

The Role of Cognitive Frames in Societal Rigidity and Change

Karla Hoff and Joseph E. Stiglitz*

This draft: March 20, 2011

Abstract. All economists agree that perceptions (beliefs) affect actions (choices). We argue here that perception is shaped by cognitive frames that are collectively held and malleable over time and are themselves constrained by fundamental values. Allowing for “equilibrium bias” in perceptions may explain the existence across societies of very different belief systems and cognitive frames. This paper sketches a general approach to societal equilibrium and evolution that clarifies the critical—and unrealistic—assumptions about individual behavior and cognition that underlie what has become the dominant developmental model within the economics literature, based on rational expectations. The central departure from rational expectations in our approach is to assume that individuals start with priors (primitive beliefs) that shape perception and performance.

To illustrate this approach, the study focuses on race—categories into which individuals put themselves and others, which have had and continue to have large impacts on social interactions and economic outcomes. The study shows how an inefficient equilibrium with inequality between different “races” can emerge and persist even if the groups do not differ in any economically relevant characteristic nor face any difference in incentives. Our approach provides an interpretation of both societal rigidity and change. For instance, “racial” differentiation persisted partly because economic agents discarded information that was inconsistent with prior (racist) beliefs. But “deeper” belief systems that limit the set of beliefs that can obtain legitimacy may have a life of their own and lead to situations where commonly used cognitive frames and these deeper beliefs are out of sync. Racial differentiation was incompatible with growing beliefs in equality, with the tension contributing to major societal change.

*A short, more formal treatment of the arguments presented here appears in our paper “Equilibrium Fictions: A Cognitive Approach to Societal Rigidity,” *American Economic Review, Papers & Proceedings*, 100 (May 2010), pp. 141-146. We are grateful to Roland Bénabou, Cristina Bicchieri, Varun Gauri, and seminar participants at Brown University (Economics) and University of California-Berkeley (Sociology) for valuable comments on earlier drafts. Special thanks go to Alana Bevan and Rebecca Wieters for outstanding research assistance.

Table of Contents

- 1. An historical pattern: The creation and entrenchment of racial categories**
 - A. From economics to politics
 - i. The construction of race to justify slavery in North America
 - ii. The British imperial narrative in India
 - iii. The emergence of a racial idiom for Native Americans
 - iv. The “expulsion” of Aborigines from the human race
 - v. Slavery without race in antiquity
 - B. From politics to ontologies
- 2. Early literature on self-confirming racial constructs**
- 3. Psychology: Going beyond the economists’ usual models**
 - A. Confidence-enhanced performance
 - i. Mood
 - ii. Having or not have power
 - iii. Stereotype threat
 - B. Cues and categories
 - C. Confirmatory bias
- 4. A model of inequality based on fictions**
- 5. Equilibrium belief systems and institutional change**
- 6. Conclusion**

At least since the publication of Gunnar Myrdal's *Asian Drama* (1968), it has been recognized that development economics would be greatly enriched by bringing in the notion of what would today be called *cognitive frames* –“largely unconscious perceptual screens that determine how one views or understands a situation” (Elkins and Simeon 1979). Myrdal argued that people in poor countries could be trapped by their attitudes toward life and work, such as the assignment of shame to certain forms of manual work, or the “psychological limitations on [the choice of] policies seeking to equalize opportunities” (cited in Meier, 1976, p. 689-90). In effect, Myrdal argued that *certain social constructs affected behavior and could be part of what we might call a dysfunctional economic equilibrium*. However, he did not address the question of the mechanisms by which such social constructs are created, evolve, or collapse. Nor did he ask, How do we reconcile such constructions of the developmental process with the usual approaches taken by economists? This paper attempts to sketch a general approach to societal evolution that clarifies the critical—and unrealistic—assumptions about individual behavior and cognition that underlie what has become the dominant developmental model within the economics literature.

To illustrate the general approach and to see how it contrasts with more conventional approaches, we focus on the issue of race—categories into which individuals put themselves and others, which have had, and continue to have, large impacts on social interactions and economic behavior. We ask, How can an inefficient equilibrium with high inequality between groups emerge and persist if the groups do not differ in any economically relevant characteristic? This provides a concrete context in which we can show how a more accurate representation of

individuals' shared "mental models" can help explain institutional persistence and change and economic development.¹

Theories of discrimination in the conventional economics literature, based on rational expectations, are theories of statistical discrimination. They do not take into account that the categories taken as given in statistical discrimination models are products of history and that the process of constructing these categories shapes their content, which in turn shapes behavior. Taking these factors into account provides a new approach to explaining the persistence of dysfunctional inequality. We will present a model that is a particular example of the general psychological insights about confirmatory bias and the impact on behavior of social identity.²

But what we are really interested in is a theory of social organization, *i.e.*, of institutions. We will use the term "social construct" to mean anything that is a collective man-made good. The words might suggest a constructing enterprise. However, we do not mean that any such enterprise exists; collective beliefs can emerge endogenously from individual behaviors. Both institutions and collective beliefs ("ideologies") are social constructs. The persistence of an institution depends on the persistence of the beliefs of individuals—beliefs both about how others behave *and* about the legitimacy of the institution; *i.e.* about what is acceptable. *We are concerned with what collective beliefs are an equilibrium, and how they change.*

¹North (2005) also argues that shared mental models distort perceptions and suggests the promise of applying cognitive science to economic development. He argues that the problems posed by the transition from one belief system to another "are at the core of the problems of economic development. There is nothing automatic about such a transition being successful" (p. 44).

² Economists (typically outside the "mainstream") have, of course, been grappling with the problem of beliefs that are *individual* constructs. For example, efficiency wage models show how perceptions of fairness (an individual construct) can affect morale, morale can affect behavior, and this can explain the persistence of dysfunctional inequality (Akerlof 1982 and Stiglitz 1987). Bénabou and Tirole (2002) present a theory of the strategic construction of beliefs about the self, and Bénabou (2008a,b) presents a theory of the construction of beliefs about the role of markets *versus* government.

Economists have well-defined theories of the evolution of other state variables that describe an economic system, such as human and physical capital stocks. Collective beliefs change too, but we do not have a well-formulated theory of how they change. Classical game theory does not provide this; it shows instead how a set of beliefs can be self-confirming because they are consistent with outcomes. Such beliefs constitute a rational expectations (henceforward, RE) equilibrium. Within the standard statistical approaches, Bayesian theory provides a well-formulated framework within which we can assess, in some highly idealized models, whether multiple RE equilibria exist and which ones are stable.³ In general (though not always), there are very few possible equilibria and each is sensitive to any change in parameters. All equilibria depend on only the current economic variables. History does not matter (except in variants of the model in which there are multiple equilibria, in which case history may determine which of the equilibria is selected). Thus these models cannot explain the rigidities that Myrdal described.

Recently, discussions of macroeconomics have taken another approach. They have assigned a role to Keynesian “animal spirits”—emotions that influence confidence—that give almost unfettered scope to changes in beliefs. Anything can happen.

We will argue that while RE equilibria may tie our hands *too much*—the set of possible equilibria is unrealistically restrictive—theories of “animal spirits” tie our hands *too little*. When there is too large a disparity between reality and beliefs, beliefs will change; and the change in beliefs will change behaviors and, thus, the reality. Using the example of racial categories, we illustrate a theory that allows us to demonstrate that the set of equilibria is larger and more robust

³ See *e.g.*, Piketty 1995. Bénabou and colleagues have augmented the standard Bayesian approach with a framework in which individuals strategically choose the probability that they will remember certain signals. Agents take that filtering process into account *ex post* in assessing the reliability of their beliefs. If the gain to the individual from holding distorted beliefs depends on the incentive properties of the institutions that agents face, there will be interactions across agents in cognitive choices that produce ideologies and “groupthink” (see Bénabou 2008b). For example, because of hedonic anticipatory utility, the size of public sector capital will affect individuals’ incentives to hold certain beliefs about government vs. markets. Because capital is a state variable, this model can also explain continuities in beliefs.

than the set of RE equilibria that have been used to explain the unequal treatment of equals. Our approach has implications for both societal rigidity and change.

In this paper, we combine experimental work in psychology and behavioral economics and the economic analysis of behavior with an historical analysis. To set the stage, we begin with an extensive historical discussion that demonstrates that racial categories are a social construct that have had significant economic consequences. Racial categories were created at particular points in time and have persisted. What is remarkable is the similarity in some critical respects of the historical patterns that we relate, which are drawn from different periods and different places. For social scientists, such repeated patterns necessitate explanation. But the patterns that we show are repeated over and over again in history are virtually impossible to explain with an RE model. We know, for example, from experiments with subjects who are junior high school students in India that boys in castes that were traditionally Untouchable can solve puzzles as well as high-caste boys (Hoff and Pandey 2005, 2006, 2011). Yet when students' caste identities are publicly revealed, the performance of the low-caste boys declines in both absolute terms and relative to the performance of the high-caste boys. We infer that in other possible worlds, the Untouchables could have been an equal or dominant group. There are no intrinsic differences in ability. It is clear that a social construct has affected behavior.

In summary, several features distinguish the approach to societal equilibria and change that we develop here from the neoclassical point of view.

1. *Beliefs play an augmented role.* Individual behavior is based on beliefs that are more complex than those reflected in judgments about probabilities of the occurrence of different states of nature. As in an RE equilibrium, in the societal equilibria that we consider beliefs are self-confirming; when they are not, beliefs change. Beliefs affect

behavior, which affects outcomes, which affect beliefs. But unlike beliefs in an RE model, *beliefs in our approach also affect what is perceived*: the categories into which information is placed and how information is absorbed and filtered. We recognize that individuals are not so fine-tuned as to perceive everything. Individuals' presuppositions influence the constructive process entailed in perception, widen the set of possible equilibria, and make them history-dependent.

2. *Many beliefs, including the categories into which information is placed, are social constructions.* Individuals do not choose their "software" in isolation, but within a social context. A set of collective beliefs (ideologies) that serves a society (or some group in society) well under one set of economic circumstances may serve it less well under another. The historical examples below show how, in such situations, beliefs change. In these examples, the beliefs that emerge are advantageous to some dominant group. But it is not as if that group got together and devised a set of beliefs that would accomplish what they sought. Our theory is thus incomplete. But we believe it is a step forward to break out of the mold of rational expectations, in which the variables described above play no role.
3. *The general beliefs about the world are a state variable that determine the set of other beliefs that are acceptable.* A third difference from the neoclassical perspective is that redistributive institutions function because they have "legitimacy," because they are accepted. Acceptance and legitimacy are a function, too, of collective beliefs. An institution, like Jim Crow, may be accepted at one time and not at another. It may be part

of an equilibrium at one time, and not at another.⁴ We recognize that institutions depend not only on economic variables, but also on the set of general beliefs about the world and the social order, which are a state variable. Collective beliefs that emerge in one period shape the set of possible institutions in the next.

An institutional arrangement is an equilibrium if the outcomes lead to beliefs that support that institutional arrangement. *We want to explain with respect to the societal belief structures what the statistical discrimination literature explained with respect to structures governing interactions between firms and workers.* There can be multiple such institutional equilibria, each of which is stable and robust. History can affect the selection of the equilibrium and changes in “reality” (in, say, technology) may eventually upset a longstanding equilibrium.

The next section discusses the evolution of racial categories in four different historical contexts in the post-Reformation period. We identify similar elements in each case and then argue that these elements were not central in antiquity, despite the presence of slavery in that period. We follow that by a brief discussion of the standard literature in economics attempting to explain inequality by statistical discrimination. This sets the stage for our more extended theory, which is divided into three parts. In Section 3, we discuss insights from experiments in psychology. In Section 4, we present a simple model based on the work of Comte and Postlewaite (2004) that shows how we can capture these ideas at a moment in time. In doing so, we provide a new answer to the question, Why is group inequality so persistent? In Section 5, we discuss ideological change. In the conclusion (Section 6), we discuss related literature.

⁴ At one time, some economists suggested that institutions have a simple role in society—to fill in the “holes” in markets, that is, to remedy market failures (see, *e.g.*, North and Thomas 1973). Arnott and Stiglitz (1991) and Hoff and Sen (2006), as well as others, show that non-market institutions purportedly resolving a market failure (like incomplete insurance markets) could, in this sense be dysfunctional—they could lead to Pareto inferior outcomes. More recent literature has highlighted the role of institutions in preserving inequalities, *e.g.*, Sokoloff and Engerman (2000) and Acemoglu et al. (2002).

1. An historical pattern: The creation and entrenchment of racial categories

A. From economics to politics

Whereas *ethnic* differences are conceived as cultural, *racial* differences are conceived as immutable. In many of the examples of the construction of race that we review in this section, we show how the phrase “all men are created equal” was suffused with particular meaning: *e.g.*, all Anglo-Saxon males are created equal. The examples illustrate what Louis Hartz described as a common pattern of racialism in sites of European colonization after the Protestant Reformation. A belief system cannot be imposed, but history shows that meanings do change. As one scholar notes in the context of her work on Native Americans, “belief in the dominance of the white body was not a given but a labored creation of racial ideology” (Chaplin 1997, p. 231).

i. The construction of race to justify slavery in North America

Skin color was not initially an organizing principle in the colonies that became the United States. There were multiple categories of coerced labor, and freedom and slavery were not yet associated with persons of white and black skin. In the 17th century,

“a substantial number of Virginia’s Negroes were free or became free. And all of them, whether servant, slave, or free, enjoyed most of the same rights and duties as other Virginians. There is no evidence during the period before 1660 that they were subjected to a more severe discipline than other servants. They could sue and be sued in court. They did penance in the parish church for having illegitimate children. They earned money of their own, bought and sold and raised cattle of their own. Sometimes they bought their own freedom...As Negroes took their place in the community, they learned English ways, including even the truculence toward authority that has always been associated with the rights of Englishmen. Tony Longo, a free Negro of Northampton, when served a warrant to appear as a witness in court, responded with a scatological opinion of warrants, called the man who served it an idle rascal, and told him to go about his business.” (Morgan, p. 18).

This anecdote illustrates that in mid-17th century Virginia, it was possible for a black to have no sense that his color was consequential or meaningful.

Racist beliefs about blacks emerged as the governing elite in the colonies reacted to new problems and perceived new opportunities. In Bacon's Rebellion in 1676, Virginia's poor (primarily former indentured servants) turned against the elite—the class of men who owned the land and ran the government. Similar rebellions occurred in other colonies, as well. In response to the threat from the landless, governments abridged the liberties of servants and lengthened their terms of indenture. A second development in the late 17th century with important effects on the development of racial ideology was that Virginia's heavy mortality rate fell. Between 1625 and 1660, Virginia was a death trap; its mortality rate was “comparable only to that found in Europe during the peak years of a plague” (Morgan, p. 19). Once that rate fell,

“an investment in slave labor was much more profitable than an investment in free labor; and the planters bought slaves as rapidly as traders made them available...Slaves probably already constituted a majority or nearly a majority of the labor force by 1700.” “As the number of slaves increased, the [Virginian] assembly passed laws that carried forward with much greater severity the trend already under way in the colony's labor laws. *But the new severity was reserved for people without white skin.* The laws specifically exonerated the master who accidentally beat his slave to death, but they placed new limitations on his punishment of ‘Christian white servants’” (Morgan, pp. 25, 26, emphasis added).

An advantage of slave labor compared to English-born indentured servants was that slaves had no rising expectations since “no one had told them they had any rights.” African slave labor provided an opportunity for repression that imported labor from England did not:

“discontent [by African slaves]...could be repressed by methods that would not have been considered reasonable, convenient, or even safe, if applied to Englishmen. Slaves could be deprived of opportunities for association and rebellion. They could be kept unarmed and unorganized. They could be subjected to savage punishments by their owners without fear of legal reprisals. And since their color disclosed their probable status, the rest of society could keep close watch on them. It is scarcely surprising that no slave insurrection in American history approached Bacon's Rebellion in its extent or in its success.” (p. 27)

Penal codes support this assessment. In 1712 South Carolina passed laws that became the “model for slave codes in the South.” They forbade schooling, travelling without permission,

church attendance, land ownership, and marriage across the color line (Fishel and Quarles, 1976, p. 21). As the former slave and abolitionist Frederick Douglass wryly noted, “the ‘inferior’ group was outlawed from bettering itself.”

The idea of race came into its own “at the very point in time when large numbers of men and women were beginning to question the moral legitimacy of slavery” (Lacsh, cited in Fields 1982, p. 152; see also Faust 1981). Out of this process, two fictions emerged. The first was the biological inequality of human beings with black and white skin. The second was that there were only two groups—bright lines between categories were defined (and redefined); a continuum was never recognized.⁵

ii. The British imperial narrative in India

We next use the account in Dirks (2006) of the nine-year trial of Warren Hastings to consider the emergence of a racial idiom for the native population of India during the period when the East India Company consolidated its control over broad parts of the subcontinent. The East India Company in the 18th century was supposed to have been a trading company, with an eastern monopoly vested by Parliament. However, it “had become a rogue state: waging war [almost constantly]..and collecting revenue over Indian territory” that produced massive private fortunes for the traders (Dirks, pp. 9, 13). The activities of the company’s employees contributed to famine conditions in Bengal and “successive crises around the fundamentals of English politics, culture, and society.” The private trade was based on the misuse and abuse of grants, treaties, agreements, and understandings between the Company and rulers in India. In his opening speech at the trial of Hastings in 1788, Edmund Burke declared, “ ‘I impeach [Hastings] in the name of

⁵ Explaining why it was “useful” to have only two groups would take us beyond this limited discussion, but see Rothman (2003).

the English constitution, which he has violated and broken, --I impeach him in the name of Indian Millions whom he has sacrificed to injustice—I impeach him in the name, and by the best rights of human nature, which he has stabbed in the heart’ ” (p. 105). Burke protested against Hastings’ “geographical morality”: “ ‘We think it necessary in justification of ourselves to declare that the laws of morality are the same everywhere’ ” (p. 107).

But over the course of the nine-year trial, perhaps the major spectacle in England in the late 18th century, something happened: “the well-known scandals of the East India company ... became either forgotten or subsumed with the imperial narrative of an exhausted land that virtually invited the British to conquer it....the social, political, cultural and economic buoyancy of India in the 18th century was not just forgotten but suppressed” (p. xii). After the trial, “*scandal itself became institutionalized* [;] it was displaced onto a civilizational map [entailing...] a [broadly] shared condemnation of the Indian present” (p. 24). “Scandal became normalized in the *assumptions and categories of modernity itself*” (emphasis added). The British now controlled not only broad swathes of India, *but also Indian history*. “Built on fabrication, colonial history [used] imputed barbarism [of the people of India] to justify, and even ennoble, imperial ambition” (p. 5). The British came up with the notion that Indians were children in need of supervision. Burke’s rhetoric in the trial exposed the strains of empire-building in the post-Reformation West and contributed to the emergence of “new vocabularies of legitimation and political right to overcome the scruples Burke so eloquently raised when dominion abroad had to be justified” (p. 206). After the trial, there emerged the “justificatory logic that the good despotism provided was much better than the bad despotism Indians had known before conquest” (p. 207). “Drawing a veil across the oppressions of history was necessary in order to sustain the legitimacy of governance as well as of civil society” (p. 350 *n.*

61). The trial ended with the repression of all ambivalence around the colonial guilt that gave rise to the trial in the first place (pp. 85, 355). What emerged was a “race theory that cast Britons and Indians in a relationship of absolute difference” (Dirks, 2001, p. 133).

iii. The emergence of a racial idiom for Native Americans

The examples above illustrate a pattern of racialism in sites of European colonization that entailed the expropriation of labor and land rents. A similar pattern emerged to justify the wholesale expropriation of land in the case of Native Americans and Australian Aborigines.

A racial idiom emerged for Native Americans beginning in the mid-17th century. At that time, it was still common to argue that “Nature knows no difference between Europe and Americans in blood, birth, bodies, &c” (*The tears of the Indians*, 1656, cited by Chaplin 1997, fn. 37). But by the 18th century, the cartographer Bernard Romans argued that it was “by no means reasonable” to designate the Indians as Americans (Chaplin, p. 230). “The English contended that the Indians were ill-adapted to *any* region”—they were a physically inferior race (p. 248, emphasis added). The English

“used that explanation to usurp the very habitat of Americas’ pre-Columbian natives..... English discourses of nature and the human body were fundamental to their imperial project. The proof was in the body.”

“These beliefs [of Indians’ physical inferiority] were not universal, nor did they fail to elicit disagreement...But a good measure of the persuasiveness of the argument for English physical superiority was the fact that *it appeared even in the writings of colonists who were sympathetic to Indians* and who otherwise argued for native hardiness” (Chaplin, pp. 232-233, 250, emphasis added).

iv. The “expulsion” of Aborigines from the human race

In the 18th century, the indigenous people of Australia were classified as British subjects. “The early governors wanted to see them converted to Christianity and farming so that they could be

absorbed, socially if not genetically, into the lower class of the colony—an idea loathed and resisted by every white [in Australia], no matter what his class” (Hughes, 1987, p. 275). The conflict over land rights between settlers and Aborigines resulted in incidents in which settlers shot Aborigines indiscriminately. In 1805, the King’s judge-advocate ruled that because “Aborigines had no grasp of such basics of English law as evidence, guilt, or oaths, they could neither be prosecuted nor sworn as witnesses...And so the best course would be to ‘pursue and inflict such punishment as they may merit’ without the formalities of a trial.” The restrictions on Aboriginal testimony in court meant that evidence from a white witness had to be acquired to prosecute the killing of an Aborigine. Such policies ushered in a new wave of dispossessions and violence. One cattle proprietor remarked that “there was no more harm in shooting a native than in shooting a dog” (Hughes, p. 277). Over time, “Force and punitive expeditions became an *institutionalized* way of dealing with the Aborigines” (Elkin, 1938, p. 364, emphasis added).

For an institution to be legitimate, it needs to have some justification beyond mere convenience. An institution may claim legitimacy on the grounds of its importance to the good functioning of society (Bicchieri, 2006, p. 21). Such claims were commonly made by white settlers in Australia. “The settlers...asserted that they [the Aborigines] could not be civilized” (Elkin 364). The emerging belief system regarded “Aborigines as inferior, unintelligent, lazy, irresponsible grown-up children, acceptable and even lovable, as long as they were ‘kept in their place’” (Hartwig 1972, p. 14). Tatz (p. 97-109) describes over 30 statutes that ascribe particular characteristics to Aborigines, including the inability to manage their own affairs and property, unworthiness for inclusion in industrial awards, inability in relation to certain rules of evidence in courts, and the inability to govern their communities. One historian has characterized the efforts to justify the policies towards the Aborigines as leading to the expulsion of the

“Aborigines from the human race and ... the development of a *Herrenvolk* democracy in Australia” (McQueen, p. 115).⁶

The racial ideology that emerged in Australia, like that emerging in the colonies that became the United States, can be viewed as a response to the tension created by treating humans merely as means to someone else’s ends in the post-Enlightenment period. David Hume (1751, p. 45) was representative of writers in the Age of Revolution when he argued that post-Enlightenment beliefs meant that oppression required a justification and narrowed the basis on which such justification might rest to one based on innate differences in worthiness. In the Age of Revolution, the “presumed incapacity for freedom [of blacks] ...crystallized into a racial one with all its subsequent pseudobiological trappings” (Fields, 1982, p. 162). Montesquieu famously quipped that “It is impossible for us to suppose that these beings should be men; because if we suppose them to be men, one would begin to believe that we ourselves were not Christians” (cited in David, 1999, p 302). This dilemma did not occur in the West before the Protestant Reformation. For instance, as we discuss below, slavery in antiquity was not generally justified on the basis of innate differences between groups of people.

v. Slavery without race in antiquity

Classical sources explained bodily variation in nonracial, environmental terms. In these sources, according to Chaplin (p. 230), “[a]ll humans are potentially the same, and adult specimens differ owing to the physical and cultural milieux in which they take final form. Natural philosophy stressed an underlying, universal human similarity.” That slavery in antiquity did not imply occupational segregation—many slaves served in positions of status and responsibility

⁶ The notion of “pollution” arose as well. For instance, a Chief Justice who later became a Governor-General warned against the “contaminations and the degrading influence of inferior races” (Encel, p. 32).

commensurate with their education and talents—is consistent with that view.⁷ Slaves were obtained through warfare and colonization. Thucydides’ Melian dialogue famously provided this justification for slavery: After colonizing the island of Melos, killing all the men, and selling all the women and children as slaves, the Athenians argued that “right, as the world goes, is only in question between equals in power, while the strong do what they can and the weak suffer what they must (Thucydides, [c. 431 BC; 1951], book 5.89)”.

B. From politics to ontologies

As illustrated above, groups of politically powerless people in the post-Reformation West came to be categorized as innately inferior races. The rules (by creating vast distance between groups and by miscegenation laws) ensured that statistical differences across the groups would persist. The political roots of racial categories surface today in censuses. For instance, Telles (2004) reports that the probability of an interviewer racially classifying a Brazilian woman in the same way she classifies herself is only 22 percent for self-identified black females, which led one reviewer to conclude that “any attempt to make sense of the Brazilian racial landscape must rely heavily on the background factors used to invent race in the first place [and not on phenotype]” (Lieberman, 2008, p. 11).

But how did members of the inferior races view these categories? Historians have suggested a process by which the social constructs were entrenched in individuals’ psychology through symbolism, ritual, and protocol. In Australia, for example,

“More than simply a technique of control and governance, the racialised designation of space [in the 1911 Aboriginals Act] expressed and reproduced social and ontological categories on which colonial society was founded.” (Cowlshaw, 1999, p. 63)

⁷ Russian serfdom provides another example in which members of an oppressed group could be highly educated and serve in high-status occupation and in which oppression was not justified on the basis of innate differences between groups of people.

Rituals reinforced the ideology by humiliating the “inferior” race. Individuals enacted the ideology as a lived practice (Connerton, 1989), and the cultural baggage prevented psychological distancing. No one was allowed to forget who he was in the racial hierarchy. Race was always a salient category. In the Jim Crow South, the unwritten rules that governed day-to-day interactions across the color line have been described “not only as a form of social control but also as a script for the performative creation of culture and of ‘race’ itself (Ritterhouse, 2006). As Charles Evers (brother of the murdered civil rights activist Medgar Evers) explained in his autobiography, “ ‘Our mothers began telling us about being black from the day we were born. The white folks weren’t any better than we were, Momma said, but they sure thought they were’—which Evers implied, amounted to much the same thing” (cited in Ritterhouse, p. 5). Violations of racial etiquette in the US South could lead to lynchings and massacres (Dailey 1997, p. 561). In Brazil, “the freedom of a former slave could be revoked for disrespect towards a former master” (Schwartz, 1987, p. 138; see also Hartwig 1972, p. 14, on Australia).⁸

Part of the power of racial categories depends upon their not seeming to be socially constructed (a related discussion is in Lessig, 1995, p. 951). Although, as we have shown, racial beliefs evolved over time, they came to be represented as fixed and objective. As the saying goes, history is written by the conquerors. One method was to invent history. Hobsbawm and Ranger (1983) describe the invention of “traditions” to give content to the “natural” differences between groups (see also Dirks 2001). Another method was to repress history. For instance, whites in the US South repressed the history of biracial parties in the 1870s-80s:

⁸ The insistence of elites on compliance with rules of etiquette between races can also be understood in the context of Laibson’s cue theory of behavior (Laibson 2001). In his theory, cues are an indirect means to control preferences/behavior and so elites would have an incentive to try to control cues. Strictly enforced codes of etiquette can be understood as active cue-management. See the discussion of related literature in Hoff and Pandey (2011).

“[I]n real political terms an idea dies, in this case the idea of black political participation and rights. ..Clearly, limiting subsequent generations’ worldview by encouraging them to ‘forget’ alternatives is a subtler form of political repression than shooting one’s political opponents. But it is not a trivial form of repression...Here it is worth thinking about the word ‘repression’ in its psychological, as well as political, sense, for to repress successfully is to forget what one is repressing” (Ritterhouse, 9-10).

2. Early literature on self-confirming racial constructs

In economics, a large literature attempts to explain as part of an RE equilibrium the persistence of racial discrimination when there are no inherent differences among groups.⁹ In the statistical discrimination literature, fully articulated RE models have multiple equilibria. There is an equilibrium in which firms believe that individuals grouped into different categories differ in ability and those beliefs shape incentives in ways that make them self-fulfilling. Given everyone else’s beliefs, there are no gains to be reaped from improving the use of existing information.

There is a coordination failure of beliefs. Institutions play no part in this explanation.

The central limitation of that literature is that *it takes the cognitive frame of race as given*, *i.e.*, by assumption, all individuals are members of a racial category and that category is salient.

This literature does not explain why this might be. It thus ignores the question of category formation to which these historical episodes call attention.

If the categories were platonic, and/or if the world were rich in data and the costs of processing were zero, then everyone should have the same beliefs. However, common knowledge is not an accurate description of society. Because people do not have rich enough data (see *e.g.*, the literature on the one-armed bandit problem), several mutually exclusive explanations may be almost

⁹ In one of the seminal works on discrimination in economics, Becker (1957) argued that discrimination could not persist in a competitive economy. By inference, any persistent wage differences reflected inherent differences in ability. The extreme assumptions underlying the analysis served one salutary purpose—eliciting a literature explaining what was wrong with the theory (see, *e.g.*, Phelps 1972, Spence 1973, and Stiglitz 1974).

equally plausible with the data that people have or are likely to obtain. There are multiple possible cognitive frames. Bayesian priors may determine which of those beliefs come to be held.

People build ideologies to resolve these indeterminacies. These ideologies shape both the Bayesian priors and the way that individuals process information. They shape the categories into which the infinite possible sets of information are placed. They even shape what information people “receive” and what they “rationalize away,” which may lead individuals to maintain prior beliefs in the face of disconfirming data (see Rabin and Schrag 1999, discussed further below).

There is an even deeper issue. It is not meaningful to say that an individual conditions his actions on everything that is observable, since he doesn’t know what is observable. What is observable interacts with an *infinite* data set and a *finite* set of socially constructed categories. The category system gives rise to another kind of bias.

The story we tell in this paper goes beyond the literature on statistical discrimination in several ways. First, it observes that in general people do not condition on all relevant variables. *There does not exist a unique, natural set of frames or categories.* Frames are a convention and a consequence of social interaction. A complete model needs to explain the choice of cognitive frames. *Our approach differs from RE in that it assumes that individuals have priors, which might be called belief systems, whose content, persistence, and evolution we try to explain.*

Second, the cognitive frames should be viewed as an equilibrium outcome. There are often ideologies that could lead to equilibria that are favorable to the dominant group, but the elites cannot simply “choose” the cognitive frames that work best for themselves (nor can the nonelites simply choose the beliefs that might work best for themselves). Those in power typically control only some of the determinants of the evolution of beliefs. Moreover, the set of admissible ideologies is constrained by a set of intrinsic values, perhaps based on some

unquestioned truth, to which an ideology must conform in order to be legitimate. This set of values might be called an *überideology*. It may evolve more slowly than the ideology, just as the ideology itself may evolve more slowly than the beliefs that attempt to reconcile observations of reality with the ideology.¹⁰

Race served as a way to structure hierarchy beginning in the post-Reformation period, when the idea emerged that all individuals should have some control over their destiny. In the context of our discussion of race, the *überideology* was that humans could not be treated only as means for someone else's ends. After the spread of Enlightenment ideals, the *überideology* tightened the constraints on the scope for one group to legitimately appropriate the labor or land of another group, since the ideal became that equals should be treated equally. These constraints, combined with the drive for domination by powerful groups, created incentives for these groups to create an ideology that identified primal differences among groups. As Davis (1999, p 303) argues, race became “the central excuse for slavery.” In the model presented in the next section, doctrines of innately unequal groups take on a life of their own, affecting behavior in a way that makes the ideology self-sustaining after discrimination has ended. Because individuals do not process all signals and because ideologies focus attention on certain phenomena, those beliefs will be more stable than would be the case under RE.

3. Psychology: Going beyond the economists' usual models

There is ample evidence that individuals' beliefs about themselves affect their behavior. The question is why. The economics literature suggests several explanations—for example, individuals who think they are being unfairly treated will work less hard, a variant of the

¹⁰ As we explain below, the *überideology* influences the “logic” of analysis.

“efficiency wage” literature (see, e.g., Akerlof 1982 and Stiglitz 1987). Here, we draw on the psychological literature that identifies several other effects.

A. Confidence-enhanced performance

Why would confidence affect performance? Experimental work has identified effects of mood, feelings of having or not having power, and stereotype threat.

i. Mood

Experiments summarized in Compte and Postlewaite (2004) demonstrate that psychological states can affect performance. A good or bad mood is induced in experimental subjects in a variety of ways, *e.g.* by having subjects read aloud sequences of depressed statements like “I feel a little down today” vs. reading aloud neutral statements like “Santa Fe is the capital of New Mexico”; or seeing a sad vs. a happy film clip. The induced negative mood impairs speech articulation, physical reaction time, and the ability to detect contradictions.

ii. Having or not have power

In three experiments in psychology, Pamela Smith *et al.* (2008) find that invoking in an experimental subject the feeling that he has little power impairs his performance in complex cognitive tasks. One way that the subjects were primed to feel more or less powerful was by asking them to unscramble a set of sentences cuing different feelings. In one treatment, the scrambled sentences included words related to lack of power (*e.g.*, subordinate, obey). In another treatment, the sentences included words relating to having power (*e.g.* authority, dominant). In the control treatment, the sentences included no words related to power. Cueing feelings of lack of power in subjects reduced their ability to plan tasks, hold information in

working memory, and inhibit goal-irrelevant information.

iii. Stereotype threat

A robust experimental effect is that cuing an identity associated with a stereotype, or cuing a condition that could confirm a negative stereotype, shifts an individual's performance in the direction of the stereotype (Steele, 2010 provides a survey of the literature; recent work is Hoff and Pandey 2011 and Afridi, Lin and Rei 2011).

There are many possible reasons. Suppose, for example, that there is a norm about performance. An agent chooses a level of effort to exert to accomplish a task that is rewarded. A belief that one is of low ability has a substitution effect (reducing effort) and an income effect (increasing effort). The experimental findings show that the substitution effect of a stereotype dominates.

However, even when there is no norm about how difficult a task is, making stereotypes salient affects performance. Why? One way to explain this experimental finding is by adopting a model of individuals with multiple identities, the salience of each of which can be manipulated.¹¹ In doing that, an individual's prior beliefs about the difficulty of the task and the rewards to accomplishing it change. *Each identity has its own priors, its own heuristics.* In contrast, in a Bayesian model, reminding a person of his race/gender/caste would change

¹¹ This is one of the fundamental differences between the standard economic model and models that have increasingly been found useful in other social sciences. To an economist, an individual has, at any moment of time, a well-defined set of preferences. (Of course, preferences may change, *e.g.* as a result of addiction, and this presents challenges for defining welfare norms.) The well-defined preferences provide all the information that is relevant for describing his behavior, i.e. his choices. By contrast, in some of the psychological and sociological literature, an individual is conceived of as having multiple identities (possible sets of preferences), which can be "expressed" at any point of time. Which is the "real" person? It is possible to describe more fully the behavior of the individual by describing the circumstances in which one kind of behavior or another is exhibited. That is, behavior depends on context. See, for example, Swidler (1986), Glaeser (2004), and Hoff and Pandey (2011).

nothing. Experimental results on stereotype threat show how race/caste/gender/caste salience affects behavior and provide evidence against the unitary self model.

While ideas can give comfort to or legitimize “poor” behavior and affect individual psychology, our concern in this paper is to explain institutional stability and change. *We argue that beliefs should be treated as state variables.* This means that beliefs instilled at one point of time may persist. A set of beliefs designed to extract rents from one group (as in the case of race) affects the future equilibrium *after* the extractive regime is removed. Two features of how the mind works, described below, enable us to link such beliefs and the future equilibrium.

B. Cues and categories

What individuals perceive is not, or not always, the unmediated effect of sensory impressions. A classic experiment in psychology by Bruner and Potter (1964) suggests that pre-conceived ideas serve as unconscious filters of sensory impressions. Experimental subjects were asked to identify an ordinary object, such as a bird or dog, in a video that slowly came into clearer focus. Subjects were divided into three groups. The first group was shown the full-length video. The second group was shown only the latter part of the video, which began with the object in medium blur. The third group was shown only the last part of the video, which began with the object in light blur. At the end of the video, the slides had the identical degree of blur for all three groups, and subjects were asked to identify the object. The results showed that correct identification of the object depends on the starting blur, but not in the way one might expect. Correct identification of the objects was about 20 percent in the group that watched the full video, about 40 percent for the group that started watching the video midway, and about 60 percent for the group that watched only the last part of the video. Bruner and Potter concluded that very blurred slides allowed the perceiver to make an incorrect inference about what the object was and *the*

perceiver held onto the incorrect perception when better information (a more focused picture) became available. Psychologists have replicated the Bruner-Potter result in many domains.

Many presuppositions are broadly shared in a society. The findings of Bruner and Potter suggest that such presuppositions might act as unconscious filters of what is perceived in an entire society. This has implications for societal rigidity. Further, the information that individuals collect is structured. For instance, individuals don't collect data on race if they have a prior that race is irrelevant. Or if they have the data, they don't use it. Agents need a model to put their perceptions into categories and to decide what things are "similar" and important. Agents take for granted certain categories and this affects even what information they gather. Someone who did not think that race was a salient characteristic would not gather or observe information about the relationship between race and performance. There are an infinite number of possible correlations between observables, *e.g.* between performance and eye color, hair color, or shoe size. Individuals do not "gather" information about all these possible correlations because, in a sense, the way we see the world suggests that that they are irrelevant. If we came to believe that they were relevant, they possibly would be.¹²

C. Confirmatory bias

Confirmatory bias is the tendency to search for, interpret, and remember information in a way that supports one's initial beliefs. Experiments demonstrate "how providing the *same* ambiguous information to people who differ in their initial beliefs on some topic can move their beliefs *farther apart*" (a survey is in Rabin and Schrag 1999). Rabin and Schrag also show formally how such *confirmatory bias* can lead to persistent statistical errors, potentially making an initial

¹² Loury (2002) has a related discussion, and Fryer and Jackson (2008) develop a model of biases that can emerge from categorization.

false belief difficult to dislodge even if one saw the outcomes of an optimally designed experiment, repeated infinitely.

A widely shared belief among people is to think too highly of themselves (*e.g.*, Weinberg 2006). For example, after taking a test, experimental subjects who predict that their answers are 90 percent correct typically have only a 60 percent hit rate. Far more than 50 percent of automobile drivers, judges, and stock market analysts rank themselves above average. However, systematic factors—psychological state, gender, and beauty, among other things— influence self-serving bias. Depressed people do not have it (Alloy and Abramson 1979, Ackermann and DeRubeis 1991). High school girls are less (more) likely to have a personal conception of themselves as skilled at mathematical (verbal) tasks than are those high school boys with equal math- (verbal-) relevant grades and test scores (Correll, 2001). This is consistent with the hypothesis that cultural beliefs about a group’s ability in a particular domain influence individuals’ perceptions of their own ability in the domain. Physically unattractive people demonstrate less self-confidence than attractive people, and others also perceive them (wrongly) to be more able (Mobius and Rosenblat 2006).

Since widespread status beliefs may change more slowly than legal and material conditions (Ridgeway 1997), individuals, implicitly acting on group status beliefs, may rewrite group inequality into new conditions (Ridgeway and Smith-Lovin 1999, p. 209). In the following passage, Huey Newton (1968, p. 16), co-founder of the Black Panther Party, describes the affects of the low status of his social identity on the lower socio-economic American black male in the 1960s:

“[He is] a man of confusion. He faces a hostile environment and is not sure that it is not his own sins that have attracted the hostilities of society. All his life he has been taught (explicitly and implicitly) that he is an inferior approximation of humanity...” “[H]is

attitude is that he lacks innate ability to cope with the socio-economic problems confronting him... The parents pass it on to the child and the social system reinforces [it].”

4. A model of inequality based on fictions

This section uses the model of confidence-enhanced performance in Compte and Postlewaite (2004) to show how a fiction can be maintained in equilibrium.

Race. There are two “races,” reds and greens. Every individual is categorized as a red or a green. There is a belief that the reds are a generally inferior.

Technology. Individuals undertake a series of projects at each of which they either fail or succeed. Confidence, based on a person’s perception ϕ of his empirical frequency of past success, increases the probability of success ρ in future attempts in ways that may be beyond his conscious control. Figure 1 shows the relationship, $\rho = \rho(\phi)$, with $\rho' > 0$.

Rational expectations equilibrium. Under the assumptions of RE, a long-run equilibrium would occur at a point of intersection between the 45 degree line and the solid curve $\rho(\phi)$. At such a point, the individual’s actual probability of success is equal to his perceived frequency of success. If the effect of confidence on performance is large enough, there could be multiple equilibria. An equilibrium is stable if the slope of the solid curve in Figure 1, $\rho(\phi)$, is less than the slope of the 45 degree line at that point. Figure 1 shows the case of a unique stable equilibrium (point RE).¹³

Biased perceptions. The equilibrium at the point RE assumes unbiased perceptions. However, as discussed above, it is human for an individual to “forget” – or rationalize as uninformative of future success—a fraction of his failures, while remembering well his

¹³ The comparative statics results for biased perceptions, which we discuss below, hold in a neighborhood of any stable equilibrium.

successes. Let γ_j denote the fraction of failures for a person of race j that are not recollected. The critical hypothesis that we make is that reds are less able than greens to “forget” experiences of failure: $\gamma_{\text{red}} < \gamma_{\text{green}}$. The interpretation of this hypothesis is as follows. Racial categories are salient—the reds see themselves as red, the greens as green (*categories shape the information that individuals collect*). The reds have historically been treated as inferior, and this affects how they interpret their experiences of failure. The reds see failures as confirming the (socially constructed) view of them as inferior. The greens compared to the reds dismiss a larger fraction of their failures because such failures are not consistent with their prior beliefs (*there is confirmatory bias*).

Frequencies of success are random variables, but in the long run a true frequency of success ρ will translate into a perceived frequency of success whose distribution is concentrated around a single value, denoted $\phi(\rho; \gamma_j)$. Figure 2 shows the relationship between the observed frequency of success (on the vertical axis) and the individual’s long-run belief about his probability of success (on the horizontal axis), given his race.

Equilibrium with biased perceptions. In equilibrium an individual holds beliefs that are consistent with the data he observes. This is a kind of “almost RE”: *beliefs match what is perceived given the agent’s perceptual screen*. Figure 3 superimposes the graphs of Figures 1 and 2. Equilibrium is the intersection of the performance function, $\rho(\phi)$, and a perception function given one’s race, $\phi(\rho; \gamma_j)$. In the equilibrium, beliefs generate a level of performance that is consistent with those beliefs. The result is that greens will perform better than reds. A social construct—that there are two hierarchically ranked races—will be self-fulfilling.

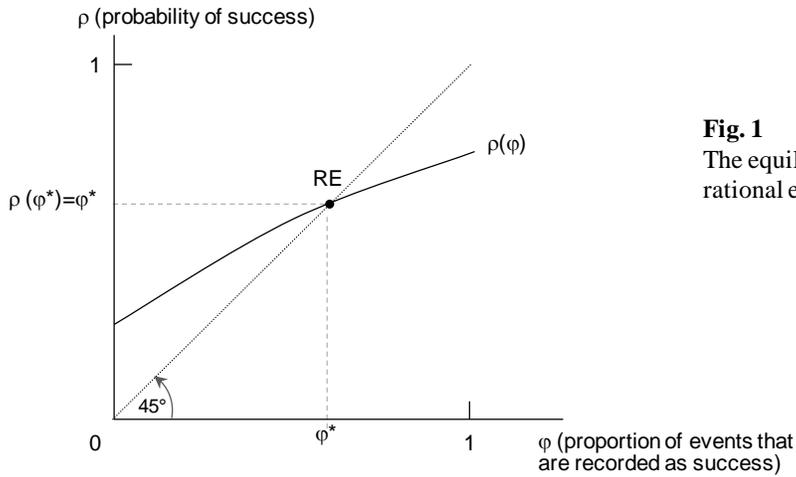


Fig. 1
The equilibrium with rational expectations

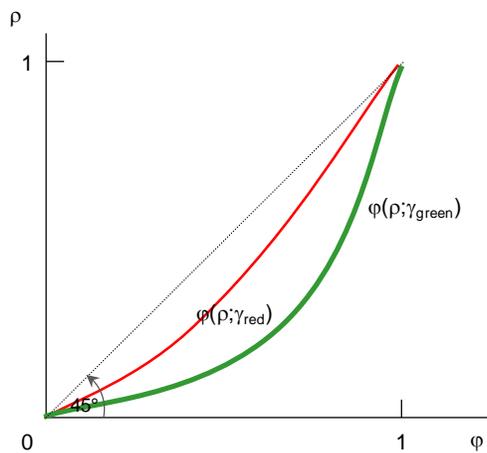


Fig. 2
Biased perceptions

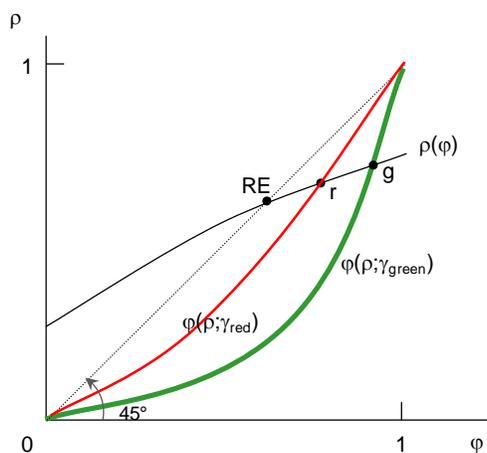


Fig. 3
The equilibrium with biased perceptions at **r** and **g**

Robustness. We now deal with a further question. Is this racial fiction robust? If there is a fixed cost to changing it, we the greens may not want to change it, and the reds will have limited ability to change it. Thus we can explain the persistence of race and racial inequality

long after the legal barriers to equal opportunity have been removed and the conditions that gave rise to them have changed.¹⁴

In the model constructed so far, there are no economic interactions. The reds and greens could be solving puzzles on their own. The only benefit from racially disparate performance is associated with another attribute of individual psychology: individuals' sense of well-being is affected by their relative position. The greens may feel better about themselves if they are doing better than the reds. This alone would give them a reason for reminding the reds of their race.

To construct a robust discriminatory equilibrium, we embed the model into an environment in which now there is a production function that requires skilled and unskilled labor. The more powerful group says to the less powerful group, "you're red." By calling them reds, the first group necessarily calls itself greens, and there is nothing the reds (being the momentarily less powerful group) can do about it. If the resulting loss of self-confidence makes the reds more likely to be complements to greens' labor than substitutes, the greens benefit by reminding the reds that they are reds even after slavery is over and a competitive labor market is reinstated.

What is striking is that this is what happened in the Jim Crow South. van den Berghe (1978, p. 30) describes two types of race relations: (i) a paternalistic system (a master-servant model) and (ii) a competitive system associated with industrial and urban societies. Segregation occurs only in the second type of race relations. The dominant group's image of the subordinate group changes from "one of backward ... children to ... dishonest, underhanded competitors for scarce resources... virulent hatred replaces condescending benevolence."

¹⁴ Note that if the racial system had allowed an overlapping distribution of abilities between reds and greens, all reds could view themselves as relatively superior reds, equal to greens. In that case, differences between reds and greens might not be self-confirming.

It was perhaps not inevitable after the end of slavery in the US that type (i) race relations morph into type (ii). There were periods in the US South before 1900 in which biracial parties existed and blacks exercised equal political rights with whites (Dailey 1997). Jim Crow was an institution that, in effect, forced (or led) even those who might have been reluctant to act in a discriminatory way to do so. What was remarkable was how long the institution lasted. It was a dysfunctional institution—overall societal productivity would arguably been far higher in its absence. But it persisted nonetheless for more than six decades.

5. Equilibrium belief systems and institutional change

Institutions in modern economics are often conceptualized as the “rules of the game” that govern cooperation. The focus is on economic equilibrium: rules, together with technology and information sets, shape outcomes. But clearly, *institutions also shape how individuals see the world, how they receive information, and the categories they use*—they affect beliefs, and those beliefs also affect behavior. We presented above an example of a short-run equilibrium in which, given the lens through which individuals view the world, people generate outcomes that are consistent with their beliefs. This equilibrium departs from RE in just one respect: It posits biased perceptions of which an individual is unaware.

In the long run, there is an equilibrium process that determines what the screen and the institutions that maintain it will be. One can think of institutional behavior and dynamics much as we have thought of individual behavior. In both cases, beliefs are central. Standard theory has emphasized the role of beliefs (or expectations) about the behavior of others. But beliefs also matter because, recognizing that non-conformity will lead to changes in the rules (or in the punishments for rule violators), individuals must decide whether or not to conform to a particular

set of rules. Whether individuals do or not depends on their legitimacy—which is a matter of both habituation and beliefs—often what we might call higher-order beliefs, *i.e.*, ideology.

In judging the legitimacy of a set of rules (an institution), individuals may look at either the outcomes or the processes. A set of outcomes, *e.g.*, the market, that was viewed as legitimate under one technology might be viewed as unacceptable under another if it led, for instance, to greater inequality than current ideologies viewed as acceptable. Something would then have to change: the institutional arrangement or the belief systems.

There is a dynamic process in which institutions at time t together with beliefs and technology at time t shape outcomes at time t , which shape beliefs and institutions at $t + 1$.¹⁵ Institutions shape how we see the world, and this can give them robustness. But the interdependence between institutions and worldviews can also give rise to “institutional breaks” (of which revolutions are one manifestation). Not all individuals process information the same way. Institutions lead some individuals to process information in ways that lead them to support the existing institutions; but others increasingly see discrepancies between the underlying ideology (beliefs) and outcomes that make the institution unacceptable.

Interests play a large role. Those whose interests are served by the current institution are less likely to perceive these disparities.¹⁶ But while technology is one driver of institutional change, beliefs themselves can have their own dynamic. The ideology that “all men are created equal” had overwhelming consequences, especially, of course, in conjunction with changes in technology.

¹⁵ We emphasize in our formalization of this idea (Hoff and Stiglitz 2010) emphasize that beliefs exist at two levels: individuals have beliefs about the nature of society (this is methodological individualism), but those beliefs are turned into social constructs (“belief systems”) through the kinds of mechanisms illustrated above in our example of race.

¹⁶ Unlike some strands in Marxian analysis, institutions cannot be explained *only* in terms of interests. And our theory does not explain changes in institutions *just* in terms of changes in technology. We have argued that beliefs can have a (partial) life of their own.

The story we have told for the development of the institutions surrounding race is a case in point. In time t , there were differences in power between groups. As violence and exploitation became institutionalized, categories of people were defined; each individual was assigned to a category; each category was made subject to a different legal order. *Given the belief that “all men are created equal,” the violence and exploitation endogenously gave rise to a change in cognitive conventions.* In time $t + 1$, each category was represented as *essentially different* from any other. These constructs came to shape how individuals categorized others and themselves, and the confidence that they had in themselves.

Over time, the outcomes of this system increasingly were perceived as disjoint from prevailing ideologies of social justice, at least by those who were not direct beneficiaries of this system. The ideologies of social justice were so ingrained that they could not change; and that meant that eventually the institutions themselves had to change. Yet even after Jim Crow and other institutional arrangements had changed, behaviors based on those institutions persisted because of the persistence of certain beliefs (which were at least partially justified by observed behaviors.)

This helps explain why societal change is so difficult. One can legislate changes in formal institutions, but one cannot legislate change in social status and in other beliefs that support behavioral patterns. If it's the case that mindsets—classification systems, grammars, logics—are state variables of an economy, then institutional change faces a whole new set of obstacles. In our model of persistence, for example, racial identities are a constraint.

This model of institutional change also identifies new opportunities for change that do not exist in the standard model: *e.g.*, changes in perceptions can create ideological strains that

trigger change; what was an equilibrium may cease to be an equilibrium. A politician or a pundit that reframes the way people see the world can help bring about a new institutional dynamic.

The coherence of an institution arises from how it shapes the way people think about new problems. Developing societies constantly confront new problems. Institutions can be thought of as providing a “logic” by which new problems are analyzed and resolved. But besides this “logic” (what might be called the grammar or language within which discourse occurs and differences are resolved), institutions are founded on a story or analogy that lends legitimacy to the institution; it is why individuals come to accept it. The importance of legitimacy has been enhanced over time as members of society have grasped that a given set of institutions is not fixed by nature: there are alternative institutional arrangements that might govern their interactions. Those who are disadvantaged by a particular set of institutional arrangements may thus question their legitimacy.

This paper is, in part, about societal rigidities—non-development—and societal change. Change can occur when the logic of an institution leads to an outcome inconsistent with the underlying “story” or ideology. Robust institutions make it difficult to see these disparities. Worldviews underlying the institutions filter out information that is inconsistent. To use our earlier vocabulary, obstacles arise from *categorization* that shapes the way information is collected and *confirmatory bias* that shapes the way it is registered and remembered. But sometimes, the disparities are so large that at least for certain segments of society, the evidence breaks through these cognitive barriers. *The broad story of history is one of changes in beliefs.* These can drive change in individual behavior both directly and indirectly through the institutions that govern the interrelations among individuals.

We have critiqued other models that try to explain how groups that are inherently the same can obtain different economic outcomes. We showed that, historically, powerful people create a myth (but a myth that, in some sense, turns out to be self-confirming) to sustain power. Why not just create a norm, as in the models of Akerlof (1976) and Abreu (1988)? One objection is this: the norm is self-enforcing only if the interest rate is low, and the equilibrium is not robust to information asymmetries (*e.g.*, not getting caught for not enforcing the norm). A second objection is that, given their beliefs, in the Abreu model it would be in the interest of the reds to secede from their own society. That they do not do so makes the model internally inconsistent.¹⁷

Our analytical approach is more robust. We replace particular assumptions about information asymmetries with particular assumptions about beliefs and behavior. Both, of course, capture elements of the truth. Their effects might be complementary. *But the source of robustness of racial inequality would seem to be better explained in our model than in the existing ones. Neither the existing models (nor any other RE model) would seem to be able to account for the historical pattern summarized in Section 1.* Self-policing in our model occurs at a deeper and more abstract level than that of the enforcement of discrimination, which is the subject of Akerlof's model of caste inequality (Akerlof 1976).

While the approach taken here is outside the norm in economics, it has important antecedents in Myrdal (1944) and in institutional sociology. Mary Douglas (1986) argues that what institutions teach people is how to “see,” how to “think,” what sets of things are “similar,” and what are the important categories. She argues that the theory of rational behavior by which “each thinker is treated as a sovereign individual” has insuperable difficulties. It cannot account

¹⁷ Even if there were economic interdependencies (*e.g.* arising out of complementarities), the threat of secession would affect the extent of equilibrium discrimination that was observed.

for the set of cognitive classificatory frameworks within which choosers choose, which are themselves collective. Institutions are legitimated by some persuasive fiction or analogy: “the entrenching of an institution is essentially an intellectual process as much as an economic and political one...To acquire legitimacy, every kind of institution needs a formula that founds its rightness in reason and in nature” (p. 45).

This tradition of work leads to a richer view of institutions. An institution is something that categorizes and gives meaning to—or provides an interpretation of—action. Thus, an institution operates both as a symbolic system and as a pattern of activity (Friedland and Alford 1991). This view suggests a broader paradigm in studying institutional change than that conventionally followed in economics.

6. Conclusion

The usual economic approach to institutions—focusing on self-enforcing equilibria of behaviors of rational agents—is inadequate to explain institutional change. As Greif and Laitin (2004) argue,

“Institutions influence factors such as wealth, identity, ability, knowledge, beliefs, ... that are usually assumed as parametric in the rules of the game. ... Such influence implies a dynamic adjustment of variables that, if this influence had been ignored, would have been considered as parameters in the stage game.”

Our paper can be viewed as amplifying the research agenda that Greif and Laitin have laid out, focusing on the dynamic adjustment of beliefs, and noting that beliefs affect even what information is gathered. And it is not just the parameters in the game that change over time, but the game itself.

Sociologists (the “New Institutionalists”) also emphasize the cognitive elements in institutions and point to implications for a theory of institutional change (Powell and DiMaggio

1991). For instance, Friedland and Alford (1991, p. 232) write that “[in a given society] institutions are potentially contradictory and hence make multiple logics available.” Taking account of ideology can help explain persistence and change in institutions.

We agree with the institutional sociologists that the cognitive element is critical. People have choices, and the way they perceive their choices depends on social constructs. Economists have left out social construction. But the sociologists have left out that social constructs need to be understood as equilibria. The challenge is to understand the path of social constructs over time.

The overarching theme of this paper is the use of ideology (in particular, in the post-Reformation West) to naturalize grossly unequal relations between groups. We presented a simple model, based on Compte and Postlewaite (2004), in which biased cognition made group inequality persistent even though abilities and opportunities were the same between groups. The ideology took on a life of its own, shaping the equilibrium. In future work we will develop other examples of institutional overhang that depend on biased cognition. Understanding the relationship between cognitive frames (ideologies) and behavior will advance models of development.

References

- Abreu, Dilip.** "On the Theory of Informally Repeated Games with Discounting," *Econometrica*, 1988, 56: 383-396.
- Acemoglu, Daron; Johnson, Simon and Robinson James A.** "Reversal of Fortune: Geography and Institutions in the Making of the Modern World Income Distribution." *Quarterly Journal of Economics*, 2002, 117(2): 1231-1294.
- Ackermann, Ruby; DeRubeis, Robert J.** "Is Depressive Realism Real?" *Clinical Psychology Review*, 1991, 11(5): 565-584.
- Afridi, Farzana; Li, Sherry Xin and Ren, Yufei.** "Social Identity and Inequality: The Impact of China's Hukou System," University of Texas at Dallas, manuscript, 2010.
- Akerlof, George A.** "The Economics of Caste and of the Rat Race and Other Woeful Tales." *The Quarterly Journal of Economics*, November 1976, 90: 599-617.
- Akerlof, George A.** "Labor Contracts as Partial Gift Exchange." *The Quarterly Journal of Economics*, November 1982, 97: 543-69.
- Akerlof, George A.; Yellen, Janet L.** *Efficiency wage models of the labor market*. Cambridge University Press, 1986.
- Alloy, Lauren B.; Abramson, Lyn Y.** "Judgment of Contingency in Depressed and Nondepressed Students: Sadder but Wiser?" *Journal of Experimental Psychology: General*, 1979, 108 (4): 441-485.
- Arnott, Richard; Stiglitz, Joseph E.** "Moral Hazard and Nonmarket Institutions: Dysfunctional Crowding Out or Peer Monitoring?" *American Economic Review*, 1991, 81(1): 179-190
- Becker, Gary S.** *The economics of discrimination*. Chicago: University of Chicago Press, 1957.
- Bénabou, Roland.** "Ideology." *Journal of the European Economic Association*, 2008a, 6(2-3): 321-352.
- Bénabou, Roland.** "Groupthink: Collective delusions in organizations and markets," manuscript, Princeton University, 2008b.
- Bénabou, Roland and Jean Tirole.** "Self-Confidence and Personal Motivation." *Quarterly Journal of Economics*, 2002, 117: 871-915.
- Bicchieri, Cristina.** *The grammar of society: The nature and dynamics of social norms*. Cambridge University Press, 2006.
- Chaplin, Joyce E.** "Natural Philosophy and an Early Racial Idiom in North America: Comparing English and Indian Bodies." *William and Mary Quarterly*, 1997, 44(1): 229-252.
- Compte, Oliver; Postlewaite, Andrew.** "Confidence-Enhanced Performance." *American Economic Review*, 2004, 94(5): 1535-1557.
- Connerton, Paul.** *How societies remember*, Cambridge: Cambridge University Press, 1989.
- Correll, Shelley J.** "Gender and the Career Choice Process: The Role of Biased Self-Assessments," *American Journal of Sociology*, 106 (6), 2001: 1691-1730.

- Cowlshaw, Gillian.** *Rednecks, eggheads and blackfellas.* The University of Michigan Press, 1999.
- Dailey, Jane.** "Deference and Violence in the Postbellum Urban South: Manners and Massacres in Danville, Virginia." *Journal of Southern History*, 1997, 63(3): 553-590.
- Davis, David Brion.** *The problem of slavery in the age of revolution 1770-1823.* New York, Oxford University Press, 1999.
- Dirks, Nicholas B.** *Castes of mind: Colonialism and the making of modern India.* Princeton and Oxford, Princeton University Press, 2001.
- Dirks, Nicholas B.** *The scandal of empire: India and the creation of imperial Britain.* Cambridge and London, the Belknap Press of Harvard University Press, 2006.
- Douglas, Mary.** *How institutions think.* Syracuse University Press, 1986.
- Elkin, A.P.** *The Australian Aborigines.* Sydney, Angus & Robertson, 1938, 4th ed., 1964.
- Elkins, David J.; Richard E. B. Simeon.** "A Cause in Search of its Effect, or What Does Political Culture Explain?" *Comparative Politics*, 1979, 11(2).
- Encel, Sol.** "The Nature of Race Prejudice in Australia," in F.S. Stevens, eds., *Racism: The Australian experience.* New York, Taplinger Publishing Co., Inc., 1971.
- Fields, Barbara J.** "Ideology and Race in American History," in J. Morgan Kousser and James M. McPherson, eds., *Region, race and reconstruction: Essays in honor of C. Vann Woodward.* New York, Oxford University Press, 1982: 143-177.
- Faust, Drew Gilpin.** *The ideology of slavery: Proslavery thought in the antebellum South, 1830-1860.* Baton Rouge, Louisiana State University Press, 1981.
- Fishel, Leslie H.; Benjamin Quarles,** *The black American: A documentary history.* Glenview, Ill.: Scott, Foresman, 1976.
- Friedland, Roger; Alford, Robert R.** *The new institutionalism in organizational analysis.* University of Chicago Press, 1991: 232-263.
- Fryer, Roland; Jackson, Matthew O.** "A Categorical Model of Cognition and Biased Decision Making," *The B.E. Journal of Theoretical Economics*, 2008, 8(1), Article 6.
- Glaeser, Edward L.,** "Psychology and the Market," *American Economic Review, Papers and Proceedings*, May 2004, 94(2): 408-413.
- Greif, Avner; Laitin, David.** "A Theory of Endogenous Institutional Change," *American Political Science Review*, 2007.
- Hartwig, M.C.** "Aborigines and Racism: An Historical Perspective," in F. S. Stevens (ed.), *Racism: The Australian experience*, vol. II, Sydney, 1972: 9-24.
- Hobsbawm, Eric; Ranger, Terence.** *The invention of tradition.* Cambridge University Press, 1983.
- Hoff, Karla; Pandey, Priyanka.** "Discrimination, Social Identity, and Durable Inequalities," *American Economic Review, Papers and Proceedings*, 2006, 96: 206-211.
- Hoff, Karla; Pandey, Priyanka.** "Opportunity is Not Everything: How Belief Systems and Mistrust Shape Responses to Economic Incentives." *Economics of Transition*, July 2005, 13 (2), Special issue on *Institutions and Economic Performance*: 445-472.

- Hoff, Karla; Pandey, Priyanka**, “Names Can Hurt You (by Cueing your Response to Incentives): Experimental Evidence on Identity and Development,” manuscript, World Bank, 2011.
- Hoff, Karla; Sen, Arijit**. “The Kin System as a Poverty Trap?” in *Poverty Traps*, Samuel Bowles, Steven Durlauf, and Karla Hoff (eds.), Princeton University Press, 2006: 95-115.
- Hoff, Karla; Stiglitz, Joseph E.** “Equilibrium Fictions: A Cognitive Approach to Societal Rigidity.” *American Economic Review, Papers and Proceedings*, 2010, 100: 141-146.
- Hughes, Robert**. *The fatal shore: The epic of Australia’s founding*. New York, Alfred A. Knopf, 1987.
- Hume, David**. *An enquiry concerning the principles of morals*. London, A. Millar, 1751.
- Laibson, David**. “A Cue-Theory of Consumption.” *Quarterly Journal of Economics* 2001, 116(1): 81-119.
- Lessig, Lawrence**. “The Regulation of Social Meaning.” *University of Chicago Law Review* 62, 1995, 943: 968-73.
- Lieberman, Evan S.** “The Analytic Pathologies of Distinguishing Race from Ethnicity.” Unpublished manuscript, Princeton University, 2008.
- Loury, Glenn C.** *The anatomy of racial inequality*. Cambridge and London, Harvard University Press, 2002.
- McQueen, H.O.** ““Racism and Australian Literature,” in F.S. Stevens, eds. *Racism: The Australian experience*, Vol. 1, New York, Taplinger Publishing Co. 1971: 115-122.
- Meier, Gerald M.** *Leading issues in economic development*. New York, Oxford University Press, 1976.
- Mobius, Markus M.; Rosenblat, Tanya S.** “Why Beauty Matters.” *American Economic Review*, March 2006, 96(1): 222-235.
- Morgan, Edmund S.** “Slavery and Freedom: The American Paradox.” *Journal of American History*, 1972, 59(1): 5-29.
- Myrdal, Gunnar**. *An American dilemma, Vol 1: The Negro problem and modern democracy*. Harper and Row, Publishers, 1944.
- Myrdal, Gunnar**. *Asian drama: An inquiry into the poverty of nations*. New York, Pantheon Books, 1968, Vol. 1, 2, 3.
- Newton, Huey**. *Essays from the minister of defense*, Black Panther Party, Oakland, CA (Pamphlet), 1968.
- North, Douglass**. *Understanding the process of economic change*. Princeton and Oxford, Princeton University Press, 2005.
- North, Douglass; Thomas, Robert**. *The rise of the Western world*. Cambridge, UK: Cambridge University Press, 1973.
- Phelps, Edmund S.**, “The Statistical Theory of Racism and Sexism,” *American Economic Review*, 1972, 62(4): 659-661
- Powell, Walter W. and Paul J. DiMaggio**. *The new institutionalism in organizational analysis*. Chicago and London, the University of Chicago Press, 1991.

- Piketty, Thomas.** "Social Mobility and Redistributive Politics." *Quarterly Journal of Economics*, 1995, 110: 551-583.
- Rabin, Matthew; Joel L. Schrag.** "First Impressions Matter: A Model of Confirmatory Bias." *Quarterly Journal of Economics*, February 1999: 37-82.
- Ridgeway, Cecilia L.** "Interaction and the Conservation of Gender Inequality: Considering Employment," *American Sociological Review*, 62, 1997: 218-235.
- Ridgeway, Cecilia L.; Smith-Lovin, Lynn.** "The Gender System and Interaction," *Annual Review of Sociology* 25, 1999: 191-216.
- Ritterhouse, Jennifer.** *Growing up Jim Crow: How black and white Southern children learned race*. Chapel Hill, University of North Carolina Press, 2006.
- Rothman, Joshua D.** *Notorious in the Neighborhood: Sex and Families across the Color Line in Virginia*. Chapel Hill: University of North Carolina Press, 2003.
- Schwartz, Stuart B.** "Plantations and Peripheries, c. 1580-c. 1750." In Bethell, Leslie (ed.), *Colonial Brazil*, Cambridge University Press, 1987.
- Smith, Pamela; Nils B. Jostmann; Adam D. Galinsky; Wilco W. van Dijk.** "Lacking Power Impairs Executive Functions." *Psychological Science*, 2008, 19(5): 441-47.
- Sokoloff, Kenneth L.; Engerman, Stanley.** "Institutions, Factor Endowments, and Paths of Development in the New World," *Journal of Economic Perspectives*, 2000, 14(3): 217-32.
- Spence, Michael.** "Job Market Signaling," *Quarterly Journal of Economics*, 1973, 87 (3): 355-374.
- Steele, Claude M.**, *Whistling Vivaldi and other clues to how stereotypes affect us*, Norton, 2010.
- Stiglitz, Joseph E.** "Theories and Policies for the Labor Market," in George M. von Furstenberg, Ann R. Horowitz, Bennett Harrison, eds., *Patterns of racial discrimination: Volume II*, Lexington, MA, Lexington Books, 1974.
- Stiglitz, Joseph E.** "The Causes and Consequences of the Dependence of Quality on Price," *Journal of Economic Literature*, 1987.
- Swidler, Ann.** "Culture In Action: Symbols And Strategies," *American Sociological Review* 1986, 51: 273-286.
- Tatz, Colin M.** "Aborigines: Law and Political Development." In *Racism: The Australian experience*, ed. Frank S. Stevens, Vol 2: *Black versus White*, New York, Taplinger Press, 1972, 8.
- Telles, Edward Eric.** *Race in another America: The significance of skin color in Brazil*. Princeton, Princeton University Press, 2004: 22-23.
- Thucydides.** *The Peloponnesian War*. Trans. Richard Crawley. New York, Modern Library, 1951.
- Van den Berghe, Pierre L.** *Race and racism: A comparative perspective*. New York, John Wiley and Sons, 1957, 2nd ed., 1978.
- Weinberg, Bruce A.** "A Model of Overconfidence," manuscript, Ohio State University, May 2006.