American Economic Review: Papers & Proceedings 100 (May 2010): 141–146
http://www.aeaweb.org/articles.php?doi=10.1257/aer.100.2.141

Equilibrium Fictions: A Cognitive Approach to Societal Rigidity

By Karla Hoff and Joseph E. Stiglitz*

Psychologists, sociologists, and anthropologists have emphasized that the cognitive frames within which people view the world are both collectively held and malleable over time. Category systems have cultural roots and influence what attributes people perceive (Eric Margolis and Stephen Laurence 1999). The cognitive frames operative in a culture unconsciously influence, as well, how people interpret whatever information they register (see, e.g., Marianne Bertrand, Dolly Chugh, and Sendhil Mullainathan 2005). Yet economists have generally neglected the role played by socially constructed cognitive frames. In rational expectations (hereafter, RE), each individual is assumed to use *all* relevant information in an unbiased way.

Recently, economic theorists have investigated the construction of ideologies (see, e.g., Edwin L. Glaeser 2005 and Roland Bénabou and Jean Tirole 2006). They have modeled individuals as trading off the benefits of subscribing to a particular ideology or of suppressing certain kinds of information, against the costs that that entails. Recent discussions of macroeconomics have assigned a role to Keynesian "animal spirits"—emotions that influence confidence—giving almost unfettered scope to changes in beliefs.

All economists agree that perceptions (beliefs) affect actions (choices). We argue here that perception is shaped by cognitive frames. The infinite set of potentially observable data and the infinite ways in which that data could be processed are limited by the finite set of socially constructed categories that are a part of what are called *ideologies*. Incorporating this perspective helps explain why institutional change can be so difficult and societies so rigid. A set of beliefs that may have been functional at one time, but is no longer so, can persist after the economics/technology that had led to the adoption of the beliefs has changed. We show that allowing for "equilibrium bias" in perceptions may explain the existence throughout history and across societies and persistence across time of very different ideologies. This approach allows for a larger and more robust set of equilibria than can be supported by a RE model. On the other hand, the set of equilibria in our approach is much more constrained than the "animal spirits" equilibrium, which presumes that virtually any set of beliefs could be sustained.

While this paper does not provide a fully articulated model of the evolution of ideologies, it does present a framework that suggests some key determinants. Because belief systems affect the equilibrium, e.g., by shaping perceptions, elites have a strong incentive to influence people's beliefs. In contrast, in a RE equilibrium, this is not relevant—cognitive frames play no role. But the elites cannot simply "choose" the cognitive frames that work best for themselves (nor can nonelites simply choose the beliefs that might work best for themselves). The task of "choosing" for themselves and imposing on others' cognitive frames is more complicated and is itself constrained by higher order beliefs. Those in "power" typically do not control all the determinants of the evolution of beliefs. Cultures are always contested.

Incorporating "cognitive frames" (ideologies) as state variables provides part of a general theory of societal change that is markedly different from traditional theories, in which only capital and the distribution of power and wealth are state variables. If ideologies change, the equilibrium can change, with no change in the "fundamentals."

We use the example of the construction of racial categories to illustrate the idea of an equilibrium social construct (ideology). Section

* Hoff: Development Research Group, World Bank, 1818 H St. NW, Washington, DC 20433 (e-mail: khoff@worldbank.org); Stiglitz: Columbia Business School, Columbia University, Uris Hall, 3022 Broadway, New York, NY 10027 (e-mail: jes322@columbia.edu). We thank Roland Bénabou and Varun Gauri for valuable comments, and Alana Bevan and Rebecca Wieters for excellent research assistance. This paper is a much shortened version of an earlier paper.

I discusses historical instances. Section II presents a model in which differentially biased cognition helps sustain the idea of race. Section III formalizes the notion of an equilibrium social construct.

I. Three Examples

A. *The Construction of Racial Ideology to Justify Slavery in North America*[1]

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 seventeenth 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." (Edmund S. Morgan, 1972, p. 18)

Racist codes emerged in the late seventeenth century, in response to changes in the economics of slavery. After 1660, Virginia's death rate, which had been "comparable only to that found in Europe during the peak years of a plague," fell sharply. Two consequences were that "an investment in slave labor was much more profitable than an investment in free labor," and the engrossment of Virginia's lands kept many former servants landless (Morgan pp. 19, 25). In Bacon's Rebellion of 1676, Virginia's "Poore Endebted Discontented and Armed" turned against the elite in a plundering expedition that spread across the entire state (p. 22). The fear of unrest contributed to the decline in the reliance on indentured servants and to the abridgement of the liberties of Africans. "To keep as slaves black men who arrived in that condition *was* possible and apparently regarded as plain common sense" (p. 25). In 1712, South Carolina passed laws that became the model for slave codes in the South. They forbade schooling, church attendance, land ownership, and marriage across the color line. Given emerging Enlightenment ideologies of equality and human rights, oppression needed justification. Out of this process, two fictions emerged. The first fiction was the biological inequality of human beings. The second was the natural discontinuity between the categories "white" and "black."

B. *The British Imperial Narrative in India*

The East India Company in the eighteenth century "had become a rogue state: waging war … and collecting revenue over Indian territory" that produced private fortunes and contributed to famine in Bengal in 1770 (Nicholas B. Dirks 2006, p. 13, 252). Parliamentary inquiries brought the scandal to national attention. They culminated in the trial of the governor of India, Warren Hastings. In his opening speech, Edmund Burke declared, "I impeach him in the name of the English Constitution, which he has violated and broken,—I impeach him in the name of Indian Millions, whom he has sacrificed to injustice." "We think it necessary in justification of ourselves to declare that the laws of morality are the same everywhere" (Dirks, pp. 105, 107). But over the nine years of the trial, the idea that British law applied to agents of Britain in India was salvaged not by finding Hastings guilty (he was acquitted) but instead by inventing a new interpretation of India. "Built on fabrication, colonial history imputed barbarism to justify, and even ennoble, imperial ambition." "Scandal became normalized in the assumptions and categories of modernity itself" (Dirks, pp. *xii,* 5, 29). A "race theory emerged that cast Britons and Indians in a relationship of absolute difference" (Dirks 2001, p. 133).

C. *The "Expulsion" of Australian Aborigines from the Human Race*

Aborigines were classified as British subjects. "The early governors wanted to see them converted to Christianity and farming …—an idea loathed and resisted by every white [settler], no matter what his class" (Robert Hughes 1987, p. 275). One historian characterizes the arguments used by the settlers to justify their policies towards the Aborigines as entailing the expulsion of the Aborigines from the human race (Humphrey McQueen 1971, p. 115). Australia

[1] 1 A more detailed discussion is contained in our Web Appendix http://www.aeaweb.org/articles.php?doi=10.1257/aer.100.2.141.

enacted statutes that characterized Aborigines as unable to manage their own affairs and property, unable to understand rules of evidence, and unable to govern their communities.

D. *Ontologies*

Given the beliefs in Europe that emerged with the Protestant Reformation and the Enlightenment, oppression both called for a justification and narrowed the basis on which such a justification might rest to one based on the slaves' or the natives' innate inferiority (David Hume 1751, p. 88), a "presumed incapacity for freedom." This did not occur, or at least not to the same extent, in earlier periods and for other regions. In antiquity, natural philosophy in general recognized an underlying, universal human similarity (Chaplin, p. 230) but saw no wrong in enslaving prisoners of war and colonized peoples. When the Athenians colonized the island of Melos, they killed all the men, sold the women and children as slaves, and justified their action this way: "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, Melian dialogue).

But after the fifteenth century, racial categories were constructed and hierarchized and given precedence over other aspects of belief systems in which oppression might be viewed as unacceptable. The racial categories were entrenched and embedded in individuals' minds through ritual and protocol. In Australia, "the racialized designation of space expressed and reproduced social and *ontological categories* on which colonial society was founded" (Gillian Cowlishaw 1999, p. 63). In the Jim Crow South, the unwritten rules that governed day-to-day interactions across race lines have been seen "not only as a form of social control but also as a script for the performative creation of culture and of 'race' itself" (Jennifer Ritterhouse 2006, p. 4). 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 in practice" (cited in Ritterhouse, p. 5). Histories in which there had been an alternative to white supremacy were repressed.

II. A Model of Inequality Based on Fictions

This section expands on Olivier Compte and Andrew Postlewaite (2004)'s model of confidence enhanced performance and biased perception to show how a fiction can be maintained in equilibrium. The model rests on three hypotheses, each supported by empirical evidence. The first is called *confirmatory bias*: People tend to misread evidence as additional support for initial hypotheses. 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*" (see Matthew Rabin and Joel L. Shrag 1999). This bias can also explain how individuals with favorable prior beliefs about themselves can maintain an inflated view of themselves.

We call the second hypothesis *preconfirmatory bias*: When people draw inferences, they begin with a specification that posits certain categories, e.g., racial categories, that they do not individually choose (Glenn Loury 2002). The third hypothesis is that self-confidence boosts performance. For example, in three separate experiments, Pamela K. Smith et al. (2008) find that the perception that one is low in the social hierarchy lowers performance in complex tasks.

A recent set of experiments with junior high school boys in India illustrates the interaction among these hypotheses. Children in castes that were traditionally Untouchable can solve mazes as well as high caste children (Hoff and Priyanka Pandey 2006). Yet when students' caste identities are publicly revealed, the performance of the low caste is reduced both absolutely and relative to the performance of the high caste. We infer that in other possible worlds, the Untouchables could have been an equal or dominant group. It is clear that a social construct has affected behavior.

In our model, individuals undertake a series of projects, at each of which they can fail or succeed. Confidence, based on a person's perception φ of his empirical frequency of past success, affects the probability of success ρ in future attempts, in ways that may be beyond his conscious control. Figure 1 shows the relationship $\rho = \rho(\varphi)$. Under the assumptions of RE, there would be a unique equilibrium at the point marked RE, where the 45 degree line intersects the curve $\rho(\varphi)$.

However, this will not be an equilibrium if an individual "forgets," or rationalizes as uninformative of future success, his failures with probability $\gamma > 0$. 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 value denoted $\varphi(\rho; \gamma)$.

There are two races, red and green. The critical hypothesis is that reds are less able to "forget" experiences of failure: $\gamma_{\text{red}} < \gamma_{\text{green}}$. The interpretation of this hypothesis is as follows. *Preconfirmatory bias* means that racial categories are salient—the reds see themselves as red, the greens as green. The reds have historically been treated as inferior, and this affects how they interpret experiences of failure. The reds see failures as confirming their (socially constructed) perceptions of inferiority, while the greens dismiss a large fraction of their failures because such failures are not consistent with their prior beliefs (*confirmatory bias*). Equilibrium is the intersection of the production function, $\rho(\varphi)$, and a perception function, $\varphi(\rho; \gamma)$. In the equilibrium, beliefs generate a level of performance (g or r) that is consistent with those beliefs. The equilibrium is stable since $\rho < \varphi$.

III. The Dynamics of Changes in Beliefs

The preceding section modeled a short run equilibrium, given a social construct, namely, a racial ideology. Here we formalize the idea of an *equilibrium social construct*. There are three levels of "cognition" or belief systems in the model: (i) a "lens" S (called ideology), through which individuals process data; (ii) using that lens, a set of beliefs about particulars, e.g., I am competent, which we denote by A_i for agent i; and (iii) an überideology (U) that affects individuals' judgments about the adequacy of their ideology. Letting boldface letters denote vectors (so, e.g., A corresponds to the vector $(A_1, A_2, A_3, \ldots)$), the structure at time t is

$$\boldsymbol{A}_t = F(\boldsymbol{Q}_t, S_t),$$

$$\boldsymbol{Q}_t = G(\boldsymbol{X}_t, \boldsymbol{A}_t),$$

$$\text{and } \boldsymbol{X}_t = H(\boldsymbol{A}_t, S_t).$$

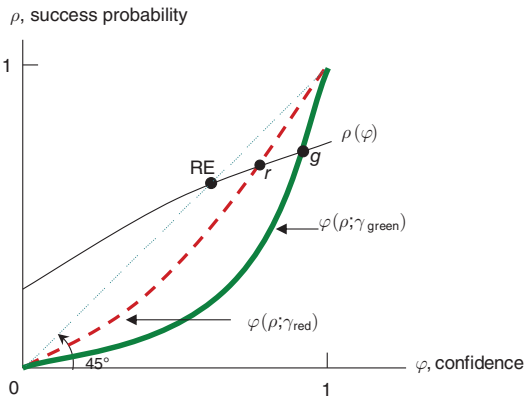

FIGURE 1. EQUILIBRIUM WITH RATIONAL EXPECTATIONS (AT RE) AND WITH BIASED PERCEPTIONS (AT POINTS R AND G).

The *short run* equilibrium modeled in the previous section takes the state variable S_t as fixed. Dropping the subscript t on S, an individual's particular beliefs at time t ($A_{i,t}$) are a function of outcomes $\boldsymbol{Q}_t$ and the "lens" S. Outcomes are a function of behaviors by each agent (denoted $X_{i,t}$) and particular beliefs. And individual behaviors are a function of the particular beliefs as mediated through the lens. For example, in the Indian experiments in puzzle solving, the "behavior" may have been effort. Particular beliefs $A_{i,t}$ may affect outcomes in other ways, too: Low caste individuals may generate worse outcomes, e.g., number of puzzles solved, when their caste identities are publicly revealed, because of anxiety or other psychological factors.

The second and third equations can be combined to give $\boldsymbol{Q}_t = G[H(\boldsymbol{A}_t, S), \boldsymbol{A}_t] \equiv Z(\boldsymbol{A}_t, S)$. *A short run equilibrium* (e.g., *an equilibrium fiction*) *is defined by a value of* A *that satisfies, for a particular S*:

$$\boldsymbol{Q}^* = Z(\boldsymbol{A}^*, S) \text{ and } \boldsymbol{A}^* = F(\boldsymbol{Q}^*, S),$$

that is, people will generate outcomes that are consistent with their beliefs. This equilibrium departs from rational expectations in just one respect: *It posits biased perceptions of which an individual is not aware.*

There is a dynamic process,

$$S_{t+1} = \chi(S_t, U_t, T_t, V_t),$$

where U_t is the überideology, T_t is the "economic regime" (technology and wealth distribution),

and $\mathbf{V}_t$ is a vector of other variables, e.g., encounters with other civilizations and their belief systems. We will not model in detail the evolution of ideologies (S_t). That process, and even more so the evolution of the überideology, are idiosyncratic.[2] Losing a war might lead to a change in views about racial superiority. An encounter with a different civilization—people who see the world through a different lens—might make individuals aware of the possibility that their own lens is distorted. There may be pressures for ideologies to change when the elites discover that there is another belief system that does a better job in maximizing their well-being. In a democratic society, pressures for change may be affected by the "median voter" as much as by elites. But there is no reason to believe that belief systems evolve *as if* they were chosen by a critical decision maker. They may be part of an inefficient Nash equilibrium, *one that does not serve well the interests of any group.*

Ideologies and individual observations of outcomes, themselves mediated through ideology, can provide a set of "conceivable" belief systems, among which individuals choose: $S_{i,t+1}$ is chosen by individual i at time $t + 1$ from a set $\Omega(S_t, \boldsymbol{A}_t, U_t, \Lambda(\boldsymbol{Q},\boldsymbol{A},S))$, where the set is defined by the previous ideology and beliefs, the überideology, and the perception of outcomes Λ as mediated through ideology and beliefs. This is methodological individualism. These "choices" are affected by the social construct. But these individual "choices" get aggregated into a new social construct: $S_{t+1} = \hat{\Phi}(S_{1,t}, S_{2,t}, S_{3,t}, \ldots S_{i,t}, \ldots, U_t, T_t, V_t)$. There is a *medium term* equilibrium in which there is no tendency for society's belief system to change. The beliefs that we model here *occur at two levels*: individuals have beliefs about the nature of society, racial differences, etc., but those beliefs are turned into social constructs ("belief systems") through the kinds of mechanisms illustrated in Section I, including legislation and rituals. What individuals can conceive of is affected by the prevailing ideology. To consider this in the context of race, it is difficult in a racist society not to use "racial categories." But one can still believe that racial categories should be outlawed and that those who are disadvantaged by them would "choose" such a frame. The *societal* frame may, however, not reflect the views of the "reds"—the relevant frames are dictated by laws and systems of information gathering, and those may be controlled by the greens. Alternatively, the vast mass of individuals who neither benefit nor lose from racism may "switch" their beliefs in a process of contagion—when a racist marries a liberal, their children may become liberal. The überideology of equality may make racist categorizations unacceptable to these individuals. That could lead to a system in which there is a sudden tilting of the societal perception when those who find a racist categorization unacceptable become a majority.

[2] For instance, notions of equality may have been championed by new economic interests to promote free markets, but the evolution of these ideas may also have been affected by the evolution of deeper philosophical or religious ideas. The framework we set forth is consistent with these alternative interpretations.

IV. Conclusion

Ideologies (belief systems, including those that determine the admissible categories, which structure individuals' perceptions) are constrained by "higher order beliefs." Beliefs about the acceptability of slavery were affected by the überideology of equality. Notions of equality made it necessary to invent and institutionalize various racial categories to justify certain economic relations. But changed notions of equality eventually made the notion of race as a basis of slavery—and eventually as an admissible category for other purposes—unacceptable.

We don't explain changes in überideology, just as we don't explain changes in technology. Changes arise partly from internal dynamics—a dialectic of ideas—and partly from other changes in economics and society. "Motivated beliefs" may help explain change, but not in the standard way. The Protestant Reformation emphasized individuals' relationship with God, not mediated by authority; and the equality of men at least arguably evolved out of these religious conceptions. Ideas have their own dynamic. For economists to ignore the factors that affect how we process information as part of the interpretation of economic change would be as wrong as to ignore the evolution of technology itself.

We agree with the sociologists and anthropologists on the need to incorporate social constructs (belief systems) into our models of institutions and development. But these scholars

have left out that social constructs need to be understood as an equilibrium. Ideology is a state variable in our theory of societal change and development. The theory helps explain societal rigidities. No one (including the government) can manage changes in belief systems. Untouchability, for example, has persisted, at least in some parts of India, despite its abolition. At the same time, the approach of this paper provides elements of a theory of societal change. Technology, contacts with the outside world, endogenous changes in power and wealth, matter not just directly but because they can lead to changes in ideology.

REFERENCES

Bénabou, Roland, and Jean Tirole. 2006. "Belief in a Just World and Redistributive Politics." *Quarterly Journal of Economics*, 121(2): 699–746.

Bertrand, Marianne, Dolly Chugh, and Sendhil Mullainathan. 2005. "Implicit Discrimination." *American Economic Review*, 95(2): 94–98.

Compte, Olivier, and Andrew Postlewaite. 2004. "Confidence–Enhanced Performance." *American Economic Review*, 94(5): 1536–57.

Cowlishaw, Gillian. 1999. *Rednecks, Eggheads, and Blackfellas: A Study of Racial Power and Intimacy in Australia.* Ann Arbor, MI: University of Michigan Press.

Dirks, Nicholas B. 2001. *Castes of Mind: Colonialism and the Making of Modern India.* Princeton, NJ: Princeton University Press.

Dirks, Nicholas B. 2006. *The Scandal of Empire: India and the Creation of Imperial Britain.* Cambridge, MA: Harvard University Press.

Glaeser, Edward L. 2005. "The Political Economy of Hatred." *Quarterly Journal of Economics*, 120(1): 45–86.

Hoff, Karla, and Priyanka Pandey. 2006. "Discrimination, Social Identity, and Durable Inequalities." *American Economic Review*, 96(2): 206–11.

Hughes, Robert. 1987. *The Fatal Shore: The Epic of Australia's Founding.* New York: Knopf.

Hume, David. 1751. *An Enquiry Concerning the Principles of Morals,* ed. Tom L. Beauchamp. Oxford: Oxford University Press, 1998.

Laurence, Stephen, and Eric Margolis, eds. 1999. *Concepts: Core Reading.* Cambridge, MA: MIT Press.

Loury, Glenn C. 2002. *The Anatomy of Racial Inequality.* Cambridge, MA: Harvard University Press.

McQueen, Humphrey. 1971. "Racism and Australian Literature." In *Racism: The Australian Experience, Vol 1,* ed. Frank S. Stevens, 115–22. New York: Taplinger Press.

Morgan, Edmund S. 1972. "Slavery and Freedom: The American Paradox." *Journal of American History*, 59(1): 5–29.

Rabin, Matthew, and Joel L. Schrag. 1999. "First Impressions Matter: A Model of Confirmatory Bias." *Quarterly Journal of Economics*, 114(1): 37–82.

Ritterhouse, Jennifer. 2006. *Growing up Jim Crow: How Black and White Southern Children Learned Race.* Chapel Hill, NC: University of North Carolina Press.

Smith, Pamela, Nils B. Jostmann, Adam D. Galinsky, and Wilco W. van Dijk. 2008. "Lacking Power Impairs Executive Functions." *Psychological Science*, 19(5): 441–47.