Ethics Inf Technol (2010) 12:29–42
DOI 10.1007/s10676-009-9198-6

ORIGINAL PAPER

Engaging rational discrimination: exploring reasons for placing regulatory constraints on decision support systems

Oscar H. Gandy Jr.

Published online: 10 June 2009

Abstract In the future systems of ambient intelligence will include decision support systems that will automate the process of discrimination among people that seek entry into environments and to engage in search of the opportunities that are available there. This article argues that these systems must be subject to active and continuous assessment and regulation because of the ways in which they are likely to contribute to economic and social inequality. This regulatory constraint must involve limitations on the collection and use of information about individuals and groups. The article explores a variety of rationales or justifications for establishing these limits. It emphasizes the unintended consequences that flow from the use of these systems as the most compelling rationale.

Keywords Race · Discrimination · Surveillance · Ambient intelligence · Privacy · Insurance · Technology assessment

Introduction

Although I am not a philosopher, I am going to attempt to place my thinking about emergent sociotechnical systems within frameworks that are common to discussions about ethics and technological choice. At the heart of my comments is a concern about the role that decision support systems should be allowed to play in our lives.

I believe that much of what we are concerned about as we imagine a not too distant future is the role that high-tech sentinels will play at the various nodes and access points we will encounter as we navigate an increasingly integrated network infrastructure.[1]

We understand the role of a sentinel at an outpost asking "who goes there; friend or foe?" We understand the sentinel makes use of a technology: it asks for a code, and it asks documentation. The sentinel is in control.

We already envision a time when this sentinel will be a kind of an ambient intelligence (AmI),[2] so it shouldn't take much for us to understand that the sentinel doesn't need to query us directly, but will access our identity systems automatically.[3]

It is already obvious that the bulk of this surveillance infrastructure will not be contained within us, or carried entirely on our person, despite the kinds of technological developments that Ray Kurzweil perceives on a not too distant horizon.[4] The sentinel will query a networked system of distributed intelligence that will evaluate our identities not only in terms of who we are biologically and historically, but also in terms of the state of the environment and its readiness to engage with the likes of us.[5] The sentinel will make use of Bayesian systems processing continuous streams of transaction-generated information to routinely update and adjust the system's assessments of risk.[6] Of course, these estimates will be of interest to a continually changing network of interested parties, and

O. H. Gandy Jr. (✉)
Annenberg School for Communication, University of Pennsylvania, Philadelphia, PA, USA
e-mail: ogandy@asc.upenn.edu

[1] Committee on Networked Systems of Embedded Computers (2001).

[2] Patrick (2007).

[3] Hildebrandt and Meints (2006).

[4] Kurzweil (2005).

[5] Lafuente-Rojo et al. (2007).

[6] Haddawy (1999).

autonomous agents that will have claimed a right to be informed about who goes where.

While our own interests and desires will certainly play some role in determining just how our encounters with these sentinels will unfold, none of us should really assume that we, rather than the network will actually be in control.[7]

Of course, my comments in this paper may not stray as far as some who have actually begun to suggest that we have pretty much lost all control over the determination of just who, or what we will actually come to be.[8] Instead, I will attempt to draw your attention to more mundane considerations about what it means to develop and implement systems that discriminate between people in ways that affect their life chances.

Although my concerns are framed in terms of a technological future when these systems are likely to be ubiquitous and therefore not only inescapable, but also not subject to identification as individually responsible actors or agents, I must necessarily express my concerns in terms of already existing practices and emerging trends.

Because it is at the heart of my concerns, I will begin this discussion with an attempt to define the meaning of discrimination in ways that link it to statistical analysis (SA) as a form of information technology (IT) that finds its use in decision support systems (DSS).[9] Next, I will introduce a set of considerations that have been brought to bear in debates about limiting the use of information about members of groups within a DSS. After exploring a subset of these arguments in greater detail, I turn to what I consider to be the most compelling rationale for limiting the use of particular kinds of DSS—the unintended consequences that we are beginning to recognize as cumulative disadvantage (CD). I will conclude with an assessment of the possibility that a regulatory response can be framed within the limits of routine technology assessment.

Technology assessment and discrimination

I think of discrimination as the outcome of a technologically enhanced process that begins with identification, proceeds through classification, and gathers momentum at the point of evaluation. Increasingly this evaluation is forward-looking, based on predictions about what some target of interest is likely to do at some point in the future.[10]

The routine, but increasing investment of resources in the identification, classification and evaluation of people, places and things is meant to produce actionable intelligence or guidance about choices that have to be made. These choices often result in the reinforcement or exacerbation of inequality within society.[11] The choice between X and Y, where Y is everything other than X, is to be understood as discrimination. The choice of X generally implies discrimination against Y. Of course, in an ideal case where X and Y are two applicants for a mortgage, the decision to grant the mortgage to X might be based on the flip of a coin, and Y then, would be the victim of bad luck, rather than discrimination.

Although the subprime mortgage crisis in the US seems to indicate that the dominant technology in use was often no more sophisticated than that of a coin flip,[12] a financial decision in favor of X increasingly would have been made on the basis of a rather complex underwriting program. Although the use of this kind of decision support technology may avoid some of the distortions that racial, and other group-centered biases might have played in the past, it remains likely that the inclusion of particular variables within many of the standard underwriting models will still work to produce a disparate effect that many of us see as both unfair and unfortunate.[13]

Because routine decision making systems become normalized quite rapidly, especially when they operate outside our fields of vision, it becomes especially important for us to consider that the discriminatory technologies that will be fully integrated into the ubiquitous networks of AmI should be subject to routine, if not continuous assessment and regulatory control.[14]

Far too often, the regulatory control of technology is motivated by concerns about threats to competitive advantage. Far less often, assessment and control is motivated by a desire to minimize or mitigate the societal harms that flow from its use. Most public technology assessments are retrospective, often initiated in response to a crisis, rather than beginning as an integral component of research and development. Still, public discussion and debate often accompanies the rise and fall of every technology's star, although regulatory interventions rarely succeed in banning the introduction of an emergent technology.[15] Nevertheless, it is quite common for legislatures to introduce controls over the nature and scope of its application, including the identification of uses and users that should be disallowed.

Initially, at least, regulations based on prospective and retrospective technology assessments may establish limits

[7] Brey (2005).

[8] Rouvroy (2008).

[9] Gandy (1995).

[10] Gandy (2006a).

[11] Graham (2005).

[12] ACORN Fair Housing (2007).

[13] Chandler (2002).

[14] Friedewald et al. (2006).

[15] Baumgartner and Jones (1993).

on the kinds of materials and resources that can be used. On occasion, regulatory initiatives may impose impact assessment and reporting requirements that enhance the possibility that unanticipated and unintended consequences will be identified and addressed before a critical point of no return has been reached.[16] Automated discrimination by AmI systems is precisely the kind of rapidly developing technology that has to be brought under the control of a global regulatory regime, and these controls have to be established now, rather than later.[17] In order for such a regime to succeed, however, it will need to have an appropriate identity.

Considerable time and effort has been spent in trying to control information-intensive activities under the banner of privacy or data protection.[18] I include myself among those who have come to realize that neither privacy, nor surveillance frameworks are adequate to the task of managing a system whose purpose is discrimination.[19] At the heart of the problem is the difficulty we face in controlling the use of information about people. Existing privacy and data management regimes are severely handicapped by a set of seemingly intractable problems related to the establishment and defense of property interests in information.[20] From my perspective, negotiating ownership claims is little more than an annoying distraction in ongoing debates about the establishment of regulatory limits on the use of "personal" or "personally-identifiable" information by private and public organizations.[21] I believe that it is far more important to find a way to engage the threat of computer-assisted discrimination more directly.

Limiting the use of information

There are a number of logical and ethical reasons for limiting the use of information that are not based on ownership, but on considerations of appropriateness or legitimacy. They include technical concerns associated with the data and the underlying analytical models that may generate inaccurate and biased recommendations. They also include moral and ethical concerns related to assumptions about causality and social responsibility. Finally, they include concerns about unanticipated, or unintended consequences and the nature of their distribution. Without suggesting that the unintended consequences that flow from the use of DSS are limited to any specific group, most of the examples of the effects that I will associate with CD in this article apply to African Americans.

We assume that rational decisionmaking involves the use of information within the context of some structured process that determines which information is both reliable and relevant to the task at hand. We don't always agree about the extent to which those standards have been met.

The first, and least controversial reason for questioning the use of data for critical decisionmaking is that the data or information is simply wrong. In the case of information about an individual, a clerical or transcription error may mean that a unique identifier has been associated with the wrong individual. Similar errors would mean that attributes ascribed to, or associated with a particular individual are incorrect or false. Many of these errors will be random, but criminal intent as well as structural constraints also tends to be reflected in the data that get recorded and distributed for later use.[22]

A more difficult assessment involves an evaluation of the completeness or correctness of the models or routines that generate the recommendations. Assuming the accuracy of the data used, poor system performance generally reflects an inappropriate choice of variables or indicators, or a misunderstanding of the underlying mechanisms that condition a behavioral response. Often, many of the variables used in a model are only weakly related to the outcomes of interest. Sometimes this may be the result of poorly specified outcome or criterion measures.

A somewhat different, but closely related concern arises in the case of some model or DSS in which the statistical relationship between predictors and outcome variables may be strong. Reliance on these systems for critical or consequential decisions are open to challenge because one or more of the predictor variables are merely correlated with the criterion, but no causal argument, rationale, or theory has been provided.

For somewhat different reasons, some of those who oppose the use of a particular program or system may argue that some variables or types of data should not be used because of philosophical or ethical concerns regarding fairness, or moral relevance, despite the fact that these variables actually perform quite well as predictors.

Some of the most troublesome candidates for regulatory exclusion or control are variables that have a strong historically generated structural linkage with other measures that we have already agreed to ban. Here I refer primarily to the measures of socioeconomic status and attainment that are closely associated with indicators of race, ethnicity and gender.

16 Tenner (1996).

17 Hildebrandt et al. (2005).

18 Waldo et al. (2007).

19 Gandy (2006b).

20 Landes and Posner (2004).

21 Hildebrandt (2006).

22 Schneier (2003).

Before we explore some of these reasons in a bit more detail, we need to come to terms with some fundamental distinction between two kinds of deciders: human beings and computer software, or machines.

Questioning the underlying model

Although we will ultimately have to determine the basis on which we should distinguish between choices made by machines or autonomous agents, choices made by humans with the assistance of machines, and choices made entirely by human beings, we should start with first things first. What does it mean to be rational, and to make a rational choice on the basis of a meaningful and relevant distinction?

Defining a concept by means of its opposition is rarely satisfactory, but it is a place to begin. Irrational decision-making is commonly associated with emotional or habitual, responses, informed by broad generalizations, rather than by careful weighing of the relevant facts. Rational decision-making generally refers to the process, rather than the outcome or results of any decision, although we understand that a carefully considered decision arrived at following a process of extensive search, reflection, and analysis, can still produce unsatisfactory results.

A realization that there are constraints on the ability of humans to access and incorporate all relevant information has led to the suggestion that the process is not necessarily irrational, but merely constrained or "bounded."[23] Most often, the concept of bounded rationality is focused on the limits of human information processing, rather than on limitations on access to information that may reflect strategic misdirection. But, as Giddens reminds us, some of the more important constraints on human agency are those blind spots we have regarding the motivations and goals of other interested parties that may be involved in some aspect of our decision-making.[24]

There is a tendency to think about rationality in terms of a continuum: one that moves from an idealized intelligence toward a comedy of errors. On one end of the continuum we might imagine a difference engine that engages in rapid computation, without errors in calculation, and more critically, without any systematic bias introduced by irrational emotional distractions. On the other end of the continuum we find the sometimes slow, sometimes fast, error prone, easily distracted, and routinely distorted and inconsistent information processing by humans.

Emotionalism

A substantial part of the problem of human decisionmaking seems to be associated with the role that emotions are believed to play in the choices that we make. There are, of course, a great many scholars who suggest that emotional responses play an important, and functionally necessary role in our decisions. They suggest that our emotional sensors are the products of long-term evolutionary development, and almost by definition, they have served some of us well. Indeed, some developers of artificial intelligence systems (AI) are actively pursuing the integration of emotional components into their machines.[25] Of course, it is not exactly clear whether those efforts are driven by a desire to be more rational, or to be more human.

On the other hand, a substantial literature, and a well-documented history of civil rights and anti-discrimination legislation has been developed in an effort to correct the bias and distortion that prejudice, disregard, animus, and own-group favoritism by humans often introduce into the calculus of social choice. Because these threats to rationality are so tightly woven into our individual assessment routines, attempts to minimize their influence on routine and more critical decisions have only been partially successful. In part, this is because key aspects of our decision-making routines are largely automatic. As a result, cognitive habits or shortcuts often shape decisions before, or instead of more careful reflection and analysis. Arguably, "only a small percentage of what we think of as judgments about situations with potential moral implications is the product of reasoned analysis. Instead, most judgments reflect immediate intuitive reactions, which individuals then justify post hoc by recourse to what they regard as socially acceptable reasons."[26] A variety of approaches to understanding this process have been taken.

Social psychologists have focused on the ways in which moral reasoning is shaped by structural conditions that we might associate with institutional and organizational cultures.[27] Neuroscientists have sought to understand this process at the molecular level through techniques based on nuclear imaging of the brain, while cognitive psychologists have developed less technology-intensive approaches toward understanding the activation of biased information processing routines.

One measure, the Implicit Association Test (IAT) has been used extensively to reveal the role that association between category membership and evaluative assessments plays in our unconscious reactions to people. Differences in response rates to a paired sorting task are interpreted as a

23 Elster (1990).

24 Giddens (1984).

25 Belavkin (2001).

26 Belavkin (2001, p. 950).

27 Regan (2007).

reliable index of an implicit preference for the members of one group over those of another.[28] A number of studies that make use of the IAT challenge simple-minded theories about racial prejudice and the way it shapes our decisions. Own-group preferences explain some, but not all of the racial biases at work. While bias against African Americans certainly reflects a racial influence in that more whites than blacks reveal an anti-black bias, upwards of 50% of African Americans also reveal anti-black bias when measured by the IAT.[29]

Although the distinction is not well defined, there is a tendency to treat preference-based discrimination as irrational, while granting belief-based discrimination the benefit of the doubt. The fact that preferences can be based to some degree on identifiable beliefs or cognitions, while an aversive response is emotional and automatic is the primary basis for maintaining a distinction between the two. However, it is also clear that stereotypical beliefs about the characteristics of disfavored groups may be used to reinforce, or justify a hostile or intuitively negative response.[30]

Because unconscious group biases, and intuitive reactions have such a powerful influence on the routine decisions made by individuals at critical points of interaction with persons at risk, many have applauded the introduction of automated, or computer-assisted decision-making strategies. While such systems may reduce the impact of biased individuals, they may also normalize the far more massive impacts of system-level biases and blind spots with regard to structural impediments that magnify the impact that disparities in starting position will have on subsequent opportunities.

Data quality and relevance

In my own work, I have tried to call attention to cases where the quality and character of the data used by policy makers has been influenced by the success of policy entrepreneurs in providing direct and indirect information subsidies.[31] Yet, there are a variety of alternate explanations for the fact that incomplete, or inaccurate data come to be used so often in making important public policy decisions.[32] Reliance on biased, or inaccurate data seems especially likely when the data are probabilistic, or estimates of uncertain futures. This means, somewhat ironically, that as more and more public and private decisions are made in the context of uncertainty and risk, the probability that these decisions will be made on the basis of inaccurate, incomplete, or biased information will only increase.[33]

Although the community of scholars who argue against the use of non-causal variables in explanatory or predictive models is relatively small, I believe their arguments are worthy of our attention.[34] The claim that categorical variables are inherently non-causal is not as widely accepted as we might assume. It shouldn't require an elaborate argument to convince most people that a woman's race cannot be the cause of her behavior. Some people can be convinced by argument and statistical evidence that suggests that the apparent relationship between two variables is actually caused by a third variable that had somehow been excluded from the model. If the magnitude of an estimated correlation moves toward zero upon the introduction of this additional variable in the statistical model, some observers would be willing to conclude that the relationship must have been spurious, rather than causal.

Addressing more fundamental concerns about reliance on highly correlated, but non-causal variables requires engagement with ongoing epistemological debates regarding the nature of causality, and the criteria necessary for making a causal claim. The primary consideration is to be found at the heart of the revered status of the experimental model. For many, the controlled experiment with random assignment to treatment conditions is the standard against which all causal claims must be assessed. Unfortunately, at least for supporting a causal claim, it is simply not possible for a person's race, or gender, or even her age to be manipulated by the researcher. The absence of manipulability thereby denies such categories the possibility of achieving the status of cause.[35]

Other arguments regarding the causal status of variables can be seen in the context of challenges to the leakage or spread of indicators, or criteria beyond the domains in which they were initially developed. A prime example is the rapid diffusion of credit scores as an index of trustworthiness.[36] The Fair Credit Reporting Act[37] is quite generous in its identification of the permissible uses of a credit report, and by extension, their representation in a summary statistic. Not only does this list include insurance underwriting, licensing, and employment, but it also includes a seemingly unbounded range of possibilities within the residual category of "legitimate business need."[38]

Opposition to many of these uses has emerged for a variety of reasons over time. Public opposition to the use of

28 Nosek et al. (2006).

29 Kang and Banaji (2006).

30 Quillian (2006).

31 Gandy (1982).

32 Koomey et al. (2002)

33 Kleindorfer et al. (1993)

34 Holland (2008).

35 Zuberi (2001).

36 Altman and Saunders (1998).

37 Public Law No. 90-321, (1970).

38 Gandy (1996).

credit scores for risk rating by insurance companies has been focused primarily on the absence of a demonstrated causal link between financial irresponsibility and the filing of claims for accidental loss. The fact that they are both indicators of a person's financial risk status explains their statistical correlation, but insurers have been pressured to find a more compelling explanatory model.[39] The fact that the leading causal hypotheses favored by the industry rely on the identification of a personality type (sensation seekers) that may have roots in genetic predispositions, would suggests that this particular battle is far from over.[40]

Ian Ayres and his colleagues identify three kinds of discriminators: [41] those who are irrational, and make decisions entirely based on the race of the other; the rational discriminator who includes race as one among many sources of information that may help to shape a decision; and the hyper-rational discriminator, a rapid calculator, perhaps an "intelligent Bayesian"[42] who continually adjusts her expected value estimates on the basis of new information. While Bayesian probability theory makes an important contribution to the design of decision systems, serious problems arise at the point in which prior probabilities are specified.[43] All of the factors that represent challenges to the accuracy or correctness of facts also operate with regard to the selection of prior probabilities.

There are also a host of problems associated with the quality of the information used in updating probability estimates. These problems are especially troublesome when they take the form of summary statistics. Summary statistics are statements about attributes of a population or an aggregate. While the statistics are only supposed to refer to an aspect of the population as a whole, it is common for people to treat such measures as an indication of something held in common by every individual within the population. Data that are supposed to provide information about a group becomes a representation of each member of the group.[44] Summary statistics are stereotypes by a different name, and we have good reasons for questioning their use.

The use of stereotypes has been characterized as an adaptive response to cognitive demands that arise each day as people negotiate rapidly changing environments and interactions with strangers. Stereotypes can be thought of as laborsaving devices that enable the realization of cognitive efficiencies.[45] In the context of decisions that involve the assessment of risk, stereotypes may also be short hand indicators of the relative, or criterion level of risk a stranger represents. It is in this context that the role of stereotypes becomes especially problematic. Some argue that in the absence of individuating information about the intentions of the young black male coming your way, it is perfectly rational, and in no way racist for you to grip your purse more securely, or even to consider crossing the street. Arguably, the young man's race provides access to "base-rate" information about criminal victimization that many believe we would be foolish to ignore.[46]

Yet, the use of statistical indicators derived from a convenience, or strategic sampling of information about some population subgroup has the same potential for generating the same kinds of invidious distinctions reinforced by the stereotypes that many of us have learned to mistrust and have tried to set aside.

Ethical concerns

Ethical concerns arise because group, or categorical data is ordinarily understood to be a representation, or a statement about members of that group. These same representations are often implicated in the reproduction and reinforcement of inequality between groups in society. These statistical measures become even more powerful as they become linked with culturally weighted labels.

Labeling is a routine activity for analysts who use multivariate statistical techniques, such as factor analysis, to discover unmeasured relations between variables. In order to facilitate sense making and communications about the factors that emerge from their efforts, analysts develop distinctive and meaningful labels for each one. These names are often idiosyncratic to the researchers, their organizations, and the goals of their investigations. On occasion, however, those factors become constructs that become part of the marketing discourse used in promoting the commercial application of the evaluative distinctions enabled by the classification. The impact spreads further as the labels slip into common usage.

A classic example can be seen in the names that have been assigned to the geodemographic clusters developed by the Claritas Corporation for its PRIZM Cluster products. These trademarked names help to reinforce the corporate construction of community identities among marketers of consumer goods and policy initiatives. The name "Black Enterprise" may not have been any more problematic as a label for the middle class African Americans living in the suburbs than "Money & Brains" was for the young intellectual elites who lived in key cities around the US. I suspect, however, that the people living in communities characterized as "Shotguns and Pickups," "Norma Rae-

39 Birnbaum (2003).
40 Brockett and Golden (2007).
41 Ayres et al. (2005).
42 Armour (1997).
43 Florescu et al. (1997).
44 Carusi (2008).
45 Schneider (2004).
46 Kahn and Lambert (2001).

ville," "Tobacco Roads," or "Hard Scrabble," were probably not as pleased.[47] In addition to expanding the number of clusters they have identified, Claritas and its competitors have modified the names assigned to these communities so that the markers of race and class were not as disparaging and disrespectful as those used initially.[48]

Of course, ethical concerns about the use of categorical data are not limited to the questionable use of group names that have emerged from the lab, or the back office.[49] In some discussions, critics seek to set these variables aside because they are said not to be "morally relevant," either because they are immutable, or because they are outside the limits of effective individual choice or control. In the US, five categories or classes (race, ethnicity, religion, sex, and national origin) have been banned for use in a number of decisional domains.[50] Legal scholar Jack Balkin suggests that the best reason for deciding that the use of an immutable trait is unjust is because of the ways in which its use establishes, or reinforces a status hierarchy "that helps dominate and oppress people."[51] In his view, other traits that are not immutable, and do reflect autonomous choice, can still be seen as "morally irrelevant" and therefore inappropriate for use as a selection criterion. He offers religion as a prime example. Beyond mutability, choice and control, the reasons for opposing most forms of categorical discrimination are complex, and not very well understood.[52]

From time to time, successful mobilization of public opinion by activist organizations has added additional categories, such as sexual orientation to the list of banned, or suspect classes. The public debates surrounding the passage of these bills have tended to cover the same familiar ground. Most recently we have seen legislation passed in the US that would ban the use of genetic information in decisions about employment and insurance. This bill was passed, and then signed [53] after decades of both scholarly reflection and somewhat more rancorous debate.[54] Legislators reported having been influenced by constituents' suggestions that people shouldn't be penalized "because they happened to be born at a higher risk for a given disease."[55] This is a concern about fairness.

Considerations of fairness are commonplace, and people show a willingness to reward behaviors seen as fair, and to punish those that appear unjust. The importance of "reciprocal fairness" to social behavior has been demonstrated repeatedly in carefully constructed "games" where self-interested economic rationality is seen to be limited by actions designed to punish unfairness by others even at a relatively high personal cost.[56] Researchers in the emerging field of "neuroeconomics" suggest that this response may not be an expression of moral rectitude, but a desire to avoid being perceived as unfair.[57]

Unfortunately, at least with regard to its usefulness as a criterion against which to evaluate the use of information for decision-making, not all definitions of fairness are based on the same moral, ethical, or even evolutionary foundation. The insurance industry, for example, has done quite well in establishing its own self-serving definition of fairness as a regulatory and judicial norm.[58] Its evaluative standard sets as ideal a match between the total value of premiums paid, and the expected value of the legitimate claims made by an individual. However, because it is not possible for the insurer to actually know what an individual's future claims will actually be, it is said to be "actuarially fair" for them to set prices for individuals on the basis of the claims histories of the persons within the risk class, or population to which an individual is determined to belong. Actuarial fairness has little to do with considerations of justice or the moral relevance of the categories that are actually used in the establishment of rates.[59]

There is growing body of evidence demonstrating that human beings vary in the extent to which their decisions are influenced by forms of "moral sensitivity" to the correctness of their actions.[60] Although moral sentiments are associated with cultural norms and values, it is obvious that within all complex societies, there will be a great many points at which differences between people's views about some action will reflect an underlying disjuncture in ideology, value system, or world-view.

We will not be surprised to see that supporters and opponents of discriminatory techniques line up along ideological lines because of the ways in which these techniques affect concerns at the heart of an ideological divide. Milton Rokeach,[61] among others,[62] has demonstrated that those who hold either liberal or conservative political views can be reliably identified on the basis of their relative ranking of two terminal values: freedom and equality. Liberals consistently place a higher value on equality, while conservatives tend to favor freedom. To the

47 Weiss (1988).
48 Weiss (2000).
49 Bowker and Star (1999).
50 Wortham (1986).
51 Balkin (1997).
52 Amegashie (2008).
53 The Genetic Nondiscrimination Act (2008).
54 Van Hoyweghen et al. (2005).
55 Harmon (2008).
56 Knoch et al. (2006).
57 Sanfey et al. (2003).
58 Thiery and Van Shoubroeck (2006).
59 Lemmens and Thiery (2007).
60 Moll et al. (2008).
61 Rokeach (1973).
62 Lakoff (2002).

extent that discrimination produces, reproduces, or exacerbates inequality within society, liberals are more likely to oppose it, and to support public policies that restrict the use of discriminatory techniques.

Other social theorists who have studied the nature of the ideological divide associate conservatives with opposition to change or a preference for stability, without regard to its impact on their freedom to choose.[63] Although there are other analytical models that emphasize the role of different value complexes,[64] the impact of any discriminatory technology on social inequality will be an inescapable aspect of its evaluation within the public sphere.

Economic concerns

Cost is also a factor that plays a role in justifying the use of some information within decision systems. The cost of acquiring, or making use of information helps to determine its cost-effectiveness, or efficiency. Economic or efficiency guidelines suggest that we should use information as long as the costs of acquiring and using the information are less than the costs that would be faced in its absence. Unfortunately, as we will explore, there are important distinctions to be drawn between private costs, or the costs to the decider, and the social costs, or costs to society as a whole. Problems regarding the distribution of both private and social costs arise because the ease with which we can acquire categorical information, such as gender, and membership in certain racial and ethnic groups means that this information is likely to be over-used.[65]

However, economists are quick to suggest that to the extent that group membership provides reliable information about the group, and by extension about any individual group member, routine use of such information may be economically efficient.[66] There is an explicit assumption that placing constraints on the efficient use of information is irrational.

On the basis of this claim, arguments that proceed from an economic perspective suggest that a DSS that makes use of race, or gender, or other categories or classes into which people may be assigned, are legitimate, because they are efficient. The use of these categories is characterized as "statistical discrimination" and its use is routinely justified on the basis of the "information" or signal that group membership provides. Defenders of statistical discrimination are not deterred by the absence of any demonstrable causal linkage between these attributes and outcomes, or criterion measures; all that matters is that the association is strong enough to stand as a reliable indicator or predictor.

Group-based discrimination in pursuit of profit is readily justified on the basis of economic efficiency. A variety of techniques that target members of readily identifiable groups are designed to exploit existing knowledge about what makes some members of that group especially vulnerable to a specific marketing ploy. There are other ways in which knowledge of the constraints that affect members of disadvantaged groups can be used to the advantage of those with opportunities on offer. Courts have historically been willing to accept the use of market criteria that have a disparate impact on members of historically burdened groups as long as these methods can be shown to meet a legitimate business purpose. The enhancement of a firm's bottom line is one such purpose.[67]

As we have noted, critical distinctions are often drawn between forms of discrimination that are motivated by racial animus or prejudice, and forms that are motivated by what we think of as rational efforts to maximize pleasure and minimize pain. Those who engage in the presumptively irrational forms of discrimination can be said to have a "taste for discrimination."[68] There is a substantial economic literature that argues that this irrational taste will be punished by the market, and ultimately extinguished. Far less attention has been paid to consideration of an equally imaginable "taste for fairness" by which individuals would forgo advantages that they might otherwise realize because of the discriminatory tastes of those they might have to deal with.[69] The fact that received economic theory assumes that people are governed more by consideration of personal gain than by moral or ethical concerns no doubt explains the relative lack of attention that a taste for fairness, or justice has received.

Although Rokeach's list of terminal values did not include efficiency, economists and those who rely upon economic rationales have focused on the extent to which efficiency and equality are incompatible objectives or values.[70] Beginning with the articulation of the idea that tradeoffs between equity and efficiency were inescapable if the benefits of economic growth were to be achieved,[71] many economists have tried to subordinate concerns about equality in debates over a broad range of public policy matters that include the regulation of technology. Attempts to formalize the analysis of these tradeoffs through benefit/cost analysis largely failed to include measures of equity

63 Jost et al. (2008).
64 Wilson (2005).
65 Schauer (2003).
66 Norman (2003).
67 Ayres (2007).
68 Becker (1971).
69 Case (2002–2003).
70 Hausman and McPherson (1996).
71 Okun (1975).

that liberal policy advocates could recognize as adequate to the task.

The fact that as values, neither equity, nor efficiency could be “traded-off” or exchanged, one against the other because they are fundamentally incommensurable has not kept proponents of market-place solutions from attempting to justify them on the basis of poorly specified indicators of efficiency.[72] On the other hand, numerous critics have provided arguments and examples intended to demonstrate that increasing equality need not result in declines in either efficiency, or economic productivity.[73] Indeed, they have argued that inequality in general, and racial inequality in particular has been a substantial constraint on economic growth and development.[74] When the measure of inequality is focused on individual capabilities, as suggested by Nobel laureate Amartya Sen,[75] is easy to see how reducing inequality by enhancing the capabilities of the poorest and most disadvantaged members of the society, would contribute to the overall productivity of an economy. This expectation also forms the basis for Rawl’s suggestion that the only inequality that should be tolerated is an inequality in distribution that favors those with the most limited resources.[76]

Despite the steady flow of position papers that emphasize the importance of economic efficiency as an basis for policy evaluation, the general public still appears to regard considerations of equality in particular, and fairness more generally as considerations that matter for many of their decisions.[77] There is also substantial evidence that considerations of equality and fairness also influence aggregate decisions, including those made by markets and by political institutions.[78]

Because of the evidence that there remains a substantial preference for more, rather than less equality within society, it makes sense to call public attention to the ways in which emergent technology and social practices that depend upon them are likely to worsen inequality.

Unintended consequences: cumulative disadvantage

We regulate the use of technological systems or their standard inputs because of their unintended consequences. Economists invite us to think about these impacts as negative externalities. The most familiar examples can be seen in our attempts to control pollution and to limit the impact of carbon-based fuels on our environment by restricting their use by a variety of means. We have paid far less attention to readily identifiable forms of social pollution, but there are clear parallels to be seen. For example, the routine use of scantily clad and unimaginably svelte models as aids to the marketing of commodities is believed to contribute to sexism among males and body image problems among women.[79] Reliance upon advertising revenue to support the production and distribution of news also generates a host of negative externalities that are believed to threaten the operation of the democratic public sphere.[80]

I am arguing here that the use of categorical information, in this case information about racial group membership in decision support systems contributes to the cumulative disadvantages that reproduce and exacerbate multiple disparities that place African Americans at the bottom of the societal hierarchy.

We can think about cumulative disadvantage in terms of the distribution of life chances, or opportunities to enjoy the rewards of one’s efforts.[81] The life chances of some people, African Americans in particular, have been minimized by a complex of factors, that bear much in common with the burden of stigmatization.[82] The cumulative impact of stigmatization that is amplified through the utilization of race, and race-linked measures within panoptic discrimination systems[83] is most clearly seen in comparative measures of inequality.[84]

Income inequality is a contribution to CDs that persist across generations. This intergenerational burden is quite substantial among African Americans.[85] Examinations of inequality as a social problem describe the ways in which income inequality comes to be reflected in other critical disparities across groups, and across generations. Income is linked to quality of life including health, and in terms of other measures of satisfaction. Relative status is also linked to levels of political involvement, which reinforces the tendency of the political system to respond to the demands of affluent citizens who tend to be more politically engaged.

The absence of neighbors and friends with the kinds of knowledge, contacts, or confidence that social theorists talk about in terms of “social capital” also works to reinforce the impact of concentrated poverty within hyper-

72 Le Grand (1990).
73 Klasen (2006).
74 Ramirez (2004).
75 Sen (1992).
76 Rawls (1971).
77 Pirttila and Uusitalo (2007).
78 Camerer and Fehr (2006).
79 Yamamiya et al. (2005).
80 Baker (2002).
81 Dahrendorf (1979).
82 Link and Phelan (2001).
83 Gandy (1993).
84 DiPrete and Eirich (2006).
85 DiPrete and Eirich (2006, p. 339).

segregated communities.[86] In part, it is the geographic isolation of African Americans that reproduces and extends the CDs that have a basis in the burdens of poverty.[87] Geographic isolation reinforces the impact of deficits in social capital that are common within African American communities. Not only are black members of these communities less likely to trust their neighbors, or to have much contact with them, they are also less likely to be involved in community projects or in the political process more generally.

Among the many aspects of an individual's identity that determine the quality of life, the communities in which they live have begun to emerge as a vitally important indicator of current position and risk status. While it is not known how much of the disadvantage that African Americans face in the housing market can be attributed to predation and other abuses within the market for consumer credit,[88] it should be clear that the impact of statistical discrimination in the housing market can not be limited to decisions made by lenders.

The development of spatial analysis techniques that transform geographic coordinates into points of comparison for units ranging in scale from individual homes, blocks and neighborhoods, to ZIP codes and Census tracts, has played a central role in the socioeconomic development of these units. Geographic information systems (GIS) combine the ability of mapping technologies with the resources of advanced datamining systems to enable the visual representation of forward-looking scenarios about population growth, economic development, or about the decay and deterioration of community resources.[89]

The racial composition of neighborhoods continues to be a very powerful predictor of the socioeconomic trajectory of those communities. Continuing residential segregation by race has a variety of causes, but racial discrimination remains its primary cause. What is not clear, of course, is the extent to which this discrimination is primarily rational or statistical, rather than being based on more emotional sources of distain. The fact that African Americans are subject to hyper-segregation at rates that far exceed those of other minority group members lends some support to the belief that aversive racism is still a major component of this process.[90]

Another critical component of the CD that burdens African Americans is the association between race and crime. As we have suggested with regard to the impact that biased or inaccurate data have on the recommendations made by both human and computer-based guidance systems, self-confirming hypotheses operate to validate and reinforce initial impressions. Such is the case with regard to crime statistics and the arrest and imprisonment of African Americans for drug-related crimes.

Racial profiling is the name applied to a discriminatory technology that justifies a higher than average number of traffic stops of black drivers because of a belief that those drivers are more likely to be carrying contraband drugs.[91] If African Americans are actually no more likely to be carrying contraband than other drivers, then a higher stop rate will still produce a greater number of good hits, convictions, and sentences. The crime statistics that result will serve as proof of the correctness of the model, and may even be used to justify increasing the rate of such stops.[92]

There is also great deal to be said about the ways in which the use of this particular technology generates a host of unintended, cumulatively disastrous consequences. Growing disrespect for the law, the police, and the courts is just one small thread that threatens to unravel the social fabric of our nations.

There is a special irony in the fact that even within the civil justice system, we find significant barriers to access to resources that might otherwise help to reduce the level of inequality in society.[93] People burdened by poverty have a sense of powerlessness, and because of this, they tend to believe that there is little chance that they will succeed in pursing legal claims against more powerful others. As a result, the poor are far less likely to pursue legal action even when they believe that their claims are just. Often, their own experiences, or those of their friends and neighbors reinforces the sense that they will encounter barriers every step of the way toward what will ultimately turn out to be a negative decision by the court. Because personal injury lawyers also estimate their chances of recovering the costs of pursuing these cases, they are far more likely to avoid taking the cases of victims most in need of their help.[94]

The fact that public policies establish punishments and rewards for behaviors that reflect ideological biases more than they reflect objective indicators of societal value also means that these policies contribute to cumulative disadvantages for members of some groups. The examples are numerous. Consider, for example, the consequences that flow from treating particular kinds of drug offenses as being more dangerous, or problematic than others. If those offenses happen to be associated with particular groups, such as African Americans, then, over time, the burdens on

86 Saguaro Seminar on Civic Engagement in America (2000).

87 Sampson et al. (2002).

88 Medina (2004).

89 Lejano (2008).

90 Charles (Charles 2003).

91 Gandy and Baruh (2006).

92 Harcourt (2007).

93 Sandefur (2008).

94 Sandefur (2008, p. 10).

African Americans will cumulate, and spread throughout their communities.[95]

Conclusion

This article has covered a lot of ground, especially with regard to CD and other unintended consequences that flow from the use of discriminatory technologies. Other examples from other domains could be easily added to those we have examined so far. The process will operate in essentially the same way, with the same results, whenever the purpose, or the primary function of the DSS is to assist discrimination on the basis of group or category membership. For this reason, the continued development and implementation of gatekeepers, monitors or sentinels that rely on automated profiling cannot be expected to produce socially optimal results, simply because these systems lack the particular sources of bias that we have associated with decisions made by human beings.

It must be recognized that certain kind of biases are inherent in the selection of the goals or objective functions that automated systems will designed to support. We must also consider that fact that as system objectives more routinely come to be framed in terms of the identification, minimization, or management of risks, rather than the achievement of objectively measured goals or achievements, the consequences of systematic error will be more difficult to observe and to control. When the goals of risk management are in the service of profit maximization and the maintenance of structures and relations of power, the introduction of this kind of bias is all but guaranteed.

A common feature of our discussions of the performance of computers, or computer-based systems, has been the extent to which these systems generate errors. A great many believers in the promise of artificial intelligence (AI) are confident that we will reach a stage in the development of autonomous systems when they will engage in routine self-correction, thereby eliminating concerns about computational errors attributable to exogenous threats like radiation, network failures, or malicious attacks.[96] A smaller number of AI enthusiasts believe that adaptive learning and autonomous alternations of code will reduce the frequency and impact of errors that are generated endogenously by the models, subroutines, and agents that govern the processing of information.

Certainly, our evaluation of these systems has to take error rates into account. However, not all kinds of errors are the same. Depending upon the nature of the goals of the system controllers, their orientation toward Type I and Type II errors will be quite different.[97] In the case of security systems, where the sentinel is determining whether the request or the transaction is a legitimate request or an attack, the primary goal is not to miss or ignore any attacks, that is, to minimize passive failures. However, because the overwhelming majority of interactions with automated systems are legitimate, a goal of minimizing passive failures will necessarily mean that there will be a high number of active failures, or the denial of access, or the assignment of burdens on clients, customers, or citizens who were actually making legitimate requests.

The determination of the error rates that are acceptable will not be left to the machine. These will be economic or political choices. These are critical choices that should be subject to routine review and assessment in the interest of the public at large. These reviews should be especially sensitive to distributional impacts that too often escape the regulatory gaze.[98]

Our regulatory assessment and response to the performance and impact of these systems must consider the distribution of errors and unintended consequences that we cannot assume will be random. It will be essential for us to ensure that those who are assigned the responsibility for technology assessments along these lines are genuinely committed to the reduction of inequality within and between nations and regions of the world.

And finally, our assessment will have to be especially sensitive to the relative scales at which different discriminators operate. It will not make sense to ignore the decisions made by autonomous intelligent sentinels even though their error rates might in fact be quite remarkably low. The fact that the number of decisions made by such sentinels may be several orders of magnitude greater than those made by human deciders suggests that cumulative, as well as catastrophic harms would emerge simply as a function of scale.

As I suggested earlier on, a privacy or data protection regime will not be adequate to the task of reducing the cumulatively disparate impact of autonomous discriminators. Privacy and data protection regimes focus primarily on the correctness of the "personally-identified information" stored within a database. It simply will not be economically or even technically feasible for data subjects to assess and then challenge the "correctness" or accuracy of the data or analytical models used by sentinels, agents, or environments.

[95] Chin (2002).

[96] Whitby et al. (2004).

[97] Type I errors refers to a conclusion claiming distinction, or difference, when there is no difference; Type II errors refer to a conclusion of no difference, when a difference, or distinction actually exists. An active failure is a Type I error; claiming a threat when there is none.

[98] Hahn and Tetlock (2007).

Most of the time, persons who have been victimized by a routine system error will not know precisely if, when, or how they have been discriminated against. If they somehow suspect, or become aware of an erroneous assessment, they will not know where to begin to gain access to the data, or the algorithms that were used to adjust their profile or status. Even with access, and with the aid of an advocate, they will likely discover that the offending result has been generated by factually correct data used within an operationally validated system that performs as well, or better than its competitors in avoiding passive failure. They won't be able to bring claims on the basis of maldistributed burdens because those data will not exist.

Framing a demand for an adjustment of such systems might be possible on a case-by-case basis, but in general, success would be so unlikely that the most rational response will be to ignore it. At the end of the day, it will be the nature of the organizational or institutional purpose or goals that will determine whether the use of any particular variable, measure, or model can be successfully challenged.[99]

Only a well-organized and highly motivated social movement is likely to succeed in challenging the elevation of a set of institutional goals and objectives beyond the reach of individuals and small groups. No such movements appear to be massing at the horizon. That doesn't mean there is no basis for hope. Perhaps a merger between egalitarian movements for civil rights and environmental movements for sustainability[100] may generate sufficient energy and intellectual resources to carry the struggle against automated and autonomous discrimination forward.

References

ACORN Fair Housing. (2007). *Foreclosure exposure: A study of racial and income disparities in home mortgage lending in 172 American cities*. Chicago: ACORN Housing Corporation.

Altman, E., & Saunders, A. (1998). Credit risk measurement: Developments over the last 20 years. *Journal of Banking and Finance*, *21*:1721–1742. See also, Mester, L. (1997). What's the point of credit scoring? *Business Review*, October/November, pp. 3–16.

Amegashie, J. A. (2008). *Socially-tolerable discrimination*. MPRA paper no. 8543, Munich Personal RePEc Archive. http://mpra.ub.uni-meunchen.de/8543.

Armour, J. (1997). *Negrophobia and reasonable racism* (pp. 35–67). New York: New York University Press.

Ayres, I. (2007). Market power and inequality: A competitive conduct standard for assessing when disparate impacts are unjustified. *California Law Review, 95*, 669–719.

Ayres, I., Vars, F., & Zakariya, N. (2005). To insure prejudice: Racial disparities in taxicab tipping. *Yale Law Journal, 114*, 1640.

[99] Thiery and Van Shoubroeck (2006).

[100] Barnes (2006).

Baker, C. E. (2002). *Media, markets, and democracy*. New York: Cambridge University Press.

Balkin, J. M. (1997). The constitution of status. *Yale Law Journal, 106*, 2366.

Barnes, P. (2006). *Capitalism 3.0: A guide to reclaiming the commons*. San Francisco: Berrett-Koehler.

Baumgartner, F., & Jones, B. (1993). *Agendas and instability in American politics*. Chicago: University of Chicago Press.

Becker, G. (1971). *The economics of discrimination* (2nd ed.). Chicago: University of Chicago Press.

Belavkin, R. (2001). *The role of emotion in problem solving*. In C. Johnson (ed), Proceedings of the AISB '01 Symposium on emotion, cognition and affective computing (pp. 49–57). York: Heslington. See also, Canamero, D. (2001). Modeling motivations and emotions as a basis for intelligent behavior. Agents '97, ACM, 1997, and Aaron Sloman. Beyond shallow models of emotion. *Cognitive Processing*, *2*:177–198.

Birnbaum, B. (2003). *Insurers' use of credit scoring for homeowners insurance in Ohio*. Columbus: Ohio Civil Rights Commission.

Bowker, G. C., & Star, S. L. (1999). *Sorting things out: Classification and its consequences*. Cambridge: The MIT Press.

Brey, P. (2005). Freedom and privacy in ambient intelligence. *Ethics and Information Technology, 7*, 157–166.

Brockett, P. L., & Golden, L. L. (2007). Biological and psychobehavioral correlates of credit scores and automobile insurance losses: Toward an explication of why credit scoring works. *The Journal of Risk and Insurance, 74*, 23–63.

Camerer, C., & Fehr, E. (2006). When does "Economic Man" dominate social behavior? *Science, 311*, 47–52.

Carusi, A. (2008). Data as representation: Beyond anonymity in e-research ethics. *International Journal of Internet Research Ethics, 1*, 37–65.

Case, M. A. (2002–2003). Developing a taste for not being discriminated against. *Stanford Law Review*, *55*:2273–2291.

Chandler, S. (2002). Visualizing adverse selection: An economic approach to the law of insurance underwriting. *Connecticut Insurance Law Journal, 8*, 435–503.

Charles, C. Z. (2003). The dynamics of racial residential segregation. *Annual Review of Sociology, 29*, 167–207.

Chin, G. J. (2002). Race, the war on drugs, and the collateral consequences of criminal conviction. *The Journal of Gender, Race & Justice, 6*(25), 3–275.

Committee on Networked Systems of Embedded Computers. (2001). *Embedded, everywhere: A research agenda for networked systems of embedded computers*. Washington: National Research Council, National Academy Press.

Dahrendorf, R. (1979). *Life chances: Approaches to social and political theory*. Chicago: University of Chicago Press.

DiPrete, T., & Eirich, G. (2006). Cumulative advantage as a mechanism for inequality: A review of theoretical and empirical developments. *Annual Review of Sociology, 32*, 271–297.

Elster, J. (1990). When rationality fails. In K. S. Cook & M. Levi (Eds.), *The limits of rationality* (pp. 19–51). Chicago: University of Chicago Press.

Florescu, D., Koller, D., & Levy, A. (1997). *Using probabilistic information in data integration*. Proceedings of the 23rd VLDB Conference. Athens: Greece, pp. 216–225.

Friedewald, M., Lindner, R., & Wright, D. (eds.), (2006). Policy options to counteract threats and vulnerabilities in ambient intelligence, SWAMI. Deliverable D3: A report of the SWAMI consortium to the European Commission under contract 006507. http://swami.jrc.es.

Gandy, O. H., Jr. (1982). *Beyond agenda setting: Information subsidies and public policy*. Norwood: Ablex. See also, Gandy, O. H., Jr. (2003). Public opinion surveys and the formation of privacy policy. *Journal of Social Issues*, *59*:283–299.

Gandy, O. (1993). *The panoptic sort: A political economy of personal information*. Boulder: Westview Press.
Gandy, O. (1995). It's discrimination stupid. In J. Brook & I. Boal (Eds.), *Resisting the virtual life: The culture and politics of information* (pp. 35–47). San Francisco: City Lights Books.
Gandy, O. (1996). *Legitimate business interest: No end in sight? An inquiry into the status of privacy in cyberspace* (pp. 77–137). Chicago: University of Chicago Legal Forum.
Gandy, O. (2006a). Quixotics unite! Engaging the pragmatists on rational discrimination. In D. Lyon (Ed.), *Theorizing surveillance: The Panopticon and beyond* (pp. 318–336). Portland: Willan Publishing.
Gandy, O. H., Jr. (2006b). Data mining, surveillance, and discrimination in the post-9/11 environment. In K. Haggerty & R. Ericson (Eds.), *The new politics of surveillance and visibility* (pp. 363–384). Toronto: University of Toronto Press.
Gandy, O., & Baruh, L. (2006). Racial profiling: They said it was against the law!. *University of Ottawa Law & Technology Journal, 3*, 297–327.
Giddens, A. (1984). *The constitution of society* (pp. 5–14). Cambridge: Polity Press.
Graham, S. (2005). Software-sorted geographies. *Progress in Human Geography, 29*, 1–19.
Haddawy, P. (1999). An overview of some recent developments in Bayesian problem solving techniques. *The AI Magazine, 20*, 11–19.
Hahn, R., & Tetlock, P. C. (2007). *Has economic analysis improved regulatory decisions? Working paper 07–08*. Washington: AEI-Brookings Joint Center for Regulatory Studies.
Harcourt, B. (2007). *Against prediction: Profiling, policing, and punishment in an actuarial age*. Chicago: University of Chicago Press.
Harmon, A. (2008). Congress passes bill to bar bias based on genes. In *New York Times Online*, http://www.nytimes.com/2008/05/02/health/policy/02gene.html.
Hausman, D., & McPherson, M. (1996). *Economic analysis and moral philosophy*. New York: Cambridge University Press.
Hildebrandt, M. (2006). Profiling: From data to knowledge. *Datenschutz und Datensicherheit, 30*, 548–552.
Hildebrandt, M., & Meints, M. (eds). (2006). D7.7: RFID, Profiling, and AmI. Future of identity in the information society (FDIS Consortium).
Hildebrandt, M., Gutwirth, S., & De Hert, P. (2005). Implications of profiling practices on democracy and the rule of law. Deliverable D7.4, FIDIS Consortium.
Holland, P. W. (2008). Causation and race. In T. Zuberi & E. Bonita-Silva (Eds.), *White logic, White methods* (pp. 93–109). Lanham: Rowman & Littlefield.
Jost, J., Nosek, B., & Gosling, S. (2008). "Ideology: Its resurgence in social, personality, and political psychology. *Perspectives on Psychological Science, 3*(2), 126–136.
Kahn, S., & Lambert, A. (2001). Perceptions of rational discrimination: When do people attempt to justify race-based prejudice? *Basic and Applied Social Psychology, 23*, 43–53.
Kang, J., & Banaji, M. (2006). Fair measures: A behavioral realist revision of affirmative action. *California Law Review, 94*, 1063–1118.
Klasen, S. (2006). *The efficiency of equity*. Discussion paper no. 145, Goettingen: Ibero-America Institute for Economic Research.
Kleindorfer, P., Kunreuther, H., & Shoemaker, P. (1993). *Decision sciences: An integrative perspective* (pp. 86–104). New York: Cambridge University Press.
Knoch, D., Pascual-Leone, A., Meyer, K., Treyer, V., & Fehr, E. (2006). Diminishing reciprocal fairness by disrupting the right prefrontal cortex. *Science, 314*, 829–832.
Koomey, J., et al. (2002). Sorry wrong number: The use and misuse of numerical facts in analysis and media reporting of energy issues. *Annual Review of Energy and the Environment*, *27*:119–158. See also, Best, J. (2001). *Damned lies and statistics: Untangling numbers from the media, politicians, and activists*. Berkeley: University of California Press.
Kurzweil, R. (2005). *The singularity is near: When humans transcend biology*. New York: Viking.
Lafuente-Rojo, A., Abascal-Gonzalez, J., & Cai, Y. (2007). Ambient intelligence: Chronicle of an announced technological revolution. *UPGRADE, 8*, 8–12.
Lakoff, G. (2002). *Moral politics: How liberals and conservatives think*. Chicago: University of Chicago Press. See also, Wilson, M. S. (2004). Values and political ideology: Rokeach's two-value model in a proportional representation environment. *New Zealand Journal of Psychology 23*.
Landes, W. M., & Posner, R. A. (2004). *The political economy of intellectual property law*. Washington: The AEI Press.
Le Grand, J. (1990). Equity versus efficiency: The elusive trade-off. *Ethics, 100*, 554–568.
Lejano, R. P. (2008). Technology and institutions: A critical appraisal of GIS in the planning domain. *Science, Technology and Human Values, 33*, 653–678.
Lemmens, T., & Thiery, Y. (2007). Insurance and human rights: What can Europe learn from Canadian anti-discrimination law? In H. Cousy & C. V. Schoubroeck (Eds.), *Discrimination in insurance*. Maklu: Academia-Bruylant.
Link, B., & Phelan, J. C. (2001). Conceptualizing stigma. *Annual Review of Sociology, 27*, 363–385.
Medina, H. (2004). *Consumer protection: Federal and state agencies face challenges in combating predatory lending*. Washington: United States General Accounting Agency. See also, Engel, K., & McCoy, P. (2002). A tale of three markets: The law and economics of predatory lending. *Texas Law Review*, *80*:1255–1367.
Moll, J., Oliveira-Souza, R., & Zahn, R. (2008). The neural basis of moral cognition: Sentiments, concepts, and values. *Annals of the New York Academy of Sciences, 1124*, 161–180.
Norman, P. (2003). Statistical discrimination and efficiency. *The Review of Economic Studies, 70*, 615–627.
Nosek, B., Greenwald, A., & Banaji, M. (2006). The implicit association test at age 7: A methodological and conceptual review. In J. A. Bargh (Ed.), *Social psychology and the unconscious: The automaticity of higher mental processes* (pp. 265–292). London: Psychology Press.
Okun, A. (1975). *Equality and efficiency: The big tradeoff*. Washington: The Brookings Institution.
Patrick, G. (2007). *The locus opus: Playing with privacy in a world of ambient intelligence*. Manuscript online at http://ssrn.com/abstract=1091116.
Pirttila, J., & Uusitalo, R. (2007). *Leaky bucket in the real world: Estimating inequality aversion using survey data*. Paper presented to the Conference on Public Sector Economics. Munich.
Quillian, L. (2006). New approaches to understanding racial prejudice and discrimination. *Annual Review of Sociology, 32*, 299–328.
Ramirez, S. A. (2004). Bearing the costs of racial inequality: Brown and the myth of the equality/efficiency trade-off. *Washburn Law Journal, 44*, 87–104.
Rawls, A. J. (1971). *Theory of justice*. Cambridge: Harvard University Press.
Regan, M. C., Jr. (2007). Moral intuitions and organizational culture. *Saint Louis University Law Journal, 51*, 941–988.
Rokeach, M. (1973). *The nature of human values*. New York: Free Press.

Rouvroy, A. (2008). Privacy, data protection, and the unprecedented challenges of ambient intelligence. *Studies in Ethics, Law, and Technology, 2*, 1–51.

Saguaro Seminar on Civic Engagement in America. (2000). *Better together*. Cambridge: John F. Kennedy School of Government, Harvard University.

Sampson, R., Morenoff, J., & Gannon-Rowley, T. (2002). Assessing 'neighborhood effects': Social processes and new directions in research. *Annual Review of Sociology, 28*, 443–478.

Sandefur, R. (2008). Access to civil justice and race, class and gender inequality. *Annual Review of Sociology, 34*(16), 1–20.

Sanfey, A., Rilling, J., Aronson, J., Nystrom, L., & Cohen, J. (2003). The neural basis of economic decision-making in the Ultimatum game. *Science, 300*, 1755–1758.

Schauer, F. F. (2003). *Profiles, probabilities and stereotypes*. Cambridge: Harvard University Press.

Schneider, J. D. (2004). *The psychology of stereotyping* (pp. 197–206). New York: The Guilford Press.

Schneier, B. (2003). *Beyond Fear: Thinking sensibly about security in an uncertain world*. New York: Copernicus Books. See also, Wall, D. S. (2007). *Cybercrime: The transformation of crime in the information age*. Cambridge: Polity.

Sen, A. (1992). *Inequality reexamined*. Cambridge: Harvard University Press.

Tenner, E. (1996). *Why things bite back: Technology and the revenge of unintended consequences*. New York: Vintage Books.

The Genetic Nondiscrimination Act (2008). Public Law No: 110–233, 21 May.

Thiery, Y., & Van Shoubroeck, C. (2006). Fairness and equality in insurance classification. *The Geneva Papers*, *31*:190–211. See also, Marmor, T., & Mashaw, J. (2006). Understanding social insurance: Fairness, affordability, and the 'modernization' of Social Security and Medicare. *Health Affairs*, *25*:114–134.

Van Hoyweghen, I., Horstman, K., & Schepers, R. (2005). 'Genetics is not the issue': Insurers on genetics and life insurance. *New Genetics and Society, 24*, 79–98.

Waldo, J., Lin, H., & Millett, L. (Eds.). (2007). *Engaging privacy and information technology in a digital age*. Washington: National Academies Press.

Weiss, M. J. (1988). *The clustering of America*. New York: Harper & Row.

Weiss, J. M. (2000). *The clustered world: How we live, what we buy, and what it all means about who we are*. Boston: Little, Brown & Co.

Whitby, A., Josang, A., & Indulska, J. (2004). *Filtering out unfair ratings in Bayesian reputation systems*, Proceedings of the autonomous agents and multi agent systems conference. New York.

Wilson, M. S. (2005). A social-value analysis of postmaterialism. *The Journal of Social Psychology, 145*(2), 209–224.

Wortham, L. (1986). Insurance classification: Too important to be left to the actuaries. *Journal of Law Reform, 19*, 349–423.

Yamamiya, Y., Cash, T., Melnyk, S., Posavac, H., & Posavac, S. (2005). Women's exposure to thin-and beautiful media images: Body image effects of media-ideal internalization and impact-reduction interventions. *Body Image*, *2*:74–80. See also, Johnson, P., McCreary, D., & Mills, J. (2007). Effects of exposure to objectified male and female media images on men's psychological well-being. *Psychology of Men and Masculinity*, *8*:95–102.

Zuberi, T. (2001). *Thicker than blood: How racial statistics lie*. Minneapolis: University of Minnesota Press, pp. 126–130. See also, Freedman, D. A. (1997). From association to causation via regression. In V. McKim & S. Turner (eds), *Causality in crisis? Statistical methods and the search for causal knowledge in the social sciences* (pp. 113–162). Notre Dame: University of Notre Dame Press.