A Theory of Social Motivation

Robert J. Akerlof*

October 5, 2008

Abstract

This paper develops a theory of social norms: what they are, how they form, and how they change. The theory also makes predictions about group formation, categorization, and discrimination, and it can be extended to model leadership and fairness. The paper accounts for the existence of "oppositional culture," where minority groups disparage the majority and are disparaged in turn. An explanation is suggested for the rise of an African-American oppositional culture of this sort in the late 1960s and early 1970s. Other implications of the model are also elaborated. According to the theory, social norms can form because they support cooperation between individuals. An aspect of how people categorize themselves and others—called "comparative fit"—is explained by the theory.

JEL Codes: D00, D60, D70, D80, J71.

Keywords: Authority, categorization, discrimination, fairness, ideals, identity, leadership, norms, oppositional culture, peer effects, status.

*Department of Economics, Harvard University, 1875 Cambridge Street, Cambridge, MA, 02138. email: akerlof@fas.harvard.edu.
I am especially grateful to Philippe Aghion, George Akerlof, Oliver Hart, Rachel Kranton, Andrei Shleifer, and Janet Yellen for their encouragement and suggestions. I would also like to thank Daniel Benjamin, Edward Glaeser, Bob Gibbons, Richard Holden, Lawrence Katz, Ted O'Donoghue, and seminar participants at Harvard University and Cornell University for helpful comments and discussions.

1 Introduction

There are many economic phenomena that cannot be explained without an understanding of social norms: what they are, how they form, and how they change. While the other social sciences–particularly sociology–have devoted considerable attention to social norms, economists are only beginning to appreciate their importance.

Social norms are crucial to the theory of the firm, for example. They potentially constitute a method for incentivizing workers: a boss may be able to manipulate the norms of the firm, making workers feel that they *should* work.[1] Norms can mitigate the moral hazard problem. As Alvin Gouldner writes in his case study of a gypsum plant, "There is...one basis on which the supervisor could feel confident that workers would do their jobs even when he was *not* around; that is, if the supervisor believed that workers themselves wanted to do what was expected of them."[2][3]

This paper attempts to build a theory of social norms. The focus is on explaining sociological phenomena. But, the hope is that the framework developed in this paper can be used to address *economic* phenomena. Some economic implications will be discussed later in the introduction and in the body of the paper.

Some of the sociological phenomena that are explained can also be explained by other models, such as peer effect models, identity models, and motivated belief models (see the discussion of related literature below). For example, the theory predicts that social interaction will usually lead individuals to behave in a more similar manner (see Proposition 6): a prediction of peer effect models. The model presented here actually nests versions of these other models.

There are also predictions that are unique to this theory. The paper attempts to highlight these predictions. One of the unique predictions is that "oppositional cultures" can form. An oppositional culture is a minority of a population with norms of behavior and beliefs that differ from the norms and beliefs of the majority, minimizes its social interaction with the majority, looks down on the majority, and is looked down on by the majority in turn (see Sections 3.1 and 5). Another unique prediction is

[1] See Akerlof and Kranton (2005), and Besley and Ghatak (2005).

[2] Gouldner (1954), p.160.

[3] There are additional reasons that social norms can reduce moral hazard, which will be discussed in Section 3.2. Workers are usually in a better position to monitor their fellow workers than the boss. To the extent that the boss can get workers to monitor and punish one another, this mitigates moral hazard. If a worker feels that workers *should* work, she will feel that a shirker is behaving inappropriately and deserves to be punished. As a result, she might choose to discipline a shirker or bring a shirker to the boss's attention. She is also motivated to monitor other workers because if she is the only person not shirking, she will feel like a "chump."

that social norms may arise because they support cooperation between individuals in a population (see Section 3.2). This relates to Hobbes' notion of a "social contract."

Principles of the theory. The theory is based upon three principles. Section 2 will build a model that incorporates these principles. They are as follows: (1) A person chooses her beliefs. In particular, a person chooses beliefs about how she should and should not behave. We will call such beliefs "ideals."[4] And, we will refer to a widely held ideal as a "social norm." To the extent that a person fails to live up to her chosen ideals, she loses utility. (2) A person gains utility when her beliefs are "confirmed" and loses utility when her beliefs are "disconfirmed." Confirmation can take place in two ways: (a) *Information-based confirmation*: a person's beliefs are confirmed when they are relatively consistent with the information she possesses and disconfirmed insofar as they are inconsistent.[5] (b) *Social confirmation*: a person's beliefs are confirmed by interaction with people with similar beliefs and disconfirmed by interaction with people with dissimilar beliefs. A person generally has some ability to choose the people with whom she interacts.[6] (3) A person gains utility from thinking of herself as high status relative to others. A person's view of her status depends upon how well she meets her ideals relative to how well others meet her ideals.[7]

According to these principles, an individual's choice of beliefs, behavior, and social interactions depend upon the social environment. The social environment also changes when people choose different beliefs, behavior, and social interactions. Therefore, the beliefs, behavior, and social interactions of a population can be in or out of equilibrium. We will refer to a stable configuration as a "social

[4] Most of the time, people are hardly aware of following ideals or applying ideals to others. However, these codes are much more constraining than one might think given our general lack of awareness of them and the ease with which we meet them. "Breaching" experiments are designed to show the importance of such codes, such as a classic one in which Harold Garfinkel asks a class of students to act as lodgers when they return to their parents' homes. See Garfinkel (1963).

[5] It will be assumed that there is no information that indicates the "right" ideals.

[6] A deeper view of social confirmation will be offered in Section 6.1. It will be suggested that social confirmation can be thought of as a concern about the status assigned to one by those with whom one interacts.

[7] George Homans writes that: "the person of the highest social rank [in a group] is the person who comes closest to realizing in his behavior the norms of the group." (Homans (1950), p. 426.) Judson Mills (1958) provides some experimental evidence for principle 3. Mills measured the attitudes of a group of sixth graders toward cheating and then had them take a competitive exam, with prizes promised for the winners. The exam was set up so that it was nearly impossible to win without cheating. Some of the students cheated while others did not. The following day, Mills again measured attitudes toward cheating. He found that those children who cheated developed a more lenient attitude while those who resisted the temptation became more critical.

equilibrium."[8][9] The paper will attempt to describe the social equilibria that can arise.

The reader may wonder whether it is necessary to incorporate all of these principles. In fact, the most interesting results of the paper require the incorporation of all three. This is shown formally in Appendix E.

Some implications of the principles. In contrast to peer effects models, which view individuals as having a desire to conform to the behavior of others, the model developed in this paper views individuals as having a desire to conform to the *beliefs/ideals* of others (principle 2b).[10] Peer effects models generally assume that people conform to average or median behavior. But, reflection suggests that people do not want to be "average." Indeed, the term is a pejorative. Rather, people want to be stars, which is achieved by meeting an ideal.

It might seem worrisome to assume that people conform to what others believe, since beliefs cannot be observed. But, there is a wealth of evidence suggesting that people have a remarkable capacity to glean the beliefs of others from their behavior. This ability is often referred to as a "theory of mind." An example will help to illustrate this capacity. An American child whose parents speak English with an accent will learn to speak without an accent, even though much of a young child's social interaction is with its parents. This finding suggests considerable sophistication on the part of the child. It suggests that the child appreciates that there is an ideal way to speak in her social context and is aware that the parents subscribe to this ideal but have difficulty meeting it.[11]

In addition to the desire to *conform* to the beliefs and ideals of others, there is also a desire to *rebel*

[8] Sociologists have understood that "structure"–essentially equivalent to beliefs in our terminology–forms through such a process. Anthony Giddens and Pierre Bourdieu are the sociologists most associated with this idea (see Bourdieu (1977) and Giddens (1984).) George Homans (1954, ch. 12) gave an early formulation of the concept. He writes that one school of sociology (associated with Durkheim) had focused on the response of individuals to existing social codes and beliefs. Another school (associated with Hobbes) had instead focused on the construction of codes. He explains why these two schools are perfectly compatible–two sides of the same coin.

[9] It is potentially interesting to view social interaction as affecting an individual's "frame." One can view the process described in this paper as describing equilibrial frames. A critic of framing effects might be inclined to discount the economic importance of framing on the grounds that it only adds variance to decision-making and does not affect mean behavior. Framing as discussed in this paper does more than add noise to decision-making. The mathematical reason for this is that there is feedback between individual and environment.

[10] See, for example, Jones (1984), Bernheim (1994), Akerlof (1997), Becker and Murphy (2000), Glaeser and Scheinkman (2000), and Glaeser, Sacerdote, and Scheinkman (2002).

[11] See Judith Harris (2006), p. 129. Also see Baron-Cohen (1995), who discusses the talent that people have for forming "theories of mind." He argues that autism is a disorder that impairs the ability to figure out what others believe, think, and feel. Consistent with this theory, autistic children tend to speak with their parents' accents. Recent brain imaging studies have identified several regions which appear to play crucial roles in forming theories of mind. See Saxe (2006) for a review.

against them. For example, suppose everyone in a population subscribes to the ideal of being a good athlete. An individual who is not an exceptional athlete will be unable to think of herself as high status (principle 3). Suppose this individual deviates, choosing instead the ideal of being a good scholar. The individual might find it easy to become a better scholar than others since others are not focused on this goal. An individual who cares about scholarship and is a better scholar than others can think of herself as high status relative to others (principle 3). Therefore, rebelling against the commonly held ideal allows the individual to improve her opinion of her status.[12][13] This explains the existence of oppositional cultures.[14][15]

There are two rather unusual aspects of the theory that are worth highlighting. The first is that ambiguity tends to have social value because it allows people to maintain high opinions of themselves. For example, the theory predicts that considerably more than a majority of people will view themselves as above average with respect to desirable traits when there is a paucity of information to indicate otherwise. This is empirically confirmed by numerous studies (see, for example, Svenson (1981)).

The second unusual aspect of the theory is the existence of network effects. Principle 2b suggests that an individual may hold a belief or an ideal because others hold it. For this reason, there may be a dense set of equilibria.

There are two empirical reasons to believe in network effects. First, network effects yield the prediction that people will tend to clump on particular ideals and beliefs. The model does not predict complete uniformity of belief, but rather that, often, subsets of the population will hold an identical belief or ideal. This is commonly observed. For example, in Roethlisberger and Dickson's classic study of the Bank Wiring Observation Room at the Hawthorne Works in Chicago, all of the wiremen expressed a belief that the ideal output was 6,600 connections per day.

While the model predicts that people will often choose the same beliefs, it does not predict that people will choose to *behave* identically. People may choose to hold the same ideals, but some will choose to meet the ideals better than others. In the Bank Wiring Observation Room, people agreed

[12] This is one of the central ideas in Bourdieu's work on tastes. See Bourdieu (1984).

[13] Readers might think that a focus on scholarship can simply be explained by comparative advantage. The theory makes at least two predictions that distinguish it from comparative advantage in this context. First, the theory predicts an individual might pursue scholarship at an economic cost in order to distinguish herself from others. Second, the theory predicts that, if interaction with athletes is unavoidable, an individal will be less inclined to become a scholar.

[14] This account for oppositional culture has similarities to Oxoby (2003, 2004).

[15] Allen (2006) finds that, among Albertan children, math scores do not go up uniformly with age. The younger students in a grade have higher scores. This may be the result of this story.

about the ideal output, but actual output in the room varied. The least productive workers turned out roughly 5,400 connections per day and the most productive turned out roughly 7,200. The variation in output should not cause us to conclude that workers did not feel the pull of the ideals. In particular, the most productive workers were very careful not to turn out too many connections. Those who turned out too much work were labeled "rate-busters;" those who turned out too little were labeled "chiselers."[16][17]

The second reason to believe in network effects is that it is hard to imagine that some of the customs that we observe in groups could be singularly predicted. Consider, for example, the following cleansing rite traditionally practiced by the Carrier Indians of British Columbia. Called "the burying alive," at first menstruation, a girl was covered in a headdress of tanned skin and sent to live alone in the wilderness for three or four years. She was believed to be a danger to the community and in danger herself.[18] It is hard to imagine exogenous parameters that would predict this rite *exactly*—parameters that would predict not only the headdress, but even the length of the girl's stay in the wilderness (not one or two years, but three or four).

The ability of a model with network effects to make predictions. The model produces a dense set of equilibria, but this does not mean that any beliefs or ideals can arise. The model still makes relatively sharp predictions. Some of the most interesting results, in fact, concern comparative statics over the equilibria.

For example, the Women's Movement and Civil Rights Movement can arguably be viewed as events that changed certain ideals.[19] It is interesting to ask what effect these events may have had on *other*

[16] See Homans (1950), Chs. 3-6.

[17] The Bank Wiring Observation Room study also provides evidence that individuals with similar ideals tend to interact more than individuals with different ideals (Principle 2b). Roethlisberger and Dickson found that workers within the room divided themselves into two groups. To be more precise, individuals who were in group 1 interacted considerably more with other individuals in group 1 than with individuals in group 2 (and vice-versa.) When we say that the workers formed two groups, we mean that the pattern of interaction took this form.

Roethlisberger and Dickson found that, while some social norms applied to everyone in the room, such as a belief that output should not be much above or below 6,600 connections per worker per day, other norms applied specifically to group 1 or group 2. For example, gambling games occurred in group 1 but not group 2. Group 1 bought and shared small quantities of chocolate candy while group 2 bought and shared large quantities of an inexpensive candy. Group 1 tolerated less noise and engaged in less horseplay than group 2.

Additional support is given by Weinberg (2005). Looking at a sample of roughly 45,000 high school students, Weinberg finds that individuals tend to interact with people who possess similar characteristics and exhibit similar behaviors.

[18] See Benedict (1934), p.28.

[19] Goldin (2006) shows that the Women's Movement had a profound effect on female labor supply elasticities. More generally, Goldin shows that changing beliefs about how women should and should not behave over the course of twentieth century can be seen in changes in labor supply elasticities.

ideals. In Section 5, it will be suggested that an oppositional culture may have developed among a subset of African Americans because of the Civil Rights Movement (which increased the desire of many African Americans to be viewed as high status) and deindustrialization of cities (which thwarted the attempts of some to meet the ideals of the majority culture).[20]

In Section 7, we will explore predictions made by the model about the categories that people will find salient. These predictions relate to an existing literature in social identity theory and the principle of "comparative fit." It will be argued that leaders often exploit the principle to their advantage, and this may explain why politicians' political ends might be furthered by spreading hate.[21]

The existing literature. This paper is most closely related to two literatures in economics. One is the literature on identity (see Davies (2004) and Hill (2006) for reviews.) One strain of this literature that is particularly relevant views individuals as attempting to meet "prescriptions," which is similar to the concept of ideals in this paper (see, especially, Akerlof and Kranton (2000)). Akerlof and Kranton (2000) take as exogenous the utility that individuals receive from holding particular prescriptions. In this sense, they do not explain why particular prescriptions arise. A second strain of the identity literature that is particularly relevant is Rabin (1994), Oxoby (2003, 2004), and Konow (2000) which argue that an individual's values are driven by a desire to think well of oneself (which closely relates to Principle 3). In Benabou and Tirole (2006b) and Horst, Kermin, and Teschl (2007), assumptions are made that create a desire to conform.

A second relevant literature relates to cognitive dissonance, motivated beliefs, and imperfect memory (see, in particular, Akerlof and Dickens (1982), Koszegi (2001), Benabou and Tirole (2002, 2003, 2004, 2006a, 2006b), Mullainathan and Shleifer (2005), and Brunnermeier and Parker (2005)). These papers allow for the possibility that the beliefs that people hold are not perfectly reflective of the information

[20] Montgomery (1994) and Austen-Smith and Fryer (2005) have also modeled African American oppositional culture. These models provide valuable insights. The model presented in this paper, however, is able to account for findings that are not accounted for by these other models. For example, in a neighborhood where there is a strong oppositional culture, the most talented youngsters are *not* necessarily the least oppositional. This is predicted by the model developed in this paper (see Section 4.2.1) but not by Montgomery (1994) or Austen-Smith and Fryer (2005). Akerlof and Kranton (2008) provide an example of this phenomenon from Whyte's Classic ethnography *Street Corner Society*. They relate the story of Doc from Boston's North End circa 1940 who, despite his great talent, chooses not to go to college. Akerlof and Kranton write that "Whyte was much impressed with Doc's intelligence and asks why Doc, the leader of the Corner Boys, had not gone to college. This is Whyte's answer, largely spoken through Doc's own words. Doc felt he should not go to college. He was the leader of the Corner Boys....Doc sees himself in contrast to the College Boys. In Doc's eyes their upward mobility is a betrayal. Instead of fulfilling the American dream they are deserting their roots in the neighborhood, and not living up to who they should be." (p. 9, Akerlof and Kranton (2008))

[21] This is an idea discussed by Glaeser (2005).

they possess. In particular, the view of information-based confirmation of belief presented in Section 6.3 relates closely to Brunnermeier and Parker (2005).[22]

This paper also relates to the literature on social networks (for a review, see Jackson (2005)). Papers in this literature have generally modeled social networks as arising randomly or for economic reasons. This paper is useful in providing an important additional set of motivations for social interaction.

The implications for economics. While the paper is primarily focused on explaining sociological phenomena, the hope is that the framework will be useful in elucidating economic phenomena. Some economic implications of the theory will be pointed out in the body of the paper and indicate the potential of the framework to address economic questions. For example, it will be suggested that social motivation can facilitate cooperation (see Section 3.2), can cause unemployment (Section 5) and discrimination (Section 6.3), and may explain some instances of money illusion (Section 6.3, Footnote 82).

One of the main reasons that individuals' ideals have economic consequences is that ideals determine what people consider to be *fair*. This paper does not present an explicit model of fairness, but to do so would only require a small additional step. In particular, when an individual fails to behave as others believe she *should* behave and, in so doing, the individual harms others, this is viewed as unfair.

Concerns about fairness can potentially cause unemployment. In particular, a culture where individuals feel that they deserve to be well paid but have poor economic prospects will generally experience high unemployment. In Section 5, it will be argued that the nonemployment rates of African Americans living in the inner city can be explained, in part, by the prevalence of oppositional culture.[23] The deindustrialization of cities, which began in the 1960s (see Wilson (1996)), would not seem to explain the persistence of high nonemployment rates in the inner city. But, as mentioned above, deindustrialization may have played a role in spawning an oppositional culture, and this oppositional culture may explain

[22] Leon Festinger (1957) suggests that individuals attempt to avoid dissonance and choose beliefs and behaviors accordingly. Aronson (2004, p. 146) defines dissonance as "a state of tension that occurs whenever an individual simultaneously holds two cognitions (ideas, attitudes, beliefs, opinions) that are psychologically inconsistent." (Aronson (2004), Chapter 5, provides a superb review of the cognitive dissonance literature.) For example, Aronson suggests that dissonance arises when a cigarette smoker reads a report linking smoking to lung cancer. To reduce dissonance, the smoker might choose to view the evidence as inconclusive.

Therefore, dissonance can be viewed as a state which arises when an individual holds a belief that is unlikely to be true given the information the individual possesses. For example, the cigarette smoker wants to believe that it is safe to smoke (since she does), and this belief is disconfirmed by the information in the report. Hence, information-based disconfirmation is a very similar notion to dissonance.

[23] See Holzer and Offner (2002) and Holzer and Freeman (1986).

why nonemployment rates have remained high.

Ideals are not just beliefs about what people should wear, how people should speak, and how a table should be set. Ideals are a source of *rules*: as we will see, fairness is key to understanding why this is. Classical economic theory takes rules as exogenous. But, understanding how rules form has been a problem of interest to economists for a long time. Attempts to answer the question have largely taken two approaches: either focusing on economic incentives to comply with and enforce rules (resulting in Folk theorems) or a conception of institutions as mechanisms of communication.[24]

One might be disinclined to view internalized ideals as a source of rules for the following reason: when the economic incentives are right, it is natural to think that people will violate their ideals and even change them. However, this argument fails to appreciate that ideals usually spawn economic incentives that bolster the social incentives.[25] In particular, an individual who violates a widely held ideal is seen as behaving *unfairly* (as long as she harms others by so doing). There is a desire to *punish* an individual who behaves unfairly. So, there are likely to be economic consequences to violating an ideal. The importance of ideals in motivating behavior can easily be overlooked because fairness tends to align economic incentives with social incentives.[26]

The theory developed in this paper is also useful for thinking about leadership and authority. Leadership will be mentioned, but will not be treated formally. We can think of a follower as a person who adopts the ideals set by a leader. A follower has as an ideal that she *should* do what the leader tells her to do. This definition corresponds exactly to the standard notion of authority in sociology. For example, George Homans and David Schneider write that "a person holds jural authority over others when, according to the stated norms of his group, he has the right to give them orders and they have the duty to obey."[27] This view of leadership has important economic consequences–particularly for the theory of the firm–because it means that individuals may be able to *set* ideals and use ideals as a way of motivating others.

[24] Kandori (1992), Cole, Mailath, and Postlewaite (1992), and Akerlof (1967) are examples of models of the former type. For a discussion of the latter approach, see Myerson (2006). Models of leadership have often taken the latter approach (see, for example, Komai, Stegeman, and Hermalin (2007)).

[25] It should be noted that social incentives can also be shaped by economic incentives. This paper hopefully gives some insight into this process.

[26] In firms, for example, workers who meet ideals better also tend to earn higher wages.

[27] Homans and Schneider (1955), p. 21. French and Raven (1959) and Emerson (1962) refer to this notion as "legitimate power." French and Raven write: "Legitmate power [of O over P] is here defined as that power which stems from internalized values in P which dictate that O has a legitimate right to influence P and that P has an obligation to accept this influence....Cultural values constitute one common basis for the legitimate power of one individual over another." (p. 265)

The theory and definitions presented in this paper help to build an understanding of what we mean when we refer to an "institution." A considerable economic literature has developed in the last fifteen years arguing that the quality of institutions has a considerable impact on the strength of an economy.[28] This literature has lacked a clear definition of an "institution" and a theory for how institutions form. The framework developed here may prove to be useful to this endeavor, giving some understanding of how to define institutions and how to think about when they will prosper and when they will perish.

The paper will proceed as follows. Section 2 builds a model that incorporates principles 1, 2, and 3. Section 3 gives an example of the model to provide intuition. Section 4 explores the model's implications under certain restrictive assumptions. Section 5 suggests a reason for the rise of an African American oppositional culture in the late 1960s, applying results developed in Section 4. Section 6 extends the model developed in Section 2 and explores additional implications. Section 7 considers predictions made by the theory about the categories that people will find salient and relates these predictions to the concept of "comparative fit."

2 The Model

We will now build a model that captures our three principles. We assume an individual i is a member of a population P ($i \in P$). Individual i has three choices to make. (1) Individual i must choose her behavior, a_i, from a set of possible behaviors, A. (2) She must choose a belief about the ideal way to behave, I_i, from A.[29] (3) Finally, she must choose the subset of the population with whom she wishes to interact, $G_i \subseteq P - \{i\}$.[30] For simplicity, we are assuming that interaction is an individual choice. Individual i can choose to interact with individual j or avoid interaction with j regardless of what individual j chooses.

In the more general version of the model, developed in Appendix B, individual i has a fourth choice: she also chooses beliefs about the state of the world, μ_i. For the time being, we will assume that

[28] For a review, see Helpman (2004), Chapter 7. Also see Greif (1993), Glaeser and Shleifer (2002), North, Summerhill, and Weingast (2000), and Acemoglu, Johnson, and Robinson (2001, 2002).

[29] Individal i applies the same ideal to all members of the population. Individual i does not believe that she should follow ideal $I_i(i)$ herself while individual j should follow ideal $I_i(j)$. It would be worthwhile to generalize the model to allow individual i to apply different ideals to different people–this captures the sociological notion of "roles"–but this is beyond the scope of the paper.

[30] It will be assumed that $i \notin G_i$.

individual i has correct beliefs about the state of the world. In particular, this means that individual i has correct beliefs about the behavior and ideals of others (a_{-i} and I_{-i}).[31]

There are four components to individual i's utility function: (1) an "innate utility function", (2) a cost associated with deviation from the ideal behavior, (3) status utility, and (4) utility from confirmation or disconfirmation of belief.

Innate Utility. The innate utility function, $u_i(a_i, a_{-i})$, reflects the motivations of individual i that are *not* socially influenced. It reflects the innate ease or difficulty–pleasure or displeasure–of behaving in one way or another. Innate utility might also be affected by the behavior of other members of the population. We will refer to innate utility as individual i's "economic motivation." We will refer to the other parts of the utility function as "social motivation."

Cost of Deviation from the Ideal. To the extent that individual i's actual behavior a_i deviates from her ideal behavior I_i, she loses utility. She loses $d(I_i, a_i)$ where d is some distance metric. We will take this metric (and therefore the amount of utility lost from deviation) as exogenous. However, the distance metric can be endogenized (it will be shown how in Section 6). Importantly, when the distance metric is endogenous, the cost of deviating from the ideal is also nonzero (or is virtually always nonzero).

Status Utility. Individual i gains utility from thinking of herself as high status relative to other members of the population. The utility she receives is given by $\Phi(s_i(i))$ where $s_i(i)$ is individual i's view of her own status and Φ is an increasing function.

Principle 3 says that the status individual i assigns to individual j, $s_i(j)$, depends upon how well individual j meets individual i's ideals relative to others in the population. To capture this, we define $s_i(j)$ as follows:

$$s_i(j) = -d(a_j, I_i) + \frac{1}{|P|} \sum_{k \in P} d(a_k, I_i)$$

As principle 3 demands, $s_i(j)$ increases as $d(a_j, I_i)$ decreases. Observe that, if i's view of j rises, i's views of others fall since: $\sum_{j \in P} s_i(j) = 0$. Status is a relative concept. When $I_i = I_j$, i and j have

[31] Readers may question whether individuals know (or even have a good idea) what others believe. However, there is considerable evidence that humans constantly theorize about what others believe and are remarkably good at it. As mentioned above, this ability is referred to as "theory of mind." See Baron-Cohen (1995), Harris (2006), and Saxe (2006).

the same views about status ($s_i(k) = s_j(k)$ for all k), but this need not be the case when $I_i \neq I_j$.[32]

Confirmation of Belief. Principle 2 says that a person's beliefs can be confirmed or disconfirmed. An individual gains utility when her beliefs are confirmed and loses utility when they are disconfirmed. In general, we will think of individuals as holding two types of beliefs, both of which can be confirmed or disconfirmed: (1) beliefs about the ideal way to behave (I_i) and (2) beliefs about the state of the world (μ_i). In the current setting, though, μ_i is taken as exogenous.

Principle 2 also says that there are two ways in which confirmation of belief can take place. There is information-based confirmation and social confirmation. We will assume that confirmation utility, $C(\cdot)$, can be separated into utility from these two sources. Therefore: $C(\cdot) = C^I(\cdot) + C^S(\cdot)$. We will discuss these two components of confirmation utility in turn.

Information-based confirmation of belief:

Principle 2a says that individual i loses utility to the extent that her beliefs are inconsistent with the information that she possesses. It will be assumed that there is *no* information indicating the "right" or "wrong" ideals. Therefore, ideals are *not* subject to information-based confirmation.[33]

Beliefs about the state of the world (μ_i) are subject to information-based confirmation. But, for the time being, we are taking μ_i as exogenous. So, for now, we will simply ignore information-based confirmation. We will assume that $C^I = 0$.

Social confirmation of belief:

According to Principle 2b, individual i's beliefs are socially confirmed when they are similar to the beliefs of the group with whom she interacts, G_i. Individual i's beliefs are socially disconfirmed when they are dissimilar. This leads us to assume that the social confirmation function takes the following form:

$$C^S(I_i, I_{-i}, G_i) = \theta_{|G_i|}(\{m(I_i, I_j)\}_{j \in G_i})$$

where m is some distance metric. $\theta_{|G_i|}$ is assumed to be decreasing in all of its arguments.

We assume that, if individual j's ideal is sufficiently close to i's ($m(I_i, I_j)$ sufficiently small), individ-

[32] Note that individual i forms status beliefs, $s_i(j)$, about everyone in the population P—not just members of her interaction group, G_i. This captures the observation that an individual i might be able to gain utility from a feeling of superiority to individual j without giving much consideration to how j views i in turn, or much desire to hold beliefs similar to those of j.

[33] Section 6.3 discusses this assumption further.

ual i gains social confirmation from interaction with j ($C^S(I_i, I_{-i}, G_i \cup \{j\}) > C^S(I_i, I_{-i}, G_i)$ for $j \notin G_i$). We further assume that, if individual j's ideal is sufficiently far from i's ($m(I_i, I_j)$ sufficiently large), individual i loses social confirmation from interaction with j ($C^S(I_i, I_{-i}, G_i \cup \{j\}) < C^S(I_i, I_{-i}, G_i)$ for $j \notin G_i$).[34]

Social confirmation of belief may establish network effects. That is, it may cause individual i to choose ideal I because individual j chooses ideal I and cause individual j to choose ideal I because individual i chooses ideal I. The existence of network effects has extremely important implications (such as the possibility of an uncountable number of equilibria). Social confirmation of belief establishes network effects if it creates an incentive for individuals to clump at the same ideal. The following is the condition under which C^S instills a desire to clump (and hence establishes network effects).

Definition 1 *If* $\frac{\partial}{\partial x_k}\theta_{|G_i|}(x_1, ..., x_{|G_i|})\Big|_{x_k=0^+} = 0$ *for all* k*, we will say that the clumping condition fails to be satisfied. Otherwise, we will say that the clumping condition is satisfied.*[35]

The Individual's Problem. We can now state the individual's entire choice problem. The individual chooses her behavior ($a_i \in A$), beliefs ($I_i \in A$) and social interactions ($G_i \subseteq P$) to maximize[36]

$$\begin{aligned} U_i(a_i, I_i, G_i, a_{-i}, I_{-i}) &= u_i(a_i, a_{-i}) - d(a_i, I_i) + \Phi(s_i(i)) + C(I_i, I_{-i}, G_i) \\ &= u_i(a_i, a_{-i}) - d(a_i, I_i) + \Phi(s_i(i)) + C^S(I_i, I_{-i}, G_i) \end{aligned}$$

Social Equilibria. The choices made by every individual in population P depend upon the choices made by every other individual in the population P. This gives us the notion of a "social equilibrium." This paper will focus on the pure strategy Nash equilibria of this game, as this is sufficient to convey the

[34] Section 6 will propose a way to endogenize C_i^S as well as $d(\cdot)$.

[35] To illustrate the importance of the clumping condition, consider what happens when it fails to hold in the following peer effects model. Suppose individuals in population P have utility functions: $U_i = -F(|a_i - x_i|) - C \cdot g(\{|a_i - a_j|\}_{j \in P})$ where $C \geq 0$, $F'(0) = 0$, $F'(x) > 0$ for $x > 0$, F'' is a positive constant, and g_k,$g_{kk} \geq 0$ for all k. Suppose the clumping condition does not hold: for all k, $g_k(\{|x_i - x_j|\}_{j \in P}) = 0$ whenever $x_i = x_k$. Then, a pure strategy Nash equilibrium has the property that: $\frac{1}{n}\sum_k a_k = \frac{1}{n}\sum_k x_k$. Therefore, the mean behavior does not depend upon C (the importance of conforming)–only the variance is affected. If the clumping condition *does* hold, there may be an uncountable number of equilibria with different means.

[36] Readers may wonder whether inclusion of $d(a_i, I_i)$ in the utility function is necessary since individual i's desire to meet her ideal is also captured by a desire for a high $s_i(i)$. When the model is extended slightly (see Appendix B), we find that there *are* reasons to include $d(a_i, I_i)$. For example, imagine a situation in which individual i has very little information about how she compares to others in the population and has very little interaction with others in the population. In such a situation, it is easy for individual i to think well of herself no matter how she behaves (see Section 6.3). But, one imagines that she would still feel a desire to meet her ideals, which begs the inclusion of $d(a_i, I_i)$. $d(a_i, I_i)$ reflects an internalized moral sense that is not always equivalent to an individual's view of herself ($s_i(i)$).

ideas presented here. Other equilibrium concepts might also be employed profitably. The pure strategy Nash equilibria are those sets $\{(a_i, I_i, G_i)\}_{i \in P}$ such that $(a_i, I_i, G_i) \in \underset{(a_i', I_i', G_i')}{\operatorname{argmax}} U_i(a_i', I_i', G_i', a_{-i}, I_{-i})$ for all i.

While, it is possible to give assumptions under which mixed-strategy Nash equilibria will exist, such equilibria are not of particular interest. In general, we cannot ensure the existence of pure strategy Nash equilibria.[37] Therefore, when we analyze the model in Section 4, we will take the approach of considering properties of pure strategy Nash equilibria when they exist.

3 An Example

In order to develop an intuition for the model, we will consider an example. This is meant to give a feeling for the model before we analyze the model more generally in Section 4. The results given in Section 4 nest this example as a special case.

We will consider a population consisting of just two individuals with identical utility functions (twins): $U_1 = U_2$.[38] We will see that, despite having identical utility functions, the twins may behave differently because they can potentially increase their opinions of themselves ($s_i(i)$) by differentiating from the other person. Judith Harris has recently suggested that the desire to differentiate to obtain status may explain why identical twins raised apart seem to be roughly as similar as identical twins raised together.[39]

While the twins have a desire to differentiate, the twins also have a desire to choose the same ideals because doing so gives them social confirmation. Social equilibria can also arise where the twins behave identically.

After exploring the twins' desires to differentiate and conform, we will amend the example slightly to illustrate another aspect of the theory. We will add a term to the innate utility functions of the twins so that they could benefit from cooperating. Social motivation potentially makes cooperation possible that would be impossible if the twins only possessed economic motivation ($U_i = u_i$).

[37] We can give plausible assumptions that ensure existence of pure strategy Nash equilibria when we take the ideals as given. This, also, is not of much interest.

[38] Readers may be interested in what happens when there are n individuals with the same utility function rather than just two. Appendix C characterizes the social equilibria that exist when there are n-tuplets.

[39] See Harris (2006).

3.1 Differentiation and conformation

We will make the following assumptions.

Assumptions

(1) $P = \{1, 2\}$
(2) $A = \mathbb{R}$
(3) $u_i(a) = -|a|^3$
(4) $d(a_i, I_i) = 2(a_i - I_i)^2$
(5) $C^S(I_i, I_j, G_i) = \begin{cases} C, \ I_i = I_j, \ G_i = \{j\} \\ -C, \ I_i \neq I_j, \ G_i = \{j\} \\ 0, \ G_i = \varnothing \end{cases}$ with $C \geq 0$ for $i = 1, 2$
(6) $\Phi(s_i(i)) = s_i(i)$[40]

The twins lose innate utility (u_i) if they deviate from behavior $a_i = 0$. If individuals i and j choose the same ideal ($I_i = I_j$), they will choose to interact with one another, which gives them social confirmation utility ($C^S(I_i, I_j, G_i) = C$). If i and j choose different ideals ($I_i \neq I_j$), they prefer not to interact in order to avoid social disconfirmation. Individual i might choose to adopt j's ideal (and vice-versa) in order to acquire social confirmation utility C.[41]

The utility of individual i is as follows:

$$\begin{aligned} U_i(a_i, I_i, G_i, a_j, I_j) &= u_i(a_i) - d(a_i, I_i) + s_i(i) + C^S(I_i, I_j, G_i) \\ &= -|a_i|^3 - 2(a_i - I_i)^2 + \frac{1}{2}\left[2(a_j - I_i)^2 - 2(a_i - I_i)^2\right] + C^S(I_i, I_j, G_i) \\ &= -|a_i|^3 - 3(a_i - I_i)^2 + (a_j - I_i)^2 + C^S(I_i, I_j, G_i) \end{aligned}$$

The first term reflects the desire of the twins to choose a_i (behavior) close to 0 in order to avoid a loss of innate utility. The second term reflects the desire to choose a_i close to I_i (behavior close to ideal behavior), in order to avoid deviation from the ideal and in order to gain status utility. The third term reflects the desire to choose an ideal behavior (I_i) that differs from j's behavior ($|a_j - I_i|$ large), which may allow individual i to look down on individual j. This desire may lead the twins to differentiate. The final term is social confirmation, which gives the twins a desire to conform.

Types of social equilibria.

[40] Assumptions (3) and (4) allow the math to work out cleanly. The assumption of a discontinuity in C^S (Assumption (5)) makes the analysis simpler, but it is in no way key to the results that are obtained.

[41] Network effects can arise even when C_i^S is continuous as a function of I_i. The discontinuity in C_i^S in this example simplifies analysis but it is not necessary. According to the clumping condition (see Definition 1), what is necessary is that, when i interacts with j, there is a first-order loss in social confirmation utility for individual i from deviating from $I_i = I_j$.

Presently, we will characterize the set of social equilibria that exist for different values of C (the desire for social confirmation). While C is not a parameter that we necessarily want to think of as variable, in looking at how the set of social equilibria changes with C, we will develop a better understanding of the properties of the model.[42]

Before characterizing the set of equilibria, we will describe the two types of social equilibria that can arise in this example. In a "conforming equilibrium," the twins choose the same ideal, behave identically, interact with one another, and view each other as equals. In a "differentiating equilibrium," the twins choose different ideals, behave differently, do not interact, and look down on one another.

Conforming equilibrium:	**Differentiating equilibrium:**
$I_i = I_j = I$	$I_i = -I_j = 4$
$a_i = a_j = \left(-1 + \sqrt{1 + 2\lvert I\rvert}\right) sign(I)$	$a_i = -a_j = 2$
$G_i = \{j\}$, $G_j = \{i\}$	$G_i = G_j = \varnothing$
$s_i(i) = s_j(j) = 0$	$s_i(i) = s_j(j) = 16$
$s_i(j) = s_j(i) = 0$	$s_i(j) = s_j(i) = -16$
$U_i(I) = U_j(I) = C - 2\lvert I\rvert(1 + \lvert I\rvert - \sqrt{1 + 2\lvert I\rvert})$	$U_i = U_j = 16$

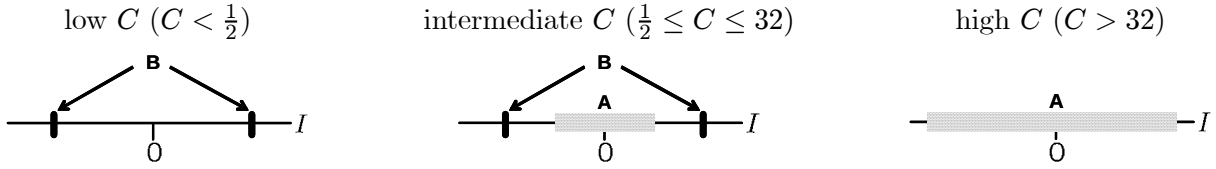
In a conforming equilibrium, individual i chooses to hold the ideal I held by individual j and interact with j ($G_i = \{j\}$) in order to obtain social confirmation utility (C). Individual i chooses a behavior (a_i) between 0 (which maximizes innate utility) and I (which minimizes deviation from the ideal). Unless $I = 0$, individual i will not meet her ideal perfectly. Individuals meet ideal I equally well, so they will view each other as equals ($s_i(i) = s_j(j) = s_i(j) = s_j(i) = 0$). When the ideal I is harder to meet, individual i's utility will be lower ($U_i(I)$ is decreasing in $|I|$).

In a differentiating equilibrium, individual i chooses a different ideal from individual j ($I_i \neq I_j$): they have different beliefs about how people should and should not behave. Choosing a different ideal allows individual i to look down on individual j ($s_i(i) = -s_i(j) = 16$). In fact, individual i chooses an ideal that is difficult to meet ($I_i \neq 0$) rather than an ideal that is easy to meet ($I_i = 0$) because this increases her feeling of superiority to j. Because individual i chooses a different ideal from j, individual i is unable to obtain social confirmation of her ideal. Individual i avoids interaction with j ($G_i = \varnothing$) in order to avoid social *disconfirmation* of her ideal. Individual i chooses a behavior (a_i) between 0 (which maximizes innate utility) and I_i (which minimizes deviation from the ideal). Individual i does not meet her ideal perfectly.

[42] Section 5 (and Appendix A) give a case where a parameter of interest changes.

The differentiating equilibrium relates to the concept of oppositional culture, which is discussed in Section 5. An oppositional culture is a minority of a population with values that differ from those of the majority, considers itself superior to the majority, and is considered inferior by the majority. In a differentiating equilibrium, two cultures form (albeit of equal size), that look down on one another.

Characterizing the equilibria.



A: conforming equilibria, B: differentiating equilibria

If there is little desire to conform to the ideal of the other person (C low), only the differentiating equilibrium exists. For intermediate values of C, conforming and differentiating equilibria both exist. As C increases, the individuals are more willing to conform to difficult ideals in order to obtain social confirmation. So, the set of conforming equilibria gets larger as C increases. If C is sufficiently large, the desire to conform overwhelms the desire to differentiate and the differentiating equilibrium disappears.

To illustrate, these are the equilibria that exist for a few selected values of C:

$C = 0$	only the differentiating equilibrium exists.
$C = 1$	conforming equilibria exist with $\|I\| \leq 0.16$ and the differentiating equilibrium exists.
$C = 10$	conforming equilibria exist with $\|I\| \leq 1.86$ and the differentiating equilibrium exists.
$C = 50$	conforming equilibria exist with $\|I\| \leq 5.19$ and the differentiating equilibrium does not exist.

3.2 Ideals as a means of achieving cooperation

Social motivation can facilitate cooperation that would not take place if individuals only possessed economic motivation ($U_i = u_i$). It can make it equilibrial for individuals to cooperate by giving individuals social confirmation for holding the ideal that one *should* cooperate.

Often, multiple social equilibria will exist: in some, individuals will feel that they should cooperate, and in others, individuals will not feel obliged to cooperate. The equilibrium that arises among the many that exist is perhaps likely to be one that encourages cooperation. Since cooperation makes everyone better off, the social equilibria that fail to encourage cooperation are likely to be Pareto-dominated by

those that do. And, Pareto-dominated equilibria are perhaps unlikely to be focal.[43]

If we amend our example slightly, giving the twins a reason to cooperate, we can see the way in which social motivation can enable cooperation. Let us add a second term to the innate utility function of the twins:

$$u_i(a_i, a_j) = -|a_i|^3 + a_j$$

As before, the twins maximize their innate utility functions by setting $a_i = 0$. But, if $a_i = a_j = a$, the innate utility of individual i is maximized when $a = \frac{\sqrt{3}}{3}$ rather than $a = 0$.

This amendment to the innate utility function does not change the set of social equilibria that exist. But, it changes the utility that the twins receive in a given equilibrium. Previously, in a conforming equilibrium in which the twins conform to ideal I, the utility of individual i was:

$$U_i(I) = C - 2|I|(1 + |I| - \sqrt{1 + 2|I|})$$

This is a decreasing function of $|I|$. Hence, the conforming equilibrium with $I = 0$ Pareto-dominated the other conforming equilibria.

With this amendment, it is no longer the case that the conforming equilibrium in which $I = 0$ Pareto-dominates the other conforming equilibria. In a conforming equilibrium with ideal I, the utility of individual i is:

$$U_i(I) = C - 2|I|(1 + |I| - \sqrt{1 + 2|I|}) + (-1 + \sqrt{1 + 2|I|}) * sign(I)$$

Because $I > 0$ encourages cooperation, $U_i(I)$ is increasing at $I = 0$ and is largest at $I^* = \frac{5}{8} < \frac{2}{3}$ (see Figure 1). A conforming equilibrium with $I = \frac{5}{8}$ exists if there is a sufficient desire for social confirmation: $C \geq 2.83$. In the conforming equilibrium with $I = \frac{5}{8}$, $a_i = a_j = \frac{1}{2} < \frac{\sqrt{3}}{3}$.

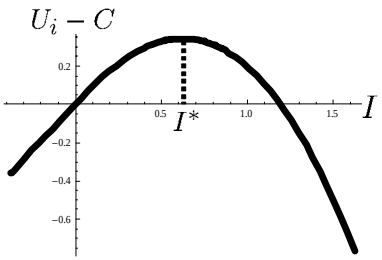


Figure 1

[43] Kuran and Sandholm (2008) also see the benefits arising from cooperation and coordination as potential cultural determinants.

Our amendment to the twins' innate utility does not eliminate noncooperative equilibria: if a conforming equilibrium with $I = \frac{5}{8}$ exists, a conforming equilibrium with $I = 0$ also exists. But, the $I = 0$ equilibrium is Pareto-dominated when there is a reason to cooperate and a Pareto-dominated equilibrium is unlikely to be focal. In this sense, the amendment may change the ideals that arise.

An aside: ideals may breed economic reasons to cooperate.

The introduction discussed the idea that ideals determine what individuals consider to be fair and individuals' concerns about fairness create economic incentives to meet ideals that bolster the non-economic incentives to meet ideals. Extending the model to capture this formally is a topic for another paper. But, it is worthwhile to briefly discuss the economic incentives to cooperate that would form in an extended model as a result of a social norm that people should cooperate.

It is reasonable to assume that an individual i will view an individual j as having behaved unfairly if j does not behave as individual i feels j *should* behave and this causes i harm. If i feels that j has behaved unfairly, she will have a desire to see j punished. Therefore, if a social norm exists that people should cooperate, an individual who fails to cooperate will be viewed as having acted unfairly and deserving of punishment.[44][45]

This may give an interpretation of the motivation individuals hold when they follow tit-for-tat strategies. Individual i may feel compelled to punish j if j defects today because of i's sense of what is fair. Defecting in the future is one way of punishing individual j for her defection today.[46]

In addition to an incentive to punish those who fail to meet a social norm, there is also an incentive to *monitor* others' compliance to a norm. We will discuss the reason momentarily. What this means is that, when a social norm exists that people should cooperate, individuals in the population will be looking for people who are not cooperating and will be eager to punish those who fail to cooperate.

While the model developed in Appendix B could be used to formally derive the existence of an incentive to monitor, we will just give an informal explanation of the reason. The reason is as follows. An individual does not want to be considered a "chump." Monitoring others helps an individual avoid being perceived as a chump. A chump is an individual who holds an ideal that one should cooperate

[44] This relates to Rabin (1993). Rabin (1993) attempts to incorporate fairness into economic games.

[45] Ideals may also create economic incentives in another way. If individual i fails to cooperate, this may signal that individual i has the wrong ideals, which may suggest to others that it would be unwise to cooperate with individual i in the future. Individual i will be seen as someone who cannot be trusted. Signalling is not captured by the model developed in Section 2, but it is captured by the model presented in Appendix B.

[46] See Axelrod (1981, 1984) and Kreps, Milgrom, Wilson, and Roberts (1982).

and cooperates herself but fails to realize that others only pay lip service to an ideal of cooperation. Those who get away with only paying lip service to the ideal will view a chump as low status. As we will see in Section 6.1, individuals want to be perceived as high status by those with whom they interact, so it is unpleasant to be a chump.[47]

4 Analyzing the Model

Let us analyze the model developed in Section 2 more thoroughly. This section will proceed in the following way. Section 4.1 describes the best response function of an individual i. Section 4.2 uses the description of the best response function to partially describe the set of social equilibria. Section 4.3 explains why ideals are likely to arise that are difficult to meet (cannot be met without losing innate utility). Section 4.4 considers the perceptions that individuals will have of one another. Appendix D gives some further analysis of the model that is not included in this section.

Assumptions.

In this section, we will make certain restrictive assumptions. It will be assumed that $A = \mathbb{R}$ (the set of possible behaviors and ideals). Functional form assumptions are also made. The assumptions are:

(1) $A = \mathbb{R}$
(2) $u_i(a) = -f_1(\frac{1}{\alpha_i}|a|)$ with $\alpha_i > 0$.
(3) $d(a_i, I_i) = f_2(|a_i - I_i|)$.
(4) $C^S(I_i, I_{-i}, G_i) = C \cdot q\left(|\{j \in G_i : I_j = I_i\}| - |\{j \in G_i : I_j \neq I_i\}|\right)$ with q nondecreasing and $C \geq 0$.[48]
(5) f_1 and f_2 are differentiable, with $f_1'', f_2'' \geq 0$, $f_1'(0) = f_2'(0) = 0$ and $f_1'(x), f_2'(x) > 0$ for $x \neq 0$.

[47] One of the themes of Ostrom (1990) is that the imposition of incentives to cooperate by a principal (such as by a government or a boss within a firm) may fail to achieve cooperation between agents. Such impositions are likely to fail unless the principal creates or fails to destroy a sense among the agents that they *should* cooperate. The reason is that a principal is likely to find it difficult to monitor agents, which makes it difficult to impose punishment in an effective way. On the other hand, agents are in a good position to monitor other agents. When the agents have the right ideals, they will have a desire to engage in monitoring and a desire to punish shirkers or bring them to the attention of a principal. One example cited by Ostrom is the nationalization of forests in Third World countries, such as Thailand, Niger, and India (p. 23). Nationalization was intended as a way of limiting use of forest resources to prevent soil erosion and decreased long run productivity. Prior to nationalization, small villages had, in fact, treated the forests as communal assets. Villagers had some sense that they should not overuse forest assets. With nationalization, villages lost the sense that the forests belonged to them and villagers lost the feeling that they should police the use of other villagers. The governments found it difficult to enforce regulations concerning use because villagers lacked a feeling that they should comply, and these nationalization experiments ended up contributing to the problem they were intended to solve.

[48] The assumption that q is a function of $|\{j \in G_i : I_j = I_i\}| - |\{j \in G_i : I_j \neq I_i\}|$ simplifies analysis. It means that, unless individual i sets $I_i = I_j$, which gives i social confirmation, individual i's desire is to differentiate from individual j rather than conform. The assumption on q ensures that inclusion of an individual j in G_i for whom $I_j = I_i$ will increase C_i^S or leave it unchanged.

(6) $\Phi(s_i(i)) = \beta s_i(i)$ with $\beta > 0$.[49]

Assumption 2 says that, as the distance between individual i's behavior (a_i) and 0 (her "bliss point") increases, her innate utility declines. Individual i's innate utility declines more quickly if she has low ability rather than high ability (α_i low rather than high). If individual i only had economic motivation ($U_i = u_i$), she would not choose to deviate from her bliss point (she would set $a_i = 0$).

According to assumption 4, social confirmation (C^S) increases when individual i interacts with someone with the same ideals ($I_i = I_j$) and decreases when i interacts with someone with different ideals ($I_i \neq I_j$). C^S satisfies the clumping condition (defined in Section 2) as long as q and C are nonzero, which means that network effects are present in the model: individual i might choose ideal I because it is held by j, and j might hold ideal I because it is held by i.

β parameterizes individual i's concern for status. C parameterizes i's concern for social confirmation utility.

Under these assumptions, individual i's utility is

$$\begin{aligned} U_i &= u_i - d(a_i, I_i) + s_i(i) + C^S(I_i, I_{-i}, G_i) \\ &= -f_1(\frac{1}{\alpha_i}|a_i|) - f_2(|a_i - I_i|) + \beta \left[-f_2(|a_i - I_i|) + \frac{1}{|P|} \sum_{k \in P} f_2(|a_k - I_i|) \right] + C^S(I_i, I_{-i}, G_i) \end{aligned} \tag{1}$$

Method of Analysis.

Unfortunately, it is not possible to give an explicit characterization of all of the social equilibria as we did in Section 3. This is due to the complexity that arises in a model with network effects and more than two agents.

We will therefore use the following approach to analyze the problem. We will characterize the best response functions of the agents. We will use these results about the best response functions to say something about the social equilibria that exist for a given set of parameters. For example, we will be able to give a partial characterization of the set of equilibria where everyone in the population holds the same ideal. We will be able to describe the effect of changes in parameters on this set.

While this style of analysis leaves something to be desired, it is powerful enough that we can make some useful predictions. For example, in Section 5, we will apply some of the results from this section to

[49] This assumption simplifies analysis because it makes the choice of a_i given I_i independent of a_j for $j \neq i$. Therefore, taking the choice of the I_i's as given, the problem is not game theoretic.

get a sense of why the Civil Rights Movement caused some African Americans to adopt an oppositional culture and why those affected by the deindustrialization of cities in the 1960s were especially likely to become oppositional.

4.1 The Best Response Function

In our attempt to characterize an agent's best response function, we will proceed as follows. First, we will consider individual i's optimal choice of interaction group (G_i) given her choice of behavior (a_i) and ideal (I_i). Then, we will consider individual i's optimal choice of behavior (a_i) given her choice of ideal (I_i). Finally, we will discuss individual i's optimal choice of ideal (I_i).

4.1.1 The best response function: the choice of G_i

Our first step in characterizing the best response function is to consider individual i's optimal choice of interaction group (G_i) given individual i's choice of behavior (a_i) and ideal (I_i).

Social interaction (G_i) only affects one term in the utility function: social confirmation (C^S). As mentioned above, social confirmation (C^S) increases when individual i interacts with someone with the same ideals ($I_i = I_j$) and decreases when i interacts with someone with different ideals ($I_i \neq I_j$). Therefore, it is optimal for individual i to interact with and only with those people who hold the same ideals. To be more precise:

$$G_i^{BR}(I_i, a_i) = \arg\max_{G_i} U_i = \{k \neq i : I_k = I_i\}$$

Since G_i^{BR} does not actually depend upon i's choice of behavior (a_i), we can write G_i^{BR} as a function of i's ideal (I_i) alone: $G_i^{BR}(I_i)$.

From expression 1, we conclude that individual i's utility conditional on the optimal choice of G_i is:

$$U_i|_{G_i=G_i^{BR}(I_i)} = -f_1(\frac{1}{\alpha_i}|a_i|) - f_2(|a_i - I_i|) + \beta \begin{bmatrix} -f_2(|a_i - I_i|) \\ +\frac{1}{|P|}\sum_{k\in P} f_2(|a_k - I_i|) \end{bmatrix} + C \cdot q(|\{k \neq i : I_k = I_i\}|) \quad (2)$$

The final term is social confirmation. We observe that it only depends upon the number of people who share ideal I_i.

Remark 1 *For simplicity, we have chosen to denote G_i^{BR} as a function of I_i. In fact, G_i^{BR} also depends upon other individuals' ideals (I_{-i}). So, it would be more appropriate to write: $G_i^{BR}(I_i, I_{-i})$. G_i^{BR} does not depend upon the behavior (a_{-i}) or social interactions (G_{-i}) of others.*

4.1.2 The best response function: the choice of a_i

Our second step in characterizing the best response function is to consider the optimal choice of behavior (a_i) conditional on the choice of ideal (I_i).

We will let a_i^{BR} denote individual i's optimal choice of behavior conditional on her choice of ideal ($a_i^{BR}(I_i) = \arg\max_{a_i} U_i|_{G_i=G_i^{BR}(I_i)}$). From expression 2, we find that a_i^{BR} can be written as follows:

$$\begin{aligned} a_i^{BR}(I_i) &= \arg\max_{a_i} \left\{ -f_1(\frac{1}{\alpha_i}|a_i|) - f_2(|a_i - I_i|) + \beta \left(-\frac{|P|-1}{|P|} f_2(|a_i - I_i|) \right) \right\} \\ &= \arg\max_{a_i} \left\{ -f_1(\frac{1}{\alpha_i}|a_i|) - \frac{\beta(|P|-1)+|P|}{|P|} f_2(|a_i - I_i|) \right\} \end{aligned}$$

The two terms of this expression represent the concerns that individual i trades off in choosing her behavior (a_i^{BR}).

Remark 2 *Individual i trades off two concerns in choosing her behavior (a_i^{BR}):*

(1) the desire to choose a behavior (a_i) close to her bliss point (0) to increase innate utility.

(2) the desire to choose a behavior (a_i) close to her ideal (I_i) in order to reduce deviation from the ideal and to gain status utility.

The following proposition gives some results about the optimal behavior.

Proposition 1 *Let $a_i^{BR}(I_i) = \arg\max_{a_i} U_i|_{G_i=G_i^{BR}(I_i)}$.*

(1) Individual i optimizes by choosing a behavior between her bliss point and her ideal ($a_i^{BR} \in (\min(0, I_i), \max(0, I_i))$ for $I_i > 0$ and $a_i^{BR} = 0$ for $I_i = 0$).

(2) An increase in individual i's ability (α_i) decreases her deviation from her ideal ($|a_i^{BR} - I_i|$), and increases her deviation from her bliss point ($|a_i^{BR}|$).

(3) An increase in the distance between the bliss point and the ideal (an increase in $|I_i|$) increases individual i's deviation from her bliss point ($|a_i^{BR}|$) and increases her deviation from her ideal ($|a_i^{BR} - I_i|$).

(4) An increase in individual i's concern about status (β) increases individual i's deviation from her bliss point ($|a_i^{BR}|$) and decreases her deviation from her ideal ($|a_i^{BR} - I_i|$).

As part (2) of the proposition indicates, individuals i and j may hold the same ideal but, because of differences in innate utility, choose different behaviors. For example, if i and j share the same ideal ($I_i = I_j = I$) and bliss point but j is less able than i ($\alpha_i > \alpha_j$), individual i will come closer to meeting the ideal ($|I - a_i| < |I - a_j|$).

As noted in the introduction, network effects may be present in the model even if we do not observe people behaving identically: individuals i and j may clump at the same ideal ($I_i = I_j = I$) but behave differently ($a_i \neq a_j$) because of differences in innate utility.

Remark 3 *Unlike G_i^{BR}, which depends upon the ideals of others (I_{-i}), a_i^{BR} only depends upon individual i's choice of ideal (I_i).*

4.1.3 The best response function: the choice of I_i

We turn now to characterizing individual i's optimal choice of ideal: $I_i^{BR} = \arg\max_{I_i} U_i|_{a_i = a_i^{BR}(I_i), G_i = G_i^{BR}(I_i)}$. It follows from expression 2 that the optimal choice of ideal is:

$$I_i^{BR} = \arg\max_{I_i} \{\Delta(|I_i|, \alpha_i) + S(I_i, a_{-i}, \alpha_i, \beta) + C \cdot q(|\{k \neq i : I_k = I_i\}|)\} \tag{3}$$

where

$$\begin{aligned} \Delta(I_i, \alpha_i) &= -f_1(\frac{1}{\alpha_i}|a_i^{BR}(I_i)|) - f_2(|a_i^{BR}(I_i) - I_i|) \text{ (note: } \Delta(I_i, \alpha_i) = \Delta(|I_i|, \alpha_i)) \\ S(I_i, a_{-i}, \alpha_i, \beta) &= \beta \left[-\frac{|P| - 1}{|P|} f_2(|a_i^{BR}(I_i) - I_i|) + \frac{1}{|P|} \sum_{k \neq i} f_2(|a_k - I_i|) \right] \end{aligned}$$

The following proposition describes Δ and S, giving us a better sense of the concerns that individual i trades off in choosing an ideal.

Proposition 2 *(1) $\Delta(|I_i|, \alpha_i)$ is decreasing in $|I_i|$: it is maximized at $I_i = 0$.*

(2) $S(I_i, a_{-i}, \alpha_i, \beta)$ is the status utility of individual i ($\Phi(s_i(i))$) if she chooses $(I_i, a_i^{BR}(I_i), G_i^{BR}(I_i))$.

(3) S is generally not maximized at $I_i = 0$.

Part (3) says that individual i's status is generally not maximized by setting $I_i = 0$ (the bliss point), which might seem surprising. It is true that the first term of S is maximized by setting $I_i = 0$. The first term reflects the status to be gained from meeting one's ideal well and $I_i = 0$ is the easiest ideal to meet. But, the second term of S reflects i's desire to choose an ideal that differentiates her from others in the population. This term is generally not maximized by setting $I_i = 0$.

Remark 4 *Individual i trades off three concerns in her choice of ideal (I_i^{BR}), each corresponding to a term of expression 3:*

(1) the desire to hold an ideal that is easy to meet (close to 0).

(2) the desire to be viewed as high status.

(3) the desire to hold a popular ideal to obtain social confirmation.

The importance of concern (1) relative to (2) and (3) depends upon the ability of individual i (α_i). As the following proposition indicates, more able individuals will find concern (1) somewhat less important.

Proposition 3 *Suppose $|I| \leq |I'|$. An increase in individual i's ability (α_i) increases the utility of choosing I' relative to the utility of choosing I. More precisely: $U_i|_{I_i=I',a_i=a_i^{BR}(I'),G_i=G_i^{BR}(I')} - U_i|_{I_i=I,a_i=a_i^{BR}(I),G_i=G_i^{BR}(I)}$ is increasing in α_i.*

Conforming and rebelling.

It is either in individual i's interest to choose an ideal held by others ($I_i^{BR} = I_j$ for some j) because of concern (3) or to become a "rebel," eschewing concern 3 and choosing a unique ideal–one that trades off concerns (1) and (2) only. The following remark states this more formally.

Remark 5 *Let I_i^{reb} be defined as follows*

$$I_i^{reb} = \arg\max_{I_i} \Delta(|I_i|, \alpha_i) + S(I_i, a_{-i}, \alpha_i, \beta)$$

Then, $I_i^{BR} \in \{I_j\}_{j \neq i} \cup \{I_i^{reb}\}$.

We will now define carefully what it means for individual i to choose to "conform to I_j" or choose to "rebel."

Definition 2 *(i) If* $I_i = I_i^{reb}$, $a_i = a_i^{BR}(I_i^{reb})$, *and* $G_i = G_i^{BR}(I_i^{reb})$, *we will say that individual* i *has chosen to "rebel."*

(ii) If $I_i = I_j$, $a_i = a_i^{BR}(I_j)$, *and* $G_i = G_i^{BR}(I_j)$,*we will say that individual* i *has chosen to "conform to* I_j*."*

The status of a rebel.

If individual i chooses to rebel, the two concerns that define her choice of ideal are (1) the desire to hold an ideal that is easy to meet and (2) the desire to gain status. Individual i's view of herself ($s_i(i)$) would be *lower* if she chose an ideal based upon concern (1) alone. If individual i were to choose an ideal based upon concern (1) alone, she would choose $I_i = 0$. The optimal behavior would be $a_i^{BR}(0) = I_i = 0$. In this event, individual i meets her ideal perfectly, so she will consider herself above average ($s_i(i) \geq 0$.). Therefore, in the event that individual i rebels, she will consider herself above average.

Proposition 4 *If individual* i *chooses to rebel, she will consider herself above average:* $s_i(i) \geq 0$. *It need not be the case that individual* i *considers herself superior to everyone in the population* ($s_i(i) \geq s_i(j)$ *for all* $j \in P$*).*

Individual i might be willing to view herself as below average (choose an ideal such that $s_i(i) \leq 0$). But, she will only be willing to do so if she receives social confirmation from so doing. As the following proposition indicates, the more individual i cares about status, the less willing she will be to hold an ideal that makes her feel inferior.

Proposition 5 *Suppose that, if individual* i *conforms to* I_j, *she will consider herself below average* ($s_i(i) \leq 0$). *Then, an increase in individual* i*'s concern about status* (β) *increases the utility of rebelling relative to the utility of conforming to* I_j. *More precisely:* $U_i|_{I_i=I_i^{reb},a_i=a_i^{BR}(I_i^{reb}),G_i=G_i^{BR}(I_i^{reb})} - U_i|_{I_i=I_j,a_i=a_i^{BR}(I_j),G_i=G_i^{BR}(I_j)}$ *is increasing in* β *if* $\frac{1}{\beta}S(I_j, a_{-i}, \alpha_i, \beta) \leq 0$.

The optimality of choosing an ideal behavior distinct from actual behavior.

Generally, individual i's best response is to choose a behavior and an ideal behavior that are distinct ($a_i^{BR} \neq I_i^{BR}$). The following remark gives the reason.

Remark 6 *It may be optimal for individual i to choose behavior a_i and ideal I_i such that $a_i \neq I_i$ for the following reasons:*

(1) If individual i chooses an ideal that is difficult to meet ($I_i \neq 0$), it may be optimal for individual i to choose a behavior (a_i) that differs from her ideal (I_i) in order to avoid a loss of innate utility (Remark 4, concern 1).

(2) Individual i might choose an ideal that is difficult to meet in order to gain status (Remark 4, concern 2) or in order to obtain social confirmation (Remark 4, concern 3).

4.1.4 The best response function: forced social interaction.

We have considered individual i's optimal choice of ideal (I_i) and behavior (a_i) when individual i chooses her social interactions optimally ($G_i = G_i^{BR}(I_i)$). It is worthwhile to consider the effect of *forced* social interaction on individual i's choice of ideal and behavior.[50]

If individual i interacts with individual j, i has a desire to hold ideals similar to j's in order to obtain social confirmation and avoid social disconfirmation (C^S). Therefore, if individual i is forced to interact with individual j, this increases i's desire to hold ideals similar to j's.

Proposition 6 *Let $(I_i^{forced}(G), a_i^{forced}(G))$ denote i's optimal choice of ideal and behavior given that i is interacting with group G (for whatever reason): $(I_i^{forced}(G), a_i^{forced}(G)) = \arg\max_{(a_i,I_i)} U_i|_{G_i=G}$. We can conclude the following:*

(1) $|I_i^{forced}(G \cup \{j\}) - I_j| \leq |I_i^{forced}(G) - I_j|$.

(2) $a_i^{forced}(G) = a_i^{BR}(I_i^{forced}(G))$.

(3) If $a_j = a_j^{BR}(I_j)$ and $\alpha_i = \alpha_j$:

(i) $|a_i^{forced}(G \cup \{j\}) - a_j| \leq |a_i^{forced}(G) - a_j|$.

(ii) If individual i holds $(I_i^{forced}(G \cup \{j\}), a_i^{forced}(G \cup \{j\}))$ rather than $(I_i^{forced}(G), a_i^{forced}(G))$, i's opinion of j relative to herself ($s_i(j) - s_i(i)$) and j's opinion of i relative to herself ($s_j(i) - s_j(j)$) are both greater.

Part (1) of the proposition says that, if individual i is forced to interact with individual j, she will

[50] We have assumed that there are no economic reasons for social interaction (u_i does not depends upon G_i). But, social interaction may, in fact, take place for economic reasons. This section is useful in examining the effect of interaction that takes place for economic reasons on individuals' behavior and ideals.

choose an ideal that is closer to j's ideal than she otherwise would.[51] This may lead individual i to choose a behavior (a_i) that is closer to individual j's behavior than she otherwise would. Part (3) of the proposition gives a condition under which forced interaction with j leads i to behave in a manner more similar to j. Under this condition, forced interaction also increases the individuals' regard for one another.

4.2 The Social Equilibria

Having characterized individual i's best response function, we turn now to examining social equilibria. We will proceed as follows. First, we will show that it is possible to view the game individuals are engaged in as equivalent to a game in which individuals are only choosing ideals (I_i). This greatly simplifies analysis.

We will then discuss one type of social equilibrium that can arise: an equilibrium in which everyone holds the same ideals. We will give a partial characterization of the set of such equilibria that exist and will discuss how this set changes when parameters change.

After discussing equilibria where everyone holds the same ideals, we will turn to cases where more than one ideal is held in equilibrium. Due to the complexity of this topic, it requires some finesse. We will discuss the impact of the rebellion by one individual in the population against a common ideal on the choices made by the other individuals. This gives some general insight into equilibria where multiple ideals are held. We will use this analysis to discuss the relationship between group affiliation and ability in equilibrium.

We will conclude this section with a discussion of the equilibria that can arise when individuals are identical ($\alpha_i = \alpha$ for all i), as was the case in Section 3.

Social equilibria.

Individuals are involved in a game in which they choose an ideal, behavior, and social interaction group (I_i, a_i, G_i). It will now be argued that this game is *equivalent* to a game in which individuals only choose an ideal I_i. It simplifies analysis greatly to think of individuals as making a single choice rather than three.

[51] With C_i^S defined as it is in Section 4, it is actually the case that $I_i^{forced}(G \cup \{j\}) = I_j$ or $I_i^{forced}(G) = I_i^{forced}(G)$. But, if C_i^S is a continuous function, forced interaction might cause i to choose an ideal closer to but not identical to j's ideal.

First, we observe that a social equilibrium can be described by the ideals that individuals hold alone. The following remark elaborates this point.

Remark 7 *If $\{(I_i, a_i, G_i)\}_{i\in P}$ is a social equilibrium, then: $a_i = a_i^{BR}(I_i)$ and $G_i = G_i^{BR}(I_i, I_{-i})$ for all i. Therefore, $\{(I_i, a_i, G_i)\}_{i\in P} = \{(I_i, a_i^{BR}(I_i), G_i^{BR}(I_i, I_{-i}))\}_{i\in P}$. Since a social equilibrium is fully specified by the ideals that the individuals hold, we can speak of the $(I_1, ...I_{|P|})$ equilibrium, and it will be clear what is meant.*

Definition 3 *If we refer to $(I_1, ...I_{|P|})$ as a potential equilibrium, we mean that individuals in the population make the following choices: $\{(I_i, a_i^{BR}(I_i), G_i^{BR}(I_i, I_{-i}))\}_{i\in P}$.*

Secondly, we observe that, if there is a profitable deviation for individual i from $(I_1, ...I_{|P|})$, that deviation must take the following form: $(I_i', a_i', G_i') = (I_i', a_i^{BR}(I_i'), G_i^{BR}(I_i', I_{-i}))$.

Definition 4 *(1) If we refer to individual i as "deviating from I_i to I_i'," we mean that individual i deviates to: $(I_i', a_i^{BR}(I_i'), G_i^{BR}(I_i', I_{-i}))$.*

(2) Let $D_i(I_1, ..., I_{|P|}, I_i')$ denote the change in individual i's utility from a deviation to I_i' More precisely:

$$\begin{aligned} D_i(I_1, ..., I_{|P|}, I_i') &= U_i(I_i', a_i^{BR}(I_i'), G_i^{BR}(I_i', I_{-i}), I_{-i}, a_{-i}^{BR}(I_{-i})) \\ &\quad -U_i(I_i, a_i^{BR}(I_i), G_i^{BR}(I_i, I_{-i}), I_{-i}, a_{-i}^{BR}(I_{-i})). \end{aligned}$$

(3) D_i can be broken up into three components reflecting concerns (1), (2), and (3) of Remark 4. $D_i = D_i^1 + D_i^2 + D_i^3$, where:

$$D_i^1 = \Delta(|I_i'|, \alpha_i) - \Delta(|I_i|, \alpha_i) \qquad D_i^2 = S(I_i', a_{-i}^{BR}(I_{-i}), \alpha_i, \beta) - S(I_i, a_{-i}^{BR}(I_{-i}), \alpha_i, \beta)$$

$$D_i^3 = C \cdot \left(q(|\{k \neq i : I_k = I_i'\}|) - q(|\{k \neq i : I_k = I_i\}|)\right)$$

If a profitable deviation exists, it must be the case that $I_i' = I_i^{BR}(I_{-i})$ is a profitable deviation.[52] Since $I_i^{BR} \in \{I_j\}_{j\neq i} \cup \{I_i^{reb}(I_{-i})\}$, if a profitable deviation exists, some $I_i' \in \{I_j\}_{j\neq i} \cup \{I_i^{reb}(I_{-i})\}$ must be a profitable deviation. This gives us the following criterion for $(I_1, ...I_{|P|})$ to be a social equilibrium.

Remark 8 *$(I_1, ...I_{|P|})$ is a social equilibrium if and only if no individual i benefits from deviating to some $I_i' \in \{I_j\}_{j\neq i} \cup \{I_i^{reb}(I_{-i})\}$. More precisely, $(I_1, ...I_{|P|})$ is a social equilibrium if and only if:*

$$D_i(I_1, ..., I_{|P|}, I_i') \leq 0 \text{ for all } I_i' \in \{I_j\}_{j\neq i} \cup \{I_i^{reb}(I_{-i})\},\ i \in P$$

[52] It should be noted that, when we write $I_i^{BR}(I_{-i})$, we mean: $I_i^{BR}(I_{-i}) = I_i^{BR}(I_{-i}, a_{-i}^{BR}(I_{-i}))$. Similarly, by $I_i^{reb}(I_{-i})$, we mean: $I_i^{reb}(I_{-i}) = I_i^{reb}(I_{-i}, a_{-i}^{BR}(I_{-i}))$.

We will use this criterion in the sections that follow.

4.2.1 Equilibria with a common ideal

We will now examine one type of social equilibrium that might exist: an equilibrium in which everyone holds the same ideal I. We will describe the set of such equilibria that exist.

It is useful to define individual i's desire to rebel against ideal I.

Definition 5 *Let $R_i(I)$ denote individual i's desire to rebel against a common ideal I: $R_i(I) = D_i(I, ..., I, I_i^{reb}(I, ..., I))$.*

From Remark 8, it follows that it is equilibrial for everyone in the population to hold ideal I if no individual gains from rebelling against it: $R_i(I) \leq 0$ for all i.

We can evaluate i's desire to rebel in terms of the three concerns that i trades off in choosing an ideal: (1) holding an ideal that is easy to meet, (2) gaining status, and (3) holding a popular ideal to obtain social confirmation. In rebelling, individual i loses social confirmation utility $q(|P| - 1) - q(0)$. Therefore the reason to rebel is either to hold an ideal that is easier to meet, to gain status, or both.

Suppose everyone in the population conforms to ideal I. As the ideal becomes harder to meet ($|I|$ increases), everyone in the population has an increased desire to rebel because of concern (1).

But, the desire to rebel because of concern (2) potentially falls for some individuals as the ideal becomes harder to meet. As the ideal becomes harder to meet, the status of high ability individuals potentially *rises* relative to the status of low ability individuals (see part 3 of the next proposition). An ideal that is difficult to meet affords high ability individuals an opportunity to show off. This gives high ability individuals an added incentive to conform.

Therefore, low ability individuals have an increased desire to rebel as the ideal becomes harder to meet and high ability individuals might have a decreased desire to rebel as the ideal becomes harder to meet. The following proposition states this more precisely.

Proposition 7 *Suppose everyone in population P conforms to ideal I.*

(1) If i is at least as able as j ($\alpha_i \geq \alpha_j$), individual i's desire to rebel is (weakly) greater than j's when the ideal is easiest to meet ($R_i(I) \geq R_j(I)$ when $I = 0$).

(2) If individual i has the least ability of anyone in population P ($\alpha_i \leq \alpha_j$ for all $j \in P$), her desire to rebel (weakly) increases as the ideal becomes harder to meet ($\frac{dR_i}{d|I|} \geq 0$).

(3) If $f_2'''(x) \leq 0$, $2f_1''(x) + xf_1'''(x) \geq 0$, *and* $(f_1''(x))^2 - f_1'(x)f_1'''(x) \geq 0$ *for all* x, *an increase in* $|I|$ *(weakly) increases* i*'s status relative to* j*'s* $(s_i(i) - s_j(j))$ *when* i *is at least as able as* j $(\alpha_i \geq \alpha_j)$. *These conditions on* f_1 *and* f_2 *are satisfied when* $f_1(x) = x^n$ *and* $f_2(x) = x^m$, *with* $n \geq 1$ *and* $m \leq 2$.

Cases can arise where high ability individuals would rebel against $I = 0$ (the easiest ideal to meet) but another ideal I' ($|I'| > 0$) exists that high ability types would not rebel against and is sufficiently easy to meet that the low ability individuals are willing to conform. Therefore, it is possible that $I = 0$ will not be an equilibrium but $|I'| > 0$ (an ideal that is harder to meet) will be an equilibrium. This is stated in the following proposition.

Proposition 8 *Let* $S = \{I :$ *an equilibrium exists where everyone conforms to* $I\}$. *Then:*

(1) If S *is nonempty, it need not be the case that* $0 \in S$.

(2) If everyone in the population has the same ability $(\alpha_i = \alpha$ *for all* $i)$, S *will be an interval symmetric about* 0 *(such as, for example,* $[-\pi, \pi]$*).*

(3) (i) If S_1 *denotes the set* S *when* $C = C_1$ *and* S_2 *denotes the set* S *when* $C = C_2 > C_1$ *(all other parameters the same), then* $S_1 \subseteq S_2$.

(ii) If S_1' *denotes the set* S *when* $\beta = \beta_1$ *and* S_2' *denotes the set* S *when* $\beta = \beta_2 < \beta_1$ *(all other parameters the same), then* $S_1' \subseteq S_2'$ *when* $\alpha_i = \alpha$ *for all* i.

4.2.2 Does rebellion stabilize or destabilize?

Suppose an individual in a population *does* rebel against a common ideal. A question of some interest is whether this inclines others to join the rebellion. We might imagine that a rebel sets off an avalanche of rebellion: one rebel leads to two, two to three, and three to four, causing the abandonment of the original ideal. Consider, for example, the effect a "class clown" may have on a class. A class clown's lack of respect for a teacher's authority may become contagious, threatening a teacher's ability to control the class.[53]

Put more formally, we would like to understand the myopic best response dynamics that take place when one individual rebels against an ideal held by everyone else in the population.

To examine this question, we will consider a population in which every individual holds ideal I except possibly individual j, who may rebel against the common ideal and hold ideal I' instead. We

[53] Recall that it has been suggested that authority is generally based upon ideals. If a teacher has control over her class, students feel that they *should* do what the teacher tells them to do.

will consider whether individual i's desire to deviate from ideal I increases or decreases when j holds ideal I' rather than I.

An individual j rebelling against a common ideal I would maximize by choosing an ideal I' on the opposite side of the bliss point 0 from I ($sign(I) = -sign(I')$). Therefore, we will focus on this case.

There are two ways in which ideal I could be destabilized by j's rebellion. Individual i could be inclined to join j and conform to I' rather than I. Individual i might also be more inclined to rebel, choosing ideal I_i^{reb} rather than ideal I. We will consider individual i's incentive to conform to I' and i's incentive to rebel in turn.

The desire to conform to I'.

The utility that i receives from deviating from I to I' is: $D_i(I, ..., I, I_j, I')$. So, the effect of a rebellion by j on i's desire to deviate to I' is: $D_i(I, ..., I, I', I') - D_i(I, ..., I, I, I') = \sum_{n=1}^{3} \begin{bmatrix} D_i^n(I, ..., I, I', I') \\ -D_i^n(I, ..., I, I, I') \end{bmatrix}$. The following proposition examines the three terms of this expression, which correspond to the three concerns of Remark 4.

Proposition 9 *Suppose* $sign(I) = -sign(I')$.

(1) $D_i^1(I, ..., I, I', I') - D_i^1(I, ..., I, I, I') = 0$.

(2) $D_i^2(I, ..., I, I', I') - D_i^2(I, ..., I, I, I') \leq 0$.

(3) $D_i^3(I, ..., I, I', I') - D_i^3(I, ..., I, I, I') \geq 0$.

A rebellion by j increases i's status from holding I and decreases i's status from holding I', which explains part (2). A rebellion by j decreases the social confirmation of I and increases the social confirmation of I', which explains part (3). (1) is equal to zero because the disutility that i experiences from choosing an ideal that differs from 0 (concern 1 of Remark 4) only depends upon her own choice of ideal and not j's (see Definition 4).

We conclude that, if a rebellion by j to I' increases i's desire to deviate to I', it is because of i's concern about social confirmation. If a rebellion by j to I' decreases i's desire to deviate to I', it is because of i's concern about status.

The desire to rebel.

Let $R_i(I, I_j) = D_i(I, ..., I, I_j, I_i^{reb})$ and $R_i^n(I, I_j) = D_i^n(I, ..., I, I_j, I_i^{reb})$. R_i is i's desire to rebel. The

effect of a rebellion by j on i's desire to rebel is: $R_i(I, I') - R_i(I, I) = \sum_{n=1}^{3} [R_i^n(I, I') - R_i^n(I, I)]$. The following proposition examines this expression.

Proposition 10 *Suppose* $sign(I) = -sign(I')$.

(1) $R_i(I, I') - R_i(I, I)$ *has ambiguous sign.*

(2) If β *is sufficiently low,* $R_i(I, I') - R_i(I, I) \geq 0$.

(3) $R_i^3(I, I') - R_i^3(I, I) \geq 0$.

If j rebels, i receives less social confirmation from holding I, which increases i's desire to rebel. This explains part (3) of the proposition. But, j's rebellion can make it easier or more difficult for i to obtain status from rebelling. Therefore, unless i does not care about her status (β low), j's rebellion can increase or decrease i's desire to rebel. This explains parts (1) and (2) of the proposition.

We conclude from Propositions 9 and 10 that a rebellion by j can be either stabilizing or destabilizing. A rebellion by j is destabilizing with respect to concern (3) of Remark 4: i's desire to obtain social confirmation. But, a rebellion by j is potentially stabilizing with respect to concern (2) of Remark 4: i's desire to obtain social status.

4.3 The existence of ideals that are difficult to meet

One might, naively, be inclined to think that an equilibrium will generally exist in which everyone chooses ideal $I = 0$ and that this equilibrium will generally Pareto-dominate other equilibria that might exist (since this maximizes innate utility). If one is further inclined to think that equilibria that are Pareto-dominated tend not to arise, one might conclude that the model tends to predict that $a_i = I_i = 0$ for all i. Put another way, one might conclude that the introduction of social motivation into the utility function does not lead people to behave differently from how they would behave if they only had economic motivation (u_i).

But, this thinking is incorrect. Let us suppose Pareto-dominated equilibria will not arise. There are two reasons that we are unlikely to see everyone choose $I = 0$: (1) individuals have a desire to differentiate to gain status, and (2) individuals may benefit from cooperation. We will discuss these ideas in turn.

4.3.1 The desire to differentiate

We have seen that, because of the desire to differentiate, it may not be equilibrial for everyone to choose $I = 0$. The individuals who have the highest ability are especially inclined to rebel rather than conform to $I = 0$ (see Proposition 7). Proposition 8 implies that it is possible for an equilibrium to exist in which everyone conforms to an ideal $I \neq 0$ while an equilibrium does not exist in which everyone conforms to an ideal $I = 0$. The example developed in Section 3.1 (see also Proposition 14 in Appendix D) shows that the desire to differentiate can lead to the existence of a differentiating equilibrium in which two identical individuals take on different costly ideals ($I_i = \pi$, $I_j = -\pi$). This may be the only equilibrium. But, if a conforming equilibrium also exists in which $I_i = I_j = 0$, it may be the case that the conforming equilibrium is Pareto-dominated by the differentiating equilibrium.

Furthermore, when individuals have different abilities, it need not be the case that an equilibrium in which everyone conforms to $I = 0$ Pareto-dominates an equilibrium in which everyone conforms to $I' \neq 0$. If $\alpha_i = \alpha$ for all $i \in P$, the I equilibrium will Pareto-dominate the I' equilibrium. But, if individuals in the population have different abilities, a high ability individual achieves greater status in an I' equilibrium. A move to an I equilibrium may be undesirable for high ability types. Therefore, an I' equilibrium need not be Pareto-dominated.

4.3.2 Cooperation

If individuals benefit from cooperation, an equilibrium in which everyone conforms to $I = 0$ may be Pareto-dominated by an equilibrium in which everyone conforms to $I' \neq 0$. Section 3.2 gave an example. A Pareto-dominated equilibrium is perhaps unlikely to be focal. Hence it is somewhat less likely that an $I = 0$ equilibrium will arise when individuals have a reason to cooperate than when they do not. The following proposition shows that, when there is a reason to cooperate, an $I = 0$ equilibrium is likely to be Pareto-dominated.

Proposition 11 *Suppose, we amend Assumption 2 by assuming that:* $u_i = -f_1(\frac{1}{\alpha_i}|a_i|) + \sum_{j \neq i} g(a_j)$ *for all* i. *Suppose* g *is twice differentiable and* $g'(0) \neq 0$. *The set of social equilibria is identical to the set when* $u_i = -f_1(\frac{1}{\alpha_i}|a_i|)$. *Let* $S = \{I :$ *an equilibrium exists where everyone conforms to* $I\}$. *If a neighborhood of* 0 *is contained in* S *and* $f_2''(0) > 0$*:*

(1) The social equilibrium in which everyone conforms to $I = 0$ *is not Pareto-dominant.*

(2) If everyone has the same ability ($\alpha_i = \alpha$ for all i), the equilibrium in which everyone conforms to $I = 0$ is Pareto-dominated by an equilibrium in which everyone conforms to $I' \neq 0$ for some $I' \in S$.

We saw an application of this theorem in Section 3.2, where the introduction of a g function caused the $I = 0$ equilibrium to be Pareto-dominated by an $I' \neq 0$ equilibrium.

4.4 The perception of others

There are two important observations to make about the way in which individuals judge one another. The first observation is that, generally, as individuals' ideals diverge, their opinions of one another fall. This idea will be relevant in our discussion of oppositional culture in Section 5. The model predicts some exceptions to this rule, but the following proposition gives conditions where it holds.

Proposition 12 *Suppose individuals i and j choose ideal and behavior $(I_i, a_i^{BR}(I_i))$ and $(I_j, a_j^{BR}(I_j))$ respectively. If $sign(I_i) = -sign(I_j)$ or $\alpha_i = \alpha_j$: a change in I_i or I_j that increases the distance between the ideals ($|I_i - I_j|$) decreases i's opinion of j relative to herself ($s_i(j) - s_i(i)$) and j's opinion of i relative to herself ($s_j(i) - s_j(j)$).*

The example from Section 3 illustrates this proposition. In the conforming equilibria, i and j held the same ideals and i considered j an equal. In the differentiating equilibria, $I_i = -I_j = 4$ and i considered herself superior to j.

The second observation concerns an interesting feature of the model: the status individual i assigns to individual j depends not only upon the behavior of individuals i and j but upon the behavior of the entire population. Therefore, i's opinion of j ($s_i(j)$) is greater when individual k meets individual i's ideal less well. This gives us the following proposition.

Proposition 13 *If a change in individual k's behavior (a_k) increases k's deviation from both i's ideal and j's ideal ($|a_k - I_i|$ and $|a_k - I_j|$ both increase), individuals i and j increase their opinions of one another ($s_i(j)$ and $s_j(i)$ both increase).*

In Section 6.1, we will propose that social confirmation is better viewed–and better modeled–as the utility that individuals receive from considering how they are viewed by those with whom they interact.

In other words, the social confirmation individual i receives from interacting with individual j is a function of individual i's belief about $s_j(i)$.

When social confirmation is viewed in this way, Proposition 13 suggests that as $|a_k - I_i|$ and $|a_k - I_j|$ both increase, individuals i and j have stronger desires to interact. This has important implications that will be discussed in Section 7.

5 Oppositional Culture

The term "oppositional culture" refers to a minority of a population that holds different ideals from those of the majority, is viewed by the majority of the population as low status, and looks down on the majority in turn. We also tend to see minimal social interaction between members of such a minority and members of the majority. Oppositional cultures are a prediction of the model. For example, in Section 3, we saw that twins might adopt different ideals, look down on one another, and avoid social interaction (of course, in this case, the groups are of equal size).

James Coleman gives us an example of an oppositional culture in his classic 1961 study of schools in Northern Illinois. He describes a rebellious group of girls in Elmtown (one of the schools). They were considered to be part of the "rough crowd." They wore black "rock and roll" jackets, which, as Coleman puts it, were "a symbol in this school of orientation to a good time, cars, music, the skating rink, and unconcern with school." They listened to Elvis Presley instead of Pat Boone, and liked to smoke and drink. Coleman writes that "they are so oriented to out-of-school activities that even in this culture, where adult values have little currency, they do not have status among the adolescents as a whole."[54]

It is probably reasonable to say that an oppositional culture exists among a subset of African Americans and that this culture is particularly strong in the inner city. This case has been made by numerous scholars (see, in particular, Fordham and Ogbu (1986), Fordham (1996), Austen-Smith and Fryer (2005)). Austen-Smith and Fryer note that "acting white" is a pejorative used by some African Americans. It means that a person is meeting "white" ideals instead of the ideals they are supposed to meet (oppositional ideals). Austen-Smith and Fryer suggest that African American oppositional culture became particularly strong by the late 1960s, when the Black Power Movement was on the rise.

[54] Coleman (1961), pp. 205-6

Later in this section, we will suggest a possible reason for the rise of an African American oppositional culture consistent with this timeline.

The existence of gangs in inner cities may be one consequence of this culture. Martin Sanchez Jankowski argues that gangs rely upon support from their communities in order to survive. This is suggested by the experience of Comet, a sixteen-year-old member of a Los Angeles gang: "Two years ago we did some things that the community was very angry about. We had definitely messed up, and we paid for it. In fact, we almost died as an organization because we couldn't get people to join us....the community was so down on us and they told people not to join."[55]

They are supported, according to Sanchez Jankowski, for two reasons. First, they help to protect the community against other gangs. But, perhaps more importantly, the ideals of the gang are, to some extent, the oppositional ideals of the community. Wayne, a fifty-seven-year-old African American machinist and father of two girls, captures this sentiment: "I know people think gangs are terrible and that the kids are just disgusting individuals, but that just ain't what's going down with these kids. Hell, they ain't bad kids, sure you gots some, but most ain't bad, they just not willing to quietly sit back and let the society take everything and give them nothing."[56]

One of the things that we observe in Wayne's remarks in addition to his sympathy with gang members is an anger at society and a sense of injustice. We also see this in the Black Power Movement and in the race riots of the 1960s. This reflects the relationship between ideals and fairness: because oppositional groups have different ideals, they have different views about how they should and should not be treated.

In Section 4, it was shown why equilibria can arise in which oppositional cultures exist. Oppositional cultures form in the model because the desire to have a good opinion of oneself relative to others can lead to rebellion against commonly-held ideals. The adolescents of Elmtown, for example, may look down upon the "rough crowd" girls for their ideals and behavior, but because of their ideals and behavior, the "rough crowd" girls are also able to look down upon their fellow students.

Of course, status utility comes at a cost for the girls in the "rough crowd." They are part of a small group and inevitably must interact, to some extent, with people outside this group. This results in considerable discomfort (which corresponds to social disconfirmation in the model). Since there is a

[55] Sanchez Jankowski (1991), p. 194
[56] Sanchez Jankowski (1991), p. 182

cost as well as a benefit to rebellion, not everyone will choose to do it. Some girls will be willing to suffer low status to be part of the majority (trading status for social confirmation).

The Rise of African American Oppositional Culture. The Civil Rights Movement changed views about the status of African Americans. It clearly changed views among whites about African Americans. Interestingly, it also seems to have changed African American views concerning their own status. For example, in 1947, Clark and Clark performed an experiment in which they presented African American children with a black doll and a white doll. They found that 67 percent of the children chose to play with the white doll. Also, the majority of the children thought that the white doll rather than the black doll looked nice and had a nice color. When Hraba and Grant repeated the experiment in 1970, they found that a clear majority of African American children preferred the black doll.

It may be right to view the Civil Rights movement as having propagated an ideal that African Americans *should* be viewed as high status, both by others and by themselves. The increased status assigned to African Americans by whites and the increased pride that African Americans felt in themselves may have been, in part, the product of such an ideal.

Prior to the Civil Rights movement many surely felt that African Americans should *not* view themselves as high status. To have too high an opinion of oneself as an African American was considered "uppity."

We have remarked upon the frequent alignment of social and economic incentives because social incentives affect what is seen as fair. There was certainly a strong economic incentive not to be "uppity" prior to Civil Rights. The consequences were especially grave in the South. We see this in the case of Emmett Till. Till, a fourteen-year-old African American from Chicago, was warned by his mother before visiting Mississippi relatives in 1955 to "mind his manners" with white people and "if you have to get on your knees and bow when a white person goes past, do it willingly." While the precise events are unclear, Till allegedly whistled at a white woman, Carolyn Bryant, in a grocery story. To punish Emmett Till for his "uppity" behavior, Ms. Bryant's brother and half-brother brutally beat and then murdered him.[57]

[57] A theoretical understanding of some of the subtleties of these events requires a model of fairness and a model in which people can hold different roles, both of which are beyond the scope of this paper. For example, a model in which people can hold different roles makes clearer why status might relate to the right to whistle at Carolyn Bryant. A model of fairness would make clearer why Till's behavior (or alleged behavior) created a desire to punish him.

If we accept that the Civil Rights Movement heightened the feeling among African Americans that they should have pride in themselves–that they should consider themselves high status–it is interesting to ask what effect this might have had on the behavior of African Americans and on their other ideals.

We observe that instilling in individuals a belief that they should have pride in themselves is very similar to increasing the parameter β (concern about status, defined in Section 4) of these individuals. Proposition 1, part 4, said that, an increase in β increases an individual's desire to meet her ideals well. So, one prediction of the theory is that the Civil Rights Movement should have caused African Americans to meet their ideals more closely. Prior to Civil Rights, there may have been a careful attempt to *avoid* meeting some ideals too closely to avoid seeming "uppity."

It would be wrong to say that, prior to the Civil Rights Movement, there were no discernible differences between white ideals and African American ideals. But many ideals surely were held in common. The theory predicts that African Americans would have felt compelled to meet ideals shared with whites more closely after Civil Rights than they had before *if* these ideals were maintained.

This brings us to the second prediction of the theory. Proposition 5 says that an individual who has difficulty achieving high status if she holds an ideal I will be more inclined to *rebel*, taking on an oppositional ideal, if she becomes more concerned about status (β increases).

The theory therefore suggests that those African Americans who were thwarted in their attempts to meet ideals shared with whites–thwarted in their attempts to move into the middle class–would have become inclined to rebel. African Americans who were unable to move into the middle class would still have felt a need to view themselves as high status. And, under the circumstances, only one means was available: a change of ideals. By adopting an oppositional culture, African Americans who were unable to move into the middle class could still justify a belief in high status.[58]

Appendix A provides a numerical example that illustrates this argument. In the example, before Civil Rights, African Americans choose to hold the same ideals as whites. Following Civil Rights, it is only those African Americans who find it difficult to meet white ideals who become oppositional.

[58] As mentioned earlier, the race riots of the 1960s seem to be related to the formation of an African American oppositional culture. The National Advisory Commission on Civil Disorders (the Kerner Commission), charged by President Johnson with the task of investigating the riots, came to conclusions that are consistent with those posited in this section. The commission's report concluded that: "The expectations aroused by the great judicial and legislative victories of the civil rights movement have led to frustration, hostility and cynicism in the face of the persistent gap between promise and fulfillment. The dramatic struggle for equal rights in the South has sensitized Northern Negroes to the economic inequalities reflected in the deprivations of ghetto life." (p. 204) The report also noted that: "Self esteem and enhanced racial pride are replacing apathy and submission to 'the system.'" (p. 205)

The prevalence of oppositional culture in inner cities may be explained by the special impediments to advancement faced by African Americans living in urban areas. In addition to the impediments to advancement faced by all African Americans, urban African Americans were forced to confront major deindustrialization in the 1960s. William Julius Wilson writes that "of the changes in the economy that have adversely affected low-skilled African American workers, perhaps the most significant have been those in the manufacturing sector."[59][60]

Fryer and Levitt (2004) give a nice piece of evidence for both predictions of the theory. They find that, in the early 1960s, there was little difference between the types of names chosen by African Americans and whites for their children. But, a major shift took place in the late 1960s and early 1970s. The median African American female in a segregated neighborhood in California went from receiving a name that was twice as likely to be given to African Americans as whites to receiving a name that was twenty times as likely to be given to African Americans as whites. Fryer and Levitt find that, at the same time, a subset of African Americans, comprising roughly one quarter of all African Americans and one half of African Americans living in predominantly white areas, moved towards names that were more white than those they had chosen previously.

As mentioned earlier, the race riots of the 1960s seem to be related to the formation of an African American oppositional culture. The Kerner Commission, charged by President Johnson with the task of investigating the riots, came to conclusions that are consistent with those posited in this section. The commission's report concluded that: "The expectations aroused by the great judicial and legislative victories of the civil rights movement have led to frustration, hostility and cynicism in the face of the persistent gap between promise and fulfillment. The dramatic struggle for equal rights in the South has sensitized Northern Negroes to the economic inequalities reflected in the deprivations of ghetto life."[61] The report also noted that: "Self esteem and enhanced racial pride are replacing apathy and submission to 'the system.'"[62][63]

[59] Wilson (1996), p. 29.

[60] The model also suggests that the emergence of oppositional cultures in urban areas would have been facilitated by the concentration of urban African Americans in ghettos and their isolation from whites. According to the theory, social interaction with whites increases the cost to becoming oppositional (it creates social disconfirmation). The Kerner Commission concluded that segregation and the concentration of urban African Americans in ghettos were important reasons for the riots (see pp. 203-4). An outmigration of wealthier African Americans from the inner city also took place after the Civil Rights Movement (see Footnote 64), which may have contributed to the formation of an oppositional culture.

[61] *Report of the National Advisory Commission on Civil Disorders*, p. 204.

[62] *Report of the National Advisory Commission on Civil Disorders*, p. 205.

[63] We see evidence for enhanced racial pride and its connection to oppositional culture in the following remark made by

William Julius Wilson points to deindustrialization of cities as important in creating an oppositional culture among African Americans living in the inner city. We have suggested here that the interaction between deindustrialization and Civil Rights may have been crucial.[64]

Deindustrialization alone does not seem to be sufficient to understand why nonemployment rates in the inner city have remained so high—particularly among African American men.[65] A change in ideals, however, may have been important in sustaining high rates of unemployment. While a full theory of fairness is a topic for a further paper, it is a reasonable view that higher status is likely to increase the wage that one feels that one deserves. Inner city African Americans, with views of themselves as high status but possessing few skills, may consider a fair wage to be well above the market wage, resulting in high unemployment.[66][67]

Another important aspect of the culture of the inner city seems to be a belief, somewhat widely held, that opportunities do not exist to enter the middle class. Or, at least, one cannot enter the middle class by conventional means: working hard and getting a good education.[68] These beliefs may have some merit since the quality of education in the inner city is generally poor and seeking a good education may be viewed as "acting white." But, these beliefs are probably also part of an oppositional culture, motivated by a desire for status. Sanchez Jankowski writes of the people he interviewed: "most of these people did not make this point explicit, but in numerous conversations they implied that they

Malcolm X, a leader of the Black Muslims. He writes, "It's a crime, the lie that has been told to generations of black men and white men both. Little innocent black children, born of parents who believed that their race had no history. Little black children seeing, before they could talk, that their parents considered themselves inferior. Innocent black children growing up, living out their lives, dying of old age–and all of their lives ashamed of being black. But the truth is pouring out of the bag now." (X and Haley (1964), p. 185.)

[64] Wilson, in fact, suggests another reason–besides increased pride–why the Civil Rights Movement may have contributed to the formation of an oppositional culture. He writes that "especially since 1970, inner-city neighborhoods have experienced an outmigration of working- and middle-class families previously confined to them by the restrictive covenants of higher-status city neighborhoods and suburbs. Combined with the increase in the number of poor caused by rising joblessness, this outmigration has sharply concentrated the poverty in the inner city." (p. 12, Wilson (1993).) He sees the social isolation of the poor as a factor contributing to the emergence of an oppositional culture. See, especially, Wilson (1987) for an elaboration of these ideas.

[65] See Holzer and Offner (2002) and Holzer and Freeman (1986). Holzer and Offner (2002) find that employment rates and real wages of young African American men remained low throughout the 1990s in spite of the booming economy.

[66] See Akerlof and Yellen (1988, 1990) for a discussion of fair wages.

[67] There is some evidence suggesting that African American women are less oppositional than African American men, which may explain why the nonemployment rate of African American women responded more to the economic boom of the 1990s than the nonemployment rate of African American men (see Holzer and Offner (2002)). For example, Kling, Ludwig, and Katz (2005), examining the Moving to Opportunity experiment, found that when low-income African American families relocated to lower-crime, lower-poverty neighborhoods, the girls tended to integrate better into their new social context than the boys. Perhaps African American women have a lower opinion of their status than the men and therefore have less of a need to adopt oppositional ideals. It would be interesting to test this hypothesis.

[68] There is some recent work in economics concerning whether people believe the world to be fair and just. See Alesina and Angeletos (2005) and Benabou and Tirole (2006a). Whether the world is fair and just has been a major issue in the debate over Welfare. See Ellwood (1988).

found some comfort that their children were having the same difficulty and facing the same prospects as they had, because it confirmed that they were not the only losers and relieved them of personal feelings of failure. Thus, they responded to their children's involvement in gangs with an absence of discouragement (which acted as a form of encouragement). They knew that gang involvement would eventually result in difficulty with the law, but their 'encouragement' was based on an ambivalence about wanting a better life for their children."[69]

6 Extensions of the Model

This section explores three ways in which the model developed in Section 2 can profitably be extended. First, we will develop an alternative way to model and think about social confirmation of belief. In Section 7, this way of thinking about social confirmation will be applied to develop an understanding of the way in which people categorize themselves and others. Second, we will endogenize the cost associated with deviating from an ideal. Finally, we will show how the model can be extended to develop an understanding of how people choose their beliefs about the state of the world. Appendix B develops a model that incorporates these extensions.

6.1 Endogenizing Social Confirmation

In Section 2, we assumed that the social confirmation individual i receives depends upon the difference between individual i's beliefs and the beliefs of those with whom individual i interacts.

$$C^S(I_i, I_{-i}, G_i) = \theta_{|G_i|}(\{m(I_i, I_j)\}_{j \in G_i})$$

An alternative way of viewing social confirmation will now be proposed. It will then be shown how the new formulation can be reconciled with the old.

The alternative view of social confirmation is as follows: social confirmation is the pleasure or displeasure individuals receive from thinking about how they are viewed by others. When we say that "individual i interacts with individual j," what we mean is that individual i is cued to think about how j views i.[70] With this in mind, we will define social confirmation in the following way.

[69] Sanchez Jankowski (1991), p. 182.

[70] George Homans observes that "persons who interact frequently with one another tend to like one another." (Homans (1950), p.111) This would seem to be consistent with this view of social confirmation.

$$C^S(\{s_j(i)\}_{j\in P}, G_i) = \lambda_{|G_i|}(\{s_j(i)\}_{j\in G_i})$$

$\lambda_{|G_i|}$ is assumed to be increasing in its arguments.[71]

It may not be immediately obvious how the new definition relates to the old. In order to relate the two, it is necessary to assume that individual j judges individual i not just based upon i's behavior (as we have assumed in the past) but also upon i's *beliefs*. Therefore, we need to assume that there are ideals pertaining to belief as well as behavior.

It is natural to assume that there are ideal things to believe as well as ideal ways to behave. Consider two examples. According to Roman Catholicism, we can sin through our thoughts as well as our actions. It is considered sinful, for example, to wish evil upon another. A second example is the importance of intentions as well as actions in determining how individuals are judged. Marc Hauser finds, for example, that people give far less praise to a CEO whose actions unintentionally help the environment than one who helps the environment intentionally.[72]

If there are ideals over beliefs, when individual i interacts with individual j, individual i gains utility from holding the beliefs that individual j believes i should hold and loses utility from holding beliefs that j believes are inappropriate. Because individual j can view herself well relative to others if she believes that others should believe what she believes herself, individual j will generally look down upon those who hold different beliefs. It is for this reason that individual i will generally have a desire to believe what j believes when i interacts with j. This shows how the new formulation of social confirmation relates to the old formulation.

This version of social confirmation gives us a very precise understanding of the motivations in Asch's classic line experiment. When a subject has different beliefs from the confederates or gives a different answer (as subjects often did in the experiment) this creates the uncomfortable feeling that the confederates have a low opinion of the subject.[73]

[71] We probably also want to assume that, if $k \notin G_i$ and $b_i(s_j(i))$ is sufficiently large, $C_i^S(G_i \cup \{k\}) > C_i^S(G_i)$. And, if $k \notin G_i$ and $b_i(s_j(i))$ is sufficiently small, $C_i^S(G_i \cup \{k\}) < C_i^S(G_i)$.

[72] Hauser (2006), p. 51.

[73] In a classic study, Asch (1951) asked subjects to match the length of a given line to one of three other lines. The success rate is very high when subjects are asked to perform this task individually. Asch compared the individual results to performance when there were seven confederates in the room, all of whom gave the same wrong answer. In this case, Asch found that subjects gave the correct response in only two thirds of the cases. (Sherif (1936, 1965) performed a similar experiment involving the movement of a spot of light.)

Subsequent work concluded that beliefs and not just responses were being affected (Ono and Davis, 1988). An important question is whether the changes in belief took place because the subjects thought the answers of the confederates contained information or because they wanted to avoid social disconfirmation.

The simplicity of the question and the high success rate in answering the question when confederates were absent

The new formulation suggests that, even in cases where an individual's ideals are not fundamentally altered by social interaction, some attempt may be made to conform to the ideals of others in order to make a favorable impression (whether or not there are economic reasons to do so).[74] The model predicts that a change in ideals might take place but lag behind a change in behavior.[75] Often, a new member of a group does not entirely share the group's ideals but will try to fit in by behaving in a way that is consistent with the group's ideals (to obtain social confirmation). In time, as the new member becomes good at meeting the ideals of the group, she finds it worthwhile to adopt them (which gives her additional social confirmation).

An individual might even distort her record in order to make a favorable impression (whether or not there is an accompanying economic reason to do so). It is common for people to try to make good impressions on others by selectively presenting aspects of themselves or even, sometimes, making fabrications. A lieutenant, for example, in speaking with a general, is likely to bring up his military accomplishments but actively avoid discussing his passion for heavy metal music, which, he suspects, the general does not share.[76]

6.2 Endogenizing the cost of deviation

In Section 2, the cost of deviating from an ideal (the d function) was taken as exogenous. The cost of deviation can easily be endogenized. It can be endogenized by treating d as a belief in the same way that I_i is a belief. Therefore, we assume that individuals not only have beliefs about the ideal way to behave but also beliefs about how people should be judged for given deviations from the ideal. Like ideals, we can assume that beliefs about the cost of deviation are socially confirmed but are not informationally confirmed.

With this endogenization, it is natural to ask why it might be equilibrial for there to be a nonzero cost of deviating from an ideal. This is an important assumption of the model. If it were violated, the model would break down. This question is very similar to a question asked and answered in Section

hints that the answer is social confirmation rather than information. Perhaps the best piece of evidence in favor a non-informational story is that subjects who answered correctly in the presence of confederates reported afterwards that they were extremely anxious about giving an answer that differed from that of the group. It is unlikely that so much anxiety would be aroused by the decision itself. It seems rather likely that the anxiety is due to social disconfirmation.

[74] As we have discussed, economic and social incentives are frequently aligned.

[75] Social psychologists have observed that behavioral changes often precede changes in attitude and belief. See, for example, Fishbein and Ajzen (1975).

[76] This is an aspect of what Erving Goffman calls "face-work." See Goffman (1967), pp. 5-45 for a further discussion.

4: why do ideals arise that are costly to meet? And, the answer is virtually identical. The desire to obtain status (look down on others) can lead to a nonzero cost of deviation. It is also possible, as in Section 4, that there could be a nonzero cost of deviation because this allows cooperation to take place that is in the general interest.

This endogenization, in addition to giving a more complete model, also tells some new stories. For example, the endogenization allows individuals to hold an ideal more or less strongly in the sense that the cost of deviation can be high or low. Therefore, we can tell a story about when an individual will hold an ideal strongly and when an individual will hold an ideal weakly.

Suppose an ideal is widely held, such as the ideal of being a good scholar. An individual who is not a good scholar can gain status by rebelling against the ideal, as we explained earlier. But, there is another way of gaining status: an individual can choose to believe that the cost of deviating from the ideal of being a good scholar is low. It may be preferable to assume a low cost of deviation rather than adopt an oppositional stance when the ideal of being a good scholar is very widely held.

Therefore, when an ideal is widely held, we might expect people who find it easy to meet the ideal to hold the ideal strongly while those who find it hard to meet the ideal to hold the ideal weakly. This can accentuate differences in performance across people of different abilities: those who perform well are highly motivated to perform well (they assume a high cost of deviation) which makes them perform better while those who perform less well are unmotivated and so perform worse than they are capable of doing.

6.3 Extending the Framework to Include Other Beliefs

In the model developed in Section 2, individuals only chose beliefs about ideal behavior. The framework can be extended to consider how people choose beliefs about the state of the world. Extending the model to understand other beliefs is highly profitable. It tells us why, for example, a white employer might favor a white job applicant to a black job applicant even though, objectively, the black applicant is more qualified. The white employer may have an incentive to hold beliefs about the white and black applicants that do not correspond perfectly to the objective reality. This gives a reason for discrimination beyond pure statistical discrimination.[77]

[77] This is not the only reason that the model gives for discrimination. The ideals that a person holds can also lead to discriminatory behavior.

One of the most important reasons to tailor one's beliefs is the desire to think well of oneself (the desire to have high $s_i(i)$). This may underlie the bias of our white employer. We see an example of this type of belief tailoring in the "minimal group studies." In these experiments, Henri Tajfel and his colleagues showed that even random assignment of subjects who are well acquainted prior to the experiment into two groups (A and B) can create bias. Members of group A tend to believe that they are kinder, more trustworthy, and better looking than members of group B (and vice-versa). This is true *even* when subjects are aware that the assignment is random.[78]

The theory developed in this paper suggests the following interpretation of the minimal group studies. An individual in group A has a desire to see membership in group A as something that gives status. Individuals in group A choose beliefs that differ from the information-based beliefs because of this desire. Individuals in group A try to believe that membership depends upon superiority along a number of dimensions. They view members of group A as superior because this is consistent with the belief that selection for group A is based upon superiority.[79]

In order to model the choice of other beliefs, we need to give a specification of information-based confirmation of belief. We need to consider how beliefs about the state of the world are informationally confirmed or disconfirmed. The implementation suggested here is similar to the model given by Brunnermeier and Parker (2005).

Let $\omega \in \Omega$ denote a possible state of the world. We consider a set M of probability measures over Ω. The true probability of each state of the world is given by the probability measure $\mu^* \in M$. Individuals do not know the true function μ^*, but they have an opinion about what the probability function looks like. We will denote the opinion of individual i by $\mu_i \in M$. This is individual i's prior. We can also think of it as individual i's "system of belief." It tells us how individual i will interpret the information that she possesses about the state of the world. Let us denote the information held by individual i about the state of the world by $X_i \subseteq \Omega$. A "belief" of individual i is the probability individual i assigns to some event $A \subseteq \Omega$ given the information she holds: $\mu_i(A|X_i)$.

We will now make an important definition:

[78] The orginial study is Tajfel, Flament, Billig, and Bundy (1971). For a review of the minimal group literature, see Brown (1986), pp. 543-551, or Haslam (2001), pp. 27-31.

[79] See Sidanius and Pratto (1999) Chapter 2 for a discussion of beliefs that exist to support a group's view of its status. Sidanius and Pratto refer to these beliefs as "legitimizing myths."

$$\begin{aligned} \Psi(\mu_i, X_i) &= \frac{\mu_i(X_i)}{\mu_{ML}(X_i)} \\ \text{where } \mu_{ML} &= \arg\max_{\mu' \in M} \mu'(X_i) \end{aligned}$$

$\Psi(\mu_i, X_i)$ answers the following question: is there a different system of beliefs, μ_{ML}, that does a better job of explaining the information held by individual i? We can think of μ_{ML} as the maximum likelihood estimate. Thus, Ψ tells us how much individual i's estimate, μ_i, differs from the maximum likelihood estimate, μ_{ML}.

Information-based confirmation. We can now provide a specification of information-based confirmation. We will assume that information-based confirmation is given by: $C^I = \Gamma(\Psi(\mu_i, X_i))$ where Γ is an increasing continuous function. Therefore, when μ_i is close to the maximum likelihood estimate, individual i's beliefs are confirmed. When μ_i is far from the maximum likelihood estimate, individual i's beliefs are disconfirmed.

Knightian Uncertainty. Knightian uncertainty means that there is little or no information available to an individual that indicates the probability of an event. When there is Knightian uncertainty, many beliefs can explain the information one possesses almost as well as the maximum likelihood beliefs. In fact, we could take this as a definition of Knightian uncertainty. According to the notation we have developed, $\Psi(\mu_i, X_i)$ is close to one for a large set of μ_i when there is uncertainty. Hence, uncertainty implies that there are many beliefs for which there is little or no information-based disconfirmation.

Beliefs are not pinned down by information when there is uncertainty. Just like ideals, the only type of confirmation that is relevant is *social* confirmation. The model therefore predicts that social forces play a much stronger role in belief formation when there is uncertainty than when people have real information to anchor them. This idea is one of the principles of Festinger's social comparison theory: "To the extent that objective, non-social means are not available, people evaluate their opinions and abilities by comparison respectively with the opinions and abilities of others."[80]

We suggested earlier that ideals are beliefs for which there is no relevant information and therefore complete uncertainty. It is for this reason that social confirmation is the only type of confirmation that is relevant to determining ideals in the model in Section 2.

[80] Festinger (1954), Hypothesis 2, p. 118.

As we noted in the introduction, ambiguity (or a lack of information) potentially has value for individuals. When there is uncertainty, individuals' beliefs are not constrained by information-based confirmation. This makes it easier for individuals to believe that they are above average. Indeed, we find empirically that people tend to consider themselves to be above average on almost every positive trait.[81] Ambiguity helps to eliminate the zero-sum nature of status that exists when everyone holds the same, information-based beliefs.

For example, suppose that the ideal in a population is to perform as well as possible on a standardized math test. If individuals have little information about the distribution of test scores, they can choose to believe that they performed well relative to others—even if they did not—with little information-based disconfirmation. Therefore, when there is a paucity of information, everyone can believe that they are high status.[82] But, the more information individuals have about the score distribution, the greater the information-based disconfirmation experienced from beliefs that are not information-based. Therefore, the better the information individuals have, the less status there is to go around and the more status becomes a zero-sum game.

Ideals often develop that promote such ambiguities.[83] Consider the following comic example. In speaking to Susan, Jane happens to spit slightly in Susan's face. The appropriate response for Susan is to act as if it did not happen, even if it is obvious to both parties that it did. Susan should not even wipe the spittle away until Jane turns her gaze. We also see such ideals in the workplace. When a boss rebukes an employee, it is frequently done in the most minimal terms to preserve the employee's self esteem.

7 Categorization

Proposition 13 says that, an increase in the deviation of an individual k from the ideals of i and j increases i's opinion of j and j's opinion of i. Section 6.1 argued that individuals with higher opinions of one another are more inclined to interact. These ideas have important consequences for the way in

[81] See, for example, Svenson (1981).

[82] This may be a reason for money illusion. When an individual is uncertain about the rate of inflation, she is inclined to interpret a nominal wage increase as a real wage increase (since a real wage increase makes an individual feel good about herself). Furthermore, there is a desire to maintain uncertainty about the inflation rate in order to feel good about oneself. This story appears to be consistent with evidence on money illusion presented by Shafir, Diamond, and Tversky (1997).

[83] See Goffman (1959).

which individuals categorize themselves and others. In this section, we will discuss these consequences and relate them to an existing literature in social identity theory.

An important idea in this section is the notion of a "perceived difference" between individuals. A "perceived difference" is the difference in *status* associated with an objective difference.[84]

For example, suppose individuals i and j have blue eyes and believe that people should have blue eyes while individual k has brown eyes and believes people should have brown eyes. If individuals i and j assign considerably lower status to k for having brown eyes, i and j perceive having brown eyes rather than blue to be an important difference.

A perceived difference depends upon two things. First, it depends upon objective differences. Therefore, i and j perceive the difference between themselves and k to be smaller if k has blue eyes rather than brown eyes. But, perceived differences also depend upon individuals' beliefs about the loss in status associated with an objective difference. The d functions that individuals i and j hold determine precisely how much status individual k loses for having brown eyes rather than blue eyes (and, hence, their perception of the difference between themselves and k). Recall that, in Section 6.2, we made the d function a choice, arguing that it was a belief in the same way that an ideal is a belief.

From Proposition 13 we can conclude that the greater the *objective* difference between i's and j's blue eye color and k's eye color, the greater will be i's opinion of j and j's opinion of i. A more general statement is: the greater i and j *perceive* the difference to be between their blue eye color and k's eye color, the greater will be i's opinion of j and j's opinion of i.

Section 6.1 concluded that an individual has a desire to interact with those who view her well. Hence, the greater i's and j's perception of the difference between their eye color and k's eye color, the more inclined i and j will be to interact. If i and j start to interact, they are likely to *increase* the degree to which they perceive each other to be similar: if they choose to perceive one another as more similar, they can generally reduce social disconfirmation.

This has important consequences that we can see by adding a second trait to our example: height. Suppose individuals i and k are tall while j is short. Suppose i, j, and k view their own combination of traits (eye color and height) to be ideal. The more weight individuals put on eye color as a determinant of status, the more i and j will choose to interact to the exclusion of k and the *less* weight i and j will

[84] Arguably, the way in which people define themselves and the categories that people consider to be important depend more upon perceived differences than objective differences.

put on height as a determinant of status. Similarly, the more weight individuals put on height as a determinant of status, the more i and k will interact to the exclusion of j and the less weight i and k will put on eye color as a determinant of status.

We see that the importance of eye color as a category is related to the importance of height as a category. When one category is more important, the other is likely to be less important.

"Comparative fit" is a term used by social identity theorists that corresponds closely to the idea conveyed in this example. Alexander Haslam (2001) gives the following definition of comparative fit: "a principle of category fit which suggests that a given category is more likely to become salient to the extent that differences between members of that category are perceived to be smaller than the differences between members of that category and comparison others."

Numerous experiments offer support for the idea of comparative fit. Haslam and Turner (1992), for example, have looked at the importance of context in the way people define themselves. They find that workers will differentiate themselves from one another in an all-worker context. But, when the context is extended to include managers as well, workers will downplay differences among workers. Hogg and Turner (1987) have shown that individuals in a same-sex group were less likely to define themselves in terms of gender than individuals organized in coed groups.

The use of the principle by leaders. Glaeser (2005) has argued that politicians sometimes spread hate in order to further their political ends. We will suggest in this section that, when politicians spread hate to further their political ends, politicians are often exploiting the principle of comparative fit. More generally, a leader in an organization may be able to use comparative fit to bolster her authority.

Comparative fit says that, the more there is a perceived difference between a set of people, S, and $P-S$, the more individuals in S will choose to interact (we could say that they become more "cohesive" or "groupy") and the less will be the perceived difference within S.

While leadership is not modeled formally in this paper, it has been suggested that leadership is a kind of ideal. A "follower" holds an ideal that she *should* do what the leader tells her to do. As with all ideals, it is a choice to be a follower, so a leader's power extends only so far. It might be in a leader's interest–in the sense that she might be able to push her followers further–to have followers

who are groupy and perceive themselves to be similar.[85] There are probably many reasons that groupy followers could be useful to a leader: one reason will be given momentarily.

A leader, exploiting comparative fit, can make her followers more groupy if she has a means to increase the perceived difference between her followers and others. Politicians in the United States often speak about what makes Americans special. This is probably a subtle attempt to create a perceived difference. Spreading hate is somewhat less subtle.

The most sinister invocations of comparative fit seem to have a common form, exploiting comparative fit to enhance a leader's power in a very particular way. Stalin was arguably using this form when, "against the wishes of Lenin and his family, Stalin orchestrated the effective deification of the leader and his embalming like an Orthodox saint in a Mausoleum in Red Square."[86] Hitler's commemoration of the Beer Hall Putsch and its sixteen "blood martyrs" also exploited the principle.[87]

The common form is as follows. A leader uses her position to increase concern about a shared value, perhaps a value that is threatened by those outside the group.[88][89] The leader thereby increases cohesion among followers and instills a sense of groupiness. The leader now sets ideals for the group (put another way, gives orders) that, prior to her manipulation, would have faced considerable resistance. But, in the current environment, an individual who resists indicates a lack of groupiness, and therefore a lack of concern for the shared values of the group.[90] So, the group is more inclined to look down upon an individual who fails to capitulate to the leader.

At the Nuremberg Trials, Hermann Goering gave a description of the technique: "It is always a simple matter to drag the people along, whether it is a democracy, or a fascist dictatorship, or a parliament, or a communist dictatorship....Voice or no voice, the people can always be brought to the bidding of the leaders. That is easy. All you have to do is to tell them they are being attacked, and

[85] Murphy and Shleifer (2004) have argued that a strong network can give power to a politician.

[86] Sebag Montefiore (2003), pp.35-6.

[87] A ceremony, attended by Hitler, was held every year in Munich to commemorate the event. It usually included a staging of the march from the Burgerbräukeller to the south side of the Feldherrnhalle. Every administrative region (Gau) held a ceremony as well.

[88] In commemorating the Beer Hall Putsch, for example, the Nazis were indicating their dream of a unified Greater Germany that would celebrate a particular conception of what it meant to be German. The Putsch represented the threat to this dream (in the form of Bavarian independence) as well as the likelihood of eventual Nazi triumph.

[89] How a leader is able to increase concern about some value is a topic of some interest. In part, the answer may be that a campaign to expose followers to the leader's values can create a network. Creation of a network is likely facilitated by existing feelings among followers that they *should* believe what their leader believes.

[90] Here is the relation to our original example. A person who indicates that she considers eye color to be important is implicitly saying that she considers height to be unimportant. Therefore, a group that chastises people for viewing height as unimportant will also chastise people for viewing eye color as important.

denounce the pacifists for lack of patriotism and exposing the country to danger. It works the same in any country." Implicit in Goering's statement is the idea that those who disagree with the program *can* be attacked for lack of patriotism. The reason that a populace might be willing to tolerate such persecution is that, in such an environment, those who object are viewed as violating an important ideal.[91]

8 Conclusion

This paper has built a model that provides micro foundations for various concepts in sociological theory, such as social norms, status, and groups. We considered various implications and applications of the model. In particular, we showed that social motivation can lead individuals to behave in ways that differ from how they would behave if they only possessed economic motivation. There are two main reasons: there is a desire to differentiate to gain status and ideals may arise to facilitate cooperation (see Section 4.3). In Section 5, we used the theory to suggest an explanation for the emergence of an oppositional culture among a subset of African Americans in the late 1960s. Section 7 considered what the model predicts about the categories that people will find salient. These predictions were related to an existing literature in social identity theory and the principle of "comparative fit."

This paper also provides an agenda for future work: utilizing the current framework to account for leadership and fairness.

[91] Both the desire not to violate an ideal (a social incentive) and the desire to avoid persecution (an economic incentive) drive compliance in such circumstances. But, the persecution is viewed as "fair" (or is, at least, better tolerated) by the populace because of the violated ideal. In this sense, the social incentives–not just the economic–are key to understanding this phenomenon.

References

Acemoglu, Daron, Simon Johnson, and James Robinson, "The Colonial Origins of Comparative Development: An Empirical Investigation." *American Economic Review,* 2001, 91, 1369-1401.

Acemoglu, Daron, Simon Johnson, and James Robinson, "Reversal of Fortune: Geography and Institutions in the Making of the Modern World Income Distribution." *Quarterly Journal of Economics,* 2002, 117, 1231-1294.

Akerlof, George, "The Economics of Caste and of the Rat Race and Other Woeful Tales." *Quarterly Journal of Economics,* 1976, 90(4), 599-617.

Akerlof, George, "Social Distance and Social Decisions." *Econometrica,* 1997, 65(5), 10005-1027.

Akerlof, George and William Dickens, "The Economic Consequences of Cognitive Dissonance." *American Economic Review,* 1982, 72(3), 307-319.

Akerlof, George and Rachel Kranton, "Economics and Identity." *Quarterly Journal of Economics*, 2000, 115(3), 715-753.

Akerlof, George and Rachel Kranton, "Identity and the Economics of Organizations." *Journal of Economic Perspectives*, 2005, 19, 9-32.

Akerlof, George and Janet Yellen, "Fairness and Unemployment." *American Economic Review Papers and Proceedings,* 1988, 78(2). 44-49.

Akerlof, George and Janet Yellen, "The Fair Wage/Effort Hypothesis and Unemployment." *Quarterly Journal of Economics*, 1990, 105(2), 255-283.

Axelrod, Robert, "The Emergence of Cooperation Among Egoists." *American Political Science Review,* 1981, 75, 306-318.

Axelrod, Robert, *The Evolution of Cooperation.* New York, NY: Basic Books, 1984.

Alesina, Alberto and George-Marios Angeletos, "Fairness and Redistribution: US versus Europe." *American Economic Review,* 2005, 95(4), 960-980.

Allen, Jeremiah, "A Modification of the Akerlof/Kranton Model of Schooling; a mystery and a solution which tests the model." mimeo, 2006.

Aronson, E., *The Social Animal, Ninth Edition.* New York, NY: Worth Publishers, 2004.

Asch, S., "Effects of Group Pressure upon the Modification and Distortion of Judgments" in H. Guetzkow (Ed.), *Groups, leadership, and men.* New Brunswick, NJ: Rutgers University Press, 1951.

Austen-Smith, David and Roland Fryer, "An Economic Analysis of 'Acting White.'" *Quarterly Journal of Economics*, 2005, 120(2), 551-583.

Baron-Cohen, Simon, *Mindblindness: An Essay on Autism and Theory of Mind.* Cambridge, MA: MIT Press, 1995.

Basu, Kaushik and Jörgen Weibull, "Punctuality: A Cultural Trait as Equilibrium," in R. Arnott, B. Greenwald, R. Kanbur and B. Nalebuff (Eds.), *Economics for an Imperfect World: Essays in Honor of Joseph Stiglitz.* Cambridge, MA: M.I.T. Press, 2003.

Becker, Gary S. and Kevin M. Murphy, *Social Economics: Market Behavior in a Social Environment.* Cambridge, MA: The Belknap Press, 2000.

Benabou, Roland and Jean Tirole, "Self Confidence and Personal Motivation." *Quarterly Journal of Economics*, 2002, 117(3), 871-915.

Benabou, Roland and Jean Tirole, "Intrinsic and Extrinsic Motivation." *Review of Economic Studies*, 2003, 70(3), 489-520.

Benabou, Roland and Jean Tirole, "Willpower and Personal Rules." *Journal of Political Economy*, 2004, 112(4), 848-886.

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

Benabou, Roland and Jean Tirole, "Incentives and Prosocial Behavior." forthcoming, *American Economic Review*, 2006, 96(5), 1652-1678.

Benabou, Roland and Jean Tirole, "Identity, Dignity, and Taboos: Beliefs as Assets." Mimeo, 2006.

Benedict, Ruth, *Patterns of Culture.* New York, NY: Mariner Books, 1934.

Bernheim, Douglas B., "A Theory of Conformity." *Journal of Political Economy*, 1994, 102(5), 842-877.

Besley, Timothy and Maitreesh Ghatak, "Competition and Incentives with Motivated Agents." *American Economic Review*, 2005, 95(3), 616-636.

Brown, Roger, *Social Psychology: The Second Edition.* New York, NY: The Free Press, 1986.

Bourdieu, Pierre, *Outline of a Theory of Practice.* Cambridge, UK: Cambridge University Press, 1977.

Bourdieu, Pierre, *Distinction: A Social Critique of the Judgment of Taste.* Cambridge, MA: Harvard University Press, 1984.

Brunnermeier, Markus and Jonathan Parker, "A structural model of 'optimal' belief distortions due to anticipatory utility." *American Economics Review*, 2005, 95(4), 1092-1118.

Clark, K. and M. Clark, "Racial identification and preference in Negro children" in T. Newcomb and E. Hartley (Eds.), *Readings in Social Psychology.* New York, NY: Holt, 1947, 169-178.

Cole, Harold, George Mailath, and Andrew Postlewaite, "Social Norms, Savings Behavior, and Growth." *Journal of Political Economy*, 1992, 100 (6), 1092-1125.

Coleman, James, *The Adolescent Society: The Social Life of the Teenager and Its Impact on Education.* New York, NY: The Free Press, 1961.

Davies, J., "Identity and Commitment." Tinbergen Institute Discussion Paper 055/2, University of Amsterdam, 2004.

Ellwood, David, *Poor Support: Poverty in the American Family.* New York, NY: Basic Books, 1988.

Emerson, R., "Power-dependence relations." *American Sociological Review,* 1962, 27, 31-41.

Festinger, Leon, "A Theory of Social Comparison Processes." *Human Relations,* 1954, 7, 117-140.

Festinger, Leon, *A Theory of Cognitive Dissonance.* Stanford, CA: Stanford University Press, 1957.

Fishbein, M., and I. Ajzen, *Belief, Attitude, Intention, and Behavior: An Introduction to Theory and Research.* Reading, MA: Addison-Wesley, 1975.

Fordham, Signithia, *Blacked Out: Dilemmas of Race, Identity, and Success at Capital High.* Chicago, IL: Chicago University Press, 1996.

Fordham, Signithia and John Ogbu "Black Students' School Success: Coping with the 'Burden of Acting White.'" *Urban Review,* 1986, 18, 176-206.

Freeman, Richard and Harry Holzer, *The Black Youth Employment Crisis.* Chicago, IL: Chicago University Press, 1986.

French, J. and B. Raven, "The Bases of Social Power" in D. Cartwright (Ed.), *Studies in Social Power.* Ann Arbor, MI: Institute for Social Research, 1959.

Fryer, Roland and Steven Levitt, "The Causes and Consequences of Distinctively Black Names." *Quarterly Journal of Economics,* 2004, 119(3), 767-806.

Garfinkel, Harold, "A conception of, and experiments with, 'trust' as a condition of stable concerted actions," in O. J. Harvey (Ed.), *Motivation and Social Interaction.* New York, NY: Ronald Press, 1963.

Giddens, Anthony, *The Constitution of Society.* Oxford, UK: Polity Press, 1984.

Glaeser, E., "The Political Economy of Hatred." *Quarterly Journal of Economics,* 2005, 120(1), 45-86.

Glaeser, E., and J. Scheinkman, "Non-Market Interactions." NBER Working Paper No. 8053, 2000.

Glaeser, E., B. Sacerdote, and J. Scheinkman, "The Social Multiplier," forthcoming in *The New Palgrave*, 2nd edition, 2002.

Glaeser, E. and A. Shleifer, "Legal Origins." *Quarterly Journal of Economics*, 2002, 117,1193-1229.

Goffman, Erving, *The Presentation of Self in Everyday Life.* New York, NY: Anchor Books, 1959.

Goffman, Erving, *Interaction Ritual.* New York, NY: Pantheon Books, 1967.

Goldin, Claudia, "The Quiet Revolution That Transformed Women's Employment, Education, and Family." *American Economic Review Papers and Proceedings,* 2006, 96(2), 1-21.

Gouldner, Alvin W., *Patterns of Industrial Bureaucracy.* New York, NY: The Free Press, 1954.

Greif, Avner, "Contract Enforceability and Economic Institutions in Early Trade: The Maghribi Traders' Coalition." *American Economic Review,* 1993, 83, 525-548.

Harris, Judith, *No Two Alike: Human Nature and Human Individuality.* New York, NY: W.W. Norton, 2006.

Haslam, S. Alexander, *Psychology in Organizations: The Social Identity Approach.* London, UK: Sage Publications, 2001.

Haslam, S. A. and J. C. Turner, "Context-dependent variation in social stereotyping 2: The relationship between frame of reference, self-categorization and accentuation." *European Journal of Social Psychology,* 1992, 22, 251-277.

Hauser, Marc D., *Moral Minds: How Nature Designed Our Universal Sense of Right and Wrong.* New York, NY: HarperCollins, 2006.

Helpman, Elhanan, *The Mystery of Economic Growth.* Cambridge, MA: Belknap Press, 2004.

Hill, C., "What The New Economics Of Identity Has To Say." University of Minnesota Legal Studies Research Paper No. 05-46, 2006.

Hogg, M. A. and J. C. Turner, "Intergroup behavior, self-stereotyping and the salience of social categories." *British Journal of Social Psychology,* 1987, 26, 325-340.

Holzer, Harry and Paul Offner, "Trends in Employment Outcomes of Young Black Men, 1979-2000." Mimeo, 2002.

Homans, George C., *The Human Group.* New York, NY: Harcourt, Brace, & World, Inc., 1950.

Homans, George C., and David Schneider, *Marriage, Authority, and Final Causes: A Study of Unilateral Cross-Cousin Marriage.* Glencoe, IL: The Free Press, 1955.

Horst, Ulrich, Alan Kirman, and Miriam Teschl, "Changing Identity: The Emergence of Social Groups." Mimeo, 2007.

Hraba, J. and G. Grant, "Black is beautiful: a re-examination of racial preference and identification." *Journal of Personality and Social Psychology,* 1970, 16, 368-402.

Jackson, Matthew, "The Economics of Social Networks," in R. Blundell, W. Newey, and T. Person (Eds.), *Advances in Economics and Econometrics, Theory and Applications: Ninth World Congress of the Econometric Society (Vol. 1).* Cambridge, UK: Cambridge University Press, 2006.

Jankowski, Martin S., *Islands in the Street: Gangs and American Urban Society.* Berkeley, CA: University of California Press, 1991.

Jones, Stephen R. G., *The Economics of Conformism.* New York, NY: Basil Blackwell Inc., 1984.

Kandori, Michihiro, "Social Norms and Community Enforcement." *Review of Economic Studies,* 1992, 59, 63-80.

Kerner, Otto (Chairman), *Report of the National Advisory Commission on Civil Disorders.* New York, NY: Bantam Books, 1968.

Kling, Jeffrey, Jens Ludwig, and Lawrence Katz, "Neighborhood Effects on Crime for Female and Male Youth: Evidence from a Randomized Housing Mobility Experiment." *Quarterly Journal of Economics,* 2005, 120, 87-130.

Komai, Mana, Mark Stegeman, and Benjamin Hermalin, "Information and Leadership." *American Economic Review,* 97(3), 944-947.

Konow, J., "Fair Shares: Accountability and Cognitive Dissonance in Allocation Decisions." *American Economic Review*, 2000, 90(4), 1072-1091.

Koszegi, Botond, "Ego Utility and Information Aquisition." Mimeo, 2001.

Koszegi, Botond, "Utility from Anticipation and Personal Equilibrium." Mimeo, 2007.

Koszegi, Botond and Matthew Rabin, "A Model of Reference-Dependent Preferences." *Quarterly Journal of Economics,* 121(4), 1133-1165.

Kreps, D.M., P. Milgrom, J. Roberts, and R. Wilson, "Rational Cooperation in the Finitely Repeated Prisoner's Dilemma." *Journal of Economic Theory,* 1982, 27, 245-252.

Kuran, Timur, and William Sandholm, "Cultural Integration and Its Discontents." *Review of Economic Studies*, 2008, 75, 201-228.

Mills, Judson, "Changes in Moral Attitudes Following Temptation." *Journal of Personality,* 1958, 26, 517-531.

Montefiore, Simon Sebag, *Stalin: The Court of the Red Tsar.* New York, NY: Vintage Books, 2003.

Montgomery, James, "Revisiting *Tally's Corner:* Mainstream Norms, Cognitive Dissonance, and Social Growth." *Rationality and Society,* 1994, 6(4), 462-488.

Mullainathan, Sendhil and Andrei Shleifer, "The Market for News." *American Economic Review*, 2005, 95(4), 1031-1053.

Murphy, Kevin M. and Andrei Shleifer, "Persuasion in Politics." *American Economic Review Papers and Proceedings*, 2004, 94(2), 435-439.

Myerson, Roger, "Fundamental Theory of Institutions: A Lecture in Honor of Leo Hurwicz." Leo Hurwicz Lecture, 2006.

North, D., W. Summerhill, and B. Weingast, "Order, Disorder, and Economic Change: Latin America versus North America" in Bruce Bueno de Mesquita and Hilton Root (Eds.), *Governing for Prosperity.* New Haven, CT: Yale University Press.

Ono, K. and J. Davis, "Individual Judgment and Group Interaction: A Variable Perspective Approach." *Organizational Behavior and Human Decision Processes,* 1988, 41, 211-232.

Ostrom, Elinor, *Governing the Commons: The Evolution of Institutions for Collective Action.* Cambridge, UK: Cambridge University Press, 1990.

Oxoby, Robert, "Attitudes and Allocations: Status, Cognitive Dissonance, and the Manipulation of Attitudes." *Journal of Economic Behavior and Organization*, 2003, 52, 365-385.

Oxoby, Robert, "Status Seeking, Cognitive Dissonance, and Growth of the Underclass." *The Economic Journal,* 2004, 114, 727-749.

Rabin, Matthew, "Incorporating Fairness into Game Theory and Economics." *American Economic Review*, 1993, 83(5), 1281-1302.

Rabin, Matthew, "Cognitive Dissonance and Social Change." *Journal of Economic Behavior and Organization*, 1994, 23, 177-194.

Saxe, Rebecca, "Uniquely Human Social Cognition." *Current Opinion in Neurobiology*, 2006, 16, 235-239.

Shafir, Eldar, Peter Diamond, and Amos Tversky, "Money Illusion." *Quarterly Journal of Economics*, 1997, 112(2), 341-374.

Sherif, Muzafer, *The Psychology of Social Norms.* New York, NY: HarperCollins, 1936.

Sherif, Muzafer, "Formation of Social Norms: The Experimental Paradigm" in H. Proshansky and B. Seidenberg (Eds.), *Basic Studies in Social Psychology.* New York, NY: Holt, Rinehart, and Winston, 1965.

Sidanius, Jim and Felicia Pratto, *Social Dominance.* Cambridge, UK: Cambridge University Press, 1999.

Svenson, Ola, "Are We All Less Risky and More Skillful Than Our Fellow Drivers?" *Acta Psychologica,* 1981, 47(2), 143-148.

Tajfel, H., Flament, C., Billig, M. G.,and Bundy, R. F., "Social Categorization and Intergroup Behavior." *European Journal of Social Psychology,* 1971, 1, 149-177.

Weinberg, Bruce, "Social Interactions and Endogenous Association." Mimeo, 2005.

Wilson, William Julius, *The Truly Disadvantaged: The Inner City, the Underclass, and Public Policy.* Chicago, IL: Chicago University Press, 1987.

Wilson, William Julius, *When Work Disappears: The World of the New Urban Poor.* New York, NY: Vintage Books, 1996.

X, Malcolm and Alex Haley, *The Autobiography of Malcolm X.* New York, NY: Ballantine Books, 1964.

9 Appendix A

Section 5 suggests a reason why a bifurcation seems to have taken place among African Americans in the late 1960s, some moving into the middle class following Civil Rights and others becoming oppositional. This appendix presents an example that illustrates the argument made in Section 5.

In Section 5, we argued that the Civil Rights Movement led most African Americans to feel that they *should* have greater pride in themselves. That is, most African Americans seem to have adopted an ideal that they should view themselves as high status. Section 5 also argued that, following Civil Rights, some African Americans faced greater obstacles than others to entering the middle class. Those African Americans who faced greater obstacles were especially likely to become oppositional.

The example presented in this appendix will capture this idea in the following way. We will assume that a population is composed of one African American and $n-1$ white people. We will consider four different conditions:

	low pride (before Civil Rights)	high pride (after Civil Rights)
fewer obstacles	Condition A: not oppositional	Condition B: not oppositional
more obstacles	Condition C: not oppositional	Condition D: oppositional

In conditions A and C, we will assume that the African American in the population has low pride, representing a pre-Civil Rights state. In conditions B and D, we will assume that the African American has greater pride, representing the effect of Civil Rights. In conditions A and B, we will assume that there are relatively few obstacles faced by the African American to meeting an ideal held by whites. In contrast, in conditions C and D, we will assume that there are significant obstacles faced by the African American. In the example, we will find that the African American only becomes oppositional following Civil Rights when she faces significant obstacles to meeting the ideal held by whites: only in condition D (the post-Civil Rights, significant obstacle condition) will the African American adopt a different ideal from whites. This corresponds to the story told in Section 5.

Representing obstacles in the example

In the example, we will assume that individual i has innate utility $u_i(a) = -\frac{|a|^3}{\alpha_i}$. We will assume that whites in the population all have the same ability: $\alpha_i = \alpha_w$ for i white. But, the African American in the population potentially has a different ability parameter: $\alpha_i = \alpha_b$ for i African American. In the "fewer obstacles" conditions (A and B), we will assume that $\alpha_b = \alpha_w$. In the "more obstacles" conditions (C and D), we will assume that $\alpha_b < \alpha_w$.[92]

Suppose the whites in the population choose to hold some ideal I. In the "more obstacles" conditions, the African American finds it more difficult to meet ideal I than the whites in the population. In

[92] In the example developed in this section, we do *not* assume that the Civil Rights Movement reduces the obstacles faced by African Americans: we assume $\alpha_b^A = \alpha_b^B$ and $\alpha_b^C = \alpha_b^D$ (α_b is equal in conditions A and condition B and α_b is equal in conditions C and D). We *do* assume that some African Americans face more obstacles than others: $\alpha_b^A > \alpha_b^C$ and $\alpha_b^B > \alpha_b^D$. This keeps the example simple, but it is possible to build a more realistic example. We could capture the idea that Civil Rights removes obstacles by assuming $\alpha_b^B > \alpha_b^A$ and $\alpha_b^D > \alpha_b^C$ in addition to assuming that $\alpha_b^A > \alpha_b^C$ and $\alpha_b^B > \alpha_b^D$. It is possible to construct an example in which $\alpha_b^A > \alpha_b^C$ and $\alpha_b^B > \alpha_b^D$ (Civil Rights removes obstacles for all African Americans), $\alpha_b^A > \alpha_b^C$ and $\alpha_b^B > \alpha_b^D$ (obstacles are greater for some African Americans than others both before and after Civil Rights), and only in condition D do African Americans become oppositional (African Americans only become oppositional following Civil Rights and then only when they faced more obstacles than other African Americans).

the "fewer obstacles" condition, the African American and whites find it equally difficult to meet ideal I.

It is important to emphasize what $\alpha_b < \alpha_w$ is meant to reflect. When we assume $\alpha_b < \alpha_w$ in conditions C and D, this is meant to reflect, in a stylized way, difficulties faced by African Americans in meeting ideals held by whites that arise from *circumstances* faced by African Americans (such as deindustrialization of cities). Despite the term "innate utility," in assuming $\alpha_b < \alpha_w$, we do *not* mean to suggest innate differences between African Americans and whites.

Representing pride in the example

In the post-Civil Rights conditions, we want to think of the African American in the population as holding an ideal that she should think of herself as high status.

The model developed in Section 2 is not able to capture the notion that an individual might hold an ideal that she should have pride in herself. The reason is that such an ideal is not an ideal about how one should behave but an ideal about what one should *believe*. The model developed in Section 2 only allows an individual to hold an ideal about how she should behave. The model developed in is capable of capturing an individual holding such an ideal.

However, if we make a slight amendment to the Section 2 model, it can be used to capture in a stylized way the adoption of such an ideal by African Americans. Observe that, if an individual adopts an ideal that she should have pride in herself, this is effectively equivalent to a heightened concern about her social status.

Therefore, we can capture the effect of Civil Rights in our example by assuming: (1) individual i's status utility is given by $\Phi_i(s_i(i)) = \beta_i s_i(i)$, (2) β_i is high for i white, and (3) β_i is low before Civil Rights and high after Civil Rights for i African American. This is an *amendment* to the Section 2 model because, in Section 2, it is assumed that everyone has the same status utility function: status utility is given by $\Phi(s_i(i))$ as opposed to $\Phi_i(s_i(i))$.

It should be noted that the model developed in allows us to capture heterogeneity in pride across individuals *without* assuming heterogeneity in status utility ($\Phi(s_i(i))$). The model enables heterogeneity in pride to be captured by heterogeneity in the ideals held by individuals.

Basic Assumptions

We will assume that individual i's utility function is:

$$U_i = u_i(a_i) - d(a_i, I_i) + \Phi_i(s_i(i)) + C^S(I_i, I_{-i}, G_i)$$

where

$$s_i(i) = \frac{1}{|P|} \sum_{j \in P} (d(a_j, I_i) - d(a_i, I_i))$$

We will also assume that:
(1) $P = \{1, 2, ..., n\}$
(2) $A = \mathbb{R}$ (the set from which individuals choose a_i and I_i)
(3) $u_i(a) = -\frac{|a|^3}{\alpha_i}$
(4) $d(a_i, I_i) = 2(a_i - I_i)^2$
(5) $C^S(I_i, I_{-i}, G_i) = C\left(|\{j \in G_i : I_j = I_i\}| - |\{j \in G_i : I_j \neq I_i\}|\right)$
(6) $\Phi_i(s_i(i)) = \beta_i s_i(i)$

We will think of individual 1 as being African American. We will assume that $\alpha_1 = \alpha_b$ and $\beta_1 = \beta_b$. We will think of individuals $i > 1$ as white and will assume that $\alpha_i = \alpha_w$ and $\beta_i = \beta_w$ for $i > 1$.

In the pre-Civil Rights conditions (A and C), we will assume that $\beta_b = \beta_b^{beforeCR} < \beta_w$. In the post-Civil Rights conditions (B and D), we will assume that $\beta_b = \beta_b^{afterCR} = \beta_w$.

In the "fewer obstacles" conditions (A and B), we will assume that $\alpha_b = \alpha_b^{fewer} = \alpha_w$. In the "more obstacles" conditions (C and D), we will assume that $\alpha_b = \alpha_b^{more} < \alpha_w$.

Parametric Assumptions

We will make the following assumptions about the values of the parameters:

$\beta_b^{beforeCR} = 0.5$	$\alpha_w = \alpha_b^{fewer} = 0.2$	$C = 14$
$\beta_w = \beta_b^{afterCR} = 1$	$\alpha_b^{more} = 0.05$	$n = 10$

Results

As is generally the case, multiple social equilibria exist in conditions A, B, C, and D. We will restrict attention to equilibria in which whites in the population hold a particular ideal: $I_i = 8$ for $i > 1$. In each condition (A, B, C, and D), a unique social equilibrium exists in which the whites hold ideal $I_i = 8$. In restricting attention to equilibria where $I_i = 8$ for whites, we are examining a case where Civil Rights does not effect white ideals but potentially has an effect on African American ideals.

The following is a characterization of the equilibria that exist in conditions A, B, C, and D for which $I_i = 8$ for $i > 1$:

Fewer Obstacles, Before Civil Rights (A)
$I_i = 8$ for all i
$a_1 = 1.58$
$a_i = 1.78$
$s_1(1) = s_i(1) = -4.55$ for $i \neq 1$
$s_i(i) = s_1(i) = 0.51$ for $i \neq 1$
$G_i = P - \{i\}$ for all i

Fewer Obstacles, After Civil Rights (B)
$I_i = 8$ for all i
$a_1 = 1.78$
$a_i = 1.78$
$s_1(1) = s_i(1) = 0$ for $i \neq 1$
$s_i(i) = s_1(i) = 0$ for $i \neq 1$
$G_i = P - \{i\}$ for all i

More Obstacles, Before Civil Rights (C)
$I_i = 8$ for all i
$a_1 = 0.83$
$a_i = 1.78$
$s_1(1) = s_i(1) = -22.74$ for $i \neq 1$
$s_i(i) = s_1(i) = 2.53$ for $i \neq 1$
$G_i = P - \{i\}$ for all i

More Obstacles, After Civil Rights (D)
$I_1 = -2.57$ and $I_i = 8$ for $i \neq 1$
$a_1 = -0.51$
$a_i = 1.78$
$s_1(1) = 26.35$ and $s_i(1) = -60.64$ for $i \neq 1$
$s_i(i) = 6.74$ and $s_1(i) = -2.93$ for $i \neq 1$
$G_1 = \varnothing$ and $G_i = P - \{1, i\}$ for $i \neq 1$

Observe that it is only in condition D (More Obstacles, After Civil Rights) that the African American in the population becomes oppositional, adopting a different ideal from whites ($I_1 = -2.57$ rather than 8). In Condition D, the African American looks down on whites and vice-versa ($s_1(i) = -2.93$ and $s_i(1) = -60.64$ for $i \neq 1$).

African Americans bifurcate in the example in the following sense: the behavior of a "fewer obstacles" African American and the behavior of a "more obstacles" African American differ more after Civil Rights than before. Before Civil Rights, $a_1^{fewer} = 1.58$, $a_1^{more} = 0.83$, and $|a_1^{fewer} - a_1^{more}| = 0.74$. After Civil Rights, $a_1^{fewer} = 1.78$, $a_1^{more} = -0.51$, and $|a_1^{fewer} - a_1^{more}| = 2.29 > 0.74$.

To conclude, let us recall the intuition behind these results. Civil Rights increases the desire of African Americans to view themselves as high status. An African American can improve her view of her

status either by maintaining the white ideal but meeting that ideal better (we see this in the movement of a_1^{fewer} from 1.58 to 1.78 following Civil Rights) or by adopting an oppositional ideal (we see this in the movement of I_1^{more} from 8 to -2.57 following Civil Rights). The latter option is more attractive when an African American has difficulty meeting the white ideal. Hence, an African American who finds it more difficult to meet the white ideal is more inclined to rebel against the white ideal than an African American who finds the white ideal easier to meet.

10 Appendix B

This appendix presents a version of the model that formalizes the ideas presented in Section 5. We begin by giving the individual's utility function and a definition of a social equilibrium. This will be followed by an explanation of the notation, a statement of the functional form assumptions, and a discussion of the terms of the utility function.

The Individual's Problem:

It is assumed that individual i's utility function takes the form

$$U_i(a_i, G_i, I_i, \mu_i) = \mu_i(u_i(\omega)|X_i) - \mu_i(I_i(\omega, \upsilon, i)|X_i) + \mu_i(\Phi(s_i(i))|X_i) + C^I(\mu_i, X_i) + C^S(\mu_i, X_i, G_i)$$

where

$$\begin{array}{ll} (1) & s_i(j) = -I_i(\omega, \upsilon, j) + \frac{1}{|P|}\sum_{k\in P} I_i(\omega, \upsilon, k) \\ (2) & C^I(\mu_i, X_i) = \Gamma(\Psi(\mu_i, X_i)) \\ & \text{with } \Psi(\mu_i, X_i) = \frac{\mu_i(X_i)}{\max_{\mu'\in M}\mu'(X_i)} \\ (3) & C^S(\mu_i, X_i, G_i) = \mu_i(\lambda_{|G_i|}(\{s_j(i)\}_{j\in G_i})|X_i) \\ (4) & X_i = X(\omega, i) \end{array}$$

Normally, we would assume that individual i chooses (a_i, G_i, I_i, μ_i) to maximize U_i. However, we might alternatively wish to think of there being a "personal equilibrium" (see Koszegi and Rabin (2006) and Koszegi (2007) for a discussion).

The Social Equilibria:

We could define a social equilibrium as a set $\{(a_i, G_i, I_i, \mu_i)\}_{i\in P}$ such that :

$$(a_i, G_i, I_i, \mu_i) \in \underset{(a_i', G_i', I_i', \mu_i')}{\text{argmax}} U_i(a_i', G_i', I_i', \mu_i') \text{ for all } i.$$

Alternatively, we might define a social equilibrium as a set $\{(a_i, G_i, I_i, \mu_i)\}_{i\in P}$ such that :

$$(a_i, G_i, I_i, \mu_i) \text{ is a personal equilibrium for all } i.$$

Notation:

a_i is the behavior of individual i.

G_i is the set of people with whom individual i chooses to interact.

$\omega \in \Omega$ is a state of the physical world. The behavior of individuals in the world constitutes part of ω. So, a_i and G_i can be determined from ω. Individual i's choice of how to behave has an effect on ω.

$\upsilon \in N$ is a state of the beliefs of individuals, including individual i. I_i, and μ_i can be determined from υ.

$(\omega, \upsilon) \in \Omega \times N$ is a state of the world.

μ_i is a probability measure over $(\Omega \times N, \mathcal{F})$, where $\mathcal{F}$ is a sigma algebra over $\Omega \times N$. μ_i is the "system of beliefs" of individual i. M (which appears in the Ψ function) denotes a set of probability measures over $(\Omega \times N, \mathcal{F})$. We assume that individual i chooses μ_i from the set M.

$X_i \in \mathcal{F}$ is the information that individual i possesses about the state of the world. We assume that the information i possesses about the state of the world is a function of the state of the world: $X_i = X(\omega, i)$. Since individual i's behavior is an aspect of ω, individual i can potentially choose to acquire or choose to avoid acquiring information. Individual i may also be able to affect the information possessed by individual j (X_j).

P denotes the set of all individuals in the world.

$I_i(\cdot)$ is individual i's belief about how individuals should and should not behave. $I_i(\omega, \upsilon, j)$ answers the question: how well does individual j live up to the ideals of individual i? How well individual j lives up individual i's ideals depends mainly upon the behavior of individual j (which is part of ω) and the beliefs of individual j (which is part of υ). But, the appropriate way to behave (and the appropriate thing to believe) may depend upon other aspects of the state of the world, and therefore other aspects of ω and υ.

$s_i(j)$ is the belief of i about the status of individual j.

Functional form assumptions:

$\lambda_{|G_i|}$ is a real-valued function, increasing in all of its arguments.

Γ is an increasing continuous function.

I_i is assumed to be a nonnegative real-valued function.

Φ is assumed to be an increasing real-valued function.

Terms of the utility function:

(1) $\mu_i(u_i(\omega)|X_i)$

The first term is expected "innate utility."

(2) $-\mu_i(I_i(\omega, \upsilon, i)|X_i)$

This term is the cost to individual i of deviating from her ideal.

(3) $f(s_i(i))$

The third term is status utility.

(4) $C^I(\mu_i, X_i)$

The fourth term is information-based confirmation of belief. It depends upon how close individual i's beliefs about the state of the world, μ_i, are to the maximum likelihood beliefs.

(5) $C^S(\mu_i, X_i, G_i)$

The fifth term is social confirmation. It depends upon individual i's beliefs about how she is viewed by the individuals with whom she interacts.

11 Appendices C, D, E, and F

Appendix C characterizes the social equilibria in a case where there are n individuals with the same innate utility function (n-tuplets). Appendix D gives some results and analysis omitted from Section 4. Appendix E suggests that the model developed in Section 2 can be viewed as making three additions to a classical utility function. This appendix discusses the impact of removing one or more of these additions. Appendix E proves the results given in Section 4 and Appendix D. These appendices are available upon request. Please email: akerlof@fas.harvard.edu.