Conformity, Consistency, and Cultural Heterogeneity

Jenna Bednar, Aaron Bramson, Andrea Jones-Rooy, and Scott Page
University of Michigan*

September 18, 2006

Abstract

We analyze two disparate forces on behavior: *consistency* and *conformity*. Individuals want to be consistent across situations and also want to conform with the behaviors of those around them. We construct a model that shows that these two forces contribute to the formation of distinct cultures. Although each force alone creates a simple dynamical system, the combined forces produce tension that slows time to convergence. This may provide an explanation for the simultaneous existence of both intra- and inter-cultural heterogeneity. We also find that the tension between the two forces is more pronounced in societies where one force dominates the other. When we expand the model to include errors (random attribute assignment), we reveal the non-linear additivity of the system. When we consider conformity and consistency forces together, the whole differs from the parts. This final result emphasizes the pitfalls of studying individual forces in isolation and extrapolating to their combined effects. Our findings may be applied to understand heterogeneity within other organized groups, including firms and political parties.

1 Introduction

The empirical research on cultural differences reveals three main findings. First, intercultural differences exist. People who belong to distinct cultures act differently: they possess different belief systems and exhibit distinct behaviors and mannerisms. Second, cultures have signature characteristics that are far from idiosyncratic collections of attributes. Individuals within a culture exhibit consistency among their behaviors that allow others to anticipate and predict responses based on cultural affiliations.

*This research was funded by AFOSR-MURI Grant 57100001867. We thank Len Sander and Casey Schneider-Mizell for help with the analytic sections of this paper and Jennifer Miller, Paul Hartzog, Maggie Penn and John Patty for helpful comments on earlier versions of this paper. Contact jbednar@umich.edu or spage@umich.edu.

For example, members of some cultures, such as French people, tend to be more risk averse than are members of other cultures (Hofstede 1991). People can also differ culturally in how much emphasis they place on the individual relative to the collective (Inglehart 1997). Third, this consistency notwithstanding, people within cultures differ. Not all French people are the same; nor are all members of the Itza' or the !Kung. Despite ample empirical evidence that it exists (as discussed below), this third finding of intra-cultural heterogeneity—differences in behavior within a cultural group—has not been the focus of much theoretical work.

In this paper we describe a model that produces all three of these empirical regularities: inter-cultural heterogeneity, intra-cultural consistency, and intra-cultural heterogeneity. To do this, the model relies on the interplay between two forces that drive a person's behavior: a desire for *social conformity* and a desire for *individual consistency*. These two forces, which operate on beliefs, behaviors, and other attributes, can also be considered to be incentives. Our interest here, though, is less in the microfoundations of these forces, but rather in how these forces aggregate (Schelling 1978), particularly in the extent to which these two forces create opposing pressures on individual beliefs, behaviors, and attributes, as well as the implications of these pressures for how we understand cultural, organizational, or group behavior.

We first consider each force separately in two models (a *pure conformity model* and a *pure consistency model*) and then together in a *combined conformity/consistency model*. For the purposes of this paper, we consider the individual interacting within a given community as belonging to a culture. Our models produce several results. First, we show that when individuals are driven only by the desire for consistency, the model produces consistent individuals but no intra-culture homogeneity; i.e. a society consisting of individuals who exhibit behavior consistent only with themselves. Next, we show that if people ignore the desire for consistency and choose instead only to conform to those around them, then all individuals converge to the same set of inconsistent attributes. Cultural differences arise from randomness (the odds of coordinating on the same attributes are low), from different initial conditions (the attributes that are most prevalent initially tend to become the dominant cultural attributes), or from preferential interaction (Axelrod 1997, Rogers 1983, Hannan 1979, Barth 1969, Simmel 1955, Homans 1950). In brief, when we include only intra-cultural conformity within disjoint communities we can explain inter-cultural heterogeneity as distinct equilibria of the coordination dynamic. Such a model, however, cannot produce intra-cultural heterogeneity. Finally, when we consider both forces in combination, we can produce all three of the afore-stated empirical regularities at issue.

Our model contributes to the literature on cultural differences in several ways. Our main finding connects two well established individual level behavioral assumptions from psychology—the individual desires to exhibit consistent and conforming behavior—with aggregate level empirical regularities long noticed by sociologists and political scientists: inter-cultural heterogeneity, intra-cultural consistency, and intra-cultural heterogeneity. Second, contrary to more intuitive expectations, an imbalance between the desire to be conforming or consistent does *not* speed convergence; rather, it slows it. Third, *when we include errors* in our model, the model with both forces creates far more heterogeneity than either of the single force models. In other words, there is no linear additivity: the combined effect of the forces differs from the sum of the effects of the individual forces. Thus, in addition to the substantive contributions we highlight here, the paper makes a methodological contribution by emphasizing the danger of carving out individual effects and studying them in isolation.

In the remainder of this section we provide an overview of the literature on why and how individuals face pressure to conform and to be consistent. We also present empirical evidence of both inter- and intra-cultural heterogeneity and discuss the importance of a better understanding of the latter.

The rest of the paper is organized as follows. In Part 2 we introduce our models of conformity and consistency. In Part 3 we analyze the results by considering (1) the time to convergence, and (2) the equilibrium distribution in models with errors. In Part 4 we use numerical experiments to test our analytic results. In Part 5 we summarize our findings and consider alternative explanations for the empirical observation of intra-cultural heterogeneity. We close the paper with a discussion of potential applications of the theory to understand heterogeneity in organizations, including firms and political parties.

The Two Forces

Conformity

Most models of cultural formation emphasize conformity. The force for conformity—the idea that people become like those around them—can be unbundled into four distinct individual-level desires and incentives (i) the need to fit in with others (ii) the strategic benefit from coordination (iii) the incentive to free ride on the information of others, and (iv) the desire to interact with people similar to oneself. The first force has long been a staple of social psychology: people often mimic the behaviors, beliefs, and attributes of those with whom they interact. Social pressure can impart desire to fit in with others (Bernheim 1994, Kuran 1995). If others positively reinforce conforming behavior, then conformity can become a conditioned response (Pavlov 1903, Skinner 1974). In brief, people who interact frequently become similar. They act similarly, they dress similarly, they reveal similar preferences (Axelrod 1997), and when confronted with a new situation, they copy the behaviors of others rather than charting their own course (Simon 1982).[1]

Conforming behavior need not be divorced from incentives. When copying, individuals often do so selectively. They look to the behaviors of their more successful neighbors (Kennedy 1988). People who face similar problems may construct similar

[1] Banduras' (1977) bobo doll experiments demonstrated that children imitate behavior they view on TV. More recently, Huesmann (1988, 1998) has shown that in the short term children copy behavior that they observe, which in the long run becomes encoded into their behavioral schemas.

solutions without imitating just as students who enroll in the same class and take identical exams may produce similar answers without copying. Seminal works in psychology by Pavlov (1903) and Skinner (1974) connect positive reinforcement and the conditioning of learned responses.

However, imitation only partly explains within-culture conformity. Institutions exert major influence, as they can create a common set of incentives and constraints on behavior (North 1990, Bednar and Page 2006). In particular, institutions create incentives for coordination, the second type of force for conformity (Young 1998). If everyone else shakes hands upon greeting, drives on the left side of the road, and speaks English, an individual will benefit from doing the same. Coordination problems are often modeled as two-by-two games in which the players attempt to choose the same action. We distinguish between conformity and coordination in terms of the measurability of the benefit. People coordinate on the side of the road they drive their cars. This decision has measurable costs and benefits. Alternatively, people can conform by wearing a certain type of pants. In this case, the benefits that come from wearing a particular style of pants, while more difficult to measure economically, are nevertheless real.

The third type of force for conformity arises in situations where people take actions contingent on their beliefs. If people see that everyone else has taken some action, they cannot help but draw certain inferences about the beliefs of others. This tendency creates what has been called *herd behavior* (Banerjee 1992) and *information cascades* (Bikchandani, et al 1993), or what is more colloquially referred to as "jumping on the bandwagon".

The final force that can lead to conformity is sorting. It has long been shown, both empirically and casually, that people prefer being around others who behave, dress, and think similarly and can satisfy this preferences by isolating themselves in cliques of like minded people.[2] This creates conformity within interacting groups and implies differences between those groups. If these differences did not exist, the two groups may as well merge and form a larger group.[3] In our model, we take the interacting groups as fixed and rule out the possibility of subcultures of this sort, acknowledging that the possibility of subcultures would create further intra-cultural heterogeneity.

Consistency

Our second fundamental force that motivates individual behavior is the desire to be consistent: to act according to a common set of principles and guidelines in different situations. This force can be explained using either of two lines of argument: one cognitive and one based on cost-benefit analysis. Psychological research shows that personal uneasiness with cognitive dissonance creates within individuals a desire for consistency. People find acting differently in every situation difficult (Festinger 1957).

[2]See, for example, Schelling's (1971) discussion of preferences and racial segregation.

[3]Sorting creates no end of empirical problems related to disentangling sorting effects from conforming effects (Brock and Durlauf 2006).

Psychologists generally agree that individuals can overcome cognitive dissonance by either restricting their behavior to be consistent with their attitudes or by changing their attitudes to match their inconsistent behavior.[4] The cognitive argument rests on current understandings of the physiology of the brain. Research shows that repeated behaviors create cognitive pathways (Gazzaniga 1999). For this reason, when confronted with a new situation, individuals often choose a behavioral response that belongs to their existing repertoire, especially if that response has been reinforced in the past (March 1991, Cavalli-Sforza and Feldman 1981).

The cost-benefit analysis logic relies on informational advantages: an individual's consistency in behavior allows others to predict his/her next moves. Accurate predictions like this grease the wheels of economic and political institutions. In fact, one broadly-accepted role of culture is to help coordinate on equilibria. Some equilibria may be more focal than others based on their relationship to the wider culture (Calvert and Johnson 1997).[5]

To summarize, empirical evidence shows that individuals exhibit tendencies toward both consistency and conformity, each up to a point. The evidence also shows that circumstance plays a decisive role in when individuals adopt which strategy. People exhibit greater consistency in situations that are common, patterned, or part of a social role; in situations where individuals confront more radically new situations they become increasingly likely to turn to others for behavior clues (Tittle and Hill, 1967, as appears in Liska 1975). At the same time, the normative environment within which people make behavioral decisions also matters: in general, the more observable one's behavior is to others, the more likely one is to conform to the majority behavior and/or the standing social norm (Liska 1975, Ajzen and Fishbein 1969, DeFleur and Warner 1969, Bowers 1968).

Inter- and Intra-Cultural Heterogeneity

As mentioned, empirical studies and casual observations reveal substantial differences between cultures. Different patterns of behavior exist between the German and French cultures as well between the cultures of the Inuit and the !Kung. No one disputes these stylized facts. In fact, this inter-cultural variation provides a foundation for nearly all social scientific comparative studies. The very nature of area studies research implies the assumed existence of recognizable and important differences between behaviors of peoples in different geographical regions, be they informal societies, communities, cities, or countries. Inglehart, in summarizing The World Values Survey data, concludes that "cultural variation is . . . relatively constant within a given society, but shows relatively great variation between different societies" (Inglehart 1997, p. 166).

[4]The latter seems to have substantial empirical support. People tend to adopt attitudes to make their behavior seem consistent.

[5]We do not mean to imply that all strategic environments include incentives to coordinate in some way. Certainly, constant sum games do exist. But growth and progress hinge on positive interactions. To be successful, cultures must exploit those interactions.

In other words, Danish attitudes about well-being can be consistently distinguished from French, Italian, or Portuguese attitudes. He bases this conclusion on evidence gathered over many years.[6] Experimental data corroborates these survey findings. Henrich et al (2001) conducted an extensive comparative study of fifteen small-scale societies across five cultures. They also found substantive evidence of inter-cultural variation in behavior. In another study Henrich (2000) finds that the economic behavior of Peruvian communities varies widely from the behavior of a Los Angeles control group, which suggests that "economic reasoning may be heavily influenced by cultural differences—that is, by socially transmitted rules about how to behave in certain circumstances (economic or otherwise) that may vary from group to group as a consequence of different cultural evolutionary trajectories" (Henrich 2000, p. 973).

The existence of inter-cultural differences does not imply people within cultures are the same. . IAnalyses of data from cross-cultural studies reveal substantial intra-cultural heterogeneity to be substantial (Au 1999, Pelto and Pelto 1975, Thompson 1975, Graves 1970). In Au's (1999) measurement of Intra-Cultural Variation (ICV), he takes six variables from The World Values Survey and compares the standard deviations for each country on each variable. He finds that some countries that share similar cultural means exhibit substantial differences in ICV. He also uncovers some surprises: Contrary to popular lore, American culture is far more homogenous than Japanese culture. Hofstede (1991) offers possible reasons for this observed ICV as differences in colonial inheritance, language, ethnicity, and sub-regional customs. Pelto and Pelto (1975) refer to Harris (1970) when they summarize the ICV logic nicely: "the degree to which behavior is rule-bound varies a good deal from one situation to another" (Pelto and Pelto 1975, p. 10). We can think of behavioral regularities within a population as part of a culture. In a given context, in a given society, behavior can vary quite widely. Hence, when presented with a survey question about their behavior, people within a culture may respond differently.

To capture this heterogeneity within cultures, we build a model that relies on agents who possess vectors of *attributes*. We use the term attributes as a catchall: attributes might include behavior, dispositions for behavior, customs, attire and so on. However, they should not be confused with immutable characteristics; by assumption the attributes we consider are plastic. Because we are interested in internal consistency, our model extends Axelrod' s (1997) model of cultural formation. In his model, values have no explicit meaning. In our model the attributes' values are comparable. A value of one on the first attribute means something similar to a value of one on the second attribute. We capture consistency by attaching meaning to these values.

We analyze this model in two ways: First, we consider the time it takes for a population to converge to an equilibrium. We do this for three models: a *pure conformity model*, a *pure consistency model*, and a *combined conformity /consistency*

[6]On most variables he finds significant variation between country means. On cross cultural differences in life satisfaction over 64 countries, for example, the United States life satisfaction mean is 7.7, based on a ten-point scale; across all 64 societies the means range from as low as 3.7 to as high as 8.2.

model. In the combined model, we vary the weight of the two forces. A priori, we have no reason to believe that conformity and consistency matter equally. If anything, our reading of the literature suggests that the weight on conformity may be larger, at least in the cultural context. Applications of our model to other contexts may require adjusting the relative weights on the two forces.

We calculate the time to convergence mathematically for a simple model and also perform computational experiments. Both approaches show that the time to convergence to equilibrium increases dramatically in the combined model. This increase in time to convergence can result in sustained intra-cultural heterogeneity even if an equilibrium exists. One might argue that if an equilibrium exists, then the system would eventually reach it. This argument rests on false optimism. When the number of interactions required to attain an equilibrium is sufficiently large we would expect that in real cultures other factors would intervene before the equilibrium could be attained. Furthermore, missteps along the path can result in substantial and perpetual deviations from the equilibrium. To test for the effect of such mistakes, we also run the models with some small probability that agents randomly change an attribute's value.

Under this assumption, the population of agents does not converge to full conformity and consistency, but instead to a distribution over attribute values. We compare these limiting distributions in the single force models and the double force model and find that errors in the pure conformity and the pure consistency model create limiting distributions that lie close to the error-free equilibria. However, when we include both forces, the equilibrium distribution becomes much more disperse (see Table 2). The mathematical analysis reveals that the interplay between the two forces creates an effect at least as large as each of the two forces on its own. In effect, one plus one equals three.

The existence of intra-cultural variation and our proposed explanation leave open the question of whether it plays any significant role. Within-culture variation has empirical relevance for prediction. Durham (1991) demonstrates variety in types of marriage custom within Tibetan culture. Thompson (1975) provides evidence of significant intra-cultural variation in willingness to accept delayed economic gratification between three communities in Uganda. A study of a series of six cultures across four continents by Mintun and Lambert (1964) and Whiting (1963) found that all but one variable on child rearing behavior was better captured by intra- rather than inter-cultural variation. Pelto and Pelto (1975) cite a study of Minnesota to illustrate their argument that "even in supposedly homogenous communities there is a wide range of variation in most aspects of belief and behavior" (Pelto and Pelto 1975, p. 6). Moreover, even ritual and ceremonial practices, which are usually treated as encapsulating the most salient elements of inter-cultural variation, exhibit variation between members of a single culture. Adler and Graham (1989) demonstrate that businessmen negotiate more differently with people from within their own culture than with people from other cultures.

Though we focus in this paper on ethnic and national cultures, we might also apply our model to corporate and organizational cultures. Within corporations and organizations, people face incentives to conform as well as to be consistent, though for reasons that differ slightly from those we described above. Finally, our model could be apply to the creation of party ideology. Members of a political party also desire conformity and consistency, and these two desires may result in the analogous effects: differences within and between parties as well as party ideology. Relatedly, we might apply our model to the question of cultural integration (Kuran and Sandholm 2003). When people from two cultures interact in a common society, they face these two pressures, a desire to be consistent with their culture and a desire to conform to the wider culture.

Normatively, the existence of intra-cultural, or intra-organizational, heterogeneity may be a plus. It may promote innovation in the form of cultural evolution. The tension between conformity and consistency maps to related tensions between "exploiters versus explorers", "conformers versus nonconformers", and "scroungers versus producers" and may balance stability and variation (Kameda and Nakanishi 2002, Boyd and Richerson 2001, Rogers 1995, Nisbett and Ross 1980, Tindall 1976, March 1991, Weick 1969, Campbell 1965, Roberts and Zuni 1964). As individuals have incentives to conform with the behavior of the most successful actors in a system, successful strategies persist. Likewise, because individuals also have incentives to be consistent, deviance also persists; thus allowing for the discovery of new and better strategies that will in turn be adopted by others in the system. Furthermore, systems with both conforming and consistent individuals can both transmit and produce learned knowledge. Cognitive diversity may result in productive and robust societies (Wallace 1991, Page 2006).

2 Models of Conformity and Consistency

We now describe our general framework for modeling conformity and consistency. There are N agents, each of which has a vector of M attributes. Each attribute takes one of A values. Given this setup, we can characterize an agent as a vector $(a_1, a_2, \ldots a_M)$, where each $a_i \in \{0, 1, \ldots A\}$. We assume that all agents interact with each other with equal likelihood, what is called *random mixing*. Including network-structured interactions would complicate the analysis without providing any obvious benefit; an investigation into network effects is therefore left for future consideration.

Modeling Agent Behavior

We assume that the agents follow *behavioral rules*. These behavioral rules can be seen either as descriptions of what people do or as learning algorithms in a model in which agents attempt to maximize their payoffs. [7]

[7] A learning model begins from a payoff function and assumes that agents' responses depend on payoffs and expectations of payoffs. Ultimately, a learning model becomes a behavioral rule.

We rely on simple characterizations of the two forces. The desire to conform leads agents to match their value on attributes with that of another agent. Examples of conforming behavior rules are: *If my neighbor shares, then I will share* and *My neighbor wears a hat, so I will wear a hat.* The desire to be consistent leads agents to match their values on one attribute to their value on another attribute. Examples of consistency-enhancing behavior are: *I punish deviators in this context, I will do so in this other context as well* or *I keep a clean office, so I will keep a clean house.* In the combined model, agents try to conform *and* they try to be consistent. An agent applies the consistency rule with probability p.

Payoff Functions

Though not explicitly based on payoff functions, our behavioral rules converge to payoff maximizing configurations given natural characterizations of payoff functions. The desire for consistency can be captured in payoff form as an incentive to have as many attributes as possible take on the same value. Formally, let $s(a^j)$ equal the number of times the most common attribute appears in agent j's vector of attributes. We can write

$$s(a^j) = max_{a\ inA}\{|\ i\ |:\ a_i^j = a\}$$

This function offers a crude measure of consistency. We can think of it as a *consistency payoff function.* Given this payoff function, an optimizing agent would set all attributes to the same value. But, if the agent is not cognizant of all of its attributes' values that agent might randomly align its attribute values; we assume this behavioral rule. Such an assumption makes sense if we imagine attributes becoming activated and agents recognizing internal inconsistencies. Note that this behavioral rule would result in a consistent set of attributes.

To capture the payoff from conformity, we can let $f(a^j, a^{-j})$ equal the percentage of other agents whose attributes match those of agent j averaged across all attributes.

$$f(a^j, a^{-j}) = \frac{\sum_{k \neq j} \sum_{i=1}^{M} \delta(a_i^j, a_i^k)}{NM}$$

where $\delta(a_i^j, a_i^k) = 1$ if and only if $a_i^j = a_i^k$. We can think of this as the *conformity payoff function.* If agents cared only about conformity, then they would choose to acquire the most common value for each attribute. That sort of coordination is not easy to accomplish. A reasonable behavioral rule would be for an agent to switch one

However, the game form and the assumption of payoff maximization as a goal constrain the set of possible behavioral rules. Rule-based models need not be based on payoff functions. Rule based behaviors are easier to write but they need not be consistent with any reasonable assumptions about the underlying game form and the learning rule. This critique may not be as damning as it sounds. Nothing constrains human behavior to be consistent with an underlying game form and a learning rule. People do use simple behavioral rules that are inconsistent with the rational underpinnings associated with formal game theoretic models.

of its attribute values to match that of some other agent. This rule converges to full conformity if the agents update asynchronously (Page 1997).

For the combined model, we can write the payoff function to agent j, π_j as a convex combination of these two functions.

$$\pi_j(a^j, a^{-j}) = \alpha s(a^j) \; + (1-\alpha)\, f(a^j, a^{-j})$$

where $\alpha \in [0,1]$ denotes the relative weight on consistency.

In this combined model, the optimal solution would be for the agents to all choose the same values for each attribute. Consistent conformity at the societal level is far easier said than done. A natural question to ask is whether a given behavioral rule locates consistent conformity as an equilibrium. The rule we choose – a probabilistic combination of the consistency rule and the conformity rule – does. A myopic best response adjustment process in which an agent only switches an attribute's value if it leads to a higher payoff need not. That rule produces inefficient local optima. For example, an agent might be in a position such that it cannot become more consistent without reducing its payoff from conformity (Kuran and Sandholm 2003). Thus, our behavioral rule may make it harder for our model to support heterogeneity.

The Consistency Model

We now describe our models, beginning with the consistency model. In this model, the agents' only behavior is to adopt consistent values on attributes. In each period, we randomly select an agent. This agent then applies an *internal consistency rule.*

Internal Consistency Rule: *The agent randomly chooses two random distinct attributes and changes the value of the first attribute to match the value of the second.*

This rule produces an unbiased random walk where the probability of moving depends upon the state. To start we restrict attention to the case of binary attribute values. The extension to non binary attributes is notationally burdensome but straightforward. Let x denote the number of an agent's attributes with value one, so that $M - x$ attributes have value zero. The variable x can take on any value in the set $\{0, 1, 2, .., M\}$. If $x = 0$ or $x = M$, we say that the agent is *consistent.* For the moment, we assume no noise. We can think of an agent whose attributes all take the same value either as in an *equilibrium* or as in an *absorbing state*; we use these two terms interchangeably throughout this paper.

Consider the special case where $x = 1$. When we apply the internal consistency rule, the variable x could 1) fall to zero, 2) it could remain at one, or 3) it could increase to two. For x to fall to zero, the first attribute chosen must be the only attribute with value one. The probability of choosing this attribute equals $\frac{1}{M}$. The second attribute chosen necessarily has value zero (since "x = 1"). The probability of choosing the one and a zero therefore equals $\frac{1}{M}$. The variable x remains at one if and only if both attributes chosen have value zero. The probability of the first attribute having value zero equals $\frac{M-1}{M}$. The probability that the second attribute has value zero equals $\frac{M-2}{M-1}$. Thus, the probability of both events occurring equals $\frac{M-2}{M}$. Finally, x increases to two only if the first attribute selected has value zero and the second attribute selected has value one. The probabilities of these two events equal $\frac{M-1}{M}$ and $\frac{1}{M-1}$ respectively. So the probability of both events occurring equals $\frac{1}{M}$. Therefore, the probability of x increasing equals the probability that it decreases when $x = 1$. We can now state the following claim, whose proof relies on an extension of this logic.

Claim 1 *The internal consistency rule applied to M attributes that take on binary values produces an unbiased random walk in which the probability of movement slows near the two absorbing states. Let $x \in \{1, 2, ..M-1\}$, denote the number of attributes whose values equal one. The probability that x increases or decreases by one equals.*

$$\frac{(M-x)x}{M(M-1)}$$

pf. For x to increase, the first attribute must be one of the x attributes with value 0. This occurs with probability $\frac{x}{M}$ and the second attribute must belong to one of the $M-x$ attributes with value 1. This occurs with probability $\frac{M-x}{M-1}$. The proof for the case in which x decreases follows the same logic.

This claim implies that the internal consistency rule produces a random walk with two absorbing states. Moreover, the probability of movement decreases as the state approaches an absorbing state.

The Conformity Model

We next consider a model in which agents want to conform. In each period, we randomly choose a pair of agents. The first agent chosen applies the *external conformity rule.*

External Conformity Rule: *The first paired agent randomly chooses an attribute and sets the value of that attribute equal to the value that the other agent assigns to that attribute.*

The external conformity rule also creates a random walk. The next claim applies to a single attribute version ($M = 1$) of the model. The extension to the more general case is trivial.

Claim 2 *If $M = 1$ and if Y of the N agents assign value 0 to the lone attribute, then the probability that Y decreases after applying the external conformity rule equals $\frac{(N-Y)Y}{N(N-1)}$ which also equals the probability that Y increases.*

The proof of this claim follows from the fact that this process is equivalent to the one described in the consistency model. This suggests a deeper symmetry that can be made formal.

Observation: *The internal consistency model applied to N agents with M attributes is equivalent to the external conformity model applied to M agents with a N attributes.*

It follows that in the Conformity Model, the time it takes for the process to converge increases with the number of agents in the population just as in the Consistency model the time it takes for the process to converge increases with the number of attributes.

The Consistent Conformity Model

In the Consistent Conformity Model, agents apply both updating rules. We create a single parameter family of rules $CC(p)$ where p denotes the probability that the agent applies the internal consistency rule. Note that the consistency and conformity models are just special cases of this model, where $CC(1)$ is the consistency model and $CC(0)$ is the conformity model.

Consistent Conformity Rule CC(p): *An agent is chosen at random and with probability p the internal consistency rule is chosen and with probability $(1-p)$ the external conformity rule is chosen.*

Describing the dynamics of $CC(p)$ models are far more complicated. The only equilibria (absorbing states) of this model require that every agent assign the same value to every attribute. Let S_i equal the number of agents who assign value 0 to attribute i. The next claim describes the dynamics in the $CC(p)$ models for $p \in [0,1]$ from the perspective of an agent with x attributes equaling one. Without loss of generality, we assume that these are the first x attributes.

Claim 3 *Consider a population of N agents with M binary attributes, and an agent whose first x attributes take value one. Let S_i equal the number of other agents in the population who have value one on attribute i, the probability that x increases by one equals*

$$p\frac{x(M-x)}{M(M-1)} + (1-p)\frac{1}{M}\sum_{i=x+1}^{M}\frac{S_i}{N-1}$$

and the probability that x decreases by one equals

$$p\frac{x(M-x)}{M(M-1)} + (1-p)\frac{1}{M}\sum_{i=1}^{x}\frac{N-1-S_i}{N-1}$$

The proof follows directly from Claims 1 and 2.

A Simple Example

Before presenting our analytic results, we construct an example that demonstrates the tension between consistency and conformity. Suppose that two members of a society interact in three distinct contexts. In each context, a person can take a *fair* action, F, that equally splits resources or take a *utilitarian* action, U that produces a higher total payoff. These actions play the role of the values in our more general model.

Given these assumptions, we can describe an agent by a vector of length three consisting of F's and U's. Let's call these people George and Laura. Suppose that they start from the following initial behavioral vectors.

George (F, F, U)
Laura (F, U, U)

Assume first that George and Laura apply the internal consistency rule. . Under this assumption, George may switch his third attribute so that his vector of attributes becomes (F, F, F). Laura, in contrast, may switch her first attribute so that her vector becomes (U, U, U). George and Laura both achieve internal consistency and do so quickly.[8]

Next suppose that George and Laura apply the external conformity rule. If we pick George first, and further pick his second attribute, then George switches his second attribute to U so that his vector becomes (F, U, U) The two quickly conform.

Finally, assume that George and Laura desire both consistency and conformity. George may first switch to (F, F, F). He may then meet Laura and switch to (F, U, F). However, he may then realize that he is being inconsistent and switch back to (F, F, F). Laura meanwhile may switch to (U, U, U) and then, hoping to conform, switch back to (F, U, U). Eventually, both George and Laura will be consistent and conform with one another but it can take much longer. The desires to conform and to be consistent can pull in different directions thereby increasing the time required to attain an equilibrium.

3 Analytic Results

Our analytic results consider two questions. The time to convergence and the equilibrium distribution in models with errors. In computer science and physics the standard question to consider is the rate at which the time to convergence changes as you increase the number of states or variables. In our case, the analog would be the number of attributes in the consistency model and the number of agents in the conformity

[8] Note that George could also change to (F, U, U) or (U, F, U) given the internal consistency rule, but at some point, he would have all three of his attributes taking the same value.

model. Using techniques developed by Bouchaud et al (1999), it can be shown that the time to convergence is of order M^2.[9]

Claim 4 *The expected time to convergence for the consistency model with binary values and M attributes for a random starting point is of order M^2 periods.*

pf: see appendix.

We can state a similar result for the conformity model.

Corollary 1 *The expected time to convergence for the conformity model with binary values and N agents converges for a random starting point is of order N^2 periods.*

pf: follows from our earlier observation of equivalence and the previous claim.

In a conformity model with more than one attribute, we can think of each attribute converging independently of the others. There are no interactions between the attributes. Therefore, the time it takes for conformity should increase linearly in the number of attributes.

The time to convergence for the Consistent Conformity Model can be shown to increase in order N^2M^2 for $p = 1/2$ (Sander, Schneider-Mizell, and Page 2006) One reason that we should expect the Consistent Conformity Model to take substantially longer to converge is that it has far fewer equilibria given the behavioral rules we assume. We capture this fact in the next three claims.

Claim 5 *The number of equilibria in the Consistency Model equals A^M, where A equals the number of values per attribute and M is the number of agents.*

Claim 6 *The number of equilibria in the Conformity Model equals A^N, where A equals the number of values per attribute and N is the number of attributes.*

Claim 7 *The number of equilibria in the Consistent Conformity Model equals A, the number of values per attribute.*

The Two Agent Two Attribute Model

In our analytic model, we consider the simplest interesting case: two agents, two attributes, and two values per attribute (A=2, N=2, M=2). This model proves sufficient to show our two main results: that the Consistent Conformity model takes longer to converge than either of the other two models and that its equilibrium in the model with errors has greater dispersion.

[9]We thank Len Sander for this proof.

Time to Convergence

We first analyze time to convergence in the three modelsby calculating the time required for each of the three processes to converge. Given our asssumptions, the two agents in this model can together be in any one of sixteen states which can be sorted into five categories. These categories correspond to the two agents being in conformity and internally consistent (C&C), consistent but not conforming (CON), conforming but not consistent (CRD), one agent consistent but the other not – what we call off by one (OBO), or both inconsistent and lacking conformity (NOT). Using the letters a and b to denote distinct attribute values, we can define each category and its probabilities as in Table 1:

Table 1: States of the System

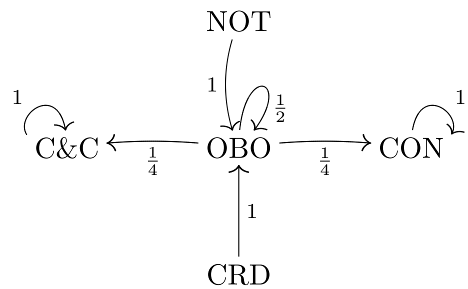
State	*Agents*	*Prob*
Conformed & Consistent (C&C)	(a,a) (a,a)	$\frac{1}{8}$
Consistent Not Conformed (CON)	(a,a) (b,b)	$\frac{1}{8}$
Conformed Not Consistent (CRD)	(a,b) (a,b)	$\frac{1}{8}$
Off By One (OBO)	(a,b) (a,a)	$\frac{1}{2}$
Not Conformed Not Consistent (NOT)	(a,b) (b,a)	$\frac{1}{8}$

For the internal consistency rule, the probability that $x = 0$ equals the probability that $x = 2$ which is $\frac{1}{4}$. The other half of the time $x = 1$. If $x = 1$, then in the first period the two attributes are selected and one matches the other and as a result, the agent becomes consistent. By the symmetry argument the expected time to equilibrium in the Consistency Model must equal the expected time to equilibrium in the conformity model. Nevertheless, making the calculation in both models is instructive.

Consistency Model

We first calculate expected time to equilibrium in the Consistency Model. In this model, any configurations in the sets $C\&C$ and CON are equilibria. We must first calculate the probability that any one of the other states moves to those states. If the initial state is in OBO, then the probability of staying in OBO equals one half, and the probability of moving to $C\&C$ or to CON equals one fourth. If the initial state is NOT or CRD, then it moves into OBO with probability one. We can write this information diagrammatically as shown in Figure 1:

Figure 1: The Dynamics of the Internal Consistency Rule



We can use the information in this diagram to calculate the expected time to convergence.

Claim 8 *The expected time to equilibrium for the Internal Consistency Rule equals* $1\frac{3}{4}$ *interactions.* [10]

pf: Let T_S denote the time (or expected time) to get to equilbrium from a given state. First, note that $T_{CON} = T_{C\&C} = 0$, since $C\&C$ and CON are absorbing states. Second note that the time to reach an absorbing state from a state in CRD or NOT equals one plus the time it takes to reach an absorbing state from OBO.

$T_{CRD} = T_{NOT} = 1 + T_{OBO}$

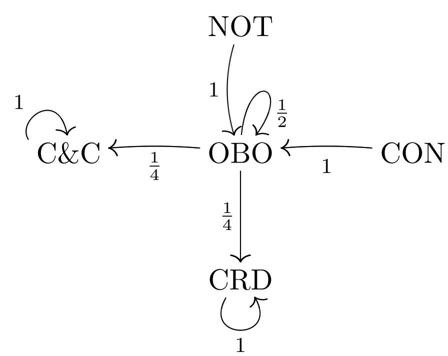
We calculate the expected time to reach an absorbing state from OBO as follows. With probability one half, it takes only one time period. The other half of the time, the process remains in OBO, which means the time to an absorbing state equals one plus the time to an absorbing state. We can write this as follows:

$T_{OBO} = \frac{1}{2}(1) + \frac{1}{2}(1 + T_{OBO}) \ = \ 1 + \frac{1}{2}T_{OBO}$

Solving for T_{OBO} yields that $T_{OBO} = 2$. Therefore $T_{CRD} = T_{NOT} = 3$, so applying the internal consistency rule, the expected time to attain an absorbing state, T^{ICR}, equals $T^{ICR} = \frac{1}{8}(0) + \frac{1}{8}(0) + \frac{1}{8}(3) + \frac{1}{8}(3) + \frac{1}{2}(2) \ = \ 1\frac{3}{4}$

[10] time is measured by the number of interactions (an interaction is one application of a rule) with each interaction taking one time step. Hence time is really a measure of the iterations of the model irrespective of the computational complexity of the iteration.

Figure 2: The Dynamics of the External Conformity Rule



Conformity Model

We can construct a similar diagram for the dynamics created by the external conformity rule (see Figure 2). Notice that this diagram is the same as the one above, with the only difference being that the states CRD and CON have changed places. Therefore, by symmetry the expected time to an absorbing state in this model is also $1\frac{3}{4} interactions$.

Claim 9 *The expected time to equilibrium for the External Conformity Rule equals* $1\frac{3}{4}$ *interactions.*

pf: follows from above.

CC(p) Model

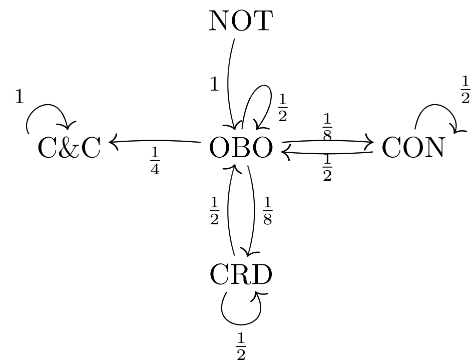
Next, we consider the $CC(p)$ model. In the diagram below, we show the case where $p = \frac{1}{2}$. The diagram for this model, Figure 3 combines the diagrams for the previous two models so that the only absorbing state is $C\&C$.
Using Figure 3, we can state the following claim.

Claim 10 *The expected time to equilibrium for the CC(p) Rule equals* $1\frac{7}{8} + \frac{1}{p(1-p)}$

pf: see appendix.

We can compare the expected time to equilibrium in the three models graphically. Figure 4 shows the expected time to equilibrium as a function of the probability of applying the consistency rule. Note first that the expected time to equilibrium is far shorter in the conformity model and the consistency model than in the $CC(p)$ model.

Figure 3: The Dynamics $CC(\frac{1}{2})$



Note also that the expected time to equilibrium is *minimized* in the $CC(p)$ model at $p = \frac{1}{2}$. For comparison, the time to convergence at $p = \frac{1}{2}$ is $5\frac{7}{8}$ interactions;this value is more than three times the time to convergence in the other two cases.

The three flow diagrams reveal the two reasons why consistent conformity takes longer to converge than either the conformity model or the consistency model. First, as we already proved, the consistent conformity model has fewer absorbing states. Whereas Figures 1 and 2 both have two categories of absorbing states, Figure 3 has a single category of absorbing states. Second, the two individual processes both head directly to the two absorbing states. The only possible delay occurs if the systems remains in state OBO. The consistent conformity model can move away from the lone absorbing state. It is possible for the process to go from CRD to OBO to CON and back to OBO. This can cause the system to take longer to converge.

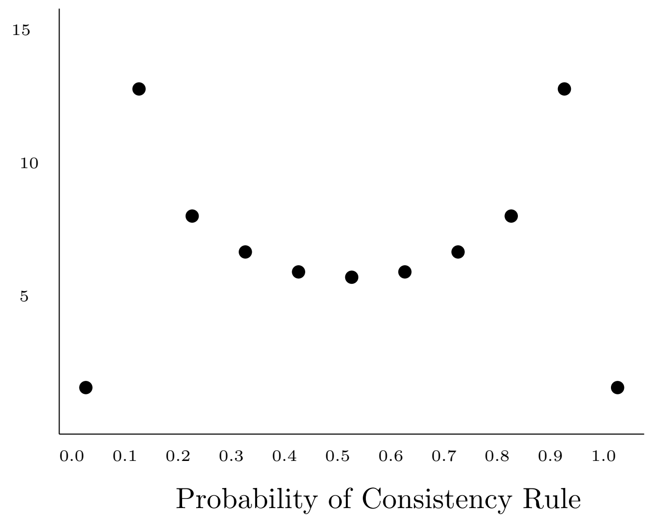
In the appendix, we also solve for time to convergence in the three binary attribute, two agent model. We find that the Consistency Model takes approximately twice as long as the Conformity model and that the time to convergence in the $CC(\frac{1}{2})$ model is more than double that of the Consistency Model and five times that of the Conformity model.

Equilibrium Distributions in a Model With Errors

We now further show the tension between consistency and conformity by considering the equilibrium distribution in models that include errors. The inclusion of errors is a standard assumption in learning and conformity models because they create *ergodicity* which guarantees a unique limiting distribution (Young 1998). That will also be the case here. In our models with errors, we obtain unique equilibrium distributions.

To capture errors, we assume that with some small positive probability, ϵ, an agent randomly changes an attribute's value rather than applying its behavioral rule. We

Figure 4: Expected Time To Equilibrium: Two Person Model



are interested in how the two forces singly and jointly magnify these errors. We might expect that by adding noise at a level ϵ creates an equilibrium distribution in which approximately ϵ of the agents are out of equilibrium. In the Consistency Model and the Conformity Model, we find something close to that. In the Consistent Conformity Model, however, the behavioral rule can magnify the noise term substantially.

Consistency Model

First, we look at the consistency model. It suffices to consider a single agent, which allows us to reduce our five states to three. We can let CNS denote the union of the states CON and $C\&C$. These represent the states where the agents are consistent. We can then combine the NOT and CRD into the state NCN. In this state, neither agent is consistent. This gives a Markov Process defined over three states CNS, NCN, and OBO. We can write the Markov Transition Matrix as follows:

$$
\begin{array}{cc|ccc}
 & & & T+1 & \\
 & & CNS & OBO & NCN \\
\hline
 & CNS & 1-\epsilon & \epsilon & 0 \\
\text{T} & OBO & \frac{1}{2} & \frac{1-\epsilon}{2} & \frac{\epsilon}{2} \\
 & NCN & 0 & 1 & 0
\end{array}
$$

This gives the following system of equations that characterize the equilibrium.

$P_{CNS} \; = \; (1-\epsilon)P_{CNS} + \frac{1}{2}P_{OBO}$

$$
\begin{aligned}
P_{OBO} &= \epsilon P_{CNS} + \tfrac{1-\epsilon}{2} P_{OBO} + P_{NCN} \\
P_{NCN} &= \tfrac{\epsilon}{2} P_{OBO}
\end{aligned}
$$

Solving these equations gives

$$
\begin{aligned}
P_{CNS} &= \tfrac{1}{1+2\epsilon+\epsilon^2} \\
P_{OBO} &= \tfrac{2\epsilon}{1+2\epsilon+\epsilon^2} \\
P_{NCN} &= \tfrac{\epsilon^2}{1+2\epsilon+\epsilon^2}
\end{aligned}
$$

Conformity Model

To analyze the the conformity model, we also combine states. Let CDC equal the union of the two states in which the two agents have confromed, CRD and $C\&C$, and let NCD equal the union of the states in which they have not, NOT and CON. We can write the Markov Transition Matrix as follows

		$T+1$		
		CDC	OBO	NCD
T	CDC	$1-\epsilon$	ϵ	0
	OBO	$\frac{1}{2}$	$\frac{1-\epsilon}{2}$	$\frac{\epsilon}{2}$
	NCD	0	1	0

This matrix is identical to the one for the Consistency Model up to a relabeling of the states. Therefore, the equilibrium equals

$$
\begin{aligned}
P_{CDC} &= \tfrac{1}{1+2\epsilon+\epsilon^2} \\
P_{OBO} &= \tfrac{2\epsilon}{1+2\epsilon+\epsilon^2} \\
P_{NCD} &= \tfrac{\epsilon^2}{1+2\epsilon+\epsilon^2}
\end{aligned}
$$

CC(p) Model

For the Consistent Conformity Model, we require all five categories of states. We can write the Markov Transition matrix between those states as follows

		$T+1$				
		$C\&C$	OBO	CRD	CON	NOT
T	$C\&C$	$1-\epsilon$	ϵ	0	0	0
	OBO	$\frac{1}{4}$	$\frac{(1-\epsilon)}{2}$	$\frac{p+\epsilon-\epsilon p}{4}$	$\frac{1-p-\epsilon p}{4}$	$\frac{\epsilon}{4}$
	CRD	0	$1-p+\epsilon p$	$p-\epsilon p$	0	0
	CON	0	$p+\epsilon+\epsilon p$	0	$1-p-\epsilon+\epsilon p$	0
	NOT	0	1	0	0	0

The following system of five equations characterizes the equilibrium.

$$\begin{aligned}
P_{C\&C} &= (1-\epsilon)P_{C\&C} + \tfrac{P_{OBO}}{4} \\
P_{OBO} &= P_{OBO} + \tfrac{(1-\epsilon)P_{OBO}}{2} + (1-p+\epsilon p)P_{CRD} + (p+\epsilon-\epsilon p)P_{CON} + \tfrac{\epsilon P_{OBO}}{4} \\
P_{CRD} &= \tfrac{(p+\epsilon-\epsilon p)P_{OBO}}{4} + (p-\epsilon p)P_{CRD} \\
P_{CON} &= \tfrac{(1-p+\epsilon p)P_{OBO}}{4} - (1-p-\epsilon+\epsilon p)P_{CON} \\
P_{NOT} &= \tfrac{\epsilon P_{OBO}}{4}
\end{aligned}$$

Solving gives the following

$$\begin{aligned}
P_{C\&C} &= \tfrac{1}{1+4\epsilon+\epsilon^2+\alpha\epsilon+\alpha^{-1}\epsilon} \\
P_{OBO} &= \tfrac{4\epsilon}{1+4\epsilon+\epsilon^2+\alpha\epsilon+\alpha^{-1}\epsilon} \\
P_{CRD} &= \tfrac{\alpha\epsilon}{1+4\epsilon+\epsilon^2+\alpha\epsilon+\alpha^{-1}\epsilon} \\
P_{CON} &= \tfrac{\alpha^{-1}\epsilon}{1+4\epsilon+\epsilon^2+\alpha\epsilon+\alpha^{-1}\epsilon} \\
P_{NOT} &= \tfrac{\epsilon^2}{1+4\epsilon+\epsilon^2+\alpha\epsilon+\alpha^{-1}\epsilon}
\end{aligned}$$

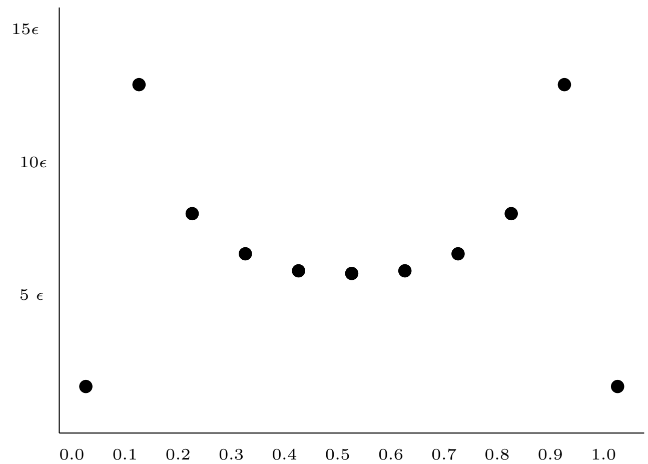
Where $\alpha = \frac{(p+\epsilon-\epsilon p)}{(1-p+\epsilon p)}$ which equals the ratio of the probability of moving from OBO to CON to the probability of moving from OBO to CRD. The higher α, the more time the system will spend in CON. The lower α, the more time that the system will spend in CRD. Setting $p = \frac{1}{2}$ maximizes the time spent in the consistent conformity state ($C\&C$). Figure 5 shows the percentage of the time the system spends outside of state $C\&C$ as a function of p for a given error level ϵ. If we let p go to 0 then α converges to ϵ and the system spends half of the time outside of the state $C\&C$. Similarly, if we let p go to 1 then α converges to $\frac{1}{\epsilon}$, and the system again spends half of the time outside of the state $C\&C$.

Note that except for the units on the y-axis, this figure matches figure 4 exactly. The equivalence, modulo a rescaling, of the time to equilibrium and the distance to the perfectly conformed and consistent equilibrium is an artifact of our assumptions. But the correlation between the two generally hints at an important insight: the longer the time to equilibrium, the more complex the dynamics. The more complex the dynamics, the larger the potential effects of error.

4 Numerical Experiments

To test whether the results from our simple model extend to larger numbers of agents, attributes, and attribute values, we ran numerical experiments in which we varied the number of agents from two to one thousand, the number of attributes from two to

Figure 5: Distance to Conformed and Consistent Equilibrium: Error Model



ten, and the number of values per attribute from two to six.[11] The results that we found were consistent with the analytic results from our two-person, two-attribute model.

We present here two sets of experiments. In the first set, we measure the time to convergence. In the second set, we measure levels of consistency and conformity in the models with errors.

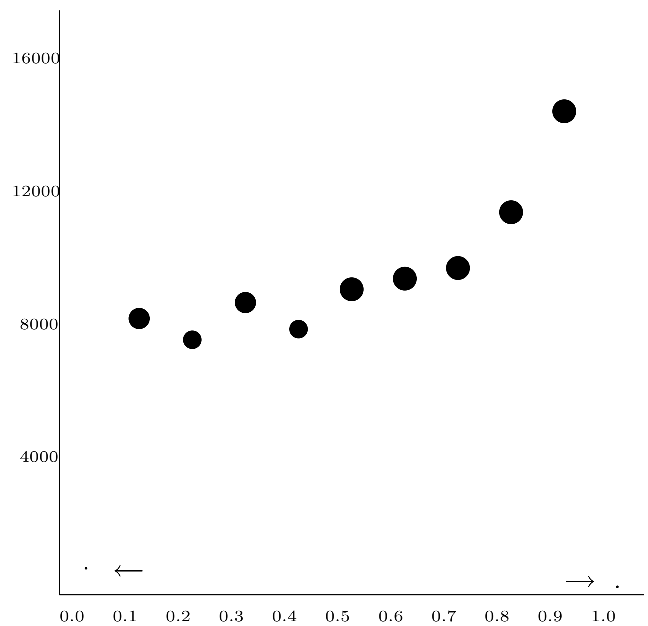
Time To Convergence Experiments

Figure 6 shows the time to convergence as a function of p for a model with one hundred agents, ten attributes, and six values per attribute. The results are averages of over fifty trials. All of the differences are statistically significant. The arrows point to the values for $p = 0$ and $p = 1$, which are otherwise easy to overlook.

Our theoretical results suggested that the time to convergence should increase as p approaches zero and one. Here, we only see that phenomenon as p approaches one. The asymmetry can be explained by the fact the number of agents is far larger than the number of attributes. The probability of applying the consistency rule must be very small before we would expect to see the time to convergence to increase given the greater need for conformity.

[11]We wrote two separate programs, one in C and one in Repast (a java-based modeling toolkit). We used the faster C program to sweep the attribute values, and the Repast program to generate the graphs that you see in the paper. We also tested our models against the analytic results presented in the paper.

Figure 6: Time to Convergence in Number of Periods



Even though all of the models converge, the magnitude of these differences appears meaningful. In the Conformity Model and the Consistency Model, the system converges in a few hundred periods. The Consistent Conformity model can take more than fifteen thousand periods to converge. Our model is abstract enough that we need not attach any specific span of time to a period; time is simply the number of agent actions. However, if periods are a day, then this difference is between a few months and more than forty years.

Experiments in Models with Error

In the next set of experiments, we test to see whether errors have a much larger effect in the Consistent Conformity Model. To make this comparison, we need some measures of consistency and conformity. In constructing these measures, we refer back to notation we used in constructing possible utility functions. Recall that $s(a^j)$ equals the number of times the most common attribute appears in agent j's vector of attributes. We can write

$$s(a^j) = max_{a\ inA}\{|\ i\ |:\ a_i^j = a\}$$

$$\text{pconsistent} = \frac{\sum_{j=1}^{M} s(a^j)}{AM}$$

Thus, *pconsistent* takes on values between 0 and 1. We define *pconformity* to be the average of the *conformity payoff functions*. Recall that the conformity payoff function equals the average number of agents who agree with the agent's attribute values.

$$f(a^j, a^{-j}) = \frac{\sum_{k \neq j} \sum_{i=1}^{M} \delta(a_i^j, a_i^k)}{NM}$$

where $\delta(a_i^j, a_i^k) = 1$ if and only if $a_i^j = a_i^k$

$$\text{pconformity} = \frac{\sum_{j=1}^{M} f(a^j, a^{-j})}{M}$$

Thus, if the entire population has conformed, then the value of *pcoordinate* equals one. The table below gives the values of *pconsistent* and *pconformity* for each of the three models under various levels of agent error for a model with ten attributes and five values per attribute and 100 agents.

		probability of consistency check					
		p = 0.0		p = 0.5		p = 1.0	
		pconformity	pconsistent	pconformity	pconsistent	pconformity	pconsistent
noise	0	1	0.360	1	1	0.200	1
	0.005	0.736	0.373	0.354	0.556	0.199	0.970
	0.01	0.585	0.376	0.299	0.510	0.200	0.946
	0.02	0.482	0.376	0.269	0.483	0.201	0.904

Table 2: **Consistency and Conformity Environments.** Average percent values and standard deviations of inter-agent value difference (pconformity) and intra-agent value difference (pconsistent) over the last 1000 interactions of 100 runs with 100 agents, 10 attributes, 5 values per attribute and a total run time of 5,000,000 interactions per run.

Notice that with no errors, the CC($\frac{1}{2}$) model converges to a consistent and coordinated state as we expect. Further, for the $CC(\frac{1}{2})$ model, the introduction of even the tiniest bit of noise (0.005) leads to substantial heterogeneity both between agents (0.354) and within agents (0.556), far more than in the other two models (0.736 and 0.970 respectively). A little noise has a much larger effect when both forces operate.[12]

[12]Comparing results for cases with noise = 0.005, the p-value for a test of the difference of means for conformity for the p=0.0 and p=0.5 models is 2.23×10^{-304} and the p-value for a test of the difference of means for consistency for the p=0.5 and p=1.0 models is 1.61×10^{-831}.

5 Discussion

In this paper we have shown how forces for consistency and conformity slow convergence in a model of cultural formation. In addition, when both forces are present, varying the dominance of one over the other resulted in even slower convergence. Finally, in a model with errors the equilibrium distribution includes substantial heterogeneity. The two forces of consistency and conformity magnify the effects of error and maintain the system away from equilibrium. Thus, our model offers an explanation for the substantial within-culture heterogeneity observed in the real world.

Of course, other candidate explanations exist for intra-cultural heterogeneity. Within a culture people differ in their preferences, experiences, and ambitions, which, too, can create behavioral differences within a culture. Moreover, just as people choose to conform their attributes to be like others, they also try to distinguish themselves. And, as we already mentioned, people typically interact in small groups, so within-group conformity and consistency need not necessarily lead to conformity and consistency at the societal level. We do not dispute or challenge these other causes of within-culture heterogeneity, nor do we think that their inclusion would have straightforward implications. Such an inference would run counter to the main methodological contribution of this paper: the *non-linear* additivity of dynamical systems. That methodological insight—that the whole differs from the parts when we look at forces—may be as important as its application within our model to culture.

Our analysis suggests some testable hypotheses and a possible rethinking of how we measure intra-cultural heterogeneity. The greater the likelihood of error, the less conformity we should see. Informational systems might provide a crude proxy for the transmission error of cultural traits. If so, we might expect less economically advanced cultures to be more heterogeneous.

In a society in which the relative tendency to conform is high relative to the tendency to be consistent, people may be less consistent but more similar. Thus, whether one culture appears more or less heterogeneous depends on the type of questions asked. If the questions ask about an existing behavior, we'd expect a higher conforming society to appear less heterogeneous. However, if the questions are hypothetical, the lack of consistency may give respondents a variety of possible behaviors to apply in the novel context. Thus, a less individualistic society, like Japan, could appear more heterogeneous than a highly individualistic society like the United States.

As mentioned in the introduction, the model also applies to corporate and organizational cultures and to the formation of political parties. In the party model, attribute values could represent participants' ideal points in policy or preference space. The internal consistency rule would capture the individual desire for a consistent ideology, and the external conformity rule would capture the collective desire for party cohesiveness. The model suggests that a consistent cohesiveness would not emerge quickly without top down encouragement or even enforcement.

We can push this insight even further. Within any organization or collection of people, be it an interest group, a community organization, or an academic department these same two forces may operate. People seek common ground—they want to conform—and they also want to be consistent. Yet, we've seen that even with a little bit of error, these two forces do not result in a coherent, consistent set of attributes. This finding agrees with what we see in the real world. Few, in any, groups and organizations converge to a state of consistent conformity.

We must be careful not to attach normative significance to reducing intra-group heterogeneity. This lack of convergence, be it in a society, a political, party, or an organization, may, on balance, be a good thing. Intra-group heterogeneity allows for experimentation. It allows a collection of individuals to balance exploration with exploitation by maintaining the variation necessary for further exploration (March 1991, Axelrod and Cohen 2000) and better problem solving and prediction (Page 2007). Thus, diversity may make societies more robust by providing the potential to adapt to changing circumstances (Bednar 2006). Societies that lack intra-cultural diversity may be prone to collapse (Diamond 2005). The persistence of diversity in the face of two homogenizing forces can therefore be seen as not only counterintuitive but serendipitous.

Appendix

Claim 4 *The time to convergence for the consistency model with binary values and M attributes of order M^2.*

pf:[13] Let x denote the number of attributes with value 1. Let T_x be the time to convergence if at location x. Let m_x be the probability of increasing or decreasing the number of attributes with value 1. By the previous claim, these probabilities are equal. After one time period, the expected time has to be one period less. Therefore, we have the following equation:

$$T_x - 1 = m_x T_{x+1} + m_x T_{x-1} + (1 - 2m_x)T_x$$

This reduces to

$$-1 = m_x[(T_{x+1} - T_x) - (T_x - T_{x-1})]$$

Recall from Claim 1 that $m_x = \frac{(M-x)x}{M(M-1)}$. For large M we can approximate this as $m_x = \frac{(M-x)x}{M^2}$. Let $p(x) = \frac{x}{M}$, so that $m_x = p(x)[1 - p(x)]$. We then can rewrite $T_{x+1} - T_x$ as

$$\frac{1}{M} \cdot \frac{(T(p(x+1)) - T(p(x))}{\frac{1}{M}}$$

For large M, this converges to $\frac{\partial T(p(x))}{\partial p}$. It follows that we can write the following approximation:

$$(T_{x+1} - T_x) - (T_x - T_{x-1}) \sim \frac{1}{M}\left[\frac{\partial T(p(x))}{\partial p} - \frac{\partial T(p(x-1))}{\partial p}\right]$$

Which in turn we can approximate as

$$\frac{1}{M^2}\frac{\partial^2 T(p(x)}{\partial p}$$

We can therefore approximate our initial difference equation as

$$-1 = p(x)[1 - p(x)]\frac{1}{M^2}\frac{\partial^2 T(p(x))}{\partial p}$$

Rearranging terms and simplifying notation gives

$$\frac{\partial^2 T(p)}{\partial p^2} = -\frac{M^2}{p(1-p)}$$

[13]We provide a somewhat loose proof here that requires approximations. The result has been verified in simulations frequently by computer scientists and physicists.

We also have that $T(0) = T(1) = 0$. The solution to this differential equation is

$$T(p) = M^2 \left[p \log(\frac{1}{p}) + (1-p)\log(\frac{1}{1-p}) \right]$$

which completes the proof.

Claim 10: *The expected time to equilibrium for the CC(p) Rule equals* $1\frac{7}{8} + \frac{1}{p(1-p)}$
pf: We can write the following equations.

$T_{C\&C} = 0$
$T_{OBO} = 1 + \frac{1}{4}T_{C\&C} + \frac{1}{2}T_{OBO} + \frac{p}{4}T_{CON} + \frac{(1-p)}{4}T_{CRD}$

$T_{CON} = 1 + (1-p)T_{OBO} + pT_{CON}$

$T_{CRD} = 1 + pT_{OBO} + (1-p)T_{CRD}$

$T_{NOT} = 1 + T_{OBO}$

By substitution, these equations imply that

$T_{CON} = \frac{1}{1-p} + T_{OBO}$

$T_{CRD} = \frac{1}{p} + T_{OBO}$

These in turn imply that

$T_{OBO} = 1 + \frac{1}{2}T_{OBO} + \frac{p}{4(1-p)} + \frac{(1-p)}{4p} + \frac{1}{4}T_{OBO}$

This reduces to

$T_{OBO} = 4 + \frac{(1-2p+2p^2)}{p(1-p)}$

Substituting back into the other equations gives

$T_{CON} = 4 + \frac{(1-p+2p^2)}{p(1-p)}$

$T_{CRD} = 4 + \frac{(2-3p+2p^2)}{p(1-p)}$

$T_{NOT} = 5 + \frac{(1-2p+2p^2)}{p(1-p)}$

Therefore the average time to convergence equals

$$\frac{1}{2}\left(4 + \frac{(1-2p+2p^2)}{p(1-p)}\right) + \frac{1}{8}\left(4 + \frac{(1-p+2p^2)}{p(1-p)} + 4 + \frac{(2-3p+2p^2)}{p(1-p)} + 5 + \frac{(1-2p+2p^2)}{p(1-p)}\right)$$

Which reduces to

$1\frac{7}{8} + \frac{1}{p(1-p)}$

For the special case $p = \frac{1}{2}$, these equations become

$T_{OBO} = 1 + \frac{1}{4}T_{C\&C} + \frac{1}{2}T_{OBO} + \frac{1}{8}T_{CON} + \frac{1}{8}T_{CRD}$

$T_{CON} = 1 + \frac{1}{2}T_{OBO} + \frac{1}{2}T_{CON}$

$T_{CRD} = 1 + \frac{1}{2}T_{OBO} + \frac{1}{2}T_{CRD}$

$T_{NOT} = 1 + T_{OBO}$

By substitution, these equations imply that $T_{CON} = T_{CRD} = 2 + T_{OBO}$. Which in turn implies that $T_{OBO} = 1 + \frac{1}{2}T_{OBO} + \frac{1}{2} + \frac{1}{4}T_{OBO}$. This is an equation in a single variable, T_{OBO}. Solving gives equation gives $T_{OBO} = 6$. Substituting back into the other equations gives $T_{CON} = T_{CRD} = 8$ and $T_{NOT} = 7$. Therefore the average time to convergence equals $\frac{1}{2}(6) + \frac{1}{8}(8 + 8 + 7) \ = \ 5\frac{7}{8}$

Three Attribute Model

We can extend our model to include a third attribute. This increases the number of possible states from sixteen to sixty four. To describe the dynamics of this system, we create eight categories:

State	*Agents*	*Prob*
Coordinated & Consistent (C&C)	(a,a,a) (a,a,a)	$\frac{1}{32}$
Consistent Not Coordinated (CON)	(a,a,a) (b,b,b)	$\frac{1}{32}$
Coordinated Not Consistent (CRD)	(a,b,b) (a,b,b)	$\frac{3}{32}$
Off By One (OBO)	(a,a,b) (a,a,a)	$\frac{6}{32}$
Off By Two (OBT)	(a,b,b) (a,a,a)	$\frac{6}{32}$
Two By Two (TWO)	(a,a,b) (a,b,b)	$\frac{6}{32}$
Switch Two (SWI)	(a,a,b) (a,b,a)	$\frac{6}{32}$
Mirror States (MIR)	(a,a,b) (b,b,a)	$\frac{3}{32}$

We can then prove similar claims for time to convergence.

Claim 11 *With two agents and three binary attributes, the expected time to equilibrium for the Internal Consistency Rule equals* $7\frac{5}{16}$.

pf: If we apply the internal consistency rule we get the following system of equations

$T_{OBO} = 1 + \frac{2}{3}T_{OBO} + \frac{1}{6}T_{OBT}$

$T_{OBT} = 1 + \frac{2}{3}T_{OBT} + \frac{1}{6}T_{OBO}$

$T_{CRD} = 1 + \frac{1}{3}T_{CRD} + \frac{1}{3}T_{TWO} + \frac{1}{3}T_{OBO}$

$T_{MIR} = 1 + \frac{1}{3}T_{MIR} + \frac{1}{3}T_{SWI} + \frac{1}{3}T_{OBT}$

$T_{TWO} = 1 + \frac{1}{3}T_{TWO} + \frac{1}{6}T_{CRD} + \frac{1}{6}T_{SWI} + \frac{1}{3}T_{OBT}$

$T_{SWI} = 1 + \frac{1}{3}T_{SWI} + \frac{1}{6}T_{MIR} + \frac{1}{6}T_{TWO} + \frac{1}{3}T_{OBO}$

The solution to this set of equations equals $T_{OBO} = 6$, $T_{OBT} = 6$, $T_{CRD} = 9$, $T_{TWO} = 9$, $T_{SWI} = 9$, $T_{MIR} = 9$. Plugging these back into the probabilities of each initial state gives the result.

Claim 12 *With two agents and three binary attributes, the expected time to equilibrium for the External Conformity Rule equals* $3\frac{1}{2}$.[14]

pf: If we apply the external conformity rule we get the following system of equations

$T_{OBO} = 1 + \frac{2}{3}T_{OBO}$

$T_{TWO} = 1 + \frac{2}{3}T_{TWO}$

$T_{CON} = 1 + T_{OBT}$

$T_{OBT} = 1 + \frac{1}{3}T_{OBT} + \frac{1}{3}T_{TWO} + \frac{1}{3}T_{OBO}$

$T_{SWI} = 1 + \frac{1}{3}T_{SWI} + \frac{1}{3}T_{OBO} + \frac{1}{3}T_{TWO}$

$T_{MIR} = 1 + \frac{2}{3}T_{SWI} + \frac{1}{3}T_{OBT}$

The solution to this set of equations equals $T_{OBO} = 3$, $T_{TWO} = 3$, $T_{CON} = \frac{11}{2}$, $T_{OBT} = \frac{9}{2}$, $T_{SWI} = \frac{9}{2}$, $T_{MIR} = \frac{11}{2}$. Plugging these back into the probabilities of each initial state gives the result.

Our final claim considers the CC(p) model. Here, we provide a numerical result for the case $p = 1/2$.

Claim 13 *With two agents and three binary attributes, the expected time to equilibrium for the CC(*$\frac{1}{2}$*) equals approximately 17.91.*

pf: Using the notation from above, we can write the equations for the time to convergence as follows:

[14]We thank Casey Schneider-Mizell for correcting an earlier proof an providing the matrix representation

$$
\begin{aligned}
T_{CON} &= 1 + (1-p)T_{CON} + pT_{OBT} \\
T_{CRD} &= 1 + (1-p)T_{CRD} + p(\frac{1}{3}T_{OBO} + \frac{1}{3}T_{TWO} + \frac{1}{3}T_{CRD}) \\
T_{OBO} &= 1 + (1-p)(\frac{2}{3}T_{OBO} + \frac{1}{6}T_{CRD}) + p(\frac{2}{3}T_{OBO} + \frac{1}{6}T_{OBT}) \\
T_{OBT} &= 1 + (1-p)(\frac{1}{3}T_{OBT} + \frac{1}{3}T_{OBO} + \frac{1}{3}T_{TWO}) + p(\frac{2}{3}T_{OBT} + \frac{1}{6}T_{CRD} + \frac{1}{6}T_{OBO}) \\
T_{TWO} &= 1 + (1-p)(\frac{2}{3}T_{TWO} + \frac{1}{3}T_{CRD}) + p(\frac{1}{3}T_{OBT} + \frac{1}{6}T_{CRD} + \frac{1}{6}T_{SWI} + \frac{1}{3}T_{TWO}) \\
T_{SWI} &= 1 + (1-p)(\frac{1}{3}T_{SWI} + \frac{1}{3}T_{OBO} + \frac{1}{3}T_{TWO}) + p(\frac{1}{3}T_{SWI} + \frac{1}{3}T_{OBO} + \frac{1}{6}T_{MIR} + \frac{1}{6}T_{TWO}) \\
T_{MIR} &= 1 + (1-p)(\frac{1}{3}T_{OBT} + \frac{2}{3}T_{SWI}) + p(\frac{1}{3}T_{MIR} + \frac{1}{3}T_{SWI} + \frac{1}{3}T_{OBT})
\end{aligned}
$$

We can write this in matrix form as follows:

$$
\begin{pmatrix} T_{CON} \\ T_{CRD} \\ T_{OBO} \\ T_{OBT} \\ T_{TWO} \\ T_{SWI} \\ T_{MIR} \end{pmatrix}
=
\begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}
+
\begin{pmatrix}
1-p & 0 & 0 & p & 0 & 0 & 0 \\
0 & 1-\frac{2}{3}p & \frac{1}{3}p & 0 & \frac{1}{3}p & 0 & 0 \\
0 & \frac{1}{6}(1-p) & \frac{2}{3} & \frac{1}{6}p & 0 & 0 & 0 \\
0 & \frac{1}{6}p & \frac{1}{3}-\frac{1}{6}p & \frac{1}{3}+\frac{1}{3}p & \frac{1}{3}(1-p) & 0 & 0 \\
0 & \frac{1}{3}-\frac{1}{6}p & 0 & \frac{1}{3}p & \frac{2}{3}-\frac{1}{3}p & \frac{1}{6}p & 0 \\
0 & 0 & \frac{1}{3} & 0 & \frac{1}{3}-\frac{1}{6}p & \frac{1}{3} & \frac{1}{6}p \\
0 & 0 & 0 & \frac{1}{3} & 0 & \frac{2}{3}-\frac{1}{3}p & \frac{1}{3}p
\end{pmatrix}
\begin{pmatrix} T_{CON} \\ T_{CRD} \\ T_{OBO} \\ T_{OBT} \\ T_{TWO} \\ T_{SWI} \\ T_{MIR} \end{pmatrix}
$$

Solving these equations and multiplying each T_x by the probability of starting in state x produces the result.

		probability of consistency check											
		p = 0.0				p = 0.5				p = 1.0			
		pconformity		pconsistent		pconformity		pconsistent		pconformity		pconsistent	
		mean	stdev	mean	stdev	mean	stdev	mean	stdev	mean	stdev	mean	stdev
noise	0	1	0	0.360	0.082	1	0	1	0	0.200	0.016	1	0
	0.005	0.736	0.064	0.373	0.044	0.354	0.081	0.556	0.067	0.199	0.012	0.970	0.009
	0.01	0.585	0.052	0.376	0.030	0.299	0.037	0.510	0.033	0.200	0.012	0.946	0.012
	0.02	0.482	0.044	0.376	0.023	0.269	0.017	0.483	0.017	0.201	0.012	0.904	0.017

Table 3: **Consistency and Conformity Environments with Error.** Average percent values and standard deviations of inter-agent value difference (pconformity) and intra-agent value difference (pconsistent) over the last 1000 interactions of 100 runs with 100 agents, 10 features, 5 values per feature and a total run time of 5,000,000 interactions per run.

References

[1] Adler, Nancy J. and John L. Graham. 1989. "Cross-Cultural Interaction: The International Comparison Fallacy?" *Journal of International Business Studies* 20(3): 515-537.

[2] Au, Kevin Y. (1999). "Intra-Cultural Variation: Evidence and Implications for International Business." *Journal of International Business Studies* 30(4): 799-812.

[3] Axelrod, Robert. (1997). The Dissemination of Culture: A Model with Local Convergence and Global Polarization. *Journal of Conflict Resolution* 41:203-226.

[4] Axelrod, Robert and Michael D. Cohen. (2000) *Harnessing Complexity: Organizational Implications of a Scientific Frontier.* Free Press. New York

[5] Bandura, A. (1977). *Social Learning Theory.* New York: General Learning Press.

[6] Banerjee, Abhijit (1992). A Simple Model of Herd Behavior. *Quarterly Journal of Economics,* 107(3): 797-817.

[7] Barth, Frederick, ed. (1969). *Ethnic groups and boundaries: The Social Organization of Culture Difference*. Boston: Little, Brown.

[8] Bednar, Jenna. 2006. *The Robust Federation.* University of Michigan book manuscript.

[9] Bednar, Jenna and Scott E. Page. 2006. "Can Game(s) Theory Explain Culture? The Emergence of Cultural Behavior within Multiple Games." Forthcoming in *Rationality and Society* 18(4), November 2006.

[10] Bernheim, Douglas (1994). A Theory of Conformity. *Journal of Political Economy,* 102(5): 841-877.

[11] Bikhchandani, Sushil, David Hirshleifer, and Ivo Welch (1992). A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades. *Journal of Political Economy,* 100(5): 992-1026.

[12] Brock, William A. and Stephen N. Durlauf (2006). "Identification of binary choice models with social interactions." *Journal of Econometrics.*

[13] Bouchaud, J-P, A. Georges, J. Koplik, A. Provata, and S. Redner (1990): Superdiffusion in Random Velocity Fields", *Physical Review Letters*: 64, 2503-2506

[14] Boyd, Robert and Peter J. Richerson. (1985). *Culture and the Evolutionary Process.* Chicago: The University of Chicago Press.

[15] Calvert, Randall and James Johnson. 1997. "Interpretation and Coordination in Constitutional Politics." University of Rochester manuscript.

[16] Campbell. (1963). "Variation and Selective Retention in Socio-Cultural Evolution." In *Social Change in Developing Areas*, pp. 19-49. Edited by H. Barringer et al. Cambridge, MA: Schenkman.

[17] Diamond, Jared. (2005)*Collapse: How Societies Choose to Fail or Succeed.* New York: Viking.

[18] Durham, W. H. (1991). *Coevolution: Genes, Culture, and Human Diversity.* Stanford, CA: Stanford University Press.

[19] Gazzaniga, M. S., Ed. (1999). (editor) *Conversations in the Cognitive Neurosciences.* MIT Press.

[20] Hannan, Timothy H. (1979). "Expense-Preference Behavior in Banking: A Reexamination." *Journal of Political Economy* 87(4): 891-95.

[21] Henrich, Joseph. (2000). "Does Culture Matter in Economic Behavior? Ultimatum Game Bargaining Among the Machiguenga of the Peruvian Amazon." em The American Economic Review 90(4): 973-979.

[22] Henrich, Joseph, Robert Boyd, Samuel Bowles, Colin Camerer, Ernst Fehr, Herbert Gintis, and Richard McElreath. (2001). "Cooperation, Reciprocity, and Punishment in Fifteen Small-Scale Societies." Working Paper, Santa Fe Institute.

[23] Hofstede, Geert. (2001). *Culture's Consequences, Comparing Values, Behaviors, Institutions, and Organizations Across Nations.* Thousand Oaks, CA: Sage Publications.

[24] Homans, G.C. (1950). *The Human Group.* New York: Harpers.

[25] Huesmann, L.R. (1988). "An Information Processing Model for the Development of Aggression." *European Journal of Personality*, 3(2): 95-106.

[26] Huesmann, L.R. (1998). "The Role of Social Information Processing and Cognitive Schema in the Acquisition and Maintenance of Habitual Aggressive Behavior" In R. G. Geen & E. Donnerstein (eds.), *Human Aggression: Theories, Research, and Implications for Policy*, pp. 73-109.. New York: Academic Press.

[27] Inglehart, R. (1977). *The Silent Revolution: Changing Values and Political Styles Among Western Publics.* Oxford: Princeton University Press.

[28] Kameda, Tatsuya and Daisuke Nakanishi. (2002). *Does Social/Cultural Learning Increase Human Adaptibility? Roger's Questions Revisited.* Japan: Hokkeido University.

[29] Kennedy, Rodney J. (1988). "The Intermediary and Social Distance in Western Torres Strait." In Marvin K. Mayers and Daniel D. Rath (eds.), *Nucleation in Papua New Guinea cultures*, pp. 87-103. Dallas: International Museum of Cultures Publication, 23.

[30] Kuran, Timur (1995). *Private Truths and Public Lies: The Social Consequences of Preference Falsification* Cambridge, MA: Harvard University Press.

[31] Kuran, Timur and William Sandholm (2003) "Cultural Integration and Its Discontents" USC CLEO Research Paper No C02-14

[32] March, J. G. (1991). "Exploration and exploitation in organizational learning." *Organization Science*, 2(1), 71-87.

[33] Nisbett, Richard. E., and Ross, Lee. (1980). *Human Inference: Strategies and Shortcomings of Social Judgment.* Englewood-Cliffs: Prentice-Hall.

[34] Page, Scott (1997) On Incentives and Updating in Agent Based Models. *Computational Economics.* 10 pp 67-87.

[35] Page, Scott (2007) *The Difference: How the Power of Diversity Creates Better Groups, Firms, Schools, and Societies.* Princeton University Press (forthcoming, December 2006).

[36] Pavlov, I.P. (1903.) "The Experimental Psychology and Psychopathology of Animals." I.P. Pavlov: Selected Works. Honolulu, Hawaii: University Press of the Pacific, 2001.

[37] Pelto, Gretel H. and Pertti J. Pelto. (1975) "Intra-Cultural Diversity: Some Theoretical Issues." *American Ethnologist* 2(1), Intra-Cultural Variation: 1-18.

[38] Roberts, John Milton. (1964). "The Self-Management of Cultures." In *Explorations in Cultural Anthropology: Essays in Honor of George Peter Murdock*, ed. W. H. Goodenough, pp. 433-54. New York: McGraw-Hill Book Company.

[39] Rogers, Everett M. (1983). *Diffusion of Innovations* (Fourth Edition), New York: Free Press.

[40] Schelling, Thomas. (1978). *Micromotives and Macrobehavior.* New York: Norton.

[41] Simmel, G. (1955). *Conflict and the Web of Group Affiliations.* New York: The Free Press.

[42] Simon, Herbert Alexander. (1982). *Models of Bounded Rationality.* Boston: MIT Press.

[43] Skinner, B. F. (1974). *About Behaviorism.* New York: Random House, Inc.

[44] Thompson, Richard W. (1975.) "Gratification Patterns in Buganda: An Explanation of Intra-Cultural Diversity." *American Ethnologist 2* (1): 193-206.

[45] Tindall, B. Allan. (1976). "Theory in the Study of Cultural Transmission." *Annual Review of Anthropology* 5:195-208.

[46] Weick, K. E. (1969). *The Social Psychology of Organizing.* Reading, MA: Addison-Wesley.

[47] Whiting, Beatrice B., ed. (1963) *Six Cultures: Studies of Child Rearing.* New York: John Wiley.

[48] Young, Peyton (1998) *Individual Strategy and Social Structure: An Evolutionary Theory of Institutions.* Princeton University Press.