A Critical, Theoretical, and Empirical Review". In C. L. Frisby, R. E. Redding, W. T. O'Donohue, and S. Lilienfeld, *Ideological and Political Bias in Psychology* (New York: Springer, 2022).

John Ioannidis. "Why Most Published Research Findings Are False". *PLoS Medicine* 2 (2005): 696–701. [CrossRef]

Stewart Justman. "What if Perceived Discrimination Isn't the Same as Real Discrimination?" *Journal of Controversial Ideas* April 2024. [CrossRef]

Ivy Maina, Tanisha Belton, Sara Ginzburg, et al. "A Decade of Studying Implicit Racial/Ethnic Bias in Healthcare Providers Using the Implicit Association Test". *Social Science and Medicine* 199 (2018): 219–29. [CrossRef]

Paul Musey and Jeffrey Kline. "Do Gender and Race Make a Difference in Acute Coronary Syndrome Pretest Probabilities in the Emergency Department?". *Academic Emergency Medicine* 24 (2017): 142–51. [CrossRef]

M. Norman Oliver, Kristin Wells, Jennifer Joy-Gaba, Carlee Beth Hawkins, and Brian Nosek. "Do Physicians' Implicit Views of African Americans Affect Clinical Decision Making?". *Journal of the American Board of Family Medicine* 27 (2014): 177–88. [CrossRef]

Eric Peterson, Steven Wright, Jennifer Daley, et al. "Racial Variation in the Use of Coronary-Revascularization Procedures". *New England Journal of Medicine* 336 (1997): 480–85. [CrossRef]

Janice Sabin. "Tackling Implic.it Bias in Health Care". *New England Journal of Medicine* 387 (2022): 105–7. [CrossRef]

Janice Sabin and Anthony Greenwald. “The Influence of Implicit Bias on Treatment Recommendations for 4 Common Pediatric Conditions: Pain, Urinary Tract Infection, Attention Deficit Hyperactivity Disorder, and Asthma”. *American Journal of Public Health* 102 (2012): 988–95. [CrossRef]

Janice Sabin, Brian Nosek, Anthony Greenwald, and Frederick Rivara. “Physicians’ Implicit and Explicit Attitudes About Race by MD Race, Ethnicity, and Gender”. *Journal of Health Care for the Poor and Underserved* 20 (2009): 896–913. [CrossRef]

Janice Sabin, Frederick Rivara, and Anthony Greenwald. “Physician Implicit Attitudes and Stereotypes about Race and Quality of Medical Care”. *Medical Care* 46 (2008): 678–85. [CrossRef] [PubMed]

Kevin Schulman, Jesse Berlin, José Escarce, et al. “Letter to the Editor”. *New England Journal of Medicine* 341 (1999): 286.

Kevin Schulman, Jesse Berlin, William Harless, et al. “The Effect of Race and Sex on Physicians’ Recommendations for Cardiac Catheterization”. *New England Journal of Medicine* 340 (1999): 618–26. [CrossRef] [PubMed]

Lisa Schwartz, Steven Woloshin, and H. Gilbert Welch. “Misunderstandings about the Effects of Race and Sex on Physicians’ Referrals for Cardiac Catheterization”. *New England Journal of Medicine* 341 (1999): 279–83. [CrossRef]

Herman Taylor, John Canto, Bonnie Sanderson, et al. “Management and Outcomes for Black Patients with Acute Myocardial Infarction in the Reperfusion Era”. *American Journal of Cardiology* 82 (1998): 1019–23. [CrossRef]

Lionel Trilling. *Sincerity and Authenticity* (Cambridge, MA: Harvard University Press, 1972).

Michelle van Ryn and Jane Burke. “The Effect of Patient Race and Socio-economic Status on Physicians’ Perceptions of Patients”. *Social Science and Medicine* 50 (2000): 813–28. [CrossRef] [PubMed]

Michele van Ryn and Steven Fu. “Paved with Good Intentions: Do Public Health and Human Service Providers Contribute to Racial/Ethnic Disparities in Health?”. *American Journal of Public Health* 93 (2003): 248–55. [CrossRef]