14.773 Political Economy of Institutions and Development.
Lecture 13: Culture, Values and Cooperation

Daron Acemoglu

MIT

March 21, 2013.

Introduction

- How do we model the effects of culture and values on social and political outcomes—and through which mechanisms?
- Why do these values persist?
- How they interact with institutions?
- In this lecture, an overview of some related research.

Intergenerational Transmission: Basic Models

- Cavalli-Sforza and Feldman (1981) and to Boyd and Richerson (1985), based on models of evolutionary biology applied to the transmission of cultural traits.
- Suppose that there is a dichotomous cultural trait in the population, $\{a, b\}$. Let the fraction of individuals with trait $i \in \{a, b\}$ be q^i.
- Focus on a continuous time model with "a-sexual" reproduction where each parent has one child at the rate λ and is replaced by the child.
- Two types of cultural transmission:
  1. *direct/vertical* (parental) socialization and
  2. *horizontal/socialization* by the society at large.

Intergenerational Transmission (continued)

- Suppose that direct vertical socialization of the parent's trait, say i, occurs with probability d^i.
- Then, if a child from a family with trait i is not directly socialized, which occurs with probability $1 - d^i$, he/she is horizontally/obliquely socialized by picking the trait of a role model chosen randomly in the population (i.e., he/she picks trait i with probability q^i and trait $j \neq i$ with probability $q^j = 1 - q^i$).
- Therefore, the probability that a child from family with trait i is socialized to have trait j, P^{ij}, is:

$$\begin{aligned} P^{ii} &= d^i + (1 - d^i) q^i \\ P^{ij} &= (1 - d^i)(1 - q^i). \end{aligned} \tag{1}$$

Intergenerational Transmission (continued)

- Now noting that each child replaces their parent in the population (at the rate λ), we have that

$$\dot{q}^i = \lambda \left[\left(d^i + (1 - d^i) q^i \right) q^i + (1 - d^j) q^i \left(1 - q^i \right) \right] - \lambda q^i .$$

- Simplifying this equation, we obtain:

$$\dot{q}^i = \lambda q^i (1 - q^i) \left(d^i - d^j \right) . \tag{2}$$

- This is a version of the replicator dynamics in evolutionary biology for a two-trait population dynamic model—i.e., a logistic differential equation.
- If $\left(d^i - d^j \right) > 0$ cultural transmission represents a selection mechanism in favor of trait i, due to its differential vertical socialization.
- However, this selection mechanism implies that there will not be cultural heterogeneity, i.e., a steady-state with $0 < q^{i*} < 1$.

Intergenerational Transmission (continued)

- The following result is now immediate.
- Let $q^i(t, q_0^i)$ denotes the fraction with trait q^i at time t starting with initial condition q_0^i. Then:

Proposition

Suppose $d^i > d^j$. Then, steady states are culturally homogeneous. Moreover, for any $q_0^i \in (0, 1]$, $q^i(t, q_0^i) \to 1$. If instead $d^i = d^j$, then $q^i(t, q_0^i) = q_0^i$, for any $t \geq 0$.

Intergenerational Transmission: Bisin-Verdier Model

- Bisin and Verdier (2000, 2001) introduce "imperfect empathy" into this framework, whereby parents look at the world with their own preferences and thus wants to socialize their offspring according to their preferences.
- Formally, suppose that individuals choose an action $x \in X$ to maximize a utility function $u^i(x)$, which is a function of the cultural trait $i \in \{a, b\}$. Suppose that this utility function is strictly quasi-concave.

Intergenerational Transmission (continued)

- Let V^{ij} denote the utility of a type i parent of a type j child, $i, j \in \{a, b\}$. Then clearly, we have

$$V^{ij} = u^i(x^j)$$

And

$$x^j = \arg\max_{x \in X} \; u^j(x)$$

- This implies the "imperfect empathy" feature:

$$V^{ii} \geq V^{ij}$$

holding with $>$ for generic preferences (i.e., in particular when the maximizers for the two types are different).

Intergenerational Transmission (continued)

- Suppose also that parents have to exert costly effort in order to socialize their children. In particular, parents of type i choose some variable τ^i, which determines

$$d^i = D\left(q^i, \tau^i\right).$$

  The dependence on q captures other sources of direct transmission working from the distribution of traits in the population.
- The cost of τ^i is assumed to be $C\left(\tau^i\right)$.
- Suppose that D is continuous, strictly increasing and strictly concave in τ^i, and satisfies $D\left(q^i, 0\right) = 0$, and C is also continuous, strictly increasing and convex. Moreover, suppose also that $D\left(q^i, \tau^i\right)$ is nonincreasing in q^i.
- Parents of type i will solve the following problem:

$$\max_{\tau^i} -C(\tau^i) + P^{ii}V^{ii} + P^{ij}V^{ij},$$

  where P^{ii} and P^{ij} depend on τ^i via d^i.

Intergenerational Transmission (continued)

- Let us say that the *cultural substitution property* holds if the solution to this problem d^{i*} is is strictly decreasing function of q^i and takes a value $d^{i*} = 0$ at $q^i = 1$. Intuitively, this implies that parents have less incentives to socialize their children when their trait is more popular/dominant in the population.
- This cultural substitution property is satisfied in this model.
- Then, the dynamics of cultural transmission can be more generally written as

$$\dot{q}^i = \lambda q^i (1 - q^i) \left(d^i \left(q^i \right) - d^j \left(1 - q^i \right) \right). \tag{3}$$

- We can also verify that this differential equation has a unique interior steady state, q^{i*}, and moreover,

Proposition

The steady states are now culturally heterogeneous. In particular, $q^i(t, q_0^i) \rightarrow q^{i*}$, *for any* $q_0^i \in (0, 1)$.

Intergenerational Transmission (continued)

- Intuition: the cultural substitution property implies that parents put more effort in socializing their children, i.e., passing on their traits, when their traits are less common in the cooperation.
- The proof of this result follows from the following observations:
  1. Clearly, an interior steady state satisfies
  $$d^i\left(q^i\right) - d^j\left(1 - q^i\right) = 0,$$
  and since both d^i and d^j are strictly decreasing, there can at most be one such steady state q^{i*}.
  2. Moreover, since $d^i\left(1\right) = 0$, existence is guaranteed.
  3. Global stability then follows from the fact that this pattern implies that $\dot{q}^i > 0$ whenever $q^i \in \left(0, q^{i*}\right)$ and at $\dot{q}^i < 0$ whenever $q^i \in \left(q^{i*}, 1\right)$.

Culture, Values and Cooperation

- Tabellini (2009) considers the following variation on the static prisoners' dilemma game.
- Individuals incur a negative disutility from defecting, but the extent of this disutility depends on how far their partner is according to some distance metric.
  - The most interesting interpretations of this distance are related to "cultural distance" or "kinship distance". For example, some individuals may not receive any disutility from defecting on strangers, but not on cousins.
  - This captures notions related to "generalized trust".

Model

- A continuum of one-period lived individuals, with measure normalized to 1, is uniformly distributed on the circumference of a circle of size $2S$, so that the maximum distance between two individuals is S.
- A higher S implies a more "heterogeneous" society—in geography, ethnicity, religion or other cultural traits.
- Each individual is (uniformly) randomly matched with another located at distance y with probability $g(y) > 0$, and naturally

$$\int_0^S g(y) = 1.$$

Model (continued)

- A matched pair play the following prisoners' dilemma:

	C	D
C	c, c	$h - l, c + w$
D	$c + w, h - l$	h, h

- Naturally, $c > h$ and $l, w > 0$. Let us also suppose that $l \geq w$, so that the loss of being defected when playing cooperate is no less than the reverse benefit.

Model (continued)

- In addition, each individual enjoys a non-economic (psychological or moral) benefit
$$de^{-\theta y}$$
whenever she plays "cooperate" (regardless of what her opponent plays) but as a function of the distance between herself and the other player, y, with the benefit declining exponentially in distance.
- Let us assume that
$$d > \max\{l, w\},$$
which ensures that this benefit is sufficient to induce cooperation with people very close.

Model (continued)

- Finally, suppose that there are two types of player indexed by $k = 0, 1$, "good" and "bad," modeled as having different rates at which the benefit from cooperation declines. In particular,

$$\theta^0 > \theta^1.$$

- This captures the idea that what varies across individuals (and perhaps across societies) is the level of "generalized trust".
- The fraction of good $(k = 1)$ types in the population is the same at any point in the circle is $1 > n > 0$.

Equilibrium

- Consider a player in a match of distance y.
- Let $\pi(y)$ denote the probability that her opponent will play C.
- We can express the player's net expected *material* gain from defecting instead of laying C as:

$$T(\pi(y)) = [l - \pi(y)(l - w)] > 0 \quad (4)$$

- This is strictly positive, as it is always better not to cooperate given the prisoners' dilemma nature of the game.
- Note also that cooperation decisions are strategic complements, since, given the assumption that $l \geq w$, the function $T(\pi(y))$ is non-increasing in $\pi(y)$

Equilibrium (continued)

- The temptation to defect will be potentially balanced by the non-economic benefit of cooperation, $de^{-\theta^k y}$, as a function of a player's type.
- To simplify the analysis, let us suppose that

$$\frac{\theta^0}{\theta^1} > \frac{\ln(l/d)}{\ln(w/d)} \tag{A0}$$

  and also focus on "best" (Pareto superior) and symmetric (independent of location on the circle) equilibria.
- Then a player of type $k = 0, 1$ will be indifferent between cooperating and not cooperating with a partner of distance $\tilde{y}^k$ defined as

$$T(\pi(\tilde{y}^k)) = de^{-\theta^k \tilde{y}^k}, \tag{5}$$

  Or as

$$\tilde{y}^k = \left\{ \ln d - \ln \left[(w - l)\, \pi(\tilde{y}^k) + l \right] \right\} / \theta^k. \tag{6}$$

Equilibrium (continued)

- Thus given the equilibrium probability of cooperation $\pi(y)$ (for all y), each individual will cooperate with players closer than $\tilde{y}^k$ ($y < \tilde{y}^k$) and defect against those farther than $\tilde{y}^k$ as a function of her type k.
- Note that if $l > w$, then the right hand side of (6) is increasing in $\pi(y)$, and there are multiple equilibria, though we are ignoring this by focusing on best equilibria.
- Now consider a bad player, $k = 0$, and suppose that she/he expects the opponent always to cooperate, so that $\pi(y) = 1$ (which will be true, since both types of players will cooperate whenever this player is choosing to operate along the equilibrium path).
- Then (6) reduces to:

$$Y^0 = [\ln d - \ln w] / \theta^0, \tag{7}$$

and player $k = 0$ will cooperate up to distance $y \leq Y^0$.

Equilibrium (continued)

- The problem of a good player is a little more complicated.
- She will necessarily cooperate up to distance $y \leq Y^0$. But beyond that, she recognizes that only other good players will cooperate, and thus $\pi(y) = n$.
- Using this with (6)

$$Y^1 = [\ln d - \ln [(w - l)\, n + l]] \,/\theta^1. \tag{8}$$

- And with players cooperate up to Y^1 (which is strictly greater than Y^0 given the assumption above).

Equilibrium (continued)

- Thus summarizing:

Proposition

In the Pareto superior symmetric equilibrium, a player of type k cooperates in a match of distance $y \leq Y^k$ and does not cooperate if $y > Y^k$, where Y^k is given (7)-(8), for $k = 0, 1$.

- This proposition captures, in a simple way, the role of "generalized trust" in society.
- It also highlights the strategic complementarity in trust, as Y^1 is increasing in n: thus good players trust others more when there are more good players. Interestingly, this does not affect bad types, given the simple structure of the prisoners' dilemma game coupled with the assumption that $l \geq w$.

Endogenous Values

- Values can now be endogenized using the same approach as Bisin and Verdier.
- Parents choose socialization effort τ at cost

$$\frac{1}{2\varphi}\tau^{2},$$

  and as a result, their offspring will be over the "good type," i.e., $\theta^k = \theta^1$, with probability $\delta + \tau$.
- As in Bisin and Verdier, they evaluate this with their own preferences, i.e., there is "imperfect empathy".

Endogenous Values (continued)

- Let V_t^{pk} denote the parent of type p's evaluation of their kid of type k's overall expected utility in the equilibrium of the matching game.
- Since the probability of a match with someone located at distance z is denoted $g(z)$, we have

$$V_t^{pk} = U_t^k + d \int_0^{Y_t^k} e^{-\theta^p z} g(z) dz, \tag{9}$$

where $U_t^k = U(\theta^k, n_t)$ denotes the expected equilibrium material payoffs of a kid of type k, in a game with a fraction n_t of good players. The integral gives the parent's evaluation of their kid's expected non-economic benefit from their offspring's cooperating in matches of distance smaller than Y_t^k.

- This is where imperfect empathy comes in, as this integral term uses the parent's value parameter, θ^p, rather than with the kid's value.

Endogenous Values (continued)

- With the same argument as in Bisin and Verdier, we have that whenever $k \neq p$, then
$$V_t^{pp} > V_t^{pk}$$
where recall that, given the assumptions, $Y^1 > Y^0$.
- The fact that parents of bad type, according to their values, have nothing to gain from exerting effort to socialize their children to be good (as they do not internalize the "moral" benefit from cooperation with farther away partners), and the fact that the marginal cost of exerting effort at zero is zero, implies the following simple result:

Proposition

A "good" parent $(p = 1)$ exerts strictly positive effort $\tau_t > 0$. A "bad" parent $(p = 0)$ exerts no effort.

Endogenous Values (continued)

- Therefore, the law of motion of types in the population follows the following difference equation:

$$n_t = n_{t-1}(\delta + \tau_t) + (1 - n_{t-1})\delta = \delta + n_{t-1}\tau_t. \tag{10}$$

- It can also be shown that the optimal level of effort for with type parents is

$$\tau_t = F\left(Y_t^1\right) \equiv \varphi d[-e^{-\theta^1 Y_t^1} + E[e^{-\theta^1 y} \mid Y_t^1 \geq y \geq Y^0]] \Pr(Y_t^1 \geq y \geq \tag{11}$$

where intuitively the benefit to good parents depends on the likelihood that their children will play against an opponent of good type, again highlighting the strategic complementarities. The right-hand side of (11), $F\left(Y_t\right)$, is as a result strictly increasing in Y_t^1.

Endogenous Values (continued)

- This means that (10) can be written as

$$n_t = \delta + n_{t-1} F(Y_t^1) \equiv N(Y_t^1, n_{t-1}), \tag{12}$$

with the date t equilibria value of Y_t^1 being defined as:

$$Y_t^1 = [\ln d - \ln [(w-l) n_t + l]] / \theta^1 \equiv Y(n_t).$$

- Now using the fact that n_t itself is a function of n_{t-1} and Y_t^1 from (10), we can express endogenous value dynamics as in two equations system:

$$Y_t^{1*} = G^Y(n_{t-1}) \tag{13}$$

$$n_t^* = G^n(n_{t-1}) \tag{14}$$

- Strategic complementarities now imply multiple steady state are possible.

Endogenous Values (continued)

- Naturally, additional conditions ensure uniqueness. One such condition would be

$$\frac{1}{\varphi} > l - w \tag{A1}$$

which ensures that the marginal cost of effort, $1/\varphi$, rises sufficiently rapidly, relative to the strategic complementarity captured by $(l - w)$.

- Given uniqueness, global stability of dynamics can also be ensured. The following proposition gives one sufficient condition

Proposition

Suppose (A1) holds and $\varphi > 0$ is sufficiently small. Then the equilibrium is unique and is globally stable, i.e., it asymptotically reaches the unique steady state (Y_s^{1*}, n_s^*). Moreover, adjustment to steady state is monotone, i.e., the fraction of what types, n_t^*, and the cooperation threshold, Y_t^{1*}, andmonotonically increase or decrease along the adjustment path.

Effects of Institutions

- Let us introduce institutional enforcement of cooperation simply by assuming that there is a probability $\chi(y)$ that defection gets detected when it takes place in a match of distance y and it gets punished.
- We can think of different types of shifts up the schedule $\chi(y)$ as corresponding to different types of changes in institutions.
- In particular, we can imagine that χ increases for high y. This will encourage more broad-based cooperation and it will also incentivize parents to socialize their children to be of the "good" type. As a result, both n_t^* and Y_t^{1*} will increase.
- At the other extreme, we can think of an improvement in local enforcement, with no change in enforcement for faraway matches. This could be considered as a family- or clan-based enforcement, or what the Mafia achieves in southern Italy. This would increase Y^0, so its static effect is good. However, it would also reduce the parental efforts for good socialization, so ultimately it would reduce n_t^* and Y_t^{1*}.

Endogenous Institutions

- One could also endogenize enforcement through a voting or political economy process.
- In this case, one can obtain richer dynamics, where parental socialization interacts with political economy. For example, more with types today leads to greater enforcement, which then encourages more would socialization.
- Multiple steady states are again possible, this time resulting from the interaction of culture and institutions.

Modeling Dynamics of Cooperation

- Different focus: how does "cooperation" (or "solution to collective action problem") emerge, and why does "history" affect the outcome of such cooperation games? Could this be an important "mechanism of persistence"?
  - Why does a history of distrust leads to distrust? How do we understand "social norms" and why do they persist?
  - Why does a society sometime break out of a history of distrust and change social norms?
  - Why does "collective action" differ across societies and why does it seem to change abruptly from time to time?
  - What is the role of leadership and "prominence"?
- Simple model based on Acemoglu and Jackson (2011).

Model

- Consider an overlapping-generations model where agents live for two periods. We suppose for simplicity that there is a single agent in each period (generation), and each agent's payoffs are determined by his interaction with agents from the two neighboring generations (older and younger agents).
- The action played by the agent born in period t is denoted $A_t \in \{H, L\}$, corresponding to "High" and "Low" actions (also can be interpreted as "honest" and "dishonest" actions).
- An agent chooses an action only once, in the first period of his or her life and that is played in both periods. This can be thought of as a proxy for a case where there is discretion, but also a high cost of changing behavior later in life.

Model

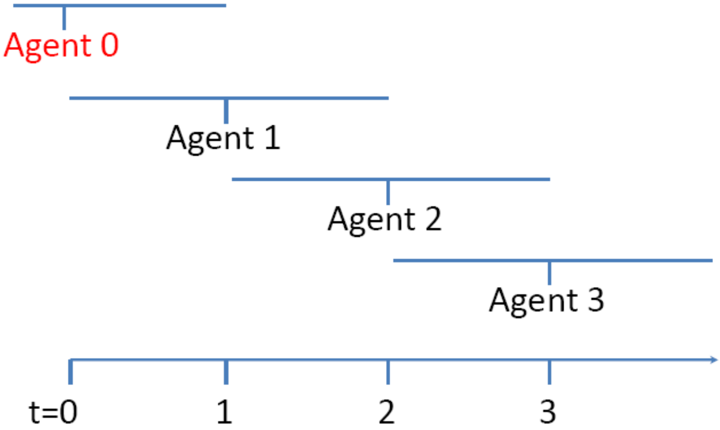
- The stage payoff to an agent playing A when another agent plays A' is denoted $u(A, A')$. The total payoff to the agent born at time t is

$$(1 - \lambda)\, u(A_t, A_{t-1}) + \lambda u(A_t, A_{t+1}), \tag{15}$$

where A_{t-1} designates the action of the agent in the previous generation and A_{t+1} is the action of the agent in the next generation.

- Implicit assumption: choose single "pattern of behavior" A_t against both generations
    - $\lambda \in [0, 1]$ is a measure of how much an agent weighs the play with the next generation compared to the previous generation.
    - When $\lambda = 1$ an agent cares only about the next generation's behavior, while when $\lambda = 0$ an agent cares only about the previous generation's actions. We do not explicitly include a discount factor, since it is subsumed by λ.

Demographics



Model (continued)

- The stage payoff function $u(A, A')$ is given by the following matrix:

	H	L
H	β, β	$-\alpha, 0$
L	$0, -\alpha$	$0, 0$

  where β and α are both positive.
- This payoff matrix captures the notion that, from the static point of view, both honesty and dishonesty could arise as social norms—i.e., both (H, H) and (L, L) are static equilibria given this payoff matrix. (H, H) is clearly the Pareto optimal equilibrium, and depending on the values of β and α, it may be the risk dominant equilibrium as well.

Endogenous and Exogenous Agents

- There are four types of agents in this society.
- First, agents are distinguished by whether they choose an action to maximize the utility function given in (15). We refer to those who do so as "endogenous" agents.
- In addition to these endogenous agents, who will choose their behavior given their information and expectations, there will also be some committed or "exogenous" agents, who will choose an exogenously given action.
    - This might be due to some irrationality, or because some agents have a different utility function.

Endogenous and Exogenous Agents (continued)

- Any given agent is an "exogenous type" with probability 2π (independently of all past events), exogenously committed to playing each of the two actions, H and D, with probability $\pi \in (0, \frac{1}{2})$, and think of π as small.
- With the complementary probability, $1 - 2\pi > 0$, the agent is "endogenous" and chooses whether to play H or D, when young and is stuck with the same decision when old.
- Any given agent is also "prominent" with probability q (again independent). Information about prominent agents will be different. Thus:

	non-prominent	prominent
endogenous	$(1-2\pi)(1-q)$	$(1-2\pi)q$
exogenous	$2\pi(1-q)$	$2\pi q$

- Let us refer to endogenous non-prominent agents as *regular.*

Towards Equilibrium Behavior

- Let ϕ_{t-1}^{t} be the the probability that the agent of generation t assigns to the agent from generation $t-1$ choosing $A = H$
- Let ϕ_{t+1}^{t} be the probability that the agent of generation t assigns, conditional on herself playing $A = H$, to the agent from generation $t+1$ choosing $A = H$.
- Payoff from L: 0
- Payoff from H: $(1-\lambda)\left[\phi_{t-1}^{t}\beta - (1-\phi_{t-1}^{t})\alpha\right] + \lambda\left[\phi_{t+1}^{t}\beta - (1-\phi_{t+1}^{t})\alpha\right]$.
- Then an endogenous agent of generation t will prefer to play $A = H$ only if
$$(1-\lambda)\,\phi_{t-1}^{t} + \lambda\phi_{t+1}^{t} \geq \frac{\alpha}{\beta+\alpha} \equiv \gamma. \qquad (16)$$
- Parameter γ a convenient way of summarizing relative payoffs (and also "basin of attraction" of L; so the greater is γ, the more attractive it is $A = L$).

Signals, Information and Prominent Agents

- A noisy signal of an action taken by a non-prominent agent of generation t is observed by the agent in generation $t+1$.
- No other agent receives any information about this action.
- In contrast, the actions taken by prominent agents are perfectly observed by all future generations.

Information Structure

- Let h^{t-1} denotes the public history at time t, which includes a list of past prominent agents and their actions up to and including time $t-1$. We denote the set of h^{t-1} histories by $\mathcal{H}^{t-1}$.
- We write $h_t = (T, a)$ if at time t the agent is both prominence type $T \in \{P, N\}$ and has taken action $a \in \{H, L\}$ if $T = P$ (if $T = N$, his action is not part of the public history).
- In addition to observing $h^{t-1} \in \mathcal{H}^{t-1}$, an agent of generation t, when born, receives a signal $s_t \in [0, 1]$ about the behavior of the agent of the previous generation (where the restriction to $[0, 1]$ is without loss of any generality). This signal has a continuous and distribution described by a density function $f_H(s)$ if $A_{t-1} = H$ and $f_L(s)$ if $A_{t-1} = L$.

Information Structure (continued)

- Without loss of generality, we order signals such that higher s has a higher likelihood ratio for H; i.e., so that

$$g(s) \equiv \frac{f_L(s)}{f_H(s)}$$

  is nonincreasing in s.
- Suppose also that it is strictly decreasing, so that we have *strict Monotone Likelihood Ratio Property (MLRP)* everywhere.
- Suppose, without loss of any generality, that $s \in [0, 1]$, so that 0 is the worst signal for past H and 1 best signal for past H.
- Let $\Phi(x, s)$ denote the posterior probability that $A_{t-1} = H$ given $s_t = s$ under the belief that an endogenous agent of generation $t-1$ plays H with probability x. This is:

$$\Phi(x, s) \equiv \frac{f_H(s)\,x}{f_H(s)\,x + f_L(s)\,(1-x)} = \frac{1}{1 + g(s)\,\frac{1-x}{x}}. \qquad (17)$$

  We assume that the game begins with a prominent agent at time

Strategies

- Let us use N to denote regular agents and P to denote prominent agents.
- With this notation, we can write the strategy of an endogenous agent of generation t (who may or may not be regular) as:

$$\sigma_t : \mathcal{H}^{t-1} \times [0,1] \times \{P,N\} \rightarrow [0,1],$$

written as $\sigma_t(h^{t-1}, s, T)$ where $h^{t-1} \in \mathcal{H}^{t-1}$ is the public history of play, $s \in [0,1]$ is the signal of the previous generation's action, and $T \in \{P,N\}$ denotes whether or not the current agent is prominent.

- The number $\sigma_t(s, h^t, T)$ corresponds to the probability that the agent of generation t plays H.

We denote the strategy profile of all agents by the sequence

$$\sigma = \left(\sigma_1(h^0, s, T), \sigma_2(h^1, s, T), ..., \sigma_t(h^t, s, T), ...\right).$$

Semi-Markovian Strategies

- For the focus here, the most relevant equilibria involve agents ignoring histories that come before the last prominent agent (in particular, it will be apparent that these histories are not payoff-relevant provided others are following similar strategies).
- Let us refer to these as *semi-Markovian* strategies.
- Semi-Markovian strategies are specified for endogenous agents as functions $\sigma_{\tau}^{SM} : \{H, D\} \times [0, 1] \times \{P, N\} \rightarrow [0, 1]$, written as $\sigma_{\tau}^{SM}(a, s, T)$ where $\tau \in \{1, 2, \ldots\}$ is the number of periods since the last prominent agent, $a \in \{H, D\}$ is the action of the last prominent agent, $s \in [0, 1]$ is the signal of the previous generation's action, and again $T \in \{P, N\}$ is whether or not the current agent is prominent.
- Let us denote a semi-Markovian by the sequence $\sigma^{SM} = \left(\sigma_1^{SM}(a, s, T), \sigma_2^{SM}(a, s, T), ..., \sigma_t^{SM}(a, s, T), ...\right)$.
- With some abuse of notation, write $\sigma_t = H$ or D to denote a strategy (or a semi-Markovian strategy) that corresponds to playing honest (dishonest) with probability one.

Equilibrium Definition

- Perfect Bayesian Equilibrium or Sequential Equilibrium.
- Only need to be careful when $q = 0$.
- Define *greatest* and *least* equilibria, and focus on greatest equilibria.

Cutoff Strategies

- We say that a strategy σ is a *cutoff strategy* if for each t, h^{t-1} such that $h_{t-1} = N$ and $T_t \in \{P, N\}$, there exists s_t^* such that $\sigma_t(h^t, s, T_t) = 1$ if $s > s_t^*$ and $\sigma_t(h^t, s, T_t) = 0$ if $s < s_t^*$.
  - Clearly, setting $\sigma_t(h^t, s, T) = 1$ (or 0) for all s is a special case of a cutoff strategy.
- Cutoff strategy profile can be represented by the sequence of cutoffs

$$c = \left(c_1^N(h_0), c_1^P(h_0), \ldots c_t^N(h_{t-1}), c_t^P(h_{t-1}), \ldots \right).$$

- Given strict MLRP, all equilibria will be in cutoff strategies.
- Define greatest equilibria using the Euclidean distance on cutoffs.

Equilibrium Characterization

Proposition

1. *All equilibria are in cutoff strategies.*
2. *There exists an equilibrium in semi-Makovian cutoff strategies.*
3. *The set of equilibria and the set of semi-Markovian equilibria form complete lattices, and the greatest (and least) equilibria of the two lattices coincide.*

Understanding History-Driven Behavior

- Look for a unique equilibrium given by history:
  - When following prominent H, will all endogenous agents play H?
  - When following prominent L, will all endogenous agents play L?
- In such an equilibrium, social norms of High and Low emerge and persist, but not forever, since there might be switches because of exogenous prominent agents.
- Related question: when is this the greatest equilibrium?

Understanding History-Driven Behavior (continued)

- Recall that an endogenous agent of generation t will prefer to play $A = H$ only if

$$(1-\lambda)\,\phi_{t-1}^{t} + \lambda\phi_{t+1}^{t} \geq \frac{\alpha}{\beta+\alpha} \equiv \gamma. \tag{18}$$

- H is a *unique* best response for all if

$$\begin{aligned} (1-\lambda)\,\phi_{t-1}^{t} + \lambda\phi_{t+1}^{t} &\geq \gamma \\ \gamma_H^* \equiv (1-\lambda)\,\Phi\,(1-\pi, 0) + \lambda\pi &\geq \gamma. \end{aligned}$$

- L is a *unique* best response for all if

$$\begin{aligned} (1-\lambda)\,\phi_{t-1}^{t} + \lambda\phi_{t+1}^{t} &< \gamma \\ \gamma_L^* \equiv (1-\lambda)\,\Phi\,(\pi, 1) + \lambda(1-\pi) &< \gamma. \end{aligned}$$

Understanding History-Driven Behavior (continued)

Proposition

1. *If $\gamma < \gamma_H^*$, then following $a = H$ at date $t = 0$, the unique continuation equilibrium involves all (prominent and non-prominent) endogenous agents playing H.*
2. *If $\gamma > \gamma_L^*$, then following $a = L$ at date $t = 0$, the unique continuation equilibrium involves all endogenous agents playing L.*
3. *If $\gamma_L^* < \gamma < \gamma_H^*$, then there is a unique equilibrium driven by the starting condition: all endogenous agents take the same action as the action of the prominent agent at date $t = 0$.*

- Interpretation: persistent, but not everlasting, social norms.

Understanding History-Driven Behavior (continued)

- The condition that $\gamma_L^* < \gamma < \gamma_H^*$ boils down to

$$\lambda(1-2\pi) < (1-\lambda)\left[\Phi(1-\pi,0) - \Phi(\pi,1)\right]. \tag{19}$$

- It requires that λ be sufficiently small, so that sufficient weight is placed on the past. Without this, behavior would coordinate with future play, which naturally leads to a multiplicity.
- It also requires that signals are not too strong (so $\Phi(1-\pi,0) - \Phi(\pi,1) > 0$), as otherwise players would react to information about the most recent past generation and could change to High behavior if they had a strong enough signal regarding the past play and would also expect the next generation to have good information.

Understanding History-Driven Behavior (continued)

- Focusing on the greatest equilibrium:
- Let

$$\overline{\gamma}_H \equiv (1-\lambda)\,\Phi(1-\pi,0) + \lambda\,(1-\pi)\,. \tag{20}$$

- Thus relative to γ_H^*, more "optimistic" expectations about the future.

Proposition

The greatest equilibrium is such that:

(i) *following a prominent play of L, there is a low social norm and all endogenous agents play L if and only if* $\overline{\gamma}_L < \gamma$*; and*

(ii) *following a prominent play of H, there is a high social norm and all endogenous agents play H if and only if* $\gamma \leq \overline{\gamma}_H$.

Thus, endogenous players always follow the play of the most recent prominent player in the greatest equilibrium if and only if $\overline{\gamma}_L < \gamma \leq \overline{\gamma}_H$.

General Characterization of Greatest Equilibrium

- Let
$$\widehat{\gamma}_H \equiv (1-\lambda)\,\Phi(1-\pi,1)+\lambda\,(1-\pi)\,.$$
- This is the expectation of $(1-\lambda)\phi_{t-1}^t+\lambda\phi_{t+1}^t$ for an agent who believes that any regular agent preceding him or her played H and sees the most optimistic signal, and believes that all subsequent endogenous agents will play H.
- Above, this threshold, no regular agent would ever play H.

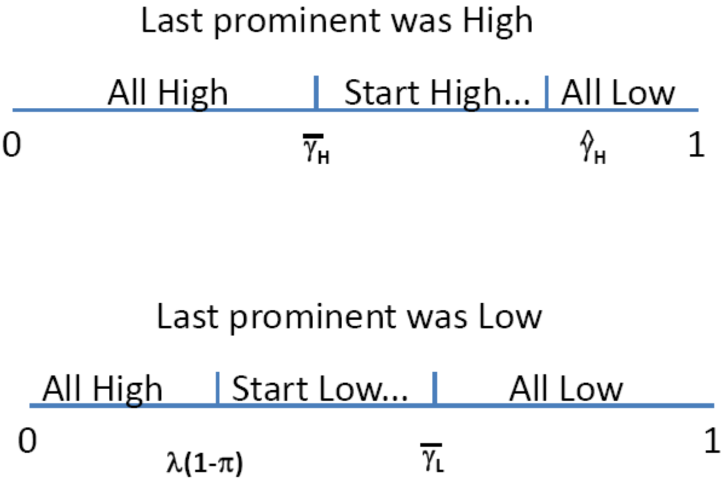
General Characterization (continued)

Proposition

In the greatest equilibrium:

1. *If $\gamma \leq \lambda(1-\pi)$, then all endogenous agents play H in all circumstances, and thus society has a stable high behavioral pattern.*
2. *If $\lambda(1-\pi) < \gamma \leq \overline{\gamma}_H$, then following a prominent play of H (but not following the prominent play of L) all endogenous agents play H.*
3. *If $\overline{\gamma}_L < \gamma \leq \overline{\gamma}_H$, then following a prominent play of L, all endogenous agents play L, and so all endogenous players follow the play of the most recent exogenous prominent player.*
4. *If $\overline{\gamma}_H < \gamma$, then endogenous agents play L for at least some signals, periods, and types even following a prominent play of H.*
5. *If $\widehat{\gamma}_H < \gamma$, then all endogenous agents who do not immediately follow a prominent H play L regardless of signals or types.*

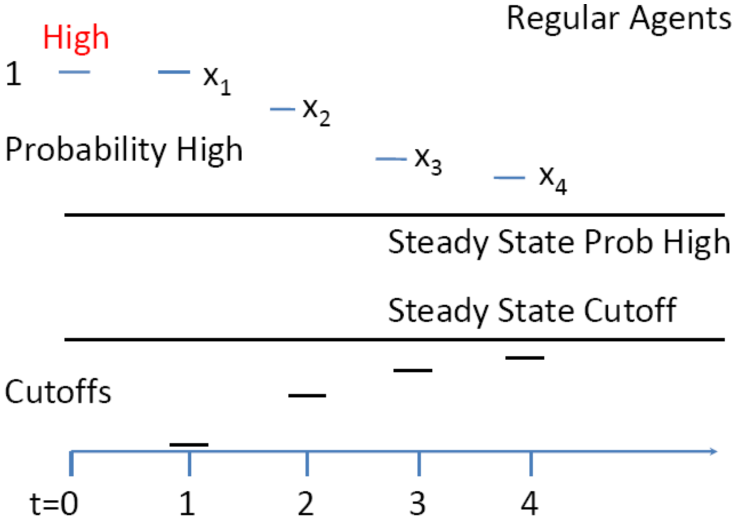
General Characterization of Greatest Equilibrium (continued)



Reversion of Play

- What happens when all High and all Low are not stable social norms?
- *Answer:* play reverts from an *extreme* (started by a prominent agent) to a steady-state distribution.
  - Start with H
  - Next player knows previous is H with probability 1
  - Next player knows previous endogenous played H, but this has probability $1 - \pi$, so *action made depend on signal*
    - In fact, even stronger, because she knows that her signals will be interpreted is not necessarily coming from H.
  - Next player knows previous play was H with probability $< 1 - \pi$.

Reversion of Play (continued)



Reversion of Play (continued)

- Let us denote the cutoffs used by prominent and non-prominent agents τ periods after the last prominent agent by c_τ^P and c_τ^N respectively.
- We say that high play *decreases* over time if $(c_\tau^P, c_\tau^N) \leq (c_{\tau+1}^P, c_{\tau+1}^N)$ for each τ.
- We say that high play *strictly decreases* over time, if in addition, we have that when $(c_\tau^P, c_\tau^N) \neq (0, 0)$, $(c_\tau^P, c_\tau^N) \neq (c_{\tau+1}^P, c_{\tau+1}^N)$.
- The concepts of high play increasing and strictly increasing are defined analogously.

Reversion of Play (continued)

Proposition

1. *In the greatest and least equilibria, cutoff sequences* $\left(c_\tau^P, c_\tau^N\right)$ *are monotone. Thus, following a prominent agent choosing* H, $\left(c_\tau^P, c_\tau^N\right)$ *are nondecreasing and following a prominent agent choosing* L*, they are non-increasing.*
2. *If* $\gamma > \overline{\gamma}_H$*, then in the greatest equilibrium, high play strictly decreases over time following high play by a prominent agent.*
3. *If* $\gamma < \overline{\gamma}_L$*, then in the greatest equilibrium, high play strictly increases over time following low play by a prominent agent.*

- But important asymmetry from switching from L to H vs from H to L
  - As we will see next, endogenous prominent agents would not like to the latter, but would prefer to do the former, so the first type of switches are driven by both exogenous and endogenous prominent agents, while the second only by exogenous prominent agents.

Reversion of Play (continued)

- The following is an immediate corollary of Proposition 10.

Corollary

As the distance from the last prominent agent grows ($\tau \to \infty$), cutoffs in the greatest equilibrium converge and the corresponding distributions of play converge to a stationary distribution. Following a choice of H by the last prominent agent, this limiting distribution involves only H by all endogenous agents if and only if $\gamma \leq \overline{\gamma}_H$. Similarly, following a choice of L by the last prominent agent, this limiting distribution involves L by all endogenous agents if and only if $\gamma \geq \overline{\gamma}_L$.

Leadership: Breaking the Low Social Norm

- Can promise breaker social norm of L?
  - Regular agents may be stuck in L for reasons analyzed so far.
  - But prominence, greater visibility in the future, can enable "leadership"
- Idea:
  - endogenous prominent agents can always break the social norm
  - when "all L" is not the unique equilibrium after prominent L, endogenous prominent agents will like to break the social norm of L and start a switch to H

Leadership: Breaking the Low Social Norm (continued)

Proposition

1. *Suppose that the last prominent agent played L and*

$$\tilde{\gamma}_L \leq \gamma < \tilde{\gamma}_H \equiv (1-\lambda)\Phi(\pi, 0) + \lambda(1-\pi). \tag{21}$$

Then there exists a fixed cutoff below 1 (after at least one period) such that an endogenous prominent agent chooses High and breaks the Low social norm if the signal is above the cutoff.

2. *Suppose that $\gamma < \tilde{\gamma}_L$ and $\gamma < \tilde{\gamma}_H$. Suppose that the last prominent agent has played L. Endogenous prominent agents have cutoffs below 1 that decrease with time such that if the signal is above the cutoff then in a greatest equilibrium the endogenous prominent agent will choose H and break a low social norm.*

3. *Moreover, in either case if $\gamma < \gamma_H^*$, the above are the unique continuation equilibrium.*

Role of Prominence

- Prominence provides greater visibility and thus coordinates future actions.
- Crucially: common knowledge of visibility.
  - Without this, prominence is less effective.

Role of Prominence (continued)

- Suppose there is a starting non-prominent agent at time 0 who plays *High* with probability $x_0 \in (0, 1)$, where x_0 is known to all agents who follow, and generates a signal for the first agent in the usual way.
- All agents after time 1 are not prominent.
- In every case all agents (including time 1 agents) are endogenous with probability $(1 - 2\pi)$.

Scenario 1: The agent at time 1 is not prominent and his or her action is observed with the usual signal structure.
Scenario 2: The agent at time 1's action is observed perfectly by the period 2 agent, but not by future agents.
Scenario 2′: The agent at time 1 is only observed by the next agent according to a signal, but then is subsequently perfectly observed by all agents who follow from time 3 onwards.
Scenario 3: The agent at time 1 is prominent, and all later agents are viewed with the usual signal structure.

Role of Prominence (continued)

- Clearly, as we move from Scenario 1 to Scenario 2 (or $2'$) to Scenario 3, we are moving from a non-prominent agent to a prominent one
- Let us focus again on the greatest equilibrium and let $c^k(\lambda, \gamma, f_H, f_L, q, \pi)$ denote the cutoff signal above which the first agent (if endogenous) plays *High* under scenario k as a function of the underlying setting.

Proposition

The cutoffs satisfy $c^2(\cdot) \geq c^3(\cdot)$ and $c^1(\cdot) \geq c^{2'}(\cdot) \geq c^3(\cdot)$, and there are settings $(\lambda, \gamma, f_H, f_L, q, \pi)$ for which all of the inequalities are strict.

Comparative Statics

Proposition

1. *An increase in λ increases $\overline{\gamma}_H$, i.e., High play following H prominent play occurs for a larger set of parameters.*
2. *There exists M^* such that an increase in λ reduces [increases] γ_H^* if $M < M^*$ [if $M > M^*$], i.e., High play being the unique equilibrium following H prominent play occurs for a larger set of parameters provided that signals favoring L play are not too precise.*
3. *There exists m^* such that an increase in λ reduces [increases] γ_L^* if $m > m^*$ [if $m < m^*$], i.e., Low play being the unique equilibrium following L prominent play occurs for larger set of parameters provided that signals favoring H play are not too precise.*
4. *An increase in λ increases the set of parameters for which an endogenous prominent agent will find it beneficial to break a L social norm.*

Comparative Statics (continued)

Proposition

1. *A lower π increases $\overline{\gamma}_H$, i.e., H play following H prominent play occurs for a larger set of parameter values.*
2. *There exists λ^* such that a lower π increases [reduces] γ_H^* if $\lambda < \lambda^*$ [if $\lambda > \lambda^*$], i.e., H play being the unique equilibrium following High prominent play occurs for a larger set of parameter values provided that agents put sufficient weight on the past.*
3. *There exists λ^{**} such that a lower π reduces [increases] γ_L^* if $\lambda < \lambda^{**}$ [if $\lambda > \lambda^{**}$], i.e., L play being the unique equilibrium following L prominent play occurs for a larger set of parameter values provided that agents put sufficient weight on the past.*
4. *A lower π reduces the set of parameter values for which an endogenous prominent agent will find it beneficial to break a L social norm.*

Comparative Statics (continued)

Proposition

1. *Suppose that the signals become more precise in the sense that the likelihood ratio changes from* $\frac{f_L(s)}{f_H(s)}$ *to* $\frac{\widehat{f}_L(s)}{\widehat{f}_H(s)}$ *such that there exists* $\bar{s} \in (0,1)$ *with* $\frac{\widehat{f}_L(s)}{\widehat{f}_H(s)} > \frac{f_L(s)}{f_H(s)}$ *for all* $s < \bar{s}$ *and* $\frac{\widehat{f}_L(s)}{\widehat{f}_H(s)} < \frac{f_L(s)}{f_H(s)}$ *for all* $s > \bar{s}$*. Then if* $\gamma > \overline{\gamma}_L$ *and* $\gamma < \overline{\gamma}_H$ *both before and after the change in the distribution of signals and the threshold signal* $\tilde{s}$ *is sufficiently close to* m*, then the likelihood that a prominent agent will break the* L *social norm increases in the greatest equilibrium.*
2. *Suppose* $\gamma < \overline{\gamma}_L$ *and* $\lambda \geq 1/2$*. Then if the distribution of signals becomes sufficiently more precise (in the same sense as in the first part of the proposition), then the likelihood that a regular agent will play* H *increases.*

Multiple Agents

- Now suppose n agents within each generation, and random matching; unless there is a prominent agent, in which case all those from previous and next generations match with the prominent agent.
- If no prominent agent, then observe a signal generated by the action of a randomly generated agent from the previous generation.
- Results generalize, except but now we can do comparative statics with respect to n.

Multiple Agents (continued)

Proposition

In the model with n agents within each generation, there exist greatest and least equilibria. In the greatest equilibrium:

1. *following a prominent play of Low, there is a Low social norm and all endogenous agents play Low (i.e., $\overline{\sigma}_{\tau}^{SM}(a = Low, s, T) = Low$ for all s, T and all $\tau > 0$) if and only if $\overline{\gamma}_L^n < \gamma$; and*
2. *following a prominent play of High, there is a High social norm and all endogenous agents play High (i.e., $\overline{\sigma}_{\tau}^{SM}(a = High, s, T) = High$ for all s, T and all $\tau > 0$) if and only if $\gamma \leq \overline{\gamma}_H^n$.*

The threshold $\overline{\gamma}_H^n$ is increasing in n and and the threshold $\overline{\gamma}_L^n$ is decreasing in n, so that both High and Low social norms following, respectively, High and Low prominent play, emerge for a larger set of parameter values.

- Intuition: signals less informative, thus history matters more.

Amnesty

- Modify the model such that agents can switch from L to H in the second period of their lives.
- Suppose that the cost of L is related to the probability of getting caught for "corrupt" behavior. Once an agent chooses L, she can get caught in both periods, so she will not want to switch to H later in life.
- An amnesty, so that past mistakes or deeds are forgiven, when announced in advance and understood by future generations, can induce a switch from a social norm of L to H.

Collective Action

- A model of collective action would be similar, except that interactions would be between many players rather than just two.
- From the New Yorker 2/28/11 quote of an Egyptian student in the context of people cleaning up Tahrir square:
  "We thought people didn't care and just threw their garbage on the street, but now we see that they just thought it was hopeless - why bother when it's so dirty. Why not be corrupt when everything is corrupted. But now things have changed, and it's a different mood overtaking. Even I can't stop smiling myself."
- Why have things changed?
- One interpretation: switches in social norm driven by prominent events (and the endogenous history that this generates).