POLITICAL ECONOMY RESEARCH INSTITUTE

University of Massachusetts Amherst

By the Content of their Character?

Discrimination, Social Identity, and Observed Distributions of Income

Paulo L. dos Santos and Noé Wiener

October 2018

WORKINGPAPER SERIES

Number 472

POLITICAL ECONOMY RESEARCH INSTITUTE

By the Content of their Character? Discrimination, Social Identity, and Observed Distributions of Income

Paulo L. dos Santos and Noé Wiener*

October 19, 2018

Abstract

This paper develops a series of information-theoretic measures to consider the systemic effects on individual incomes of complex patterns of social and economic discrimination by race, ethnicity, and gender, in the U.S. It derives coefficients of joint, conditional or incremental, and mutual information that offer non-parametric characterizations of the relative influence of economic and social-identity characteristics in the determination of individual income for different groups. It reports on estimates of those coefficients obtained using large-scale cross-sectional data from that economy. Those estimates support two sets of conclusions. First, the informational significance of social identity in the determination of incomes differs clearly and persistently across social-identity groups. For some groups social identity exerts a significant informational influence in the determination of income. Other groups enjoy greater scopes for individual differentiation by factors other than social identity. Second, the informational influence of educational attainment on income is deeply shaped by social identity. Among other expressions of this, the paper finds that some identity groups see the comparative measure of informational association between their incomes and educational attainment rise steadily with levels of educational attainment. In contrast, other groups see those comparative measures fall as educational attainment rises. These observations point to the economic effects systems of discrimination impose on certain groups, and to the relative privileges enjoyed by those not subjected to them.

Keywords: Information Theory, Income Distributions, Discrimination

JEL codes:

*Department of Economics, New School for Social Research; and Department of Economics, University of Massachusets, Amherst.

1 Introduction

Income inequality is a paramount form of social differentiation in decentralized market economies. It embodies differences in the measures of market command over social resources different individuals secure over a given time period. Distributions of income are thus an observable expression of a highly consequential social outcome in those economies: The market distribution of the social product across individuals and groups of individuals.

While most immediately determined in markets, individual incomes also express a broad range of irreducibly social relationships between groups of people.[1] This is clear in the overwhelming preponderance in the top ranges of national income distributions of individuals whose incomes primarily consist of profits (or other forms of propertied income drawing on profits), and the converse dominance of wages over everybody else's incomes. This attests to the ongoing social significance in contemporary capitalist economies of *class*—as understood in the salient contributions of Classical Political Economy.[2] It is even clearer in the continued absence of any income accruing to household reproductive labor, carried out overwhelmingly by women, which gives sharp expression to ongoing patterns of iniquity by gender in the contemporary organization of economic activity.[3]

Canonical microeconomic approaches to the distribution of income have emphasized the role of important contemporaneous individual characteristics in the market determination of income,[4] like the preferences and knowledge states of the parties involved, the stocks of financial and "human" capital individuals wish to engage productively, and the prices at which they succeed in doing so. But those current characteristics themselves are cumulative outcomes of social and economic processes.[5] Stocks of financial capital reflect past, often inter-generational wealth accumulation. Levels of education, skills, and work experience reflect much of the detail of personal, educational, and professional histories, as well as the accumulated results of much paid and unpaid labor performed by others during those histories. And complex social processes shape attitudes, expectations, and knowledge states conditioning all aspects of individuals' engagement with labor markets and its

[1] Marx, 1970; Polanyi, 2001.

[2] Smith, 1982; Ricardo, 1951; Marx, 1919.

[3] See Fraser, 1994; Folbre, 1994 and Wood, 2002, for instance.

[4] Debreu, 1959; G. Becker, 1957.

[5] Abbott, 1983; Abbott, 1995; Folbre, 2012; Cheng, 2014.

outcomes.

Patterns of social discrimination by elements of individuals' *social identity*—like their gender, race, ethnicity, or class background—exert an important influence on these processes, their outcomes, and their implications for income. Discriminatory treatment generally yields less favorable outcomes for women, minorities, and individuals with lower socio-economic status at each step along the cumulative processes shaping an individual's observable economic characteristics. It is also well understood that individuals with comparable observable economic characteristics are systematically treated differently in labor markets according to social identity.

As a result, individual incomes established in markets reflect more than just individual characteristics and contingencies capable of shaping economic trajectories and outcomes. They give observable and quantifiable expression to the functioning of social *systems* of discrimination. While this represents a potential boon for inquiry into the nature of social discrimination, drawing on those distributions to make detailed inferences about the influence of social identity on income poses a number of difficulties. The interrelationships in question are dynamic and complex. They involve non-linearities, path dependences, as well as the influence of unobservable characteristics, social interactions and processes. Efforts to draw on observation to estimate parameters in strongly specified models of particular mechanisms or specific influences involved are unlikely to yield successful characterizations.

In contrast, comparisons of distributions of individual income, observable economic characteristics, and observable elements of social identity can cast light onto the aggregate or systemic economic manifestations and effects of social discrimination. In the absence of economically meaningful patterns of discrimination by social identity, marginal distributions of income in a market economy would not differ systematically across observable social identity groups. Those distributions would be shaped in the same manner by the same sets of individual characteristics, social processes, vagaries and contingencies influencing individual incomes.[6] Similarly, in the absence of discrimination joint distributions of income and observable economic characteristics would also be comparable across social-identity groups. Observable economic characteristics would have compa-

[6] Note that this could include additional, unobservable patterns of discrimination by socially relevant individual characteristics that may have little to do with productivity.

rable aggregate effects on income for all such groups.

It is in this connection that the present paper makes a set of innovative methodological and empirical contributions. It draws on Information Theory to develop a series of indices of joint, conditional or incremental, and mutual informational association between sets of individual economic and social-identity characteristics and individual income. Those indices offer non-parametric, purely informational characterizations of the influence and interactions between these sets of characteristics in the determination of individual income. Their development leads to three general formal results that can assist inquiry into influences and interactions involving two sets of variables in the determination of a quantity of interest.

The paper applies these indices and results to large-scale cross-sectional data sets constructed from the U.S. census to cast light on the emergent influence of race, ethnicity, and gender on aggregate patterns of income in that economy.[7] This results in a series of new or newly framed conclusions about the economic effects of social patterns of discrimination. It also illustrates the usefulness of information-theoretic inquiry in seeking to make inferences about the functioning of complex socio-economic systems from observable outcomes.[8] Most broadly, the paper establishes that the pattern of differentiation among individuals and its association with observable economic characteristics differs systematically across social-identity groups. Two significant informational features of these differences stand out.

First, there are persistent differences between social-identity groups in the heterogeneity or diversity of income, as measured by the *entropy* of their distributions of individuals across income levels. For instance, data for 2010 show almost 15 percent less diversity in the incomes of Hispanic women than in the incomes of all people sampled. Put differently, observing that a person is a Hispanic woman is highly informative, in that it reduces uncertainty about her income by almost 15 percent. In contrast, the distribution for white men had five percent *more* diversity than the entire sample distribution. We know less about a given individual's income after we learn they are a white man.

This is a significant finding. It suggests some identity groups enjoy not only comparatively

[7]Unfortunately this rich and large dataset does not capture any reliable measure of class background.

[8]As such this paper builds on a tradition started by Theil, 1971 and Theil and Finezza, 1971.

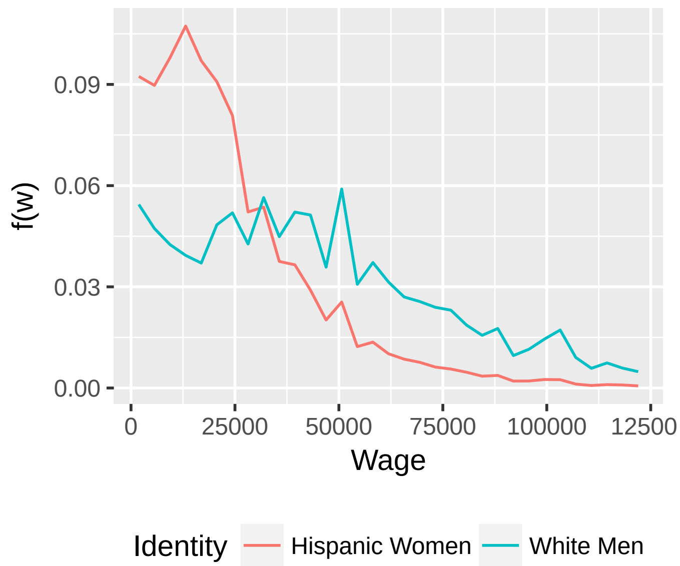

Figure 1: Identity and annual wage data for white men and hispanic women in 2010 ACS.

higher average incomes, but also greater scopes for income differentiation by characteristics other than their social identity. Others face narrower ranges of economic outcomes, as can be seen in Figure 1 and Figure 2. This is true not only for unconditional distributions of income across social-identity groups, but also for distributions of income across identity groups with the same observable economic characteristics.

The paper shows formally how differences in entropy across income distributions ensure that the comparative, incremental informational significance of social identity relative to that of *any* set of economic characteristic in the determination of income will *always* be greater among women and minorities than among white men. The extent to which the complex, dynamic socio-economic processes shaping levels of individual income effectively treat individuals on the basis of their social identity is systematically greater for some social-identity groups. Discrimination by social identity involves a comparative failure to discriminate among individuals belonging to certain identity groups on the basis of the content of their individual economic characteristics.

Second, the paper establishes important differences in the informational association educational achievement has with incomes across different identity groups. Certain combinations of educational achievement and social identity have "informational synergies" in their associations with individual income. “Being white and having a college degree” is more informative about a person’s income than the sum of the separate informational associations of “being white” and “being a college graduate” with individual incomes. Significantly, this is not the case for other university-educated groups. Conversely, being a woman and having a low level of formal education is more informative about a person’s income than the sum of the informational associations of being a woman and that of having a low level of formal education. This effect does not apply to men with low levels of formal educational achievement.

In both cases this suggests the combination of characteristics in question is statistically associated with further characteristics and processes that shape the effect of educational attainment on distributions of income. Those processes ensure high levels of educational attainment are unusually informative of incomes for whites, and low levels of educational attainment unusually informative of incomes for women. In fact, the paper shows how the comparative informational significance of educational attainment for the income of women consistently *decreases* as levels of education increase. In contrast, that comparative informational significance consistently *increases* as levels of education increase for white men. Returns on education are part of broader social processes shaped by gender, race, and ethnicity in ways that effectively generate greater rewards for high educational achievement for whites and impose greater losses on low levels of formal education on women. They appear to embody forms of discrimination by social identity, with important implications for our understanding of how individuals and societies can strive to make progress against discrimination and its economic consequences.

The paper is organized as follows. Section two offers a discussion on the complex, dynamic relationship between social identity and income and motivates the need for a *systemic* approach to its study. Section three discusses fundamental information theoretic concepts and their applications to analysis of complex socio-economic systems. Section four derives indices of joint, incremental, and mutual information and establishes a few formal results involving them that can be used in such applications. Section five reports on estimates of those indices obtained using successive cross-sections of U.S. data, and discusses their significance. Section six concludes the paper with a discussion on the understanding of the economic effects of discrimination by elements of social

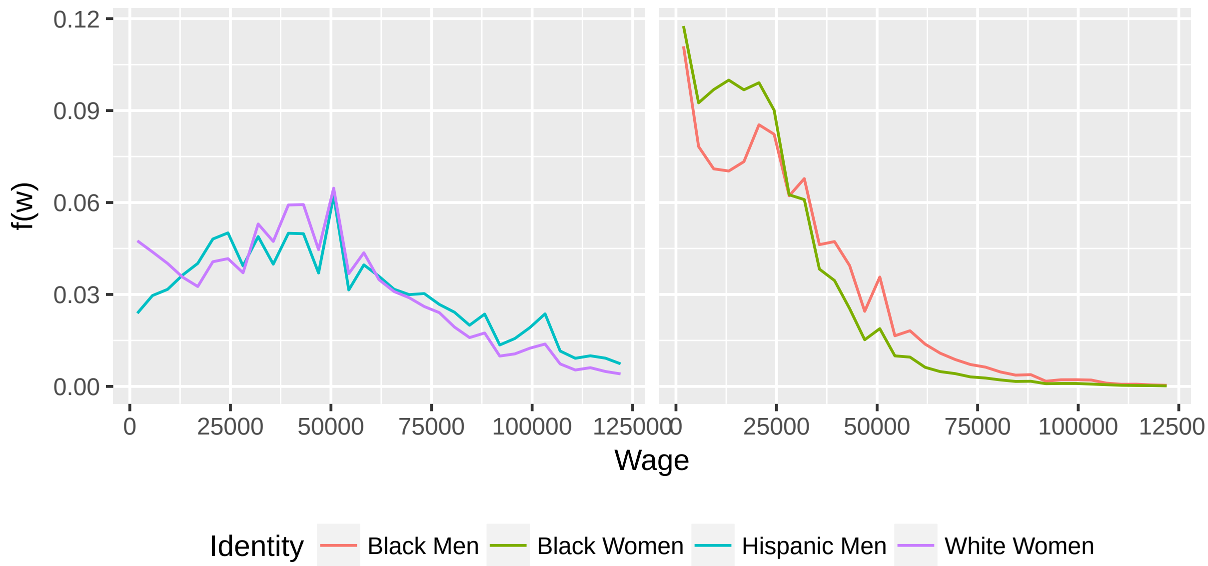

Figure 2: Marginal wage distributions by identity, 2010 ACS

identity, and of the kind of privilege some groups enjoy, suggested by the perspective offered in this paper.

2 Social Identity and Income

The influence of social identity on income is well established. Many studies have considered its effect on a broad range of labor-market outcomes among individuals with the same observed economic characteristics. Experimental contributions from sociologists and economists have established patterns of discrimination against women and minorities in hiring across several advanced economies by considering the comparative fortunes of controlled pairs of fictitious job applicants.[9] Other contributions have established that women and minorities are over (under) represented in occupations with lower (higher) average measures of pay, even after accounting for educational attainment.[10] They also tend to receive lower incomes than white, male peers performing the same jobs.[11]

[9] See Quillian et al. (2017), which reports on a large-scale meta-analysis of several such studies undertaken in the U.S. has estimated that white job applicants enjoy on average 36 and 24 percent more callbacks than otherwise comparable black and Latino applicants, respectively. See also Daniel, 1968; Jowell and Wissoker, 1970; Firth, 1981; Firth, 1982; Riach and Rich, 1987; Kenney and Wissoker, 1994; M Bendick and Reinoso, 1994; Riach and Rich, 2002.

[10] Hamilton et al., 2011; Kossek et al., 2016, etc.

[11] Joshl and Paci, 2001

Contemporary economic theory has offered two broad approaches for thinking about the influence of discrimination by social identity on incomes and other market outcomes. It is possible to consider that individual consumers, employers and workers have preferences that are biased against certain social groups; so much so that they are willing to act on those biases even in the face of pecuniary losses.[12] As a result, workers belonging to those groups are employed less often and typically at lower wages than other workers. It is also possible to consider that employers carry out "statistical discrimination," relying on perceived associations between elements of social identity and unobservable productive characteristics when making hiring decisions.[13] Both approaches are capable of drawing out a number of undesirable economic and odious ethical consequences of discrimination—including social and individual economic losses consequent to biases and stereotypes. And the idea of "statistical discrimination" gets to the very substance of social *pre-judice*: The judgement of individuals not by the content of their character, but by prior and incorrect beliefs about associations between observable markers of social identity and individual economic characteristics.

Despite this, the methodological choice of taking individual preferences and beliefs as given leaves no scope for explicit consideration of the complex, dynamic interrelationships between the attitudes, expectations, and actions that define discrimination by social identity, the evolving economic characteristics of individuals, and their employment and income outcomes. This is a significant shortcoming in any attempt to develop empirically grounded characterizations of the influence of social identity on aggregate patterns of income distribution.

This section briefly discusses the growing body of evidence documenting the impact of discrimination on a broad range of outcomes shaping individual incomes. It also discusses important methodological implications of these findings, including on why *systemic* conceptualizations of the economic effects of patterns of discrimination are far more appropriate than individualist, micro-kinetic approaches as bases for observational work.

[12] G. Becker, 1957; Arrow, 1971.

[13] Phelps, 1972

2.1 Identities and the Social Construction of Economic Characteristics

An individual's present income is an outcome of a succession of social and economic processes taking place over their lifetime.[14] Those processes condition economic characteristics like education, skills, experience, and preferences that shape that individual's present engagement with labor markets. A growing body of contributions from Psychology, Epidemiology, Pedagogy, Sociology, Development Studies, Anthropology, and Political Economy have established how at every step along the way, patterns of discrimination by socio-economic status and social identity shape individual economic outcomes. As a result, the distribution of present economic characteristics differs across socio-economic status and social identity, reinforcing—in turn—discriminatory social attitudes, expectations, and interactions.

From infancy, parents may effectively favor boys in the allocation of household resources affecting future labor-market outcomes, like food, parental attention, and educational expenditures.[15] The same may be true for children with phenotypical traits associated with more favorable racial classifications in a given society.[16] Favored children generally benefit from preferential treatment—as well as from the broader social advantages that motivate it—and achieve better average outcomes, giving real measures of self-fulfilling, social validation to parental discrimination.

In formal education, a growing literature points to the presence and academic consequences of teacher discrimination of students, in some cases as early as pre-school.[17] Patterns of low academic expectations by educators and students resulting in low educational achievement continue through all levels of education, including college.[18] Experiences with discrimination by teachers and peers can increase measures of physical stress suffered by young students and impair their cognitive development.[19] Minority students also have disproportionately high rates of poverty, which is often associated with poor nutrition and adverse health outcomes that have been linked to poor

[14] Abbott, 1983; Abbott, 1995; Cheng, 2014.

[15] See Behrman et al., 1986; Rosenzweig and Shultz, 1982, for instance.

[16] Rangel, 2015.

[17] See Gilliam et al., 2016; Gershenson and Papageorge, 2018; Rosenbloom and Way, 2004; Greene et al., 2006; Wong et al., 2003, for instance.

[18] Thomas, 2017.

[19] See Levy et al., 2016; Brown, 2015.

educational achievement and lower measures of adult productivity and income.[20]

Beyond education, processes of socialization reproduce differential expectations, preferences, and attitudes across social identities, with significant effects on labor-market outcomes. Socially accepted gender roles define different ranges of behavior deemed acceptable or appropriate, including behaviors associated with career success and higher earnings.[21] Work and career preferences, as well as attitudes toward competing family and professional demands, are deeply shaped by traditional gendered expectations and beliefs concerning what jobs and functions are appropriate for men and women.[22] The economic significance of gendered work and career expectations will be greatest in settings where wage differentials between "women's work" and other jobs are large, and in settings where good-quality child and elder care services and paid family leave are not widely accessible.

Conventional beliefs about the capacities and characteristics of individuals across different social identities shape both the kinds of jobs individuals seek and the kinds of jobs they get. Differential patterns of occupancy across jobs help perpetuate those beliefs, as they create role models for both wage earners and employers.[23] They also perpetuate iniquities as some groups are underrepresented in the kinds of jobs that create more opportunities for further career advancement.[24]

Finally, a substantial literature has established important patterns of iniquity and discrimination of "class migrants" in professional careers.[25] Despite wide evidence that people from working-class backgrounds often have unique skills and ability their peers from more privileged backgrounds may lack, they still report that their background adversely affects their career outcomes. They themselves report lower levels of belonging, and are less often seen as a "good fit" with a particular workplace as a result of their ignorance of tacit class rules and conventions.

The effects of such multifaceted patterns of discrimination on economic outcomes are profound.

[20] See Darmon and Drewnowski, 2008; Drewnowski and Darmon, 2005; Taveras et al., 2010; Larson and Story, 2011; Strauss and Thomas, 1998; Hoddinott et al., 2008; Bhargava, 2008, for instance.

[21] Udry, 1994, etc.

[22] See Kossek, 2006; Martins et al., 2002; Lavassani and Movahedi, 2014; Lawson et al., 2015; Kossek et al., 2016, for instance.

[23] See Hilton and W. von Hippel, 2012; C. von Hippel et al., 2015; Eagly and Karau, 2002; Heilman, 2012, for instance.

[24] Kanter, 1977.

[25] See the succinct summary of these contributions offered by Williams et al., 2018.

Even the manner in which cognitive and non-cognitive traits at childhood and adolescence shape future educational achievement and income is fundamentally defined by a person's background and social identity. The personality traits associated with higher rates of college completion are very different across student socio-economic status.[26] Among other characteristics, adult entrepreneurs have been found to have engaged in illicit activities in adolescence at relatively high rates.[27] As the adverse consequences for illicit behavior differ drastically across social identities,[28] the opportunity to put one's risk-seeking appetites to potentially productive uses is unequally distributed across potential entrepreneurs. Finally, associations between childhood externalizing behaviors, like aggression, a willingness to disobey rules, cheat, and steal,[29] and higher adult incomes have been found in British data for individuals with high childhood socio-economic status,[30] but not for others. Notably, *reductions* in childhood externalizing behaviors have also been found to be associated with increases in future income among disadvantaged black children in the United States.[31] Some personality traits that can help some people achieve better economic outcomes have the opposite effects for members of other socio-economic groups.

2.2 Methodological Implications

The findings of these and related contributions strongly suggest that economic discrimination by elements of social identity is not reducible to individual preferences or beliefs about unobservable characteristics of workers. Neither is it simply the heterogeneous treatment of economically equivalent individuals. It reflects the operation of *social systems* defined by self-reinforcing processes taking place both within and beyond markets. Economic outcomes and social attitudes, expectations, and actions feed back on each other, systemically reproducing patterns of iniquity by social identity.[32]

[26] Lundberg, 2013.

[27] Levine and Rubinstein, 2013.

[28] Not only has the criminal justice system especially targeted types of offences "for which black and Hispanic people often are disproportionately arrested and convicted" (Council, 2014, , p. 91), racial disparities in the severity of sentencing have also been found as the cumulative effect of "small but systematic racial differences in case processing" (Council, 2014, , p. 103).

[29] Liu, 2004.

[30] Papageorge et al., 2017.

[31] Heckman and Kautz, 2013.

[32] See Folbre, 2012 others on systems of oppression.

Within those processes, the role of social identity in conditioning incomes is multifaceted, dynamic, and complex. This has important methodological implications.

Individualist approaches, which predicate labor-market outcomes on detailed descriptions of *given* individual characteristics and the micro-level interactions they define, face both practical and conceptual difficulties in analysis of this role. Individual characteristics are ever-changing moments in a broader set of complex processes of social determination. There is no reason to expect parametrizations that take those characteristics as given will be particularly advantageous in inquiry into the full, independent effect social identities have on income. In fact, such approaches are assured to flounder in attempts to give them empirical foundations. Many of the parameters involved relate to unobservables, like subjective states. And the frequencies at which individual states may be observed are generally far lower than the frequencies at which interactions change individual economic states, and possibly individuals themselves. What can be observed are not inherently individual characteristics, but the accumulation of the outcomes of a series of unobservable social interactions.[33] To use such observation to infer parameters in models of individual behavior is to make a category mistake.

Conceptually, individualist parametrizations take as given patterns of heterogeneity and iniquity in current individual economic characteristics and behaviors that have been conditioned by previous social and economic interactions.[34] This leaves no space for explicit consideration of the processes conditioning these heterogeneities, generally resulting in an underestimation of the full measure and significance of elements of social identity in the determination of the distribution of social output.

An important case in point is the influential approach to the determination of individual incomes defined by the "human capital" framework articulated by Schultz, 1961 and G. S. Becker,

[33] See dos Santos, 2017.

[34] Economics has a long record of starting analysis by taking present states that reflect dynamic social and economic iniquities as given. This was already evident in Ricardo, 1951's advocacy of free trade base on the concept of comparative advantage—the ability to reallocate domestic resources to produce more of one good while sacrificing fewer units of other goods than other countries. In Ricardo's celebrated example, trade between Portugal and England is taken as advantageous to both countries since Portugal has a comparative advantage in producing wine, while England's advantage lies in production of cloth. The choice of "wine" for Portugal in Ricardo's thought exercise was not innocuous. It evokes inherent, climatological differences between the two countries. But cloth is different—its production in England reflected processes of mechanisation that embodied and opened possibilities for further gains in absolute and comparative advantages. Ricardo's analytical choice effectively disappeared the economic *privileges* enjoyed by the English economy in international trade.

1962. That framework, which most broadly characterizes an individual's income as a yield on their investment in skills, education, and experience, has inspired a voluminous literature seeking to account for observed individual incomes as the result of individual "human capital" investment choices.[35]

If "human capital" is simply the discounted present or capitalized value of all future earnings, the theory can become expansive to the point of being vacuous: Any distribution of wages can be attributed to a particular distribution of unobserved human capital. Seen from the perspective afforded by the contributions discussed above, the resulting approaches underemphasize important systemic, social determinants and costs of education, skills, and experience, while exaggerating the influence of "individual investment choice" in determining individual productive characteristics.

An alternative, systemic conceptualization of the economic effects of social identity can enable practical work without these difficulties. It is possible to consider that through the kinds of complex, dynamic processes discussed above, socio-economic systems reproduce macroscopic patterns of heterogeneity in economic outcomes across large social-identity groups. Those patterns should be understood as exclusively social, in that they have no foundation outside of those systems.[36] They are the systemic economic expression of discrimination by social identity. Their formal characteristics contain information about emergent, macroscopic associations between social identity and different observable economic characteristics and outcomes. Information Theory makes it possible to characterize and quantify those associations, casting new light onto the systemic expressions of discrimination in the processes through which a political economy distributes the social product.

[35]See the influential contribution by Mincer, 1974 and the synthetic review in Ashenfelter et al., 1999.

[36]Specifically, race and ethnicity are social categories with no biological bases. The genetic variability across the different sets of distinctive human populations that constitute various racial and ethnic categories are very small compared to the overall genetic variability across humanity as a whole (See Yu et al., 2002; Witherspoon et al., 2007, for instance). Sex is obviously a biological category, which very likely conditions important elements of behavior in our species (Udry, 1994). But the *economic* consequences of sex-dimorphic behaviors are expressions of gender, and reflect how the economic organization of a society generally favors one set of its members over another. There is no *a priori* reason why market incomes should be allocated differently according to sex.

3 Drawing on Information Theory

Information Theory offers distinctively useful tools enabling formal inferences about complex patterns of economic and social interaction based on their observable outcomes. This section discusses two central information theoretic concepts that can help guide observational inquiry into the associations between economic and social characteristics of individuals and their income: entropy and mutual information.

To see this, consider an economic or social system as composed of a large number N of individual members. At any given point in time, each of those members has an individual state defined over a set of $k+1$ degrees of freedom, $\mathbb{X} = \{X_0, X_1, ..., X_k\}$. Individual degrees of freedom may describe quantifiable individual characteristics as well as macroscopic quantities that take the same value across a large number of individuals in the system. They may also describe qualitative or categorical individual characteristics. Coding schemes mapping the latter characteristics onto distinct real numbers allow individual states to be represented by a vector $\mathbf{x} = \{x_0, x_1, ..., x_k\}$, with the set of all such individual states denoted by $T \subseteq \mathbb{R}^{k+1}$.

The macroscopic state of the system can be defined over a "coarse-graining" of T into s bins, as a frequency function $f(x_0, x_1, ..., x_k) = f(\mathbf{x})$ describing the normalized occupancy of each individual-state bin.[37] The functioning of a system defines a phase space Φ containing all macroscopic states $f(\mathbf{x})$ the system may actually occupy. The laws and regularities that define a system are given statistical expression in the shape of Φ.

In observational social science we typically face variations of the following analytical problem within this setting: We can observe the values taken by $v \leq k+1$ individual degrees of freedom across $n < N$ members of the system. This allows construction of frequency histograms $f(\mathbf{x_v})$ over the values taken by the vector $\mathbf{x_v}$ of observed individual states. We have limited knowledge of the micro-level interactions driving the functioning of the system, as those are either unobservable or involve non-linearities and high-frequency interactions that make it very difficult to draw on

[37] Note that coarse graining of the domain for a quantitative individual degree of freedom can be understood as a distinctive type of categorical coding: one where the quantity associated with the category defined by the occupancy of a bin—typically the central value taken by the degree of freedom in the bin—is relevant to the functioning of the system.

observation to characterize them.[38] And we generally do not know the full set $\mathbb{X}$ of relevant degrees of freedom. But we would like to draw on what we observe to infer as much as we can about the functioning of the social or economic system at hand. Formally, we want to develop increasingly accurate descriptions of the shape of Φ.

In this connection the concept of *entropy* is distinctively useful. The entropy $H(\mathbb{Y})$ for any set of degrees of freedom $\mathbb{Y}$ in a state $\mathbf{f}$, defined over m bins, is a measure of the statistical weight W_f of that macroscopic state across all possible micro-level configurations of individual members across all bins. Formally,

$$H(\mathbb{Y}) = \frac{\log W_f}{N} = -\sum_{i=1}^{m} f_i \log f_i \tag{1}$$

Note that this quantity can also be understood as a measure of the diversity or heterogeneity in the values taken by $\mathbb{Y}$. If all individuals are in the same bin, 1 ensures entropy is zero. If individuals are evenly distributed across all m bins—a state of maximum diversity or heterogeneity—entropy reaches its maximum value: $\log m$. It should be obvious that a change in the state of a single individual results in an increase in entropy if and only if the change takes that individual to a state with lower occupancy than the state it originally occupied. That is, entropy increases only when diversity or heterogeneity increases.[39]

Entropy is useful in analysis of systems with large $N >> m$ for at least two reasons. First, for those systems the combinatorial dominance of the distribution f^* achieving maximum entropy over their phase spaces is overwhelming. If we know the shape of the phase space for such a system, we should generally expect to observe macroscopic behavior in line with a state f^*.

A converse application of this observation is particularly useful in observational work in political economy.[40] Sometimes measured frequencies $f(\mathbf{x_v})$ over the system's v observable individual degrees of freedom are consistently well described by known, closed-form functional forms. Those

[38] See dos Santos, 2017.

[39] For a formal proof of this observation for changes in any number of individual states, see the Appendix in dos Santos and Scharfenaker, 2018.

[40] See Stanley et al., 1996; Bottazzi and Secchi, 2003; Bottazzi and Secchi, 2006; Alfarano and Milaković, 2008; Alfarano and Milaković, 2008; Scharfenaker and dos Santos, 2015; dos Santos and Scharfenaker, 2018, for instance.

functional forms are often entropy maxima over phase spaces defined by known moment constraints on the distribution of $\mathbf{x_v}$. It can be inferred that those constraints are good descriptions of the observable section of the system's phase space, $\Phi_v \subset \Phi$. All interactions and influences involving observed and non-observed degrees of freedom effectively resolve themselves into those constraints. They are the emergent, systemic expression of the outcome of those interactions and influences.[41] Observationally successful theories of those processes must be formally equivalent to those constraints, which give us formal clues about the macroscopic or *social* content of the micro-processes at hand.

3.1 Uncertainty and Mutual Information

Entropy is also useful in settings where observed distributions are not well described by known, closed-form functional forms. Entropy is a measure of the uncertainty we have about the exact micro-level configuration of a system at a macroscopic state $f(\mathbf{x})$. Depending on the base of the logarithm used in 1, entropy measures the average number of bits, nats, or dits necessary to enumerate all W_f micro-level configurations in which the system may find itself at that state.

This measure of what we do not know about the micro state of a system motivates the concept of *mutual information*. The mutual information between two degrees of freedom is given by,

$$I(X_i, X_j) = H(X_i) - H(X_i|X_j) = H(X_i) - \sum_{x_j} f(x_j) H(X_i|x_j) \tag{2}$$

This is a quantification of the average reduction in our uncertainty about X_i when we observe the distribution of X_j: The change in the average number of bits, nats, or dits needed to enumerate or identify uniquely each micro-level configuration compatible with observation when move from observing only X_i to observing X_i and X_j. Mutual information can also be thought of as a measure of the information shared between the two quantities, in that it quantifies how much we learn about

[41] Most generally, we are interested in settings where observed distributions are consistently well described by Lambert-W functions, which include all maxima for generalized, (c, d) entropy functionals. In cases where $(c, d) \neq (1, 1)$ we have non-Shannon entropies and systems where some interactions can be most conveniently understood to resolve themselves not into macroscopic constraints we may infer but into non-extensive dependences of the system's phase-space volume on the number of members in the system. See Hanel and Thurner, 2011.

one of them from observation of the other.

The multivariate generalization of mutual information requires careful consideration. As motivated by McGill, 1954; Fano, 1961; Han, 1980, note that $I(X_i, X_j) = I(X_i) - I(X_i|X_j)$, where the self mutual information $I(Y) = H(Y)$. By extension,

$$I(X_0, X_1, X_2) = I(X_0, X_1) - I(X_0, X_1|X_2) \tag{3}$$

where the *conditional mutual information* $I(X_0, X_1|X_2) = H(X_0|X_2) - H(X_0|X_1, X_2)$ measures the information gained about X_0 upon observation of X_1 when X_2 is already known. The tripartite mutual information in 3 is a measure of the information shared by all three variables: the information shared by X_0 and X_1, minus the part of that shared information not contained in X_2.

The general multivariate mutual information can be defined recursively,

$$I(X_0, X_1, ..., X_k) = I(X_0, X_1, ..., X_{k-1}) - I(X_0, X_1, ..., X_{k-1}|X_k) \tag{4}$$

The mutual information between all $k+1$ variables measures the shared informational content of the first k variables minus the part of that content not contained in X_k.[42]

3.2 Joint Mutual Information

In inquiry into the functioning of economic and social systems, a different measure of informational association may be more useful. We are often interested in learning not about the informational content shared among a number of variables but in how much of the uncertainty in a single degree of freedom X_0 is removed when we observe values taken by a set $\mathbb{X}_k = \{X_1, ..., X_k\}$ of other degrees of freedom. Put differently, we are often interested in the *informational account* of X_0 given by the elements in $\mathbb{X}_k$: How much do we know about individual values x_0 taken by X_0 based on knowledge or observation of $\chi_k = \{x_1, ..., x_k\}$.

This may come up as part of general inquiry into the dynamic co-determinations between all

[42]The concepts outlined so far have been recently used to guide inquiry into patterns of segregation across occupation and levels of education on the basis of gender. See Puyenbroeck et al., 2012.

these variables. It may also come up in settings where we know X_0 has no or very weak causal influence over members of the set $\mathbb{X}_n$, in which case the informational equivalence can be taken as a measure of the extent to which the latter degrees of freedom combine to determine values of X_0.

To characterize this kind of informational accounting, a measure of *joint mutual information* is more useful.[43] Defining it first for a setting with three degrees of freedom, consider,

$$I\left(X_0; \left(X_1, X_2\right)\right) = H\left(X_0\right) - H\left(X_0 | \left(X_1, X_2\right)\right) \tag{5}$$

Which measures the reduction in uncertainty about values of X_0 once values of X_1 and X_2 are taken into account. The relationship between this measure and the conditional mutual information can be easily established. Adding $H\left(X_0|X_1\right) - H\left(X_0|X_1\right) = 0$ to this definitions yields,

$$\begin{aligned} I\left(X_0; \left(X_1, X_2\right)\right) &= \left(H\left(X_0\right) - H\left(X_0|X_1\right)\right) - \left(H\left(X_0|X_1\right) - H\left(X_0| \left(X_1, X_2\right)\right)\right) \\ &= I\left(X_0, X_1\right) + I\left(X_0, X_2|X_1\right) \end{aligned} \tag{6}$$

The joint mutual information between X_0 and (X_1, X_2) is the sum of the mutual information between X_0 and X_1 and a conditional mutual information—the information gained about X_0 upon observation of X_2 when X_1 is already known. This results in a measure of the total reduction in uncertainty about X_0 arising from joint observation of X_1 and X_2.

The multivariate generalization of this measure for X_0 and a set $\mathbb{X}_k$ of n other degrees of freedom that may take individual values $\chi_n = \{x_1, x_2, ..., x_k\}$ may also be defined recursively,

$$\begin{aligned} I\left(X_0, \mathbb{X}_k\right) &= H\left(X_0\right) - H\left(X_0|\mathbb{X}_k\right) \\ &= \left(H\left(X_0\right) - H\left(X_0|\mathbb{X}_{k-1}\right)\right) + \left(H\left(X_0|\mathbb{X}_{k-1}\right) - H\left(X_0|\mathbb{X}_k\right)\right) \\ &= I\left(X_0, \mathbb{X}_{k-1}\right) + I\left(X_0, X_k|\mathbb{X}_{k-1}\right) \end{aligned} \tag{7}$$

In many applications in economic and social inquiry, we are interested in a more general decomposition, separating the variables in $\mathbb{X}_k$ into two mutually exclusive sets, $\mathbb{X}_e$ containing e of the

[43]Measures equivalent to this have been motivated and used in computational biology and in work on model feature selection. See Yang and Moody, 1999; Bennasar et al., 2015, and Ince, 2017, for instance.

k individual degrees of freedom, and its complement in $\mathbb{X}_k$, $\mathbb{X}_i$, containing the remaining $i = k - e$ ones,

$$\begin{aligned} I(X_0; \mathbb{X}_k) = (H(X_0) - H(X_0|\mathbb{X}_e)) + (H(X_0|\mathbb{X}_e) - H(X_0|\mathbb{X}_k)) \\ = I(X_0; \mathbb{X}_e) + I(X_0; \mathbb{X}_i|\mathbb{X}_e) \end{aligned} \tag{8}$$

The total, joint informational association of the degrees of freedom in $\mathbb{X}_k$ and the variable of interest X_0 is given by the joint mutual information between the latter and the variables in the set $\mathbb{X}_e$ plus the incremental information gained about X_0 upon observation of $\mathbb{X}_i$ when $\mathbb{X}_e$ is already known.

Since the degrees of freedom in each two sets $\mathbb{X}_e$ and $\mathbb{X}_i$ are being considered jointly, it is also possible to consider the tripartite mutual information,

$$I(X_0, \mathbb{X}_e, \mathbb{X}_i) = I(X_0, \mathbb{X}_e) - I(X_0, \mathbb{X}_e|\, \mathbb{X}_i) = I(X_0, \mathbb{X}_e) + I(X_0, \mathbb{X}_i) - I(X_0, \mathbb{X}_k) \tag{9}$$

The joint mutual information in 8 can be decomposed into the two measures of conditional or incremental mutual information defined by $\mathbb{X}_e$ and $\mathbb{X}_i$ and the mutual information between X_0 and the two sets.

$$I(X_0; \mathbb{X}_k) = I(X_0; \mathbb{X}_e|\mathbb{X}_i) + I(X_0; \mathbb{X}_i|\mathbb{X}_e) + I(X_0, \mathbb{X}_e, \mathbb{X}_i) \tag{10}$$

There are many settings where we are interested in this decomposition of the total association between a quantity of interest and two sets of variables into the incremental association of each set of variables, and their indistinguishable, joint association with the quantity of interest. Income and sets of economic and social-identity characteristics are a notable and important instance.

3.3 Coefficients of Informational Association

Normalized versions of the measures of informational association defined above offer useful quantifications. Consider first the coefficient of joint mutual information between X_0 and $\mathbb{X}_k$, which measures the extent to which the former is informationally equivalent to the latter,

$$A(X_0||\mathbb{X}_k) \equiv \frac{I(X_0, \mathbb{X}_k)}{H(X_0)} = 1 - \frac{H(X_0|\mathbb{X}_k)}{H(X_0)} \tag{11}$$

It should be obvious that $A(X_0||\mathbb{X}_k) \in [0,1]$, with $A(X_0||\mathbb{X}_k) = 1$ only when the informational account of X_0 provided by $\mathbb{X}_k$ is deterministic, leaving no uncertainty about all individual measures x_0 when the corresponding values χ_k taken by $\mathbb{X}_k$ are known. We may term a degree of freedom X_i in an account provided by $\mathbb{X}_k$ *independent* if $A(X_i||X_j) = 0$, $\forall X_j \in \mathbb{X}_k, j \neq i$. An account may be termed *orthogonal* if all the degrees of freedom involved are independent.

There should be no expectation that analysis of complex social systems can even approximately result in deterministic or orthogonal accounts. But in social inquiry, we can often make some progress toward understanding the influences on a degree of freedom X_0 by considering measures of its incremental and mutual informational association with two mutually exclusive subsets of $\mathbb{X}_k$, $\mathbb{X}_e$ and $\mathbb{X}_i$,

$$\mathbb{I}_{\mathbb{X}_e|\mathbb{X}_i} \equiv \frac{I(X_0; \mathbb{X}_e|\mathbb{X}_i)}{H(X_0)}; \quad \mathbb{I}_{\mathbb{X}_i|\mathbb{X}_e} \equiv \frac{I(X_0; \mathbb{X}_i|\mathbb{X}_e)}{H(X_0)}; \quad \mathbb{M}(X_0, \mathbb{X}_e, \mathbb{X}_i) \equiv \frac{I(X_0, \mathbb{X}_e, \mathbb{X}_i)}{H(X_0)} \tag{12}$$

These conventions permit a number of different ways to express the decomposition of $A(X_0||\mathbb{X}_k)$,

$$\begin{aligned} A(X_0||\mathbb{X}_k) &= A(X_0||\mathbb{X}_e) + (1 - A(X_0||\mathbb{X}_e))\, A(X_0|\mathbb{X}_e||\mathbb{X}_i) \\ &= A(X_0||\mathbb{X}_e) + \mathbb{I}_{\mathbb{X}_i|\mathbb{X}_e} = A(X_0||\mathbb{X}_i) + \mathbb{I}_{\mathbb{X}_e|\mathbb{X}_i} \\ &= \mathbb{I}_{\mathbb{X}_i|\mathbb{X}_e} + \mathbb{I}_{\mathbb{X}_e|\mathbb{X}_i} + \mathbb{M}(X_0, \mathbb{X}_e, \mathbb{X}_i) \end{aligned} \tag{13}$$

The total proportional reduction in uncertainty about X_0 can be divided into the coefficient of unconditional informational association between X_0 and one of the two sets of individual degrees of freedom, and the coefficient of incremental informational association between X_0 and the other set of individual degrees of freedom. It can also be expressed as a sum of the two coefficients of incremental informational association, minus the coefficient of mutual informational association between X_0, $\mathbb{X}_e$, and $\mathbb{X}_i$.

The sign of $\mathbb{M}\left(X_0, \mathbb{X}_e, \mathbb{X}_i\right)$ reveals an important informational relationship between these degrees of freedom. Formally, $\mathbb{M}\left(X_0, \mathbb{X}_e, \mathbb{X}_i\right) < 0 \Leftrightarrow \mathbb{I}_{\mathbb{X}_i|\mathbb{X}_e} > A\left(X_0||\mathbb{X}_i\right) \Leftrightarrow \mathbb{I}_{\mathbb{X}_e|\mathbb{X}_i} > A\left(X_0||\mathbb{X}_e\right)$. That is, the coefficient of mutual information will be negative in any and all cases of "informational synergy" between the two sets of degrees of freedom and X_0. This occurs whenever the incremental informational association of each set $\mathbb{X}_e$, $\mathbb{X}_i$ with X_0 is greater than its respective unconditional informational association with X_0. In those cases, knowledge of one set of degrees of freedom reduces more uncertainty about X_0 if the other degree of freedom is already known. There is information about X_0 in the combination of $\mathbb{X}_e$ and $\mathbb{X}_i$ that is not contained in either of those two sets individually.

Conversely, in settings where $\mathbb{M}\left(X_0, \mathbb{X}_e, \mathbb{X}_i\right) > 0$ there is a measure of redundancy between the two sets of degrees of freedom in informationally accounting for X_0. The two sets of individual degrees of freedom contain some of the same information about X_0.

4 Individual Income and Its Categorical Determinants

The concepts and approach outlined above can be applied to inquiry into the associations involved in the determination of individual income. Specifically, they enable characterizations of the informational associations between income and sets of observable economic and social-identity characteristics. This section develops part-piecewise representations of the indices developed above that allow such characterizations, within a broader analytical framework for considering the informational determinants of individual income.

4.1 Categorical Degrees of Freedom and Part-Piecewise Representations

Formally, let Y be an individual degree of freedom measuring income over a given time period, and let f be the (marginal) frequency distribution of individuals over a coarse-graining of the range of possible income values into m bins. The entropy of this distribution is given by,[44]

$$H(Y) = -\sum_{m} f_m \log f_m \tag{14}$$

As noted above, this is a measure of heterogeneity, diversity, or differentiation in individual measures of income. It takes a maximum value of $\log m$ when income is distributed with maximum heterogeneity, evenly across all possible income levels. It's minimum possible value is zero, when all incomes take on the same value. $H(Y)$ is also a measure of how much uncertainty an observer of the distribution f has about the exact distribution of income across each specific individual—given by the average number of bits needed to enumerate all permutations of individuals across income levels possible under f.

Let $\mathbb{X}_k$ be a set of observable categorical individual degrees of freedom. That set may be divided into a subset $\mathbb{X}_e$ of observable "economic" categories—specifically, level of education or age group—and another subset $\mathbb{X}_i$ of social-identity categories. Since present flows of individual income cannot be expected to have any causal influence on an individual's social identity, age, or level of education, the informational associations between Y and all elements of $\mathbb{X}_k$ can be understood as informational measures of the influence those elements have in the determination of individual measures of income.

Since all elements of $\mathbb{X}_k$ are categorical variables, it will be useful to consider a partially pointwise decomposition of the coefficient of informational association defined in 13 across all individual values χ_k taken by $\mathbb{X}_k$,

$$A(Y||\mathbb{X}_k) = \sum_{\chi_k} f(\chi_k)\, a(Y||\chi_k)\,; \qquad a(Y||\chi_k) \equiv \frac{H(Y) - H(Y|\chi_k)}{H(Y)} \tag{15}$$

[44] It is important to note that this is different than the entropy-index of income inequality proposed by Theil, 1971. That index is given by $-\sum_n y_n \log y_n$, where y_n is the share of total income received by each of the $n = 1, ..., N$ individuals in the system. While this index has a number of useful properties, its relationship to the combinatorial considerations that make entropy a useful conceptual tool in analysis of large-N economic systems is not at all clear. From the perspective developed in what follows, the convention in Theil, 1971 amounts to taking units of income as the functional units of an economy, and considering that they may take "states" corresponding to different individuals. This is an impractical basis for descriptions of economic functioning, not least because money is fungible and income is a flow (and hence a poor unit of analysis over any specific time period).

The coefficients of joint pointwise informational association $a\left(X_0||\chi_k\right)$ measure the proportional reduction in heterogeneity or observer uncertainty about Y once it is verified that $\mathbb{X}_k = \chi_k$. Put differently, it is a proportional, informational measure of the heterogeneity in values of individual income in the sub-population defined by χ_k, relative to the heterogeneity in values of individual income across the population as a whole.

Note that even though $A\left(Y||\mathbb{X}_k\right)$ is always non-negative, values of $a\left(Y||\chi_k\right)$ can be negative. This occurs when a χ_k sub-population has greater heterogeneity in income levels than the population as a whole. In those cases, the measures in $\mathbb{X}_k$ have a greater informational influence on incomes within sub-populations with $\mathbb{X}_k \neq \chi_k$ than on incomes for individuals with χ_k characteristics. Put differently, factors *other than those contained in* $\mathbb{X}_k$ have a greater informational role in shaping the heterogeneity of individual income within sub-population χ_k than within the population as a whole.

The point-wise coefficients of joint informational association inside the sum in 15 admit the same kind of decomposition into two sets of individual degrees of freedom or characteristics as above. Denoting those sets by $\mathbb{X}_e$ and $\mathbb{X}_i$ this may be formally expressed as,

$$\begin{aligned} a\left(Y||\chi_k\right) &= a\left(Y||\chi_e\right) + \mathbb{I}_{\chi_i|\chi_e} = a\left(Y||\chi_i\right) + \mathbb{I}_{\chi_e|\chi_i} \\ &= \mathbb{I}_{\chi_i|\chi_e} + \mathbb{I}_{\chi_e|\chi_i} + m\left(Y,\chi_e,\chi_i\right) \end{aligned} \tag{16}$$

Where $\mathbb{I}_{\alpha|\beta}$ and $m\left(Y,\alpha,\beta\right)$ are part-pointwise versions of the coefficients of incremental association defined in 12. The relationship between the part-pointwise coefficient of mutual association and its population-wide version follows trivially,

$$\begin{aligned} \mathbb{M}\left(Y,\mathbb{X}_e,\mathbb{X}_i\right) &= \sum_{\chi_e,\chi_i} f\left(\chi_e,\chi_i\right) m\left(Y,\chi_e,\chi_i\right) \\ &\text{where,} \\ m\left(Y,\chi_e,\chi_i\right) &= a\left(Y||\chi_e\right) - \mathbb{I}_{\chi_e|\chi_i} = a\left(Y||\chi_i\right) - \mathbb{I}_{\chi_i|\chi_e} \\ &= a\left(Y||\chi_e\right) + a\left(Y||\chi_i\right) - a\left(Y||\chi_k\right) \end{aligned} \tag{17}$$

This coefficient reflects an important aspect of the informational association between Y and pairs $\chi_k = (\chi_e, \chi_i)$ of individual characteristics. As with the population coefficient of mutual information, one broad way to characteritze what m captures is by considering its sign. Any pair with $m(Y, \chi_e, \chi_i) < 0$ can be understood to have a pointwise "synergistic" informational association with Y. For such pairs, their incremental informational association coefficients, $\mathbb{I}_{\chi_e|\chi_i}, \mathbb{I}_{\chi_e|\chi_i}$ are greater than their respective unconditional coefficients of informational association, $a(Y||\chi_e), a(Y||\chi_i)$. Equivalently, their joint coefficient of informational association with Y is greater than the sum of their respective coefficients of informational association with Y. There is information about individual income in the combination of characteristics (χ_e, χ_i) that is not contained in either of those characteristics by themselves. Conversely, characteristic pairs (χ_e, χ_i) for which $m(Y, \chi_e, \chi_i) > 0$ have measures of redundancy in their informational association with Y. Note that pairs (χ_e, χ_i) can be redundant or synergistic even when the sets $\mathbb{X}_e$ and $\mathbb{X}_i$ are synergetic or redundant, respectively.

The piecewise mutual information coefficient has a more general interpretation and significance. It is a negative measure of informational association between a set of economic characteristics χ_e and income for a group χ_i, relative to the overall informational association between those economic characteristics and income for the entire population: $m(Y, \chi_e, \chi_i) = a(Y||\chi_e) - \mathbb{I}_{\chi_e|\chi_i}$. It may be taken as a measure of the comparative informational significance of those characteristics for the income of each social-identity group.

Three general results follow from 16 and 17. First, note that 16 defines a simple measure of within-group comparative incremental informational association between income and two sets of characteristics (χ_e, χ_i) within the sub-population they define,

$$d_{e,i} \equiv \mathbb{I}_{\chi_e|\chi_i} - \mathbb{I}_{\chi_i|\chi_e} = a(Y||\chi_e) - a(Y||\chi_i) \tag{18}$$

This difference measures the independent informational association of χ_e and Y *relative* to the independent informational association of χ_i and Y for group (χ_e, χ_i). Interestingly, this difference can be computed by considering only the population-wide informational association between the set of economic characteristics χ_e and income, and the equivalent measure of the social identity χ_i. This allows estimation over a lower-dimensional coarse-graining than direct estimation of $d_{e,i}$.

Second, the difference of $d_{e,i}$ across two identity groups, χ_i^m and χ_i^w, also offers a useful quantification of the role played by social identity in the determination of income distribution. Formally, that difference is given by,

$$\Delta_{i^m,i^w} \equiv d_{e,i^m} - d_{e,i^w} = a\left(Y||\chi_i^w\right) - a\left(Y||\chi_i^m\right) \tag{19}$$

Whenever $a\left(Y||\chi_i^m\right) < a\left(Y||\chi_i^m\right)$, *any set of economic characteristics* will be more informationally significant relative to identity in the determination of the distribution of income within population χ_{i^m} than in the determination of the distribution of income within population χ_{i^w}. The simple and easily estimated difference $\Delta_{i^m,i^w} = a\left(Y||\chi_i^w\right) - a\left(Y||\chi_i^m\right)$ thus offers a quick, macroscopic measure of discriminatory treatment across all possible individual economic characteristics.

Finally, from 17 it is possible to consider the difference between the coefficient of mutual information $m\left(Y, \chi_e, \chi_i\right)$ across two social-identity groups, χ_i^v and χ_i^h,

$$m\left(Y, \chi_e, \chi_i^v\right) - m\left(Y, \chi_e, \chi_i^h\right) = \mathbb{I}_{\chi_e|\chi_i^h} - \mathbb{I}_{\chi_e|\chi_i^v} \tag{20}$$

The coefficient of mutual information for group χ_i^v will be larger than the equivalent measure for group χ_i^h if the incremental measure of informational association of the economic characteristics in question is greater for χ_i^h than for χ_i^v. Differences in values of m across two groups give us a negative measure of the independent informational influence the set of economic characteristics χ_e has on the income of each group.

The part-piecewise coefficients developed above allow formal, non-paramteric characterizations of various informational associations and interactions between economic and social-identity characteristics and individual income. Those characterizations are defined within the broader terms of an information-theoretic framework that situates those associations within the broader set of determinations of values of Y. This allows integration of inquiry into the effects of patterns of discrimination by elements of social identity and inquiry into the full set of determinants of individual income. A first step toward such an undertaking is offered below, on the basis of large-scale, repeated cross-sectional data from the U.S.

5 Observed Patterns in U.S. data

The framework can be applied to large-scale U.S. cross-sectional data to cast light onto the associations and interaction between income, social identity, and certain economic characteristics established by systems of social discrimination in that society. To do this, we considered four waves of the U.S. Census data, from 1970 to 2000, as well as the 2007-2011 pooled American Community Survey (hereafter referred to as the 2010 Census for simplicity), extracted from Ruggles et al., 2015. These surveys provide the most comprehensive and nationally representative source of data for income estimates across various subpopulations in the United States.[45]

For each respondent reporting market income, we observe their annual wage Y, economic characteristics χ_e and social identity χ_i. We then generate for each year a coarse-grained joint distribution $f(Y, \chi_e, \chi_i)$ by "binning" Y into equal-spaced income brackets for each cell defined by categorical attributes (χ_e, χ_i).[46] All indices developed above are estimated for the five annual joint distributions in question. This section reports the salient results of these exercises, starting with those involving only the two observable characteristics that can be most narrowly understood as "economic," and then considering the influence of identity and the interaction of identity and economic characteristics.

5.1 Observable Economic Characteristics

The dataset offers two "economic" characteristics widely understood to shape income: Level of education and age, which may be broadly associated with years of work experience. The informational associations between those two measures and income are interesting and illustrate well the distinctive usefulness of coefficients of informational association in the study of the determinants of individual income.

Sub-populations with higher levels of education not only have higher average wages, but they also have overall distributions of income with greater measures of heterogeneity or entropy. Put differently, those with higher levels of education enjoy greater scopes for differentiation in their

[45] See appendix A.1 for details on the construction of our sample, and appendix A.2 for sample sizes.
[46] For a robustness analysis of the binning scheme, see appendix A.3.

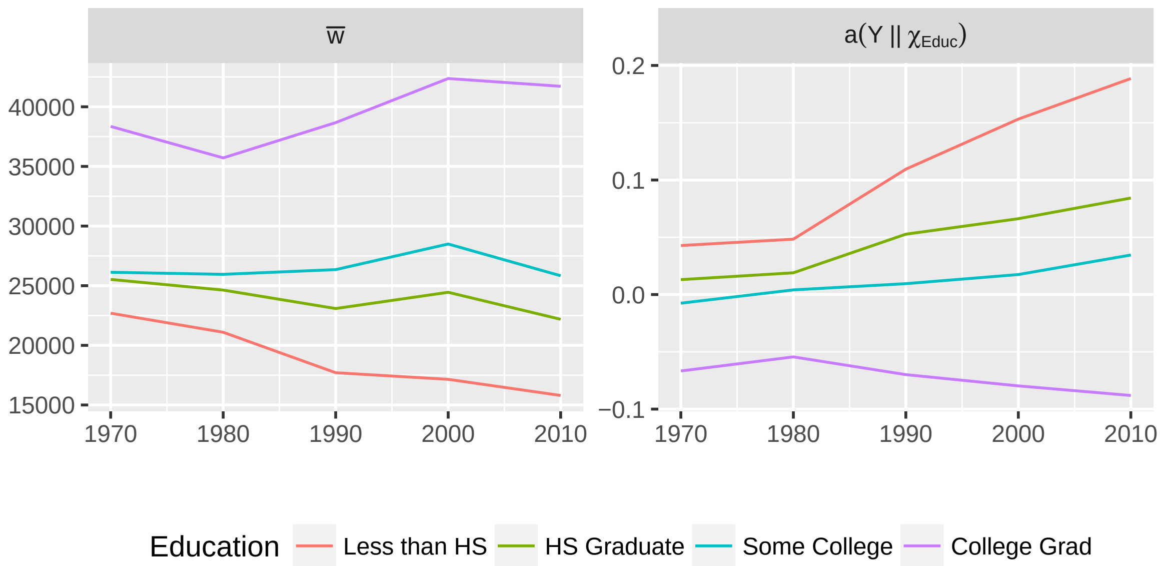

Figure 3: Average real wages and unconditional informational association between education and wages. 1970-2010 Census and ACS data.

income. In a manner of speaking, education appears to yield heterogeneity gains to the population in question, as shown in Figure 3. Those lacking a college degree have seen a steady decrease in the heterogeneity of their distribution of income—ensuring that their levels of educational attainment have become steadily more *informative* of their income. This is particularly dramatic for the well-studied category of high-school dropouts. By 2010 their educational attainment accounted for almost 20 percent of the informational content in their distribution of income. At the same time, college graduates have steadily gained scopes for income differentiation over the past 30 years.

The informational associations of income and age follow a similar pattern. Age is most informative for respondents aged 20-29, and then of decreasing importance as workers advance in their career, as shown in Figure 4. There does appear to be a reversal for ages 60-64, when workers are reaching retirement age. The informational association of young age and wage incomes appears to be increasing over the period 1970-2010. Not only are wages most strongly determined by age for the young, this is also the group with the lowest average wage.

Education and experience appear to generate heterogeneity gains for the sub-populations in question. More insights into the conjugation of the effects of education and experience can be

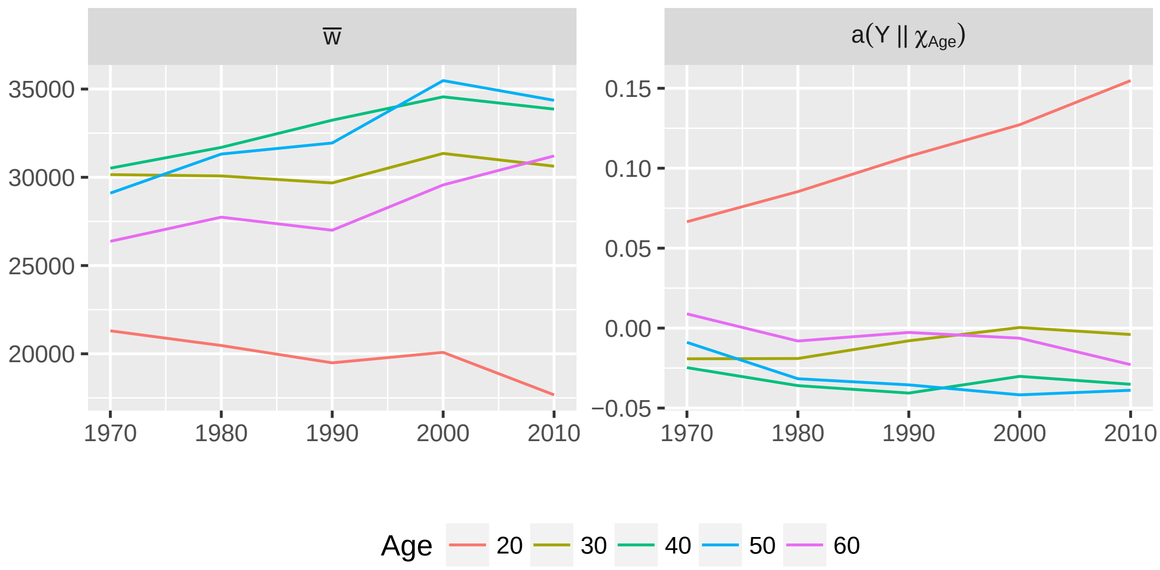

Figure 4: Average real wages and unconditional informational association between age and wages. 1970-2010 Census and ACS data.

obtained from considering the coefficient of incremental informational association of educational attainment for each age group, $\mathbb{I}_{\chi_{educ}|\chi_{age}}$. The evolution of these measures across age and education-level groups is shown for three years of observation in Figure 5.

Four features of this figure stand out. First, the informational significance of education does not change very much across age groups for those without a high-school degree. For that population, age and experience do not appear to translate into heterogeneity gains. Second, for all other groups, education results in heterogeneity gains that start to grow with age and experience, reach a peak, and then decline. Third, heterogeneity gains are greater for those with higher education, for whom gains are greater and peak later in life. Those with university degrees achieve the greatest measure of income heterogeneity, in their 50s. Finally, the overall differences in heterogeneity across education groups have increased dramatically in the past 40 years, across all age groups. These findings give a new, informational perspective on the well established finding that returns on education have increased over the past few decades in the U.S.

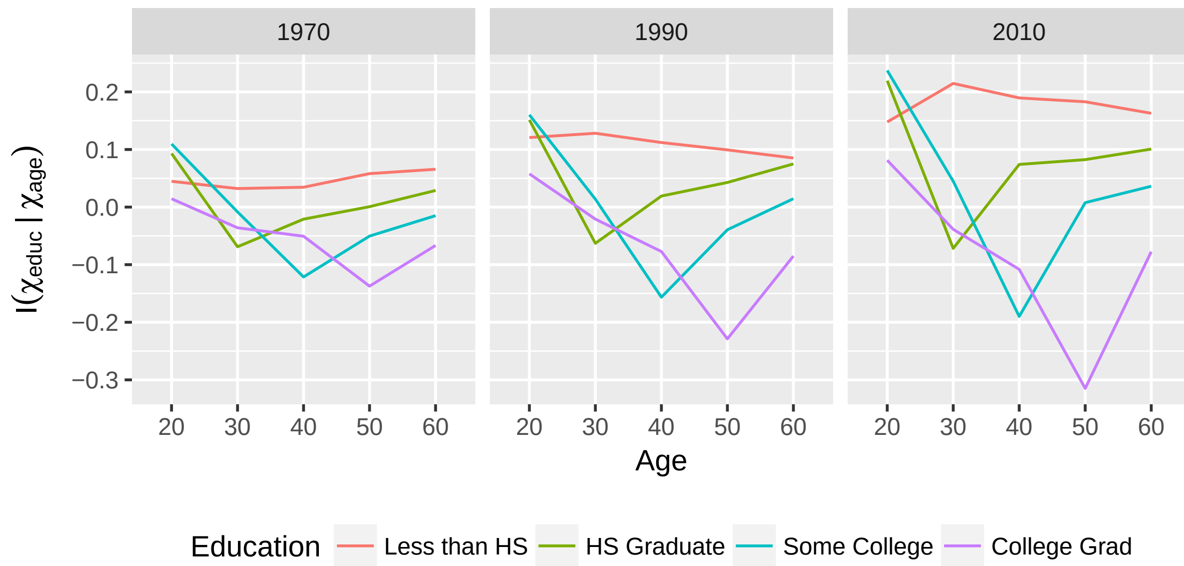

Figure 5: Incremental informational association between education and income by age group. 1970-2010 Census and ACS data.

5.2 The Effects of Social Identity

The dataset also offers fairly detailed information on respondents' social identity, by gender, race, and ethnicity. We considered six sub-populations defined by two gender and by three race/ethnicity categories: black, white, and hispanic. There is considerable variation in the distributions of income and economic characteristics across these groups. That variation tells us much about the economic effects of systemic patterns of discrimination by social identity. Three of its features stand out. First, there is a fairly persistent ordinal ranking in the heterogeneity of distributions of income across the six groups, with women and minorities enjoying smaller scopes for differentiation by income. Second, for those groups, the independent informativeness of economic characteristics relative to that of social identity is always smaller than for white men. And third, measures of the relative informativeness of educational attainment over levels of income for different groups reveal a striking pattern: for women, those measures fall as higher levels of educational attainment are considered. For white men, they increase.

Consider first the unconditional informational association between gender and income. Figure 6 reveals the strong informational impact of being female on income, compared to that of being

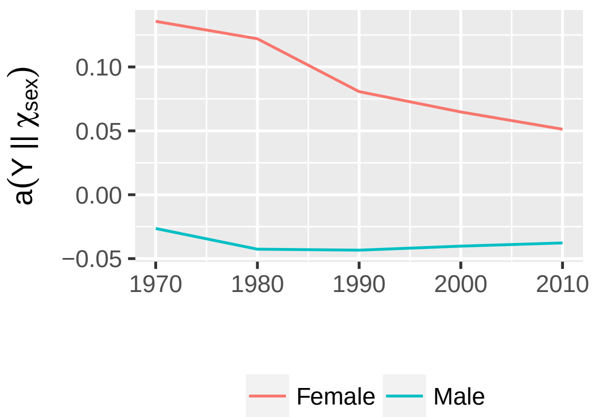

Figure 6: Unconditional informational association between gender and annual and hourly income. 1970-2010 Census and ACS data.

male. Not only do women have lower average incomes, but they also have smaller scopes to achieve differentiated income outcomes. This doubtlessly reflects in part the fact that, as a reflection of the social processes shaping expectations about care obligations discussed above, there is a higher prevalence of part-time employment among women than among men. But the comparative distributions of *hourly* earnings show a similar, if moderated, pattern of differential heterogeneity.

Evidence from a number of OECD economies has established that women are concentrated in a much smaller number of occupations than men. In 2005, that concentration ensured that, on average, half of employed women had jobs in only eleven of 110 listed occupations in European OECD countries and the U.S. In contrast, half of employed men were on average concentrated in the 21 more common occupational categories.[47] The patterns evident in Figure 6 illustrate how entropy differentials across income distributions offer a simple and robust way to measure the relative "compression" of economic outcomes for women. As will become clear below, those differentials are important economic manifestations of inequalities by gender and other dimensions of social identity. They also reveal systematic differences in the comparative informational impact of economic characteristics on income distributions across identities.

The same pattern of entropy differentials is evident across the six identity groups we considered. Figure 7 shows the evolution of the informational association between social identity and income

[47]OECD, 2005; Puyenbroeck et al., 2012.

across the four observed levels of educational attainment. In all four panels we observe the same ordinal pattern, with white men enjoying the greatest scopes for income heterogeneity, followed by other men, and with Hispanic and Black women experiencing the greatest measures of income compression or homogeneity. As with women, minorities too suffer from an effective "compression" or reduction in heterogeneity of incomes compared to white men. This finding is broadly in line with the finding of greater occupational concentration among minorities than whites.[48] But it also points to a series of further results.

The entropy differentials in Figure 7 establish that social identity plays a significantly stronger informational role in the determination of incomes for certain groups. Other groups see factors other than their social identity play stronger informational roles in the determination of their incomes. Those factors include the economic characteristics in the data we considered, whose associations and interactions in the formation of income levels point to further economic iniquities born of systems of discrimination by social identity.

In line with the results established in equations 18 and 19, those differentials imply that the measure of incremental informational significance in the determination of income of social-identity characteristics relative to that of economic ones given by $d_{e,i}$ is always greater for those groups. This is illustrated in relation to the two observable economic characteristics in the dataset under consideration, educational attainment and age, in Figure 8, which shows the evolution of $d_{e,i}$ for those two characteristics across social-identity groups.

The understanding of the coefficients $m\left(Y, \chi_e, \chi_i\right)$ as negative, relative measures of the informational significance of economic characteristics χ_e in the determination of incomes for group χ_i, and the result expressed in equation 20 lead to a final set of revealing observations. The evolution of these coefficients for each of the six identity groups we considered is depicted in Figure 9.

In each panel in that figure the ordinal ranking of different identity groups gives us an inverse ranking of the relative strength of the informational influence of a given level of education in the determination of incomes across groups. A striking pattern emerges from this. Black and white women have the lowest measures for those coefficients for low levels of educational attainment. For

[48]See Grodsky and Pager, 2001 and Kochhar, 2005.

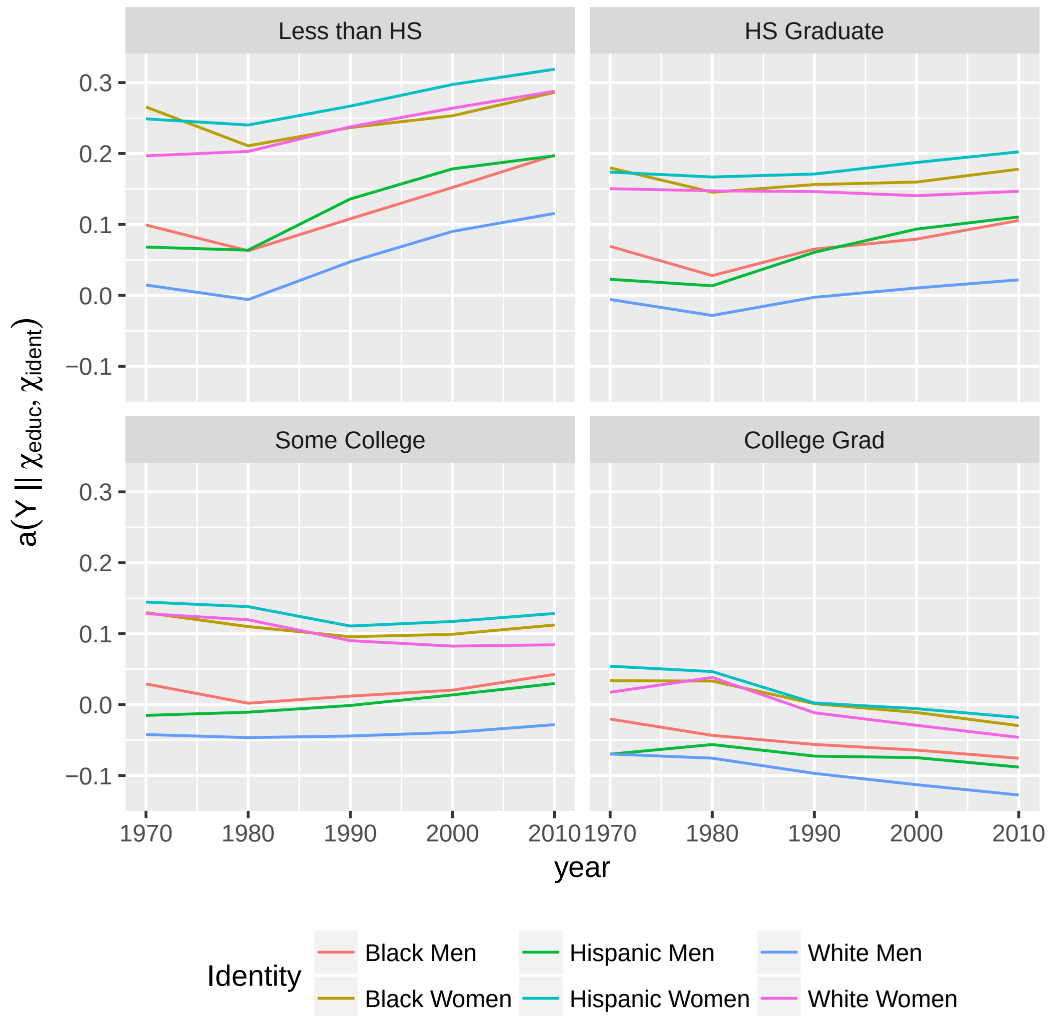

Figure 7: Joint informational association of educational attainment and social identity on income. 1970-2010 Census and ACS data.

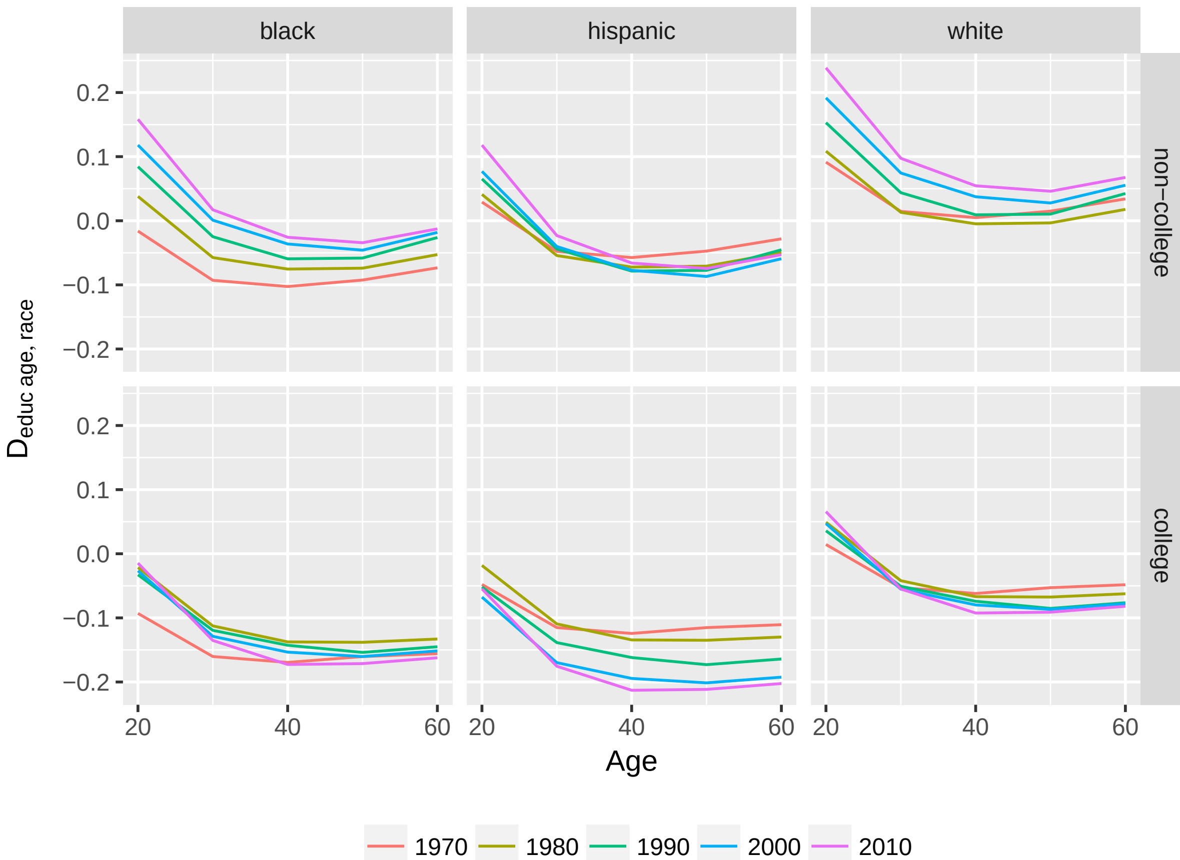

Figure 8: Relative independent informational association between observed economic characteristics and income, and between race/ethnicity and income. 1970-2010 Census and ACS data.

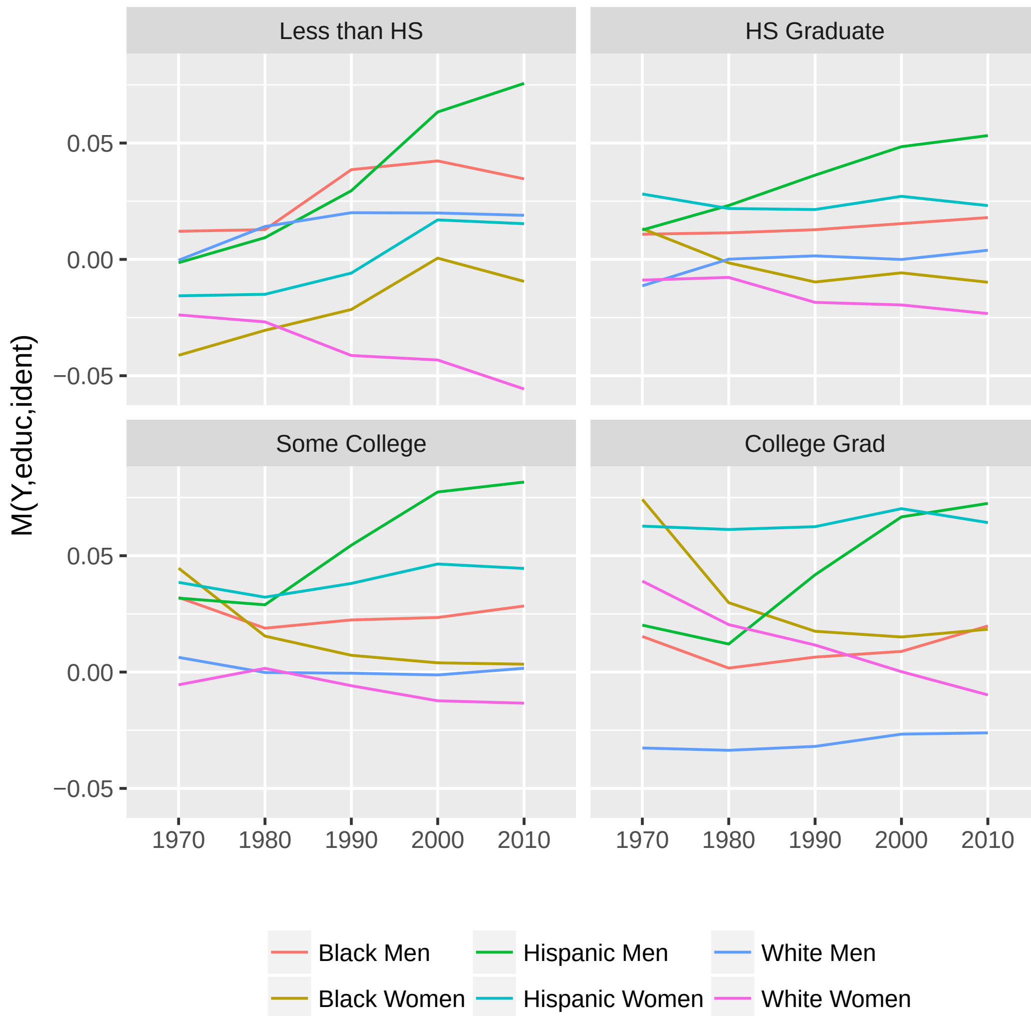

Figure 9: Tripartite mutual information between income, education, and identity. 1970-2010 Census and ACS data.

college graduates, white men, and in 2010 white women, have the lowest measures. In fact, in all these cases, $m\left(Y, \chi_e, \chi_i\right)$ is negative, establishing the presence of informational synergies between education and identity in the formation of income levels in those cases.

The combination of low educational attainment and being a woman is associated with a further set of determinants that exert a strong influence in the determination of income. Conversely, the combination of being white and having a college degree is associated with a further set of determinants that exert a strong influence in the determination of income. It seems clear that the effects of different educational attainments on income are very much mediated by a broader array of processes during which social realities of race and gender condition rather iniquitous outcomes.

A stronger and broader version of this result and other findings can be established with the aid of a final graphical representation of our findings. Let $S_{e,i} = \frac{\bar{w}_{e,i}}{\bar{\bar{w}}}$ be the ratio between the mean wage of group $\left(\chi_e, \chi_i\right)$ and the mean population wage. Now consider the relationship between this ratio and $m\left(Y, \chi_e, \chi_i\right)$, which provides a sense of how relative incomes vary as relative measures of the informational significance of educational level on income vary. This results in the associations reported in Figure 10.

The figure conveys a number of well-understood features, iniquities, and developments in the distribution of income in the U.S. over the past four decades. First, the growing range in each panel for the values of $S_{e,i}$ over time reflects the significant growth in income differentials across levels of education. Throughout these processes, that range also reveals stark differentials in average pay for the most and least formally qualified groups across social identity. Second, the relative average income of almost all groups of people without university degrees has been steadily falling over this period. White women are the only exception—those with high-school diplomas and some college have achieved some average increases in relative levels of pay. Third, the average relative income of college educated groups has grown unevenly across social-identity, with white, college-educated women securing the most significant relative gains in this period.

The figure also conveys a series of patterns involving differentials in the measures of $m\left(Y, \chi_e, \chi_i\right)$. First, the curves for all groups of women have a distinctive "tilt," ensuring they are almost always upward sloping in this space. They exhibit lower measures of $m\left(Y, \chi_e, \chi_i\right)$ for low levels of education,

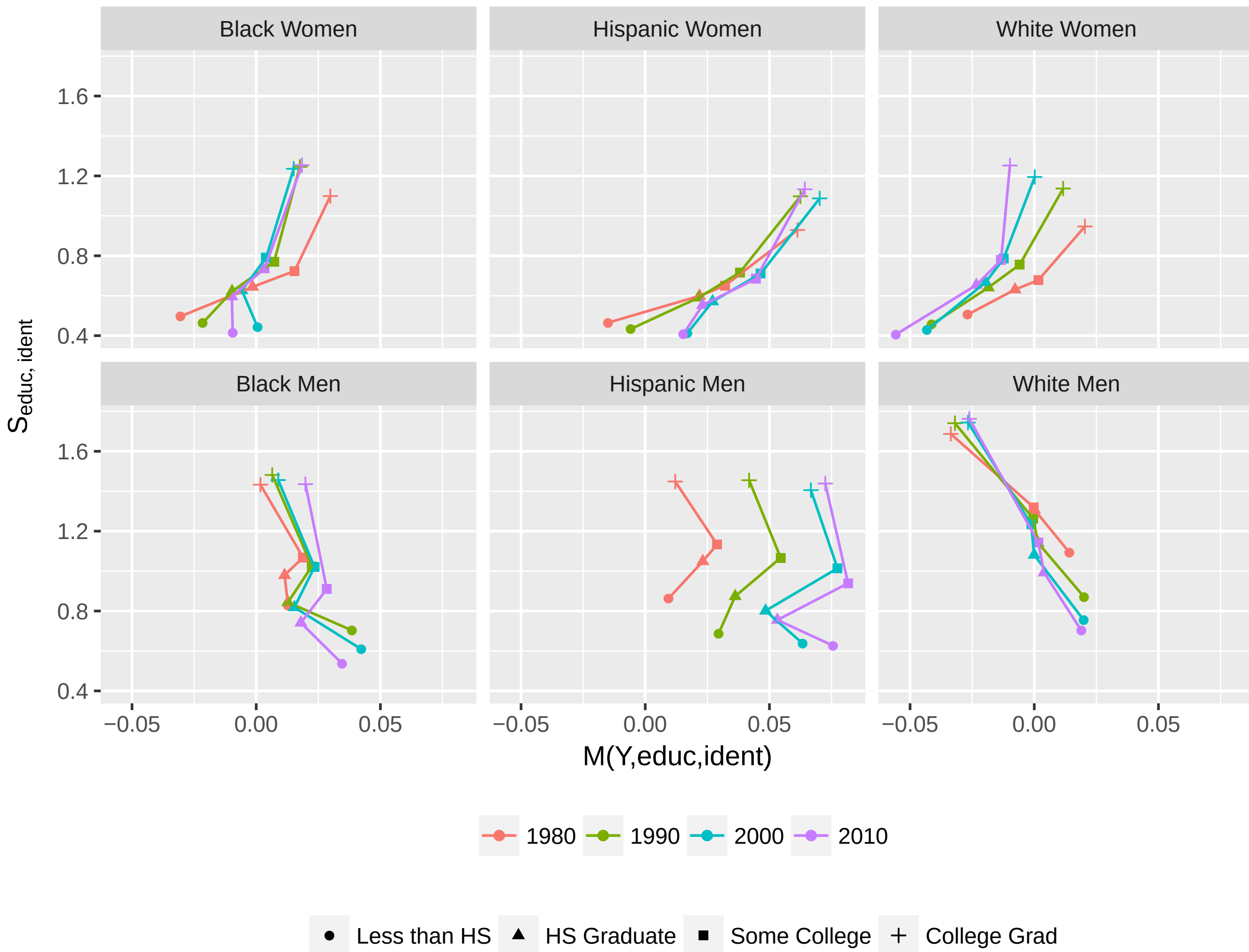

Figure 10: Mutual information between income, education, and identity, and relative average income of education level in question. 1970-2010 Census and ACS data.

which correspond to low average levels of pay. There is a comparatively strong relative incremental informational association between their low levels of education and their incomes. In contrast, all groups of women have higher measures for that coefficient of mutual information for high levels of education and pay. They enjoy comparatively weak relative incremental informational associations between their high levels of education and their incomes. Put differently, their educational level becomes a comparatively weaker informational predictor of their incomes as the level of education rises. It is a stronger informational predictor as it falls.

Two developments over the past forty years have tended to moderate the strength of this iniquitous reality. There has been a noticeable and steady increase in the independent informational significance of education in the determination of incomes for college-educated women. In 2010, college-educated white women joined college-educated white men as the only groups of college-educated people for whom we observe informational synergies between education and identity in accounting for income. This has not been accompanied by a similar improvement for black and hispanic women. At the same time, black and hispanic women with no high-school degree have seen an uneven but discernable decrease in the informational association between their level of education and their income over the past few decades. Interestingly, white women without high-school degrees have seen an increase in the informational significance of their low level of education in the determination of their income.

The curves for white men have also have a "tilt," but one with a negative slope. As their educational attainment levels increase, educational achievement becomes a comparatively *stronger* informational predictor of their incomes. The patterns for black and hispanic men are different. Across all educational levels, the comparative incremental informational significance of education for the incomes of hispanic men has been decreasing—resulting in the marked shift to the right in their curves. The curves for black men are also in part shifting to the right, with the most significant shift involving a steady decrease in the strength of education as an informational predictor of the income for college-educated black men.

At the broadest level, these observations strongly suggest that social patterns of discrimination ensure the returns on educational achievement are very unevenly distributed across social-identity groups. This has important implications for our understanding of the possibilities and limits of individual and social interventions seeking to reduce the economic consequences of discrimination.

6 Discrimination and a Privilege

The discussion above opens new perspectives on the economic consequences of systemic patterns of discrimination by social identity. It also helps identify a distinctive form of social privilege enjoyed by those not subject to it.

Systems of discrimination by elements of social identity do not merely result in lower average incomes for discriminated groups. They ensure that in the market allocation of the social product, social identity plays a pervasive and iniquitous role. It conditions important differences in the group-wide effects economic characteristics like educational attainment have on income. And most broadly, it exerts a stronger influence in the formation of individual incomes within discriminated groups. The economic, political, and social processes shaping the incomes of women and minorities effectively leave individual members of those groups with significantly smaller scopes for differentiation in economic outcomes on the basis of underlying, economically relevant individual characteristics. In varying but persistently significant measures, their social identities play an outsized role in the determination of their levels of income. Their economic outcomes reflect the extent to which they are treated by their gender. By their ethnicity. And by the color of their skin.

The contrast with the treatment received by other groups, as evident from their patterns of income, is striking. In addition to receiving higher average incomes, white men in the U.S. as a group enjoy a distinctive and subtle privilege: significantly greater scopes for differentiation by income among themselves. While that differentiation is doubtlessly shaped by non-economic factors that limit the possibilities available to many individuals—such as class or regional background, sexuality, physical traits, and many other characteristics we have been unable to consider here—*as a group*, white males are clearly treated by the socio-economic processes shaping incomes in a more individually discerning manner. This is a privilege not generally enjoyed by other groups.

It is also a difficult privilege to address, because it may not appear as such to those enjoying it. Within the privileged group, high measures of differentiation in income by contemporaneous, observable individual economic characteristics can lend support to "meritocratic" inferences. The group-wide comparisons offered by this paper point to the problems with such conclusions: Differentiation by "merit" or by any other characteristic is not as widely available to members of other social identity groups. They also point to further exercises that may help highlight the significance of other forms of discrimination in the definition of this privilege and of who enjoys it exactly: exercises that may establish how the narrowing of economic opportunities experienced by women and minorities may also by experienced by groups of white men defined by class background or other characteristics.

We hope that in this sense the paper's methods and findings contribute to further work on the economic consequences of all forms of social discrimination, and help inform debates on the kinds of economic outcomes a society that genuinely values all its members"by the content of their character" would attempt to promote and achieve.[49]

References

Abbott, A. (1983). "Sequence of Social Events: Concepts and Methods for the Analysis of Order in Social Processes." *Historical Methods: A Journal of Quantitative and Interdisciplinary History* 16.4, pp. 129–147.

Abbott, A. (1995). "Sequence Analysis: New Methods for Old Ideas." *Annual Review of Sociology* 21, pp. 93–113.

Alfarano, S. and M. Milaković (2008). "Does classical competition explain the statistical features of firm growth?" *Economics Letters* 101.3, pp. 272–274.

Arrow, K. J. (1971). "The Theory of Discrimination." *Industrial Relations Section, Princeton University Working Paper* 30A.

Ashenfelter, O., C. Harmon, and H. Oosterbeek (1999). "A review of estimates of the schooling/earnings relationship, with tests for publication bias." *Labour Economics* 6.4, pp. 453–470.

Becker, G. (1957). *The Economics of Discrimination*. University of Chicago Press.

[49] As put by Martin Luther King Jr. in his 1963 speech to the March on Washington.

Becker, G. S. (1962). “Investment in Human Capital: A Theoretical Analysis.” *Journal of Political Economy* 70.5, Part 2, pp. 9–49.

Behrman, J., R. Pollak, and P. Taubman (1986). “Do Parents Favor Boys?” *International Economic Review* 27.1, pp. 33–54.

Bennasar, M., Y. Hicks, and R. Setchi (2015). “Feature Selection Using Joint Mutual Information Maximinzation.” *Expert Systems with Applications* 42, pp. 8520–8532.

Bhargava, A. (2008). *Food, Economics and Health*. Oxford University Press.

Bottazzi, G. and A. Secchi (2003). “Why Are Distributions of Firm Growth Rates Tent-shaped?” *Economics Letters* 80, pp. 415–420.

Bottazzi, G. and A. Secchi (2006). “Explaining the Distribution of Firm Growth Rates.” *The RAND Journal of Economics* 37.2, pp. 235–256.

Brown, C. (2015). “The Educational, Psychological, and Social Impact of Discrimination on the Immigrant Child.” *Migration Policy Institute Report*.

Cheng, S. (2014). “A Life Course Trajectory Framework for Understanding the Intracohort Pattern of Wage Inequality.” *American Journal of Sociology* 120.3, pp. 633–700.

Council, N. R. (2014). *The Growth of Incarceration in the United States: Exploring Causes and Consequences*. Ed. by J. Travis, B. Western, and S. Redburn. The National Academies Press. ISBN: 978-0-309-29801-8. DOI: 10.17226/18613.

Daniel, W. (1968). *Racial Discrimination in England*. Penguin Books.

Darmon, N. and A. Drewnowski (2008). “Does Social Class Predict Diet Quality?” *The American Journal of Clinical Nutrition* 87.5, pp. 1107–1117.

Debreu, G. (1959). *Theory of Value*. Yale University Press.

Dos Santos, P. L. and E. Scharfenaker (2018). “Competition, Self-Organization, and Social Scaling—Accounting for the Observed Distributions of Tobin’s q.” forthcoming.

dos Santos, P. L. (2017). “The Principle of Social Scaling.” *Complexity* 2017.doi:10.1155/2017/8358909.

Drewnowski, A. and N. Darmon (2005). “Food Choices and Diet Costs: And Economic Analysis.” *The Journal of Nutrition* 135.4, pp. 900–904.

Eagly, H. and S. Karau (2002). “Role Congruity Theory of Prejudice Toward Felmale Leaders.” *Psychological Review* 573-598.

Fano, R. M. (1961). “Transmission of Information,” p. 57059.

Firth, M. (1981). “Racial Discrimination in the British Labour Market.” *Industrial and Labor Relations* 34, pp. 265–272.

Firth, M. (1982). “Sex Discrimination and Job Opportunities for Women.” *Sex Roles* 8.8, pp. 891–901.

Folbre, N. (1994). *Who pays for the kids? Gender and the structures of constraint.* Economics as social theory. London ; New York: Routledge. 335 pp. ISBN: 978-0-415-07564-0 978-0-415-07565-7.

Folbre, N. (2012). “The Political Economy of Human Capital.” *Review of Radical Political Economics* 44.3, pp. 281–292.

Fraser, N. (1994). “After the Family Wage: Gender Equity and the Welfare State.” *Political Theory* 22, pp. 591–618.

Gershenson, S. and N. Papageorge (2018). “The Power of Teacher Expectations.” *Education Next.*

Gilliam, W., A. Maupin, C. Reyes, M. Accavitti, and F. Shick (2016). “Do Early Educators’ Implicit Biases Regarding Sex and Race Relate to Behavior Expectations and Recommendations of Preschool Expulsions and Suspensions.” *Yale Child Study Center Research Study Brief.*

Greene, M., N. Way, and K. Pahl (2006). “Trajectories of Perceived Adult and Peer Discrimination among Black, Latino, and Asian American Adolescents: Patterns and Psychological Correlates.” *Developmental Psychology* 42.2, p. 219.

Grodsky, E. and D. Pager (2001). “The Structure of Disadvantage: Individual and Occupational Determinants of the Black-White Wage Gap.” *American Sociological Review* 66.4, pp. 542–567.

Hamilton, D., A. Austin, and W. Darity (2011). “Whiter Jobs, Higher Wages.” *Economic Policy Institute Briefing Paper* 288, pp. 1–13.

Han, T. S. (1980). “Multiple Mutual Informationa and Multiple Interactions in Frequency Data.” *Information and Control* 46.1, pp. 26–45.

Hanel, R. and S. Thurner (2011). “When do generalized entropies apply? How phase space volume determines entropy When do generalized entropies apply? How phase space volume determines entropy.” *EPL* 96.

Heckman, J. and T. Kautz (2013). “Fostering and Measuring Skills: Interventions that Improve Character and Recognition.” *NBER Working Paper*.

Heilman, M. (2012). “Gender Stereotypes and Workplace Bias.” *Research in Organizational Behavior* 32, pp. 113–135.

Hilton, J. and W. von Hippel (2012). “Stereotypes.” *Annual Review of Psychology* 47, pp. 237–271.

Hoddinott, J., J. Maluccio, J. Behrman, R. Flores, and R. Martorell (2008). “Effect of a nutrition intervention during early childhood on economic productivity in Guatemalan adults.” *Lancet* 371, pp. 411–416.

Ince, R. (2017). “The Partial Entropy Decomposition: Decomposing multivariate entropy and mutual information via pointwise common surprisal.” arXiv:1702.01591v2.

Joshl, H. and P. Paci (2001). *Unequal Pay for Women and Men: Evidence from the British Cohort Studies*. MIT Press.

Jowell, R. and D. Wissoker (1970). “Racial Discrimination and White-Collar Workers in Britain.” *Race* 11, pp. 397–417.

Kanter, R. (1977). *Men and Women of the Corporation*. Basic Books.

Kenney, G. and D. Wissoker (1994). “An Analysis of the Correlates of Discrimination Facing Young Hispanic Job-Seekers.” *American Economic Review* 84.674-683.

Kochhar, R. (2005). “The Occupational Status and Mobility of Hispanics.” *Pew Hispanic Center*.

Kossek, E. (2006). “Work and Family in America: Growing Tensions Between Employment Policy and a Changing Workforce.” In: *America at Work: Choices and Challenges*. Ed. by E. Lawler and J. O’Toole. Palgrave MacMillan, pp. 53–72.

Kossek, E., R. Su, and L. Wu (2016). ““Opting Out” or “Pushed Out”? Integrating Perspectives on Women’s Career Equality for Gender Inclusion and Interventions.” *Journal of Management* 43.1, pp. 228–254.

Larson, N. and M. Story (2011). "Food Insecurity and Weight Status Among U.S. Children and Families: A Review of the Literature." *American Journal of Preventive Medicine* 40.2, pp. 166–173.

Lavassani, K. and B. Movahedi (2014). "Developments in Theories and Measures of Work-Family Relationships: From Conflict to Balance." *Contemporary Research on Organization Management and Administration* 2.1, pp. 6–19.

Lawson, K., A. Crouter, and S. McHale (2015). "Links between family gender socialization experiences in childhood and gendered occupational attainment in young adulthood." *Journal of Vocational Behavior* 90, pp. 26–35.

Levine, R. and Y. Rubinstein (2013). "Smart and Illicit: Who Becomes an Entrepreneur and Does it Pay." *NBER Working Paper.*

Levy, D., J. Heissel, J. Richeson, and A. EK (2016). "Psychological and biological responses to race-based social stress as pathways to disparities in educational outcomes." *American Psychologist* 71.6, pp. 455–473.

Liu, J. (2004). "Childhood Externalizing Behavior: Theory and Implications." *Journal of Child and Adolescent Psychiatric Nursing* 17.3, p. 930103.

Lundberg, S. (2013). "Personality and Educational Inequality." *Journal of Labor Economics* 31.3, pp. 421–441.

M Bendick, C. J. and V. Reinoso (1994). "Measuring Employment Discrimination Through Controlled Experiments." *Review of Black Political Economy* 23, pp. 25–48.

Martins, L., K. Eddleston, and J. Veiga (2002). "Moderators of the Relationship between Work-Family Conflict and Career Satisfaction." *Academy of Management Journal* 45.2, pp. 399–409.

Marx, K. (1919). *Capital, Volume I.* Charles H. Kerr and Company.

Marx, K. (1970). *A Contribution to the Critique of Political Economy.* Progress Publishers.

McGill, W. J. (1954). "Multivariate Information Transmission." *IEEE Trans. Information Theory* 4.4, pp. 93–111.

Mincer, J. (1974). *Schooling, experience, and earnings.* Human behavior and social institutions 2. New York: National Bureau of Economic Research; distributed by Columbia University Press. 152 pp.

OECD (2005). *OECD Employment Outlook.* OECD.

Papageorge, N., V. Ronda, and Y. Zheng (2017). "The Economic Value of Breaking Bad: Misbehavior, Schooling and the Labor Market." *IZA Institute of Labor Economics* 10822.

Phelps, E. (1972). "The Statistical Theory of Racism and Sexism." *The American Economic Review* 62.4, pp. 659–661.

Polanyi, K. (2001). *The Great Transformation.* Beacon Press.

Puyenbroeck, T. V., K. D. Bruyne, and L. Sels (2012). "More than 'Mutual Information': Educational and sectoral gender segregation and their interaction on the Flemish labor market." *Labour Economics* 19.1, pp. 1–8. ISSN: 0927-5371. DOI: https://doi.org/10.1016/j.labeco.2011.05.002.

Quillian, L., D. Pager, O. Hexel, and A. Midtboen (2017). "Meta-Analysis of Field Experiments Shows no Change in Racial Discrimination in Hiring Over Time." *PNAS* 114.41, pp. 10870–10875.

Rangel, M. (2015). "Is Parental Love Colorblind? Human Capital Accumulation within Mixed Families." *Review of Black Political Economy* 42, pp. 57–86.

Riach, P. and J. Rich (1987). "Testing for Sexual Discrimination in the Labour Market." *Australian Economic Papers* 26, pp. 165–178.

Riach, P. and J. Rich (2002). "Field Experiments of Discrimination in the Market Place." *The Economic Journal* 483.F480-F518.

Ricardo, D. (1951). *On the Principles of Political Economy and Taxation.* Cambridge University Press.

Rosenbloom, S. and N. Way (2004). "Experiences of Discrimination among African American, Asian American, and Latino Adolescents in an Urban High School." *Youth and Society* 35.4, pp. 420–451.

Rosenzweig, M. and P. Shultz (1982). "Market Opportunities, Genetic Endowments, and Interfamily Resource Distribution: Child Survival in Rural India." *American Economic Review* 72, pp. 803–815.

Ruggles, S., K. Genadek, R. Goeken, J. Grover, and M. Sobek (2015). *Integrated Public Use Microdata Series: Version 6.0 [Machine-readable database].*

Scharfenaker, E. and P. L. dos Santos (2015). "The Distribution and Regulation of Tobin's Q." *Economic Letters* 137.2015, pp. 191–194.

Schultz, T. W. (1961). "Investment in Human Capital." *The American Economic Review* 51.1, pp. 1–17. ISSN: 00028282.

Smith, A. (1982). *The Wealth of Nations*. Penguin.

Stanley, M., L. Amaral, S. Buldyrev, S. Havlin, H. Leschhorn, P. Maass, M. Salinger, and H. Stanley (1996). "Scaling behavior in the growth of companies." *Nature* 379, pp. 804–806.

Strauss, J. and D. Thomas (1998). "Health, Nutrition, and Economic Development." *Journal of Economic Literature* 36.2, pp. 766–817.

Taveras, E., M. Gillman, K. Kleinman, J. Rich-Edwards, and S. Rifas-Shiman (2010). "Racial/Ethnic Differences in Early-Life Risk Factors for Childhood Obesity." *Pediatrics* 125.4, pp. 2009–2100.

Theil, H. (1971). *Economics and Information Theory*. Elsevier.

Theil, H. and A. Finezza (1971). "A Note on the Measurement of Racial Integration of Schools by Means of Informational Concepts." *Journal of Mathematical Sociology* 1, pp. 187–194.

Thomas, D. (2017). "Expectations of Educational Success as a Mediator Between Racial Discrimination and College GPA." *The New School Psychology Bulletin* 14.1.

Udry, J. R. (1994). "The Nature of Gender." *Demography* 31.4, pp. 561–473.

Von Hippel, C., D. Sekaquaptewa, and M. McFarlane (2015). "Stereotype threat among women in finance: Negative effects on identity, workplace well-being, and recruiting." *Psychology of Women Quarterly* 39, pp. 405–414.

Williams, J. C., M. Multhaup, and S. Mihaylo (2018). "Why Companies Should Add Class to Their Diversity Discussions." *Harvard Business Review*.

Witherspoon, D., S. Wooding, A. Rogers, E. Marchani, W. Watkins, M. Batzer, and L. Jorde (2007). "Genetic Similarities Within and Between Human Populations." *Genetics* 176, pp. 351–359.

Wong, C., J. Eccles, and A. Sameroff (2003). "The Influence of Ethnic Discrimination and Ethnic Identification on African American Adolescents' School and Socioemotional Adjustment." *Journal of Personality* 71.6, pp. 1197–1232.

Wood, E. M. (2002). *The Origin of Capitalism: A Longer View.* Verso.

Yang, H. and J. Moody (1999). "Feature Selection Based on Joint Mutual Information." In: *Proceedings of International ICSC Symposium on Advances in Intelligent Data Analysis*, pp. 22–25.

Yu, N., F.-C. Chen, S. Ota, L. Jorde, P. Pamilo, L. Patthy, M. Ramsey, T. Jenkins, S.-K. Shyue, and W.-H. Li (2002). "Larger Genetic Differences Within Africans than Between Africans and Eurasians." *Genetics* 161, pp. 269–274.

A Appendices

A.1 Data source and variable construction

The data for this study are drawn from the US decennial census in the years 1970-2000. After 2000, the Census Bureau replaced the more detailed "long" census form (which contains income questions *inter-alia*) with an annual nationally representative survey, the American Community Survey (ACS). We use the pooled ACS data from 2007-2011, but refer to it as the 2010 census for simplicity.

We have restricted the sample to respondents aged 18-64 not currently living in group quarters and whose primary occupation was non-military. For the overwhelming majority of working-age respondents, wages and salaries are the most important source of income in any given year. The income measure used in this study is therefore the reported annual pre-tax wages and salaries during the past calendar year (1970-2000) or the past 12 months (2010). We deflate this income measure to constant US Dollars using the 1999 Consumer Price Index. Since we focus on wage incomes, we further limit our analysis to respondents reporting at least one week of work in the reference year as well as a positive annual wage.

Measuring the differential evolution of heterogeneity in group outcomes requires a definition of the relevant social space over which agents can differentiate. Some groups may not currently occupy certain wage income ranges, even though they would be accessible under alternative configurations. We address this problem by considering the range of wages bounded between zero and the highest observed value in the data for the sample as a whole. This range will serve as a proxy for the social space potentially accessible to each subgroup.

A second, related problem concerns alterations of the data to preserve the anonymity of respondents. The most significant such modification consists in the censoring of very high earners, which means that wages of respondents above a certain threshold are replaced with a common value. Although relatively small in number, these so-called "top-coded" observations introduce distortions into the wage distribution. In particular, top-coding will tend to unduly homogenize wage incomes. As a first attempt at dealing with this problem, we choose to remove the top two percentiles of observations in each year.

We consider wages ranging between 0 and a year-specific maximum value based on the 98th percentile. Within this range, we generate 34 equally spaced histogram cells to construct the coarse-grained distribution $f(y, \chi_e, \chi_i)$.[50] The complex survey design of the Census and ACS data requires the use of sampling weights. These weights inflate the importance of individual observations to make the sample representative of the total population. The weights are summed up for of all observations within each cell defined by an income range $(y, y + \delta)$ and a combination of (χ_e, χ_i).

Social identity χ_i is analyzed through a very simple scheme with two gender categories, men and women, and four race categories: non-hispanic blacks and whites, hispanics, and a residual category of other races. The race variable is constructed by Ruggles et al., 2015 in an attempt to create consistent categories on the basis of the original racial and ethnic self-identification questions, which varied across census years. Notably, individuals with multiple responses were assigned a single response based on other demographic attributes of the individual and their area of residence. While hispanic origin was asked directly of respondents in 1980 and later, it was imputed in 1970 on

[50]See appendix A.3

the basis of parents' and grandparents' birthplace among other information (Ruggles et al., 2015). The residual "other races" category combines diverse groups such as American Indian and Asian, making the results hard to interpret. We therefore omit this category from the plots, although it was used in the calculations.

Economic characteristics χ_e are limited to education and experience. We use age grouped by decade to proxy for experience. Our education measure is based on the harmonized 'EDUC' variable, which represents either years of completed schooling or attained degree. We then collapse this variable into four broad education groups: Less than High School (Nursery school to Grade 11), High School Graduate (Grade 12), Some College (up to three years of college), College Graduate (four or more years of college).

A.2 Sample size

The census data sets used in this study are 1% or 5% samples of the national population. After application of the selection rules outlined in appendix A.1, the final sample sizes per year remain substantial. A final adjustment was made to exclude cases with missing values on either social identity, economic characteristics or wage information. Table 1 shows the sample sizes for calculations involving both social identity and educational achievement.

A.3 Binning scheme

Discrete frequency distributions of annual wages were formed by grouping observations into 34 equally spaced histogram bins. In order to assess the robustness of the key findings in this paper, some calculations were repeated using a binning scheme with 24 bins instead. The results are qualitatively identical to those presented in the main body of the paper, as the following comparisons show.

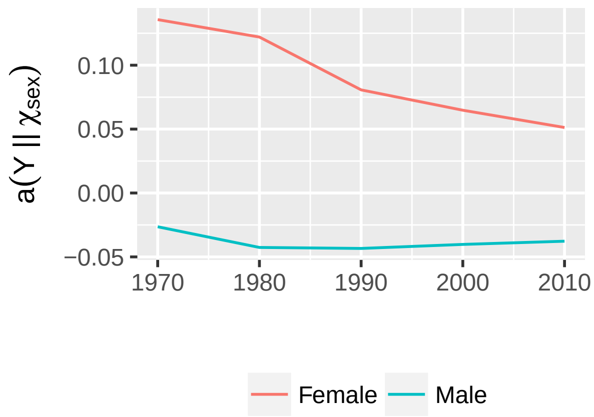

Figure 11: Unconditional informational association between gender and annual and hourly income. Compare to Figure 6

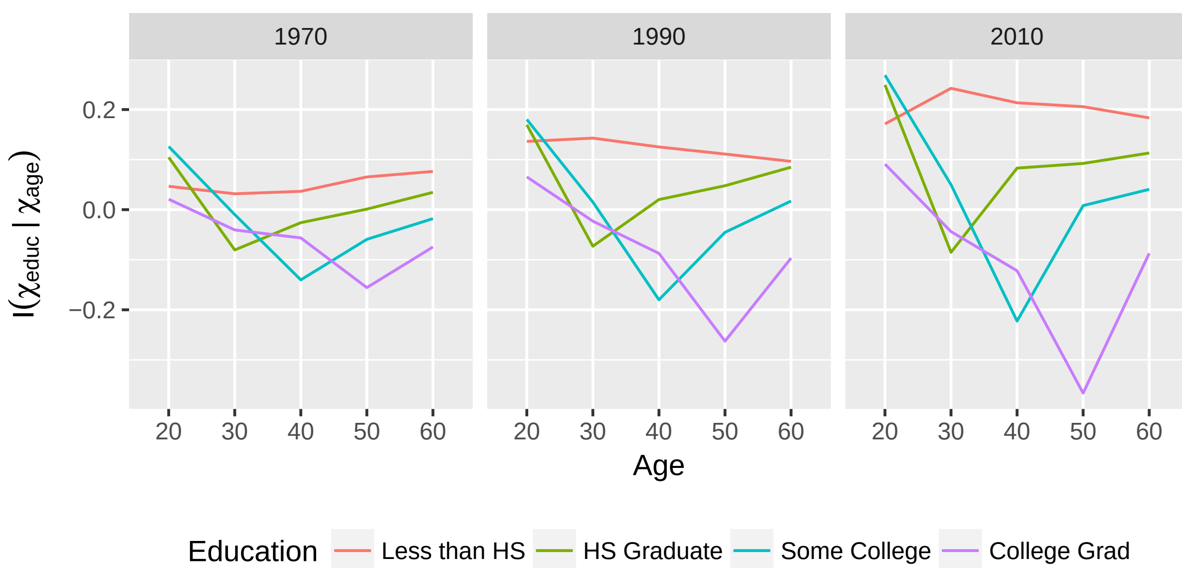

Figure 12: Incremental informational association between education and income by age group. Compare to Figure 5

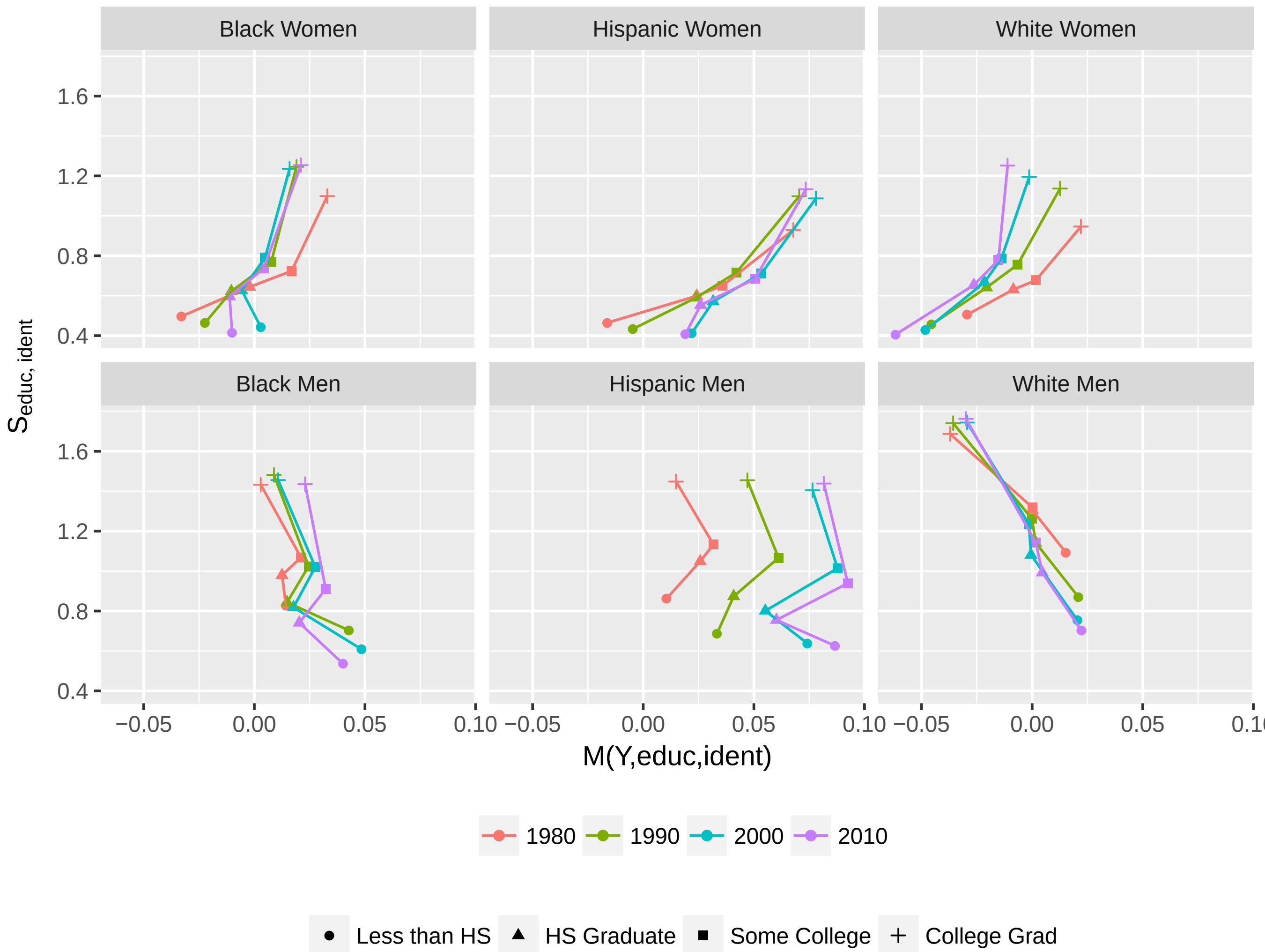

Figure 13: Mutual information between income, education, and identity, and relative average income of education level in question. Compare to Figure 10

Year	Identity	Less than HS	HS Graduate	Some College	College Grad	Total
1970	Black Men	47597	20317	6019	3308	77241
1970	Black Women	37740	23659	6744	4654	72797
1970	Hispanic Men	19538	7868	3279	2067	32752
1970	Hispanic Women	11108	6833	2180	1081	21202
1970	White Men	261573	248456	106876	96718	713623
1970	White Women	154769	231591	78829	58314	523503
1980	Black Men	88590	82399	40975	20711	232675
1980	Black Women	71874	95486	49931	26944	244235
1980	Hispanic Men	76963	43020	24223	12643	156849
1980	Hispanic Women	47700	40274	19368	8566	115908
1980	White Men	451383	804133	417535	424329	2097380
1980	White Women	296628	792951	366706	288578	1744863
1990	Black Men	49054	92850	61153	26591	229648
1990	Black Women	41797	102665	88411	38923	271796
1990	Hispanic Men	97168	71513	50378	20206	239265
1990	Hispanic Women	56353	60227	49585	18198	184363
1990	White Men	275601	794572	668303	526815	2265291
1990	White Women	184978	766198	692206	462415	2105797
2000	Black Men	37835	135743	66406	40522	280506
2000	Black Women	37379	158380	97717	61822	355298
2000	Hispanic Men	152599	142383	58402	33944	387328
2000	Hispanic Women	81853	115708	60884	35562	294007
2000	White Men	196825	978455	555043	615719	2346042
2000	White Women	131188	900990	585423	618360	2235961
2010	Black Men	24582	118256	72780	51198	266816
2010	Black Women	24699	130839	112855	85907	354300
2010	Hispanic Men	135238	167596	82872	55484	441190
2010	Hispanic Women	74481	132403	92845	67616	367345
2010	White Men	121329	893388	591341	744587	2350645
2010	White Women	76263	779487	652105	830034	2337889

Table 1: Sample sizes by year, social identity and educational achievement group. Census and ACS data.