JUSTICE, HEALTH, AND STATUS:
MORAL THEORY AND THE NEW EPIDEMIOLOGY OF HEALTH DISPARITIES

Andrew Courtwright

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Philosophy.

Chapel Hill
2008

Approved by:
Advisor: Professor Thomas E. Hill, Jr.
Reader: Professor Rebecca Walker
Reader: Professor Geoffrey Brennan
Reader: Professor Jeanette Boxill
Reader: Professor Bernard Boxill
Health is distributed unequally. Compared to whites in my home state of North Carolina, the infant mortality rate among blacks is 2.4 times greater, Native American children are hospitalized for asthma 2.8 times more frequently, and black adults are 40% more likely to die from strokes. Ethical and political discussions of these and other health disparities have focused on justice. The general strategy philosophers like Norman Daniels and Jennifer Ruger employ is to identify a certain state of affairs—equality of opportunity and equality of capabilities, respectively—and then to argue that health disparities limiting an individual’s or group’s access to that condition are unjust, demanding intervention. The most common corrective in this context is to improve access to health care.

Recent work in epidemiology, however, has highlighted the importance of socioeconomic factors outside of access to care in creating health disparities. I explore the ways in which theories of justice have been expanded in light of this information, particularly data on the affects of social stigma on health. I suggest that further work in a broader moral territory—including the appropriateness of interpersonal attitudes like disesteem and contempt—is required if such theories are to provide an adequate framework for addressing health disparities. I conclude by considering an alternative approach, focused on a normative theory of socioeconomic status, for thinking about health disparities outside of the context of justice.
“To become effective in practice, every universalist morality has to make up for the loss of concrete ethical substance, which is initially accepted because of the cognitive advantages attending it.”

Jurgen Habermas, *Discourse Ethics*
# TABLE OF CONTENTS

LIST OF FIGURES ........................................................................................................ viii

Chapter

I. INTRODUCTION TO THE HEALTH DISPARITIES LANDSCAPE ........1

II. JUSTICE AND HEALTH...............................................................................9

Daniels and Health Care ........................................................................................12

Criticisms ...............................................................................................................19

III. THE SOCIAL DETERMINANTS OF HEALTH .......................................29

Access to Health Care ............................................................................................33

Health Behavior .....................................................................................................38

Income Inequality ..................................................................................................41

Other Candidates....................................................................................................45

Reverse Causation..................................................................................................47

IV. Autonomy and Health..................................................................................50

Autonomy ..............................................................................................................52

Some Candidates....................................................................................................54
LIST OF FIGURES

Figures

1. Effect one of introducing a minimal level of socioeconomic status (SES) ...... 146
2. Effect two of introducing a minimal level of socioeconomic status (SES) ...... 147
Chapter I: Introduction to the Health Disparities Landscape

One evening while I was on call at the hospital where I work, we admitted two very different patients in the early stages of heart attacks. Mr. Johnson was a sixty-seven year old white man with multiple risk factors for heart disease. His blood pressure and cholesterol, however, were well-managed by his primary care physician in Washington, D.C., where he worked in a high-level position at the National Institutes of Health. He had recently developed shortness of breath during a trip to Chapel Hill; his wife, concerned, brought him to our emergency room where an electrocardiogram and blood work showed changes consistent with a heart attack.

Mr. Franklin was a forty-six year old black man with poorly-controlled diabetes and high blood pressure. He worked occasionally for a landscaper. In his mind, visiting the doctor for chronic disease management was an unnecessary luxury. He had woken up that morning with vague chest pain that worsened throughout the day. In the late afternoon, he went to an urgent care near his house; the physician there sent him to our hospital after an electrocardiogram showed changes consistent with a heart attack.

Both Mr. Johnson and Mr. Franklin had severe blockages in the arteries to their heart, significant enough that open-heart surgery was indicated. Both men did well during surgery. Their postoperative courses, however, differed dramatically. Mr. Johnson spent a day in the intensive care unit (ICU) recovering from surgery and then moved to a normal hospital bed. He was up and walking within 36 hours. His wife, a nurse, encouraged him to work actively

\[1\] I have modified names and clinical information for these and other patients.
with physical therapy. He was discharged after five uneventful days. His plan was to finish recovering in his mountain cabin and then return to his job. Mr. Franklin, in contrast, spent three days in the ICU because of an unanticipated difficulty weaning him from the ventilator. The day after being transferred to a normal unit, a nurse noticed that his chest incision was infected, a complication more common among diabetics. Mr. Franklin remained in bed and was uncooperative with physical therapy; he subsequently developed pneumonia and had to be reintubated and transferred back to the ICU. Even when he was finally medically stable, his discharge was postponed because the social worker did not believe he had adequate support at home, should he have a late complication. The pattern these men represent—especially the connection between race and class and cardiac morbidity and mortality—is typical of heart disease in America.²

More generally, health disparities—defined by the American National Institutes of Health (NIH) as “differences in the incidence, prevalence, mortality, and burden of disease and other health conditions that exist among specific population groups”—occur between races, occupations, ethnic groups, and individuals of differing income in the United States.³ For example, the 2005 infant mortality rate for blacks in the United States was 13.9 per 1,000 live births versus 5.8 and 5.6 for whites and Hispanics, respectively.⁴ In 2001, the life expectancy at birth for whites was 77.7 years; for blacks, 72.2.⁵ American homosexual men are more than 5 times as likely to attempt suicide

---

³ National Institutes of Health, 2005.
⁵ Center for Disease Control, 2004, p. 33.
compared to heterosexual men.\(^6\) Twelve percent of children in the United States living in low-income families have an elevated level of lead in their blood, compared with 2% of children in high-income families.\(^7\) The relative risk of death for individuals from households in the United States making less than $15,000 a year is almost 300% that of individuals from households making more than $70,000 a year.\(^8\)

Compared to whites in my home state of North Carolina, the infant mortality rate among blacks is 2.4 times greater, Native American children are hospitalized for asthma 2.8 times more frequently, and black adults are 40% more likely to die from strokes.\(^9\) The overall health of North Carolinians—measured along a number of dimensions including total mortality and premature death—was 5.6% below the national average in 2005.\(^{10}\) *The burden of disease is distributed unevenly between populations.* Given that there are steps we could take to reduce these health disparities, is it morally acceptable for them to persist? A common way of interpreting this question is to ask whether it is unfair that some groups have better health than others.

Where fairness is concerned, we often appeal to a theory of justice to help us determine what responsibilities (if any) we have to alter the arrangements that produce inequalities and which mechanisms are appropriate for instituting these changes. Among politicians, physicians, and philosophers who believe that health disparities are unjust, the growing consensus is that we should expand access to medical care as a corrective

---


\(^7\) Brody, et al., 1994, pp. 277-83.

\(^8\) McDonough, et al., 1997, pp. 1476-83.


\(^{10}\) United Health Foundation, 2005.
measure. In the political arena, considerations of fairness have been used to argue for expanding preexisting programs like Medicaid or implementing large-scale, universal health care coverage as changes meant to address health disparities. Senator Edward Kennedy, in concluding a lengthy discussion of health inequalities, suggests that “lack of access and low-quality care are two major causes of these disparities.”\textsuperscript{11} He goes on to endorse a proposal to provide universal access to adequate health care. The Healthcare Equality and Accountability Act, which was introduced to the Senate and House in 2005 and enjoys the support of Senators Hillary Clinton and Barack Obama and current Speaker of the House Nancy Pelosi, “addresses health disparities by providing quality healthcare for minority patients across the country.”\textsuperscript{12} One of the central aims of this legislation is to expand minority medical coverage so as to correct differences in the prevalence and outcomes of conditions like diabetes, asthma, and heart disease. Similarly, the Healthy People 2010 Initiative and the Office of Civil Rights has made it a priority to reduce the unequal burden of disease by addressing “disparities in access, quality, and availability of health services.”\textsuperscript{13}

Medical groups and doctors also associate the (unjust) existence and persistence of health disparities with differences in access to care. For example, the Physicians Working Group for Single Payer National Health Insurance writes,

“Abolishing financial barriers to health care is the sine qua non of reform. Only a single comprehensive program, covering rich and poor alike, can end disparities based on race, ethnicity, social class, and geographic region that compromise the health care of the American people.”\textsuperscript{14}

\textsuperscript{11} Kennedy, 2005, p. 452.

\textsuperscript{12} Staunton, 2005.

\textsuperscript{13} Campanelli, 2003, p. 1624.

\textsuperscript{14} Physicians Working Group for Single Payer National Health Insurance, 2003, p. 801.
The Journal of the American Medical Association, in a call for papers on health disparities and access to health care notes that, “those who lack access to needed care, which may include inability to obtain primary care, chronic care, specialist care, or timely emergency or urgent care, are at risk for serious health consequences.”15 Antonia Novello—the former Surgeon General—quotes Thucydides on justice in her opening address to the 2004 conference on Minorities, the Medically Underserved, and Cancer, and then argues that differences in health insurance status play a leading role in causing health inequalities.16 By far the most cited aphorism in this context, however, is from Dr. Martin Luther King, “Of all the forms of inequality, injustice in health care is the most shocking and inhumane.”

The philosophical discussion of justice and American health disparities has focused on the role of government and private institutions in addressing health inequalities. The strategy of philosophers like Norman Daniels is to identify a certain desirable state of affairs and then argue that health disparities limiting access to that condition are unjust. So, for example, Daniels—whose account and subsequent revisions I will discuss in more detail later—argues that health disparities that limit equality of opportunity are unjust. And our responsibilities in the face of this injustice? Daniels writes that his approach “supports the provision of universal access to appropriate health care—including traditional public health and preventive measures—through public or mixed public and private insurance schemes.”17

15 Drummond, 2006, p. 2182.
17 Daniels, 1985, p. 27.
The growing consensus among politicians, doctors, and philosophers who believe that health disparities in contemporary America are unjust is that we have a pressing obligation to expand access to medical care. This has certainly been the focus of the last four major government initiatives in this area—the Clinton Health Care Security Act, the expansion of the State Children Health Insurance Programs, the addition of prescription drug benefits to Medicare, and Massachusetts’ mandated health care coverage initiative.

These efforts make sense if one believes that inequalities in access to care drive health disparities, a widespread and not implausible assumption. Of the approximately 45 million Americans without health insurance coverage, a disproportionate number are also members of groups with worse health. Among the uninsured, 14.6% are white, 19.6% are black, and 32.7% are Hispanic, while these groups make up 75.1%, 12.3%, and 12.5% of the total population, respectively. Low-income families make up almost 60% of the uninsured while 40% of people who did not graduate from high school lack insurance compared to only 10% percent of those who graduated from college. There appears to be significant overlap between populations without health insurance and those with comparatively worse health. And common sense seems to indicate that, if certain groups lack access to medical care, we should expect their health to suffer.

Ultimately, I will argue that this view is not borne out by current epidemiologic research, but there is an intuitive appeal to the idea that if Mr. Franklin had had the resources necessary to see a physician, his diabetes and blood pressure would have been

---

better controlled and his hospital course would have resembled Mr. Johnson’s. He might not even have had heart disease in the first place, especially at such a young age.

Empirical evidence aside, approaching health disparities from the perspective of justice may also lead us to emphasize the importance of access to care. The subject of justice is usually taken to be the arrangement of institutions that distribute certain goods and resources. If the poorer health of various groups is unjust, it is therefore important to change the institutions that distribute goods so as to reduce health disparities. The most obvious candidates for change, in this context, are the systems—primarily the Federal and state government and insurance companies—that determine who has access to health care. Since justice deals with the way that institutions distribute goods, it is not surprising that the political, medical, and philosophical work on justice and health disparities has demanded the reallocation of health care resources.

In the chapters that follow, I argue that excessive focus on justice has produced a misleading picture of the origins of health disparities and our obligations to reduce them. These disparities are not merely the product of institutional arrangements and the manipulation of goods and services like health care, but arise out of complex interactions including the effects of social stigma, negative interpersonal attitudes, and marginalization. As such, I suggest that an adequate normative approach to health inequalities should consider a broader territory beyond the traditional domain of distributive justice. I begin with a more detailed discussion of the ways in which theories of justice—particularly John Rawls’—have been applied to the problem of health disparities in the United States and other developed countries. This is the focus of the second chapter. In the third and fourth chapters I examine recent developments in the
epidemiology of health disparities, specifically studies suggesting that these disparities are rooted in a number of factors beyond access to care. In the fifth chapter I explore the ways in which philosophers have reexamined the requirements of justice in the light of these data, becoming more sophisticated in their approach. I argue in the sixth chapter that, thus far, these changes have been insufficient to adequately address the moral status of health disparities. In the seventh chapter I argue that work in a broader moral territory, focused on a normative theory of socioeconomic status, is a viable alternative to justice-based approaches. I conclude by showing how a specific conception of socioeconomic status, generated from Jurgen Habermas’ discourse ethics, can provide recommendations for reducing health disparities.
In 1997 Congress authorized spending on the State Children’s Health Insurance Program (SCHIP), a joint federal and state Medicaid-like initiative to provide health care access to approximately 5 million children from low-income families without health insurance. Enrollment was initially slow, forcing states to turn to local community organizations, which could better identify eligible families and encourage them to participate. In 2003 I joined the SCHIP initiative sponsored through the University of North Carolina Hospitals’ clinics, working with low-income parents to enroll their children in the program. Although these families still faced substantial barriers in access to health care—transportation, work, and sometimes language-related—the cost of care and prescription medication was removed as a significant concern. During the time I worked in the UNC clinics, I saw a dramatic increase in the number of SCHIP children seen for both emergent and preventative care. By 2007, when the SCHIP bill came to Congress for reauthorization, over 6.6 million children were enrolled.

Rather than simply reauthorize the program, however, Congress passed legislation to expand SCHIP by between 3 to 4 million participants. Funding for the expansion was provided by a combination of cuts to Medicare Advantage and a 61 cent tax increase on cigarettes, which was necessary to raise the 35 billion additional dollars the program would require. Critics of the SCHIP expansion argued that it represented unfair taxation, designed to take resources from two groups—the elderly and smokers—and use them to provide
another good (health care access) to a third group, low-income children. Is such an action unfair or does justice require the expansion of SCHIP, as the bill’s supporters argued?

The relationship between health, medical care, and justice is complex. No theory of justice has been explicitly introduced to address health disparities, so most work in this area has involved applying preexisting frameworks to questions of health and health care distribution. Norman Daniels has done this with John Rawls’ justice as fairness; the epidemiologists Michael Marmot and Jennifer Ruger with Amartya Sen’s justice as equality of capabilities. Before I turn to these accounts I want to point out several features they have in common. First, justice approaches begin by identifying a state of affairs as being of particular moral importance. In Daniels’ case, this is fair equality of opportunity; for Ruger, it is equality in the space of certain capabilities. They then argue that access to health resources, including public health interventions and medical care, is a necessary step toward achieving this condition, concluding that we ought to provide such access. For these theorists, health matters because it leads to some other normatively favorable condition (equal opportunity; equal capabilities) and access to health care matters because it leads to health. For example, Ruger writes, “Due to its influence on health, society must guarantee healthcare so that health improves overall and health inequalities attributable to health care diminish.”

---


24 Daniels indicates that there is a close conceptual connection between the role disease plays in limiting equality of opportunity in his position and the impact of poor health on capabilities within Sen’s account. The emphasis for both authors is on protecting health to establish a space of exercisable opportunities and may make the two positions distinct in terminology only. Daniels, 2001, p. 5-6. Daniels, et al., 2000, pp. 85-94. Ruger has argued that there is a substantive difference between Sen and Rawls on this issue. Ruger, 2006, pp. 408-10.

This is not to suggest that health, by itself, is unimportant within their frameworks. It is not even a comment on what sort of value health has in their systems; both accounts are compatible with health having non-instrumental value (i.e. being valuable without contributing to or producing another condition).26,27 Rather, for both Ruger and Daniels, health is a target for justice only insofar as it contributes to another state of affairs, be it equality of opportunity or capabilities.

Second, because theories of justice cover a range of distributable goods, the direct and opportunity costs of pursuing certain health/health care goals have to be weighed against policies involving other goods that also contribute to equality of opportunity or capabilities. For example, the political and economic capital expenditures required to expand health care coverage in contemporary America have to be considered against gains in equality obtained through other mechanisms like increasing differential tax gradients or expanding education programs. This is not a criticism of using theories of justice to address health disparities; rather, it is a reminder that, where justice is concerned, health is only one variable that contributes to (in)equality.28 As such, any theory of justice—including Sen’s and Rawls’—will make specific recommendations about health care only within the context of distributing other goods.

26 The same is not true for most utilitarian philosophers who discuss health disparities. Within the utilitarian system, health only matters insofar as it is an element of another state of affairs (e.g. well-being, absence of pain, presence of pleasure) that has value. In this context health necessarily has non-intrinsic value. See Stern, 1983, pp. 339-62. Savulescu, 2001, pp. 39-41.

27 Ruger suggests that the capabilities approach “values health intrinsically” and criticizes Daniels for treating health as a non-intrinsically valuable good. Ruger, 2004a, pp. 1092-3.

28 The philosophers Sudhir Anand and Fabienne Peter do take this point to be a criticism of Daniels. Their concern is that other aspects of Rawls’ theory will come in conflict with what Daniels sees as the demands of the fair equality of opportunity principle with regard to health care. See Anand and Peter, 2000, pp. 48-52.
With these general comments in mind, I want to turn to Daniels’ early discussion of justice and health disparities. I have chosen to focus on Daniels rather than on the political or medical debate on these issues because he fully articulates a common way of arguing for redistributing health resources. He has also worked to refine his theory as new epidemiological evidence has become available, a project that will later allow us to examine what moves are available when reconsidering the demands of justice in the face of changing epidemiological information about health disparities.

Daniels and Health Care

In 1985 Daniels published *Just Health Care*, which was, then, one of the few book-length treatments of the relationship between justice and health care distribution. Daniels makes two claims: first, a certain subset of health disparities is unjust and, second, we should change our institutions to ensure a more equal distribution of health care. Daniels finds support for the first point in John Rawls’ *A Theory of Justice*. It is important, therefore, to sketch a brief overview of the Rawlsian framework to explicate Daniels’ starting assumptions. In *A Theory of Justice*, Rawls is concerned with the basic structure of social institutions that distribute certain types of goods. In broadest outline, he argues that justice as fairness, a conception of justice derived from a contractarian procedure centering on hypothetical agreement, generates principles whose application best matches with our considered judgments about distributive justice. Institutions governed by these principles, he suggests, would also be plausibly stable if established in a society willing to follow them.

---

29 The other common approach is simply to assert a right to health care (or, even less coherently, to health) and argue that it is inappropriate to treat health care as a commodity where market or ability-to-pay mechanisms should determine access. This approach is unsatisfactory given that rights are not *de novo* moral entities, brought into existence by merely asserting them. They must be grounded in a framework of principles. Thus, if we are to make sense of rights talk, we must first settle on the appropriate moral principles (e.g. principles of distributive justice) and then decide on the existence of subsidiary rights like the right to health care. See Daniels, 1985, pp. 4-9. For a critique of efforts to ground a right to health care in Rawls’ principles of justice see, Flic, 2007, pp. 51-72.
Rawls sees the principles of justice needed for determining the fair distribution of goods as being generated from an ahistorical, hypothetical agreement between different parties within what he calls the original position. The original position is, most importantly, a device for representing agents as free and equal persons—situated in identical positions of power and abstracted from morally irrelevant features—given the goal of understanding the substantive results of such a conception. Some of the most important characteristics of agents in the original position include: first, they are behind a veil of ignorance, which prevents them from knowing what their place will be in the society ruled by the principles on which they agree. Second, Rawls assumes that they are mutually disinterested with the goal of maximizing the amount/number of primary goods for themselves and the generations they represent. Primary goods, as Rawls understands them, fall into four groups: basic rights (freedom of speech, assembly, etc.); non-basic rights and prerogatives (especially those attached to specific social positions); wealth and income; the social bases of self-respect. Third, the contractors are rational in that they have a capacity for a conception of the good, which, while it does not have specific content, serves to direct their decisions with respect to primary goods. The primary goods, in turn, are normally useful for pursuing any specific conception of the good. Rationality also implies, according to Rawls, the ability to revise one’s conception of the good as well as a commitment to pursuing the necessary means toward our ends. Fourth, Rawls’ contractors are thought to have a capacity for a sense of justice (they are “reasonable”), which matters for certain questions concerning institutional stability.

Given this set of conditions—first, an idea of persons as mutually disinterested with a capacity for a sense of justice, a capacity for a conception of the good, and the goal of
maximizing their share of primary goods; second, a position in which participants are veiled from knowing what place they will occupy in the resulting society—Rawls considers what principles for the distribution of primary goods would emerge. He suggests two, with a lexical ordering such that the second does not come into play until the first has been satisfied (excepting certain conditions in which there are not enough resources). The first is the liberty principle: each person has an equal right to a fully adequate scheme of basic rights and liberties, which scheme is compatible with a similar scheme for all. The second is the equality of fair opportunity/difference principle: social and economic inequalities are permissible so long as, first, they must be attached to offices and positions open to all under conditions of equal opportunity and, second, they must be to the greatest benefit of the least advantaged. There is also a lexical ordering between the two parts of this principle.

Rawls’ conception of justice is explicitly tailored to set aside issues of health and health care, at least at the level of the highest principles. Partly, this is because he thought of health as a natural—as opposed to social primary good—which has different implications for its distribution. Rawls writes, “health and vigor, intelligence and imagination are natural goods; although their possession is influenced by the basic structure [of institutions], they are not so directly under its control.”30 While Rawls acknowledges the importance of health in the pursuit of a conception of the good, because social institutions at best indirectly influence its distribution, health was not a focus of his account, at least compared to the social primary goods. As the economist Robert Sugden notes, within the Rawlsian framework, “what has to be distributed justly—or fairly—are the benefits and burdens of social cooperation . . . as natural primary goods are not the product of social cooperation, the question of whether they

30 Rawls, 1999, p. 54.
are distributed fairly or unfairly does not arise within a theory of fair cooperation."31 In later writings, Rawls was concerned that focusing on individuals with poor health or disability might “distract our moral perception by leading us to think of people distant from us whose fate arouses pity and anxiety.”32 Rawls goes on to suggest that we postpone questions involving, for example, special health needs until we address certain higher-order questions concerning the principles of justice.33

For his initial purposes, then, Rawls thought of the contractors as a healthy population without disease or significant disability. The concern with this simplifying assumption, however, is that differences in health in our world translate into differences in equality of opportunity, even when the other primary goods are distributed fairly. Or, to put it a different way, in ignoring health in constructing his principles of justice, Rawls’ runs the risk of generating a theory that fails to account for a central injustice, health disparities.34 The most obvious solution to this problem—including health or health care to the list of primary goods—has been subject to a number of criticisms, most importantly from the economist Kenneth Arrow.35

First, Arrow points out that, since the difference principle requires that inequalities in primary goods work to the advantage of the worst-off, very sick individuals with chronic diseases would be entitled to the resources necessary to keep them alive, even if doing so represented a great cost to the interests of others. This is the problem of the so-called “health

32 Rawls, 1975, p. 96.
34 For an extensive development of this criticism, see Nussbaum, 2006, pp. 15-34.
care sinkhole.”

Second, Arrow argues that any attempt to draw a line in the amount of health care to provide would require weighing health against the distribution of other primary goods (wealth, income, etc.), necessitating interpersonal comparisons of utility, which Rawls hoped to avoid. A number of responses have been made to Arrow’s second point, and, since it rests more on the fact that the list of primary goods contains multiple items and less on any specific problem with adding health, it is less relevant to the discussion here. Arrow is, however, right to insist that, no matter how important health is to our overall well-being, adding it directly into the theory as a resource for distribution creates unnecessary problems for the Rawlsian approach. One of Daniels’ central aims in *Just Health Care* is to incorporate health and health care at a different level of Rawls’ framework.

Daniels begins his account by distinguishing a certain set of needs that should receive priority in any distributory scheme involving social goods. Such needs must meet two criteria: first, they must be *objective ascribable* in that they are true of an individual even if she refuses to recognize them. For example, it may be true of me that I need cholesterol-lowering medicine even if I explicitly deny it or have a strong preference to not take such medicine. The second characteristic of Daniels’ privileged needs is that they are *objectively important*—“we attach a special weight to claims based on them in a variety of moral contexts.” In cashing out this criterion, Daniels rejects a definition of importance whereby prioritized needs are those that are required to achieve our specific goals (i.e. things we need no matter what else we need). Instead, he suggests that objectively important needs are those required to maintain normal species function over time. Species typical functioning, in turn,

---

36 See Kitcher, 1996.


38 Daniels, 1985, p. 25.
is defined in terms of an evolutionary account of the “design” of human beings including facts about our physiology as well as our “acquisition of knowledge, linguistic communication, and social cooperation in the broad and changing range of environments in which we live.”

Daniels’ argument, then, is that those resources necessary for species typical functioning ought to be thought of as distinct from other sorts of goods, especially ones whose distribution we leave largely in the hands of free-market structures. What identifies these resources as different is that “impairments of normal species functioning reduce the range of opportunity open to the individual in which he may construct his ‘plan of life’ or ‘conception of the good.’” Life plans—an idea that Daniels adapts from Rawls—represent goals around which we schedule our behavior so as to satisfy a certain range or even all of our desires. When our normal species function is impaired, we are denied the opportunity to pursue certain life plans to which we would otherwise be suited. In other words, restrictions in normal species function curtail the normal opportunity range available to us within our specific social situation.

At this point, Daniels and Rawls begin to dovetail. Recall that Rawls’ fair equality of opportunity principle requires institutions to distribute goods in such a way that the occupations to which these goods attach are open under conditions of fair equality of opportunity. As Daniels interprets this principle, its primary purpose is to minimize the effects that fortunate accidents of birth have in shaping individual opportunity. Such “natural” advantages include the set of social conditions into which one is born—being a White Afrikaan in 1970s South Africa or being a member of the royal family in 17th century

---

39 Daniels, 1985, p. 29.
40 Daniels, 1985, p. 27.
Spain—as well as certain genetic facts—one’s height or number of fast-twitch muscle fibers. It is not enough, in this regard, to eliminate selection processes that are based solely on, for example, race, ethnicity, or gender so as to remove barriers from equal opportunity employment. Adopting such a negative constraint would only meet what Rawls calls *formal* equality of opportunity. Rather, we are also obligated to adopt institutional arrangements that redistribute resources like education, wealth, and income in such a way as to counter the disadvantage that social/genetic factors like race, class, family, or physiology can produce in a given context. As Daniels points out, the fair equality of opportunity account does not require institutions to level all differences between persons or even to give everyone *equal* shares in the normal opportunity range. Rather, wherever possible we are to allocate resources so as to guarantee that “all persons are equally spared certain kinds of impediments to opportunity.”\(^{41}\) Any consequent inequalities based on differences in, for example, talents or skills are permissible within the constraints of the difference principle. That is, inequalities that arise from variations in talents or motivations—assuming fair equality of opportunity—are permissible within the Rawlsian framework so long as they work to the advantage of the least well off.

Given this account of fair equality of opportunity, if a specific subset of goods is necessary to have a chance at achieving certain positions, then they must be made universally available within the constraints of the liberty principle. All that remains to complete Daniels’ argument is to establish health care resources as belonging to this special subset of goods. He does so by defining health as “the absence of disease,” and disease as “deviations from

---

\(^{41}\) Daniels, 1985, p. 52.
the natural functional organization of a typical member of a species.” Thus diseases impair normal species functioning, which, in turn, Daniels argues, limits the range of opportunity. Since health is best secured through adequate health care resources, these goods should be distributed in such a way as to restore conditions of fair equality of opportunity. As with limitations in our space of opportunity due to racism or classism, fair equality of opportunity requires us to restore or prevent disease-caused impairments in normal species function. Doing so helps to ensure that we are given the opportunity to pursue certain life plans to which we would otherwise be denied because of ill health.

To summarize: based on the Rawlsian fair equality of opportunity principle and an account of the importance of a certain set of goods—which includes health care resources—for equal opportunity to pursue a life plan within a specific social context, Daniels argues that health care ought to be distributed in such a way so as to reduce the unequal burden of disease. Doing so will prevent differences in normal species functioning (i.e. differences in health) from being translated into inequality of opportunity. As already mentioned, he concludes that we have an obligation to design institutions that provide universal access to appropriate health care, where that includes chronic- and acute-care medicine as well as preventative and public health services.

Criticisms

There are three principle areas where critics have pressed Daniels’ argument. From narrowest to broadest they include: Daniels’ definition of disease, his account of equality of opportunity, and his use of the Rawlsian framework. First, Daniels relies on his definition of disease along with the assumption about the connection between diseases and health resources to place health care goods in a special category for distributive purposes. Because

---

42 Daniels, 1985, p. 28.
his account of disease plays such a critical role, it is worth pausing to note just how controversial it is. Daniels understands disease based on the ‘biomedical’ model offered by Boorse in a series of articles written in the 1970s and refined subsequently. The problem with any of these strictly scientific approaches is that they ignore (or, more charitably, deemphasize) what many have come to see as an essential component of diseases; namely, their evaluative aspects. Part of what makes something a disease, the argument goes, is that it is an unwanted or undesirable state viewed either in terms of a social or personal judgment.

This is not an appropriate venue for taking on the vast and quagmireish literature that has sprung up around various definitions of diseases, including Daniels’. Part of the problem is that the whole enterprise seems to be an ill-conceived application of ordinary-language philosophy where necessary and sufficient conditions for the concept ‘disease’ are plotted out by repeated rounds of example and counterexample. At best, this approach has revealed our intuitions about what sorts of things are diseases to be immature and confused. Assuming, however, that Daniels can defend his strictly scientific approach to disease, and it is hard to imagine anyone still willing to take on that task, there are other problems in his definition. He will, for a start, owe us an account of what it is to be a member of a species, no small project considering the ongoing controversy in philosophy of biology as to whether there are even such natural kinds as species. In addition, he will have to generate an account of function that does not depend on evaluative judgments, a goal that has also frustrated

---

43 See, for example, Boorse, 1975, pp. 49-68. Boorse, 1977, pp. 542-73.


45 For an excellent summary, see Peterson, 2001.
philosophers of biology. Finally, he will have to address a host of questions about the relationship between the treatment of disease to restore species typical function versus enhancing that function.

These problems are not limited to Daniels’ original work. In his most recent reevaluations of his position, which I will discuss later in more detail, the concept of disease still plays an essential role and still looks much the same. He writes, “Disease and disability, both physical and mental, are construed as adverse departures from or impairments of species-typical functional organization or ‘normal functioning,’ for short.” Whether or not Daniels can support this claim, several of his critics have taken its implausibility to be prima facie evidence against the position.

The second criticism of Daniels’ project concerns his interpretation of equality of opportunity. An early critique by Allan Buchanan ascribed to Daniels the position that we should allocate social resources so as to guarantee the normal opportunity range within a social system, regardless of an individual’s skills or talents. The worry was that equality of opportunity either requires us to bring everyone down to the skill/talent level of the lowest member or devote excessive resources toward bringing the most untalented up to everyone else. As I have already indicated, however, Daniels’ extension of the equality of opportunity principle applies only to certain impediments to opportunity; namely, the way that disease (or poor health) limits normal species functioning, reducing an individual’s normal

---

46 Daniels himself explicitly declines to provide such an account. See Daniels, 2001, p. 3 fn1.


48 Daniels, 2001, p. 3 fn1.


50 See Buchanan’s discussion of “Principle D” in Buchanan, 1984, pp. 55-78.
opportunity range. Being untalented is compatible with having perfectly typical species function; in fact, it is the condition in which most of us find ourselves. Similarly, it would be difficult to argue that unskilled laborers are species atypical or that individuals with ho-hum motivational states are malfunctioning from an evolutionary perspective. To a certain extent, this challenge returns to the question of whether being untalented is a kind of disease on Daniels’ definition, but on a reasonable interpretation of that view, the impact of talents, motivations, or skills fall outside of his purview. Such characteristics, although they impact opportunity, are not, according to Daniels, the proper subject of redistributive aims.

More recently, Francis Kamm has challenged Daniels’ interpretation of equality of opportunity and its utility in addressing health disparities. Although some of her concerns are about the lexical ordering of the equality of opportunity and difference principles—a discussion too far a field for the purposes of this chapter—her primary question is whether ensuring normal species function (or health) is necessary for equality of opportunity.51 Kamm begins by arguing that Daniels’ position is compatible with a social system that guaranteed equal unhealth to all its citizens; that is, even if everyone had the same range of poor species function, they would still satisfy the requirements of equality of opportunity. If this is true, then, using equality of opportunity as a standard, we should be indifferent between social arrangements in which citizens are guaranteed equal health or unhealth. But, Kamm argues, our intuitions about justice suggest that the former is preferable. Insofar as Daniels’ account fails to meet this expectation, it shows the inability of equality of opportunity to adequately address health disparities.

The problem with Kamm’s argument is that it misplaces the reason why health matters within Daniels’ scheme. As the discussion of Buchanan’s “leveling down” objection

illustrates, Daniels’ concern is with whether institutions should allow certain factors to influence our ability to pursue life plans. Ideally for Daniels—and Rawls, as Daniels understands him—our skills, talents, and motivations determine the kind of person we become, not discriminatory practices in hiring or education (as would exist without formal equality of opportunity) or the absence of compensatory redistributions for certain kinds of historical discrimination (as would exist without fair equality of opportunity). The argument is that equality of opportunity is established when ill health (among other variables) ceases to be an impediment to our normal range of species function, a standard based on our membership in a certain natural kind, not our specific social context. In a society bent on egalitarian unhealth, we would all be equally unable to secure that range because we were not functioning within our normal species range. This is, however, not the same thing as having equality of opportunity. Rather, Daniels’ account would require us to redistribute health care resources so as to ensure—as far as possible and consistent with the liberty principle—that the citizens of that society enjoy the same range of species typical function. Or, to put the point slightly differently, equally absent from disease.

Third, a number of philosophers have challenged the appropriateness of the Rawlsian framework for addressing health disparities. These have included communitarians, who have argued for distribution schemes based on shared local values,\textsuperscript{52} libertarians, who focus on free-market mechanisms to satisfy preferences for varying degrees of health care,\textsuperscript{53} and Aristotelians, who emphasize the importance of virtues outside of justice in confronting inequalities.\textsuperscript{54} The most persistent objection to Daniels’ project, however, has come from


\textsuperscript{53} Englehardt, Jr., 1997, pp. 180-96.
philosophical proponents of the “decent minimum” approach to health care. Several authors have thought that the chief aim of health care ethics should be to articulate and defend a right to a decent minimum of medical care rather than examine the effects that health has on a condition like equality of opportunity. Daniels, in contrast, explicitly denies that his argument establishes a universal right to health care, at a minimal level or otherwise.

Allen Buchanan offers three reasons why we should prefer the decent minimum approach. First, a decent minimum of health care is society relative; that is, the demandingness of the right varies with the resources available in a specific social context. This allows the content of the right to be determined depending on the environment in which it is established. Second, a decent minimum avoids the problem of trying to guarantee access to the best health care available. It assures a certain level of health services but allows citizens to purchase care given their preferences. Third, the idea of a decent minimum encompasses enough health care to support a tolerable life within one’s specific social context. To this list, we might also add a political consideration; namely, rights to health or health care are entrenched in popular discourse through documents like the United Nations Declaration on Human Rights. A defense of a decent minimum of health care would be able to draw on this background in generating support for its practical realization.

The two primary arguments for a right to a decent minimum have come from utilitarianists like Julian Savulescu, and from Buchanan, who has offered a pluralistic approach.

56 Daniels, 2001, pp. 53-4.
57 See also Hessler and Buchanan, 2002, p. 84-101.
58 Gruskin and Braveman, 2006, pp. 405-417.
to deriving the right. Savulescu’s arguments turn on his contention that the goal of medical services is well-being and not to cure disease or promote health (normal species functioning) as Daniels and others have thought. Health care is important because it is necessary to achieve a minimally decent life (i.e. one with a certain level of well-being). Given the centrality of health and health care for our well-being, Savulescu argues, medical resources should be distributed so as to ensure that the maximum number of people have a decent minimum of care. To put this idea in more traditional utilitarian terms, we should allocate health care resources so as to maximize the number of people who enjoy a decent level of well-being. Further differences in well-being are tolerated, on Savulescu’s view, so long as they do not interfere with maximizing the decent minimum.

There are two primary problems with this account. First, Savulescu confuses an outcome of medicine—it promotes well-being—with its goal. If the purpose of medicine were to advance patient well-being, then doctors would be obligated, among other things, to find dates for their patients. This is too expansive. Even if we accept that the goal of medicine is not to cure disease, it must be more circumscribed than Savulescu acknowledges to avoid collapsing the distinction between health care and other services. A more sophisticated account might argue that the purpose of medicine is to promote patient well-being over a certain range of hindrances (i.e. diseases) although I suspect that this will bring us uncomfortably close to Daniels’ position. If the hindrances in question are those that inhibit species typical functioning, then the only difference between the accounts will be Daniels’ focus on the ability to pursue one’s life plan versus Savulescu’s focus on procuring a life of minimally decent well-being.
A separate worry with Savulescu’s account is that it fails to support a universal right to a decent minimum. Given that the goal of his account is to maximize the number of people who enjoy a decent level of care, this leaves open the possibility that not all citizens will have this minimum. This is a specific instance of a more general problem with utilitarian derivations of rights; namely, they cannot account for the fact that rights seem to exist, in part, to protect us from certain utilitarian calculations. If an individual has a right to a decent minimum of health care, this is commonly thought to entail that his health care ought not to be sacrificed for the sake of establishing a decent minimum for some larger aggregate of others. The force of this criticism largely depends on how committed Savulescu is to the language of rights; his project is still coherent without the attempt to derive a right to a decent minimum of health care and he articulates it, in at least some places, without explicitly doing so.\textsuperscript{59}

Allen Buchanan rejects Rawlsian and utilitarian approaches to justifying a decent minimum in favor of grounding the right in a collection of distinct principles.\textsuperscript{60} First, under the principle of reparation, he notes that some groups (African Americans, Native Americans) are due a decent minimum of health care because of past harms/injustices. Individuals who have been harmed by the government or corporations may be similarly owed health care along with individuals, including soldiers, who have made sacrifices for the general social welfare. Under the principle of preventing harm, Buchanan argues that traditional public health services ought to be universally available as a corrective for the harms (e.g. poor sanitation and infectious disease) that occur when groups of people live together. Under a principle of prudence, Buchanan notes that it is in the interests of a social

\textsuperscript{59} Savulescu, 2001, p. 39.

\textsuperscript{60} Buchanan, 1984, pp. 67-72.
system to have more productive citizens, both for economic and defense purposes, and that a decent minimum of health care is important for such a goal. Finally, Buchanan notes that, if we consider health care to be a public good, a principle of enforced beneficence would be justified to introduce taxation with the goal of providing a decent minimum of care for all citizens.

Putting aside the question of whether health care is a public good, whether it is actually in the economic interests of a social system to provide a decent minimum of health care, and the conditions under which it is permissible for states to enforce moral duties of beneficence, the most general problem with Buchanan’s approach is its mixed derivation. He employs a selection of principles that include both negative and positive duties—obligations not to interfere/harm and to help/benefit, respectively—where the former are traditionally understood to be stronger than the later. Given limited medical resources, we require a prioritization scheme between various citizens whose rights are derived within different contexts. For example, if a limited resource is necessary for a minimally decent life—for example, neonatal erythromycin eye drops to prevent vertically-transmitted, *Chlamydia trachomatis*-associated blindness—should it go to the children of soldiers or to those who contribute to the enforced beneficence scheme? By deriving the right to a minimum of care from multiple sources, Buchanan will have to supplement his account with a weighing mechanisms for when the demands of his subsidiary principles conflict. This may be achieved, but it complicates both the philosophical and practical picture significantly. All else being equal, a position with a single starting place—like Daniels’—may be preferred, even if it means leaving behind the decent minimum framework.
Despite these objections to his approach, Daniels’ account has remained a touchstone for the philosophical debate over the requirements of justice in addressing health disparities. In 2001, however, he offered a modified version of his position in the article “Justice, Health, and Healthcare,” which was subsequently expanded into a new book on the topic, *Just Health: Meeting Health Needs Fairly*. The impetus for his revisions was not any definitive argument against the Rawlsian approach; rather, developments in the epidemiology of health disparities had called into question the relationship between access to medical care and the persistence of health inequalities. In order to understand and assess Daniels’ revised position, it is necessary to review this epidemiological research, a task to which I turn in the next chapter.
Chapter III: The Social Determinants of Health

I began the last chapter with a discussion of the State Children’s Health Insurance Program (SCHIP) and recent Congressional efforts to increase the number of enrollees. Opposition to SCIP expansion focused on whether it represented a step toward government-funded universal health care and, if so, whether that was an appropriate role for the state to take. Daniels, we can surmise, would support such a move. The SCHIP program seems designed to promote equality of opportunity by providing American children from low-income families similar access to health care as their higher-income counterparts. Critics of the program, however, contended that health care access should be determined within a free market system even if that means some individuals would be without insurance. The central question in the debate was whether it was fair for government institutions to redistribute resources in the form of tax dollars so as to provide another good—health care—to people with less resources. Few commentators, however, asked what the expected impact of the SCHIP expansion would be on the health burdens that American children from low-income families face later in life.

One of the central questions in recent public health research has been whether inequalities in access to medical care drive health disparities. The implementation of universal health care programs, particularly in Europe, has created a kind of social experiment to evaluate changes in health disparities with more equal medical access. Research in this area has suggested that there are a number of factors that contribute to health
disparities beyond access to care. These include health behavior; absolute, relative, and overall distribution of income; genetic differences in disease susceptibility; available social capital; and racism, classism, and other -isms. In this section I want to review some of the data concerning the broader determinants of health.

Let me begin with a few notes about the research on which I am focusing. First, the disparities that concern me are those that occur in the developed world. There is no question about the importance of access to health care and adequate public health measures in the developing world. I will argue, however, that, beyond a certain point, access to health care ceases to drive disparities when compared with broader social determinants. Thus, while improved (or any) public health or health care would dramatically influence the health disparities between the world’s richest and poorest nations, the effect on the difference between the uninsured and the insured in, for example, the United States would be much less significant.

Second, I will generally focus on studies whose emphasis is on all-cause mortality rather than inequalities in specific diseases, which often have unique explanations for why they are more common among certain groups. For example, the Multiple Risk Factor Intervention Trial (MRFIT) reports that men from higher income areas have an increased relative risk of dying in flying accidents compared to lower-income men.\textsuperscript{61} This is not surprising given that air travel is an expensive undertaking and itself follows an income gradient. In contrast, the relative risk of developing stomach cancer for low-employment-grade Britains approaches twice that of the upper grades.\textsuperscript{62} The primary etiological agent of stomach cancer—\textit{Helicobacter pylori}—is acquired more

\textsuperscript{61} Smith, et al., 1996, pp. 486-96.

\textsuperscript{62} Smith, 2003, p. xxiv.
frequently and earlier in life among lower-grade families, which largely explain the
differential rate of this disease.

Focusing on trends in all-cause mortality helps to wash out the differences that air
travel or *H. pylori* infection produce, allowing us to consider the broader patterns of
health inequalities. One exception to this general point is coronary heart disease (CHD or
CAD). In addition to being a leading cause of mortality in the United States and world-
wide, there are known physiologic risk factors (smoking, dyslipidemia, hypertension,
family history, age, and diabetes) and medical interventions (medication, stenting,
bypass, etc.). This means that CHD is an ideal target for assessing the role that
expanding access to care—in terms of public health measures designed to address
modifiable risk factors as well as traditional medical care in the form of pharmacologic
and surgical interventions—has on inequalities in the incidence, prevalence, morbidity,
and mortality from the disease.

Third, my main focus will be on the Whitehall I and II studies in Great Britain
and similar studies in the United States, where available. The Whitehall studies, which
were conducted on members of the British civil service, are notable in several respects:
first, they were large-scale, prospective, and longitudinal, with Whitehall I following over
10,000 men for a decade and Whitehall II following over 10,000 men and women for
eleven years. Second, the population investigated had universal access to medical care
through the United Kingdom’s National Health Insurance. Third, participants were
essentially pre-stratified by education and income in virtue of their position (grade) in the
Finally, data from these studies have shown that health inequalities correlate with the civil service hierarchy such that any given individual has, on average, better health than someone immediately below him in the hierarchy and worse health than a person just above. In other words, while there is a marked difference in health between the top and the bottom—between the haves and have-nots—there is also a step-wise gradient in health for all levels in between. Absolute deprivation may translate into poor health, but so does relative deprivation compared to other social groups. Evidence exists for a social gradient in health in a number of countries, including places thought to be more broadly egalitarian than Great Britain, such as the United States. All of these factors make the Whitehall studies an important tool for analyzing the relative impact of conditions outside of access to medical care on health.

Finally, a few comments about the statistics found in this chapter. In general, I will be relying on data that report relative risk of disease and mortality for different members of the socioeconomic strata. One caveat on relative mortality risk, which is the ratio of the incidence of death in a population exposed to a risk factor (e.g. lacking health insurance) to incidence of death in an unexposed population (e.g. having health insurance): it is usually important to also know the absolute risk of death for the overall population. If the absolute risk (or the incidence) is low, then the relative risk measures may not reflect a substantive clinical difference. For example, a 289% increase relative risk of death may sound impressive unless it really translates, in the context of a very low

---

63 Of the five social classes, Grade I includes professionals; Grade II, managerial and technical occupations; Grade III, skilled manual and non-manual occupations; Grade IV, partly-skilled occupations; Grade V, unskilled workers.

absolute risk, into 289 less seconds of life. Information on the absolute risk of death for various diseases in different countries can be found through the World Health Organization. I will not be reporting 95% confidence intervals along with relative risk; that information is generally available in the primary sources. Finally, this chapter is not intended as a formal review paper or meta-analysis of the broader determinants of health. I have no explicit search strategy or methodological criteria for included some papers and not others, although I have tried to rely on peer-reviewed, anthologized studies. My goal is to give an overall impression of the current state of research in this area.

Access to Health Care

Let me begin, then, with data on the relationship between increased access to care and the reduction of health disparities.\textsuperscript{65} There are three lines of evidence that universal access to medical care is inadequate to address health disparities. First, disparities persist in countries that adopt even the most generous government-sponsored medical coverage. For example, research on relative differences in mortality between non-manual and manual male workers ages 45-65 in various European countries has shown that, despite extensive health coverage networks, manual workers in the Nordic countries have between a 1.33 and 1.52 increased relative risk of death compared to non-manual workers.\textsuperscript{66} Furthermore, Finland has a higher relative inequality in death rates between manual and non-manual workers (relative risk 1.52) than countries like England and Wales (relative risk 1.45) that have a less extensive national health service. European health disparities—and not just in mortality but in all-cause morbidity and self-reported

\textsuperscript{65} For an overview of this subject see Lurie and Dubowitz, 2007, pp. 1118-21. Adler, et al., 1993, pp. 3140-5.

\textsuperscript{66} Mackenbach, 2006, pp. 223-50.
health as well—reappear when different indicators for socioeconomic status besides occupation are used, including income, housing and unemployment rates, and education.\textsuperscript{67} Morbidity data consistently show similar degrees of relative health disparities throughout Europe, regardless of health care system. Countries that enact legislation expanding health care access across their populations continue to have marked health inequalities. Even Japan, which has an expansive health care system and whose citizens enjoy one of the highest age-adjusted life expectancy on Earth, has a significant degree of health inequality. The lowest income class is 1.54 times more likely to report poor health compared to the highest class; the second lowest, 1.22 times.\textsuperscript{68}

Second, health gradients in Western European countries are as steep as or steeper than those in countries with less extensive coverage, like the United States or Russia. For example, a cross-sectional study comparing the relative risk of death from heart disease in men in manual versus non-manual occupations showed a relative risk of 1.50 for English men between the ages of 45-59 and 1.25 for American men in the same age group.\textsuperscript{69} In other words the relative morality rate from ischemic heart disease in these populations was twice as high in Britain—which had universal health care access—as in the United States. Similar differences in relative risk of death have also been identified in post-Soviet Russia, which is somewhere in between the United States and Britain with regard to medical coverage within its health care system. For example, Russian men age 25-64 who have only completed primary school are 1.23 times more likely to die from.

\textsuperscript{67} It can be misleading, as Davey Smith emphasizes, to think of income and education as measuring equivalent aspects of one underlying condition—socioeconomic status—since separate life events, which themselves represent variable disease exposures, contribute to each. See Smith and Hart, 1998, pp. 1744-5.

\textsuperscript{68} Shibuya et al., 2002, pp. 16-9.

\textsuperscript{69} Kunst, et al., 1999, pp. 47-53.
cardiovascular disease than university educated men, adjusting for age, smoking, total cholesterol, systolic blood pressure, frequency of drinking, and body mass index.\textsuperscript{70} Similar disparities have been reported for Russian men of lower compared to higher income.\textsuperscript{71} These data suggest that health disparities are as significant in countries without universal health care as in those with such systems.\textsuperscript{72}

Third, inequalities in all-cause mortality have also increased over time—despite overall improvements in population health—in countries with universal health service like the United Kingdom. Between 1972 and 1992 the difference in male age-adjusted life expectancy between the highest and lowest social classes in England and Wales increased from 5.5 years to 9.5 years.\textsuperscript{73} Between 1990 and 1999, the British relative index of inequality—the relative mortality rate of the (hypothetically) poorest person compared to that of the (hypothetically) richest one—increased from 1.68 to 1.85.\textsuperscript{74} In the United States, between the years 1960 and 1986 the difference in mortality between the most and least educated men ages 25-64 doubled.\textsuperscript{75} From the years 1980-1992, the overall male age-adjusted life expectancy at birth in Britain increased from 70.8 to 73.2.\textsuperscript{76} In the United States, it increased from 70.0 to 72.3.\textsuperscript{77} Given similar

\textsuperscript{70} Malyutina, et al., 2004, pp. 244-9.

\textsuperscript{71} Nicholson, et al., 2005, pp. 2345-54.

\textsuperscript{72} There is a separate question as to whether universal access to health care is important for better absolute population health (i.e. age-adjusted life expectancy of the population as a whole), which may improve even while the degree of health inequality stays constant. See the Black Report, 1992.

\textsuperscript{73} Drever and Whitehead, 1997: pp. 1-257.

\textsuperscript{74} Smith, et al., 2002, pp. 434-5. Part of the explanation for the widening gap may have to do with increased smoking cessation in the highest classes compared to the lowest during this time period.

\textsuperscript{75} Pappas, et al., 1993, pp. 103-9.

\textsuperscript{76} National Statistics Online. Great Britain, 2007.

35
improvements in British and American population health and similar rates of expansion in disparities in all cause mortality, it is difficult to conclude that the universal health care available in Great Britain has served as a corrective to the growth of health disparities.

It is important, however, not to overstate the case against expanding access to care. First, there are data to suggest that regional expansions of health insurance/coverage translate into modest reductions in health inequalities for some diseases.78 For example, Federal and state broadening of Medicaid programs to cover antiretroviral therapy for low-income HIV patients has been credited with decreasing the rate of AIDS-associated opportunistic infections.79 The key point, however, is that, contrary to the aim of most American political efforts in this area, merely expanding health care access will not alleviate health inequalities, nor will it prevent them from worsening.80 In this regard, however, we should be careful not to confuse access with utilization, especially since persons with poor health prior to the expansion of coverage may be in a worse position to take advantage of it. For example, low socioeconomic status Canadians are less likely to use specialist services than those with higher status, despite identical coverage under the same single-payer health system.81

77 National Center for Health Statistics, 2002.


80 Geoff Brennan has suggested that the expansion of health care can create incentives for poor health behavior (so-called moral hazards) such that people may take up risky or imprudent behaviors once they are guaranteed medical care. If poorer, less educated individuals were more likely to act on these incentives then it would appear that health disparities persist in the face of equal access to health care even though expanding access—considered independently of its effects on moral hazards—has a significant impact on these disparities. This should be kept open as a theoretical possibility, although it is difficult to argue that individuals at the lower end of the socioeconomic spectrum would take up poor health behaviors so as to exactly balance (or somewhat worsen) health disparities that would have otherwise decreased.

In this case, the persistence of health disparities even in the face of adequate access to care might be explained because of different preferences or abilities to receive that care. Data from the Whitehall II study, however, suggest that South Asian (predominately Indian) civil servants receive more cardiovascular care (in the form of resting electrocardiograms and coronary angioplasty) than their white counterparts, but are still at a two-fold higher risk of coronary-related morbidity and mortality. Other studies concerning health-seeking behavior among British South Indians report that they are more likely to seek care for (hypothetical) anginal symptoms than whites.

There is also a question regarding the quality of care low status individuals receive. A commonly-cited Institute of Medicine study suggests that minorities and the poor in the United States who have access to care receive worse chronic and acute disease treatment as well as lower-quality diagnostic care. This effect persists when adjusting for disease severity, patient age and gender, type of health system in which care was received, co-morbid disease, and type of insurance/payment for care. Thus, even within a universal health care framework, some patients may receive worse care than others, which could also explain the persistence of health disparities. Although the utilization and quality of care data are limited in scope, the points about expanding access are still highly relevant.

---


84 Institute of Medicine Committee on Understanding and Eliminating Racial and Ethnic Disparities in Health Care, 2001. See also American College of Cardiology Foundation, 2002.

85 It is important to note that, even if expanding access to health care did, in fact, significantly reduce health disparities, other socioeconomic factors, insofar as they determined the willingness of a society to provide equal access to care in the first place, might still deserve moral attention. For example, Kawachi and Kennedy, 1999, pp. 215-27 have argued that relative income inequality erodes social cohesion and capital, which affects the willingness of countries to introduce safety nets in areas like health. If this is right,
Health Behavior

An alternative explanation for health disparities other than differential access to care is health behavior. Former Senator William Frist, a physician, writes, in the context of racial and class health inequalities, “[patients] must stop smoking, eat right, exercise, take their medication, and monitor their blood sugar, based on their own volition and usually outside of the clinical setting. Public policies must encourage patients to embrace personal responsibility.” As Senator Frist correctly points out, the poor are more likely to lead a sedentary lifestyle, to be obese, to smoke regularly, to be regularly drunk, and to use illicit drugs than the middle-class and wealthy. They also have a worse diet. Should we be surprised, given this accumulation of risk factors for disease, that they have worse health? Data like these are cited in a number of contexts, most commonly as part of an argument that, because poor health behavior is a matter of personal choice, correcting the resulting disease disparities is an inappropriate target for social resources.

For example, Rudolf Klein, a British political scientist, writes:

“If differences in health status are the result of individuals making autonomous choices—for example, by smoking or refusing to take exercise—then it is difficult to argue that this can be described as inequitable. To the extent that society has to pick up the costs of such choices, there may still be a case for trying to change the behaviour of those individuals, but it cannot be argued on equity grounds.”

Putting aside the issue of the social and genetic context in which these choices are made, the complex factors that lead the poor to persist in these behaviors, and the philosophical relationship between desert, responsibility, and the distribution of social

regardless of that policy’s specific effect on health disparities, we might have an obligation to redistribute income so as to make it more likely that our social system would take steps to provide universal coverage.

86 Frist, 2005, p. 447.
88 Klein, 2000, p. 569.
resources, there is still a question as to how much health behaviors actually contribute to health disparities. For example, the Whitehall I study indicates that, while higher employment grades have lower mortality from coronary heart disease (CHD), this effect persists, although not as dramatically, when controlling for the chief risk factors for CHD (relative risk of 1.77 between the lowest and highest grades). Taking plasma cholesterol concentration, blood sugar, systolic blood pressure, height, and smoking into account, the lower employment grades have a 50% higher mortality rate from CHD (relative risk of 1.50, reduced from 1.77) than those at the highest employment grade. Almost two-thirds of the difference in mortality between the highest and lowest employment grades remains unaccounted for when the above factors are controlled.

The British Regional Heart Study (BRHS) reported a less significant residual disparity in age-adjusted CHD between manual and non-manual class men—20% or a relative risk of 1.20—after controlling for cardiac risk factors including blood cholesterol, blood pressure, body mass index, cigarette smoking, alcohol, physical activity, height, and lung function. Among the explanations for the different residual relative risks (1.50 vs 1.20) between the BRHS and Whitehall studies is that the latter used a more finely-graded hierarchy, which allowed them to further identify CHD disparities among non-manual workers, producing a higher relative risk from top to bottom. In both cases, however, a significant difference in mortality was due to factors


90 Marmot, 2004, p. 44.

91 Marmot, 2004, p.44.

92 Emberson, et al., 2004, pp. 289-96. Coronary heart disease was defined in this study using a composite endpoint of non-fatal myocardial infarction or coronary death in contrast to the Whitehall studies, which focused on CHD-related mortality alone.
outside of health behavior and access to care (which all subjects in the studies had). Furthermore, to put a 20% contribution in context, consider that smoking—the single most important contributor to the global burden of disease—accounts for only 12.2% of lost disability-adjusted life years (DALYS) in the developed world.\textsuperscript{93} High blood pressure accounted for 10.9%; alcohol use for 9.2%.

The Longitudinal Study on Socioeconomic Health Differences (LSSHD), which stratified 27,070 Dutch men and women by education, reported a hazard ratio of 1.64 (a statistical measure similar to relative risk) for all-cause mortality between the most and least educated.\textsuperscript{94} Controlling for alcohol consumption, smoking, body mass index, and physical activity—the four behavioral factors most associated with increased mortality in the study population—resulted in a 53% reduction in hazard ratio (to 1.30), leaving half of the mortality difference unexplained. Other studies looking at the relationship between health behavior and health disparities in the United States have reported residual disparities in ranges similar to the Whitehall and LSSHD data.\textsuperscript{95} The take-home message is that, while poor health behavior is associated with lower socioeconomic status and with worse health, adjusting for its effects leaves a significant difference in mortality between the top, middle, and bottom unexplained.

\textit{Income Inequality}

When looking at the relationship between income and health disparities, it is important to identify several different ways in which income may affect health, including

\textsuperscript{93}Ezzati, et al., 2002, pp. 1347-60.

\textsuperscript{94}Schrijvers, et al., 1999, pp. 535-40. Level of education was broken down into four categories and calculated adjusting for age, gender, marital status, religious affiliation, and degree of urbanization.

differences in absolute income, percent of income spent on health care, relative income, the form income takes, and overall inequalities in income distribution. First, GDP per capita does not correlate well, beyond possessing a basic minimum, with standard measures of population health like life expectancy or infant mortality.\textsuperscript{96} For example, the 2004 infant mortality rate in the United States, whose GDP per person was $40,100, was 6.5 deaths/1,000 live births. The 2004 infant mortality rate in Japan, whose GDP per person was $29,400, was 3.26 deaths/1,000 live births.\textsuperscript{97} Considering developed countries more generally, the correlation between per capita GDP—converted at purchasing power parity to allow comparison of absolute wealth—and infant mortality is weak (r=0.08).\textsuperscript{98} This may not be surprising given that GDP, by itself, does not capture how wealth is distributed in a nation. Having a very high GDP is compatible with a country having a large impoverished population, which, in turn, would increase infant mortality rates despite a significant GDP.

Second, the percent GDP spent on health care does not correlate well with standard population health measures. In 2004 the United States, with a life expectancy of 77.7 years, spent approximately 14% of its GDP on health care while Japan, with a life expectancy of 81.1 years, spent less than 10% of its GDP.\textsuperscript{99} In 2001, the United Kingdom spent 7.5% of its GDP on health care and had a life expectancy of 77.9 years;

\begin{footnotesize}
\textsuperscript{96} United Nations Development Programme, 2003.
\textsuperscript{97} CIA World Factbook, 2005a.
\textsuperscript{98} Wilkinson, 1997, pp. 591-5. An r value of zero implies no correlation; negative r values (0 to -1), indirect correlation; positive r values (0 to 1), direct correlation.
\textsuperscript{99} CIA World Factbook, 2005b.
\end{footnotesize}
Germany spent 10.8% with a life expectancy of 78.0 years.\textsuperscript{100} Importantly, this argument assumes that health expenditures in different countries are equally proportioned between administration and actual health care; in other words, that percent GDP spent on health actually goes to medical care. There is evidence that the United States spends a significantly greater amount of its health budget on administrative costs compared to other countries.\textsuperscript{101} Physicians in America also have a higher salary (and higher malpractice insurance rates) than other countries, factors that may decrease the fraction of the health-related GDP that actually goes toward patient care. Even correcting for these ‘overhead’ expenditures, however, the return in American life expectancy on GDP investment in medical care is comparatively poor.

Third, relative income \textit{within} a country does seem to have an impact on health. For example, in the United States, household income correlates with relative risk of death after stratifying for age, sex, race, family size, and time period. As noted in the first chapter, households making less than $15,000 per year have a relative risk of death almost 3 times that of households making greater than $70,000 per year. Households making between $50,000 and $70,000 are at a 1.34 increased relative risk of death. Stratification for education reduces the steepness of the difference, but a significant gap remains between each group and between the highest and lowest income households.\textsuperscript{102} In England, almost 40% of the poorest quintile of 50-54 year old men report poor health

\textsuperscript{100} World Health Organization, 2004.

\textsuperscript{101} United States Congress, Office of Technology Assessment, 1994.

\textsuperscript{102} McDonough, et al., 1997, pp. 1476-83.
compared to less than 10% of the richest quintile, an effect that reoccurs in older age
groups.\textsuperscript{103}

Fourth, differences in the form income takes as a household asset—for example, whether money earned is held as cash or invested in stock or mutual funds—may translate into differences in health. United States Census Bureau data from 1990 indicate that black Americans are significantly less likely than whites to have their assets in a financial institution, own stock, or have an IRA account, which affects their ability to translate income into commodities like home ownership.\textsuperscript{104} In 1990, 67.3% of white households owned their own home, compared to 43.8% of black household.\textsuperscript{105} By 2006, those numbers were 75% and 47%, respectively.\textsuperscript{106} Home ownership, in turn, is an important determinant of environmental exposures—including lead and air pollution—that contribute to health status. In this context, it is not just how much income you have that matters, but what form it takes.

A different idea is that income inequality \textit{as such} is creating health differences such that what matters is not so much absolute deprivation or wealth, but the magnitude of the relative differences.\textsuperscript{107} This is a different marker than relative income, as discussed above. Here, the focus is on the degree of difference between the richest and poorest members of society. If relative position in an income hierarchy tells how much you have compared to other people, income inequality measures how big the gap in

\textsuperscript{103} Marmot, et al., 2002.
\textsuperscript{104} For a review of this data see Oliver and Shapiro, 1995.
\textsuperscript{105} Wilkerson, 1990, p. A1
\textsuperscript{106} U.S. Census Bureau, 2006.
\textsuperscript{107} See, for example, Wilkinson, 1996.
income is between the top and the bottom. Since it is possible for someone to have the same relative income in two societies with very different degrees of income inequality, this is a different candidate for explaining health disparities. In this context, a country’s Gini index, a measure of income inequality between zero (total income equality) and one (total income inequality; i.e. one person with everything), does inversely correlate with its life expectancy at birth, an effect that persists when controlling for total income level, education, and savings level \( (r = -0.45) \).\textsuperscript{108} These data, however, have recently been challenged; the inclusion of Gini index scores for a larger number of countries than in the original study decreased the strength of the correlation \( (r = -0.09) \).\textsuperscript{109}

Within the United States, studies have shown that, adjusting for differences in absolute income, income inequality accounts for approximately a quarter of the difference in age-adjusted mortality rates between states.\textsuperscript{110} The Robin Hood Index—a measure of income inequality representing the proportion of aggregate income that must be transferred from households above the mean to those below the mean so as to establish perfect equity—is strongly correlated to state by state variation in all-cause mortality \( (r = 0.54) \).\textsuperscript{111} The mechanism by which income inequality generates health disparities is unclear. Some commentators argue that relative inequality erodes social cohesion and capital, which affects the willingness of countries to introduce safety nets in areas like health.\textsuperscript{112} Others believe that declining social capital directly affects health.\textsuperscript{113} In

\begin{footnotes}
\end{footnotes}
contrast, Angus Deaton, an econometrician, has argued that income inequality and health disparities are merely correlated; social systems that tolerate a substantial inequality in income are likely to take a similar view of health disparities.\textsuperscript{114}

Other Candidates

Another factor thought to influence health disparities relates to various prejudices on the part of health care providers, including racism, sexism, and classism.\textsuperscript{115} Although this is likely a real phenomenon—as I indicated in the earlier discussion of the Institute of Medicine review—it does not, by itself, offer a compelling explanation for health disparities. For example, an American College of Cardiology meta-analysis of racial differences in cardiac interventions found that minority patients are somewhat less likely to undergo percutaneous cardiac angiography or coronary artery bypass surgery even after adjusting for age, insurance, co-morbidities, and/or disease severity. Although a number of factors underlie these differences, discrimination may be a contributor.\textsuperscript{116} Low income patients are less likely to have blood pressure or cholesterol checked by their primary care providers, which may be evidence of class bias, although it may also have to do with the quality of physicians that care for financially poor patients.\textsuperscript{117} The various “-isms” may have more of an effect on health disparities via the way they shape

\begin{footnotes}
\footnote{113 For a review, see Berkam and Melchior, 2006, pp. 55-72.}
\footnote{114 Deaton, 2003, pp. 113-58.}
\footnote{115 See, for example, van Ryn and Burke, 2000, pp. 813-28.}
\footnote{116 Schulman, et al., 1999, pp. 618-26.}
\footnote{117 Das and Gertler, 2007, pp. 296-309. Unfortunately there are little reliable data available on the quality of care rich vs. poor or urban vs. rural or Medicare vs private insurance patients receive. If there is a significant difference in the quality of health care providers that serve these populations, it may again emphasize the point that expanding access to medical care cannot alone reduce health disparities.}
\end{footnotes}
institutional arrangements rather than through the attitudes and actions of health professionals. At this higher level of abstraction, however, the specific effect of prejudices can be difficult to entangle from the more general ways that social arrangements—without being overtly bigoted—lead to health disparities.

Finally, there is research indicating that variables like education; employment status; job stress; neighborhood and housing conditions; low birth weight; early environmental exposures to lead, allergens, and infection; use of city spaces, including recreational areas and forms of transport; family structure; patient preferences for intervention; and health and religious beliefs about the nature of disease contribute to health disparities.\(^\text{118}\) Most—if not all—of these factors distribute along a socioeconomic gradient, although their relative contribution to health differences has been difficult to accurately measure because of confounders like income. Some of them—neighborhood conditions and available recreation space, for instance—act partially through behavioral risk factors and are, as such, already reflected in the earlier discussion. For example, predominately black American neighborhoods are less likely to have recreational facilities and grocery stores and more likely to have fast food restaurants, which discourages certain health behaviors.\(^\text{119}\)

**Reverse Causation**

So far, the theories I have been considering use various features of the socioeconomic hierarchy—income, differences in health care access, etc.—to explain why there is a corresponding gradient in health. I want to conclude this chapter by considering an alternative theory. Do people with worse health fall to the bottom of the

\(^{118}\) For a discussion of most of these factors, see the papers in Marmot and Wilkinson, 1999.

hierarchy or is there something about being there that affects their health, as I have thus far suggested? Several authors have argued that low status groups are in that position because they have poor health, which prevents them from achieving upward social mobility.\textsuperscript{120} And the same argument can be made for high status groups. There are several lines of evidence that lead away from this conclusion. First, premature death (<65 years old) of one’s parents is generally considered a marker of susceptibility to disease as well as early life exposure to certain risk factors (e.g. second-hand smoke, poor diet, etc.). If this is true, and if one’s background health/disease risk determine social class, then populations whose parents did not die prematurely should show relative uniformity in disease morbidity and mortality. In other words, if you were not exposed to genetic or environmental health risks via your parents, then you should not, in general, face health-related impediments to higher social status. Data from the Whitehall studies show, however, that the social gradient in CHD is as steep among individuals whose parents did not suffer premature death.\textsuperscript{121} This provides some evidence against health-related mobility.

Second, while it is true that individuals who move up the social gradient are healthier than other members of the groups they leave behind, they are still less healthy than people in the class they join. For example, epidemiologists Mel Bartley and Ian Plewis looked at changes in social class and health status between 1971 and 1991 in a sample of 1% of the English and Welsh male population ages 15-40.\textsuperscript{122} Men who were stably located in the highest social class (based on occupation) were 30% less likely to

\textsuperscript{120} West, 1991, pp. 373-84.
\textsuperscript{122} Bartley and Plewis, 1997, pp. 376-86.
report limiting long-term illness (LLLT) than men upwardly mobile into that social class. The same is true in reverse for those who fall. Men who were downwardly mobile into the lowest social classes were still 40% less likely to report LLLT than men who were stable in those groups. Blane et al. performed a similar study using data on the social mobility and health of over 46,000 English men ages 45-65. Approximately one-third of their population experienced a change in class over the time period. Men who were upwardly mobile (moved from lower to higher social classes, again based on occupation) had a 1.15 hazard ratio of mortality compared to those who remained stable in the higher class. Similar data have emerged from research done on social mobility from childhood onward. Together these studies suggest that, if anything, health-related social mobility acts to constrain the health disparities between socioeconomic classes. The fact that men who change classes end up with mortality rates intermediate between the class of origin and the one they join suggests that a diluting effect takes place when considering the overall difference in mortality between classes.

I have barely scratched the surface of the epidemiological research on health disparities. Looking at all of this information, one may despair at finding a unifying explanation or point of intervention. For every potential contributor we hold up—take wealth and education, for example—we have to consider whether it is actually a cause or merely a correlative of health disparities. Do the financially poor have worse health because of their poverty or is poverty a marker for some underlying factor that leads to poor health? Does lower education in adults predict worse health because of its effects on medical literacy or do people fail to achieve better education because of childhood

illnesses, which themselves also affect adult health? The answers to these questions may require multiple explanatory frameworks, each focusing on different factors. In the next chapter, however, I will examine one potentially unifying theory, centered on the relationship between autonomy and health.
Chapter IV: Autonomy and Health

In 2001 the *Annals of Internal Medicine* published an article examining the relationship between winning an Academy Award and longevity.\(^{125}\) The study authors compared the age at death for actors and actresses who had won an Academy Award to two groups: gender-matched individuals who had appeared in the same movie and had not been nominated and other actors and actresses who had been nominated for a different film but had not won. After adjusting for birth year, ethnicity, birth country, name change, age at first film, and total films in career, individuals who had won an Academy Award lived almost four years longer than those who had been merely nominated and those not recognized at all; individuals who won multiple times lived six years longer.

The Academy Awards phenomenon illustrates a health gradient: winners of multiple awards lived longer than those who had only won once, who lived longer than those merely nominated and those not recognized at all. As I noted in the previous chapter, health gradients are a pervasive feature of developed countries with and without universal health care. When stratifying individuals by occupation, income, or level of education, each person has, on average, better health than someone immediately below him and worse health than a person just above. Health disparities between the riches and poorest represent the extremes of a phenomenon identified in the Whitehall studies, the step-wise social hierarchy of health. The British epidemiologist who led the Whitehall research, Michael Marmot, is more concerned with how regional disparities like, for example, those between whites and blacks

in North Carolina fit into this larger pattern of health inequalities. He writes, “There is a
social gradient in health in individuals who are not poor: the higher the social position, the
better the health. I have labeled this ‘the status syndrome.’”

Under the status syndrome, the problem of why North Carolina’s poor die younger
than its rich turns out to be the same as why Academy Award winners live longer than
nominees and why people with PhDs live longer than those with Master’s degrees. This
is, Marmot acknowledges, a surprising conclusion, but he believes he can provide a unifying
explanation. More important than, for example, a person’s access to health care or health
behaviors is his or her autonomy. He writes, “The lower individuals are in the social
hierarchy, the less likely it is that their fundamental human needs for autonomy and to be
integrated into society will be met. Failure to meet these needs leads to metabolic and
endocrine changes that in turn lead to increased risk of disease.” The poor, ethnic
minorities, and people with Master’s degrees have less autonomy and are less integrated into
society than those with PhDs and so have worse health.

Marmot draws on an extensive epidemiologic literature—some of which I touched on
in the previous chapter—to argue for autonomy and social participation as causes of the
status syndrome. His interpretation of these data has been challenged elsewhere. Although much turns on the empirical debate, there are several conceptual points that need
clarification. Here, I will focus primarily on the relationship between autonomy and health

---

126 Marmot, 2006, pp. 1304.


129 Among the principle challenges is whether low control causes poor health or if poor health produces low
with less emphasis on the role of social participation.130  This discussion is especially pressing given Marmot’s recent call for the American medical profession to take aim at health disparities by championing policies—“education, skills training, better working and living conditions, and support for older individuals”—to improve patient autonomy.131

*Autonomy*

It is important to distinguish Marmot’s explanation for the status syndrome from a more widely-held theory that traces health disparities to differences in *socioeconomic* status (SES). For example, one might think that poverty causes a subset of North Carolinians to have worse health. Without money to pay for medical care, drugs, better food, or to protect them from harmful environmental exposures, the poor end up with worse health. Marmot explicitly denies this idea. He writes, “The social gradient in health is not due to differences in medical care, or primarily to differences in health behaviors, or to differences in material circumstances.”132 In support of this claim, Marmot notes that within groups low on life stressors and social isolation education no longer correlates with mortality. The same is true for groups high on life stressors and social isolation.133 For Marmot, the traditional determinants of SES like income, education, or occupation merely correlate with better or worse health. What underlies health disparities is differences in status *simplicitur*.

Initially, this is a difficult claim to understand. SES—or relative social position, a term Marmot uses synonymously—is just a conglomeration of factors like distributable, material goods; our place, time, and location of birth; our grammar and accent; our gender;


the house and neighborhood we live in; the attitudes that others adopt toward us; the color of our skin, our general appearance, and the clothes and accessories we wear; our talents; our education; the kind of job we have and how well we perform it; whether and who we marry and, if we have children, how they turn out. So what causal (or explanatory) power does SES have independently of these components? SES, it turns out, is just useful shorthand for talking about what all these elements have in common; namely, their ability to promote our autonomy. Being well-educated, having a good job, and possessing material goods do not themselves directly contribute to better health; rather, Marmot argues, they promote the autonomy of the person who has them, which, in turn, leads to health. So better SES correlates with improved health because it is a marker for greater autonomy. It is, however, important to note that, while Marmot’s goal is to explain the ultimate causes of health disparities, his practical recommendations for changing gradients in autonomy will be very similar to those of a person who believes that such disparities are mediated through SES. This is because SES is a proximate cause for Marmot and one that is more easily adjusted than autonomy.\(^{134}\) At the end of the day, even if Marmot is wrong about the relationship between autonomy and health, it may still be true that differences in SES are a leading contributor to health disparities.

That said, Marmot’s link between autonomy and health travels via the neuroendocrine pathophysiology of the sympatho-medullary and the hypothalamic-pituitary-adrenal axes, an extended discussion of which is too far afield for my purposes.\(^{135}\) Instead, I will focus on the role autonomy plays in his framework. Despite—or maybe because of—being a central

\(^{134}\) Marmot, 2006, p. 1304.

talking point in biomedical ethics, moral psychology, and normative and metaethics, autonomy remains a slippery concept. In a recent survey of the autonomy literature, the philosopher Nomy Arpaly identifies at least eight different uses of the term.\textsuperscript{136} The concern is that agreement or disagreement with Marmot’s position may occur because of grammatical similarity with whatever idea of autonomy the reader already possesses. Equivocation, intentional or not, is a significant worry in this context.

Some Candidates

Marmot’s definition of autonomy is “how much control you have over your life.”\textsuperscript{137} If addressing individual autonomy is going to drive our approach to health disparities, this definition requires clarification. Most importantly, we need to know what sort of control Marmot has in mind.

Control as Desire Congruence

On a common view in philosophy, having control over your life means being motivated by or acting on only desires that you reflectively endorse.\textsuperscript{138} The absence of control occurs when we find ourselves being motivated by desires that we do not want to have. For example, the addict who sees his desire for drugs as alien yet nevertheless shoots up lacks control in this first sense. The addict could reasonably insist that he is not being himself—that he is out of control—when he uses drugs. The closeted religious leader who has sex with other men despite his alienation from this desire may feel a similar absence of control. The writer, Tom Wolfe, describes a character, Charlotte Simmons, who lacks this kind of control, finding her life stressful and complicated when she cannot bring herself to

\begin{footnotes}
\item[136] Arpaly, 2003, 117-48. I will be following Arpaly’s structure in the subsequent discussion of autonomy.
\item[137] Marmot, 2004, p. 11
\item[138] For example, Frankfurt, 1988, pp. 11-26.
\end{footnotes}
love the man, Adam Gellin, she believes is right for her. Wolfe writes, “The truth was Charlotte wanted to fall in love with Adam. If only she could! How much tidier life would be! . . . And she wanted to want Adam! She wanted to kiss Adam good night in a deeply committed way.”139 Incongruence between one’s first and second order desires can lead to increased stress and neuroendocrine activation, as happens to Charlotte, making this type of control a possible candidate for Marmot’s definition. On this position, then, having maximal control means acting (or being motivated) only by those desires that you want to have. Call this type of control desire-congruence.

Control as Independence of Mind or Independence of Means

Consider the philosopher Thomas Hill’s description of the Deferential Wife,

“This is a woman who is utterly devoted to serving her husband. She buys the clothes he prefers, invites the guests he wants to entertain, and makes love whenever he is in the mood . . . she tends not to form her own interests, values, and ideals; and, when she does, she counts them as less important than her husband’s.”140

Let us suppose that the Deferential Wife has never made an explicit decision to subordinate her projects to her husband’s. She was raised in a social system in which she is simply fulfilling the expected role for women. In this context, the Deferential Wife is not in control of her life. She may have perfect desire-congruence control—she may want to want to serve her husband—and may be happy doing so. But, insofar as her husband runs her life, telling her who to see, what to do, and what to believe, she lacks independence of mind. It is easy enough to imagine a feminist encouraging the Deferential Wife to seize control of her life; to reevaluate her commitment to her husband’s desires and to give some priority to (developing) her own projects. Being maximally in control, in this context, means

140 Hill, 1991, p. 5.
developing and exercising one’s capacity to think for oneself, where that includes deciding what things to value and to what degree without undue outside influence. Call this type of control *independence of mind*.

A related understanding of control is *independence of means*. In this sense, you are in control of your life if you are self-sufficient. The elderly couple that has saved wisely for their retirement and is able to live on their own has independence of means. The cancer patient who is confined to bed because of intractable bone pain lacks control; he is dependent on others to fulfill his most basic needs. The parents who encourage their adult child still living at home to get a job and go out on his own are promoting his self-sufficiency. Like independence of mind, however, independence of means or self-sufficiency is separate from the structure of one’s desires. You may or may not want to be dependent on others and you might not identify with what you want. Central to independence of means, however, is whether, within your specific context, you are able to accomplish your goals with minimal or no support from others.

*Control as Freedom*

On another view, control means having the resources and ability to get, in the broadest terms, what you want. This is control as *freedom*. There are several ways in which a person might exhibit this type of control, depending on how we cash out “what you want.” At the broadest level, we can imagine an individual—call her Lynn—who has the resources and ability to turn any of her *desires* into reality. Whenever Lynn desires something, she is able to get it. If she wakes up and the room is too cold, she can make it warmer. If she misses an exhibit from her favorite artist, she can have the museum reopen the galleries, or, better yet, arrange for a private showing. More dramatically, if she wants it to stop raining, it
does, and if she needs an extra hour of sunlight, she gets it. From moment to moment, throughout her life, Lynn has the resources and abilities to obtain what she desires. According to psychologist Ellen Skinner, this type of control, which is commonly employed in the empirical literature on self-efficacy, “refers to the extent to which an agent can intentionally produce desired outcomes and prevent undesired ones.”\textsuperscript{141} Of course, how believable Lynn is as an actual person depends on what she desires. Unless she maintains a very practical set of desires, having this type of control over her life will require extra-human abilities.

A more narrow type of freedom would involve the kind of control that comes with the ability and resources to turn one’s \textit{projects} into reality. Let me pause to broadly distinguish projects from desires. First, although our projects are often a subset of the things we desire, we are more deeply committed to our projects. This commitment is illustrated in the way that we identify ourselves wholly or partly with our projects, in contrast to our less central desires. I do not think of myself as someone who likes coffee, although this is true of me. I am a runner and someone who wants to learn to play Chopin Preludes perfectly; these things are among my projects, not just passing desires. They form part of my conception of who I am. I would lose a part of my identity if I stopped running, but not if I gave up coffee. Second, we organize our other, less central desires around our projects. We are more concerned with fulfilling them than our passing fancies. For example, a desire for bourbon might strike me late at night while I am working on my project of becoming a doctor. I consider how going out for a drink will affect my ability to finish studying that evening. I stay home. We sacrifice or postpone fulfilling our immediate desires when they come into

conflict with our projects. When we act on a desire that threatens our projects, we feel regret; we may believe we have made a poor decision.

If we have the resources and ability to fulfill our projects then we are able to exercise a certain kind of control over our lives. Consider Scott as an exemplar of this type of control. Unlike Lynn, Scott cannot obtain the object of every desire that he has. There are no gallery re-openings or late sunsets in his life. There are things he wants but does not get. Instead, he is able do those things necessary to complete his projects. Suppose that Scott has always wanted to be a mechanical engineer. This is one of his projects and it is part of his identity; he thinks of himself as someone who wants a career in engineering. Because he has the resources and ability to achieve his projects, he is able to do well in high school, enter college, finish his coursework, and get a job with an engineering firm. Along the way, he always has money for tuition, the kind of intelligence and drive necessary to pursue mechanical engineering, and the personality traits needed to work on complex projects with others. Sometimes there are vacations he wants to take but does not have the time or money; there are houses he wants but cannot afford; from time to time he wishes he were in a band, nothing serious, just something to do on the weekends. But Scott does not adopt these things as projects; they are passing desires and he lacks the kind of control over his life to make them a reality. Like Lynn, however, Scott’s plausibility as a real person largely depends on the kinds of projects he has. If his life-goal is to be the first person to land a spaceship on the sun, he will require the same extra-human abilities as Lynn in order to maintain the kind of control—the ability to fulfill his projects—we imagine him having.

Both Scott and Lynn exhibit a kind of freedom that allows them to translate certain internal states—projects and desires—into reality. There are other variations on this theme;
perhaps the resources and abilities to obtain those things necessary for further fulfillment of
one’s desires represents another kind of control. A different possibility might involve having
the resources and ability to get what one values, if that is a distinct category from one’s
projects and desires. Another sort of freedom, however, has to do with control over certain
external exposures. There are some people who have the resources and ability to structure
their lives such that they avoid a range of negative experiences without any explicit desire or
project to do so. As such, they exhibit have a different kind of freedom than Lynn and Scott.

During the course of our lives we are exposed to a number of environmental changes.
Most of these are small—is it raining today? Is the pool going to be open an hour later?
Should I bring a salad for lunch because the microwave in my building is broken?—but we
can also run up against larger events—a hurricane that destroys our house, a gram of heroin
laced with talc, a drunk driver on a Friday night. The resources and abilities we have
available to us over the course of our lives determines, to a significant extent, what changes
we encounter. While we may intentionally structure our lives to avoid or ensure certain
exposures—I choose this school because of the quality of the faculty; I go to that market
because the tomatoes are always fresh; I stay away from his house because the floors are
sticky—we also avoid many events as a byproduct of our choices and not because of any
specific desire in this regard.

For example, I might have the money and financial savvy necessary to live in the
suburbs, thereby avoiding the higher crime rates in the city, but I might choose to live in the
suburbs because I like the idea of having a front yard. In this case, the desire that motivates
me to enter a particular environment means that, as a side-effect, I avoid a certain exposure
(being robbed, etc.), even if doing so never explicitly enter my calculus. I exhibit a kind of
freedom in this regard that is not present in someone who lacks the resources to move out of a high-crime area. In this case, the advantage of having resources is not that you get what you desire or fulfill your projects—although this may be true as well—but that your life is structured so that you do not have certain negative experiences.

In a similar vein, one of the historical prerogatives of wealth has been a relative insulation from infectious diseases (polio being a notable exception). The basic infrastructure present in most of the United States means that, in virtue of being born into a certain life, Americans are not exposed to yellow fever or cholera. We live our lives in such a way that—without ever having specific desires related to avoiding these diseases—we are not subject to them. What makes living in the developing world so treacherous from an infectious disease standpoint is that it is relatively impossible to structure one’s life in such a way to avoid these exposures; this is one reason why visitors have to get so many vaccines.

Imagine two contemporary residents of Wilmington, North Carolina: Adele and Charles. Adele is an upper-class socialite who summers in the coastal city, pursuing a season of beach-related parties and events. Charles, who has been in and out of prison because of a crack-cocaine habit, is a year-round resident of Wilmington, living in crowded housing in the poorest section of the city. Charles contracts tuberculosis (TB) because of his multiple incarcerations and living conditions—both risk factors for the disease. Adele’s worse medical problem is mild hayfever. Charles, unlike Adele, lacks the resources and abilities necessary to avoid exposure to TB. What makes Adele different from Charles is not a question of desire fulfillment. Neither of them has any explicit desires related to avoiding TB. In fact, they do not even know that it still affects people in the United States. Adele has the freedom to lead a life structured in such a way that she never encounters TB; Charles
does not. Being unable to determine the events to which one is exposed is a significant limitation to having control over one’s life.

Importantly, control as a species of freedom—on any of the above definitions—is independent of higher-order psychological states. Looking back at Lynn, her control is over the fulfillment of her first-order desires; she may, for example, wish that she was not always asking for more sunlight. But Lynn does not have the resources or abilities to alter what she desires to desires. That would require a different kind of control. Similarly with Scott, who may not want to want to be a mechanical engineer. Like the addict, having the resources and abilities to get what we want is independent of whether we want to want those things.

Additionally, you can have the kind of control that comes with freedom in one of two ways. You may be able to get what you desire, fulfill your projects, determine your environmental exposures, etc because no one interferes with you or you may be able to do those things because of support from others. In the first case, which might be thought of as negative control, what allows you to get what you want is the fact that others are not preventing you from doing so. They are not helping you achieve your ends; rather, they are merely staying out of your way, whether intentionally or because their own pursuits do not hinder yours. For example, Adele is able to structure her life because no one else is forcing her to behave in a certain way. If her husband demanded that she take a trip to Thailand, where she was exposed to TB, this would represent a reduction in the negative control she has over her life (i.e. the control she has because others do not interfere with her).

In the second case, if you have freedom because of support from others then you exhibit positive control. What allows you to get what you want is the fact that others are assisting you. This may be through direct contributions to your ends—parents allowing a
grown child to live at home while he looks for a job—or through indirect mechanisms like relying on Social Security or other tax-funded government programs for a regular source of income. For example, if Scott can only become a mechanical engineer because of federal student loans, then he would have positive control over his projects. This, incidentally, means that people who have positive control necessarily lack independence of means. They require help from others to get what they want.

Finally, it is important to distinguish control as freedom from control as independence of mind. The Deferential Wife, as long as she has the resources to do what her husband wants, has control in the freedom sense, even if she lacks independence of mind. If his desires are hers, if his projects are hers, and she is able to fulfill them, then she has a kind of control (freedom) over her life. In all of these scenarios, she has control in the freedom sense, even if she is not involved in deciding what desires, projects, or events to pursue.

*Moral Autonomy*

In this brief survey, I have left aside a number of more technical definitions of autonomy, especially those used in the normative context of “respecting autonomy.” Most of these—for example, autonomy as self-legislation—are used in Kantian-style arguments connecting practical reason and morality or in various forms of action theory. For example, Thomas Hill describes Kantian autonomy as, “a property of the wills of all adult human beings insofar as they are viewed as ideal moral legislators, prescribing general principles to themselves rationally, free from causal determinism, and not motivated by sensuous desires.” Autonomy, in this largely non-empirical sense, is almost universally possessed and is not distributed according to a social gradient. Given that the property of self-legislation does not vary widely between persons, it is an unlikely candidate for a principle.

---

142 Hill, 1991, p. 44.
cause of health disparities. This should, however, serve as a reminder that autonomy is a morally loaded word. Merely demonstrating that someone is more or less autonomous on a descriptive understanding of that concept—for example, more or less independent of means—is insufficient to establish any specific normative conclusion. We require an additional argument for why we should value that kind of autonomy and work toward its redistribution.

*Marmotian Control*

That said, which idea of control—and corresponding understanding of autonomy—is closest to Marmot’s? Consider his paradigmatic cases of absence of control: a petty officer of the Royal Navy who succumbs on an Artic expedition “lacked control over his own destiny”; a factory worker in rural Massachusetts who may be fired because of a shift to overseas operations is subject to forces “totally outside of her control”; a foreman with a damaged car lacks “the resources to take control of the situation rather than have events control him.” These examples indicate that Marmot is not primarily concerned with motivational states. His characters are not torn between what they desire and what they want to desire. Desire-congruence is not part of his concept of control.

Marmot notes that autonomy matters because it determines whether people are “able to lead the lives they most want to lead.” This is a difficult phrase to parse, but it seems to locate the central question as whether people are able to get what they want. This is, I think, closest to understanding control as a matter of freedom. What characterizes Lynn, Scott, and Adele is their ability to have what they want in their lives. But is Marmot concerned with getting what one desires, fulfilling one’s projects, or controlling one’s exposure? First, there is little suggestion that Marmot’s characters lack control because they are unable to fulfill

---

their projects. Although the foreman might not be able to get to work given the absence of reliable transportation, he is not worried about being unable to complete his central projects. He is not threatened with a loss of identity in the way that a concert pianist might be if she began going blind; the foreman’s concerns are much more about being able to complete certain mundane daily tasks.

Instead, what unites Marmot’s characters is that they are exposed to certain events outside of their control, and, once they enter these situations, they are unable to have them resolve as they desire. His characters lack a certain ability to navigate the world, both in terms of avoiding particular kinds of exposures—being conscripted to go to the Artic; being affected by shifting factory work overseas—and in being able to respond to those exposures—lacking the money to pay for car repairs; not having the training to find a new job quickly.

Control, for Marmot, ends up being a combination of the lives lead by Adele and Lynn. What marks someone as having control is the extent to which she can avoid certain exposures, and, when they happen, whether she can determine how those events turn out. The lower down a person is in the status gradient, the less likely it is that he will have the resources and abilities to control his exposure to stressful life events and the less likely it is that he will be able to dictate the outcome. He will be conscripted into an Artic expedition and then die of exposure when the trip goes badly. She will be fired from her job when the market shifts and then be unable to work because she has limited skills. A person with higher status—who possesses the goods and abilities that come with elevated social position—is better able to control and adjust to his environmental exposures. He will form an expedition to the Artic and then bravely make it out alive when things go badly. When his
factory becomes unprofitable due to market changes, he can move it oversees, maintaining his income. He can lead the kind of life he wants to have, thus avoiding the ill health effects of chronic neuroendocrine activation.

Is there anything more to Marmot’s definition of autonomy beyond this type of control? First, as a practical point regarding independence of means, we might wonder whether it really is possible for a person to lead the life he wants if he is entirely dependent on the resources of others. People get tired of giving, are not always available, or refuse certain requests. Even individuals who bear a special relationship to us—families and friends, for example—may be unwilling to take full responsibility for giving us everything we want. Theoretically, however, since independence of means is not necessary to control the events to which one is exposed or to dictate the outcome of such exposures, I will assume that promoting Marmotian autonomy does not include that aspect of control.

Now consider the Deferential Wife. There is a sense in which, so long as she is able to do everything her husband wants, she is leading the life she most wants to lead. Parsing the idea of “the life she wants to lead” differently, however, might encourage us to conclude that, because the Deferential Wife is laboring under a kind of ‘false consciousness,’ she does not really have the life she most wants. We might argue that it is not the life she wants to lead; it is the life her husband or some larger, patriarchal system wants her to have. Just like the petty officer who dies in the Artic, there is a real sense in which the Deferential Wife lacks control over her destiny. She does not decide what happens to her nor does she determine how her life changes in response to these events. Her husband does. The question is whether, in order to count as having control in Marmot’s sense, a person not only has to get what she wants but also has to have some involvement in what it is that she wants.
It is hard to know what Marmot would say about the Deferential Wife. On the few occasions he talks about individuals in similar situations—for example, office workers who are entirely under the thumbs of their bosses—he makes it clear that their health is worse because they lack the kind of control necessary to avoid dull, demoralizing, and repetitive work. In contrast, it is precisely because the Deferential Wife lacks independence of mind that she is not particularly stressed, and stress is the pathophysiological link that Marmot needs to connect lack of control with poor health. Let us suppose, then, that, for Marmot, the kind of control that comes with autonomy is understood to only include the kind of freedom I have been discussing. Independence of mind is not necessary for autonomy on this account.

Promoting Marmotian autonomy is, therefore, compatible with a world in which people find themselves alienated from their first-order desires (lack desire-congruence), are largely or entirely dependent on the resources of others (lack independence of means), and are able to get what they want but are minimally involved in setting their ends (lack independence of mind). This list is not intended as a criticism of Marmot’s account, only to highlight the type of autonomy he sees as linked to health. While lacking autonomy in these other senses may also contribute to poor health, they are not at the center of his attention. It is also possible, of course, that there are other grounds on which we should promote the different kinds of control represented in these cases, but these recommendations cannot be generated out of Marmot’s focus as it stands.

Given the way that Marmot defines autonomy, it is important to note that his other candidate for a principle determinant of health—degree of social participation—will be encompassed as one of the exposures to which individuals may have more or less control. Individuals with a high degree of autonomy in their lives will have the ability to shape their
social participation and access to social capital, which, in turn, will give them more resources for dealing with negative exposures than someone with less control. Marmot acknowledges this point, writing “autonomy includes opportunities for social participation and the filling of social roles.”\textsuperscript{144} Given that social participation will ultimately be mediated through autonomy, our primary focus should be on this factor.

\textit{The Illusion of Control?}

Before I turn explicitly to the relationship between Marmotian autonomy, a descriptive concept, and a normative approach to health disparities, I want to discuss the relationship between health and the \textit{belief} that one has autonomy versus the actual possession of autonomy. If we are to intervene, should we focus on increasing the control low status individuals have over their lives or should we instead try to instill the belief that they have control? Take two men, George and Franklin, both of whom earn similar wages working for a small construction company. They both finished high school, but not beyond. They used to smoke, but stopped recently because of small children at home. George’s wife is trying to get him to cut back on drinking; Franklin’s wife thinks he should probably eat less red meat. On the weekends, the two men get together with their families to grill at one or the other’s house, both located in a small, low-crime neighborhood. When they talk, George always seems stressed: he is worried about gas prices; their company does not provide health insurance and the children have to have regular pediatric visits; he has trouble making minimum payments on his credit card debt.

Privately, George admits that his life is not turning out how he wanted. If he just had a little more money or was able to take night classes at the community college and become a supervisor, things might be different. As it is, he is worried that one or two unexpected costs

\textsuperscript{144} Marmot, 2004, p. 241.
will put him in a really bad position. Franklin, who has similar problems, is, by temperament, less worried. He just cannot bring himself to get worked up in the same way as George. Sure, he does not have everything that he wants, and sometimes he and his wife have to think hard about their budget when something unexpected happens, but he would generally describe himself as satisfied. Franklin even has vague plans to move to a neighborhood with better schools once the children get older.

The scope of George and Franklin’s autonomy is limited to the same degree. They have similar abilities and resources available to avoid certain exposures—they live in low-crime neighborhoods; have steady employment—and respond to unforeseen circumstances—a broken down car might be equally problematic—but only George appreciates his absence of control. Franklin, in contrast, thinks of himself as being able to exercise a fair degree of control over how his life goes. Will this difference be reflected in their health? More generally, although the degree of class-related health inequality is at least as pronounced in America as in developed other countries, Americans are less likely to see themselves as belonging to a lower class even if, objectively they do. Most Americans below the poverty level view themselves as either having just left the middle class or about to join it. They believe that they are more autonomous than they may actually be, considering, for example, their education, job status, and income. Does this belief insulate these Americans from the stressors that would come with confronting the reality of their situation?

Almost all of the epidemiologic studies that have been done on the relationship between differences in control and health disparities have focused on control beliefs. For example, in a Dutch study of men and women age 57 and older, lower socioeconomic groups
were 1.42 times more likely to suffer from heart disease. Adjusting for control beliefs—measured through questions like “sometimes I feel that I am being pushed around in life”—reduced the relative risk of death by a third, to 1.30. The authors suggest that control beliefs account for a significant degree of the difference in heart disease incidence between low and high status individuals. Other studies have shown similar results.

Marmot, while acknowledging the difference between perceived versus actual autonomy, tends to emphasize the latter. For example, in discussing the relationship between work environment and health, Marmot writes, “If [an employee] has a high degree of control over work, it is less stressful and will have less impact on health . . . high-status civil servants have a high degree of control over their work.” He also emphasizes that his approach is not merely about what people believe they can do with what they have but what they are actually capable of doing. Finally, Marmot’s proposed interventions for reducing health disparities—adopted largely from the Acheson Independent Inquiry into Inequalities in Health—focus on promoting actual control. They recommend “reductions in poverty in women of child-bearing age”; “promoting management practice that lead to an increased level of control . . . in the workforce”; “policies which will promote the material well-being

145 The 1.42 RR is adjusted for classical cardiac disease risk factors, including hypertension, diabetes, obesity, exercise, and smoking, and represents a 4% reduction from baseline risk of 1.45. Bosma, et al., 2005, pp. 737-45.

146 For a review of these data along with recommendations for intervention, see Bosma, 2006, 153-66. For an argument that control beliefs do not contribute to health disparities see Smith and Harding, 1997, pp. 1369-70.

147 Marmot, 2004, pp. 124-5. Marmot notes that both employee beliefs about the control they exercise in their work environments and employer rankings of the amount of control different jobs provide correlate with self-reported health and with each other.

of older people.”149 Few, if any, of the Acheson recommendations focus on developing control beliefs as opposed to redistributing the resources necessary for actual control.

Of course, this is not a zero-sum game. Interventions that promote actual control among low status persons through, for example, the redistribution of income or increasing educational opportunities probably improve beliefs about control. And encouraging positive control beliefs probably empowers individuals to improve their actual control. Nevertheless, Marmot’s primary interventions are aimed at reducing differences in the resources and abilities that cause disparities in control and health. We are not to encourage people to think that they are able to lead the life they most want to live but to work to make them actually able to do so.

We are now in a position to state how Marmot understands the empirical connection between autonomy and the status syndrome. The relative ability of an individual to actually shape the events to which she is exposed as well as dictate the outcome of these exposures is a leading cause of differences in relative health. Put differently, the greater the difference in autonomy between the highest and lowest members of a society, the larger the degree of health inequality. Ultimately, we can remain agnostic regarding Marmot’s connection between actual autonomy, the neuroendocrine systems, and health disparities.150 The determinants of socioeconomic status may work through a common mechanism (i.e. autonomy) or, as I suggested in the previous chapter (and find more plausible), cause disparities through separate pathways. In either case, if we are able to affect the distribution

---


150 Although it matters for Marmot’s normative approach—based on Amartya Sen’s theory of justice—to the social determinants of health that they are mediated through autonomy and do not work through separate pathways.
of the resources and abilities that shape SES, whether or not they function by changing
autonomy, should we? And, if so, on what grounds and to what extent?
Chapter V: Status

I began this dissertation with an overview of contemporary health disparities in the United States, which exist between racial and ethnic groups and between individuals with different types of occupation and levels of income and education. The political, medical, and philosophical response to this issue has focused on expanding access to health care as a way to correct what has been perceived as an unfair distribution of disease. The philosopher Norman Daniels, in particular, has argued that, because health disparities create inequalities in opportunity, they are unjust. This way of conceptualizing the moral status of health disparities suggests that inequalities come from an unfair allocation of resources. Unsurprisingly, Daniels has concluded that we have an obligation to ensure access to medical resources to reduce the impact disease has on equality of opportunity.

Recent research in epidemiology, however, has called into question the connection between health disparities and access to health care. A number of factors—the broader determinants of health—help to create, maintain, and worsen health disparities. In fact, the epidemiologist Michael Marmot has argued that these determinants work through a common mechanism; namely, neuroendocrine responses to the diminished autonomy produced by low socioeconomic status (SES). In this context, Marmot understands autonomy as the relative ability of an individual to actually shape the events to which she is exposed as well as to dictate the outcome of those exposures. Up until this point, however, the discussion of these epidemiological facts has been entirely descriptive. The task, then, is to connect the broader determinants of health—whether mediated through autonomy or otherwise—with a
normative framework that can make recommendations as to how we should change their. There are a number of ways to approach this project. First, we might treat the current application of theories of justice as involving a flawed piece of instrumental reasoning: while the conclusion is correct—health disparities are unjust—the corrective means—equalize access to health care—are not. The solution is to insert the appropriate methods of alleviating health disparities into the equation. For example, we might maintain that health disparities are unjust and that we should focus on redistributing the broader determinants of health as a way of remedying these inequalities. Here, the unjustness of health disparities does the normative work with manipulation of the broader determinants simply as a means to achieving more equitable health.

Second, we might look outside theories of justice for suggestions as to why and how we should redistribute the broader determinants of health. For example, a welfare utilitarian might argue that the steep gradients in SES that produce health disparities are detrimental to the maximization of well-being and so ought to be reduced. Here, as in the first approach, reducing health disparities would be a way of attaining another normatively desirable goal and redistributing the social determinants of health would be the mechanism for enacting that change.

A third way is to focus, not on the unfairness of health disparities, per se, but on the relative unfairness of the processes that generate them. The so-called indirect approach, which has been developed by the philosopher Fabienne Peter, does not argue that health disparities demand intervention because, for example, they limit fair equality of opportunity.151 Rather, Peter suggests that, insofar as an unjust social process—for example, persistent inequality in income distribution—produces health disparities and, insofar as we

---

151 Peter, 2004, p. 79.
are called on to correct these processes, we have grounds on which to correct health disparities. Here, the reduction in health inequalities is a byproduct of achieving a normatively desirable state of affairs (reduction in other, unjust social processes) and not the reason why we are instituting change in the first place.

A final and related approach to the new epidemiology of health disparities is to argue that SES is itself morally important such that we should direct our efforts at promoting a more equitable distribution of it. As with the indirect approach and in contrast to the first two cases, where equalizing the broader determinants of health was seen as a means to achieve another goal (the elimination of an unjust state of affairs, improving overall welfare, etc.), here it is the primary end. In contrast to the indirect approach, however, the issue is not whether differences in SES or control are unjust, but that being able to exercise a certain amount of control over one’s life is morally important enough to serve as independent grounds for reducing SES gradients. Ultimately, I will be arguing that a variation on this fourth approach should be our preferred normative approach to health disparities. It provides a broader context for thinking about SES and health disparities outside the standard domain of distributive justice. Before I turn to this point, however, I want to locate Daniels’ position within this framework and discuss its relative merits.

Daniels Redux

The first approach is most similar to Daniels’ original arguments connecting the unjustness of health disparities to inequality of opportunity with expanding access to health care as the mechanism for correcting this state of affairs. Given the new epidemiology of health disparities, a modified version of Daniels’ account might read: first, health disparities are unjust because they limit equality of opportunity, which is a requirement of justice.
Second, gradients in SES produce health disparities. Therefore, we should design our institutions so as to reduce these gradients. The focus now is on the broader determinants of health rather than merely access to care. As I already mentioned, Daniels has made a similar response to the new epidemiological data.

In a recent series of papers and in his book, *Just Health: Meeting Health Needs Fairly*, Daniels has argued that the Rawlsian approach must be modified to focus more broadly on educational opportunities, social cohesion, and income redistribution in addition to questions of health care access when addressing health disparities. He writes, “Our health is affected not simply by the ease with which we can see a doctor—though that surely matters—but also by our social position and the underlying inequality of our society.” In his approach to the public health literature, Daniels has tended to emphasize the effects (or correlations, as he is apt to call them) that SES—as determined by income, income inequality, participation in political activity, and social capital—has on population health. He writes, “We should be looking as well to improve social conditions—such as access to basic education, levels of material deprivation, a healthy workplace environment, and equality of political participation—that help to determine the health of societies.”

Recall that Rawls’ contractors are concerned with the distribution of primary goods: basic rights, prerogatives, wealth/income, and the bases of self-respect. Included among these are some of the principle socioeconomic determinants of health: access to education and the quality of that education, wealth, income, and the responsibilities that come with certain positions/jobs. Rawls’ two principles, which the contractors are supposed to accept...

---


153 Daniels, 2001, p. 6.

given the other features of the original position, establish the framework for allocating these goods. Since, according to Daniels, these goods are driving health disparities, the principles of justice that govern their distribution also provide a normative structure for addressing the non-access causes of health disparities.\textsuperscript{155} Daniels concludes that, “We should view health inequalities that derive from social determinants as unjust unless the determinants are distributed in conformity with [Rawls’] principles.”\textsuperscript{156} In other words, health disparities in \textit{our} society are permissible only if they would still exist in a world where primary goods (i.e. the social determinants of health) are properly distributed according to the two Rawlsian principles and whatever secondary rules are generated therein.\textsuperscript{157} But how and to what extent would these principles reduce health disparities?

First, recall that the fair equality of opportunity principle holds that socioeconomic inequalities are only permissible if the offices and positions to which they attach are open to all under conditions of equal opportunity. As I noted in the second chapter, this requires not just formal equality of opportunity—for example, eliminating race-based discrimination in hiring practices—but institutional changes designed to reduce the effects of certain contingencies on opportunity. In Daniels’ original argument, the central contingency of interest was differences in disease (or health), a focus that he still maintains, writing, “the equal opportunity principle also requires extensive public health, medical, and social support

\textsuperscript{155} See also Caputo, 2003, pp. 85-112.

\textsuperscript{156} Daniels, 2001, p. 6.

\textsuperscript{157} It is important to note that primary function of the Rawlsian distributory scheme is to provide fair resource allocation along a number of dimensions, with health being only one aspect of that. It would be a mistake to suppose that Daniels is arguing that we should manipulate the principles such that they generate the distribution that \textit{maximally} removes health disparities.
services aimed at promoting normal functioning for all.”158 In his more recent work, Daniels has argued that things like public education, day care, and early childhood intervention programs are also supported under the equal opportunity principle. Such programs, which would be targeted so as to reduce socioeconomic inequalities in, for example, educational attainment, would also reduce the health disparities predicated on these inequalities.

Second, the Difference Principle, as Daniels envisions it, would limit disparities in health produced from inequalities in income and wealth. Recall that the Difference Principle, which is lexically second to the fair equality of opportunity principle, permits resource inequalities so long as such inequalities make the worst off as well off as possible. Daniels interprets this principle as providing grounds for reducing gradients in income and wealth. Given that, in a hierarchy, there is always a group that is least well off, the difference principle calls for continued redistribution so as to control the steepness of the gradient from top to bottom, placing “significant restrictions on allowable inequalities in income and wealth.”159

The Rawlsian principles thus speak to the redistribution of several key socioeconomic determinants of health. Daniels suggests that the Rawlsian system would, “probably flatten the socioeconomic gradient even more than we see in the most egalitarian welfare states of northern Europe.”160 And, as a result, would significantly reduce the steepness in health gradients. To summarize, one response to the new epidemiological data is to use the same normative framework, centered on Rawlsian distributive justice, but insert a different empirical premise concerning the causes of health disparities. On this account, justice

---

158 Daniels, et al., 2004, p. 79.
159 Daniels, et al., 2004, p. 79.
requires that we not only provide more equitable access to health care but that we also
redistribute the socioeconomic factors that help shape health. Is this adjustment adequate?

The Nature of Socioeconomic Status

For Marmot, SES matters because the things that it encompasses are what generate
autonomy. Higher SES individuals have the resources and abilities necessary to shape the
events to which they are exposed as well as dictate the outcome of these exposures. As a
result, they have better health than lower SES individuals; the greater the difference in
autonomy, the greater the health disparities. Daniels, in response to these data, has suggested
that Rawls’ account can guide us as to how to redistribute some of the key determinants of
SES and thereby reduce gradients in control and the corresponding health disparities. The
best way to evaluate the adequacy of Daniels claim is to examine the nature of SES and what
it encompasses. Are there forces that shape SES and control that fall outside of the
traditional domain of distributive justice but nevertheless demand our intervention? If so, we
may have grounds to question the adequacy of Daniels’ normative approach. This will be my
focus for the remainder of the chapter.161

Let me start with the suggestion that to have a certain status is to have a kind of value
in the eyes of others.162 SES, in this context, is the value we have in the eyes of others based

161 Other challenges to Daniels’ new approach as well as a reply from him and his co-authors can be found in
Cohen and Rogers, 2000. In particular, Marsha Angell (pp. 42-7) and Steffie Wollhandler and David
Himmelstein (pp. 79-82) argue that absolute deprivation in material goods rather than degree of overall
inequality should be the central focus for intervention; Michael Marmot (pp. 37-41) suggests that Sen’s
capabilities approach to distributive justice should be the preferred normative framework, a point that I address
in Chapter 6; Barbara Starfield (p. 67) questions some of Daniels’ (and Marmot’s) empirical assumptions,
suggesting that the primary way to reduce health disparities is to expand access to primary care.

162 I am using the term ‘value’ here and throughout in a relatively loose sense. I intend this account to be
compatible with a number of ways of specifying what it is to value something including having a pro-attitude
toward it, believing that there are reasons that make it choice-worthy, believing that your ideal self would want
it, etc. Importantly, I am interested in maintaining a close connection between the idea of valuing a person and
adopting specific interpersonal attitudes (especially esteem) toward him or her. See Frankfurt, 1988, pp. 80-94.
on our social and economic properties.\textsuperscript{163} I mean this as an entirely descriptive claim, without yet commenting on the sorts of things that people should care about. One’s socioeconomic value in the eyes of others—or SES—is largely context dependent, resting on possession of whatever factors matter within a given group. As I noted in the previous chapter, the social and economic features that determine SES amongst contemporary Americans typically include: distributable, material goods; characteristics of an individual’s parents; her place, time, and location of birth; her grammar and accent; her gender; the color of her skin and her general appearance; her friends, acquaintances, and contacts; the attitudes that others adopt toward her; where she goes to school and how educated she becomes; the kind of job she has and how well she performs it; whether and who she marries and, if she has children, how they turn out.\textsuperscript{164}

We can be valued because our social or economic properties are useful to other people: for example, having a high level of income means that we buy many goods and services, which works to the advantage of people selling those things; being well-educated means that we may have helpful species of knowledge, as is the case with physicians who use their expertise to treat patients. The converse is also true such that people can be disvalued for properties that make them useless or harmful to others: for example, people with low income who depend on social assistance may be disvalued as ‘nonproductive’ members of society; having children that grow up to be criminals may lead people to disvalue an individual insofar as this reflects on her parenting skills. In these cases, what

\textsuperscript{163} Notice that these are not judgments about the value of persons \textit{qua} persons, only about their value given who they are socially and economically. Since I am just interested in SES in this dissertation, when I discuss our ‘value’ in the eyes of others, I am only referring to socioeconomic value. I briefly touch on how the kind of value we have in virtue of our rationality might inform our socioeconomic value in Chapter 7.

\textsuperscript{164} Race and ethnicity are not usually thought of as part of one’s SES but since these are social characteristics of persons, I will be discussing them under that general rubric.
determines our (dis)value in the eyes of others is how (un)helpful our properties are to them. We can, however, be valued for social or economic properties that are not specifically useful to others: being attractive, being born in New England, or having a particular racial or ethnic background. These characteristics do not do anything for other people but nevertheless can mark us as having a certain SES.

Perhaps the best illustration of SES comes from popular depictions of social divisions in high school. The entire plot of the 2004 movie *Mean Girls* focuses on the daughter of zoologists who, after being home schooled for a number of years, is able to successfully transition to high school because she recognizes its similarities to the social status hierarchies of chimpanzees, which she has studied extensively.\(^ {165} \) The teenagers with the highest SES possess and do the things that matter most to their peers: they wear the right clothes, drive the right cars, know and associate with the right people, participate in the right activities. Being of high SES allows them both to determine certain exposures—the head cheerleader has options when it comes to choosing her prom date—and more easily react to any negative events they might encounter—if the head cheerleader’s prom date breaks up with her, she can always find someone else.

It is important to note that an individual may be of high SES within one group but nevertheless be of low SES given a broader comparison class. To return to the high school example, the president of the audio-visual club may have the best recording equipment in town and be well-respected by other club members but still be of lower status than the quarterback of the football team. While it may be difficult to make fine distinctions in SES given a sufficiently large population—does the AV club president have higher status than the

\(^{165}\) *Mean Girls*, 2004. This analogy is only somewhat glib; studies among non-human primates have shown differences in health that correspond to primate social hierarchies. See Spolsky, 2004, 393-418.
lead in the high school play?—there are still broad divisions like total income or highest level of educational attainment that are meaningful. And, in fact, the epidemiological research discussed in Chapter 3 used such broad distinctions—between for example, individuals with no education, high school, college, and advanced degrees—to capture SES across large populations. In general, an upper class, be it cheerleaders in high school or those at the top of the British civil service, will have more or a better mix of the factors that are important to the other members of their socioeconomic structure.

Status gradients do not end once we leave the high school cafeteria, although they may become more subtle, partly because one’s SES in contemporary society is not static over time. This was not always the case, especially in socioeconomic systems where the position into which one was born—nobility or the peasant class, for example—cemented, for better or worse, one’s social standing. Limited upward or downward mobility meant that SES was relatively fixed, although a person could certainly rise or fall within his specific social sphere. Given, however, that many of the things on the above list are subject to change—whether chosen by an individual or not—contemporary SES gradients are more fluid. SES varies as trends pass and the perceived value of places, other people, things, and connections change. Individuals—academics, politicians, businesswomen, etc.—rise and fall within SES hierarchies depending on their ability to adjust to variations in what matters within and outside of their respective spheres.

A central theme in J.M. Coetzee’s novel Disgrace is the way that political change and personal action can alter our SES. The protagonist, the white Afrikaner David Lurie, is a well-respected professor of Romantic literature in South Africa; following the post-Apartheid

166 For a discussion of the change from a hereditary system of establishing SES to a meritocracy and the impact that has had on contemporary society, see de Botton, A, 2004.
formation of Mandela’s government, his university is redesignated as a technical college and he is forced to teach courses in communication skills. Divorced, increasingly irrelevant to his colleagues, lecturing on a subject he perceives as beneath him, David begins to visit a prostitute, but even she eventually refuses to see him. An ill-advised sexual affair with one of his students is discovered by the university and his department, full of indignation over his behavior, forces him to leave. David feels that his status has fallen so far that, were he to run into one of his colleagues, they would shudder, “as one shudders at a cockroach in a washbasin in the middle of the night.”

Eventually he moves to a farm with his daughter, Lucy. When he and Lucy are the victim of an assault, the local police are unable or unwilling to pursue the attackers and his own efforts are useless. David’s daughter advises him to accept the humiliation. He should learn to accept that they have nothing and are nothing: “no cards, no weapons, no property, no rights, no dignity.”

By the end of the novel, David has come to occupy the social position of black South Africans under Apartheid.

My concern at this juncture is not whether it is appropriate for an individual’s SES to rise or fall based on, for example, his race or sexual indiscretions; rather, I am interested in SES gradients as a purely descriptive phenomenon. As the above examples illustrate, SES is determined by a wide range of factors, varies depending on the comparison group, and can change significantly over time. It should be granted that many of the things that work to establish SES in contemporary cultures are included amongst the primary goods, especially, as Daniels has emphasized, income (relative and absolute), educational attainment, and occupation. In redistributing these factors (or creating equality of opportunity to access

---


them), we would, undoubtedly, flatten SES gradients. If everyone in high school had the same amount of money to spend on a car then it seems likely that vehicles would fall out of the equations that govern relative SES within this group. School uniforms might be understood partly as an attempt to end status differentiation based on clothing. By iterating this effect over the whole spectrum of primary goods—and in groups beyond teenagers—we would presumably flatten the SES gradient. Given what we know from the epidemiological data, we might also expect a reduction in health disparities. And this is exactly what Daniels has in mind.

Such an approach, however, fully addresses health disparities in the developed world only if we assume that the distribution of primary goods is the sole determinant of the socioeconomic hierarchies that underlie these disparities. There are, however, many other factors on the above list that go into shaping an individual’s SES and thus contribute to her health. I want to focus on interpersonal attitudes as an example of one of these forces. I will suggest that the health disparities these attitudes generate through their effects on SES hierarchies are unjustified. They are not, however, adequately addressed by theories of justice, at least in their current form.

Interpersonal Attitudes

What motivates individuals to seek higher SES or to acquire the social and economic characteristics that would lead others value them more highly? A number of evolutionary biologists have claimed that we are, in some sense, programmed to constantly reassess our status in relation to others and seek better status, when possible.169 We do not, however, have to go so far as to embrace nativism about status-seeking in order to explain the

---

phenomena; elevated social standing conveys significant benefits that would be appealing even outside of evolutionary impulses. The goods (income, education, etc.) and opportunities (business contacts, invitations, etc.) that characterize higher SES individuals are useful both in and of themselves—a house in the suburbs reduces exposure to inner-city crime—and as resources to enhance one’s ability to obtain further SES—a house can be mortgaged to provide funds for one’s children to go to Harvard, which would raise one’s prestige. These goods create space for leisure as well as the pursuit of projects and interests beyond what is necessary for daily subsistence. At the same time, however, individuals continue to acquire these resources even when they already have more than would be needed for several life-times worth of projects. This suggests that there is something beyond material goods or opportunities that makes high SES valuable. As Adam Smith has phrased the issue,

“What is the end of avarice and ambition, of the pursuit of wealth, of power and pre-eminence? Is it to supply the necessities of nature? The wages of the meanest labourer can supply them. What then are the advantages of that great purpose of human life which we call bettering our condition?”

The answer that Smith gives is that, more important than the material or intellectual benefits of high SES, is the attitudes others adopt toward those at the top. He writes, “To be observed, to be attended to, to be taken notice of with sympathy, compliancy, and approbation, are all the advantages we can propose to derive from [bettering our condition].” The philosopher Alain de Botton notes that being of high SES means that, “Our presence is noted, our name is registered, our views are listened to, our failings are

---

170 I am not adopting the relatively implausible thesis that a desire for higher status motivates all (or even most of) our actions; rather, I am suggesting that material goods can be used in the pursuit of higher status, if that is one’s goal.

171 Smith, 1817, p. 77.

172 Smith, 1817, p. 78.
treated with indulgence and our needs are ministered to.” 173 Some of these benefits take the form of actions—being on the guest list of the most exclusive restaurants in town—but the central point of emphasis is that high SES individuals are approached with a certain set of attitudes including sympathy, deference, respect, admiration, and esteem. I want to focus on the last of these, esteem, as an illustration of the relationship between SES and interpersonal attitudes, although I intended my remarks to apply to the other attitudes as well.

The economist Geoffry Brennan and the philosopher Philip Pettit broadly characterize esteem as an attitude that involves a positive assessment of a person. 174 They identify three characteristics of esteem: first, it is evaluative, ranking individuals in various respects. These can be egoistically centered—I esteem her more than him because of what she does for me—or they can simply be judgments of merit—I esteem her for being a smart person. Esteem can be focused on people for having character traits like kindness, or for having positional properties like being the most honest person in the room. In the context of the present discussion, an individual with high SES is esteemed because she possesses the right mix of social and economic properties within her group. Second, Brennan and Pettit suggest that esteem is a comparative attitude in that it is dependent on relative rather than absolute standards. Just because someone ranks low on a particular scale—say education—does not necessarily determine whether or not I esteem him. There is an essential social context involved, for, although he may rank poorly on the absolute scale of education, relative to the community he may be fairly well-educated, and so I would esteem him given the comparison class. This feature of esteem should not be particularly surprising in the context of SES


hierarchies, where, as I pointed out earlier, an individual’s relative position depends on the social group in which she is being evaluated.

Third, Brennan and Pettit argue that esteem is a directive attitude, in that it is largely focused on evaluating areas that the subject of esteem is thought to be able to control.\textsuperscript{175} This is in contrast with an attitude like awe, which responds to certain characteristics like beauty not directly under the control of those who possess them. Esteem, on the other hand, is closer to being a reactive attitude like resentment.\textsuperscript{176} It is formed in response to circumstances in which what has happened (or how the individual was disposed to behave) was perceived to be under that agent’s control.\textsuperscript{177} Esteem is also directive in the sense that it can encourage people to make adjustments to the characteristics or actions under their control that are the target of esteem. In this context, esteem can encourage individuals to acquire the traits or abilities that would be useful in obtaining high SES. Understanding esteem as a directive attitude helps explain the way in which it motivates people to achieve higher SES.\textsuperscript{178}

To summarize, high SES individuals are approached with certain attitudes. Among the most important of these is esteem. We evaluate others with regard to specific features—in this context, the same social and economic factors that make them of high SES—and adopt an approving attitude. SES seekers are motivated to establish and maintain this approval.

\textsuperscript{175} Not all of the attitudes that are aimed at individuals with high status are directive in this sense. For example, respect may be an appropriate attitude toward individuals with characteristics like native intelligence not directly under their control.

\textsuperscript{176} See Strawson, 1972.

\textsuperscript{177} This is not to say that such judgments regarding control are always correct; disesteeming someone because he lacks kindness may make more or less sense depending on what one believes about how much of our character is under our control.

\textsuperscript{178} I will leave aside the question of whether esteem is sought for itself or for what other benefits being held in high esteem can generate. See Brennan and Pettit, 2004, p. 29-31.
The benefits of being esteemed are such that individuals will aim for higher SES even if the goods that accompany such a position are not, as Smith noted, necessary for mere survival. Of course, not all attitudes—for example, jealousy or envy—directed toward those with high SES are positive. But even an attitude like jealousy can be affirming insofar as it reinforces that one’s position or possessions are significant enough for someone else to want them.

The converse point regarding esteem is that part of what makes being of low SES so miserable is that, beyond material deprivation, one is potentially subject to the negative attitudes of others. If the benefits of high SES include respect, sympathy, admiration, and esteem then the costs of being low SES include contempt, disdain, scorn, disesteem, and condescension. As John Adams, the second American president, noted, “The desire of the esteem of others is as real a want of nature as hunger—and the neglect and contempt of the world as severe a pain as the gout or stone.”\footnote{Adams, 1805, p. 28.} Although perhaps overdramatized, Adams’ analogy is particularly telling insofar as it connects the afflictions of disesteem with poor health.

The effects of disesteem are not limited to those with the very lowest SES; it is a recurring literary theme that people with newly acquired wealth are not accorded the same esteem as old money. The whispered gossip, the contemptful dismissals, and the disdainful looks express attitudes whose purpose is to reinforce the divide between money and money properly aged. Both F. Scott Fitzgerald’s Jay Gatsby and Henry James’ Daisy Miller are ruined, in part, because of these attitudes. Their wealth is not sufficient to obtain the respect or acceptance of the upper class. Their reactions—especially Gatsby’s self-destructive pursuit of wealth to impress Daisy Buchanan—illustrate the effects that the attitudes of others can have on our actions. Just as SES is not an all or none phenomena, but exists as a
graded hierarchy, esteem and disesteem can be targeted at individuals just above or below oneself in the gradient.

If disesteem is the dark mirror of esteem, then, when we adopt this attitude toward other people based on their social and economic characteristics, we are judging that disapproval is the appropriate response to someone in their position. Individuals with low SES are disesteemed or held in contempt because of the factors that make them low SES, whether this is race, ethnicity, education, class background, or income. Another common literary theme, found in books ranging from Fitzgerald’s *The Great Gatsby* (again) to Gabriel Garcia Marquez’s *Love in the Time of Cholera* to Nicholas Sparks’ *The Notebook* is the love between a high status woman and a poor suitor, who is held with disregard by her family. In each of these cases, being the subject of disesteem motivates the lead male character to “better himself” and improve his SES.

It is tempting to think that attitudes like respect and esteem or disrespect and disesteem are entirely reactions to the perceived SES of an individual. We consider what she owns, her place in the world, how others respond to her, etc and then determine whether she is worthy of our esteem. On this view, the attitudes we adopt depend on an individual’s preexisting SES. One thing that has been generally missed in the discussion of attitudes and SES seeking, however, is that exposure to positive or negative attitudes is not merely an outcome of one’s position in the social hierarchy; these attitudes also work to establish our SES.

First, ‘having been subject to negative attitudes’ is itself a characteristic that can determine our SES. People react to the attitudes that others adopt toward us. If I am regularly treated with esteem, it is more likely that people who observe this will treat me with
esteem (and vice-versa). Thus, esteem or disesteem from others can become a self-fulfilling prophecy whereby, because we have been treated or thought of in a certain way in the past, it is more likely that people will approach us with similar attitudes in the future. Second, the internalization of positive or negative attitudes can make us more or less likely to seek other characteristics that might improve our status. Consider, for example, the situation of Southern blacks under Jim Crow laws. In reaction to certain features—being poor, rural, but most importantly, simply being black—white Southerners adopted a set of attitudes—dismissal, condescension, contempt—that produced a climate in which it was impossible for blacks to see themselves as anything like equals.

Marginalization did not occur merely through deprivation of resources; it also required a set of attitudes to undergird relative social standing. Consistent with the previous discussion, these attitudes were a response to the perception of blacks as being of low SES. Along with legal sanction and outright force, attitudes like contempt perpetuated the status quo by actively discouraging blacks from pursuing a better life. The repeal of the Jim Crow laws and the cultural changes that took place with the civil rights movement made significant inroads in dismantling the legal framework of segregation and, to a lesser extent, changing the public perception of what attitudes were appropriate to have toward blacks. Even if, however, blacks were no longer directly exposed to disesteem or contempt, the ways in which these attitudes were internalized served to reinforce their social.

In this context, negative attitudes are initially a reaction to the perceived socioeconomic value of an individual (his or her SES). Over time, a person who is the subject of these attitudes may come to believe that they reflect an accurate assessment of his value. In this way, negative attitudes initiate a self-fulfilling process whereby individuals
internalize the external assessment of their value and do not seek a higher SES—the characteristics that would give them a higher value in the eyes of others—because they are convinced that they deserve low SES. This phenomenon, whereby what an individual expects/desires from his life adjusts to his low social position, has been widely studied in economics where it has been labeled 'adaptive preferences.' As the philosopher Martha Nussbaum notes, “We are especially likely to encounter adaptive preferences when we are studying groups that have been persistent victims of discrimination, and who may as a result have internalized a conception of their own unequal worth.”

The character Uncle Tom in Harriet Beecher Stowe’s *Uncle Tom’s Cabin* is a classic illustration of the effects of adaptive preferences. Although significantly more complex as Stowe originally portrayed him, Uncle Tom now stereotypically represents an African-American who is satisfied with his low SES, believes that he deserves it, and does not take steps to better his position because of this belief. The internalization of negative attitudes and its relationship with perpetuating SES levels is also well-documented outside of literature. For example, in a 2003 study, researchers took high and low SES white and black first graders and had them rank the prestige associated with various occupations. Black children viewed jobs with only African-American workers as being of lower status than identical occupations presented with only white European workers, an effect that persisted when adjusting for SES.

From a young age, these children learned to associate being black with having low social SES, presumably not through any explicit channels—both black and white children in

---


the study said that blacks and whites should equally be able to do whatever jobs they want—but through the internalization of implicit social cues. Included in the social environment that might produce the association between low SES occupations and being black are the attitudes that have characterized race relations in the United States for the better part of two centuries. As one of the study author notes, these implicit or unconscious associations become a self-fulfilling prophecy whereby the jobs that minorities take “may, across time, be viewed as lower in status simply as a function of the race of the worker, and consequently show decreasing levels of pay and prestige.”  

The larger philosophical point is that our evaluation of our own worth, the worth of our projects, and our place in a social system is affected, in part, by other’s assessment of these things. When negative judgments are made—and the corresponding attitudes of disesteem, contempt, disdain, scorn, or condescension are adopted—our SES can suffer. We react to these attitudes. In some ways, they are as significant in determining what individuals can do to shape their lives and how they react to negative events as having the necessary resources. Attitudes like contempt can reinforce and solidify self-perceptions of what we are capable of doing and how we understand our place in the world, creating the same kinds of constraints as poverty or lack of education.

More generally, if contempt is a fixed feature of our world or if we come to believe that contempt toward us is justified or deserved—that it reflects an accurate assessment of our worth—we will be less likely to strive for the things that might give us higher SES. We


184 There is an extensive literature on contempt as a moral attitude, which attempts to answer when contempt is a proper attitude for expressing disapproval toward immoral individuals. I am not here specifically concerned with moral contempt but the more common sort, which is an attitude that regards another person as inferior or worthless without that person necessarily having done something immoral. Again, I am also not speaking to the normative question of whether this sort of contempt is ever justified (although I suspect that it is not). See Mason, 2003, pp. 234-72.
lower our expectations and a kind of defeatism settles in along with the belief that it is inappropriate for us to ask for a better life. If opportunities for advancement are offered we may hesitate to take it advantage. We will be less likely to demand our fair share of resources having been convinced that we are not worthy of them. Our SES is not determined merely in virtue of what we possess but also by our ability to acquire the social and economic characteristics that generate higher SES, an ability that can be significantly constrained by our beliefs about what we are capable of achieving.

One potential criticism of this account is that it fails to distinguish two potentially separate responses that society might have toward low SES individuals: contempt or neglect. If we merely ignore those with low SES and fail to respond to them with any attitudes (positive or negative), then it may be hard to see how negative interpersonal attitudes could work to affect SES. For example, servants in Victorian Britain were largely invisible to members of the upper-class, barely registering on the social radar. Under conditions of status neglect rather than contempt, it may be hard to argue that servants had part of their SES determined by the attitudes of their employers. I have three responses to this suggestion: first, neglect may simply be an extreme form of contempt, reinforcing the idea that one’s status is so low that one does not even register. Second, while neglect may have been the most common day-to-day experience British servants, if a member of the household were to fail to perform a required task or act in a way ‘inappropriate’ for their social class—attempt to join a dinner party rather than serve the food—it seems likely that their employers would direct hostile attitudes toward them. Third, even if the highest members of a social structure never adopted negative interpersonal attitudes toward the lowest individuals, persons in

---

185 Geoff Brennan initially pointed out this distinction in reference to Adams’ claim that both contempt and neglect prove equally motivating for individuals to seek higher status.
intermediate social positions might. Since status is a graded phenomenon, there may always be someone close enough to us in status to treat us with contempt rather than neglect. Thus, negative interpersonal attitudes are likely to be a salient determinant of low SES, even if they are not directed ‘from the top’ so to speak.

The internalization of positive attitudes can also affect SES. While it may be true that higher SES individuals have further to fall, the fact that they have been respected throughout their lives can provide a psychological cushion that may make it easier to rise again in SES gradient or at least to believe that such a change is possible. Illustrations of this phenomenon abound in the popular media. For example, in an episode of the television show, *The Simpsons*, entitled “The Old Man and the Lisa,” one of the secondary characterize, an elderly millionaire named Mr. Burns, loses his fortune and is forced to live in a retirement home.\(^{186}\) He befriends another character, Lisa Simpson, who encourages him to collect recyclables as a way to start rebuilding his fortune. Over time, Mr. Burns modifies Lisa’s idea and gets the capital—based on his old reputation—to start a recycling plant (that doubles as a fish cannery) and once again becomes the richest man in town. Of the several advantages that Mr. Burns has over an individual who has never been high SES is the fact that he has years of experience of being esteemed by others, which reinforces his belief that he can react well to negative events in his life. In other words, exposure to those attitudes helps give him a sense of control.

I have emphasized interpersonal attitudes because the responses that others have toward us, especially contempt, disdain, condescension, disrespect, or scorn, can play the same role in determining our SES as depriving us of material goods. As William James noted, “If every person we met ‘cut us dead,’ and acted as if we were non-existent things, a

---

\(^{186}\) *The Simpsons*. 1997, Episode no. 21.
kind of rage and impotent despair would before long well up in us, from which the cruelest bodily torture would be a relief.” ¹⁸⁷ Two centuries earlier, Bishop Butler wrote, “There is such a correspondence between the inward sensations of one man and those of another that disgrace is as much avoided as bodily pain, and to be the object of esteem and love as much desired as any external goods.” ¹⁸⁸ Being the object of disesteem over long periods of time may be a significant enough blow that, even if we are no longer subject to the attitude, its effects persist. We remain convinced that we are of low SES and should not try to improve our position.

In case the connection between attitudes, status, and health seems too theoretical there are also empirical data suggesting a relationship. Recent studies have argued that changes in the attitudes that supervisors have toward employees can have a positive influence on workers’ long-term health. ¹⁸⁹ The most significant improvements in health outcomes occur when supervisors “consider their employees’ viewpoint and are able to suppress personal biases.” ¹⁹⁰ Furthermore, studies have shown that workplace bullying, which includes co-workers or supervisors who adopt attitudes designed to isolate or exclude others or “to torment, wear down, or frustrate,” increases rates of sick absence, depression, and heart disease among bullied staff. ¹⁹¹ The internalization of negative racial attitudes has been correlated with higher rates of substance abuse ¹⁹² and depression ¹⁹³ among American

¹⁸⁷ James, 1950, pp. 293-4.
¹⁸⁸ Butler, 1983, Sermon 1
¹⁹⁰ Kivimaki, et al., 2004, p. 931.
blacks. Studies of black Caribbean women have shown a relationship between indices of internalized racism and abdominal obesity and insulin resistance that persists even after adjusting for potential confounders.

Similar studies on the internalization of negative social attitudes amongst gays and lesbians have shown a close correlation between the degree of internalized homophobia and rates of alcohol and drug use, high risk sexual activity, and suicide. Controlling for perceived discrimination significantly attenuates differences in psychiatric morbidity and mortality between heterosexuals and homosexuals. To put these data in Marmot’s terms, the attitudes of others help determine our SES, which, in turn, predicts how much autonomy we have and how healthy we will be. Insofar as interpersonal attitudess shape our SES, they help determine the events to which we are exposed as well as our ability to dictate the outcome of those exposures. Individuals and populations subjected to negative attitudes lose or fail to attain higher SES and their health suffers accordingly.

Implications for Daniels

What are we to say about those who have part of their SES (and health) determined in virtue of the attitudes of others? As Daniels notes, when primary goods are distributed

196 Hammelman, 1993, pp. 77–89
198 Rofes, 1983. For an overview of the research relating internalized attitudes and health outcomes amongst gays and lesbians see Williamson, 2000, pp. 97-107.
199 Mays and Cochran, 2001, pp. 1869-76.
appropriately, any residual health disparities are permissible. The implication is that, if a person is of low SES in a society that bases SES on education and income but has an unfair distribution these goods, something should be done to address his worse health. Specifically, we should arrange institutions to allocate education and income more equitably, thus flattening the SES gradient and adjusting the corresponding differences in health. If, however, that same person is of low SES because of the negative attitudes that other citizens have adopted but has a fair share of primary goods, there is nothing else to say about his worse health.

This seems too contingent a fact on which to rest our judgment about these cases. If anything, health disparities that find their origin in the negative attitudes of others are less justified than those that result from differences in income or education. A social system that creates significant disparities in the possession of material goods may be unjust, but there is something especially pernicious about arrangements in which higher SES individuals adopt attitudes designed to convince the less fortunate that they deserve their deprivation. In this context, the use of negative attitudes is a factor that shapes SES, autonomy, and health but appears to fall outside of the traditional domain of distributive justice. Insofar as it is morally impermissible to allow such attitudes to generate health disparities, however, we may have grounds to question the adequacy of Daniels’ normative approach.

How might Daniels respond to this challenge? Are there resources within Rawls’ account to address the impact that attitudes have on our SES and health? If not, are there other theories of justice that are broad enough to support interventions aimed at changing interpersonal attitudes or do we require a different kind of moral approach altogether? These will be the central questions of the next chapter.
Chapter VI: Justice and Status

Of the advances made in Human Immunodeficiency Virus (HIV) prevention over the past twenty years, one of the most important has been the reduction of social stigma surrounding infection with the disease. It has been well-documented that the stigma associated with HIV discourages individuals from being tested for the virus, and, once it has been diagnosed, from seeking adequate treatment. In the United State, the negative attitudes surrounding HIV are related, in part, to its association with marginalized groups like gay men and intravenous drug use. In the developing world, stigmatization springs from the perceived relationship between HIV and promiscuity or prostitution. In both cases, the medical consequences of having AIDS are compounded by the social stigma of the diagnosis. For example, one in three HIV positive men and women in Cape Town, South Africa reported feeling dirty, ashamed, or guilty because of their HIV status and internalization of stigma has been associated with increased risk for depression among this population.

The relationship between social stigma and the diagnosis and treatment of HIV/AIDS again illustrates the impact that attitudes like disdain or contempt can have on our health. In the previous chapter I argued that, while these attitudes can directly affect health, as is the case with HIV, they may also indirectly do so as one of the broader determinants of socioeconomic status (SES). I suggested that Daniels’ Rawlsian approach may be unable to account for the intuition that it is morally unjustified for attitudes to play this role in

---


generating health disparities. In this chapter, I will consider what resources are available to Daniels and to theories of justice, considered more broadly, to address this challenge.

*Status and Self-respect*

The first response is to insist that Daniels’ approach is sufficient, in its current form, to address interpersonal attitudes as a determinant of SES. For example, one might note that principles of justice are founded on a concept of persons as free, equal, and worthy of respect. The state has a duty to treat people accordingly, an obligation that is discharged, in part, by adopting principles that distribute primary goods fairly. For example, Rawls writes that, “A desirable feature of a conception of justice is that it should publicly express men’s respect for one another.”202 The political philosophers Ronald Dworkin and Elizabeth Anderson also emphasizes the connection between principles of justice and the obligation of the state to treat its citizens with respect; Anderson writes, “Goods must be distributed according to principles and processes that express respect for all.”203 Such principles reflect an underlying concern, not just with the goods themselves, but also with respecting persons—an attitude opposed to contempt or disdain. In adopting the principles of justice and implementing policies based on them, the state expresses respect for its citizens. Since the just state displays the same attitude toward all persons, it might be thought that this would dilute the effects that other attitudes have on SES hierarchies. If the state respects me, the thought might go, what does it matter if others hold me in contempt?

202 Rawls, 1999, p. 156.

The problem with this argument is that the attitudes in question are primarily between citizens, not between the state (or social institutions) and its citizens.\(^{204}\) In the examples above, it is the beliefs of individuals and groups that create and reinforce the attitudes that influence SES. There is no guarantee that a state whose policies reflect respect for its citizens will be populated by individuals who display respect for one another. In fact, individuals who believe that the respect shown by the state to certain groups is unwarranted may harden their own negative attitudes. For example, the initial resistance to desegregating Southern schools was not just based on the belief that it was inappropriate to educate blacks alongside whites, but that it was wrong for the federal government to impose its vision of equality on what was perceived as a clearly inferior group. This is not to dismiss the importance of what state policies express, only to suggest that more is required if our concern is with interpersonal attitudes. As Anderson insists, there is something deeply disturbing about social institutions that distribute goods fairly, but view the recipients of goods, for example, the disabled or untalented, with pity or condescension.\(^{205}\) Nevertheless, respect on the part of the state will not be sufficient to adjust SES gradients in the absence of a mechanism for addressing the attitudes between citizens that create and reinforce SES.

The second approach to the problem of SES and attitudes is to fall back on something like Rawls’ discussion of self-respect as a primary good.\(^{206}\) As Daniels notes, “Without self-respect, however, it is difficult for individuals to use their capabilities as individuals and citizens; since social structures critically support or undermine self-respect, Rawls includes

---

\(^{204}\) There are some notable exceptions to this point, including the marriage rights of gays and lesbians in the United States. In the developing world, where state-sanctioned discrimination of certain ethnic or racial groups is far more common, I suspect that changing the attitudes that government policies express toward these citizens would have a more significant impact on their relative status.


\(^{206}\) Rebecca Walker initially suggested this argument as a possible reply for Daniels.
the social basis of it on the index.”207 Thus, in addition to redistributing the determinants of SES that are connected to the other primary goods, justice as fairness would also require us to evaluate and reduce socioeconomic gradients that undermine self-respect. How might this help us address the role that interpersonal attitudes play in producing relative SES? Self-respect, according to Rawls, involves, “a person’s sense of his own value, his secure conviction that his conception of his good, his plan of life, is worth carrying out.”208 It allows us to move forward with our projects, freeing us from certain kinds of self-doubt and affirming our worth.

And this seems like exactly the sort of armor that might protect us from the attitudes of others. If I am confident in my life plan then perhaps the approval or disapproval of others will not—as they do in the earlier examples—affect my ability to carry it out. Daniels does not make this argument but it might fit within his larger framework: the two principles of justice not only distribute the liberties, material goods, and opportunities that construct social status, but, by apportioning (the social bases of) self-respect, they create the conditions necessary to protect us from the unwelcome attitudes of others. In so doing, they reduce the role these attitudes play in generating SES hierarchies.

There are two things to say about this argument. First, Rawls does not talk about self-respect as a good whose distribution the state is to direct in a manner similar, for example, wealth. Nor does he apply the two principles to self-respect in the same way as other primary goods such that societies might achieve fair allocation by adjusting pools of available self-respect in a manner analogous to material resources. Instead, his discussion of

207 Daniels, et al., 2004, p. 77.

208 Rawls, 1999, p. 386. Although the definitions of self-respect are legion, Rawls’ version is probably closer to self-esteem. For contrasting discussions see Hill, 1991, pp. 4-18 and Didion, 1968, pp. 142-8.
self-respect, which takes place largely in Part III of *A Theory of Justice* where the concern is the stability of justice as fairness, turns on the ability of the two principles, when fully realized, to create the conditions necessary for self-respect. His emphasis is on the role that respect from the state has in fostering its citizens’ self-respect with the idea that institutional respect will help to generate self-respect. Rawls writes, “The basis for self-respect in a just society is . . . the publicly affirmed distribution of fundamental rights and liberties.”209 If the state were to treat some people as less equal in this regard, providing only limited rights, the results would be potentially damaging to those individuals’ self-respect.210

For Rawls, the attitudes that public institutions express are essential for assuring citizens of their own worth. These attitudes are also necessary to assuage the drive for higher relative position that might otherwise destabilize the well-ordered society. In such a state, “the need for status is met by the public recognition of just institutions.”211 This recognition, Rawls suggests, would help to prevent the formation of socioeconomic status gradients—particularly in income—that we see, for example, in the contemporary United States. In this context, self-respect does not provide a shield against the unwelcome attitudes of others but helps ensure the stability of a society already ordered by the two principles.212 Self-respect is such an important good that individuals who have it are unlikely to engage in behavior that might jeopardize the political structure that has provided it.

If, however, self-respect can only be realized within an environment governed by the two principles, it is difficult to understand how societies like ours might make use of it to

---


reduce SES gradients. It is easy enough to see that incremental changes in, for example, the distribution of wealth in contemporary America, might affect relative SES, but it is unclear how manipulating a good that we cannot access now or in the foreseeable future might do the same. Self-respect, at least as Rawls envisioned it, is not accessible to us when surveying those things we might leverage in our society to affect the steepness of SES gradients and, thereby, health disparities. It is not helpful in this context to point out that self-respect helps check the growth of SES hierarchies in a hypothetical society without an account of how we might establish it in a very different environment.

Even given such an explanation, however, there is a deeper problem with any approach that relies solely on self-respect to address SES differences created by interpersonal attitudes. These accounts must rely on self-respect having an unrealistically strong ability to resist the attitudes of others. Without some external confirmation of our worth or that of our projects, even the most self-assured would find it difficult to continue affirming the value of either, especially in the face of hostile attitudes like contempt. Rawls recognized this point, writing that, “Our self-respect normally depends upon the respect of others. Unless we feel that our endeavors are respected by them, it is difficult if not impossible for us to maintain the conviction that our ends are worth advancing.”²¹³ Although undoubtedly important, by itself, self-respect is not sufficient to counteract the effects that interpersonal attitudes have on SES. Uncertainty about our value creeps in and, faced only with the denigrating attitudes of others, we begin to suspect that their assessments are correct. We are of lower status; internal ‘bucking-up’ may not be enough to convince us otherwise.

Rawls’ has two solutions to this problem. First, he argues that, in a society structured by the two principles of justice, there will be communities made up of individuals with

²¹³ Rawls, 1999, p. 156.
similar projects and goals. These associations will help to assuage self-doubt amongst their members and reinforce the sense that what they are doing is worthwhile. In other words, they represent groups of people whose attitudes will help ground our self-respect and potentially resist the negative attitudes of others. Like self-respect, however, these associations are not available outside of a society ordered by the two principles and so cannot be utilized as a mechanism for reducing the SES gradients that confront individuals in existing social systems. Furthermore, even granting that an individual’s community will esteem her, there is still the question of how the negative attitudes that one association has toward another would affect the SES of these groups. An individual might be highly esteemed within her association, but, if other communities disesteem that association, her overall SES would still suffer (e.g. the AV club president and the quarterback of the football team).

Rawls’ second solution to the inadequacy of self-respect in confronting the attitudes of others is to argue that the contractors in the original position would endorse a duty of mutual respect. Because their “self-respect and their confidence in the value of their own system of ends cannot withstand the indifference much less the contempt of others,” the contractors would support an obligation to treat other members of the well-ordered society with the “respect which is due to [them] as moral beings.”\textsuperscript{214} Mutual respect, in this context, includes a willingness to give reasons for actions that affect others, take up the situation from their point of view, and maintain an awareness of their feelings and aspirations. Having other citizens who adopt this attitude will bolster our self-respect, allowing us to conceive of our projects as worthwhile.

\textsuperscript{214} Rawls, 1999, pp. 297.
A few comments on this argument: first, as with his discussion of self-respect, Rawls is more concerned with the way in which a duty of mutual respect would work to stabilize the well-ordered society. Second, it is unclear whether the contractors would, in fact, adopt a duty of mutual respect, given that they start from the position of mutual disinterest. Finally, and most importantly, although the duty of mutual respect is generated in a similar manner as the two principles, it is not an obligation we have to one another because of considerations of justice. There is, in fact, a separate natural duty of justice focused on supporting just institutions and working to develop them where they do not otherwise exist. The duty of mutual respect, in contrast, is something we owe to one another because we are moral beings (i.e. we each have a sense of justice and a conception of the good). Thus, while a duty of mutual respect may significantly alleviate the effects of interpersonal attitudes on SES, it can only be invoked from a framework concerned with issues beyond the traditional scope of justice, a point I will emphasize later in this chapter and the next. Thus, the Rawlsian well-ordered society may take into account interpersonal attitudes, but only because of considerations outside of justice.

A final way in which Daniels might expand his Rawlsian approach to address interpersonal attitudes is by considering whether they might be taken into account at a different stage of Rawls’ political structure. So far, my discussion of Rawls has focused on the most abstract level of justice, where the two principles are formulated from the original position. In the second part of *A Theory of Justice*, Rawls lays out a four-tiered structure to help determine how the principles might apply given increasing information

---

215 Thomas Hill initially suggested this as possible reply for Daniels.
about the empirical circumstances in which we find ourselves. For example, he suggests that during the legislative (third) stage we are to evaluate laws and policies both with reference to the two principles and to the procedures decided in the constitutional (second) stage. It is possible that the legislature might adopt policies designed to limit the impact that negative attitudes have on SES. In this way, they could work to address all of the broader determinants of health as more empirical information became available in a way that still starts from the overall Rawlsian approach. For now, however, this is a placeholder for a real argument. In the absence of further specification in this area, the mere possibility that another argument may be available for Rawls/Daniels does not establish that the position can adequately address the effect of interpersonal attitudes on health. Finally, it is important to note that Rawls specifically excludes interpersonal attitudes from the focus of his work in *A Theory of Justice*, writing, “We also call the *attitudes* and *dispositions* of persons, and persons themselves, just and unjust. Our topic, however, is that of social justice [emphasis added].”

Social justice, which centers on the basic structure of society, has, at best, a derivative concern with interpersonal attitudes.

Insofar as these replies on behalf of the Daniels/Rawlsian project are insufficient there are, as I see it, two options. The first is to continue to explore the requirements of justice so as to broaden its domain to cover the relevant socioeconomic determinants of health, including interpersonal attitudes. The second is to look outside of justice for grounds on which to evaluate the moral permissibility of SES gradients. If we go the first

---


218 For a survey of how various theories of justice might, in their current form, approach the epidemiological data cited here see Marchand, et al., 1998, pp. 449-67.
route, which will be the topic of the remainder of this chapter, we may have to look to conceptions of justice other than Rawls’. For example, Marmot’s preferred normative framework for understanding the relationship between socioeconomic status, autonomy, and health disparities comes from the economist Amartya Sen’s capabilities approach to justice. As with Rawls, it may be helpful to briefly sketch Sen’s general position before explicitly focusing on its application to health disparities.

Capabilities and Status

Sen’s primary concern is equality in the actual freedom that an individual has to choose between the alternative lives she could lead. He argues that a person’s life is made up of a collection of activities and doings that he labels functionings. His examples of functionings include “being adequately nourished, being in good health, avoiding escapable morbidity and premature mortality . . . being happy, having self-respect, taking part in the life of the community.” When grouped together a collection of functionings makes up a particular life. For example, the life of a physician consists, in part, of performing physical exams or interventional procedures; recognizing specific diseases; having certain equipment available for these tasks; identifying the appropriate medication, etc. The life of a journalist consists, in part, of identifying stories; interviewing witnesses and participants; taking notes and reconstructing them as a written narrative. There are, of course, certain functionings that will be part of most lives—having adequate food and water; communicating; having social relationships—and others—taking a patient’s pulse; applying mens rea; working with bronze—that are more specialized. Basic and specialized functionings are closely interrelated and build off one another; interviewing a patient requires communicating and a

---


successful interview with a patient can translate into income that can be used to obtain adequate food and water.

Sen, however, is not concerned with just any sort of being; his focus is on well-being. In this regard, he argues that not all lives are equally instances of well-being. For example, picking locks, cracking safes, escaping from the police, etc. are functionings that would, in part, constitute the life of a successful thief, but the mere fact that a set of functionings can be brought together to form a life does not determine whether we should support someone obtaining or having access to those functionings. Some functionings, he argues, have an intrinsic value that makes them worth pursuing. These specific functionings form a life of well-being.

Sen does not suggest, however, that there is only one kind of life that constitutes being well; there is a plurality of options that people may pursue and their goals may change over time as they rethink the kind of life they want. Sen thus emphasizes the role of choice in deciding which set of intrinsically valuable functionings one should try to achieve. In this regard, he defines an individual’s capabilities as the set of lives—or groups of functionings—among which she can pick in deciding what particular life she wants to attempt to lead. For example, in order to be capable of becoming a physician, a person must be able to achieve a wide range of functionings including having basic needs met—food and water, adequate housing, good health—participating in political processes—having freedom

---

221 The nature of this value is somewhat obscure. At points, Sen seems to suggest that it springs from basic, cross-cultural human needs. See Sen, 1991, pp. 39-42. In other places he cashes it out in terms of functionings that we would have reason to value, writing “In the capability-based assessment of justice, individuals claims are to be assessed . . . in terms of the freedoms [individuals] actually enjoy to choose between different ways of living that they can have reason to value.” Sen, 1990, p. 115. Martha Nussbaum, in developing a variant of Sen’s position, argues that a particular set of functionings should be preferred if they can be agreed upon as part of an “overlapping census” among individuals who otherwise have significantly different positions on the good life for humans. See Nussbaum, 2000a, p. 124.
of speech and freedom from discrimination and violence—accessing non-basic resources—
education, income, medical schools and hospitals—to name a few of the preconditions. If we
limit an individual’s access to certain functionings by, for example, denying her the ability to
participate in political processes or become literate, we significantly restrict the lives she is
capable of leading and thereby reduce her opportunity to achieve well-being. In this way, a
person’s capability set represents her actual freedom to choose between the alternative lives
she could lead. A narrowed capabilities set means that an individual is unable to function
in ways required to lead certain lives. Just institutions, Sen argues, should work to establish
equality in capabilities by supporting access to intrinsically valuable functionings, which
helps provide everyone the same opportunity to obtain well-being.

There are two crucial points of contrast between Sen’s and Rawls’ conceptions of
justice. The first has to do with the basic structure of their theories. Rawls is interested in
a contractualist approach to justice. His goal is to identify fair procedures for how to
distribute certain resources while remaining neutral between the different conceptions of the
good life these resources might be used to support. Recall from Chapter 2 that the
contractors in the original position have a capacity for a conception of the good as well as the
ability to revise that conception (they are rational). They do not, however, endorse any
particular idea of the good life; the primary goods they acquire are used to support whatever

222 Freedom comes into play at two points in Sen’s theory: first, the capabilities set is an expression of the
freedom to choose among alternative lives. Second, Sen argues that freedom to choose is itself an intrinsically
valuable functioning. Exactly how these two senses of freedom (and the nature of their values) relate to one
another is a complicated issue within the theory. See Sen, 1991, p. 49-51.

223 Nussbaum’s interprets this kind of equality of opportunity as requiring the just state to ensure that its citizens

224 In general, the capabilities approach is thought to represent a conception of justice somewhere between the
economist Paul Samuelson’s revealed preference utilitarianism and Rawls’ justice as fairness. See Samuelson,
pp. 241-59.
life plan they have once the veil of ignorance is lifted. Sen, in contrast, offers a substantive theory of human good, which he believes should direct our actions when determining what functionings/capabilities we should support. He is not primarily interested in fair procedures or processes but endorses a ‘metaphysical’ conception of human well-being, as laid out in his set of intrinsically valuable functionings, which serves as a normative guide when we are assessing the relative injustice of various states of affairs.225

The second difference between Sen and Rawls, and the more important one where health disparities are concerned, is the scope of their respective approaches. Rawls is concerned with the equal distribution of primary goods, resources that provide the means for an individual to pursue a specific conception of the good. In contrast, Sen is interested in the actual capabilities into which these resources translate for specific individuals. Merely having equality in resources, Sen suggests, is not sufficient to establish equality in what people are capable of doing with those resources. This is because specific characteristics of an individual may make it more or less difficult for her to translate primary goods into functionings and thus into an adequate capabilities set. For example, given the basic function of surviving, a patient in congestive heart failure will require significantly more resources to achieve this function than one with sinusitis. Supposing equal resources (or even significantly more for the heart failure patient), there will still be marked inequality in capabilities between these two individuals.

---

225 Sen had argued that he is not providing a comprehensive doctrine or substantive idea of the good insofar as “capability reflects a person’s freedom to choose between alternative lives . . . it’s value need not be derived from one particular “comprehensive doctrine” demanding one specific way of living.” Insofar as Sen restricts the set of valuable functionings it is hard to see how he can avoid a commitment to a substantive idea of the good even if he leaves up a wide-range of options within that idea. Sen, 1990, p. 118. Unlike Sen, Nussbaum acknowledges that she is laying out a comprehensive doctrine, but also suggests that insofar as Rawls focuses on a specific set of resources in identifying the primary goods, he is doing so as well. See Crocker, 1992, p. 599.
For Sen, Rawls focuses on the “means to freedom, rather than on the extent of the freedom that a person actually has,” while he (Sen) is also concerned with, “the relevant personal characteristics that govern the conversion of primary goods into the person’s ability to promote her ends.”\textsuperscript{226} \text{It is worth stressing that this is not necessarily a criticism of Rawls. Given his starting point—exploring fair procedures for distributing goods without presupposing a particular conception of human well-being—it may emerge that there are fair procedures that, when implemented, generate inequalities in capabilities. Insofar as Sen’s project is broader in aim, it may be able to speak to these inequalities, but it is not a criticism of Rawls’ approach if it was never designed to address questions regarding the way that resources translate into capabilities.}

In summary, Amartya Sen offers a theory of justice centered on a conception of human well-being. There are certain intrinsically valuable functionings that can be combined in different ways to generate a group of possible lives. Among these functionings are some that relate to Rawls’ primary goods (having self-respect and acting freely) as well as functionings based on properties that Rawls did not directly consider (having good health). The functionings to which we have access establish what we are capable of becoming or, in different terms, our opportunity to obtain a life of well-being. Justice, Sen argues, is concerned with equality in capability sets such that states should work to ensure their citizens have access to the functionings necessary to achieve this equality.\textsuperscript{227} \text{The scope of Sen’s capabilities approach raises the possibility that he may have more to say about the broader

\textsuperscript{226} Sen, 1991, p. 81.

\textsuperscript{227} Importantly, states are not to force their citizens to choose one particular group of intrinsically valuable functionings. The emphasis in Sen’s account is on promoting capabilities. Perfectionist theories of justice, in contrast, argue that the state should promote particular ways of functioning in a more narrow sense than Sen permits. See Arneson, 2000, pp. 37-63.
determinants of health than Daniels.\textsuperscript{228} And, in fact, Marmot argues that the appropriate normative framework for thinking about health disparities is through their impact on capabilities.

Like Daniels, Marmot approaches health disparities by way of their relationship to a normatively desirable state of affairs, in his case, equality of capabilities. Recall that Marmot’s central thesis is that position within a socioeconomic status (SES) gradient is a marker for autonomy. Having more or less autonomy, in turn, causes neuroendocrine changes that lead to health disparities between higher and lower status individuals. At the same time, Marmot also suggests that differences in SES translate into inequalities in capabilities. He writes, “What matters is the degree to which inequalities in rankings [SES] lead to inequalities in capabilities—being able to lead the lives [we] most want to lead” (Marmot 2004, 240). Marmot is not, however, introducing a fourth idea here—capabilities—in addition to SES, autonomy, and health disparities. Instead, he takes his definition of autonomy to be synonymous with Sen’s definition of capabilities. Marmot understands the actual freedom that an individual has to choose between the alternative lives she could lead (i.e. her capabilities) to be the same thing as the kind of control that allows an individual to determine what exposures she encounters over the course of her life as well as dictate the outcome of any negative events that happen to her (i.e. her autonomy). He writes, “A more just distribution of capabilities—control and social engagement—will lead to a more equal distribution of health” (Marmot 2004, 249). Lacking an adequate capabilities set—not being able to lead the lives we most want to lead—causes neuroendocrine activation, which

\textsuperscript{228} It is worth reiterating that Daniels sees the difference between the Rawlsian and capabilities approaches as merely terminological, given that the emphasis in both accounts is on protecting health to establish a space of exercisable opportunities. Given that Sen’s approach may allow us to take into account a broader range of factors that affect health than Daniels allows, however, there may be a more substantial difference between the two positions. See Daniels, 2001, pp. 5-6.
produces poor health. A more just distribution of capabilities will, therefore, reduce health disparities.

If you can determine your exposures and how well events go for you, you will be able to choose among many different possible lives. You will, according to Marmot, have both significant autonomy and a broad capabilities set. By arguing that equality in capabilities is equivalent to achieving equality in autonomy and that it is inequality in autonomy that generates health disparities, Marmot is able to connect the epidemiological literature on the broader determinants of health with a normative foundation for how to think about the permissibility of health disparities.\textsuperscript{229} Justice requires equality in capabilities, which is equivalent to equality in autonomy. And, if we equalize autonomy, we will reduce health disparities. In order to achieve this end, we will have to improve access to the broader determinants of autonomy, functionings like education, income, vaccination, housing environment, etc.

The philosopher and epidemiologist Jennifer Ruger has also worked to apply the capabilities approach to health disparities. Unlike Marmot, Ruger is not convinced that the broader determinants of health work through a single mechanism like autonomy. Instead, she emphasizes that being healthy is an intrinsically valuable functioning and should be available within a capability set. She argues, therefore, that reducing health disparities would help create greater equality in capabilities. She writes, “public policy should focus on the ability to function, and health policy should aim to maintain and improve this ability by meeting health needs.”\textsuperscript{230} She is open to the possibility that there are multiple variables that

\textsuperscript{229} I actually do not think that Marmot’s version of autonomy as control maps on to Sen’s capabilities approach, preventing Marmot from drawing normative conclusions from his epidemiological data.

\textsuperscript{230} Ruger, 2004b, p. 1075.
impact health needs, possibly through different mechanisms. Thus, while Ruger does emphasize the impact of employment, education, and participation in social life on health, since her focus is on health as a kind of functioning, if it turns out that there are other, more important determinants, she can refocus her account on whatever factors the epidemiological literature produces.

There are two other features of Ruger’s extension of Sen that are worth highlighting here. First, while she is interested in the importance of health as a way of functioning, she acknowledges that there are other areas in which social change might be enacted that would also improve equality of capabilities. Thus we must weigh the political, social, and economic costs of reducing health disparities with other potential interventions that might improve equality in the space of capabilities (e.g. income redistribution). She is also concerned not to subsume all redistributive efforts under the rubric of health policy, as she believes is the case in Daniels’ work. Ruger writes, “[My] approach is also cautious about extending the traditional boundaries of health policy to include all policy domains that affect health.”

By preserving separate spheres for different functionings, Ruger acknowledges that we have to weigh the gains in equality from reducing the impact that different levels of health have on capabilities has to be weighed against other changes that might show a larger equalization of capabilities.

Second, because she is concerned with whichever factors go into shaping health, she is also open to the possibility that the most effective mechanism for reducing health disparities might involve something other than government-directed redistribution of certain resources or opportunities (a point of emphasis in Daniels and Marmot). She writes, “the capability perspective also recognizes the importance of . . . including non-governmental

---

231 Ruger, 2004a, p. 1095.
actors and market forces to achieve public policy objectives . . . the movement in global health toward better public-private partnerships reflects this view.\textsuperscript{232} This opens up a larger space for thinking about how to reduce health disparities, whether it turns out that they are the result of unequal distribution of primary goods or gradients in socioeconomic status/control. Given the current political climate, which is marked by significant disagreement between public versus private approaches to health care provision, acknowledging the availability of multiple government and non-government options may be the first step toward political progress in this area.

\textit{Capabilities and Attitudes}

Importantly, neither Marmot nor Ruger (nor the Acheson Recommendations) specifically address the role that interpersonal attitudes can play in shaping socioeconomic status and control.\textsuperscript{233} But can the capabilities approach be expanded to account for this relationship? One might, in this regard, draw inspiration from Elizabeth Anderson’s use of Sen’s framework as part of what she calls democratic equality. Anderson begins by noting that, historically, inequality was not understood so much as to involve an unfair distribution of resources but as unequal relationships between higher and lower individuals in a social order. While those of superior rank may have had more goods than the people below them, the true mark of inequality, according to Anderson, has to do with the nature of the relationship between those to groups. In an unjust system, “those of superior rank were thought entitled to . . . exclude or segregate [inferiors] from social life, to treat them with contempt, to force them to obey, work without reciprocation, and abandon their own

\textsuperscript{232} Ruger, 2004a, p. 1097.

\textsuperscript{233} Although Ruger notes that, “the capability perspective emphasizes the empowerment of individuals to be active agents of change in their own terms,” a phrase that might be interpreted to address the way that interpersonal attitudes can disempower individuals. Ruger, 2004a, p. 1094.
cultures.” The appropriate aim of justice is to oppose hierarchies that endorse and perpetuate this sort of social relationship whereby “some people dominate, exploit, marginalize, demean, and inflict violence on others.”

From these passages, it should be obvious that Anderson is concerned, in part, with whether certain interpersonal attitudes like contempt are appropriate ways for individuals in different social strata to relate to one another. She argues that it is unjustified for contempt and hostility to limit, for example, whether gays and lesbians can publicly reveal their identities and suggests that attitudes like pity are inappropriate ways for the more fortunate to relate to the less fortunate. She also emphasizes the effect that the internalization of certain negative attitudes can have, noting that norms of femininity encourage “self-abnegation, lack of confidence, and low self-esteem” and reinforce the lower SES of women, even in the face of access to similar resources as men. On what grounds, however, are we to judge whether a certain attitude or way of relating to an inferior is unjust? At this point, Anderson draws on Sen’s capabilities approach to argue that social relationships that result in reduction in equality in the space of capabilities are unjust.

Anderson identifies two sets of functionings that are of greatest concern in this context. First, she suggests that individuals should be able to access functionings that would enable them to avoid the effects of oppressive social relationships. This is a negative goal insofar as it works to prevent one group of people from interfering with another in certain

---


237 This may be a circular argument if oppressive social relationships are just defined as those interactions that reduce equality in capabilities. At some points, Anderson writes as if the wrongness of oppression exists independently of its impact on equal freedoms (capabilities), but, insofar as this is true, she will need a different framework to justify our attention toward oppression, which she does not provide.
ways. The functionings that Anderson has in mind here include living one’s life without excessive domination in personal relationships (as, she suggests, happens to women in some developing nations), having certain psychological resources like a sense of self-respect and self-esteem, and acting with confidence in one’s affairs. The importance of these functionings, Anderson argues, means that others should not adopt attitudes that would threaten to undercut their development.

The second group of functionings Anderson identifies is those necessary for participating as an equal citizen in the state. This is a positive goal insofar as it requires more than just non-interference but active steps to support an individual’s involvement in certain social and political processes. More specifically, Anderson argues that we are obligated to provide access to functionings like voting in a fair political process, engaging in free speech, petitioning the government, participating in the economy, thinking for oneself, freely associating with others, and using public spaces, insofar as involvement in these areas is necessary to stand as an equal citizen.238

Although Anderson focuses on a broad range of functionings, there are some that she excludes from being an appropriate focus of justice. For example, she argues that society is under no obligation to ensure that everyone could function as a good card player, if he or she so desired. This is because being bad at cards neither makes one oppressed nor determines one’s place in civil society in the way that, for example, being black in the 1930s rural South did. More generally, Anderson is not interested in comprehensive equality in the space capabilities, but only equality in the functionings necessary to avoid oppressive social

238 It is important to note that the goal of democratic equality is not to guarantee everyone the same actual level of functioning, but equal access to that level. She writes, “Individuals are free to choose to function at a lower level than they are guaranteed.” Furthermore, she argues that institutions are permitted to make access to certain functionings (e.g. income) contingent on appropriate behaviors on the part of a citizen (e.g. working). Anderson, 1999, p. 318.
relationships and participate in the state as an equal citizen. In this regard, Anderson’s approach may be both broader—insofar as it takes into account the role of interpersonal attitudes in shaping status—and narrower—insofar as it addresses only those aspect of control that are stunted by oppressive social relationships or unequal civil participation—than Marmot’s position.

Ruger, Marmot, and Anderson all rely on Amartya Sen’s capabilities approach to justice in evaluating certain states of affairs (health and health disparities for the first two authors; the foundations of democratic equality for the last). Given that they are engaged in somewhat divergent projects, is there a unifying account that might illustrate how the capabilities approach would approach health disparities in the developed world? While Marmot’s arguments most directly connect capabilities and health disparities, the approach with the least empirical assumptions would be to focus, like Ruger, on being healthy as an intrinsically valuable functioning. Given the goal of equality in capabilities, we are obligated to ensure a certain level of healthy functioning. And, as there are multiple other functionings that work to produce health—including, as Anderson suggests, avoiding negative interpersonal attitudes—we should also work to ensure that these functionings are also available within a capabilities set.

Criticisms

I want to identify three general concerns with the capabilities approach before I consider its specific merits in addressing health disparities. First, even if we disagree with Rawls that a theory of justice should refrain from endorsing a particular comprehensive doctrine or substantive position on human well-being, we might have specific worries about the group of functionings that Sen identifies as intrinsically valuable. Why these
functionings? Why not picking locks or getting away with murder? The concern is that Sen’s list just represents the norms (having self-respect; being literate) of the Western cultural tradition and has no more substantial claim to objectivity than that. It is difficult to decide exactly how to phrase this objection (and how seriously to take it), but the territory—cultural relativism, anti-universalism, etc—is familiar enough. Concerns about the ‘objectivity’ of any particular idea of human well-being are a well-rehearsed complaint against the capabilities approach. Both Sen and his collaborator, the philosopher and classicist Martha Nussbaum, have provided arguments regarding the culture-transcendence or universality or reason-giving force of the functionings that each defends. I mention this issue here only to note that there is a significant meta-ethical debate to be settled over the possibility of an ‘objective’ list of functionings (or transcultural theory of human well-being).

Second, even granting that there are intrinsically valuable, non-parochial functionings it is unclear why we should aim for equality in capabilities. Sen may be correct about the importance of being free to choose amongst different alternative lives, but this point is separate from the idea that everyone should have equal freedom in this regard. There are many goods whose importance we recognize, but whose distribution we do not necessarily require to be equal. For example, a libertarian might argue that, even though having food and water might be very important—even essential—for surviving, this is not enough to show that we should make sure everyone is equal with regard to these resources, especially if

239 See Rorty, 1993, pp. 111-34.


241 A more interesting question is whether or not Sen and Nussbaum’s approach makes sense without the teleological assumptions that allow Aristotle to ask meaningful questions about functioning well as a human being.
doing so would violate the rights of others. In response to this concern, Nussbaum (but not Sen) has argued that the requirement for equalizing capabilities springs from the Kantian idea that we respect human dignity (or treat each person as an end). Since dignity is a constant property across individuals, if we are going to work to improve the capabilities of any particular person or group out of respect for his/her dignity, we must do so equally for everyone else. While this may help explain the grounds for equality of capabilities, it introduces a number of other questions including the nature of respect for persons and why changing capabilities is the appropriate way of expressing that respect.

Third, where health is concerned, the historical motivation for focusing on equality in access to the resources that promote good health is that ensuring equality in health is thought to be an inappropriate (and possibly unobtainable) social goal. Given Ruger’s focus on health as a valuable functioning that should be part of a person’s capabilities set, requiring equality in these sets may have undesirable consequences. As I noted in Chapter 2, demanding equality in health would place a significant strain on the institutional resources that could be used to achieve other ends. Because a certain level of health is necessary to have even a minimal range of choices within a capabilities set, requiring equality in capabilities entails that we work to ensure citizens do not fall below this threshold. Putting aside cases where such efforts are frankly impossible—for example, there is no medical intervention that will make a neonate with anencephaly have even an approximation of the capability set of a normocephalic neonate—there are many diseases where the costs of

---

242 Nussbaum, 2000a, p. 124.

243 Nussbaum, 2000b, p. 2.

244 As Beachamp and Childress, 1994, note, egalitarian theories of justice, “propose that persons be provided an equal distribution of certain goods such as health care, but all prominent egalitarian theories of justice are cautiously formulated to avoid making equal sharing of all possible social benefits [like good health] a requirement of justice.”
providing a minimal capabilities set are prohibitive. For example, patients with end-stage heart failure (i.e. symptoms at rest) could be given a dramatic increase in the freedom to choose among alternative lives (an expanded capabilities set) if they underwent heart transplants, but this does not mean that we are obligated to find the needed hearts, artificial or otherwise. Patients are not entitled to whatever resources are necessary to maintain certain capabilities or even the resources needed to continue functioning at all.

Again, Nussbaum has offered one response to this problem, which is to suggest that the threshold at which equality of capabilities is guaranteed is sufficient high enough to support basic medical care, but not so high that it demands the use of every possible resource. In the heart failure example, we may be obligated to provide access to inotropes, oxygen, diuretics, ventricular assist devices, and pacemakers but not to transplants. More generally, Nussbaum argues that the threshold is best set for each individual government based on the current capabilities and resources of its population and how those are distributed. She does, however, leave it open as to whether justice might require further equalization above this threshold.

Assuming that these more general criticisms can be adequately addressed, I suspect that a combined Ruger/Anderson account has the resources, if appropriately developed, to address many of the broader factors that create health disparities, including interpersonal

245 In her more recent work, Ruger has also adopted a threshold account, arguing that her approach, “supports the allocation of resources to those with health needs in efforts to bring them as close as possible to a threshold level of functioning as their circumstances permit.” In dealing with the problem of “bottomless pits,” Ruger argues for an account of medically appropriate versus inappropriate or futile care that can be used to draw a line concerning the amount of resources we are obligated to provide a particular person. She also adopts a “reasonable accommodation” standard when evaluating how much resources we are to devote to providing the disabled access to the threshold level of functioning. Ruger, 2006, p. 441.


247 Nussbaum, 2000a, p. 126.
attitudes. Like Daniels, however, these writers examine the distribution of goods and status through their effect on an independent state of affairs; namely, equality of capabilities. I again want to emphasize—as I did at the beginning of Chapters 2 and 5—the basic structure of this kind of approach.

Starting with the empirical premise that differences in socioeconomic status—whether mediated through autonomy or otherwise—create health disparities, what is the appropriate normative response to these disparities? For simplicity, let us focus on one aspect of SES gradients: suppose that differences in income are a leading cause of health disparities. Daniels argues that health disparities translate into inequality of opportunity—an unjust and unjustified state of affairs—and that we should, therefore, reduce differences in income as a corrective. He does, however, have to go back and investigate whether such a move is justified within the Rawlsian framework and, if so, on what grounds and to what extent. As it turns out, Rawls’ discussion of the other primary goods does offer a reason to reduce differences in income (by way of the difference principle).

Now, let us take the same empirical starting point—differences in income create health disparities—and examine how Ruger’s approach is structured. She argues that health disparities translate into inequality of capabilities—an unjust and unjustified state of affairs—and that we should, therefore, reduce differences in income. Like Daniels, Ruger has to go back and try to determine if such a move is justified within Sen’s approach. Is having a certain degree of income a valuable functioning to which we should equalize access? If it is a valuable functioning, the capabilities approach offers a reason to reduce differences in income.
Now consider a different empirical premise: persistent stigmatization through negative interpersonal attitudes creates health disparities. Such disparities still lead to inequality of opportunity or capabilities, but now, when Daniels examines the Rawlsian framework for grounds on which we might address these attitudes, there is nothing to account for the impermissibility of adopting such attitudes, or so I have argued. In contrast, I have suggested that, when Ruger looks for guidance in this regard from the capabilities approach, she is more successful. Elizabeth Anderson has argued that we should be able to access functionings that would enable us to avoid the effects of oppressive social relationships (i.e. the stigmatizing effects of interpersonal attitudes).

These are just two examples of the factors that create SES gradients, which in turn, lead to health disparities. We can, as Ruger suggests, iterate a theory of justice across all of the things that work to produce SES (and health), but insofar as the focus is on equality of capabilities, only the determinants of SES that can be justifiably leveraged to reach this condition will fall under the domain of the theory. As Anderson notes, however, the goal of Sen’s approach is not comprehensive equality in the space capabilities. Thus it remains an open question as to whether, like Daniels, there will be some factors that shape SES and health that are not an appropriate focus of justice (conceived of as equality of capabilities) but are nevertheless morally relevant. If our concern is with which health disparities are justified then it may be worth turning our attention directly to the forces that shape health without wondering how these factors fit into a space of equality or worrying about whether they fall under the domain of justice. For the remainder of this dissertation I want to sketch a different option, which is to turn to theories that treat the creation and promulgation of socioeconomic status gradients as their primary concern.
Chapter VII: A Theory of Socioeconomic Status

In 1993, a group of social scientists conducted a study on the attitudes of women from Eastern North Carolina toward breast cancer. In particular, they were trying to understand why black women were significantly more likely than white women to present with advanced breast cancer. Besides race, the factor they found to be most correlated with late presentation was a high degree of fatalism, which was closely tied to perceived control. One of the women in the study reported that, “When you know you have cancer, then there’s nothing for it but to turn it over to God. If you have enough faith, He will heal it and you don’t need no operation. Because there is nothing a doctor can do for you—only God has the power.” The idea that these women’s lives were not under their control—that they were beholden to external forces, whether supernatural, in the form of God, or natural, coming from the demands of their families and the men around them—was a consistent theme in the study. Given that late presentation is closely related to mortality rates from the disease, this is another case in which perceived control—no doubt related to a lifetime of low socioeconomic status—directly impacts health.

251 Similar studies have been done connecting class, control beliefs, and cancer mortality. For an overview of these data and a discussion of fatalism in this context, see Balshem, 1991, pp. 152-72.
Even when black women seek timely care, however, they still have worse outcomes than white women with breast cancer. For example, the black women with breast cancer who come to the cancer clinic in North Carolina where I work are 1.2 times more likely to die than white women with similar demographics, cancer stage, and tumor characteristics.\textsuperscript{252} In terms of overall cancer rates, blacks in the United States are more likely to develop cancer than other populations, with an incidence between 1992-1999 of 526 per 100,000 compared to 480 for whites and 330 for Hispanics.\textsuperscript{253} American blacks are also more likely to die from malignant tumors, with a 30\% higher death rate than whites from all cancers.\textsuperscript{254}

I began this dissertation with a list of health inequalities like the cancer disparities between black and white patients in North Carolina and throughout the United States. The central question has been whether it is morally acceptable for these disparities to exist, assuming that action could be taken to reduce them. Based on new epidemiological data on health inequalities, I have suggested that this question is best understood as being about whether it is acceptable for there to be significant differences in socioeconomic status (or control, if that is how the effect is mediated), with the result being differences in health. Daniels and Ruger have argued that such arrangements are unacceptable because they are unjust and they are unjust because differences in health translate into significant disparities in opportunity/capabilities. At the end of the last chapter I suggested an alternative approach, which was to directly ask—without first offering a theory of justice or an explanation of the injustice of poor health—whether it is acceptable for there to be significant differences in


\textsuperscript{253} American Cancer Society, 2003.

\textsuperscript{254} American Cancer Society, 2003.
socioeconomic status. I will refer to a moral framework that attempts to address this question as a *theory of socioeconomic status*.

A Theory of Socioeconomic Status

At the beginning of *A Theory of Justice*, Rawls suggests that the *concept* of justice concerns the distribution of rights, duties, and the advantages of social cooperation within the basic institutions of society. In contrast, a *conception* of justice is a set of principles that specifies how we should arrange our institutions so as to appropriately apportion basic rights and duties and to weigh claims regarding the advantages of social life.255 Individual conceptions of justice give content to the concept ‘justice.’ For example, Rawls’ conception of justice as fairness gives us two distributive principles necessary for operating just, basic institutions. Justice is not, however, the only attribute of social arrangements. They can be efficient, stable, liberal, tolerant, or possess a range of other virtues. Each of these concepts will have particular conceptions according to which we are to arrange our institutions so as to meet a certain standard of, for example, efficiency. While this means that we need a further mechanism for weighing the various (conflicting) recommendations of these conceptions, when taken in sum, they form a conception of a social ideal or “a vision of the way in which the aims and purposes of social cooperation are to be understood.”256

Consider the following analogy to Rawls’ discussion: first, the *concept* of socioeconomic status concerns the value individuals should have based on their social and economic characteristics, and the appropriate response to that value. Second, a *conception* of socioeconomic status is a set of principles designed to identify the social and economic

---

255 Rawls, 1999, p. 5.

features according to which we should value a person and how to respond to that value.257

As with justice, a conception of socioeconomic status gives content to the concept ‘socioeconomic status.’ Importantly, a conception of socioeconomic status is, in this sense, normative rather than descriptive. It is not interested in whether, for example, level of education is adequately taken into account when evaluating the relationship between socioeconomic status and health disparities. The question of what collection of features best measures socioeconomic status in a given environment is a descriptive task best left to epidemiologists. In contrast, a normative conception of socioeconomic status is focused on whether, for example, level of education is an appropriate ground on which to value persons differently and, if so, how should we treat better (or worse) educated individuals.

As justice is not the only attribute of institutions, however, socioeconomic characteristics are not the only attributes of persons. Individuals can, for example, be more or less rational.258 Various conceptions of rational status will identify the value we have based on our degree of rationality as well as how to respond to that value. For example, creatures who display means-ends reasoning might warrant a different value and response compared to those who can reason about their ends compared to those who display no rationality whatsoever. As is the case with the various aspects of institutions, we may need a further mechanism for weighing the (potentially conflicting) recommendations of these conceptions regarding our value. Taken together, however, they represent a conception of

---

257 The idea of valuing a person based on her economic characteristics may not be as clear as what it is to treat people differently according to their social features (level of education, ethnicity, gender, etc.). The most straightforward way to understand what it would be like to value someone based on their economic characteristics is to ask whether we should esteem based on their income level (or purchasing power, etc.). Ultimately I think that these economic characteristics can be reduced to ‘social’ features if we understand that term widely enough. In order to make this discussion align as closely as possible to the epidemiological data discussed in previous chapters, I will, however, continue to use the term ‘socioeconomic status.’

258 The ability to feel pain or to suffer is another characteristic of persons (and animals) that has been at the center of ethical theories, especially utilitarianism.
status, which covers the overall value we should have in the eyes of others and what the appropriate response should be to that value.

At this point, it may be tempting to think that the right conception of rational status will settle the question of whether and to what degree we should value persons based on their social and economic characteristics, thus eliminating the need for a separate theory of SES. Ensuring that we respond appropriately to one another as rational agents may be our first responsibility—in the same way that establishing just institutions is our first obligation when producing social arrangements—but I do not believe that it exhausts the question of how we should treat one another. This is especially true if we think that very little follows from the bare fact of personhood (or rationality) and that the more important question is how we should treat people given their social features.259 Just as there are still questions of efficiency or tolerance in the society governed by Rawls’ two principles, questions not settled by his conception of justice, there will still be the matter of how to esteem or disesteem persons based on their social and economic characteristics, even after we have adequately responded to them as rational agents. Furthermore, insofar as a specific conception of rational status (or the status we have merely as persons, to put the point in a slightly less Kantian way) does make recommendations about how we should value one another’s social or economic characteristics, I believe that that part of the position should also be understood as providing a conception of socioeconomic status.

Given that the question of what follows from the basic fact of personhood (or agency, rationality, etc.) is still open, there is a significant moral territory to be explored concerning the appropriate conception of socioeconomic status. We should not assume in advance that the answer to the former debate will settle the latter, although it may lay the foundations for

259 This is a point that communitarians have pressed against Rawls, amongst others. See Sandel, 1998.
what an adequate theory of socioeconomic status might look like. My focus in this chapter, then, will be on the concept of socioeconomic status (SES). I want to begin by examining several historical conceptions of SES as a way to explicate the more general concept. My goal will be to illustrate how a moral framework focused on SES might think about health disparities and to contrast this approach with that offered by theories centered on justice. In the next chapter I will discuss one particular conception of SES, based on the idea of a minimal threshold of status, and show how it might be used to generate specific recommendations regarding health disparities in the developed world.

Conceptions of Socioeconomic Status

Consider the system of hereditary entitlements in place for most of European history as one conception of SES. The European feudal and, later, monarchial social structures were centered on the idea that the value a person should have in the eyes of others was, first and foremost, determined by the circumstances of his birth. Individuals of noble heritage—kings and queens and their Court—were accorded the highest value. People of increasingly ‘common’ origin were seen as less and less valuable. The Indian caste system, the feudal structure in Medieval Japan, and the Zhouh dynasty in China, which introduced the Mandate of Heaven, can all be understood as offering a conception of SES on this model. It is important to note, however, that the hereditary conception is not merely descriptive—it does not just identify those features that in a particular social setting do, in fact, give individuals more or less value in the eyes of others—but it is prescriptive, telling us why we should view people as having a higher or lower value.

For example, the Divine Right of Kings can be understood as a claim about why we ought to think of Kings as having a certain SES. First espoused as part of a British legal
framework regarding the authority of the King, it reached its culmination under James I and Charles I in the context of the Counter-Reformation. The Divine Right of Kings established a close connection between theological and political power such that the English King was variously seen as God’s representative on Earth or even as a direct manifestation of God (“Kings are visible gods and God an invisible king” was the phrase that opened Parliamentary debates under James I). As God had attendants and subjects in heaven, arranged from the Archangels to the Choirs of Heaven to the ordinary souls, so did the English Kings on Earth. The House of Lords, the House of Commons, the craftsmen, and the laborers—positions determined largely by the circumstances of one’s birth—represented the corresponding positions in his hierarchy.

The Divine Right of Kings, along with its broader hereditary framework, represents a conception of SES. It identifies the social and economic grounds on which we are to evaluate one another—the kind of life, whether noble or ignoble, into which we are born—and what the appropriate response is to that value. In the case of the British King, his value, as predicated on his divine status, was such that he should be allowed to rule, other citizens should pay him taxes, he should be granted rights to land, etc. In this context, the Divine Right of Kings created such legal fictions as the idea that the king could do no wrong; that he was immortal (“the king is dead, long live the king”), and that, just as God provided divine law, so the English King authored Earthly law. By themselves, the things that the King was given—power, wealth, property—did not establish his SES, but were reflections of what

---


having a certain SES entails.\textsuperscript{262} Of course, not all high SES positions within hereditary frameworks have been associated with reactions we would consider beneficial. For example, early Spanish observers of Aztec rituals noted that each year a young man would be chosen to represent the god Tezcatlipoca and would live a correspondingly high status life during that time.\textsuperscript{263} At the end of the year, he would be sacrificed in honor of the god. In this case, the appropriate response to a highly valued individual was to kill him.

On the other end of the spectrum, the grounds on which someone is ascribed low SES also identify the appropriate reactions to him. The position of the Untouchables in the Indian caste system was understood to reflect the (bad) karma they had acquired in a previous incarnation. The Law of Karma justified the low SES of the Untouchables; they deserved this SES because of their actions in a past life. As such, it was appropriate for them to work in occupations that would make them ‘contaminated’ like burying the dead; to be isolated from the rest of the community; to be forced to walk backward and sweep up their footprints so that high Caste individuals would not accidentally come into contact with any part of an Untouchable. The Karmic underpinnings of their SES also made it inappropriate for the Untouchables to have access to resources like education that would help them have a ‘better’ life.

It should be obvious from these examples that the two parts of a conception of SES—the value we should have based on our social and economic characteristics and the appropriate response to that value—are closely connected. The characteristics by which our socioeconomic value is determined help set up, for example, the resources, goods, attitudes,

\textsuperscript{262} Like esteem, these things can, over time, also be objects that determine status and not merely reflections of it. For the sake of simplicity, however, I will put this point aside.

\textsuperscript{263} See Sahagún, 1982.
and rights that accrue to such a value. The SES of European Kings was predicated on their relationship to God, giving them a value that generated absolute rights of sovereignty. In contrast, for the Untouchables, the way in which their value was ‘assigned’ through the Karmic wheel also provided the grounds on which others were not to work to improve their conditions. Such efforts would represent an interference with the Karmic cycle and so nothing should be done, for example, to educate the Untouchables. When political authority was taken away from the European Kings and social interventions were undertaken to improve the lives of the Untouchables, part of what allowed these changes to take place was a rejection of the conception of SES that underwrote the hereditary assignment of value. Kings were not God’s representatives on Earth, did not have absolute authority, and should not have a certain value merely because of the circumstances of their birth. The Untouchables were not the victims of Karmic retribution but of pernicious social conditions that justified their being deprived of basic necessities. In both cases, the rejection of the underlying conception of SES meant denying that value should be ascribed in certain ways.

In contemporary social systems, SES conceptions tend to be more meritocratic with lower or higher SES based on what goods and resources individuals have achieved using their talents and motivations. Examples of this conception of SES might include Ayn Rand’s Objectivism and Milton Friedman’s version of laissez-faire markets developed in his *Capitalism and Freedom*. Here, SES is not accorded because of the turns of Karma or because a certain arrangement fits with a divine plan but because it is the just desert of hard work and ability. In contrast to the hereditary conception of SES, the perceived worth of goods like income, education, housing, occupational position, etc. do work to establish social position insofar as they have been earned appropriately (and not merely given because of the
circumstances of one’s birth). While the specific goods that mark off high SES may vary depending on which meritocratic conception we endorse, in general, higher SES should be given to those who have used their talents and abilities to accumulate certain goods while the lazy and untalented deserve the low social position that comes with lacking such resources. The idea that value should be ascribed based on what we achieve with our talents and motivations also helps direct our response to high and low SES individuals. Because those with high SES have earned certain goods and resources, it is not appropriate to take them away and give them to low SES people (who do not deserve them). Furthermore, we are to esteem and emulate individuals who have earned their high SES and disesteem and avoid the others.

Not all conceptions of SES support as steeply a structured hierarchy as the meritocratic and hereditary positions. For example, a radical egalitarian conception of SES might hold that we ought to value one another equally regardless of differences in social and economic characteristics, including the circumstances of our birth or what goods we achieve with our talents and motivations. Such a conception might be most closely linked with certain kinds of Communism (recall the use of ‘comrade’ amongst Marxist-Leninists, regardless of rank, or the replacement of the honorifics ‘Miss’ and ‘Mister’ with ‘citizen’ amongst French revolutionaries). In its strongest form, this position would hold that there are no social or economic differences between people that should translate into their being more or less valued or subject to esteem or disesteem based on their perceived value.\(^{264}\) We should not support a hierarchy in which some individuals have higher or lower SES.

\(^{264}\) Rebecca Walker has suggested that it may be a mistake to think of the radical egalitarian as offering a conception of socioeconomic status; instead, they may be better understood as holding that our status as persons (or rational agents) settles the question of whether or not we should use social or economic characteristics to evaluate one another (we shouldn’t). In the case of Marxism, however, it is not clear to what extent their
A minimal threshold conception of SES, which I will discuss in more detail in the next chapter, might hold that, while it may be appropriate for some individuals to have higher SES than others, there is a certain level of SES below which the lowest members ought not to fall. We might, for example, think that treating an individual or group with long-standing contempt is sufficient to lower their SES below this minimal level such that we ought not to adopt this attitude toward them. A different way of expressing this point is that no one should have the SES that a person would have if she were the same person, only allowed to be treated with contempt. We could iterate this sort of move across all or a range of the potential determinants of SES, depending on what we thought a minimal level should reflect. We might believe that a minimum of SES requires that individuals not be allowed to die from easily treatable diseases. That is, no one should have the SES a person would have if it were permissible to allow her to die from easily treatable diseases. A modified egalitarian conception of SES might be silent concerning what other features we ought to take into consideration when bestowing higher SES or it might recommend some social and economic characteristics that deserve higher SES—Hard work? Attractiveness? Diligent pursuit of self-interest?—and hold that all other factors should be irrelevant. In this case, we mark out certain ways of determining SES as impermissible but allow that SES gradients can develop around other characteristics.

Each of the above conceptions also tells us which grounds are inappropriate for valuing other people. For example, the hereditary conception of SES might say that we should not value a person based on his wealth—that money is not sufficient to give someone redistributive conclusions came from facts about personhood; instead, the position was largely centered on a modified version of Ricardo’s labor theory of value. In the case of other radical egalitarian positions—Ronald Dworkin’s work, possibly—I think that, insofar as it provides an answer to what social or economic features make persons more or less valuable, it should be thought of as offering a conception of SES even if it is built from a more basic conception of personhood.
high SES if he is not born into it. In contrast, a meritocratic conception of SES might suggest that no one is to be accorded higher value based on characteristics that he did not earn, like noble birth. Similarly, it would be inappropriate to treat people as having low SES based on, for example, race, gender, ethnicity, or sexual orientation. For the radical egalitarian conception there are no grounds on which it is appropriate to assign higher or lower socioeconomic value; for the minimal threshold approach, there is a certain level of SES below which the lowest members ought not to fall.

It is important to note that I am not defending a particular conception of SES but illustrating the territory that a theory of SES would cover. Once we have settled on a particular conception of SES, we can then consider how closely the way SES is determined in our world mirrors the way in which it should be accorded. For example, a person with a meritocratic conception of SES would have extensive recommendations for change when confronted with a mid-17th century European SES system as would a person with a hereditary conception of SES when viewing contemporary America’s version of meritocracy. But how do we move from the ways in which we currently determine SES to appropriately distributing it according to the conception of SES we have chosen? There are at least two mechanisms for changing SES gradients.

The first is to convince people to stop valuing each other in certain ways. For example, if we want people to stop using race as grounds on which to treat people as having higher or lower SES, there are a number of options. We can encourage a social climate in which people who evaluate SES based on race are, themselves, disesteemed. This was (and is) a large part of the efforts to end race-based discrimination in the United Statues; namely, changing norms about how we view racists. By subjecting these people to disesteem we
reinforce the idea that it is not appropriate to treat certain people as having low SES because of their race.\textsuperscript{265} Such an approach, whether it involves the use of disesteem or taking other valued possessions from a person willing to ascribe low (or high) SES based on an inappropriate characteristic, would, over time, affect the way SES is distributed.

One only has to consider the impact that using certain slurs—Don Imus referring to the Rutgers women’s basketball team as “nappy-head hos”; Isaiah Washington calling T.R. Knight a “faggot”; George Allan referring to an Indian man as a “macaca”—has on a person’s career to see this process in action. Not only were these individuals disesteemed because of the characteristics by which they (improperly) value others, but they were also threatened with the loss of other things they cared about. Don Imus’ radio program was cancelled; Isaiah Washington was fired from the television series \textit{Grey's Anatomy}; George Allan lost his campaign for the Senate. Given sufficient negative or positive incentives people will at least refrain from publicly expressing the idea that race or sexually orientation are appropriate grounds for treating someone as having low SES.

Some beliefs about the social and economic grounds on which people should be valued will undoubtedly be very resistant to change, even in the face of disesteem or other incentives. More extreme efforts to alter these beliefs—for example, subjecting racists to prolonged public humiliation or even imprisoning them—may be inconsistent with the values at the center of the new conception of SES. For example, a minimal threshold position might hold that only those mechanisms that would still acknowledge a certain level of value can be used to change a person’s beliefs about the low SES of others. In this context, it might be

\textsuperscript{265} Brennan and Pettit have argued that the distribution of esteem can serve as an alternative mechanism to the invisible hand of the market and the iron hand of government for moderating social behavior. By changing the way we esteem certain actions—for example, the expression of contempt toward minorities—we can discourage their continuation. Brennan and Pettit, 2004.
inappropriate to imprison someone who refuses to renounce her own views on how people should be valued. With regard to highly resistant beliefs about SES, the role of education becomes critical. Even if one generation’s (inappropriate) conception of SES cannot be altered, steps can be taken to ensure that that conception is not passed on and the appropriate social and economic grounds for valuing persons is taught to the next generation.\(^{266}\)

The second mechanism for bringing existing SES gradients into closer alignment with our chosen conception of SES is not to change what people value but to change how it is distributed. For example, on the radical egalitarian conception of SES, given that there should be no difference between the highest and lowest members of the SES gradient, we should redistribute almost all the relevant determinants of SES. If Gucci watches cement high SES in our world then everyone gets a Gucci watch; if differences in income affect the value we accord one another, then we should redistribute income; and similarly with education. If everyone has the thing of value—what makes them warrant higher SES—then these things will, over time, cease to be grounds on which we different between one another since they are possessed by all. For those determinants of SES that cannot be so obviously redistributed—for example, circumstances of birth, personal traits like beauty, or scare resources like diamonds—we would presumably rely on the first type of mechanism, changing what people value, in order to alter SES gradients.

Regardless of which conception of SES we adopt, once we have arranged a system so as to properly reflect the grounds on which we should value one another’s social and

economic traits, we will be left with a SES gradient of a certain degree. Depending on the particular characteristics identified as being appropriate determinants of higher SES and how widely shared these features are, the gradient will be more or less steep. In some instances, like the hereditary system, there will be a marked difference in SES between the highest and lowest groups, between the king and the serfs. In other cases, like the radical egalitarian system, there will be little, if any, SES differential between individuals. Given the relationship between SES—whether mediated through control or otherwise—each conception of SES will also condone a certain degree of health disparities, directly correlated to the steepness of the gradient it permits.

Importantly, because SES is only one aspect of an overall theory of status, a conception of SES will not completely answer the question of which health disparities are morally acceptable. It will have to be coupled with, for example, the appropriate conception of what we owe persons as rational agents (or merely as persons, outside of their social or economic features). In the absence of a consensus on what follows morally from the fact of personhood, however, a conception of SES will still prove useful in assessing the moral status of health disparities. As a first pass, with potential modifications to follow based on our final vision of overall status, differences in health in our world would be acceptable only if they would still exist in a world in which SES is distributed appropriately. In this way, we can begin to move from a conception of SES to a judgment about which health disparities are acceptable.

267 This raises a separate question, also of concern for various conceptions of justice, as to how we ought to arrange policies and institutions so as to best ensure the stability of the status (or, in the case of justice, the distributive institutions) of groups and individuals.

268 This will not be true for theories that build their conception of SES out of a conception of rational agency (potentially like the radical egalitarian). In this case, the whole system will hang together so as to spell out how we should treat one another both in terms of our status as persons (or agents, etc.) and as persons with such-and-such social and economic features.
To summarize, while a conception of justice tells us how to distribute basic rights and liberties and how to apportion the advantages of social cooperation, a conception of SES identifies the value we should have based on our social and economic characteristics and the appropriate response to that value. Given a world in which some institutions are unjust, a conception of justice should also tell us how to remedy this situation and how to ensure the stability of the just institutions that emerge. Similarly, given a world in which individuals are valued on inappropriate grounds, a conception of SES should identify how to change this situation. I have suggested two possible mechanisms: altering peoples’ current beliefs about what characteristics give a person higher or lower SES or redistributing the (inappropriate) determinants of value such that everyone has access to them thus ceasing to be grounds on which we could distinguish between persons. When these goals have been accomplished, the result will be a SES hierarchy of a certain steepness, depending on how widely the determinants of SES are distributed. When coupled with an adequate theory of overall status, whatever health disparities this distribution produces are morally acceptable because they arise from morally acceptable ways of treating one another.

*Why Socioeconomic Status?*

I now want to highlight four reasons why we might prefer to approach health disparities through a theory of SES. First, as I suggested at the end of the last chapter, Daniels and Ruger start with the moral standing of health disparities and work their way backward to the moral standing of the determinants of health. In contrast, starting with a theory of SES, we do not have to worry about whether a particular socioeconomic determinant of health will fall under the domain of justice since we will be focusing directly on these factors. As additional epidemiological literature on the broader determinants of
health becomes available, particular conceptions of SES will be able to directly address how these determinants should be distributed without first considering their relationship to just institutions.

Second, starting from a theory of SES has the advantage of side-stepping the problem of health-care resources sinks. Recall that, given Ruger’s focus on health as a valuable functioning that should be equally part of our capabilities set, she has to consider how many resources we are required to devote to ensuring a certain level of health. In contrast, a theory of SES treats health disparities as an emergent phenomenon of whatever socioeconomic gradient is produced from the conception of SES we endorse. The focus is on the slope of certain gradients rather than on the use health has for a particular individual in determining what she is capable of becoming. Once the determinants of SES are distributed appropriately, it is justified for someone’s worse health to limit her opportunities or capabilities set and we are not required to spend excessive resources to bring her up to a certain minimum level of healthy functioning.

A third advantage of this approach is that it allows us to draw on a wider array of mechanisms to address health disparities than are available within a theory of justice. Recall that the concept of justice, as Rawls understands it, is focused on the appropriate arrangement of institutions. Daniels’ expansion of the Rawlsian framework is concerned with how to alter social institutions (public and private) so as to reduce health disparities, but leaves out the role that individuals may have—indeed independently of their role as agents of institutional change—in this regard. Daniels mechanisms for addressing the broader determinants of health are all built around institutional revision, for example, in taxation and redistribution of income, the education system, public health provision, and health care access. In contrast, a
theory of SES is not wedded to the idea that change in socioeconomic gradients must be conducted through institutional channels. For example, a radical egalitarian conception of SES might hold that, while institutions should be arranged to redistribute certain resources—for example, Prada sunglasses—so as to reduce SES gradients, it is incumbent on individuals to stop disvaluing one another based on certain attributes—for example, race. In the latter case, the responsibility is on citizens not to change their institutions but to change the way that they treat one another.

More generally, while a particular conception of SES will tell us which grounds are appropriate for valuing one another, it is not committed to any specific mechanism for encouraging the appropriate realization of that value. In fact, institutions do not have to play a role at all; it may simply be a moral obligation for people to give part of their income to others or to work to ensure equality in educational opportunities as part of an overall duty to reduce SES gradients. Putting aside the question of whether this reduction could actually be achieved without institutional arrangements to ensure that citizens fulfill their moral duties, it is still possible that a conception of SES would endorse entirely non-institutional mechanisms of change. This flexibility has a number of promising features when it comes to thinking about health disparities.

For example, the focus on non-institutional mechanisms for addressing health disparities emphasizes the role that individuals have in creating—and reducing—the SES gradients that generate differences in health. There is a tendency to think about a problem as large and complex as, for example, the relative cancer mortality rates of blacks and whites in the United States as being solvable only through institutional change. While this may be partially correct, such a view encourages citizens to distance themselves from the problem,
leaving it up to elected officials to create solutions. And, when the political process fails to move forward or gets the institutional structure wrong, this can leave us with the impression that the problem is intractable and that we have minimal responsibilities to enact change ourselves. If the government cannot adequately address health disparities, the thought goes, what chance do I have as just one person? In contrast, a conception of SES that emphasizes the role interpersonal attitudes have in creating and reinforcing health disparities can make it clear that change must be mediated not merely through institutional rearrangements but by individual citizens in their daily lives. One should not, for example, treat others as inferior because of their race or ethnicity, regardless of the current shape of our institutions. Ultimately, it is the way that individuals and groups value one another that creates differences in SES and health. We do not necessarily have to appeal to political or institutional change—although this certainly may help—to alter such methods of valuing or their outcomes.

The flexibility that the SES-based approach gives us reinforces the idea that merely changing access to health care is insufficient to adequately address health disparities. As I mentioned in the first chapter, starting with the idea that certain disparities are unjust naturally leads one to the conclusion that the appropriate corrective measure is the redistribution of resources, since this is the focus of justice. If we begin with the idea that health disparities arise from differences in the ways in which we value one another (i.e. differences in SES), it is much more difficult to assume that the solution lies solely in changing access to health care. And this is consistent with the epidemiological data on the broader determinants of health. Insofar as starting from a conception of SES reminds us that both institutional and non-institutional arrangements underlie health disparities, it makes it
more likely we will adequately address all of the relevant determinants of health without simply defaulting to changing access to care.

The fourth advantage of starting with a theory of SES is that doing so opens up a wider moral territory than merely viewing health disparities through the lens of justice. While justice is an important virtue (and the first virtue of social institutions), it does not exhaust the domain of morality. As should be obvious from the discussion thus far, the SES that we have and the SES we seek play as significant a role in our daily lives as the relative fairness of the institutions in which we find ourselves. I have suggested that the most appropriate way to understand and evaluate this phenomenon is through a theory of SES. To lend further support to this claim, I will, in the next chapter, present a more detailed a conception of SES—centered on the idea of a minimal threshold of status—and show how it might be used to generate specific recommendations about health disparities in the United States.
Chapter VIII: Minimal Thresholds and Progress on Health Disparities

Six weeks into my first year of medical school, I was sent to shadow a family practice physician, Dr. P, in a small town in rural North Carolina. Although he was a good diagnostician and had an impressive range of clinical knowledge, Dr. P was most memorable for being a model of how not to practice medicine. His goal, he said when we first met, was to introduce me to an elite occupation—a “priesthood,” he called it—that was responsible for guiding the lives of the less elevated members of society. Although affirmative action and the introduction of women to the profession had weakened its status, it still provided a place for people like us—white men—to occupy the position we deserved.

Although Dr. P presented his views about medicine in the privacy of his office, it was clear that his opinions spilled over into patient care. He refused to treat Hispanics, arguing that if they couldn’t be bothered to learn to speak English, he couldn’t be bothered to see them. He referred to the only black patient who came to the clinic as a “drug-seeker” and treated four white patients, all of whom had later appointment times, before him. Dr. P’s views on Middle Easterners were largely parroted from Rush Limbaugh, whose voice drifted from the break-room radio tuned to his program. Dr. P’s attitude toward his female patients and nurses was patronizing and dismissive. At one point he rolled his eyes at me while a patient was recounting her symptoms; he patted her on the hand before walking out halfway through one of her questions. He also made disparaging comments about gays and patients he considered to be “fairies.” After two days I switched to another practice.
Dr. P’s actions inside and outside of his clinic probably contributed to health disparities in his community. All of these were cases in which people were treated inappropriately and, as a likely consequence, their health suffered. They were not, however, victims of unjust basic institutions or policies, at least not directly. Instead, they were treated with something less than the full consideration they deserved. They were ignored, made to wait, mocked, berated, and refused to be seen. What explains the wrongness of my Dr. P’s behavior? At bottom, his actions represent a failure to accord the appropriate status to individuals. Blacks are not all drug-seekers who should wait in examination rooms until they learn not to bother the doctor; women are not all bored housewives who have no real problems and just need someone to talk with; the fact that someone does not speak English is hardly grounds on which he should be denied health care. Each of these actions—and the attitudes that frame them—represents a mistaken view about how we should treat one another.

In the previous chapter, I suggested that, because differences in socioeconomic status (SES) tend to drive health disparities, we should be asking questions about the value assumptions that underlie these gradients. SES hierarchies are built around conceptions of which social and economic characteristics are appropriate grounds for valuing individuals and how we should respond to those features. If we want to know whether, for example, the health disparities that Dr. P created because of his treatment of blacks, Hispanics, women, and gays were justified, we have to ask whether he treated them with the appropriate status. We can pose a similar question regarding health disparities formed on a larger scale like those between blacks and whites in North Carolina or rich and poor families across the United States.
As I have argued, there are a number of conceptions of SES that could answer this question. Here, I want to focus on a version that supports a minimal level of SES below which individuals should not fall. I will provide a rough sketch of the theoretical underpinnings of this approach and make an initial suggestion as to how we might set that minimal level. I will then provide a few specific applications of the conception to contemporary American health disparities, focusing especially on policy recommendations that I believe are not as easily generated from within a theory of justice. Like Daniels and Ruger, my overall aim is to use a more abstract moral theory to generate concrete solutions to health disparities.

Starting Points

The kind of moral theory I am interested in exploring is one that sets a minimal level of SES. It provides a threshold of status that establishes what counts as minimally acceptable behavior between individuals. Such a theory takes certain characteristics of persons—race and gender, for example—to be an inappropriate basis on which to assign low status. By restricting the grounds on which we should evaluate one another, this approach guarantees a minimal level of status; no one will be treated in the harmful way they would be if race or gender were appropriate reasons for disvaluing them. When my preceptor assumed that all blacks are drug seekers, he was treating his black patient as having a kind of status (in virtue of being black) that is below this minimal level, or so I will argue.

Before I endorse any specific proposal for where to set a minimal level of SES, I want to discuss the relationship between this general approach to SES and the reduction of health disparities. Figure 1 illustrates one possible effect of establishing a minimal level of SES on health disparities. World A represents conditions before the introduction of a minimal level
of socioeconomic status (SES). SES and health are correlated in a way familiar from the discussion in Chapter 3 such that each individual has, on average and other things being equal, better health than the people below him in SES and worse health than those above. World B represents conditions where behaviors, attitudes, institutions, etc. have been changed to reflect a minimal level of SES. If we assume that the status of the highest group stays constant between the two worlds, the difference in health between the top and the bottom in World B is not as large as it is in World A \((x_2 < x_1)\). Thus, while the steepness in the health/status gradient is the same in both worlds \((y_1/x_1 = y_2/x_2)\), World B nevertheless represents a reduction in health disparities compared to World A.

**Effect of Introducing a Minimal Level of Socioeconomic Status (SES)**

![Diagram](image)

**Figure 1:** Health disparities \((y_1 \text{ vs } y_2)\) decrease in World B compared to World A if the status of the highest individuals stays constant between the two.

Importantly, the minimal threshold approach to SES cannot *guarantee* a reduction in current health disparities. If the implementation of a minimal level of SES results in an increased drive for better status amongst those who already have high status, health disparities will remain constant or even worsen (see Figure 2). Again Worlds A and B are
pre- and post-implementation of the minimal level of SES. If we assume that the status of the highest group increases between the two worlds, the difference in health between the top and the bottom in World B will be the same as World A ($x_2 = x_1$) or worse ($x_2 > x_1$) depending on how much the status of the highest group increases.

**Effect Two of Introducing a Minimal Level of Socioeconomic Status (SES)**

![Diagram](image)

**Figure 2**: Health disparities ($y_1$ vs $y_2$) are the same in World B compared to World A if the status of the highest individuals increases in World B.

I do not think that we can guarantee in advance that the scenario in Figure 1 will occur compared to Figure 2. If it turns out that health disparities are worse after we implement a minimal level of SES, this may be grounds on which to prefer a different conception of SES or to combine the minimal threshold approach with a more specific position on the maximal allowable difference in status from top to bottom. Since my overall goal in this dissertation is to argue that we should explore moral territory outside of a theory of justice in thinking about health disparities, I do not think it would be devastating if it turned out that one specific approach proved inadequate. As I lay out my particular method
for setting a minimal level of SES, however, I will argue that it should constrain health disparities in a way akin to the first scenario. Before I turn to that task, I want to make two additional observations.

First, because SES and health are closely correlated, it may be tempting to think that establishing a minimal level of SES also helps to set a minimal level of health for those with the lowest SES. But SES and health are not completely correlated; we all get sick and die but we do not all drop in status. As such, this approach should not be interpreted as a guaranteeing a level of health akin to a right to health. In addition to reintroducing the problem of what to do with chronically unhealthy persons who are ‘resource sinks,’ this way of conceptualizing the minimal threshold approach misses that the focus is on health as an emergent phenomenon of SES gradients, not a primary center of moral attention. That implementing a minimal level of SES might reduce the difference in life expectancy between the top and bottom of the SES gradient because of improvements in the age-adjusted life expectancy of the worst off does not mean that any specific person of low SES should be guaranteed a certain length of life.

Second, when explaining why an individual or group should not be treated in a certain way or why certain attitudes should not be adopted toward them, the minimal threshold approach should not refer back to facts about personhood or rationality. If it is wrong to treat female patients with contempt because they are rational agents then the theory would be concerned with rational status, not socioeconomic status. What is required for a conception of SES is that it provides a threshold based on social feature of persons.269 For

---

269 It is possible that a minimal threshold might be developed around an economic feature of persons—for example, the fact that we all are capable of participating in a market might set certain limits on what we can do to one another—but I find it much more plausible that such a threshold would spring from our social characteristics and so will not pursue the economic possibilities.
example, we might hold that SES should be based on circumstances of birth but that no one should have the status that peasants or the Untouchables had within the feudal or caste systems. Such a conception would support a relatively steep socioeconomic gradient based on one’s lineage, but would have a bottom threshold below which individuals of ‘common’ birth should not fall. The people with the lowest SES should still have a higher status than, for example, the Untouchables, such that it would not be appropriate to require them to sweep up their footprints or deny them access to education. In general, regardless of the specific content, a minimal threshold must be set in virtue of or in accordance with a specific social characteristic of individuals.

Discourse Ethics

Although there are several possible candidates for a social feature of individuals that could delineate a minimally acceptable set of behaviors and attitudes, here I want to focus on the idea that we are all participants in communicative discourse. In exploring how facts about our involvement in discourse can help establish a minimal level of SES, I will draw on the work of the philosopher Jurgen Habermas, from whom the epigraph of this dissertation is taken. Of necessity, I will take many of his starting points for granted, without offering significant justification for them. While I believe that his approach to moral theory is highly defensible, my goal here is to take the later step and show how a reasonably developed version of it might be applied to a practical problem like health disparities.

270 A different starting point might be to focus on T.M. Scanlon’s claim that we are all involved in justifying ourselves to others on grounds that they cannot reasonably reject. I have chosen to focus on communicative discourse because it is unclear whether this activity is true of us simply in virtue of our being reason-responsive agents. If this is right, Scanlon’s starting point would be better associated with a theory of rational rather than socioeconomic status. See Scanlon, 1998.
In contrast to the anti-rationalist, anti-universalist Continental philosophers (Rorty, Gadamer) with whom he is often associated, Habermas is actually better understood as developing a universalist moral theory that has its roots in the Kantian tradition. In this regard, he has a number of commitments in common both with Kant and contemporary Kantian constructivists like John Rawls, Onora O’Neill, and Thomas Hill. Here I want to focus on three specific areas of overlap with these philosophers as a way of explicating Habermas’ starting assumptions, methodology, and substantive conclusions.271

Starting Assumptions

Habermas is primarily interested in defending a picture of morality that focuses on our obligations—what is required and prohibited for us to do as members of a community of interacting agents—as opposed to questions of character or virtue. In this regard, he is best understood as developing a theory of the right rather than, as with classic utilitarians, a theory of the good that generates a secondary account of the right. Like Kant, Habermas treats moral obligations as having a categorical force such that they apply to us whether or not they work toward our self-interest or desire fulfillment; there are some actions we ought to perform even if the thought of doing so currently ‘leaves us cold.’ In the same vein, Habermas rejects familiar anti-universalist claims that there is a pluralism of mutually incompatible value orientations that cannot be rationally adjudicated. Part of his goal is to identify high-level, abstract principles that both explain the categorical status of morality

271 I will be focusing on Habermas’ work in his essay, “Discourse ethics: Notes on a program of philosophical justification,” but I also draw on more recent writing to identify substantial changes in his thinking, where relevant. Habermas, 1999, pp. 43-109.
(recall the Kantian question, how is a categorical imperative possible?) and can be used to adjudicate inter-cultural debates about our specific obligations.272

Methodology

Habermas constructs his moral theory through a series of arguments against an imagined opponent, who variously endorses ethical non-cognitivism, subjectivism, and cultural relativism. Habermas begins by drawing attention to the moral phenomena of resentment, an idea that he borrows from Strawson.273 When we feel indignation or resentment at another person who has wronged us, we are not just responding to our specific situation but to the more general fact that that individual has ignored a norm meant to be valid for all competent agents. What makes resentment a distinctly moral emotion is its connection with the idea that an impersonal behavioral norm has been violated. For Strawson and Habermas, when we consider whether resentment or reproach is appropriate in a given situation, we are implicitly asking whether a particular interpersonal norm is, in fact, valid. Habermas writes, “When employing normative utterances in everyday life, we raise claims to validity that we are prepared to defend against criticism.”274 These utterances constitute communicative discourse.

Communication, for Habermas, includes any interaction (primarily, although not limited to, speech acts) in which “one actor seeks rationally to motivate another.”275

272 Unsurprisingly, Habermas also identifies himself as a cognitivist about moral sentences, arguing that their truth can be determined by reviewing the criteria for what would count as a good reason in favor of believing those sentences. Habermas, 1999, p. 56.


274 Habermas, 1999, p. 56.

275 Habermas, 1999, p. 58. Elsewhere, Habermas endorses a stronger claim; namely, that the meaning of speech acts can only be understood with relation to reason-giving. He writes, “We understand a speech act when we know the kinds of reasons that a speaker could provide in order to convince a hearer that he is entitled in the
communicative discourse that occurs over moral norms involves participants introducing reasons for accepting or rejecting particular norms. This project is similar to what we do in science when we ask which one of a number of propositions is true. Although Habermas believes that the evidential standards are different when considering whether descriptive sentences are true or false versus whether a moral norm is valid or invalid, he argues that both factual and moral inquiry have at least one commonality. Just as arguments in science involve certain formal rules—for example, principles of induction to move from particular observations to a general hypothesis—for assessing the truth of a propositional sentence, arguments in morality require formal rules if we are to identify which norms are valid. Habermas tentatively offers the following principle for determining whether a norm that governs behavior, attitudes, or institutional structures is valid:

(U) “All affected can accept the consequences and the side effects [the norm’s] general observance can be anticipated to have for the satisfaction of everyone’s interests (and these consequences are preferred to those of known alternative possibilities for regulation) [emphasis his].”276

As Habermas understands it, (U) places a number of constraints on communicative discourse. For example, he argues that fulfilling the conditions of (U) entails that discourse can only produce valid norms when: 1) no one capable of making a relevant contribution has been excluded; 2) participants have equal voice; 3) they are internally free to speak their honest opinion without deception or self-deception; 4) there are no sources of coercion built into the process and procedures of discourse.277 In his earlier work, Habermas talked about (U) as setting out an ideal speech situation that real discourse must meet in order to generate given circumstances to claim validity for his utterance.” Habermas, 1998, 232. I do not think that this stronger linguistic thesis is necessary to support Habermas’ position on the communication that occurs in discourse over moral norms.

276 Habermas, 1999, p. 65.

277 Habermas, 2005, p. 89.
valid norms. More recently, however, he has argued that the requirements (U) imposes are better understood as checks or reminders on actual discourse to ensure that there are no obvious violations of 1-4, granting that it may be impossible to guarantee that these conditions are met.

After suggesting a possible candidate for assessing the validity of the moral norms under debate in communicative discourse, Habermas considers the skeptical argument that (U) merely represents a “hasty generalization of moral intuitions peculiar to our own Western culture.” A critic voicing this complaint, Habermas suggests, is casting doubt on whether there can be any principle for adjudicating between competing moral norms that is not, itself, just another moral norm in need of validation. There are not, according to this position, any standards for regulating discourse that are not themselves up for debate. Habermas’ response to this challenge is to show that a skeptic expressing such an argument is implicitly committed to (U). He writes, “Every argumentation, regardless of the context in which it occurs, rests on pragmatic presuppositions from whose propositional content the principle of universalism (U) can be derived.” In other words, (U) does not just delineate the minimal requirements for validating a moral norm, but is also an underlying presupposition in any communicative argument.

This approach—providing a transcendental argument that aims to show that, by arguing against a particular claim, the skeptic is already committed to that claim—has a long pedigree tracing back at least to Descartes’ cogito. The useful historical comparison in ethics


279 Habermas, 2005, p. 91.

280 Habermas, 1999, p. 76.

281 Habermas, 1999, p. 82.
is with Kant, who thought that, given the connection between rationality and morality, anyone who could ask the question ‘Why should I be moral?’ already had sufficient reason to be moral. In contrast to Kant, whose argument focused on the commitments that spring from having a rational will (a property he thought the skeptic and the defender of morality had in common), Habermas is interested in the commitments we take on when we communicate with one another (a social feature of persons). The skeptic who argues that (U) is simply another norm in need of validation is also involved in communicative discourse. He is, Habermas argues, committed to providing reasons for his position. But the process of giving and receiving reasons is itself a rule-governed activity. If it turns out that one of these rules is (U), then, the skeptic, like the defender of morality, is already committed to the principle of universalism.

Thus Habermas’ transcendental approach aims to identify the presuppositions of communicative discourse—arguments in which reasons are exchange—and to show that (U) is among these presuppositions. He suggests that, at the most basic level, argumentation is only intelligible as reasoned discourse if participants follow a ‘minimal logic’; for example, they do not contradict themselves, they do not use the same expressions with different meaning, etc. Speakers who violate these rules—which do not have yet have recognizably ethical content—are not, for Habermas, engaging in argument. They cannot be understood to be offering reasons for their position if they are, for example, willing to contradict themselves. Similarly, participants in discourse who exclude or coerce or who misrepresent their opinions, either directly or through self-deception, in order to convince others to adopt their position are not actually engaged in reasoned discourse. Torturing your critics is one way to get them to agree to your position, but you have not won them over through reasoned
argument. Once you have decided to stop offering reasons and start shooting, you are no longer involved in argument, and, once you accept argument as a way to settle a question like whether (U) is an appropriate test for validating moral norms, violent, exclusionary tactics are off the table. You are committed to reasoning with everyone who has a stake in these norms and this is exactly what (U) demands. Thus Habermas suggests that the imagined critic who argues against (U) as delineating a procedure for validating norms is, at the same time, committed to (U). He writes, “The principle of universalization, which acts as a rule of argumentation, is implied by the presuppositions of argumentation in general.”

To summarize Habermas’ argument: moral debate occurs within the realm of communicative discourse. When we argue about what we ought to do, we are asking whether there are reasons for supporting one moral norm or another. This is a question about which norms are valid. Habermas tentatively suggests one criterion for assessing the validity of proposed norms: can everyone affected by the norm accept the consequences and the side effects that its general observance can be anticipated to have for the satisfaction of everyone’s interests. This is the principle of universalism (U). Habermas then argues that (U) is also a formal principle of argumentation in general such that anyone involved in communicative discourse over the status of (U) must implicitly accept the principle. Thus the moral skeptic, insofar as he provides reasons for his position, is also committed to (U).

There are a number of places to press this argument. Most importantly we might wonder whether Habermas’ imagined critic is committed to communicative discourse merely by voicing his concerns about the validity of (U). Is there a meta-discourse in which the skeptic could express his argument without thereby taking on the requirements of standard communicative discourse? We might also wonder how Habermas would respond to the

---

282 Habermas, 1999, p. 86.
person who refuses to enter discourse at all. Although he has tentative replies to these problems, this debate is of less interest here. Instead, I want to turn to Habermas’ final project, which is to show how discourse ethics can generate practical content.

*Substantive Conclusions*

Habermas uses principle (U) as the foundation for a general procedure for validating proposed, substantive moral norms, a variation on an approach familiar from contemporary Kantian constructivists like Rawls. For Habermas, like the constructivists, the validation of subsidiary moral principles must come from a process rather than from mind-independent facts, as for some moral Realists, or facts about pleasure/pain, as in classical utilitarianism. That a certain moral principle emerges from or would be validated by a constructivist procedure serves as (partial) justification for that principle. In Habermas’ case, this procedure is communicative discourse taking place under the constraints of (U).

The analogy to Rawls should be clear at this point. Rawls starts with a conception of persons as mutually disinterested with a capacity for a sense of justice, a capacity for a conception of the good, and the goal of maximizing their share of primary goods and then asks what principles of justice they would accept given that they do not know what place they will occupy in the resulting society. He argues that they would agree to a society whose basic institutions are structured by the liberty and difference principles. Habermas starts with a conception of persons as communicate agents and argues that that commits them to a

---

283 Neo-Kantians like Christine Korsgaard have a similar worry in the figure of the wanton, who refuses to regulate his will in accordance with any constant practical identity. Korsgaard, 1996, p. 142-4.


process in which each individual has an equal, non-coerced, non-self-deceived voice in considering which moral norms/principles are valid.

Unlike Rawls, however, Habermas does not speculate as to what specific principles agents situated in this context would endorse. This is because he is not interested in which rules a hypothetical group of agents under idealizing conditions would or could agree to; rather, he is concerned with what principles actual persons agree to once they are operating under a reasonable approximation of (U). Thus Habermas rejects a ‘top down’ approach to identifying and applying the most abstract principles whereby we generate secondary rules based on increasing amounts of empirical information.286 Instead, moral principles (norms) are validated when they are proposed and accepted by participants in the actual conversations that take place under the constraints of (U).

It would be a mistake, however, to think that Habermas has identified a completely formal constraint on our behavior or attitudes, a lá certain criticisms of Kant’s Formula of Universal Law. Like that much-maligned principle, (U) has substantive implications for how agents should treat one another. Granting that any practical application of discourse ethics will be somewhat speculative, what kinds of policies/principles might agents engaged in this enterprise endorse?287 I think that the most obvious candidates are practices that would work to sustain the communicative discourse itself as well as reinforce the idea that we are all equal participants in this activity. Such policies should attempt to ensure that actual discourse approximates the conditions of (U) as closely as possible.

---

286 He writes, “Practical discourse is not a procedure for generating justified norms but a procedure for testing the validity of norms that are being proposed and hypothetically considered for adoption.” Habermas, 1999, p. 103.

287 It may, in fact, be inappropriate for any single agent (like this author) to try to determine in advance what the outcome of (U)-constrained discourse might be, given Habermas’ emphasis on the essential role of dialogue. A better way to think about what follows might be to consider it a proposal open to further modification by participants in discourse ethics.
This requires a commitment to non-exclusion such that discussions of moral norms, whether regulating individual behavior or broader questions about the arrangements of institutions, involve, as much as possible, all those people who may be affected by the outcome. Participants in these conversations should be given equal voice where that may require both respectful attitudes from others and the self-respect necessary for believing that one is entitled to express one’s point of view. The discourse should have mechanisms designed to prevent coercion and deception, where that may involve limiting the amount of power any given individual can hold over the outcome of the conversation. These are only a few examples; a full specification of these sorts of policies is beyond the scope of this dissertation, but it seems relatively straightforward that individuals committed to communicative discourse—a description that, Habermas argues, even characterizes someone who doubts he is involved in that form of communication—would endorse practices that would maintain discourse in a version as close as possible to the ideas expressed in (U).

Supposing that these sorts of recommendations would be validated within the framework (U) establishes, we can explain the wrongness of Dr. P’s attitudes and behaviors. In treating his patients—and, potentially, his wider community—the way that he did, he was essentially denying them the ability to participate as equal agents in a communicative discourse. In some cases, he simply excluded them from conversation altogether, as with the Hispanics he refused to treat. In other cases, his attitudes—contempt for the women he saw; distrust of blacks; denigration of gays—worked to ensure that these patients lacked an equal voice in their care. The only people Dr. P counted as having something worth considering

288 As with many of the ideas I introduce here, we will need a further discussion of what counts as exclusion and inclusion in conversations about the norms that affect multiple individuals. In some cases, a political process like direct or representative democracy will fulfill the obligations of inclusion; in other cases, like the doctor-patient relationship, inclusion may require that each party be more substantively involved.
in communicative discourse was other educated, white men, a fact that was repeatedly illustrated by the way he actively engaged with this sort of patient. Other patients simply did not merit this status, a view that, I believe, is unjustified under the demands of discourse ethics.

Let us now turn to the broader question of how Habermas’ work on communicative action might set a minimal level of SES. Recall that a conception of socioeconomic status is charged with identifying those social and economic features of persons that are appropriate grounds for treating them with higher or lower status. A minimal threshold approach will pick out a characteristic of persons that establishes the most basic set of behaviors, attitudes, and responses we should have toward one another. For Habermas, this social feature is our ability to engage in communicative discourse. According to a conception of SES built around discourse ethics, no one should be treated in ways that would violate his or her status as participants in communicative discourse. A useful way of understanding what this threshold of status establishes is to consider the kinds of behaviors that would be permissible toward us if we were not creatures capable of participating in communicative discourse. It would be appropriate to exclude us from decisions that affect our lives; to coerce us into following policies that we had no role in developing; or to suppress our activities. As the kinds of creatures able to engage in communicative discourse, however, (U) rules out activities, behaviors, or attitudes that would treat us in these ways. For example, excluding a person because of her skin color, gender, or sexual orientation from participating in political processes that impact her life is a violation of (U). Similarly, treating an individual with long-standing contempt—whether because of his income level, place of birth, ethnicity,
etc.—insofar as this would be destructive to his or her self-respect and ability to present her opinions, is also impermissible.

It is a further question—left up to the participants in actual discourse—as to whether, once we ensure that everyone has been given the status (U) guarantees, it is permissible for us to treat one another with higher esteem based on characteristics like education or earned income. As long as everyone is granted a minimal level of consideration, there is no reason, on Habermas’ account, why it would be inappropriate to allow some people to accumulate more education or wealth than others or to treat the wise, the rich, or even those of noble birth with esteem, so long as this did not result in decreasing the status of individuals without these characteristics below the level (U) guarantees. If it turned out that unequal distribution of resources like education or wealth always translates into the marginalization of some groups from communicative discourse or always leads to one party adopting coercive policies, then (U) would limit the development of disparities in education and wealth.

With this point in mind, if we draw on discourse ethics to set a minimal level of SES, should we expect an outcome like the scenario in Figure 1 where health disparities decrease or the scenario in Figure 2 where they stay the same or worsen? I think that Habermas’ emphasis on equalizing the ability of individuals to participate in communicative discourse would work to constrain or stabilize the status-seeking efforts of the highest status individuals. The minimal level of SES will always help ensure that the lowest members of society are given an equal voice with those at the top, especially where moral norms—ranging from the shape of basic institutions to the attitudes that are appropriate when individuals of different social strata confront one another—that affect both groups are concerned. It is unlikely, given the strong notion of equality between participants in
communicative discourse, that the conclusion of these conversations will function in such a way as to drive up the status of the highest groups. There is, of course, no *a priori* argument that the scenario in Figure 2 will not occur; if health disparities worsen under a discourse-ethics derived minimal level of SES, this may be a reason to prefer a different approach.

So far I have talked about relatively abstract constraints on the shape of SES gradients. I now want to speculate as to what secondary policies individuals committed to maintaining communicative discourse might endorse. If we were to structure our institutions and behaviors in such a way as to reflect the minimal level of SES discourse ethics establishes, what sort of changes might we expect to the current state of affairs? And to what extent would health disparities be reduced as a result? In answering these questions, I want to focus on policy change in the current American health-care system as well as the potential for broader social transformation.

*Changes within the Health Care System*

It may seem odd, given this dissertation’s central emphasis on the disconnect between expanding health care and reducing health disparities, to start a discussion of policy change by focusing on medical care. While I continue to endorse the idea that addressing the broader determinants of health is a more useful way to decrease health inequalities than altering how and to whom medical care is delivered, because theories of justice have focused so much on this area, it will provide a useful contrast for applications of the minimal threshold approach. To a certain extent, the two types of theories will share similar recommendations. Thus, rather than focusing on, for example, questions about universal access to health care, I want to identify two areas—cultural competency training for physicians and the use of patient navigators in clinics—where changes could be made to
potentially effect health disparities. I will argue that these policies are more easily justified within the minimal threshold approach, centered on the commitments of discourse ethics, compared to theories of justice.

In Chapter 3 I briefly noted that the Institute of Medicine (IOM), in a report titled *Unequal Treatment: Confronting Racial and Ethnic Disparities in Healthcare*, suggested that discrimination (conscious or unconscious) or bias on the part of health care providers causes a portion of racial/ethnic health disparities. One policy-based response to the IOM report was to emphasize cultural competency training for physicians. The idea of cultural competency in medicine is to help health care providers develop a set of “academic, interpersonal, and clinical skills” to increase their understanding of cultural differences between racial and ethnic groups.289 Such training is not, however, limited to improving cultural awareness or sensitivity but involves being able to bridge cross-cultural barriers to adequate delivery of medical care. As Brach and Fraser note, cultural competency includes:

“A set of congruent behaviors, attitudes, and policies that come together in a system, agency, or amongst professionals and enables that system, agency or those professionals to work effectively in cross-cultural situations [emphasis added].”290

Hospitals and medical schools have implemented a wide variety of programs as part of their commitment to culturally competent health care provision. Examples include: interpreter services, as spelled out in the Office of Minority Health’s Culturally and Linguistically Appropriate Service Standards; cultural competency tests and training as a requirement for physician licensure, which is mandated in California and Washington; and efforts to recruit and retain minority health care staff to create a hospital and clinic

---

289 Williams, 2007, p. 55.

290 Brach and Fraser, 2000, p. 182
environment that is more welcoming and appreciative of the needs of minority communities. These policies aim to alter discriminatory behaviors and attitudes that may function in the medical environment to create health disparities.

Data on the effects of cultural competency programs, particularly the commitment to bridging language barriers through interpreter services, have demonstrated some reduction in health disparities. A 2007 review of 28 studies investigating the impact of professional interpreters on the care patients with limited English proficiency (LES)—a group with significant health disparities compared to native speakers—concluded that the availability of such services increased utilization of medical care (both outpatient and emergency) among LES patients. A 1999 study in Australia showed that the availability of a professional interpreter eliminated differences in Cesarean section rates for LES (primarily Southeast Asian) versus native-speaking patients. A 1998 study in America showed a similar reduction in differences in hemoglobin A1C, lipid, and creatinine values between LES and native speakers with type II diabetes. Little data exist, however, on the impact of other aspects of cultural competency—for example, physician training—on health disparities, although there is a growing research interest in this area.

Coming at health disparities from a different direction, in 2004, the National Cancer Institute, as part of an effort to reduce the disparities in morbidity and mortality that occur after cancer diagnosis in minority groups, made 25 million dollars available to pilot Patient

---


Navigator Programs (PNP). Patient navigators are individuals from the community who are familiar with cancer diagnosis and treatment—either through their own experiences or that of families or friends—and have been hired and trained to help newly-diagnosed patients negotiate the treatment system. Within a PNP, individuals are matched with a navigator of a similar background. This process ideally includes variables such as race and ethnicity and community location.

The goal of navigation services is to reduce barriers to treatment including language differences, emotional worries, travel and distance concerns, difficulty communicating with care providers, financial support, and misinformation about medical options. Patient navigators might arrange for translation/interpretation services, help organize transportation to and childcare during appointments, coordinate services among different providers and ensure the availability of medical records, identify forms of financial support, and correct misinformation about cancer prognosis, diagnostic testing, and treatment interventions. Perhaps most importantly, the navigator offers minority patients a member of the care team with whom they may be more comfortable communicating and who can advocate on their behalf.

Since PNP are a relatively new strategy for reducing health disparities, there is limited data on their effectiveness. Preliminary studies, however, suggest some benefit to including patient navigators on the treatment team. Randomized control trials are also underway to examine the effect of navigators on health disparities in cancer treatment and

293 National Cancer Institute, 2004.

follow-up. More generally, the use of patient navigators has been shown to increase minority participation in screening programs for breast, prostate, and colon cancer, which may impact relative morbidity and mortality from these diseases.

Both cultural competency training (CCT) for physicians and patient navigator programs (PNP) represent attempts to equalize the communicative positions in medical discourse. CCT aims to help physicians approach their patients with a set of attitudes and behaviors that make it more likely that patients will participate in the medical process and comply with treatment recommendations. This training is designed to bridge the often substantial communicative gap between physicians and patients, who may come from very different cultural backgrounds. PNP, in contrast, work to make patients more familiar with the clinic or hospital experience, empowering them to be active participants in their care. By understanding what they can expect and by having someone on the health care team with whom they can more comfortably communicate, patients are likely to have better health outcomes.

I believe that both of these programs would be endorsed (and, therefore, validated) by participants in communicative discourse, under the commitments expressed in (U). But can a Rawlsian-style theory of justice make similar recommendations? If we assume that the absence of CCT or PNP leads to health disparities, which impacts equality of opportunity/capabilities, then there is an obvious argument for implementing these programs. The inequality of opportunity/capabilities that health disparities create is unjust; therefore, we should implement PNP and CCT to fix this state of affairs. As I have emphasized in earlier chapters, however, it is unclear whether theories of justice like Rawls’ can make this sort of


recommendation. A central goal of CCT and PNP is to change the attitudes physicians and patients have toward one another. CCT aims to foster an environment of respect for cultural difference and diversity among physicians, and PNP try to combat stereotypes and untrusting attitudes some patients may have about the medical profession. If theories of justice are in the business of adjusting the basic institutions of society rather than directly altering the attitudes of citizens living under those institutions, then it is unclear whether such a position can offer appropriate justification for CCT or PNP.

Changes in a Broader Policy Context

As with recommendations regarding health care provision, there is likely to be a degree of overlap between a minimal threshold approach to SES and theories of justices where broader policy questions are concerned. This is in part because conceptions of SES are also focused on the redistribution of goods (or equalizing access to them) insofar as the ownership of these things affects status. Theories of SES are also interested in basic institutional structures insofar as these institutions are vehicles through which SES gradients are constructed, preserved, and altered. Again, however, since the more useful contrast is with recommendations that cannot be as easily generated from a theory of justice, I will focus on one specific policy change—centered on the idea of a public sphere—that the minimal threshold approach to SES better supports.

Habermas discusses the public sphere as a concrete example of how actual discourse might best approximate the idealizing (or counterfactual criteria) (U) requires. He defines the public sphere as follows:

---

297 Actually, the chronology is a bit backward. Habermas’ concern in the earlier *Structural Transformation of the Public Sphere* was with the loss of the public sphere. He developed discourse ethics, in part, to justify the preservation of this space in contemporary democracies. See Habermas, 1991, p. xi-xiv.
“The sphere of private people come together as public . . . to engage [public authorities] in a debate over the general rules governing relations in the sphere of commodity exchange and social labor.”

The conversations in the public sphere range over areas of “common concern,” including the norms that regulate interpersonal behavior and attitudes in social and market transactions. Prior to the development of the public sphere in the 18th century, state or church authorities were the principle forces in the development, promulgation, and enforcement of these norms. One function of the public sphere, then, was to make such decisions dialogical—coming from discourse among many affected individuals—rather than monological—imposed unilaterally from a central authority. Thus, while Habermas spends a considerable amount of time discussing the scope of the public sphere and identifying its boundaries, more important for our purposes is how he characterizes the interactions between people in this environment.

First, the public sphere represents a context in which “far from presupposing the equality of status, [participants] disregarded status altogether. The tendency replaced the celebration of rank with a tact befitting equals.” As Habermas notes, this idea was not always realized in earnest, but it was at least instutionalized early on within the structure of the public sphere. For example, discussions in 18th century coffee houses—one historical example of a public sphere—often involved participants from classes—the sons of the nobility and the sons of shopkeepers—whose relative social standing would otherwise preclude interaction. Within the confines of the coffee house or salon “opinion became emancipated from the bonds of economic dependence.” And, we might add, social rank.

In this context, reasoned debate amongst participants settles questions that arise in discourse

298 Habermas, 1991, p. 27.
299 Habermas, 1991, p. 36.
and not economic coercion or appeal to an individual’s social standing as providing a trump
that ends argument. Second, the public sphere is a fundamentally inclusive space: “it could
never close itself off entirely . . . for it always understood and found itself immersed within a
more inclusive public of all private people.” The issues of concern to participants in the
public sphere are, by definition, those that affect citizens in their social and economic
interactions with one another; thus it should be possible that everyone so affected can
participate.301

The connection with (U) should be clear at this point: the public sphere represents an
area where the commitments that this principle requires, while not always met, have the best
chance of being approximated in the structure of actual discourse. The public sphere is a
communicative environment governed by the exchange of reasons; it is inclusionary;
participants are entitled to an equal voice; and coercion and deception are minimized
(secondary to the way that social and economic standing outside of the public sphere is de-
emphasized). As such, the public sphere is a ‘real world’ example of an environment that
could produce valid norms through communicative discourse. Given this possibility, part of
Habermas’ project is to explain why preserving the public sphere—in the face of threats from
late capitalism—is so important. Suppose that we accept that the public sphere is worth
protecting or reinstituting. How would this fit in with the broader idea of a minimal level of
SES and what would be the expected impact on health disparities?

301 Although, as Habermas acknowledges, it is one thing to be affected by a norm and another thing to be able to
engage in reasoned discourse over it. Given the literacy rates and subsidence living among the majority of the
‘public’ in the 18th century, it is unlikely that most individuals could have participated in the public sphere. As
conditions improved—a change, Habermas argues, that took place in part because of the efforts of the those
involved in the public sphere—it become possible to broaden the number of participants. There is no in
principle reason, then, why discourse could not include all those affected by a social or economic norm.
Unlike with the health policy examples, I have no specific epidemiological evidence that preserving the public sphere would reduce health disparities. As I have suggested, however, there is a theoretical reason to think that, insofar as having a public sphere constrains the development of SES gradients through its support for a minimal level of SES, it would work to reduce health disparities. It is difficult to imagine that participants in the public sphere would endorse stigmatizing or exclusionary norms toward HIV positive individuals (or cancer patients, for that matter), which would potentially affect their health-care seeking behavior. Furthermore, the inclusionary nature of the public sphere would likely work against the internalization of negative interpersonal attitudes and the way that that translates into SES and level of control over one’s life. All of this is, of course, speculative, although it is important that there is a theoretical link between the public sphere and the reduction of health disparities. Ultimately, I believe that the creation and preservation of the public sphere is an activity that participants in (U) would endorse, not because of its potential effect on health disparities, but because it would work to sustain the broader commitments of discourse ethics and reinforce the idea that we are all equal participants in this activity. Any resulting reduction in health disparities would be a welcome side effect.

As with cultural competency training and patient navigator programs, we can ask if a theory of justice (either of the Rawlsian or Sen-ish variety) could support the development and preservation of a Habermasian public sphere. Again, there is an obvious argument: if the absence of a public sphere leads to health disparities and health disparities lead to inequality of opportunity/capabilities, we should take steps to develop and preserve a public sphere. How plausible is this argument given the constraints on the domain of justice? Since one of
the central roles of the public sphere is to create a space that works to equalize status
relations by requiring participants to approach one another with a certain set of attitudes,
theories of justice may have difficulty justifying its creation. An individual’s attitudes are
not a central focus of justice. If, however, we think of the public sphere as an institution that
would be particularly useful, for example, in ensuring the stability of a society that the
Rawlsian principles of justice structure, then there might be grounds within a theory of
justice to argue for its implementation. Alternatively, a Rawlsian might argue that the public
sphere is necessary for solving certain political disagreements in a society where justice as
fairness serves as a point of overlapping consensus. In both cases, however, we do not want
the public sphere merely because it is useful for social stability or solving certain kinds of
disagreement; rather, it plays an essential role in reinforcing our status as participants in
communicative discourse and it is this status that is worth preserving.

There is room here for an extended debate about the ways in which theories of justice
can be shaped to either directly or indirectly account for the relationship between
interpersonal attitudes, status judgments, and health. Sparking this debate has been a central
aim of this dissertation. If we are to expand theories of justice to cover the broader
determinants of health—as Daniels and Ruger have suggested—then we must make sure that
they are equipped to do so. While I suspect that we will have more luck using a set of
theories directly focused on those determinants, I do not want to preclude a conversation that
may potentially clarify the requirements of justice in this area. Thus the policy
recommendations I have presented here represent a challenge to theorists of justice. If we
think their implementation is morally appropriate (or required), can we explain how they
could be grounded within a framework focused on justice?
Chapter IX Conclusions: Political Change

In September 2007 Congress passed the SCHIP expansion bill by substantial majorities in both Houses. President Bush vetoed the legislation on the grounds that, “I happen to believe that what you're seeing when you expand eligibility for federal programs is the desire by some in Washington, D.C. to federalize health care. I don't think that's good for the country.”302 Bush’s veto was applauded by several legislators from North Carolina, the country’s leading producer of tobacco products, who argued that it was unfair that the economy of the state should bear the brunt of the tax increase necessary to provide the funding for the health insurance expansion. Again, an underlying question in the debate was whether justice required allocation of resources to children from poor families so that they could have improved access to health care. In mid-October, Congress unsuccessfully attempted to overrode the presidential veto. The legislature and the President will likely compromise, continuing funding for existing SCHIP enrollees, but not increasing the number of eligible participants.

The debate over SCHIP is both good and bad news. On the positive side, it shows a political commitment to improving the health of poor Americans, a central aim of the bill’s supporters. Unfortunately, it also illustrates that any significant change in this area is likely to take the form of expanding health insurance coverage. In fact, all of the candidates for the 2008 presidential election have endorsed some plan for expanding health insurance coverage, whether through tax credits (Giuliani), an expanded government-sponsored health-care

302 Bush, 2007
system (Edwards), or a mixed public-private insurance program (Obama) coupled with coverage mandates (Clinton). The impact of these programs on American health disparities is hard to predict, although experiences with universal coverage in Europe have not been encouraging. Despite the growing epidemiological evidence that differences in socioeconomic status may be more important for health than access to care, it seems unlikely that there will be legislation specifically targeting the broader determinants of health at any point in the foreseeable future. Rather than seeing the expansion of health care access as an occasion for hand-wringing, however, it is probably better viewed as one step in a shift toward addressing the broader determinants of health, a task now occupying the health agenda of several European countries.

Putting aside the question of where practical change will occur, on what grounds and to what extent might we seek to reduce the socioeconomic status gradients that work to produce health disparities? Ultimately, this dissertation is about a shift in emphasis. The historical debate over the moral status of health disparities has taken place as a special topic within the domain of distributive justice. The central question has been whether it is unfair that health is distributed unevenly. I have tried to suggest that this is an incomplete way to argue about health disparities. A broader moral perspective is helpful in guiding our response to the social determinants of health, especially the way that stigma and negative interpersonal attitudes can create health disparities. As I have argued, theories of justice have, at best, a derivative concern with many of the broader determinants. I have, therefore, advocated theories of socioeconomic status as an alternative approach to thinking about the moral status of health disparities.
If distributive justice is focused on how we should organize the basic institutions of society, theories of socioeconomic status are concerned with which social and economic characteristics of persons are grounds for valuing them differently—for giving them higher or lower SES. When confronted with SES gradients in which status is accorded inappropriately, these theories make recommendations for altering our current beliefs about what characteristics give a person higher or lower SES and redistributing or deemphasizing the (inappropriate) determinants of value. When these goals have been accomplished, the result will be a SES hierarchy of a certain steepness, depending on how widely the determinants of SES are distributed. Since health disparities correlate with SES, a conception of SES—when coupled with an account of overall status—will identify what constitutes an acceptable difference in health between different members of a social system.

As an example of this kind of theory, I have suggested that Habermas’ discourse ethics can help us set a minimal level of SES below which the lowest members of society ought not to fall. A minimal threshold approach picks out a social characteristic of persons that establishes the most basic set of behaviors, attitudes, and responses we should have toward one another. For Habermas, this social feature is our ability to engage in communicative discourse. According to a conception of SES built around discourse ethics, no one should be treated in ways that would violate his or her status as a participant in communicative discourse. If we did not have this status it would be appropriate to exclude us from decisions that affect our lives; to coerce us into following policies that we had no role in developing; and to suppress our activities. As creatures able to engage in communicative discourse, however, the principle of universalism (U) rules out activities, behaviors, or attitudes that would treat us in these ways. I have argued that, once this
minimal level of SES is established, the result will be a reduction in health disparities.

Whether we take this approach to SES or another, we will have to think about the ways in which factors beyond distributable goods shape health.
REFERENCES


Rofes, E. I Thought People Like That Killed Themselves: Lesbians, Gay Men and Suicide. San Francisco: Grey Fox, 1983.


Scott, J. “Life at the top in America isn't just better, it's longer.” The New York Times, May 16, 2005 Section A; Column 1; National Desk; pg. 1


The Simpsons. Episode no. 21, first broadcast 20 April 1997 by Fox. Directed by Kirkland, M. and written by Swartzwelder, J.


